ED 398 610 CS 509 304

AUTHOR	Fowler, Carol A., Ed.
TITLE	Status Report on Speech Research, July 1994-December 1995.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, MD.
REPORT NO	SR-119/120
PUB DATE	95
CONTRACT	DBS-9112198; HD-02994; MH-51230; MH-53662
NOTE	310p.; For the January-June 1994 Report, see ED 390 082.
PUB TYPE	Reports - Research/Technical (143) -- Collected Works - General (020)
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	Arabic; *Communication Skills; English; Hebrew; Language Processing; *Music; *Reading Difficulties; *Reading Processes; Reading Research; Spanish; *Speech Communication; Syntax; Word Recognition; Writing Research
IDENTIFIERS	Phonological Awareness; Speech Research

ABSTRACT

This publication (one of a series) contains 19 articles which report the status and progress of studies on the nature of speech, instruments for its investigation, and practical applications. Articles are: "Speech Perception Deficits in Poor Readers: Auditory Processing or Phonological Coding?" (Maria Mody and others); "Auditory Temporal Perception Deficits in the Reading-Impaired: A Critical Review of the Evidence" (Michael Studdert-Kennedy and Maria Mody); "Lengthened Formant Transitions Are Irrelevant to the Improvement of Speech and Language Impairments" (Michael Studdert-Kennedy and others); "Phonological Awareness in Illiterates: Observations from Serbo-Croatian" (Katerina Lukatela and others); "Attention Factors Mediating Syntactic Deficiency in Reading-Disabled Children" (Avital Deutsch and Shlomo Bentin); "Cognitive Profiles of Reading-Disabled Children: Comparison of Language Skills in Phonology, Morphology and Syntax" (D. Shankweiler and others); "Syntactic Processing in Agrammatic Aphasia by Speakers of a Slavic Language" (Katarina Lukatela and others); "Tasks and Timing in the Perception of Linguistic Anomaly" (Janet Dean Fodor and others); "Sidestepping Garden Paths: Assessing the Contributions of Syntax, Semantics and Plausibility in Resolving Ambiguities" (Weijia Ni and others); "Why Is Speech so Much Easier Than Reading and Writing?" (Alvin M. Liberman); "How Theories of Speech Affect Research in Reading and Writing" (Alvin M. Liberman); "New Evidence for Phonological Processing during Visual Word Recognition: The Case of Arabic" (Shlomo Bentin and Raphiq Ibrahim); "Lexical and Semantic Influences on Syntactic Processing" (Avital Deutsch and others); "Decomposing Words into Their Constituent Morphemes: Evidence from English and Hebrew" (Laurie Beth Feldman and others); "An Articulatory View of Historical S-aspiration in Spanish" (Joaquin Romero); "Manfred Clynes, Pianist" (Bruno H. Repp); "The Difficulty of Measuring Musical Quality (and Quantity). Commentary on R. W. Weisberg (1994). Genius and Madness? A Quasi-Experimental Test of the Hypothesis That Manic-Depression Increases Creativity" (Bruno H. Repp); "A Review of David Epstein's Shaping Time: Music, the Brain, and Performance" (Bruno H. Repp); and "The Dynamics of Expressive Piano Performance: Schumann's 'Traumerei' Revisited" (Bruno H. Repp). (RS)

Haskins Laboratories Status Report on

Speech Research

SR-119/120
JULY 1994-DECEMBER 1995

Haskins
Laboratories
Status Report on

Speech Research

SR-119/120
JULY 1994-DECEMBER 1995

NEW HAVEN, CONNECTICUT

Distribution Statement

Editor
Carol A. Fowler

Production
Yvonne Manning-Jones
Fawn Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Over the coming year we will be experimenting with different methods for publishing/distributing the *Haskins Laboratories Status Report*. The next issue (or selected portions of the next issue) will appear on our World Wide Web site (`http://www.haskins.yale.edu`) and should be available in the late Fall.

Check our Status Report page on the website:
(`http://www.haskins.yale.edu/Haskins/SR/sr.html`)
for up-do-date information.

This issue will then be followed by a more traditional printed issue of the Status Report. These plans are subject to change and future plans are still up in the air.

Carol Fowler, Editor

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511-6695

Phone: (203) 865-6163 *FAX:* (203) 865-8963
Internet: HASKINS@HASKINS.YALE.EDU

This Report was reproduced on recycled paper

4

Acknowledgment

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development

Grant HD-01994

National Institute of Mental Health

Grant MH-51230

Grant MH-53662

National Science Foundation

Grant DBS-9112198

National Institute on Deafness and Other Communication Disorders

Grant DC 00121	Grant DC 00594
Grant DC 00183	Grant DC 01147
Grant DC 00403	Grant DC 00044
Grant DC 00016	Grant DC 01247
Grant DC 02151	Grant DC 00865

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Eric Bateson*
Fredericka Bell-Berti*
Shlomo Bentin*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Dani Byrd
Claudia Carello*
Franklin S. Cooper*
Stephen Crain*
Alice Faber
Laurie B. Feldman*
Janet Fodor*
Anne Fowler*
Carol A. Fowler*
Ram Frost*
Louis Goldstein*
Carol Gracco
Vincent Gracco
Katherine S. Harris*
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Alvin M. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist
Georgije Lukatela*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Walter Naito†
Weijia Ni
Patrick W. Nye
Kenneth Pugh*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Douglas Whalen

Technical Staff

Michael D'Angelo
Vincent Gulisano
Donald Hailey
Yvonne Manning-Jones
William P. Scully
Fawn Zefang Wang

Administrative Staff

Philip Chagnon
Alice Dadourian
Betty J. DeLise
Lisa Fresa

Students*

Lawrence Brancazio
Melanie Campbell
Douglas Honorof
Pai-Ling Hsiao
Laura Koenig
Subhobrata Mitra
Joaquin Romero
Dorothy Ross
Arlyne Russo
Michelle Sancier
Brenda Stone
Mark Tiede
Qi Wang

*Part-time
†Visiting from University of Tokyo, Japan

6

Contents

7

8

Haskins Laboratories Status Report on

Speech Research

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 1-24

Speech Perception Deficits in Poor Readers: Auditory Processing or Phonological Coding?*

Maria Mody,† Michael Studdert-Kennedy, and Susan Brady‡

Poor readers are inferior to normal-reading peers in aspects of speech perception. Two hypotheses have been proposed to account for their deficits: (i) a speech-specific failure in phonological representation, or (ii) a general deficit in auditory "temporal processing", such that they cannot easily perceive the rapid spectral changes of formant transitions at the onset of stop-vowel syllables. To test these hypotheses, two groups of second-grade children (20 "good readers", 20 "poor readers"), matched for age and intelligence, were selected to differ significantly on a /ba/-/da/ temporal order judgment (TOJ) task, said to be diagnostic of a temporal processing deficit. Three experiments then showed that the groups did not differ in: (i) TOJ when /ba/ and /da/ were paired with more easily discriminated syllables (/ba/-/sa/, /da/-/ʃa/); (ii) discriminating non-speech sine wave analogs of the second and third formants of /ba/ and /da/; (iii) sensitivity to brief transitional cues varying along a synthetic speech continuum. Thus, poor readers' difficulties with /ba/-/da/ reflected perceptual confusion between phonetically similar, though phonologically contrastive, syllables rather than difficulty in perceiving rapid spectral changes. The results are consistent with a speech-specific, not a general auditory deficit.

Reading is a complex skill and there are many reasons why children may fail. The most firmly established correlate of reading disability is a deficiency in skills related to phonological processing. Phonological processing entails the segmental analysis of words for ordinary speaking and listening, as well as the metaphonological skills required for explicitly analyzing the sound structure of speech into the phonemic components represented by the alphabet. Many studies have shown poor readers to be significantly inferior to their normal reading peers in "...perceptual discrimination of phonemes, phonological awareness tasks involving the manipulation of phones within words, speed and accuracy in lexical access for picture names, verbal short-term memory, syntactic awareness and semantic processing on tasks of listening comprehension" (Olson, 1992, p.896). Many, if not all, of these weaknesses may arise, directly or indirectly, from a deficit in speech perception. Several independent lines of research point to speech perception as a source of subtle, but ramifying deficit in reading-impaired children and adults.

The nature and origin of the perceptual deficit have been a matter of debate for over fifteen years. One account sees it as purely linguistic, specific to speech and closely related to the deficit in verbal working memory, also often observed in poor readers (e.g., Bradley & Bryant, 1991; Brady, Shankweiler, & Mann, 1983; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). Accordingly, poor readers are said to have normal auditory capacities, but, for unknown reasons,

The experiments were drawn from the first author's doctoral thesis submitted in partial fulfillment of the requirements for the degree to the City University of New York. We thank the principals, teachers and school children of the Stratford (Connecticut) School system who adjusted their schedules and cheerfully gave their time to the research. We also thank Loraine Obler and Richard Schwartz for guidance and criticism; Susan Nittrouer for providing the stimuli and technical assistance for Experiment 3; Len Katz for statistical advice; Carol Fowler, Alvin Liberman, Richard Olson, Bruno Repp, Donald Shankweiler and four anonymous reviewers for useful comments on earlier versions of the paper. The research was supported by a grant to Haskins Laboratories from the National Institutes of Health and Human Development (HD-01994).

to be less efficient in transforming linguistic input, whether spoken or written, into the phonological code necessary for working and long-term verbal memory. The assumption here is that both reading and listening require lexical items to be held in working memory long enough to extract syntactic structure and meaning from the strings of which they are part. Deficient phonological access or storage may then show up as an impairment in reading and listening, perhaps even masquerading as a syntactic deficit (Shankweiler & Crain, 1986).

An alternative account also acknowledges the difficulty as phonological, but sees it as stemming from a general auditory deficit in "temporal processing". Tallal (1980), the original proponent of this account, asserts that reading-disabled children cannot easily process brief and/or rapidly changing acoustic events, whether speech or non-speech. They therefore have difficulty in judging the temporal order not only of brief, rapidly presented non-speech tones, but also of stop-consonant-vowel syllables contrasting in their initial formant transitions. The difficulty subtly interferes with overall speech perception and so with normal language development, including learning to read.

THE SPEECH-SPECIFIC HYPOTHESIS

At least three independent bodies of evidence are consistent with a speech-specific interpretation of the perceptual deficit: studies of categorical perception, of speech perception under demanding conditions, and of verbal memory span.

Categorical Perception

Reading-impaired children have characteristically deviant patterns of identification and discrimination on tests of categorical perception with synthetic speech sounds. For example, Godfrey, Syrdal-Lasky, Millay, and Knox (1981), comparing performances on two synthetic continua, /ba/-/da/ and /da/-/ga/, found that dyslexic subjects were significantly less consistent than normals in identification, even at the extremes of the continua. Other studies have reported similar results for /ba/ - /da/ (Reed, 1989; Steffens, Eilers, Gross-Glen, & Jallad, 1992; Werker & Tees, 1987), for /pʌ/ - /tʌ/ (De Weirdt, 1988), and for /sa/ - /sta/ (Steffens et al., 1992). Inconsistent identification can also give rise to deviant patterns of discrimination along synthetic continua. For example, poor readers may be less accurate than good readers on between-category, but not on within-category discrimination, suggesting that poor readers cannot easily exploit the phonological contrast which normally enhances discrimination across a phoneme boundary (De Weirdt, 1988; Godfrey et al., 1981; Pallay, 1986; Werker & Tees, 1987). In short, the poor readers' difficulties in all these studies seem to have been primarily in identifying phonetically similar, though phonologically contrastive, synthetic syllables. Such results suggest that speech categories may be, for unknown reasons, broader and less sharply separated in reading disabled than in normal children.

Speech Perception Under Demanding Conditions

A variety of tests has found poor readers to be less successful than good readers at recognizing spoken words under demanding conditions. An effective way to increase the difficulty of repetition tasks and to reveal reading-group differences is to reduce the familiarity of stimuli by using pseudowords (Apthorp, 1995; Brady, Poggie, & Rapala, 1989; Gathercole & Baddeley, 1993; Hansen & Bowey, 1994; Snowling, Goulandris, Bowlby, & Howell, 1986; Taylor, Lean, & Schwartz, 1989). A subject then necessarily relies on the phonological representation of novel input to formulate and execute a correct response. Recent results have confirmed that poor readers are less accurate at pseudoword repetition than both chronological-age matched good readers and reading-age matched controls (Stone & Brady, 1995). Reading group differences have also been reported when the perceptual difficulty of tasks has been increased by presenting time-compressed speech (Freeman & Beasley, 1978; Pallay, 1986; Watson & Rastatter, 1985), synthetic speech or speech produced by infants (Lieberman, Meskill, Chatillon, & Schupack, 1985), and by embedding words in noise (Brady et al., 1983; Snowling et al., 1986). For example, poor readers were significantly worse than good readers at repeating naturally spoken monosyllabic words presented in noise, but no worse when the same stimuli were presented without noise (Brady et al., 1983). By contrast, with non-speech sounds (e.g., environmental sounds such as clapping, knocking on a door, etc.), both groups were worse on the noisy condition, but to the same degree. These results also are consistent with the notion that phonological categories are broader and less well separated among disabled than among normal readers: Poorly defined categories would presumably be more vulnerable to signal degradation by noise than well defined categories.

In a related study of normal adults, Rabbitt (1968) found poorer recall of digits correctly perceived in noise than of digits correctly perceived in quiet. In accord with a limited capacity model of working memory, Rabbit argued that adding noise to the signal necessitated increased use of resources, so that fewer resources were available for storing items in memory. Brady et al. (1983) link their perceptual results to Rabbitt's short-term memory findings, and hypothesize that poor readers' inferior performances on both speech-in-noise and verbal working memory stem from relatively coarse-grained phonological storage.

Verbal Memory Span

Poor readers often have shorter verbal memory spans than do good readers of comparable age, and their memory deficits are specific to speech: When stimuli are neither words nor easy to represent linguistically by verbal labels, recall has not been found to be related to reading ability. (For review and discussion, see Brady, 1991; Perfetti, 1985; Wagner & Torgesen, 1987.) This lack of correlation has been observed in memory tasks with nonsense figures and unfamiliar writing systems as well as on a standard test of visual pattern recall, the Corsi Block Design Test (Gould & Glencross, 1982; Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982; Rapala & Brady, 1990; Vellutino, Pruzek, Steger, & Meshoulam, 1973). The difference is not, however, a matter of input modality (visual vs. auditory): On verbal tests both good and poor readers display essentially the same patterns of errors whether the list to be recalled is heard or read (Shankweiler et al., 1979). Therefore, rather than attribute the poor readers' memory deficit to the coincidence of low level dysfunctions in both auditory and visual systems, as does Tallal (e.g., Tallal, 1990; Tallal, Sainburg, & Jernigan, 1991), the speech-specific hypothesis attributes it to a deficit in the phonological representation common to both modes of input.

THE GENERAL AUDITORY HYPOTHESIS CRITICALLY EXAMINED

A more general account attributes the phonological deficits of reading disabled children and certain other clinical populations to an impaired rate of auditory processing: Defective children are simply slower than normals in apprehending the auditory structure of a signal. If children are slower, their deficit can show up in at least two ways: (i) They can be poor at perceiving signals that follow one another rapidly (i.e., that have short interstimulus intervals (ISIs); (ii) they can be poor at perceiving signals that are very brief. If the two properties, brief signal and short ISI, are combined, as in consonant formant transitions followed immediately by a vowel, children with this deficit will be doubly disadvantaged. Thus, rapid spectral changes, such as formant transitions at the onset of stop consonant-vowel syllables, pose a special problem for these children. Tallal draws on her findings with developmental aphasics (Tallal & Piercy, 1973, 1974, 1975) and aphasic adults (Tallal & Newcombe, 1978) to support this position.

Defining "Temporal Perception"

Before considering the hypothesis in detail, we must clarify terminology. First, we should distinguish between capacities often confounded in the literature: namely, perceiving the temporal properties of events (duration, sequence, relative timing, rhythm), on the one hand, and rapidly identifying or discriminating between very brief events, on the other. Only a deficit in the former can properly be considered a deficit in "temporal analysis", "temporal processing", or "temporal perception". Difficulties in perceiving very brief events and/or events with very brief intervals between them indicate a deficit not in temporal perception, but in the perception of rapidly presented information. We should not confuse rate of perception with perception of rate.

Perception is "temporal" if the defining property of the perceived event is temporal; it does not become "temporal" by virtue of being effected rapidly.

The second important distinction in terminology is between identification and discrimination. To identify a stimulus is to assign it a specific response, such as pressing a certain button or saying its name. To discriminate between stimuli is merely to indicate that they are different in some respect. Identification therefore entails discrimination, but not vice versa. For temporal order judgment (TOJ), discrimination alone is not enough: Identification is required. Errors are therefore ambiguous, unless we have independent evidence of correct identification. The difficulty is obvious from the standard format of TOJ tests in which pairs of stimuli (1,2) are presented for ordering in the combinations: 1-1, 2-2, 1-2, 2-1. All errors where the same stimuli are judged as different, or different stimuli are judged as the same, necessarily involve errors of identification. Only errors of reversal (e.g., 1-2 for 2-1) may be pure errors of temporal judgment. The reader should bear this ambiguity in mind throughout our discussion and reports on TOJ tests below. Arguably, all the supposed difficulties in "auditory temporal perception", or actual difficulties in perceiving rapidly presented information, so far reported for both reading- impaired and specifically language-impaired children, can be traced to difficulties in stimulus identification. Tallal has repeatedly acknowledged this fact (e.g. Tallal, 1980, p.193; Tallal & Piercy, 1973, p.396; Tallal et al., 1991, p.365), and Reed (1989, pp. 287-289) makes the same point.

Developmentally Aphasic Children and Aphasic Adults

In the first of a series of studies, Tallal and Piercy (1973) found developmentally aphasic children (n=12) to be significantly impaired, in comparison to an equal number of age-matched, normal controls, on tasks involving rapid auditory perceptual processing. The dysphasic, or language-impaired, children had difficulty with TOJ for pairs of complex tones differing in fundamental frequency (100 vs. 305 Hz). The impaired children made significantly more TOJ errors than normal controls when the tones were short (75 msec) rather than long (250 msec), and/or when the interval between the tones (ISI) was short (150 msec) rather than long (300 msec). The authors viewed these findings as evidence that developmental aphasics process auditory stimuli more slowly than normals.

On the hypothesis that, if children had trouble perceiving rapid auditory events, it would be evident in their perception of speech, Tallal and Piercy (1974) extended their research to verbal stimuli characterized by brief and rapid spectral changes. Their stimuli were: the synthetic syllables, /ba/ and /da/, for which the brief (approx. 40 msec) second and third formant transitions at onset are critical cues to consonant place of articulation; two synthetic three-formant steady-state vowels, /ɛ/ and /ae/; and the two long tones of the previous study. All stimuli were 250 msec in duration. Repeating their earlier procedure, the investigators found that only five of the twelve dysphasic subjects reached identification training criterion with /ba/ and /da/, and only two of these five reached criterion on same/different discrimination of these stimuli at a long ISI; only these two therefore were tested on TOJ at short ISIs. By contrast, all twelve controls readily reached training criterion on /ba/-/da/ and performed perfectly on discrimination and TOJ at all ISIs. On the same tasks with the 250 msec vowels and tones, the two groups did not differ significantly. We emphasize that the dysphasic children's difficulties were not shown to be with TOJ, on which only two of them were tested (both doing well), but with the identification and discrimination of /ba/ and /da/.

In a further study, Tallal and Piercy (1975) showed that these same children had no difficulty with /ba/-/da/ when their F1, F2 and F3 transitions were extended from 43 msec to 95 msec. In this study all twelve aphasic children and their matched controls reached criterion on identification, and on TOJ and discrimination at a long ISI. Even when ISI was reduced, dysphasics continued to perform as well as normals. Thus, improved TOJ at short ISIs followed improved identification, suggesting again that the difficulty was with the latter rather than the former. That dysphasics did worse when the syllables incorporated brief formant transitions was viewed as consistent with the earlier findings for brief non-speech tones. Taken together the two findings suggested to the authors that (i) "...it is the brevity not the transitional character of this component of synthesized stop consonants, which results in the impaired perception of our dysphasic children" (Tallal & Piercy, 1975, p. 73), and (ii) the impairment was due to a general auditory deficit, not specific to speech. Notice, however, that the authors did not establish the auditory basis of the dysphasic children's difficulties with /ba/-/da/ by demonstrating equivalent difficulties for appropriate non-speech control patterns with brief and rapidly changing onset frequencies. The claim that the deficit was general and auditory rather than phonetic and specific to speech was therefore unsubstantiated, and has remained so.

No one, so far as we know, has replicated the results of Tallal and Piercy (1975) with specifically reading-impaired children, but Tallal and Newcombe (1978) did observe improved identification of /ba/-/da/ with lengthened transitions (from 40 msec to 80 msec) in four out of ten adult aphasics. Attempts to replicate these results have not been successful, however. Two studies, with a combined total of 25 adult aphasic subjects, failed to find that lengthened transitions improved either identification or discrimination of stop consonant place of articulation (Blumstein, Tartter, Nigro, & Statlender, 1984; Riedel & Studdert-Kennedy, 1985). We should note, moreover, that even if the advantage due to lengthened transitions were better attested than it is, its interpretation would be uncertain. The standard finding in the acoustic phonetic literature is that the perceived manner of syllable-initial stop consonants shifts from stop to glide when formant transitions are lengthened (Borden, Harris, & Raphael, 1994, Chapter 6). Improved identification and discrimination by aphasic subjects due to lengthened transitions may then reflect either facilitated auditory processing, as Tallal and Piercy (1975) assert, or increased phonetic distance across a phonological contrast.[1]

"Auditory Temporal Perception" and Phonological Decoding

We turn next to the paper in which Tallal (1980) extended her hypothesis concerning the perceptual deficits of dysphasic children to children with reading impairments (cf., Tallal (1984) and Tallal, Miller, and Fitch (1993), for example, who also propose extending the hypothesis to reading-impaired children). She tested 20 reading-impaired children on a battery of tests, including verbal and performance IQ, five reading tests, and two non-speech auditory perceptual tests. The last were the short ISI discrimination and TOJ tests with 75 msec complex tones differing in fundamental frequency (100 vs. 305 Hz), described above. Tallal compared the performance of the 20 reading-impaired children on the auditory tests with that of 12 normal controls from a previous study (Tallal, 1976). Eleven of the reading impaired children performed normally; only nine fell below the worst control.

Despite this variability, Tallal found significant rank order correlations between nonverbal auditory perception and all five reading tests. The highest correlation was between tone TOJ and a Nonsense Word Reading Test, evaluating decoding skill (Spearman's $R=.81$, $p < .001$). Tallal argued that the children's difficulties in identifying brief, rapidly presented tones related to their difficulties in reading by appealing to her previous findings with dysphasic children who had difficulty both with brief tones and (by hypothesis) with brief formant transitions at the onset of /ba/ and /da/ (Tallal & Piercy, 1973, 1974, 1975). If we assume that the reading-impaired children of Tallal (1980) suffered from this same syndrome of deficits, we have "...a possible basic perceptual mechanism that may underlie some difficulties in analyzing the phonetic code efficiently, and ultimately in learning to read" (p.196). The supposedly defective mechanism is that engaged for "...the analysis of rapidly changing acoustic information...in formant transitions...[D]ifficulty in analyzing rapid information may lead to difficulty in analyzing speech at the phonemic level...[and so to] some of the difficulties...poor readers...have segmenting and recoding phonemically" (p.196).

Yet several questions arise. First, the reading-impaired children's performance on tone TOJ was not significantly worse than their performance on tone discrimination. From this finding Tallal (1980) inferred (as she had earlier inferred for her developmental aphasics) that their "...difficulty with temporal pattern perception may stem from a more primary [*sic*] perceptual deficit that affects the rate at which they can process perceptual information" (p.193). In other words, the difficulty may not have been with "auditory temporal perception" itself (as the title of the paper implies), but with identifying the tones correctly, when they were presented in rapid succession.

Second, the reading-impaired children's difficulties in identifying brief tones at short ISIs do not warrant the inference that they would have had similar difficulties with /ba/ and /da/. To be sure, Reed (1989), the only researcher to have tested specifically reading-impaired children on both speech and non-speech tasks in the Tallal paradigm, did find that such children performed significantly worse than normal controls on TOJ for both brief tones and stop consonants. And at least two other studies have reported poor performance by disabled readers on tone TOJ (e.g., Bedi, 1994) and on stop consonant discrimination at short ISIs (Hurford & Sanders, 1990). But we have no reason to suppose that the two weaknesses reflect the same underlying discriminative deficit, since tones and syllables contrast on entirely different acoustic dimensions. The tones are discrete, steady-state events, contrasting in fundamental frequency; the transitions of synthetic /ba/ and /da/ are continuous spectral sweeps, contrasting in spectral locus and direction. (For a critique of the equation of tone sequences with formant transitions, see Studdert-Kennedy & Mody, 1995.)

Finally, the proposal that difficulty in identifying /ba/ and /da/ should be attributed to difficulty in "the analysis of rapidly changing acoustic information" is precisely the hypothesis that Tallal and Piercy (1975) tested with dysphasic children and, as we have seen, rejected in their conclusion that "...it is the brevity, not the transitional character" (p.73) of formant transitions which causes difficulty. Yet no subsequent experiment has established, by means of an appropriate non-speech control, that rapid acoustic changes are indeed difficult to perceive for either specifically language-impaired or specifically reading-impaired children.

In fact, the only relevant study known to us does not even support the claim that reading-impaired children are likely to have difficulty with the brevity, let alone the spectral changes, of formant transitions. Pallay (1986) manipulated the duration of F1 transitions along two synthetic continua: one ranging from /ba/ (30 msec transitions) to /wa/ (100 msec transitions) and the other, a non-speech control, ranging across the isolated F1 transitions of the /ba/-/wa/ continuum. Second-grade reading-impaired children and matched normal controls identified the stimuli as /ba/ or /wa/ for the speech series, and as "long" or "short" for the non-speech series. There was no significant difference between normal and poor readers in the positions of their category boundaries on either series. In other words, poor readers did not need a longer F1 transition than normal controls to identify stimuli either as /wa/ or as "long". We should emphasize that the manner contrast, /ba/-/wa/, unlike the place contrast, /ba/-/da/, does call for a temporal judgement. For while place information is largely carried by a pattern of spectral change in F2 and F3, stop-glide manner information is largely carried by the duration of the F1 transition into the vowel (Liberman, Delattre, Gerstman, & Cooper, 1956). Here, then, in the only study ever to call on reading-impaired children to identify both speech sounds contrasting in their temporal properties and an appropriate set of nonspeech controls, the children displayed completely normal capacities, both auditory and phonetic.

In sum, no direct evidence for a temporal processing deficit in poor readers has been reported. However, a number of studies has found that poor readers may have difficulties in discriminating or identifying /b/ - /d/ (e.g., Godfrey et al., 1981; Hurford & Sanders, 1990; Reed, 1989; Steffens et al., 1992; Werker & Tees, 1987). An auditory account of that effect, attributing it to a deficit in some aspect of so-called temporal processing, has not yet been subjected to direct test. Such a test was a main purpose of the present study.

THE PRESENT STUDY

Experiment 1a served to select a group of poor readers who had difficulty with synthetic /ba/-/da/ TOJ in a test modeled after Tallal's, and a group of good readers who had no such difficulty; both groups were also tested on /ba/-/da/ discrimination. Selection was necessary to ensure that the poor readers did indeed have difficulty with /ba/-/da/ TOJ, because both Tallal (1980) and Reed (1989) found poor readers whose TOJ performance fell within the normal range. Experiment 1b determined whether the poor readers' apparent difficulties with TOJ and discrimination remained when the syllables to be judged were highly contrastive, and so readily identifiable. If the difficulties vanished for readily identifiable syllable contrasts, we could conclude that errors on /ba/-/da/ TOJ were due to difficulties in identifying the syllables rather than in judging their temporal order.

Experiment 2 was a non-speech discrimination control for Experiment 1a, using sine wave patterns corresponding to the center frequencies of F2 and F3 in synthetic /ba/ and /da/. If poor readers displayed the same effects of reduced ISI on discrimination of non-speech control patterns as they had on discrimination of the full syllables, we could conclude that their difficulties might indeed arise in the auditory processing of rapid spectral changes; on the other hand, if effects of ISI, and group differences, disappeared, we could conclude that poor readers' difficulties with /ba/-/da/ were not auditory, but phonetic, that is, specific to the perception of speech.

Yet whatever the outcome of Experiment 2, it would be of interest to know whether poor readers' difficulties extended to other phonological contrasts carried by brief formant transitions. Experiment 3 therefore compared good and poor readers on sensitivity to the frequency extent of rapid F1 transitions varying along a synthetic continuum from /seɪ/ to /steɪ/. If poor readers were less sensitive to transitional information than good readers, we would expect them to need a more extensive transition in order to detect the presence of a stop between fricative and vowel. Alternatively, we might expect, from the results of a previous study with this continuum (Nittrouer, 1992), in which 3-, 4- and 5-year olds proved more rather than less sensitive to formant transition variations than 7-year olds and adults, that this final test would afford a measure of developmental delay in the speech perception of 8-year old impaired readers. If neither of these outcomes occurred, we could conclude that poor readers did not differ from good readers in their sensitivity to the extent of formant transitions.

GENERAL METHOD

Subjects

The subjects consisted of forty second-grade children (mean grade level: 2.5 yrs), between the ages of 7;0 and 9;3 years, from a public school district in south central Connecticut. These were drawn from a pool of 220 children screened for the study. The large pool was necessitated partly by school requirements that all members of a class participate, partly by the study's stringent criteria for performance and for statistical matching of the groups.

Overview of Selection Criteria

Because two of the central goals were to study whether problems on TOJ tasks stem from difficulties with identification and whether less-skilled readers have difficulty on both speech and nonspeech tasks, it was essential to select poor readers who could meet an identification training criterion for /b/ and /d/, yet who made errors on the speech TOJ task. These demands ruled out a number of poor readers (n=10), who like over half of Tallal's (1980) subjects on tone TOJ, failed to make errors on /ba/-/da/ TOJ. For reasons detailed below, the remaining subjects reading below grade level were matched as a group (n=20) in age and in verbal and non-verbal IQ scores with a group of better readers (n=20). The combined set of criteria, and some additional screening requirements, disallowed many subjects, but resulted in a well-matched set of participants for whom differences on the experimental measures were interpretable.[2] Table 1 lists means and standard deviations on the selection criteria for the two groups.

Reading

Reading performance was assessed with the Word Attack and Word Identification subtests of the Woodcock-Johnson Reading Mastery Test-Revised (Woodcock, 1987). These measures were selected because of converging evidence that the major reading deficits for poor readers are difficulties in decoding and in word recognition (e.g., Olson, Forsberg, Wise, & Rack, 1994; Rack, Snowling, & Olson, 1992; Stanovich, 1991). To ensure appropriate classification, an individual was included as a skilled or unskilled reader only if his/her scores on the two Woodcock subtests both categorized the child as a skilled reader or as a less-skilled reader.

Table 1. *Mean (M) and standard deviation (SD) for good and poor readers on each of the selection criteria.*

	GOOD READERS		POOR READERS		
	M	SD	M	SD	*t*
AGE (months)	95.9	5.3	97.0	6.6	0.61
PPVT-R (standard scores)	98.5	7.7	97.4	7.7	0.43
WISC-R (standard scores)	105.9	9.2	106.3	14.2	0.33
WORD IDENTIFICATION+	4.4	0.8	2.1	0.3	11.97*
WORD ATTACK+	8.9	4.6	1.9	0.4	6.80*
/ba/-/da/					
TOJ (errors++)	0.0	0.0	7.4	4.4	7.59*
Discrimination (errors++)	0.0	0.0	5.8	3.9	6.65*

*$p < .001$
+Grade equivalent scores
++Errors out of 36 trials per subject, across three short ISIs combined.

Because reading ability is normally distributed, and because dyslexia is not a discrete clinical category (e.g., Shaywitz, Escobar, Shaywitz, & Fletcher, 1995; Stanovich & Siegel, 1994), our goal was to test groups of twenty subjects each, well-separated in reading level, but conforming to the other selection criteria described below. The less-skilled readers selected were all at least five months behind mid-year grade level in their reading; the good readers were at least five months above grade level. The two groups did not overlap on their reading scores as measured by the Woodcock subtests. We note, further, that these measures tend to over-estimate reading skill, as may be judged from the fact that 17 of the poor readers had been identified by their school's reading coordinator as having reading difficulties, and were receiving supplemental reading instruction.

IQ

Studies have shown that poor readers with low IQ perform worse than poor readers with normal IQ on the Tallal non-speech tone task (Jorm, Share, Maclean, & Matthews, 1986). Therefore, to avoid potential problems of IQ confounds, we selected children who differed in reading ability but whose IQs (as estimated by the Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981)) and by the Block Design subtest of the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) were in the normal range and were fairly comparable. The inclusion range was specified as scores from 80-120, although, because of difficulty in finding poor readers who met all the specified criteria, two children with Block Design scores of 135 were included. Thus the groups were closely matched on receptive language and on non-verbal IQ.

Temporal Order Judgment on /ba/-/da/

As noted above, the groups were further defined on the basis of their performance on Tallal's /ba/-/da/ TOJ task. Only poor readers who made a minimum of three errors out of 36 trials (8%) on the three short ISIs combined, and good readers who scored 100% correct on the same were included. We would have preferred a higher minimum error rate for the poor readers, but were compelled to settle for an arbitrary 8%, an average of one error in each block of 12 trials at each ISI, by the shortage of poor readers making substantial numbers of errors.[3] Nonetheless, the groups were well separated on /ba/-/da/ TOJ at short ISIs: the good readers' mean number of errors was zero, the poor readers' mean number of errors differed significantly from zero. As indicated in Table 1, the mean number of errors on /ba/-/da/ discrimination at short ISIs was also zero for the good readers and also differed significantly from zero for the poor readers.

The selection of subjects on the basis of performance on /ba/-/da/ TOJ was an essential part of the experimental design, both to ensure that the poor readers displayed behavior consistent with the deficit that Tallal (1980) proposes may underlie difficulties in segmenting and recoding speech phonemically, and to ensure that the good readers differed significantly from the poor readers in this respect. We emphasize that this aspect of the experimental design precluded either good readers or poor readers from being fully representative samples of the good and poor reading populations. Rather, they were samples from experimentally defined populations of good readers who make no errors on Tallal's task and of poor readers who make at least 8% errors on that task. Since the main goal of the study was to test a hypothesis concerning the cause of errors on the task, these were appropriate target populations.

Age

The skilled and less-skilled readers were matched on age to control for potential developmental factors on performance.

Additional criteria

All subjects came from middle-income families, had no history of emotional, neurological, or attentional disorders, and were monolingual, native speakers of English (i.e., English was the predominant language spoken in the home). They had normal hearing (25dB SPL at 500 Hz, and 20dB SPL at 1000 Hz, 2000 Hz, 4000 Hz, 6000 Hz and 8000 Hz), and no dialect-related consonant substitutions or omissions relevant to the stimuli being used in the study. Except for two good readers, who were ambidextrous, all subjects were right-handed. Handedness was determined by an abbreviated five-question checklist drawn from Annett's (1970) original set of criteria for handedness. The checklist consisted of the following questions: Which hand do you use for the following activities: writing, brushing your teeth, throwing a ball, swinging a bat, hammering a nail? Subjects were considered right handed, if they used their right hand for all five tasks. Written consent was obtained from the parents of all participants.

Procedures

All testing was done individually in a quiet room in the children's schools. After the screening tests, qualified subjects met with the experimenter three more times; each session lasted about 45 minutes. Four speech perception tasks (one not reported here) and one non-speech perception task were administered over the course of the study. The final screening test, using Tallal's TOJ task, was carried out as Part a of Experiment 1. Stimuli were presented through TDH-39 headphones at a comfortable listening level. Subjects were rewarded with colorful stickers and/or pencils.

EXPERIMENT 1

Experiment 1a was a portion of the subject selection process, intended to establish a significant difference between good and poor readers on TOJ at short ISIs for /ba/-/da/ in the standard Tallal task. Experiment 1b was designed to determine whether the apparent TOJ deficit of the poor readers in 1a might arise from difficulties in identifying /ba/ and /da/ at rapid rates of presentation due to their close acoustic-phonetic similarity rather than from a deficit in judgments of temporal order itself. If this were so, we would expect their difficulties to disappear when the syllables were presented in more easily discriminable pairs, such as /ba/-/sa/ and /da/-/ʃa/. Notice that while /ba/-/da/ differ on a single phonetic feature (place), /ba/-/sa/ and /da/-/ʃa/ differ on three features (place, manner, voicing). Experiment 1b therefore tested discrimination and TOJ for these stop-fricative combinations.

Method

Stimuli

The syllables /ba/, /da/, /sa/ and /ʃa/ were generated on the Haskins serial synthesizer on a VAX 11/780. The stimuli /ba/ and /da/, each 250 msec in duration, were composed of three formants with no release burst. Values of the stimulus parameters were identical to those used by Reed (1989) in her successful replication of Tallal's experiments (Tallal, 1980; Tallal & Stark, 1981). The two syllables had identical steady-state portions with F1 at 750 Hz, F2 at 1200 Hz, and F3 at 2350 Hz. The bandwidths of the three formants were 90 Hz, 90 Hz, and 130 Hz respectively. Each syllable started with F0 at 121 Hz rising to 125 Hz in 40 msec and falling to 100 Hz at syllable end. Whereas F1 began at 200 Hz for both syllables, reaching its steady-state value in 25 msec, F2 and F3 onsets differed for the two syllables. For /ba/, the second and third formants began at 825 Hz and 2000 Hz respectively, reaching their steady states in 35 msec. For /da/, F2 began at 1500 Hz and F3 at 2630 Hz, both reaching their steady-states in 35 msec.

The /sa/ and /ʃa/ stimuli were each 400 msec long, the frication noise lasting 150 msec and the vowel formants 250 msec (Tallal & Stark, 1981); /sa/ had a fricative noise high-pass cut-off at 4100 Hz, /ʃa/ at 2800 Hz. Relative amplitude of the frication noise rose from 20 dB to 35 dB over the first 50 msec, and then fell to 30 dB over the last 50 msec of the frication duration. The vocalic portions were identical for both sounds, the formant frequency values being the same as those of the steady-state vocalic portions of /ba/ and /da/.

Procedure

The sequence and structure of the training and test procedures exactly followed those of Tallal (1980), with minor adaptations incorporated by Reed (1989) in her successful replication of Tallal's work. Experiment 1a consisted of a TOJ task and a discrimination task with a stimulus pair also used by Tallal, /ba/-/da/ (Tallal & Piercy, 1974). In Experiment 1b subjects repeated the TOJ and discrimination tasks, this time with a different stimulus pair in which for half of each subject group the pair was /ba/-/sa/ and for the other half of each group the pair was /da/-/ʃa/. The order of presentation of TOJ and discrimination was counterbalanced across subjects within a group. Following Tallal's protocol, a written log of subjects' responses was maintained throughout the session.

Experiment 1a

Identification training. Subjects were told that they would hear two syllables, /ba/ and /da/. Their task was to identify these syllables by pointing to a red dot on the board before them if they heard /ba/, to a green dot if they heard /da/. Each child was presented with six repetitions of the syllable /ba/, followed by six repetitions of the syllable /da/ to familiarize him/her with the sounds and the correct responses. Training then continued for up to 48 trials, 24 of each stimulus quasi-randomly presented one at a time with the restriction of a maximum of three of one type in succession and an unlimited interval for responding. For this identification training only (i.e., not for any of the subsequent tasks), the experimenter switched after the first two subjects, to a "point and say" response, because this seemed to engage the child's attention more fully and to increase response consistency. When subjects reached a criterion of 12 correct out of 16 consecutive trials ($p < .001$, binomial test), they proceeded to one or other of the next two tasks.

Temporal order judgment training and test. Here the subject was first trained to respond to two stimuli presented in succession with a 400 msec ISI by pointing to the red and green dots in the correct order of presentation of the two sounds. There were four possible orders: 1-1, 1-2 , 2-1, 2-2 (where 1 and 2 represent /ba/ and /da/ respectively). Four demonstrations by the experimenter were followed by eight training trials with feedback and by 24 further trials without feedback, to a criterion of 12 correct out of 16 consecutive trails. All subjects reached criterion and were then given an additional thirty-six TOJ test trials at reduced ISIs: A series of 12 two-stimulus patterns was presented at each of three shorter ISIs, viz., 100 msec, 50 msec, 10 msec. The series were presented in the same order of decreasing ISI for all subjects.

Discrimination training and test. A same/different task was used. A subject was initially presented with six examples of identical syllable pairs (either /ba/-/ba/ or /da/-/da/), at ISIs of 400 msec and was instructed to point to two blue dots (i.e., two dots of the same color) after each trial. Then came six examples of trials with different syllables, for which the subject was instructed to point to two differently colored dots (a blue dot and a yellow dot). Next, 48 trials consisting of an equal number of identical and different syllable pairs, were presented in a quasi-random order, with the proviso that there be a maximum of three of a kind in succession. Training continued to a criterion of 12 correct out of 16 consecutive trials. All subjects reached criterion and were then tested for discrimination with the same series of 12 trials at each of the three short ISIs as were used in the TOJ task (i.e., 100, 50 and 10 msec.).

Experiment 1b

As noted above, for half the subjects in each group, the TOJ and discrimination tasks were repeated with the stimulus pair /ba-sa/, and for the other half they were repeated with the stimulus pair /da/-/ʃa/. Ideally, each subject should have completed these tasks with both stimulus pairs, but this would have made the procedure unacceptably long. Once again, the order of presentation of the TOJ and discrimination tasks was counterbalanced within the groups of good and poor readers, and the subjects had to respond by pointing.

Results and Discussion

Training

In Experiment 1a (/ba/-/da/) only four of the good readers, but 14 of the poor readers, made at least one error during one or more of the three training segments. The mean numbers of errors and the differences between the group means were very small. The training means (with standard deviations in parentheses) for good readers were 0.2(0.4) on identification, 0(0) on discrimination, and 0.1(0.2) on TOJ; for poor readers they were 1.1(1.7) on identification, 0.6(0.9) on discrimination, and 1.0(1.3) on TOJ. Since the data clearly did not meet the assumptions of normality and homogeneity of variance, we compared the groups non-parametrically across the three training segments combined. The difference between the number of good readers (4) and poor readers (14), making at least one error was significant ($\chi^2(1) = 10.10, p < .01$).

By contrast, in Experiment 1b good and poor readers performed equally well on identification and TOJ training with stimulus pairs /ba/-/sa/ or /da/-/ʃa/, making no errors at all. Performance on discrimination training was almost error-free too, with the exception of a single error by one poor reader in the /da/-/ʃa/ subgroup. Thus, there were no significant differences between the groups on any of the training segments for the stop-fricative contrast.

Perception at Short ISIs

Figure 1 displays the mean number of errors on discrimination and TOJ of /ba/-/da/ at short ISIs by the two groups. Whereas errors increase monotonically with decreases in ISI for the poor readers on both tasks, good readers were unaffected by the change in ISI, and their performance was identical on the two tasks. Although more errors might be expected on TOJ (chance: 25%) than on discrimination (chance: 50%), poor readers made more errors on TOJ than on discrimination only at 100 msec and 10 msec, not at the intermediate ISI of 50 msec. A two-way ANOVA (Task X ISI) for the poor readers yielded a significant main effect of ISI ($F(2,38)=11.26, p < .001$), an effect just short of significance for task ($F(1,19)=4.30, p=.052$), and no significant interaction between the two variables $F(2,38)=1.64, p > .05$). Thus, ISI had the same effect on both tasks.

When the syllable pair was changed to /ba/-/sa/ or /da/-/ʃa/, poor readers performed almost as well as the controls with the exception of one temporal ordering error and two discrimination errors, all at the shortest ISI and on syllable pair /da/-/ʃa/. Good readers continued to make no errors at any ISI. Overall, there was no difference between good and poor readers on discrimination and temporal ordering at short ISIs for the stop-fricative contrast.

In summary, the two groups were selected to differ significantly on overall /ba/-/da/ TOJ; they also differed significantly on overall /ba/-/da/ discrimination (Table 1). Yet there was no significant difference between their performances on the same tasks with stimulus pairs /ba/-/sa/ or /da/-/ʃa/13. These findings demonstrate that the poor readers' difficulties with /ba/-/da/ TOJ do not reflect a general problem with temporal order analysis: Poor readers judge temporal order accurately, even at rapid rates of presentation, if they can identify the items to be ordered. Perhaps, then, their difficulties with /ba/-/da/ are phonological. As noted earlier, these syllables differ on a single phonetic feature. If poor readers have broader, less well separated phonological categories than normal, such pairs may be difficult to discriminate and, for TOJ, to identify under time pressure. Nonetheless, it is still possible that their similarity is acoustic rather than phonetic. Experiment 2 was designed to resolve this issue.

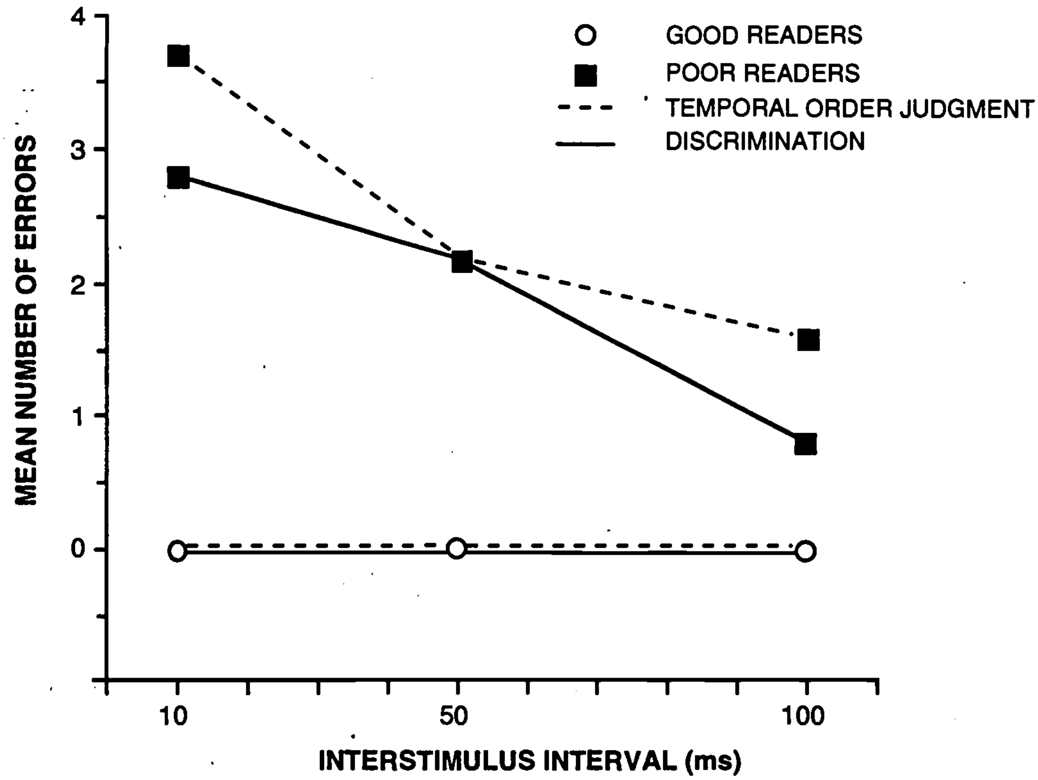


Figure 1. **Mean number of errors by good and poor readers as a function of ISI on /ba/-/da/ discrimination and temporal ordering tasks.**

EXPERIMENT 2

This experiment was a non-speech control condition for Experiment 1a, required to determine whether the poor readers' difficulties with /ba/-/da/ were indeed auditory in origin. The stimuli were frequency modulated sine wave patterns following the center frequencies of the second and third formants of synthetic /ba/ and /da/; as may be recalled from the description of the stimuli for Experiment 1, /ba/ and /da/ differed in F2 and F3, but not in F1. Because sine wave syllables composed of second and third formant analogs, without a first formant analog, are not heard as speech, even by listeners instructed to listen to the sounds as speech (Remez, Rubin, Pisoni, & Carrell, 1981), these stimuli constituted acoustically matched, but perceptually distinct, non-speech controls. The sine wave patterns were presented for identification training, discrimination training at 400 msec ISI, and discrimination at short ISIs in exactly the same sequence and numbers of stimuli as for the syllables in Experiment 1. Both for reasons of time and because (as in Tallal, 1980) there were no significant differences between discrimination and TOJ at short ISIs in Experiment 1a, we omitted TOJ for these non-speech control patterns. If the poor readers' deficit were auditory rather than phonetic, we would expect them to display the same effects of ISI on the non-speech control as on /ba/-/da/, and again to perform significantly worse than good readers. If the deficit were specific to speech, the non-speech control patterns should yield no effect of ISI and no group differences.

Method

Stimuli

Non-speech stimuli were generated on the Haskins sine wave synthesizer through a VAX 11/780. The two stimuli, each 250 msec in duration, were each composed of two sine waves with durations and frequency trajectories identical to those of the center frequencies of F2 and F3 in the synthetic /ba/ and /da/ described above. Perceptually, they did not resemble their speech models. Listeners judged them to be unfamiliar non-speech sounds and assigned them the descriptive labels of "up" and "down", respectively. The selection of these labels was based on a pilot run of the stimuli with four normal children who unanimously preferred "up"/"down" over "high"/"low."

Procedure

The procedures were those of the identification training, discrimination training at 400 msec ISI, and discrimination at short ISIs of Experiment 1a. For identification training the red dot and the green dot were replaced by an upward-pointing and a downward-pointing arrow, respectively, to match the "up" and "down" identification labels.

Results and Discussion

Training: Identification and discrimination

To facilitate comparison between speech (Experiment 1a) and non-speech (present experiment), Table 2 lists, for each group, the mean number of errors to criterion for identification training and discrimination training, under each condition. Both groups found the non-speech stimuli harder, as indicated by the increase in errors. On identification training the mean increase in errors was equal for the two groups (6.2); on discrimination training the mean increase was slightly greater for good readers (2.2) than for poor (1.4). Separate two-way analyses of variance (Group × Condition (Speech/Non-Speech)) for identification training and for discrimination training yielded the same pattern of results: A main effect of Condition for both identification ($F(1,38)=96.04, p < .001$) and discrimination ($F(1,38)=15.68, p < .01$); no effects of Group (identification: $F(1,38)=1.64, p > .05$; discrimination: $F(1,38=.17, p > .05$), and no significant interactions (identification: $F(1,38)=0, p > .05$; discrimination: $F(1,38)=0.7, p > .05$).

The main effects of condition, combined with the absence both of main effects for group and of interactions between group and condition, indicate that, while both groups found non-speech more difficult than speech, training was not significantly harder for one group than for the other. The lack of group differences in learning to identify and discriminate the sine wave patterns would not be expected, if the poor readers' difficulties with /ba/-/da/ were indeed due to a general deficit in perceiving "rapid acoustic changes".

Discrimination at short ISIs

On non-speech discrimination, the poor readers (mean errors across ISIs=1.4, SD=1.5) slightly outperformed the good readers (mean errors across ISIs=2.3, SD=1.9). The two groups performed similarly with no effect of ISI over the two longer values (100 msec and 50 msec), but good readers showed a sharp increase in errors at the shortest ISI. A two-way ANOVA (Group × ISI) found no significant difference between Groups ($F(1,38)=3.5, p > .05$), and no main effect of ISI ($F(2,76)=2.42, p > .05$), but a significant interaction between Group and ISI ($F(2,76)=4.12, p < .02$), reflecting the effect of the shortest ISI on good readers, and the lack of such an effect on poor readers. However, a *post hoc* t-test of the difference between the shortest ISI and the mean of the two longer ISIs for the good readers fell short of significance by the conservative criterion of Scheffé ($t(19)=2.75, p > .05$). Thus, ISI had no significant effect on non-speech discrimination for either group. (For evidence that the lack of group differences in errors on non-speech discrimination cannot be attributed to regression to the mean from the extreme error scores on speech discrimination reported in Experiment 1a, see the Appendix.)

Table 2. Mean (M) and standard deviation (SD) of errors to criterion on training tasks by good and poor readers, under speech and non-speech conditions.

TASK	GOOD READERS				POOR READERS			
	Speech		Non-speech		Speech		Non-speech	
	M	SD	M	SD	M	SD	M	SD
Identification	0.2	0.4	6.4	3.5	1.1	1.7	7.3	4.5
Discrimination at 400 msec ISI	0.0	0.0	2.2	2.9	0.6	0.9	2.0	2.4

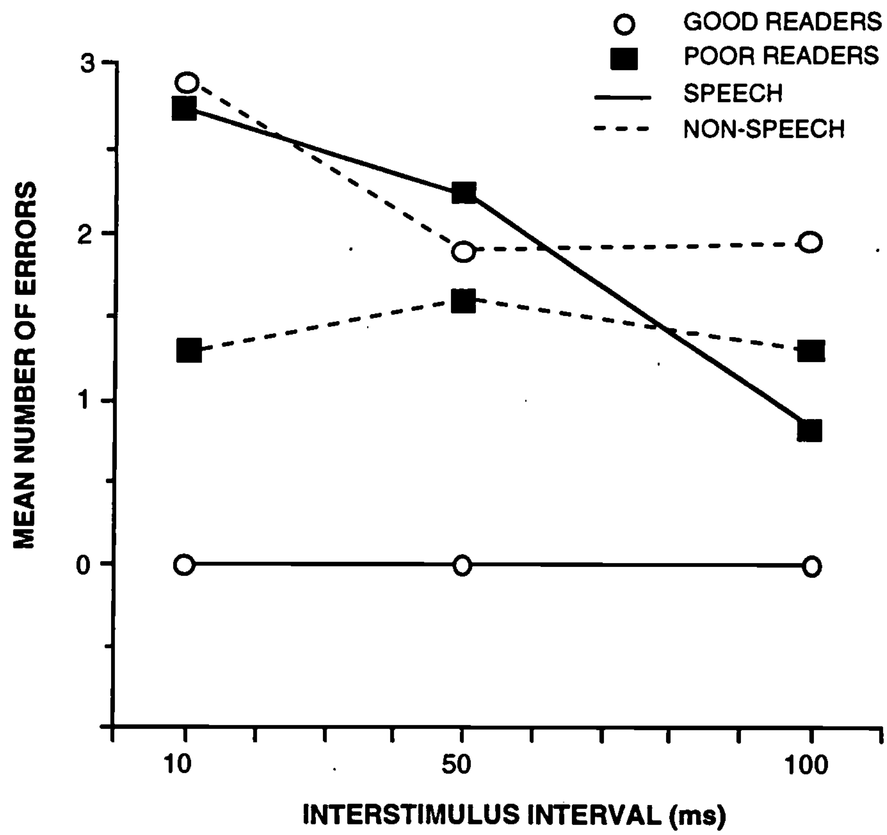


Figure 2. **Mean number of errors by good and poor readers as a function of ISI on speech (/ba/ - /da/) and non-speech discrimination.**

To illustrate the differences between speech (/ba/-/da/) (Experiment 1a) and non-speech (Experiment 2), Figure 2 plots discrimination errors of the two groups as a function of ISI for the two conditions. The most striking features of the figure are: (i) a strong effect of condition for the good readers, but a relatively weak effect for the poor readers; (ii) no effect of ISI on either condition for the good readers (see Scheffé test above); (iii) a strong effect of ISI on speech, but not on non-speech, for the poor readers. A three-way ANOVA (Group × ISI × Stimulus Condition) on the error data, with ISI and Stimulus Condition as within-subject variables, confirms this description: Significant main effects of Condition ($F(1,38)=12.33$, $p < .001$) and ISI ($F(2,76)=10.91$, $p < .0001$), but not of Group ($F(1,38)=2.56$, $p > .10$); significant two-way interactions between Group and Condition ($F(1,38)=34.52$, $p < .0001$) and between Group and ISI ($F(2,76)=3.15$, $p < .05$), and a significant three-way interaction among Group, ISI and Condition ($F(2,76)=8.45$, $p < .001$). Thus, the effects of both condition and ISI are different for the two groups. For the poor readers, the seemingly weak effect of condition and the strong effect of ISI on speech, but not on non-speech, are confirmed by a two-way ANOVA across Experiments 1a and 2 (Condition X ISI) for these readers alone: No effect of Condition ($F(1,19)= 3.43$, $p > .05$), but a significant effect of ISI ($F(2,38)=8.15$, $p < .01$), entirely due to its strong effect on speech, as indicated by a significant ISI by Condition interaction ($F(2,38)=4.90$, $p < 02$). Thus, whatever difficulties were induced in the poor readers by increasingly rapid presentation of synthetic stop-vowel syllables were not similarly induced by the non-speech control patterns.

These results demonstrate that the poor readers' difficulties with /ba/-/da/ discrimination were specific to speech, and cannot be attributed to a general auditory deficit in the perception of brief patterns of rapidly changing acoustic information. Nonetheless, it still seemed possible that the poor readers' difficulty with /ba/-/da/ might be specific to phonetic processing of brief formant transitions. Accordingly, we undertook Experiment 3.

EXPERIMENT 3

Two salient cues specify the presence of a stop consonant in a fricative-stop cluster, as in the word *split,* /splɪt/ (Fitch, Halwes, Erickson, & Liberman, 1980) or in the word *stay,* /steɪ/ (Best, Morrongiello, & Robson, 1981). First, a sharp drop in energy (i.e., silence) after the fricative indicates that the vocal tract has closed. Second, a sharp rise in F1 energy and frequency after the silence indicates that the tract is opening. The duration of the silence and the extent of the F1 frequency rise (i.e., of the F1 transition) are reciprocally related and additive in their perceptual effect. In synthetic speech they "trade": A given probability of hearing a stop can be determined either by a brief silence and a relatively extensive F1 frequency rise, or by a longer silence and a less extensive F1 frequency rise.

Morrongiello, Robson, Best, and Clifton (1984) exploited these facts in a study of 5-year-old children and adults. Their subjects identified members of two series of syllables ranging from /seɪ/ to /steɪ/. Each syllable consisted of a natural /s/ frication concatenated with a vocalic portion synthesized to sound either strongly or weakly as /deI/: The stronger /deI/ was cued by a lower F1 onset frequency, and so a more extensive F1 transition to its steady state, the weaker /deI/ was cued by a higher F1 onset and so a less extensive F1 transition. Various amounts of silence were inserted between the noise and the transitions to generate two /seI-steI/ continua, one with the high F1 onset, one with the low. Children and adults did not differ in their identification functions on the low F1 series. But on the high F1 series children needed less silence to hear a stop than adults, indicating that they weighted the less extensive F1 transition more heavily relative to the silence than the adults did. The authors concluded that five-year-olds were more sensitive to rapid intrasyllabic formant transitions than adults. Nittrouer (1992) reached the same conclusion for 3-, 4-, and 5-year-olds, compared with 7-year-olds and adults on identification of a /seɪ/ - /steɪ/ continuum. (For theoretical interpretation of such findings, see Nittrouer, Studdert-Kennedy & McGowan, 1989.)

Precisely the opposite result was reported by Tallal and Stark (1978) for a group of language-impaired children, tested on a /sa/ - /sta/ continuum, with fixed F1 transition extent and varying silent intervals between frication offset and vowel onset. These children needed a significantly longer silent interval between fricative and vowel than normal children to shift their judgments from /sa/ to /sta/; in other words, they were less sensitive to (or weighted less heavily) the extent of F1 transition than normals. A similar result was reported by Steffens et al. (1992) for adult subjects with familial dyslexia. Such results are consistent with Tallal's expectation that dyslexics and language-impaired children should display reduced sensitivity to brief acoustic events.

The present experiment was a variation of the above studies. Instead of fixing F1 onset frequency and varying silence, we followed Nittrouer (1992), fixing the silent interval and manipulating F1 onset frequency. If the 8-year old poor readers behaved perceptually like the 5-year olds of Morrongiello et al. (1984) and of Nittrouer (1992), they would require a higher F1 onset frequency (i.e., a less extensive F1 transition) to hear a stop than normals. On the other hand, if the poor readers' difficulties with /ba/-/da/ in Experiment 1 stemmed from a general deficit in the phonetic processing of brief formant transitions, they would require a lower F1 onset frequency (i.e., a more extensive F1 transition) than good readers to switch their judgments from /seɪ/ to /steɪ/. Finally, if poor readers were neither more nor less sensitive than good readers to variations in the extent of F1 transitions, we could conclude that they were neither developmentally delayed nor had difficulty in processing rapid formant transitions. Such an outcome would suggest that poor readers' difficulties with /ba/-/da/ did not stem from acoustic properties of the syllables.

Method

Stimuli

The stimuli were drawn from a /seI-steI/ continuum, each step made up of a natural sample of /s/ frication noise, followed by a synthetic vocalic portion. These stimuli, identical to those used by Nittrouer (1992), were modeled after those of Best et al. (1981) and Morrongiello et al. (1984). The duration of the frication noise was 120 msec and that of the vocalic portion 300 msec, accompanied by a falling F0 from 120 Hz to 100 Hz. F3 fell from 3196 Hz to 2694 Hz in the first 40 msec, remained there for the next 120 msec and rose to 2929 Hz over a 90 msec period where it remained steady for the last 50 msec. F2 remained constant at 1840 Hz during the first 160 msec and then rose to 2240 Hz over the next 90 msec, where it remained steady for the final 50 msec. F1 onset varied from 211 Hz to 611 Hz in 50 Hz steps, reaching 611 Hz over the first 40 msec, where it remained for the next 120 msec. Then F1 fell to 304 Hz over 90 msec where it stayed for the final 50 msec. A silent gap of 20 msec was inserted between the /s/ noise and each following vocalic segment.

Procedure

Subjects were trained to a 90 percent correct criterion with ten repetitions of good exemplars from the two categories. They were then presented with the test stimuli, one at a time, in a random order, with an unlimited interval in which to respond; there was a total of 90 trials (9 tokens X 10 repetitions per token). The stimuli were presented by a Compaq 386 portable computer (IBM-clone) through a 901F Frequency Devices filter and TDH-39 headphones. Subjects responded both by saying aloud what they heard and by pointing to a picture of a little girl with an empty blurb balloon for the word "say" and of a man appearing to admonish a dog with a raised hand for "stay". Responses were registered directly to the computer by the experimenter.

Results and Discussion

Figure 3 displays the mean probability of /steɪ/ responses as a function of F1 transition onset frequency for the two groups. The two functions are very similar, although poor readers seem to have a somewhat shallower slope. Close inspection reveals that the shallower appearance is largely due to the slope between stimuli 4 and 5 where the poor readers' function crosses the good readers'.

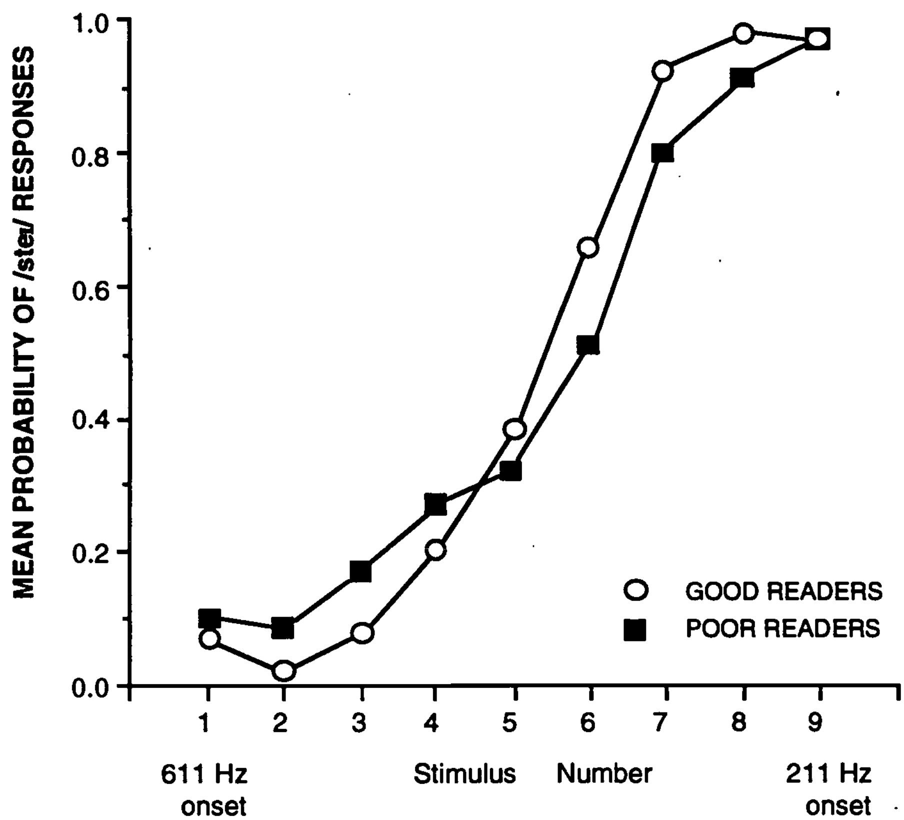


Figure 3. **Mean identification functions for /seI-steI/ in good and poor readers. Stimulus numbers refer to F1 onset-frequencies ranging from 611 Hz to 211 Hz in 50 Hz steps.**

Cumulative normal distributions were fit to individual data by the method of least squares (Finney, 1964) and yielded means and standard deviations of the individual distributions. The mean is an estimate of the phoneme boundary, the value of the F1 transition for which "say" and "stay" responses are equally likely; the standard deviation corresponds to the reciprocal of the slope of the cumulative function. Whereas the good and poor readers have almost the same mean phoneme boundaries (398 (39.1) Hz and 396 (45.5) Hz, respectively, with standard deviations in parentheses), the former have steeper mean slopes (15.5 (6.9) Hz vs. 12.4 (9.5) Hz). The differences between the groups were not significant, however, in either phoneme boundary ($t(38)=.18$, $p > .05$) or slope ($t(38)=1.16, p > .05$).

Individual data of several subjects were not well fit by the cumulative normal curve, throwing doubt on the propriety of probit analysis. An alternative, though coarser, measure of response to F1 onset variations was therefore computed, namely, the total number of "stay" responses across the continuum. Once again, the groups did not differ significantly: The mean number of "stay" responses by the good readers (42.3) was almost equal to that of the poor readers (41.1) ($t(38)=.47$, $p >.05$). Good and poor readers were also equally variable on this measure: Both groups had the same standard deviations (7.7).

These results demonstrate that, unlike the language-impaired children of Tallal and Stark (1978) and the adult dyslexics of Steffens et al. (1992), the poor readers of the present study were not less sensitive than normals to the F1 transition: They did not require a more extensive transition in order to hear a stop in a syllable-initial /s/-stop cluster. Also, unlike the poor readers of previous studies described in the introduction, the poor readers were no less consistent than good readers in identifying synthetic syllables distributed along a continuum. Perhaps this was because the contrast here was between the presence and absence of a full segment marked by several features rather than between segments differing on a single feature, as in earlier studies. This interpretation fits neatly with the results of Experiment 1, where poor readers' difficulties with a single feature contrast (/ba/-/da/) vanished for triple feature contrasts (/ba/-/sa/, /da/-/ʃa/).

In any event, the poor readers of this study, despite their demonstrated difficulties with /ba/-/da/ discrimination and TOJ, clearly did not suffer either from the general auditory deficit posited by Tallal (1980) or from a corresponding domain-specific, phonetic deficit in the perception of brief formant transitions. Nor did they exhibit the developmental delay, suggested by the findings of Morrongiello et al. (1984) and of Nittrouer (1992), in the form of heightened sensitivity to transitions. In other words, the reading-impaired children were completely normal, as sensitive to a very brief acoustic event in speech as they had proved to be in non-speech (Experiment 2).

GENERAL DISCUSSION

The general auditory account of phonological deficits in both language-impaired and reading-impaired children (e.g. Tallal et al., 1991, pp. 369-370) makes two independent claims: (i) the basic deficit is in "temporal processing"; (ii) the deficit is general rather than specific to speech. The present study lends no support to either of these claims for reading-impaired children.

We should acknowledge, however, that the study has certain limitations. First, due in part to the rigorous criteria for subject selection, the poor readers were less severely impaired than those of Tallal (1980) whose children were reading at least one year below grade level. Yet the poor readers did display precisely the difficulties with /ba/-/da/ TOJ that Tallal (1980) proposed as symptomatic of a phonological disorder and that Reed (1989) subsequently observed in some reading-impaired children. Whether the difference in reading level is a serious concern depends therefore on how likely it is that identical difficulties with /ba/-/da/ TOJ stem from different perceptual deficits in more severely impaired than in less severely impaired readers. We do not find this likely. As noted in the Subjects section, several large-scale studies converge on the conclusion that reading ability is normally distributed with no qualitative difference between those who are simply less skilled and those who meet standard criteria as reading-disabled (e.g., Shaywitz et al., 1995; Stanovich & Siegel, 1994). If this is so, the results of the present experiments can be generalized to specifically reading-impaired children who have difficulty with /ba/-/da/ TOJ, regardless of their degree of reading impairment. Whether the results can also be generalized to specifically language-impaired children, whose difficulties are not confined to reading, is a question for future research incorporating the non-speech controls which previous studies of such children have omitted (e.g., Tallal & Piercy, 1974; Tallal & Stark, 1981; Tallal, Miller, Bedi, Byma, Wang, Nagarajan, Schreiner, Jenkins, & Merzenich, 1996).

A second limitation of this study (and of any other study designed to test Tallal's hypothesis) is that its results cannot disprove the hypothesis: They can merely fail to support it where support would be expected. Yet we should not forget that the hypothesis of a general auditory deficit underlying difficulties with rapid stop-vowel syllable perception is itself purely speculative and without experimental support. No one has ever shown that either language-impaired or reading-impaired children have difficulty in discriminating both brief, rapidly changing speech sounds (formant transitions) and acoustically matched non-speech control patterns. The concurrent difficulties with both /ba/-/da/ and tone TOJ, reported by Reed (1989) for impaired readers, do not fill the gap, because rapidly presented pairs of discrete, complex tones, differing in fundamental frequency (pitch), do not qualify as acoustically matched controls for the rapid continuous sweeps of formant transitions, differing in spectral distribution (timbre). Nor can we attribute concurrent tone and stop-vowel difficulties to some more general deficit in, say, speed of auditory stimulus classification (cf. Nicholson & Fawcett, 1994). Instead, such evidence as we have in this regard seems to favor the notion of independent deficits in different aspects of speech and non-speech discriminative capacity. We briefly consider this proposal in the following sections.

Difficulty in identification stems from a deficit in discriminative capacity not in "temporal processing"

From one of her earliest papers (Tallal & Piercy, 1973, p.396), to a recent paper eighteen years later, Tallal has repeatedly acknowledged that the "...sequencing deficit...in dysphasic children is...secondary...to the more primary [*sic*] deficit in...discrimination of rapidly presented stimuli" (Tallal et al., 1991, p.365). In other words, the deficit is in stimulus discrimination, and so *a fortiori* in identification, not in TOJ itself. Experiment lb of the present study, in which apparent TOJ errors vanished for readily identifiable syllables, merely confirms therefore what Tallal herself would predict.

Yet descriptions of the deficit as one of "temporal processing" persist, maintained by a shift in the locus of the supposedly defective "temporal processing" from the judgment of sequence (TOJ) to the judgment of rapid spectral change. Thus, in the very paper from which the above quotation is drawn, we also read of "...basic temporal processing deficits which interfere...with adequate perception of specific verbal stimuli which require resolution of brief duration formant transitions, resulting in disordered language development" (p.363). The phrase "specific verbal stimuli" evidently refers to stop-vowel syllables, such as /ba/ and /da/. Here, then, perception of a contrast between syllables with identical temporal properties (duration, rate of spectral change), but differing in the frequencies and directions of spectral change at syllable onset, is deemed "temporal", simply because the critical spectral shift is of "brief duration". The passage nicely illustrates the confusion between temporal perception and rapid perception to which we referred in the introduction. (For fuller analysis than is possible here of Tallal's concept of "temporal processing", see Studdert-Kennedy & Mody, 1995.)

Yet, even if we set aside the seemingly trivial, though conceptually and substantively critical, issue of terminology, the claim that poor readers find /ba/-/da/ difficult to identify because these syllables "...require resolution of brief duration formant transitions", is not only without experimental support, but is also at odds with the results of the present study. In Experiment 3 poor readers, chosen precisely for their difficulties with /ba/-/da/ TOJ, were asked to detect the presence or absence of a stop consonant between a syllable-initial fricative and vowel; they proved no less sensitive than good readers to small variations in the frequency extent of brief F1 transitions. And in Experiment 2, where the brief transitions were in non-speech sine waves, the poor readers discriminated between them as well as good readers. Nor, finally, are /ba/ and /da/ difficult for poor readers because they are acoustically similar: The non-speech sine waves were as acoustically similar as the syllables on which they were modeled. The source of the difficulty, then, is evidently phonetic: /ba/ and /da/ are difficult to discriminate and identify at rapid rates of presentation because, although phonologically contrastive, they are phonetically similar. As earlier remarked, /ba/ and /da/, like /sa/ and /ʃa/, which language-impaired children may also find difficult to discriminate (Tallal & Stark, 1981), differ on a single phonetic feature.

How, then, is this deficit in speech perception related to the difficulty with non-speech TOJ, repeatedly reported for reading-impaired children?

Non-speech Tone Temporal Order Judgment

As we saw in the introduction, some relation seems to obtain between performance on rapid non-speech tone TOJ and phonological decoding skill in reading, but the relation is at best statistical and its functional basis is obscure. Tallal's (1980) view that difficulty with tone TOJ is one symptom of a general deficit in "temporal processing", of which difficulty with /ba/-/da/ TOJ is another, is undermined by the present finding that the latter is neither general nor "temporal", but specific to speech and a consequence of difficulty in identification at rapid rates of presentation.

That tone TOJ itself may not be a reliable measure of general temporal processing is also strongly suggested by recent work of Watson and Miller (1993). These authors studied a sample of 94 undergraduates, 24 of whom were diagnosed as reading-disabled, on a battery of 35 tests assessing nine factors potentially relevant to reading skill. They found no significant differences between normal and reading-disabled subjects in intelligence, simple auditory discrimination (pitch, loudness), or in nonverbal "auditory temporal processing", as measured by discrimination thresholds for (i) tone duration around a 100 msec, 1 kHz standard, (ii) interpulse intervals differing from a 40 msec standard, and (iii) presence of an embedded 10-200 msec tone in a nine tone sequence. They did find, however, highly significant differences on tests of speech perception, short- and long-term verbal memory, phoneme segmentation, and of TOJ for tones of 550 and 710 Hz, varying in duration from 20 to 200 msec, with no silent gap between tones.

Yet, upon entering tone TOJ into a linear structural relations analysis in combination with the three tests of temporal discrimination described above, Watson and Miller (1993) found no relation between non-verbal temporal processing and either phonological skills or reading itself. They therefore concluded: "Overall, these results indicate that the phonological processing variables are highly associated with speech perception, and that nonverbal temporal processing does not explain a significant amount of variance in the phonological variables independent of speech perception and intelligence" (p. 859).

Two points concerning Watson and Miller's study deserve emphasis. First, the three tests that call for judgments of a temporal dimension (duration), and therefore have clear face validity as measures of temporal processing, did not separate normal from disabled readers. Second, tone TOJ, a test that calls in the first instance for judgment of a non-temporal dimension (fundamental frequency), and only secondarily for judgment of temporal sequence, did separate normal from impaired readers, but did not cluster with tests of temporal processing.

We are left then with the puzzle of why tests calling for rapid identification of the relative pitch of non-speech tones separate, at least statistically, normal from impaired readers. A new approach to this puzzle comes from a recent experimental study by Nicholson and Fawcett (1993, 1994). These authors compared groups of dyslexics, aged 15 and 11 years, with chronological age (CA) and reading age (RA) controls on selective choice reaction time (SCRT) to pure tones differing in pitch, and on reaction time in a lexical decision task. They found that the dyslexics were significantly slower than CA controls (though not than RA controls) on SCRT to tones (a "quantitative deficit" in speed of decision), and significantly slower even than RA controls on lexical decision for words, but not for non-words (a "qualitative deficit" in speed of lexical access).

The SCRT results suggested a general slow-down in processing speed; but this could not account for the "qualitative" difference between words and non-words in speed of lexical decision. The authors therefore concluded that: "Phonological deficits must still be posited over and above any putative deficits in information processing speed" (Nicholson & Fawcett, 1994, p.45). In other words, they proposed that the lexical decision effect stemmed from two independent deficits: "...a phonological deficit in lexical access speed, together with a non-phonological deficit in stimulus classification speed" (p.46).

The co-occurrence of phonological impairment and other cognitive problems is not at odds, of course, with the real-world circumstance of multiple deficits in reading-disabled children. For example, it is now recognized that roughly 40% of such children may have an independent co-occurring "attention-deficit-disorder" (Shaywitz, Fletcher & Shaywitz, 1994). Presumably, phonological deficits are present in every poor reader (Shankweiler, Crain, Katz, Fowler, Liberman, Brady, Thornton, Lundquist, Dreyer, Fletcher, Stuebing, Shaywitz, & Shaywitz, 1995). Co-occurring non-phonological deficits are evidently more variable in occurrence, as shown by the repeated finding that some poor readers fall within the normal range on tone TOJ (e.g., Bedi, 1994; Reed, 1989; Tallal, 1980).

CONCLUSION

Deficits in speech perception among reading-impaired children are domain-specific and phonological rather than general and auditory in origin. Such children's difficulties with /ba/-/da/ TOJ (when they are present) arise from difficulty in identifying the phonological categories of phonetically similar speech sounds rather than from deficits either in temporal order judgment itself or in processing the brief acoustic changes of formant transitions. The full nature, origin, and extent of the perceptual deficit remain to be determined. For example, how poor readers' deficits in speech perception relate to their characteristically impaired phonological awareness, and so to reading, is a question we must leave to future research.

REFERENCES

Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology, 61*, 303-321.

Apthorp, H. (1995). Phonetic coding and reading in college students with and without learning disabilities. *Journal of Learning Disabilities, 28*, 342-352.

Bedi, G. C. (1994). *Low level visual and auditory processing in dyslexic readers.* Unpublished doctoral dissertation, City University of New York.

Best, C., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and non-speech perception. *Perception and Psychophysics, 29*, 191-211.

Blumstein, S., Tartter, V., Nigro, G., & Statlender, S. (1984). Acoustic cues for the perception of place of articulation in aphasia. *Brain and Language, 22*, 128-149.

Borden, G., Harris, K., & Raphael, L. (1994). *Speech science primer* (3rd edition). Baltimore, MD: Williams & Wilkins.

Bradley, L., & Bryant, P. (1991). Phonological skills before and after learning to read. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 37-45). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brady, S. A. (1991). The role of working memory in reading disability. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Liberman* (pp. 129-152). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brady, S. A., Poggie, E., & Rapala, M. (1990). Speech repetition abilities in children who differ in reading skills. *Language & Speech, 32*, 109-122.

Brady, S. A., Shankweiler, D. P., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35*, 345-367.

De Weirdt, W. (1988). Speech perception and frequency discrimination in good and poor readers. *Applied Psycholinguistics, 16*, 163-183.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test-revised.* Circle Pines, MN: American Guidance Service.

Finney, D. J. (1964). *Probit analysis.* Cambridge: Cambridge University Press.

Fitch, H. L., Halwes, T., Erickson, D. M.,& Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27*, 343-350.

Freeman, B., & Beasley, D. (1978). Discrimination of time-altered sentential approximations and monosyllables by children with reading problems. *Journal of Speech and Hearing Research, 21*, 497-506.

Gathercole, S., & Baddeley, A. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition. *European Journal of Psychology of Education, 78*, 259-272.

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology, 32*, 401-424.

Gould, J. H., & Glencross, D. J. (1990). Do children with a specific reading disability have a general serial-ordering deficit? *Neuropsychologia, 28*, 271-278.

Hansen, J., & Bowey, J. (1994). Phonological analysis skills, verbal working memory, and reading ability in second-grade children. *Child Development, 65*, 938-950.

Hurford, D. P., & Sanders, R. E. (1990). Assessment and remediation of a phonemic discrimination deficit in reading disabled second and fourth graders. *Journal of Experimental Child Psychology, 50*, 396-415.

Jorm, A. F., Share, D. L., Maclean, R., & Matthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology and Psychiatry, 27*, 45-55.

Katz, R., Shankweiler, D. P., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32*, 474-484.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology, 52*, 127-137.

Liberman, I. Y., Mann, V. A., Shankweiler, D. P., & Werfelman, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex, 18*, 367-376.

Lieberman, P., Meskill, R. H., Chatillon, M., & Shupack, H. (1985). Phonetic speech perception deficits in dyslexia. *Journal of Speech and Hearing Research, 28*, 448-486.

Morrongiello, B. A., Robson, R. C., Best, C. T., & Clifton, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology, 37*, 231-250.

Nicholson, R. I., & Fawcett, A. J. (1993). Children with dyslexia classify pure tones slowly. In P. Tallal, A. M. Galaburda, R. R. Llinas, & C. von Euler (Eds.). *Temporal information processing in the nervous system* (pp. 387-389). New York: Annals of the New York Academy of Sciences, Vol. 682.

Nicholson, R. I., & Fawcett, A. J. (1994). Reaction times and dyslexia. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47A*, 29-48.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics, 20*, 351-382.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32*, 120-132.

Olson, R. K. (1992). Language deficits in "specific" reading disability. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 895-916). New York: Academic Press.

Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on assessment issues* (pp. 243-278). Baltimore, MD: Paul H. Brookes.

Pallay, S. L. (1986). *Speech perception in dyslexic children*. Unpublished doctoral dissertation, City University of New York.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology, 20*, 241-248.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 28-53.

Rapala, M. M., & Brady, S. A. (1990). Reading ability and short-term memory: The role of phonological processing. *Reading & Writing: An Interdisciplinary Journal, 2*, 1-25.

Reed, M .A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology, 48*, 270-292.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947-950.

Riedel, K., & Studdert-Kennedy, M. (1985). Extending formant transitions may not improve aphasics' perception of stop consonant place of articulation. *Brain and Language, 24*, 223-232.

Shankweiler, D. P., & Crain, S. (1986). Language mechanisms and reading disorders: A modular approach. *Cognition, 24*, 1390-168.

Shankweiler, D. P., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 531-545.

Shankweiler, D. P., Crain, S., Katz, L., Fowler, A. E., Liberman, A. M., Brady, S. A., Thornton, R., Lundquist, E., Dreyer, L., Fletcher, J. M., Stuebing, K. K., Shaywitz, S. E., & Shaywitz, B. A. (1995). Cognitive profiles of reading-disabled children. *Psychological Science, 6*, 1-8.

Shaywitz, B. A., Escobar, M. D., Shaywitz, S., & Fletcher, J. M. (1994). Evidence that dyslexia may represent the lower tail of the normal distribution of reading ability. *New England Journal of Medicine, 326*, 145-150.

Shaywitz, S. E., Fletcher, J. M., & Shaywitz, B. (1994). Issues in the definition and classification of attention deficit disorder. *Topics in Language Disorders, 14*, 1-25.

Snowling, M., Goulandris, N., Bowlby, M., & Howell, P. (1986). Segmentation and speech perception in relation to reading skill: A developmental analysis. *Journal of Experimental Child Psychology, 41*, 489-507.

Stanovich, K. (1991). Word recognition: Changing perspectives. In M. L. Kamil, P. Mosenthal, and D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 418-432). New York: Longman.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24-53.

Steffens, M. L., Eilers, R., Gross-Glen, K., & Jallad, B. (1992). Speech perception deficits in adult subjects with familial dyslexia. *Journal of Speech and Hearing Research, 35*, 192-200.

Stone, B. H., & Brady, S. A. (1995). Evidence for phonological processing deficits in less-skilled readers. *Annals of Dyslexia, 45*, 51-78.

Studdert-Kennedy, M., & Mody. M. (1995). Auditory temporal perception deficits in the reading-impaired: A critical review of the evidence. *Psychonomic Bulletin & Review, 2*, 508-514.

Tallal, P. (1980). Auditory temporal perception, phonics and reading disabilities in children. *Brain and Language, 9*, 182-198.

Tallal, P. (1984). Temporal or phonetic processing deficit in dyslexia? That is the question. *Applied Psycholinguistics, 5*, 13-24.

Tallal, P. (1990). Fine-grained discrimination deficits in language-learning impaired children are specific neither to the auditory modality nor to speech perception. *Journal of Speech and Hearing Research, 33*, 616-617.

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: A case for preeminence of temporal processing. In P. Tallal, A. M. Galburda, R. R. Llinas and C. von Euler (Eds.), *Temporal information processing in the nervous system* (Vol. 682, pp. 27-47). New York: Annals of the New York Academy of Sciences.

Tallal, P., & Newcombe, F. (1978). Impairment of auditory perception and language comprehension in dysphasia. *Brain and Language, 5*, 13-24.

Tallal, P., & Piercy, M. (1973). Developmental aphasia: Impaired rate of nonverbal processing as a function of sensory modality. *Neuropsychologia, 11*, 389-398.

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia, 12*, 83-93.

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia, 13*, 69-74.

Tallal, P., & Stark, R .E. (1978). Identification of a /sa/ -/sta/ continuum by normally developing and language-delayed children. *Journal of the Acoustical Society of America, 64*, 50.

Tallal, P., & Stark, R .E. (1981). Speech acoustic cue discrimination abilities of normally developing and language impaired children. *Journal of the Acoustical Society of America, 69*, 568-574.

Tallal, P., Sainburg, R. L., & Jernigan, T. (1991). The neuropathology of developmental dysphasia: Behavioral, morphological and physiological evidence for a pervasive temporal processing disorder. *Reading and Writing: An Interdisciplinary Journal, 3*, 363-377.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 272*, 81-84.

Taylor, H. G., Lean, D., & Schwartz, S. (1989). Pseudo-word repetition in learning disabled children. *Journal of Auditory Research, 25*, 167-173.

Vellutino, F. R., Pruzek, R., Stegger, J., & Meshoulam, U. (1973). Immediate visual recall in poor and normal readers as a function of orthographic-linguistic familiarity. *Cortex, 8*, 106-118.

Wagner, R. K., & Torgesen, J. (1987). The nature of phonological processes in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Watson, B. U., & Miller, T. K. (1993). Auditory perception, phonological processing and reading ability/disability. *Journal of Speech and Hearing Research, 36*, 850-863.

Watson, M., & Rastatter, M. (1985). The effects of time compression on the auditory processing abilities of learning disabled children. *Journal of Auditory Research, 25*, 167-173.

Wechsler, D. (1974). *Wechsler intelligence scale for children-revised*. New York: The Psychological Corporation.

Werker, J., & Tees, R. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41*, 48-61.

Woodcock, R. W. (1987). *Woodcock reading mastery tests-revised*. Circle Pines, MN: American Guidance Services.

FOOTNOTES

Journal of Experimental Child Psychology, in press. (Note: The editors of JECP invited Paula Tallal to respond to this paper, but she declined to do so.)

†Albert Einstein College of Medicine, New York. Now at ENST, Dept. Signal, Paris, France.

‡Also at the University of Rhode Island.

[1]Riedel and Studdert-Kennedy (1985), who modeled their synthetic syllables on those of Tallal & Piercy (1975), found that when transitions were increased from 30 to 82 ms, identifications reported by their aphasic subjects included, in addition to /ba/-/da/: /wa/-/la/, /bwa/-/dla/, /wa/-/da/ and /ra/-/ja/. The diversity of response was probably due to the absence of systematic variations in prevoicing, normally present in natural or well-synthesized /ba/-/wa/ and /da/-/ja/ contrasts.

[2]In accordance with the criteria detailed in the Subjects section, 122 subjects were eliminated from the study for: Borderline reading scores that did not fall clearly within the skilled or less-skilled reading groups (n=28); high PPVT-R or WISC-R scores (n=60); hearing difficulty, bilingualism, or documented attention deficit disorder (n=32); school absence (n=2). The remaining 98 subjects comprised 66 good readers and 32 poor readers. From these we eliminated: 29 good readers and 2 poor readers due to lack of match on IQ or age; good readers with at least one TOJ error (n=13); and poor readers with fewer than three TOJ errors (n=10). Four good readers were then dropped at random to form two matched groups with n=20.

[3]Performance on the /ba/-/da/ TOJ task is far from normally distributed either within or across the populations of good and poor readers. Of the 40 good readers tested on 36 trials each, 27 (67%) made no errors at all, 35 (87%) made fewer than three errors, and only 5 (12%) made three or more errors; by contrast, of the 31 poor readers tested 1 (3%) made no errors, 10 (32%) made fewer than three errors, and 20 (65%) made three or more errors.

APPENDIX

An anonymous reviewer questioned the validity of the comparison between good and poor readers on non-speech discrimination because, by selecting good readers who made no errors on /ba/-/da/ TOJ, we had truncated "the variability expected in normal children" and so had invited in any test of "another sound contrast, regression to the mean ...in both groups, as is seen in the data." We should note, first, that, the most important finding of Experiment 2 was not the lack of a main effect for group, for which regression might perhaps be called to account (although see below), but the lack of a significant effect of ISI on non-speech for the poor readers. This result, taken with the highly significant effect of ISI on speech in Experiment 1a, demonstrates that the poor readers' difficulties with rapid stimulus presentation were confined to speech; a two-way ANOVA across Experiments 1a and 2 for the poor readers alone confirms this conclusion (see main text). Notice further that the regression hypothesis cannot predict the different effects of ISI across conditions, because subjects were selected by speech TOJ errors summed across ISIs, not separately for each ISI.

As for the regression hypothesis itself, comparison of the error distributions in the parent populations of good and poor readers with those in the experimental samples demonstrates that the hypothesis is implausible. The distributions of errors on /ba/-/da/ discrimination, like those on /ba/-/da/ TOJ, were positively skewed for both the 40 good readers and the 31 poor readers from whom the final selection of subjects was made. The positive skews are indicated by the fact that in each distribution the mean was greater than the median. For the good readers (n=40), the median was actually indeterminate because 30 (75%) of the subjects made no errors at all, while the mean was 1.4 (pulled up from 0.4 by two mavericks who made more errors than any poor reader, viz., 17 and 21); for the poor readers (n=31), the median was 2.7, the mean 3.9 (as compared with 6.2 and 7.4 respectively, for the poor readers selected for the experimental group (n=20)). Thus, the children selected to participate as good and poor readers were from opposite extremes of two quite differently weighted distributions of speech errors. Yet on the non-speech test, in a shift that the regression hypothesis would not predict, the two groups not only came closer together, but switched their relative positions, so that the good readers now made more errors than the poor readers. We see this in the values of the means and medians on the non-speech test: For good readers a median of 6.0 and a mean of 6.7, for poor readers a median of 2.5 and a mean of 4.0. Thus the non-speech mean for the poor readers was indeed numerically compatible with regression; that is, on the non-speech test, the mean error rate for poor readers did shift to a less extreme value, almost equal, in fact, to the mean of the distribution of speech errors on /ba/-/da/ discrimination in the parent population. Yet for the good readers (who were presumably more susceptible to regression, because drawn entirely from the floor of their distribution), the shift was not to a less extreme value closer to the population mean, but to an even more extreme value (i.e., even further from the mean than zero), at the opposite end of their distribution. Such a shift is not toward the mean, but past the mean, from the non-skewed end to the skewed end of a heavily skewed distribution. Regression is surely not a plausible explanation for a shift of this magnitude.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 25-34

Auditory Temporal Perception Deficits in the Reading-impaired: A Critical Review of the Evidence*

Michael Studdert-Kennedy and Maria Mody†

We assess evidence and arguments brought forward by Tallal (e.g., 1980) and by the target paper (Farmer & Klein, 1995) for a general deficit in auditory temporal perception as the source of phonological deficits in impaired readers. We argue that (1) errors in temporal order judgment of both syllables and tones reflect difficulty in identifying similar (and so readily confusable) stimuli rapidly, not in judging their temporal order; (2) difficulty in identifying similar syllables or tones rapidly stem from independent deficits in speech and nonspeech discriminative capacity, not from a general deficit in rate of auditory perception; (3) the results of dichotic experiments and studies of aphasics purporting to demonstrate left-hemisphere specialization for nonspeech auditory temporal perception are inconclusive. The paper supports its arguments with data from a recent control study. We conclude that, on the available evidence, the phonological deficit of impaired readers cannot be traced to any co-occurring non-speech deficits so far observed, and is phonetic in origin, but that its full nature, origin, and extent remain to be determined.

The target paper (Farmer & Klein, 1995) starts from the widely accepted assumption (with which we agree) that dyslexia, or reading impairment, is often, if not always, associated with a phonological deficit. The stated purpose of the paper is, then, to "evaluate the plausibility" of the hypothesis of Tallal (1984) that this deficit is "a symptom of an underlying auditory temporal processing deficit" (abstract). Unfortunately, this hypothesis has never been clearly or consistently framed by Tallal herself, and Farmer and Klein do nothing to clarify it. Much of what we have to say therefore entails analysis of Tallal's work no less than that of Farmer and Klein. Our remarks are limited to studies of audition because these alone bear on possible weaknesses in speech perception from which a phonological deficit might arise.

As best we can determine, Tallal's hypothesis, originally a proposal concerning the perceptual deficits of dysphasic children, has come to comprise three logically independent, but mutually reinforcing, propositions (for a recent review, see Tallal, Miller, & Fitch, 1993): (1) "Rapid auditory temporal processing" is essential to speech perception; (2) specialization of the left cerebral hemisphere for speech perception (and so for phonology), in most right-handed individuals, is grounded in a prior specialization of that hemisphere for "rapid auditory temporal processing"; (3) phonological deficits in some dysphasic children, some aphasic adults, and some impaired readers, or dyslexics, stem from deficits in "rapid auditory temporal processing." For Tallal and her colleagues, the first proposition, though far from clear, seems to be axiomatic. They have given most attention to the third proposition, rather less to the second. Farmer and Klein follow this distribution of emphasis, and we largely follow the target paper. But all three propositions seem to be required for a full statement of the hypothesis.

Preparation of the paper was supported in part by NICHD Grant HD-01994 to Haskins Laboratories. We thank Susan Brady for useful discussion and James Neely for judicious editorial advice.

BACKGROUND

We begin by distinguishing two concepts repeatedly confused both in the work of Tallal (e.g., Tallal & Newcombe, 1978; see below) and in the target paper (Farmer & Klein, 1995): temporal perception (or "processing") and rapid perception. In normal linguistic usage, temporal perception contrasts with, say, spectral perception in audition, or spatial perception in vision; it refers to perceiving the temporal properties of events (duration, sequence, relative timing, rhythm). To the extent that they identify temporal perception with sequential perception Farmer and Klein (1995) follow this usage. However, when they equate temporal perception with "processing rapidly presented stimuli" (pp. 4, 39) or with perceiving "spectral changes in the time frame of tens of milliseconds" (p. 20), they confuse rate of perception with perception of rate. Perception is "temporal" if the defining property of the perceived event is temporal; it does not become "temporal" by virtue of being effected rapidly. This distinction is not a trivial matter of terminology; it is conceptually and substantively important, because the precise nature of a speech perceptual deficit bears directly on what correlated deficits we might expect in speech production, on the nature of the underlying defective neural mechanism, and on how we might go about remediation. Let us see, then, how the confusion has arisen.

In a series of studies from which much of Tallal's later work springs (see the review by Tallal et al., 1993), Tallal and Piercy (1973, 1974, 1975) compared dysphasic children with normal controls on tests of discrimination and temporal order judgment (TOJ) for pairs of stimuli presented with "long" (428 msec) and "short" (8-305 msec) interstimulus intervals (ISI). The stimuli included (1) short (75 msec) and long (250 msec) complex tones differing in fundamental frequency (100 vs. 305 Hz); (2) long (250 msec) steady-state synthetic vowels (/ɛ/,/æ/); (3) short (43 msec) steady-state synthetic vowels (/ɛ/,/æ/) immediately followed by a longer (207 msec) steady-state synthetic vowel (/ɪ/); (4) synthetic consonant-vowel (CV) syllables (/ba/ vs. /da/, 250 msec), in which the contrasting second ($F2$) and third ($F3$) formant transitions at syllable onset were either short (43 msec) or long (95 msec) in duration. The dysphasic children performed significantly worse than normals on short tones, short vowels and short transition consonants, at short ISIs, but not on the corresponding long stimuli, nor at long ISIs. Moreover, performance on discrimination and TOJ did not differ significantly. From this last finding Tallal & Piercy (1973, p. 396) inferred that apparent deficits in auditory sequencing could be due to difficulty in discriminating and identifying stimuli rapidly, rather than to deficits in temporal perception itself. From the similar results for short steady-state vowels and short transition consonants, and from improved performance on long transition consonants,[1] they concluded that *"it is the brevity not the transitional character*...[of the formant transitions] of synthesized stop consonants which results in the impaired perception of our dysphasic children" (Tallal & Piercy, 1975, p. 73, our italics). And from the similar results for short tones and short speech sounds they concluded that the dysphasics' impairment was a general auditory deficit, not specific to speech (Tallal & Piercy, 1973, 1974). These three conclusions directly address three main issues in the target paper concerning auditory temporal perception. Although Farmer and Klein cite Tallal and Piercy (1973, 1975, but not 1974), they do not comment on these conclusions. Nonetheless, let us examine each in turn.

Discrimination, not TOJ. The most important conclusion, in the present context, is that dysphasic children's difficulties were with discrimination, not with TOJ itself. Similarly, Tallal (1980), in her only study of specifically reading-impaired children, again found that discrimination and TOJ of complex tones did not differ significantly, and again concluded that the children's "...difficulty with temporal pattern perception may stem from a *more primary* [sic] *perceptual deficit that affects the rate* at which they can process perceptual information" (Tallal, 1980, p. 193, our italics; see Tallal, Sainburg, & Jernigan, 1991, p. 365, for a recent restatement of this view.) Notice that on this account a slowed *rate of perception,* as indicated by errors on TOJ at short ISI's, *is not a general cause of the impaired child's difficulties, but a specific result of low discriminative capacity along a particular dimension.* Reed (1989) the only other researcher to extend Tallal's TOJ tests to reading-impaired children, concurs, proposing that TOJ not be viewed as a measure of temporal processing at all: "...the temporal task simply provides a setting where perceptual capabilities can be stressed, allowing the measurement of differences in the absence of ceiling performance" (Reed, 1989, p. 287). In other words, the TOJ task is primarily a diagnostic tool for picking up subtle deficits in discriminative capacity, and these deficits reveal themselves in a slowed rate of perception, specific to the dimension being tested. If this view is correct, as we believe it is, we have no experimental evidence at all for deficits either in auditory temporal perception or in general rate of auditory perception among the reading-impaired. Again, although Farmer and Klein cite Tallal (1980) and Reed (1989), they do not report these conclusions or consider how their work tempers the argument of the target paper.

Perceiving formant transitions: The brevity of transitions, not their transitional character (i.e. not acoustic changes), causes the perceptual difficulty. This conclusion contradicts the claim that supposed deficits in perception of formant transitions stem from difficulty in perceiving "rapid acoustic changes". As noted in the next paragraph, no subsequent experimental work has established that claim.

The relation between deficits in speech and non-speech auditory perception. Tallal & Piercy (1973, 1974) demonstrated that dysphasic children suffered from deficits in both speech (phonetic) and non-speech (auditory) discrimination, but not that the latter caused the former, as an auditory account of phonological deficits would require. Later papers (Schwartz & Tallal, 1980; Tallal & Newcombe, 1978), which we analyze in some detail below, tried to fill the gap by proposing that the phonological capacities of the left cerebral hemisphere rested on capacities for "auditory temporal processing", presumed deficient in certain clinical populations. Yet neither Tallal & Piercy (1974) nor anyone else has demonstrated that difficulties with /ba/ - /da/ discrimination are auditory rather than phonetic, because no one has demonstrated equivalent difficulties for matched non-speech control patterns with brief and rapidly changing onset frequencies. Similarly, we cannot attribute difficulties in discriminating /ba/ - /da/ or /ɛ/ - /æ/ to concomitant difficulties in discriminating the fundamental frequency of complex non-speech tones, because neither speech contrast depends on variations in fundamental frequency.

A CONCEPTUAL MUDDLE

In light of the foregoing, we may well be puzzled by Tallal's continued use of such phrases as "temporal processing disorder" (e.g., Tallal, Sainburg, & Jernigan, 1991, p. 363) and deficits in "the processing of rapidly changing acoustic spectra" (e.g., Tallal, Miller, & Fitch, 1993, p. 40) to describe the condition of poor readers, and by the target paper's uncritical acceptance of this terminology. To understand how this usage has come about we must turn to a paper, not mentioned by Farmer and Klein, that extended Tallal's TOJ tests to adults with brain lesions due to missile wounds (Tallal & Newcombe, 1978). We discuss the findings of this paper below (see *Aphasic studies*). Here, our immediate interest is in how the authors describe the work of Tallal and Piercy (1973, 1974, 1975). Two inconsistencies deserve note. First, despite the earlier conclusion that the dysphasic children's difficulties were with identifying stimuli at rapid rates of presentation rather than with temporal perception itself, the introduction to the new paper states that these earlier studies "...strongly support the hypothesis that some developmental language disorders may result from a primary impairment in auditory temporal analysis" (Tallal & Newcombe, 1978, p. 13). From this point on the phrases "impaired on auditory temporal processing tasks" (p. 14), "defect in temporal acoustic processing" (p. 22), and the like are used interchangeably with "impairment in responding to rapidly presented acoustic information" (p. 14). No justification is offered either for this conflation of perceiving the temporal properties of events with perceiving brief events rapidly, or for the switch in interpretation from that of Tallal and Piercy (1973).

The second inconsistency is no less serious. Despite the conclusion of Tallal and Piercy (1975) that "...it is the brevity not the transitional character" (p. 73) of formant transitions that causes difficulty for dysphasic children, Tallal and Newcombe (1978) attribute the children's difficulty to "...speech sounds that incorporate rapidly changing acoustic spectra" (p. 13). Thus, without any new evidence, they adopt the very interpretation that Tallal and Piercy (1975) tested and rejected. Tallal and Newcombe (1978) do not acknowledge this reversal, and so can hardly justify, or even explain, it. We can infer the underlying rationale, however, from their equation of discrete pitch changes in the tone TOJ test with the continuously changing spectral sweeps of stop-vowel formant transitions. The equation itself, also not explicitly acknowledged, we infer from several passing remarks. First, the authors characterize the difficulties of Tallal and Piercy's (1973) dysphasic children with tone TOJ as "...imperception of rapidly changing nonverbal acoustic material" (Tallal & Newcombe, 1978, p. 13). Second, they state that their adult subjects with left hemisphere damage were "...impaired in their ability to respond correctly to rapidly changing acoustic stimuli, regardless of whether stimuli were verbal or nonverbal" (p. 13). Yet the only verbal stimuli with which the patients had difficulty were stop consonant-vowel syllables, while the only non-verbal stimuli with which they were tested were the discrete tones of the tone TOJ test.

Finally, we infer the equation of tones and transitions from the claim, based on the performance of their aphasic subjects on the TOJ tests that: "...the left hemisphere must play a primary role in the analysis of specific rapidly changing acoustic cues, verbal and nonverbal, and...such analysis is critically involved in both the development and maintainance [*sic*] of language" (p. 19). Tallal & Newcombe support this claim by referring to Halperin, Nachshon and Carmon (1973). These authors asked listeners to label dichotic tone triads, each tone in a triad being either high(H) or low(L) (long or short in a second condition); they found that a left ear advantage for homogeneous triads (e.g., HHH or LLL) shifted increasingly to a right ear advantage as the number of "transitions" increased from one (e.g., HLL, HHL, etc.) to two (e.g., HLH, LHL). Halperin et al. (1973) concluded that "...perception of temporal patterns might be one of the underlying mechanisms in speech perception" (p. 46).

Perhaps Tallal and Newcombe were misled by the unhappy coincidence of the word "transition" being used to describe both Halperin et al.'s temporal patterns and the spectral sweeps at the onset of stop-vowel syllables. In any event, even if patterns of discrete pitch change, as in Tallal's or in Halperin et al.'s tone sequencing tasks, can properly be described as temporal, the continuous sweeps of stop-vowel transitions certainly cannot. To see this we must take a brief detour into the problem of coarticulation. Coarticulation refers to the overlapping articulation of two or more neighboring segments (consonants or vowels). A prototypical example is a consonant-vowel-consonant (CVC) syllable, such as the word *bag* (Liberman, 1970; see also Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967), in which so-called "perseveratory" effects of the initial /b/ and "anticipatory" effects of the final /g/ are distributed across the entire vowel. As a result, every portion of the syllable, both articulatorily and acoustically, carries information simultaneously about more than one segment. The syllable is then a unit of spatio-temporal interaction among articulators, and its integral acoustic form conveys information about successive segments in parallel rather than in series. Thus, the rapidly changing resonances of the vocal tract (formant transitions) at the onset of a stop-vowel syllable convey information about both consonant and vowel. Moreover, the transitions of synthetic /ba/ and /da/, used by Tallal, have identical temporal properties (duration, rate of frequency change); they differ in the loci and directions of their frequency trajectories. The contrast is therefore spectral, not temporal, and it is spectral, not temporal, sensitivity that perception of the contrast requires. (For textbook discussions of coarticulation and the problems it raises for speech perception, see Borden, Harris, & Raphael, 1994, Chapter 6; Pickett, 1980, Chapter 10).

Here then is the start of the conceptual confusion in the "temporal processing" hypothesis. The trouble begins when Tallal and Newcombe (1978) completely reverse, without evidence or explanation, the conclusions of Tallal and Piercy (1973, 1974, 1975). They do this (i) by equating "temporal perception" with rapid perception, and (ii) by attributing the dysphasic children's difficulties to the transitional character rather than the brevity of the formant transitions. They then compound the muddle by adopting such phrases as "rapidly changing acoustic...nformation" (Tallal & Newcombe, 1978, p. 19) to describe both the temporal patterns of discrete tone sequences and the continuous spectral sweeps of formant transitions. We turn now to see how Farmer and Klein, seemingly unaware of contradictions in the hypothesis they have undertaken to evaluate, handle the three issues that emerged from the work of Tallal and Piercy.

DEFICITS IN DISCRIMINATIVE CAPACITY, NOT IN TOJ

Farmer and Klein divide "sequential processing" into three components: Stimulus identification (Table 1), gap detection (Table 2), and TOJ (Table 3). Of these, only the second is necessarily temporal. Stimulus identification is, of course, prerequisite to TOJ, but would only itself be temporal if the defining property of the event to be identified were temporal. The stimulus identification studies selected for Table 1 reveal how confusion between tone sequences and formant transitions has ramified through the target paper: Farmer and Klein omit all studies of synthetic stop consonant continua because "...such phonemes are not regarded as single stimuli, but as a series of rapidly changing acoustic events" (p. 465). They do refer to some of these studies at a later point (p. 482) but, by omitting them from Table 1, Farmer and Klein lose an opportunity to focus attention on the single most important body of work concerning speech perception deficits in the reading-impaired, namely, some half dozen studies reporting less consistent identification along synthetic continua by poor or dyslexic readers than by controls (e.g., De Weirdt, 1988; Godfrey, Syrdal-Lasky, Millay, & Knox, 1981; Pallay, 1986; Reed, 1989; Steffens, Eilers, Gross-Glen & Jallad, 1992; Werker & Tees, 1987). In several of these studies where discrimination was also tested, impaired readers were significantly worse than normal controls between categories but not within, indicating that they could not easily exploit the phonological contrast that normally enhances discrimination across a phoneme boundary. Thus, their difficulties were in identifying and discriminating phonologically contrastive, but phonetically similar, synthetic syllables. Such results suggest that phonological categories may be less sharply defined in reading impaired than in normal children. Their omission from Table 1 impedes the reader's recognition of deficits in phonetic discriminative capacity that predict precisely the difficulties on /ba/ - /da/ TOJ observed by Reed (1989) in reading impaired children (Table 3), and that support the non-temporal account of TOJ deficits favored by Tallal herself (e.g., 1980) (Table 3). One further study, not listed by Farmer and Klein, also supports this account. Watson and Miller (1993) found that, although reading-disabled undergraduates made significantly more errors than normal readers on a test of non-speech tone TOJs three other tests, which (unlike tone TOJ) had clear face validity as measures of auditory temporal perception, did not distinguish between the groups. Table 3, then, lists apparent TOJ deficits among the reading-impaired as reported for non-speech tones by two studies (Reed, 1989; Tallal, 1980), for non-speech auditory clicks by one (Kinsbourne et al., 1991) and for speech sounds (/ba/ - /da/) by one (Reed, 1989). All are consistent with the view that the deficits are in discrimination/identification, not in temporal perception itself.

For the rest, only one study (McCroskey & Kidder, 1980, listed in Table 2) out of the 20 studies (5 auditory, 15 visual) listed in Tables 1, 2 & 3, reports an unambiguous deficit in auditory temporal perception in reading-impaired subjects, and it is far from clear how this deficit relates to speech perception. Yet Farmer and Klein summarily conclude: "In short, there is compelling evidence in groups of dyslexics for a deficit in TOJs in the auditory domain" (p. 23). This statement grossly misrepresents the facts.

PERCEIVING FORMANT TRANSITIONS

As we have seen, no evidence beyond the assertions of Tallal and Newcombe (1978) supports the claim that some impaired children have difficulty with the "rapid acoustic changes" of formant transitions. Nonetheless, the target paper offers a speculative account of how such a difficulty might arise. Farmer and Klein apparently accept without question that transitions are equivalent to tone sequences, and that their perception is a matter of temporal order judgment. Thus, referring to the patterns along a synthetic stop consonant continuum, they write, as already quoted: "...such phonemes are not regarded as single stimuli, but as a series of rapidly changing acoustic events (the spectral changes of formant transitions)" (p. 14), and later, to explain how a deficit in sequencing ability might affect speech perception: "The stop consonants involve spectral changes in the time frame of tens of milliseconds, and any impairment in the ability to process *the order of these changes* would result in impaired discrimination of the sounds" (p. 20, our italics). These statements include at least three points of misunderstanding.

First, a CV formant transition is not "...a series of rapidly changing acoustic events", but an integral spectral sweep reflecting the continuously changing resonances of the vocal tract, as a speaker moves from a point of closure into the following vowel. Second, as many experiments have shown (e.g., Mattingly, Liberman, Syrdal, & Halwes, 1971; Mann & Liberman, 1983), a brief formant transition removed from the speech signal is heard as a rapid, integral glissando, or "chirp", of which the parts or "spectral changes" cannot be perceptually "individuated", as Farmer and Klein's own account of TOJ would require. Third, even if a temporal order error in perception of a transition were possible, the resulting percept, since the transition begins with consonant release and ends in the vowel nucleus, would presumably reverse these segments, yielding /ab/ for /ba/ or /ad/ for /da/. Studies of speech errors never report such within-syllable metatheses in either production or perception. In short, Farmer and Klein's notion that a deficit in TOJ capacity would cause a failure to perceive formant transitions is untenable.

THE RELATION BETWEEN DEFICITS IN SPEECH AND NON-SPEECH AUDITORY PERCEPTION

Hemispheric specialization

The claim that phonological deficits are auditory in origin (Farmer and Klein, 1995, pp. 480-483) would be encouraged by evidence that the well-established specialization of the left hemisphere for speech perception is grounded in a prior specialization for aspects of auditory perception essential to speech. We now briefly review claims for such evidence from dichotic and aphasic studies.

Dichotic studies. A key paper, cited by Farmer and Klein, comes from Schwartz and Tallal (1980). In this paper a right ear advantage (REA) for synthetic stop consonant-vowel syllables was significantly reduced when the initial transitions were lengthened from 40 to 80 msec. Farmer and Klein follow the authors in interpreting this result as evidence for left hemisphere dominance in processing rapidly changing acoustic events. However, two conditions are necessary for an ear advantage: (i) hemispheric specialization, (ii) fuller access to the specialized hemisphere from the contralateral than from the ipsilateral ear (Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970, 1981). Variations in the magnitude of the REA within or between different classes of speech sound are therefore ambiguous: Are they due to differences in degree of hemispheric specialization (presumably, a more or less stable property of brain structure and function) or to differences in contralateral access (at least in part, a variable aspect of perceptual function)? Many studies have shown that the size and even direction of an ear advantage can be manipulated experimentally (see Studdert-Kennedy & Shankweiler (1981) for references). The most straight forward interpretation of Schwartz and Tallal (1980) therefore is not that they reduced the degree of left hemisphere engagement by increasing the duration of the onset transitions, but that they simply raised the salience of the ipsilateral signal, and so reduced the ear advantage. The latter is the more plausible interpretation because we do not then have to suppose that in normal listening the left hemisphere is more engaged by some portions of the speech signal than by others, or that its degree of engagement varies with the duration or intensity of the signal.

Non-speech dichotic studies purporting to show specialization of the left hemisphere for temporal processing are no less ambiguous, often because we cannot rule out covert verbal mediation (e.g., Halperin et al. (1973), cited above). The only dichotic study cited by Farmer and Klein as "...evidence of an REA for the processing of temporally complex non-speech sounds" (p. 481) was Divenyi and Efron's (1979) which actually used 100 msec, steady-state pure tones; they were judged for pitch and yielded a LEA in five out of six subjects. Finally, the notion that specialization of the left hemisphere, as indicated by dichotic studies, rests on a capacity for processing a particular type of physical stimulus is belied by studies in which identical stimuli give rise to different ear advantages as a function of their status in the listeners' language (e.g., Avery & Best, 1994; van Lancker & Fromkin, 1973).

Aphasic studies. In the study cited above, Tallal and Newcombe (1978) undertook to determine whether TOJ deficits of the kind observed in dysphasic children by Tallal and Piercy

(1973, 1974, 1975) were associated with left or right hemisphere lesions in adults. They found that : (1) left hemisphere patients were significantly worse than right hemisphere patients or normal controls on rapid TOJ for tones and synthetic /ba/ - /da/, but not on long, steady-state vowels; (2) while only 3 out of 10 left hemisphere patients could identify /ba/ - /da/ with 40 msec transitions, 7 out of 10 could do so when the syllables had 80 msec transitions; (iii) for left hemisphere patients rank order correlation between errors on tone TOJ and on a test of language comprehension was highly significant. Tallal and Newcombe (1978) therefore concluded, as already quoted, that: "...the left hemisphere must play a primary role in the analysis of specific rapidly changing acoustic cues, verbal and nonverbal, and...such analysis is critically involved in both the development and maintainance [*sic*] of language" (p. 19).

Yet, as we have seen, the supposed deficit "...in the analysis of specific rapidly changing acoustic cues, verbal and non-verbal" is based on unwarranted equation of tone sequences with formant transitions. Moreover, attempts to replicate the effect of lengthened transitions with adult aphasics have not been successful (Blumstein, Tartter, Nigro, & Statlender, 1984; Riedel & Studdert-Kennedy, 1985; cf. footnote 1, above). Finally, as often in aphasic studies, interpretation of the correlation is unsure: Was the deficit underlying difficulties with TOJ a cause or a consequence of the language deficit? In well-known related studies, Efron (1963) and Swisher and Hirsh (1972) observed non-speech TOJ deficits in aphasics, but explicitly rejected causal interpretations.

Stronger than these arguments, however, are the results of an experimental test of Tallal and Newcombe's (1978) claims by Aram and Ekelman (1988). They compared 20 left- and 12 right-brain lesioned children on Tallal's tone discrimination and TOJ tests. The test materials were prepared in Tallal's laboratory and testers were trained in the procedures of test administration by Tallal and her staff. Yet Aram and Ekelman found no differences between the lesioned children and normal controls or between the left- and right-lesioned children. Nor did they find any relation between the performances of the left-lesioned children on the tone tasks and various language tasks. They concluded that "...the higher level language deficits seen in left brain lesioned children cannot be attributed to difficulty in more preliminary analyses of the acoustic stimuli" (p. 935).

Farmer and Klein dismiss this work in a footnote on the grounds that the children represented "...quite a different population from the developmentally language impaired children studied by Tallal" (note 5, p. 492). So, of course, did the adult aphasics of Tallal and Newcombe (1978); yet those authors did not hesitate to generalize their findings to the children of Tallal and Piercy (1973, 1974, 1975). Moreover, if the left hemisphere does indeed "play a critical role" in temporal analysis, if such analysis is indeed "critical to the development and maintenance of language", and if Tallal's tests do indeed measure this left hemisphere capacity, we would surely expect that left-lesioned children with serious language deficits would have difficulty with those tests. Farmer and Klein's disclaimer therefore strikes us as less than compelling.

A Control Study

Conspicuously absent from Tallal's work, in the nearly 20 years that have elapsed since Tallal and Newcombe (1978) first made the claim, is any attempt to test the auditory basis of the supposed deficit in the perception of formant transitions by means of an appropriate non-speech control. The required control has come from Mody (1993; Mody, Studdert-Kennedy, & Brady, in press). Her reading-impaired subjects were 20 second-grade children, reading at least five months below grade level, and selected for their significant number of errors on Tallal's /ba/ - /da/ TOJ task. In Experiment 1a subjects were tested on discrimination and TOJ of synthetic /ba/ - /da/ at short ISIs: Errors increased monotonically as ISI decreased on both discrimination and TOJ, with no significant difference between tasks. In Experiment 1b, the same procedure was followed with syllable pairs /ba/ - /sa/ for one half of the group, and /da/ - /ʃa/ for the other half of the group. They made almost no errors on either task. Thus, despite their difficulties with /ba/ - /da/, they performed perfectly under time pressure, when they could clearly identify the syllables to be ordered, a result consistent with Tallal's (1980) view of TOJ. Evidently /ba/ - /da/ were difficult to discriminate and identify in rapid succession because they are very similar.

Is the similarity of /ba/ and /da/ auditory or phonetic? Experiment 2 was designed to answer this question. Non-speech control stimuli were synthesized. They consisted of two sine waves with durations and frequency trajectories identical to those of the center frequencies of F2 and F3 that carried the /ba/ - /da/ contrast in Experiment 1. Perceptually, the stimuli did not resemble their speech models. After preliminary identification training, subjects were tested for discrimination at the short ISIs of Experiment 1. Contrary to the results for /ba/ - /da/, performance was completely unaffected by decreases in ISI.

The combined results of the two experiments demonstrate that the reading-impaired children's difficulties with /ba/ - /da/ were due neither to a general deficit in rate of auditory processing nor to difficulties in processing brief patterns of rapidly changing acoustic information, but rather to difficulties in identifying similar syllables, presented in rapid succession. Since the non-speech control patterns of Experiment 2 were as acoustically similar as the /ba/ and /da/ of Experiment 1, it would seem that /ba/ and /da/ are difficult to identify at rapid presentation rates because, although phonologically contrastive, they are phonetically similar: Like speech sounds at the poles of a synthetic continuum which, as mentioned above, impaired readers often cannot readily identify, they differ on a single phonetic feature.

We cannot evade the results of this study (as Farmer and Klein hoped to evade the results of Aram and Ekelman (1988)) by arguing that the sampled population somehow differed from the population sampled in other studies. To be sure, the poor readers of this study were not dyslexic, or even as impaired as the poor readers of Tallal (1980). They did display, however, precisely the difficulties with /ba/ - /da/ TOJ that Reed (1989) observed in some reading-impaired children and that Tallal (1980) proposed as symptomatic of a phonological disorder in such children. Yet the source of those difficulties in the children of Mody (1993) was definitely not an inability to perceive brief formant transitions. Unless we are willing to suppose that perceptual difficulties in readers who read half a year behind grade level (Mody, 1993) have different causes than identical difficulties attributed to readers who read a full year behind grade level (Tallal, 1980), we have to concede that, difficulties in identifying /ba/ - /da/ at rapid rates of presentation have nothing to do with the transitional properties of their onsets (ironically, the very conclusion of Tallal and Piercy (1975)!) and are phonetic, not auditory, in origin.

CONCLUSION

The hypothesis that impaired readers' phonological deficits stem from a left hemisphere deficit in auditory temporal perception rests on conceptual confusion between temporal perception and rapid perception, and on misinterpretation of data from dichotic experiments and aphasic studies. The difficulties of some impaired readers with rapid temporal order judgments in speech and/or nonspeech seem to reflect independent deficits in discriminative capacity of unknown origin, not a general deficit in either "temporal processing" or rate of auditory perception. We conclude that, on the available evidence, the phonological deficit of impaired readers cannot be traced to any co-occurring non-speech deficits so far observed, and is phonetic in origin, but that its full nature, origin and extent remain to be determined.

REFERENCES

Aram, D. M., & Ekelman, B. L. (1988). Auditory temporal perception of children with left or right brain lesions. *Neuropsychologia, 26,* 931935.

Avery, R. A., & Best, C. T. (1994). Hemispheric differences in the perception of Zulu click consonants. *Journal of the Acoustical Society of America, 96,* 3230.

Blumstein, S., Tartter, V., Nigro & Statlender, S. (1984). Acoustic cues for the perception of place of articulation in aphasia. *Brain and Language, 22,* 128-149.

Borden, G., Harris, K. S., & Raphael, L. J. (1994). *Speech Science Primer* (3rd ed.). Baltimore: Williams & Wilkins.

De Weirdt, W. (1988). Speech perception and frequency discrimination in good and poor readers. *Applied Psycholinguistics, 16,* 163-183.

Divenyi, P. L. & Efron, R. (1979). Spectral versus temporal features in dichotic listening. *Brain & Language, 7,* 375-386.

Efron, R. (1963). Temporal perception, aphasia and déjà vu. *Brain, 86,* 403424.

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review, 2*, 460-493.

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology, 32*, 401-424.

Halperin, G., Nachshon, I. & Carmon, A. (1973). Shift of ear superiority in dichotic listening to temporally patterned nonverbal stimuli. *Journal of the Acoustical Society of America, 53*, 46-50.

Kinsbourne, M., Rufo, D. T., Gamzu, E., Palmer, R. L., & Berliner, A. K. (1991). Neuropsychological deficits in adults with dyslexia. *Developmental Medicine and Child Neurology, 33*, 763-775.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology, 1*, 301-323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431461.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14*, 211-235.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and non-speech modes. *Cognitive Psychology, 2*, 419-430.

McCroskey, R. L., & Kidder, H. C. (1980). Auditory fusion among learning disabled, reading disabled, and normal children. *Journal of Learning Disabilities, 13*, 18-25.

Mody, M. (1993). *Bases of reading impairment in speech perception: A deficit in rate of auditory processing or in phonological coding?* Unpublished doctoral dissertation, City University of New York.

Mody, M., Studdert-Kennedy, M. & Brady, S. (in press). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology.*

Pallay, S. L. (1986). *Speech perception in dyslexic children.* Unpublished doctoral dissertation, City University of New York.

Pickett, J. M. (1980). *The sounds of speech communication.* Baltimore: University Park Press.

Reed, M .A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology, 48*, 270-292.

Riedel, K., & Studdert-Kennedy, M. (1985). Extending formant transitions may not improve aphasics' perception of stop consonant place of articulation. *Brain and Language, 24*, 223-232.

Schwartz, J., & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science, 207*, 13801381.

Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental psychology, 19*, 59-63.

Steffens, M. L., Eilers, R., Gross-Glen, K., & Jallad, B. (1992). Speech perception deficits in adult subjects with familial dyslexia. *Journal of Speech and Hearing Research, 35*, 192-200.

Studdert-Kennedy, M., & Shankweiler, D. P. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America, 48*, 579-594.

Studdert-Kennedy, M., & Shankweiler, D. (1981). Hemispheric specialization for language processes. *Science, 211*, 960-961.

Swisher, L. & Hirsh, I. J. (1972). Brain damage and the ordering of two temporally successive stimuli. *Neuropsychologia, 10*, 137-152.

Tallal, P. (1980). Auditory temporal perception, phonics and reading disabilities in children. *Brain and Language, 9*, 182-198.

Tallal, P. (1984). Temporal or phonetic processing deficit in dyslexia? That is the question. *Applied Psycholinguistics, 5*, 167-169.

Tallal, P., & Newcombe, F. (1978). Impairment of auditory perception and language comprehension in dysphasia. *Brain and Language, 5*, 13-24.

Tallal, P., &Piercy, M. (1973). Developmental aphasia: Impaired rate of nonverbal processing as a function of sensory modality. *Neuropsychologia, 11*, 389-398.

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia, 12*, 83-93.

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia, 13*, 69-74.

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. In P. Tallal, A. M. Galaburda, R. R. Llinas, & C. von Euler, (eds.), *Temporal information processing in the nervous system* (Vol. 82, pp. 27-47). New York: Annals of the New York Academy of Sciences.

Tallal, P., Sainburg, R. L., & Jernigan, T. (1991). The neuropathology of developmental dysphasia: Behavioral, morphological and physiological evidence for a pervasive temporal processing disorder. *Reading and Writing: An Interdisciplinary Journal, 3*, 363-377.

van Lancker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and "tone": Evidence from Thai. *Journal of Phonetics, 1,* 101-109.

Watson, B. U., & Miller, T. K. (1993). Auditory perception, phonological processing and reading ability/disability. *Journal of Speech and Hearing Research, 36,* 850-863.

Werker, J., & Tees, R. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41,* 48-61.

FOOTNOTES

*Published in *Psychonomic Bulletin & Review,* (1995), *2,* 508-514.

† Albert Einstein College of Medicine, New York. Now at ENST, Dept. Signal, Paris, France.

[1] Lengthening formant transitions from 30 ms to 80 ms shifts the phonetic quality of the phonological contrast from stop to glide (Borden, Harris, & Raphael, 1994, Ch. 6; Pickett, 1980, Ch. 10; cf. Riedel & Studdert-Kennedy, 1985), and is therefore not purely auditory in its effect. No one, so far as we know, has replicated the findings of Tallal and Piercy (1975) with specifically reading-impaired children.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 35-38

Lengthened Formant Transitions are Irrelevant to the Improvement of Speech and Language Impairments*

M. Studdert-Kennedy, A. M. Liberman, S. A. Brady,† A. E. Fowler,‡
M. Mody,†† and D. P. Shankweiler‡‡

No experiment has ever supported the claim of Merzenich, Jenkins, Johnston, Schreiner, Miller, and Tallal (1996) and of Tallal, Miller, Bedi, Byma, Wang, Nagarajan, Schreiner, Jenkins, & Merzenich (1996) that the difficulty some impaired children and adults have in discriminating pairs of stop-vowel syllables is due to the inability of their auditory systems to cope with the rapid frequency changes (formant transitions) on which the contrast depends.[1] None of the studies cited in support of the claim (Tallal & Piercy, 1973; Tallal & Newcombe, 1978)[2] tested to see if the difficulty extended to acoustically matched non-speech controls. We cannot tell therefore whether the difficulty lay in how the acoustic information was analyzed, or in how the language system used it to form a phonetic representation. Yet exactly the latter process, not the former, was at fault in monolingual speakers of Japanese who were unable to discriminate synthetic /ra/ and /la/ that differed only in formant transitions (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975); and many other studies have confirmed that performance on discrimination of a transition depends on whether or not it cues a phonetic percept.[3] A recent and directly relevant study found that poor readers who had difficulty discriminating rapidly presented synthetic /ba/-/da/ had no comparable difficulty with acoustically matched non-speech patterns; moreover, their speech difficulty arose from the close phonetic similarity of the consonants, not from the presence in both of rapid formant transitions (Mody, 1993; Mody, Studdert-Kennedy, & Brady, in press).

Inadequate controls also render the experiments of Merzenich et al. (1996) and of Tallal et al. (1996) uninterpretable, because they manipulated at least eight independent variables, but provided only one control group.[4] Before we come to this, however, we must clarify a misunderstanding of "temporal processing" in speech perception, evident in the authors' conflation of perceiving rapid sequences of discrete tones (or syllables) with perceiving the rapid spectral sweeps of formant transitions.[5] Difficulty with the former they term a "...basic temporal segmentation deficit [that] may disrupt the normal sharpening of ...phonetic prototypes" (Tallal et al., 1996, p.82). Yet we know that coarticulation, the mechanism assuring normal rates of speaking, causes acoustic information for each segment (consonant or vowel) to overlap extensively with information for neighboring segments (Joos, 1948; Liberman, 1970; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). An utterance is not therefore a string of discrete alphabetic units, and its perception does not require identifying the order of brief "events...within...tens of milliseconds" (Tallal et al., 1996, p.82). What it does require, rather, is sensitivity to how coarticulation marks serial position by variations in acoustic shape, making the transitions for the stop consonants in /bi/ and /ib/, for example, into mirror images, and causing them to carry information simultaneously about consonant, vowel and their order (Joos, 1948; Liberman, 1970; Liberman et al., 1967; Gottfried & Strange, 1980; Jenkins, Strange, & Edman, 1983; Rakerd & Verbrugge, 1987; Strange, 1989; Strange, Verbrugge, Shankweiler, & Edman, 1976). The relevance to speech perception of temporal thresholds (or "processing" speed) for perception of discrete sounds is therefore highly questionable, and is rendered even more so by evidence that dyslexics' slow reaction time in discriminating non-speech tones is independent of their phonological deficits (Nicholson & Fawcett, 1994). In this connection, the correlations between change in temporal threshold and post-training language performance reported by Tallal et al. (Tallal et al., 1996, p.82, Figure 3), are uninterpretable, because neither pretraining language performance nor general intelligence was partialled out.[6]

As for perception of rapid spectral sweeps in formant transitions, the authors (Merzenich et al., 1996; Tallal et al., 1996) nowhere acknowledge that differences in transition duration cue the distinction between stop and semivowel (Liberman, Delattre, Gerstman, & Cooper, 1956). If some children do indeed hear syllables with lengthened transitions as stops rather than semivowels, we would expect them to confuse *we* with *be*, *you* with *do* or *goo*, and so on. The authors offer no evidence for such confusions, nor do they demonstrate that whatever difficulties the children may have had in discriminating rapidly presented /ba/-/da/ were due to difficulty in perceiving formant transitions rather than to, say, the two syllables' close phonetic similarity. The apparent effectiveness of training, with interstimulus interval (ISI) and extended/amplified transitions as adaptive parameters (Merzenich et al., 1996), is not evidence for difficulty with transitions, because, without a control group trained on normal syllables with ISI as the only adaptive parameter, we have no evidence that manipulation of the transitions had any effect at all.

Moreover, because there is no control to separate the effects of adaptive training with synthetic speech (Merzenich et al., 1996) from those of the modified natural speech (Tallal et al., 1996), nor the effects of extending and amplifying transitions from the overall slowing of speech, we cannot exclude the possibility that the language gains of the experimental group simply arose from intensive exposure to natural speech, slowed not specifically in its formant transitions, but in the overall rate at which words were delivered. Language impaired children may well find slowed speech easier to perceive than normal speech, not because the auditory system does not then have to contend with rapid transitions or with discrete sounds that follow each other within tens of milliseconds, but rather because the reduced rate allows more time for the disabled language system to form phonetic representations.

Other and simpler ways of training a child's attention to speech are already available, however. Significant improvements in discrimination of confusable phonetic units,[7] as well as in other phonological skills more directly related to reading,[8] have been achieved in reading-impaired subjects by various training procedures, without recourse to the acoustic manipulations contrived by Merzenich, Tallal and their colleagues. The issue is not whether we can help language-impaired and reading-impaired children, for we surely can, but rather how accurately we have identified the underlying impairment. Only when we have got the science right can we expect to design the best treatment.

REFERENCES

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26(1)*, 49-66.

Bentin, S., & Mann, V. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America, 88*, 64-74.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics, 29*, 191-211.

Best, C. T., Studdert-Kennedy, M., Manuel, S., & Rubin-Spitz, J. (1989). Discovering phonetic coherence in auditory patterns. *Perception and Psychophysics, 45*, 237-250.

Blumstein, S., Tartter, V., Nigro & Statlender, S. (1984). Acoustic cues for the perception of place of articulation in aphasia. *Brain and Language, 22*, 128-149.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301*, 419-421.

Brady, S., Fowler, A., Stone, B., & Winbury, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. *Annals of Dyslexia, 44*, 26-59.

Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phoneme awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*(10), 4-11.

Fowler, A., Brady, S., Yehuda, N. (1994). *The role of perception in achieving phoneme awareness.* Paper read at 46th Annual Conference of the Orton Dyslexia Society, Houston.

Gottfried, T. L., & Strange, W. (1980). Identification of coarticulated vowels. *Journal of the Acoustical Society of America, 68*, 1626-1635.

Grunke, M. E., & Pisoni, D. B. (1982). Perceptual learning of mirror-image acoustic patterns. *Perception and Psychophysics, 31*, 210-218.

Hurford, D. P. (1990). Training phoneme segmentation ability with a phonemic discrimination intervention in second- and third-grade children with learning disabilities. *Journal of Learning Disability, 23*, 564 -569.

Hurford, D. P., Johnston, M., Nepote, P., Hampton, S., Moore, S., Neal, J., Mueller, A., McGeorge, K., Huff, L., Awad, A., Tatro, C., Juliano, C., Huffman, D. (1994). Early identification and remediation of phonological processing deficits in first-grade children at risk for reading disabilities. *Journal of Learning Disability, 27*, 647-659.

Hurford, D. P., & Sanders, R. E. (1990). Assessment and remediation of a phonemic discrimination deficit in reading disabled second and fourth graders. *Journal of Experimental Child Psychology, 50*, 396-415.

Jenkins, J. J., Strange, W., & Edman, T. R. (1983). Identification of vowels in 'voiceless' syllables. *Perception and Psychophysics, 34*, 441-450.

Joos, M. (1948). Acoustic phonetics. *Language, 24(2).*

Jorm, A. F., Share, D. L., Maclean, R., & Matthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology and Psychiatry, 27*, 45-55.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology, 1*, 301-323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology, 52*, 127-137.

Lundberg, I., Frost, J., & Peterson, O.-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23*, 263-284.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14*, 211-235.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 272*, 77-80.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, 18*, 331-340.

Mody, M. (1993). *Bases of reading impairment in speech perception: A deficit in rate of auditory processing or in phonological coding?* Unpublished doctoral dissertation, City University of New York.

Mody, M., Studdert-Kennedy, M. & Brady, S. (in press). *Speech perception deficits in poor readers: Auditory processing or phonological coding? Journal of Experimental Child Psychology.*

Nicholson, R. I., & Fawcett, A. J. (1994). Reaction times and dyslexia. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47A*, 29-48.

Nygaard, L. C., & Eimas, P. D. (1990). A new version of duplex perception: Evidence for phonetic and nonphonetic fusion, *88(1)*, 75-86.

Olson, R. K. Wise, B. W. (1992). Reading on the computer with orthographic and speech feedback. *Reading and Writing, 4*, 107-144.

Pickett, J. M. (1980). *The sounds of speech communication: A primer of acoustic phonetics and speech perception.* Baltimore: University Park Press.

Rakerd, B., & Verbrugge, R. (1987). Evidence that the dynamic information in vowels is talker independent in form. *Journal of Memory and Language, 26*, 558-563.

Riedel, K., & Studdert-Kennedy, M. (1985). Extending formant transitions may not improve aphasics' perception of stop consonant place of articulation, *24*, 223-232.

Riedel, K., & Studdert-Kennedy, M. (1985). Extending formant transitions may not improve aphasics' perception of stop consonant place of articulation. *Brain & Language, 24*, 223-232.

Strange, W. (1989). Evolving theories of vowel perception. *Journal of Acoustical Society of America, 85*, 2081-2087.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America, 60*, 213-224.

Studdert-Kennedy, M., & Mody, M. (1995). Auditory temporal perception deficits in the reading-impaired: A critical review of the evidence. *Psychonomic Bulletin, 2*, 508-514.

Tallal, P., & Piercy, M. (1973). Developmental aphasia: Impaired rate of nonverbal processing as a function of sensory modality. *Neuropsychologia, 11*, 389-398.

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia, 12*, 83-93.

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia, 13*, 69-74.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 272*, 81-84

Tallal, P., & Newcombe, F. (1978). Impairment of auditory perception and language comprehension in dysphasia. *Brain and Language, 5*, 13-24.

Tallal, P., & Stark, R. E. (1982). Perceptual/Motor profiles of reading impaired children with or without concomitant oral language deficits. *Annals of Dyslexia, 32*, 163-176.

Tallal, P., & Stark, R. E., & Mellits, E. D. (1985). Identification of language-impaired children on the basis of rapid perception and production skills. *Brain and Language, 25*, 314 -322.

Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America, 81*, 755-764.

Vorperian, H. K., Ochs, M. T., Grantham, D. W. (1995). Stimulus intensity and fundamental frequency effects on duplex perception. *Journal of the Acoustical Society of America, 98*, 734-744.

Whalen, D., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237*, 169-171.

Wise, B. W., & Olson, R. K. (1992). Spelling exploration with a talking computer improves phonological coding. *Reading and Writing, 4*, 145-156.

FOOTNOTES

*This paper was submitted to *Science* as a Technical Comment, but was rejected for unspecified reasons.

†Also at the University of Rhode Island.

‡Also Wesleyan University.

††Albert Einstein College of Medicine, New York. Now at ENST, Dept. Signal, Paris, France.

‡‡Also University of Connecticut, Storrs.

1 The claim that rapid frequency changes cause difficulty contradicts the conclusion of Tallal and Piercy (1975) that: "...it is the brevity not the transitional character...[of the formant transitions] of synthesized stop consonants which results in the impaired perception of our dysphasic children" (p. 73). For an account of the origin of the claim, despite its lack of experimental support, see Studdert-Kennedy & Mody, 1995.

2 For failures to replicate Tallal, & Newcombe, 1978, see Blumstein, Tartter, Nigro, & Statlender, 1984; Riedel & Studdert-Kennedy, 1985.

3 See, for example, Bentin & Mann, 1990; Best, Morrongiello, & Robson, 1981; Best, Studdert-Kennedy, Manuel, & Spitz, 1989; Grunke & Pisoni, 1982; Nygaard & Eimas, 1990; Mann & Liberman, 1983; Vorperian, Ochs, & Grantham, 1995; Whalen & Liberman, 1987. For auditory-phonetic divergence in perception of steady-state sounds, see Tomiak, Mullenix, & Sawusch, 1987.

4 The exact number of experimental variables depends on how we count parameter repetitions under the main conditions. Experimental subjects were exposed to: (1) modified natural speech, (2) adaptive training in "temporal processing". Natural speech was modified by: (1) overall prolongation, (2) differential amplification of transitions. Adaptive training was conducted with: (1) non-speech frequency sweeps, (2) syllable pairs contrasting in stop consonants. Adaptive parameters for the non-speech were: (1) interstimulus interval (ISI), (2) frequency range. Adaptive parameters for the syllables were: (1) ISI, (2) transition duration, (3) differential amplification of transitions.

5 For the origin of the conflation of discrete tone sequences with formant transitions in Tallal's work, see Studdert-Kennedy & Mody, 1995.

6 For the role of intelligence in judgments of temporal order sequence, see Jorm, Share, MacLean, and Mathews (1986).

7 See, for example: Fowler, Brady, & Yehuda (1994); Hurford (1990); Hurford & Sanders (1990); Hurford, Johnston, Nepote, Hampton, Moore, Neal, McGeorge, Huff, Awad, Tatro, Juliano, & Huffman (1994).

8 See, for example: Ball & Blachman (1991); Bradley & Bryant (1983); Brady, Fowler, Stone, & Winbury (1994); Byrne & Fielding-Barnsley, (1993); Lundberg, Frost, & Peterson (1988); Olson & Wise (1992); Wise & Olson, (1992).

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 39-57

Phonological Awareness in Illiterates: Observations from Serbo-Croatian*

Katerina Lukatela,† Claudia Carello,† Donald Shankweiler,† and Isabelle Y. Liberman

Adult illiterate and semi-literate speakers of Serbo-Croatian were assessed on reading, writing, phonological, and control tasks. Most subjects had acquired measurable literacy skills despite a documented lack of formal instruction. Individual differences in these skills were highly specific. They were related to measures of phoneme segmentation and alphabet knowledge, but only weakly related to general cognitive abilities. Three groups, categorized with respect to subjects' ability to identify the letters of their Cyrillic alphabet, differed on phoneme deletion and phoneme counting tasks but not on syllable counting, picture vocabulary, or tone counting tasks. Alphabet knowledge was more tightly coupled with phoneme awareness than has been found in speakers of English. Cross-language similarities and differences were discussed, highlighting the role that phonological transparency of the orthography may play in the acquisition of literacy.

It has long been argued that in order to master reading and writing in an alphabetic system, the learner must become aware that words come apart into sequences of phonemes (Liberman, 1973). Lacking that insight, a beginner would be unable to grasp the alphabetic principle. A considerable body of research has borne out that a new learner's ability to segment words phonologically is the most powerful predictor of future reading and spelling skills (Lundberg, Frost, & Petersen, 1988; Mann & Liberman, 1984; Stanovich, Cunningham, & Cramer, 1984). There is evidence that apprehension of phoneme segmentation is not particularly easy for young children to grasp or, indeed, for illiterates of any age. Experience with speech is ordinarily not sufficient to make one conscious of the phonemic structure of words. Other aspects of phonological structure (syllables, onsets and rimes, rhyme and alliteration) present lesser difficulties. Liberman and her colleagues have suggested that the physiology of speech itself is one reason for the special difficulty of segmentation by phoneme (Liberman, Shankweiler, Fischer, & Carter, 1974). Because speech is highly coarticulated, the speech signal is quasicontinuous; it is not discretely partitioned into the phonemes we represent alphabetically in writing. If this circumstance does contribute to the difficulty of achieving phonemic awareness, it would cut across all languages.

Other factors may act to modulate the difficulty of achieving full awareness of phoneme segmentation, and these may vary with particular characteristics of the language and orthography. Languages are known to differ on phonological characteristics such as the complexity and variety of syllable types (DeFrancis, 1989; Mattingly, 1992). They also differ in how transparently and consistently the orthographic representations map onto the phonology

Preparation of this manuscript was supported in part by National Institutes of Child Health and Human Development Grants HD-01994 and HD-08495 to Haskins Laboratories and the University of Belgrade, respectively.

Until the time of her death in July of 1990, Isabelle Y. Liberman contributed importantly to each phase of this research.

(Liberman, Liberman, Mattingly, & Shankweiler, 1980). In so-called phonologically shallow orthographies, constraints on sound sequences are the only sources of constraints on letter sequences. The orthography can be said to more directly represent the phonology. In phonologically deep orthographies, restrictions on letter sequences derive not only from phonological constraints but also from constraints relating to the etymology and morphology of the written language. Serbo-Croatian can be said to anchor one end of an orthographic depth continuum, closely allied with the Romance languages; Hebrew and logographic Chinese are at the other extreme; English is nearer the middle of the continuum. To the extent that such differences make a difference to the course and outcome of literacy instruction, all alphabetic systems are not equally learnable. In particular, learning to read a phonologically shallow orthography may be characterized by more rapid development of phonological awareness, and correspondingly rapid development of word decoding skills.

This article reports an experiment using adult illiterates and near-literates whose language is Serbo-Croatian. Given that each grapheme has only one phonemic interpretation, and there are no silent or double letters, the orthography-phonology link is explicit. Of concern are the segmentation abilities of Serbo-Croatian speakers and, in particular, with the question of how these abilities compare with those demonstrated by speakers of languages which may (or may not) differ on this dimension. Before considering the particulars of the experiment, we will marshall evidence that lead to clear predictions about the role of language environment in the development of reading skill.

The role of phonological awareness in reading acquisition

As remarked at the outset, grasp of the alphabetic principle would seem to require awareness of the segmental structure of speech and, in fact, some measures of segmental awareness are good predictors of future reading skill. A number of tasks have been used to evaluate segmental awareness. Subjects may be asked to count the number of segments in an utterance, reverse its segments, add a segment to the front, or delete one. They might be required to choose those utterances whose relevant segments match or differ. The segments in question have included syllables as well as phonemes. Somewhat related are tasks directed at rhyme sensitivity, in which subjects are asked to supply a rhyme for a target utterance, or to choose the rhyming members from a sequence of utterances. Phonological awareness is not of a piece: Tasks involving phonemes, syllables, and rhymes reflect different levels of phonological structure, with awareness of phoneme segmentation most closely related to reading skill.

Research in several language communities has found that children who cannot yet read have great difficulties with tasks that tap awareness of phonemes. Liberman and her colleagues demonstrated that syllable segmentation develops earlier than phoneme segmentation in English-speaking American children, and continues to be easier for the young child from preschool through first grade (Liberman et al., 1974). Moreover, early phoneme analysis skills are a better predictor of later reading achievement than are syllable analysis skills (Mann & Liberman, 1984). Italian children show two of these patterns: Syllable segmentation skills develop earlier than phoneme segmentation skills, and phoneme analysis skills are a better predictor of later reading achievement than are syllable analysis skills (Cossu, Shankweiler, Liberman, Katz, & Tola, 1988). Swedish and Danish children, too, show that phoneme analysis skills in kindergarten outstrip syllable analysis and rhyme production skills in correlating with first- and second-grade reading achievement (Lundberg et al., 1988; Lundberg, Olofsson, & Wall, 1980, reanalyzed in Wagner & Torgesen, 1987). A similar failure by rhyming tasks to predict reading success in English was possibly due to a limited range of item difficulty, given that the rhyme task was performed at ceiling (Stanovich et al. 1984; Yopp, 1988). Interestingly, even though the degree of difficulty of phoneme tasks varies widely, with phoneme deletion being especially difficult, their predictive power is more or less equivalent (Stanovich et al., 1984).

Preschool children, though unable to segment words by phoneme, are nonetheless aware of some aspects of phonological structure. In addition to the evidence that syllables are identified earlier than phonemes, there is also evidence that appreciation of rhyme and alliteration precedes the development of phoneme awareness. Moreover, it has been suggested that these coarser aspects of phonological awareness may play a facilitating role in children's grasp of the alphabetic principle. On the basis of such findings, the Oxford group maintains that appreciation of rhyme may be a preliminary stage to the development of full phonological awareness (i.e., phoneme awareness; Bryant, MacLean, & Bradley, 1990; Bryant, MacLean, Bradley, & Crossland, 1990).

There is evidence from training studies with English speakers that full phonological awareness is not particularly easy for children to come by even after some literacy instruction (Byrne, 1990). For reasons discussed at the beginning, we could expect some degree of difficulty in any alphabetic system. But, as we suggest, language differences may contribute to relative ease, both in terms of the depth at which the orthography represents the phonology and in the extent to which reading instruction emphasizes this link. Because of this it is unfortunate that so little comparable data exists in languages other than English. More rapid development of phonological awareness and more rapid progress in learning the code might be expected in a language that maintains a one-to-one correspondence between graphemes and phonological units. Support for this contention comes from a comparison of Italian and American children provided by Cossu et al. (1988): Italian children performed better than their American counterparts on phoneme tasks at the pre-school level and in each of the first two school years. Their superiority on syllable tasks also reflected language differences: Italian has fewer vowel distinctions, morphophonological alternations, and syllable types than English.

Comparisons of successful and unsuccessful readers

As was noted with respect to beginning reading, phoneme awareness tasks distinguish children who have acquired the alphabetic principle from those who have not. In addition, older children categorized as good or poor readers (on the basis of teachers' evaluations, reading achievement tests, latencies and errors in decoding pseudowords) are differentiated on tasks that seem to implicate phonological abilities. For example, over and above differences in IQ, good readers are better than poor readers at remembering both printed and spoken nonsense syllables, letter strings, and words (see Mann, 1984, for a review); they do not differ in memory for faces or nonsense drawings (Liberman, Mann, Shankweiler, & Werfelman, 1982).

Perhaps most telling are manipulations that hinder the success of good readers precisely because they are phonologically analytic. For example, recall of consonant strings with phonetically confusable names (e.g., CEGVZ) is much more difficult for good readers than recall of consonant strings with phonetically nonconfusable names (e.g., OFQYX), so much so that their mean number of errors approaches that of marginal and poor readers (who have difficulty with both types of strings; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). Similarly, although good readers make fewer errors on both meaningful and semantically anomalous sentences, if the constituent words are phonetically confusable, their errors rival those of poor readers (Mann, Liberman, & Shankweiler, 1980). A parallel result has been obtained in Serbo-Croatian: Skilled readers are hindered more by phonological ambiguity[1] than are less skilled readers (Feldman, Lukatela, & Turvey, 1985).

Phonological awareness in adult literacy

Evidence from a number of languages suggests that adults who cannot read an alphabetic orthography are unable to manipulate phonemes. Results of several studies by the Brussels-based group, who explored metalinguistic abilities in adult illiterates living in rural Portugal (Morais, Cary, Alegria, & Bertelson, 1979; Morais, Bertelson, Cary, & Alegria, 1986; Morais, Content, Bertelson, Cary, & Kolinsky, 1988), indicated that illiterates performed very poorly on tasks that assessed abilities to carry out analysis of spoken words into phonemes. For example, illiterate subjects could neither add a consonant to nor delete a consonant from the beginning of a nonsense word (Morais et al., 1979), nor could they segment speech into units smaller than the syllable (Morais et al., 1986). Illiterate subjects were able to delete syllables and detect rhymes but their performance was nonetheless inferior to that of ex-illiterates (those who had participated in a course of reading instruction as adults). A picture recall task revealed that although illiterates worked with smaller sets than ex-illiterates (i.e., showed poorer short-term retention), both groups showed poorer performance on rhyming sets, reflecting the use of speech related codes. Even when the ex-illiterates were separated into good (fast, fluent, error-free) and poor readers (slower, less fluent, with occasional errors), differences between illiterates and poor readers on most tests were large compared to differences between good and poor readers (Morais et al., 1986). An exception surfaced in the control task, which required the subjects to segment melodies (reproducing the last 3 notes of a 4-note sequence), where illiterates and poor readers were equivalent to each other though inferior to good readers.

Adult quasi-illiterate speakers of English are similarly impaired on phonemic segmentation tasks. Nine men enrolled in a community literacy class (who had some schooling and who reported serious difficulties in spelling) were able to achieve only a 58% success rate in phoneme segmentation tested by consonant deletion (Liberman, Rubin, Duquès, & Carlisle, 1985). This contrasts with 75% performance by 11 and 12 year old American school children, Rosner & Simon, 1971. A large sample of men of low literacy (fifth grade reading level or below), studied by Read and Ruyter (1985), performed very poorly on segmentation tasks. As has been found with children, the difficulties were greater with phonemes than with syllables: 39% correct on phoneme counting, 48% correct on phoneme addition, 77% correct on syllable counting. The poor level of success in reading nonwords (57%) was comparable to a group of fifth graders (taken from Richardson, DiBenedetto, & Adler, 1982) who were a year or more below grade level (58%), in contrast to fifth grade good readers (93%) (Read & Ruyter, 1985).

Corroboration of the view that experience with an alphabetic orthography facilitates the acquisition of phonemic segmentation comes from a study carried out in China with logographically-literate adults. The subjects were grouped according to whether or not they had received instruction in the alphabetic pinyin orthography (which, since 1958, has been taught for four weeks in the first grade; Read, Zhang, Nie, & Ding, 1984). One group had received pin yin instruction and a second group of older Chinese readers, though also literates in reading the traditional logograms, had received no instruction in the alphabetic principle. In phoneme addition or deletion tasks (similar to those used with Portuguese subjects), all 12 alphabetic subjects got at least 70% correct whereas only 2 of 18 nonalphabetic subjects did better than 55%. Unlike literate and illiterate adults, who differ in written language experience, both alphabetic and nonalphabetic subjects encounter written language daily. It can be assumed that their language experience is nearly comparable. These results suggest that alphabetic reading instruction is a critical factor in developing phonological segmentation abilities (or, perhaps, in preserving those abilities; see Mann, 1986).

Implications of literacy training

Taken together, the findings with beginning readers, poor readers, and (alphabetically) illiterate adults suggest that all find it difficult to penetrate the internal structure of words to recover their phonemic structure. These individuals are all language users but their experience with the spoken language, in itself, has not provided them with explicit conscious awareness of phonemic structure. Morais and colleagues concluded that the differences between illiterates and ex-illiterates were tied to instruction and experience in an alphabetic system and, quite specifically, to tasks involving phonemes rather than syllables, rhymes, or alliteration. While reading instruction brings about improvement in all segment and sound-based abilities, they argued that a phonemic analysis capability seems particularly reliant on the experience gained through specific instruction. (Most often this instruction occurs in the context of teaching to read, but it may be taught independently as Byrne & Fielding-Barnsley, 1991, in press, and Lundberg et al., 1988, have shown). Presumably, when instruction makes explicit the relationship between sounds and graphemes (grapheme-phoneme correspondence rules or connections), the beginning reader becomes aware of the phonological structure of utterances—the reader has grasped the alphabetic principle (Liberman, Shankweiler, & Liberman, 1989; Mattingly, 1972, 1980).

A number of writers, including Morais and his colleagues, have maintained that this is not a one-way relationship but a complex interaction: With acquisition of an alphabetic orthography there ensues a rapid development of phonological segmentation skills; reciprocally, an advanced level of phonological awareness improves reading and writing abilities (Ehri, 1984; Liberman et al., 1974). For present purposes, we choose to emphasize only that the rapidity of the development seems to depend crucially on language. That is to say, phonological segmentation skills should develop more thoroughly and earlier in languages whose orthography is phonologically precise.

The link between phonological awareness and literacy is supported by the failures of illiterate Portuguese adults and nonalphabetic Chinese readers on phoneme segmentation tasks. But it should be noted that their failure was not complete: In phoneme segmentation, 20% of the illiterates matched or exceeded the performance of 23% of the literates in Morais et al. (1979) and 62% of the poor readers in Morais et al. (1986). These more successful illiterates were those who had some early schooling or who had been taught to identify letters by their children; they performed somewhat better than those who had received no instruction at all (Morais et al., 1979). Similar anomalies can be found in the data from the readers of Chinese: Of the nonalphabetic subjects, 11% matched or exceeded the performance of 25% of the alphabetic subjects in Read et al. (1984). While it is possible that these levels of performance mean that segmentation skill can arise outside the context of alphabetic literacy training in some people, it is reasonable to suppose that the illiterates in question gained some degree of familiarity with sound-to-letter correspondence (perhaps through incidental reading instruction from their children or, in the case of the Chinese, in noticing the pinyin transcriptions of logographic signs that are provided for foreigners in Beijing). This interpretation would suggest that a rather minimal alphabetic exposure—not necessarily enough to read whole words—is sufficient to develop phonemic awareness in linguistically mature adults, given a shallow phonology and orthography. This possibility has conceptual and empirical implications. Conceptually, it suggests that the literacy-illiteracy distinction should be viewed as poles of a continuum rather than as distinct categories. Empirically, it implies the need for control of the factor of alphabetic familiarity to allow a clean evaluation of the link between phonological awareness and literacy.

Other empirical issues are also germane. In addition to an explicit assessment of adult illiterates' degree of alphabetic familiarity, the tendency for good readers to be superior to poor readers in all tasks demands an evaluation of subjects' general verbal abilities. It remains possible that differences between literate and illiterate groups might reflect more general differences in their capacity (due to pre-existing analytic skills, intellectual or motivational level) to benefit from literacy instruction or language experience. For example, ex-illiterates could be successful in the segmentation tasks not because they acquired literacy but because they had segmentation skills before they acquired literacy. On the other hand, illiterates (or low literacy adults who have attended literacy courses but have not attained the level of proficiency required to earn a certificate) may be unable to benefit from instruction because they lack even the most rudimentary awareness of phonemic segmentation. The comparison between illiterate and ex-illiterate subjects, in and of itself, therefore, does not allow firm conclusions about the relationship between literacy training and phonological segmentation abilities.

Expectations of adult illiterate speakers of Serbo-Croatian

In Serbo-Croatian the phonology of the spoken language is closely transcribed by its orthography (partly a result of spelling reform by V. Karadjić in the early 1800s). If we assume that phonemic awareness is stimulated by encounters with the alphabet then, as remarked, we could expect more rapid acquisition in a language that maintains a one-to-one correspondence between graphemes and phonological units. Indeed, comparisons of beginning readers of English and Italian bear this out. Like Serbo-Croatian, Italian is relatively shallow, with few morphophonological alternations. Italian children are better at both syllable and phoneme segmentation tasks than their English-speaking American counterparts, and they probably advance more rapidly in early reading acquisition (Cossu et al., 1988; Cossu, Shankweiler, Liberman, & Gugliotta, in press). Such differences suggest that Serbo-Croatian provides a revealing comparison with other languages that have a deeper orthography such as English or, to a lesser extent, Portuguese. The question is: Should the metalinguistic abilities of Serbo-Croatian speakers differ from other populations whose languages do not have such a straightforward relationship to their orthographies? The goal of the present study, then, was to explore metalinguistic word-segmentation abilities of Yugoslav adult illiterate subjects drawing comparisons, where possible, to beginning readers and illiterate adults from other linguistic environments.

Special care was taken in the selection and categorization of illiterate subjects, restricting individual characteristics such as age, profession, geographic setting, and cause of illiteracy. They were Yugoslavian adult speakers of Serbo-Croatian who had been exposed to minimal, if any, reading instruction. Although they differed in their familiarity with their Cyrillic alphabet, their daily routine did not include reading and, therefore, they did not "tune up" their phonological awareness. Subjects' analytic metalinguistic abilities were assessed with segmentation tasks similar to those used with Portuguese illiterates (Morais et al., 1979) and with American children (Liberman et al., 1974): syllable counting, phoneme counting, and deletion of the initial consonant. Because failure on the metalinguistic tasks could reflect a general analytic deficiency rather than a deficiency specifically for the phonological level, two non-linguistic control tasks were introduced: counting tones and counting sticks. Subjects' verbal IQ and short-term memory span (a limitation associated with poor readers) were also assessed.

There were three major expectations. First, in light of the intimate connection between literacy and phonemic awareness, truly illiterate adults should be unable to perform phonemically analytic tasks. Second, consistent with previous results with children and adults of low literacy, it was expected that syllable tasks would be easier than phoneme tasks. And third, given the nature of the letter-sound relationship in Serbo-Croatian, it was expected that familiarity with the alphabet would establish fairly secure phonemic awareness. The first prediction concerns what would constitute evidence for a close relationship between segmentation abilities and reading skill. The second prediction derives from our understanding of the syllable as a less abstract segment, closer to the basic unit of articulation. The final prediction is in contrast to what has been found with English-speaking adults of low literacy and children who are beginning readers, in which knowing the letter names, *per se*, did not contribute much to segmental skills.

Method

Subjects. The study was carried out in a rural area of Serbia which was, at that time, a republic of Yugoslavia. The subjects, 23 adult females between the ages of 55 and 76, were tested in three villages (Selevac, Krcedin, and Nova Pazova) within a 200 km radius of Belgrade. Although two alphabets are in use in Yugoslavia (see Footnote 1), in rural Serbia all printed matter (street and shop signs, packaging) is in Cyrillic. Exposure to the Roman alphabet would be considerably less.

All subjects were active farm workers; none of them had ever been employed outside the home. They were paid for participation in this study. Most of them were identified and introduced to the Experimenter by local educational personnel (teachers and school directors). No subject had ever attended a regular school. The reason for that in all cases was poverty, combined with a traditional attitude that girls do not benefit from schooling. Some of the subjects had, as adults, attended an obligatory elementary reading instruction course that was instituted throughout the whole of Yugoslavia after World War II. None of these subjects had completed the course, spending between several days and one month in attendance; in all cases, their drop-out was for economic reasons and reasons related to culturally-defined gender roles. From questioning each of the subjects, it was ascertained that they withdrew because they could not be spared from farm and household chores (in conjunction with the prevailing village attitude that women do not need to be literate). Nonetheless, some of the subjects had learned the alphabet, presumably with help from their children and grandchildren. Subjects who knew the alphabet are identified in all analyses.

Literacy Assessment

Literacy was assessed by testing each subject on her reading and writing abilities at three levels: letters, words, and sentences. First, each subject was presented all 30 letters from the Cyrillic alphabet to identify. If a subject could identify a majority of the letters, she was asked to read individual words, all of which were concrete, frequent nouns. To conclude the reading test, a subject was given sentences and paragraphs from a second year reading text book.

The writing test resembled the reading test. A subject was asked to write individual letters, then words and, finally sentences. We found that subjects who could read some of the letters often could not write them.

Tasks and Procedures

Subjects were tested on three metalinguistic tasks that tested their ability to segment speech: (a) phoneme counting, (b) syllable counting, and (c) deletion of the initial consonant.

Phoneme counting

Materials. The test contained 40 items; 30 were words and 10 were single phonemes. There were equal numbers of one-, two-, three-, and four-phoneme items. The single-phoneme items included all the vowels in Serbo-Croatian (a, o, e, i, u) and some consonants (z, ž, s, š, f). Among the two-phoneme items, half were pronouns and half were nouns. All three- and four-phoneme items were concrete, high frequency nouns. The structure of most test words was alternating consonant-vowel (CV) sequences. The test began with four sets of training items (each set containing one-, two-, three-, and four-phoneme items ordered successively; these items offered the subject an opportunity to deduce the nature of the unit being counted.

Procedure. This test, modeled after that designed by Liberman et al. (1974), required the subjects to count the phonemes of auditorily presented items. The examiner demonstrated the task by performing the first training set: she said the word with normal speed and intonation, and then she repeated the word again slowly, tapping once with a wooden dowel for each spoken phoneme. After the demonstration, the subject was invited to imitate the examiner and participate in the practice trials. The demonstration continued through three practice sets. The test items were ordered randomly with respect to the number of phonemes. Subjects responded by repeating the item and tapping the answer without the examiner's help.

Syllable counting

Materials. The test contained 40 items. There were equal numbers of one-, two-, three-, and four-syllable words. All the words were concrete, high frequency nouns. All syllables had either the CV or CVC structure.

Procedure. The design of this task was identical to that for phoneme counting except that items varied in the number of syllables instead of in the number of phonemes. The task was to count syllables in an auditorily presented word. The procedure was the same as in phoneme counting, starting with four sets of training items.

Consonant deletion

Materials. The items were pseudowords whose initial consonants were either [p], [s], or [m]. After correct deletion of the initial consonant, half of the items (experimental and practice) would become words and half would remain pseudowords. All items were of the CVCV or CVCVC structure.

Procedure. The subject had to delete the first phoneme from auditorily presented items provided by the experimenter. Initially the subject was told that her task was to delete the first "sound" of the item presented by the examiner, but because this instruction was rather difficult for some to follow, a special example was given using the subject's name. For example, if a subject's name was Zora, she was asked to say what she would be called if she lost the first "sound" of her name (_ora). If the subject could not provide the answer, the examiner would do so and then continue the instruction by using the name of another family member. The examiner started with ten demonstration trials and provided the response if necessary. After the practice trials were completed, 20 test items were presented with no feedback.

Control tasks

In order to assess verbal intelligence, a Serbo-Croatian form of the Peabody Picture Vocabulary Test (PPVT) was administered. Short-term memory was examined by means of forward and backward digit span subtests of the Wechsler Adult Intelligence Scale. Verbal short-term memory is known to have a phonological component but, in contrast to the experimental tasks, it is not segmental. Counterparts to the Liberman et al. (1974) tapping test required the subjects to count musical tones and sticks. It should be emphasized that tone counting is, in fact, quite challenging, given that the speed of presentation was intended to mimic speech.

Results

Literacy Assessment

In spite of their almost total lack of schooling, most of the subjects managed to develop measurable literacy skills (Table 1).

Alphabet identification. Knowledge of the alphabet, even in subjects who could recognize a majority of the letters, was not necessarily complete. Thus, the percentage of letter identification provides a convenient continuous measure of alphabet familiarity. Subjects were divided into three groups according to their ability to identify single letters of the alphabet. Those with poor letter recognition (the PLR group) consisted of 7 subjects, aged 60-74 (mean 68), who identified fewer than 50% of the letters (including two subjects who could not identify any letters).

Table 1. *Letter Recognition Scores (in %), Reading Achievement, Writing Achievement, and Literacy Scores for each subject.*

Letter recognition	Reading achievement	Writing achievement	Literacy score[a]
	Poor letter recognition		
0	does not read	does not write	0
0	does not read	does not write	0
25	reads single letters	does not write	1
30	reads single letters	does not write	1
50	letter-by-letter, up to 3 letters	few letters	4
40	does not read	few letters	1
40	does not read	does not write	0
	Medium letter recognition		
80	letter-by-letter, up to 4 letter words	does not write	4
90	letter-by-letter, up to 5 letter words	does not write	5
80	letter-by-letter, up to 3 letter words	few letters, with effort	4
95	letter-by-letter, up to 5 letter words	short words, many errors	9
90	letter-by-letter, up to 4 letter words	20% of letters, no words	6
90	letter-by-letter, up to 6 letter words	50% of letters, 3-letter words	10
60	only letters, no words	no letters	1
70	letter-by-letter, up to 4 letter words	no letters	4
90	letter-by-letter, up to 8 letter words	50% of letters	11
	Good letter recognition		
100	reads paragraphs, slowly, with errors	words, many errors	13
100	reads paragraphs, slowly, with errors	words, few errors	14
100	reads paragraphs, slowly, fluently	sentences, many errors	15
100	reads paragraphs, slowly, with errors	words, many errors	13
100	reads paragraphs, slowly, with errors	words, many errors	13
100	reads paragraphs, slowly, with errors	words, many errors	13
100	reads paragraphs, slowly, fluently	sentences, few errors	16

[a]Each increment in reading or writing achievement adds one point. See text for details.

The medium letter recognition group (MLR) consisted of 9 subjects, aged 55-76 (mean 65), who could identify 50-95% of the letters. Finally, the good letter recognition group (GLR) of 7 subjects, aged 55-65 (mean 61), identified all of the letters.

Reading achievement. In general, subjects' reading achievement was predicted by the subjects' letter recognition success. A qualitative description of each subject's performance is provided in Table 1, next to their letter recognition scores.

Women in the PLR group could not read. The one who recognized 50% of the letters could make letter-by-letter sounds for words up to three letters, but she could not combine those sounds into a single word unit. Women in the MLR group read letter by letter for words between 3 and 8 letters (mean = 4.3). That is to say, they made a sound for each letter and, in many cases, were then able to blend those sounds into a single word. In a number of cases, however, the blended sound was a nonword. Women in the GLR group could all read words, sentences, and paragraphs. All were slow and most made a large number of errors. In view of the fact that the material was from a second grade reading text, it must be noted that even the best subjects were only near-literates.

Writing achievement. For every skill level, writing ability lagged behind reading ability. Figure 1 shows some writing samples and demonstrates difficulties even among those with the highest alphabet familiarity. A qualitative description of subjects' performance is provided in Table 1, next to their reading achievement scores.

Only two women in the PLR group could write any letters at all. Seven women in the MLR group could write no more than 6 letters (despite recognizing 18-29 of them), two could write half of the letters, and one of these could write some 3-letter words. For the GLR group, those who did not make reading errors were also the best at writing to dictation, even attempting sentences (one with many errors, the other with few errors). The other GLR subjects could, with many errors, write words to dictation.

Estimated literacy score. For subsequent analyses, these reading and writing achievements were assigned a score, beginning with zero, for those who could neither read nor write any letters, up to 16 for those who could read paragraphs fluently. Each increment on either the reading or writing side was worth one point. These improvised literacy scores are provided in the rightmost column of Table 1. A one-way analysis of variance (ANOVA) performed on these reading scores as a function of letter recognition group revealed a significant effect of group, $F(2, 20) = 54.18$, $p < .0001$ (PLR = 1.0 MLR = 6.0, GLR = 13.9). Post hoc tests found all comparisons to be significant, $p < .05$.

Experimental and control tasks

All subjects achieved 100% success on stick counting. Subjects had great difficulty with backward digit span, averaging only 1.2 digits (the groups did not differ in a one-way ANOVA, $F < 1$). These two tasks were not considered in subsequent analyses. Subjects' scores on each of the experimental and remaining control tasks are shown in Table 2. The data were evaluated in a number of ways. Because particular control tasks were not logically paired with particular experimental tasks separate analyses were conducted. For the experimental tasks, a 3 (good, medium, and poor letter identification) × 3 (syllable counting, phoneme counting, and phoneme deletion) ANOVA addressed whether people of varying alphabetic familiarity differed with respect to phonemic awareness. The main effect of group, $F(2, 20) = 70.11$, $p < .0001$, indicates that skill at letter identification is associated with better overall performance on the tasks (GLR = 87%, MLR = 61%, PLR = 34%). The main effect of task, $F(2, 40) = 20.96$, $p < .0001$, indicates that subjects performed better on syllable counting (72%) and phoneme counting (70%) than phoneme deletion (40%). The interaction of group with task was significant, $F(4, 40) = 6.46$, $p < .0004$, and qualifies both of these interpretations. In particular, planned comparisons between letter recognition groups on each of the tasks revealed no group differences on syllable counting but all comparisons (GLR-MLR, MLR-PLR, and GLR-PLR) were significant for both phoneme counting and phoneme deletion (Tukey, $p < .05$).

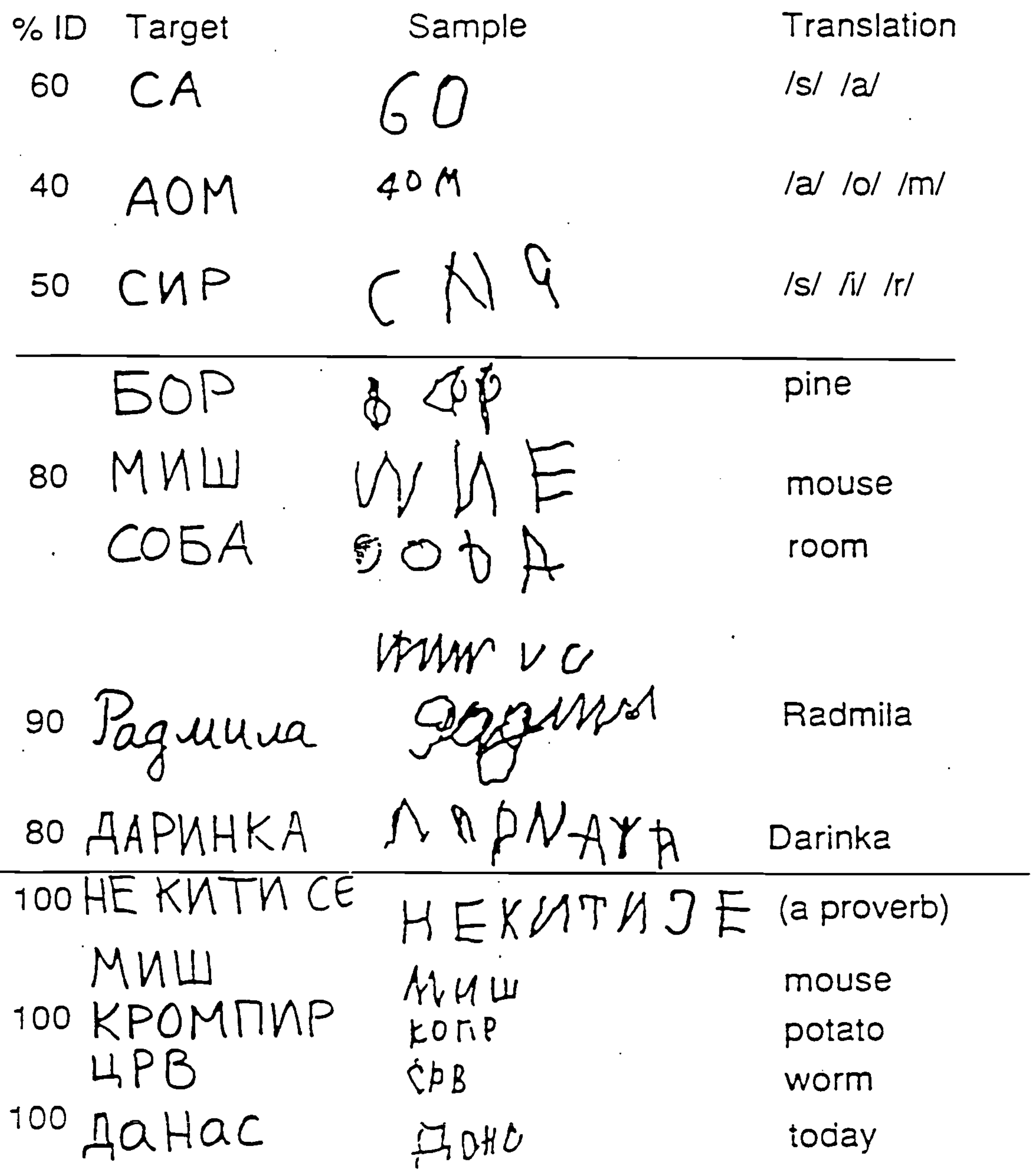


Figure 1. **Writing samples from subjects at different levels of letter recognition skill (indicated as %ID). Note that, even among those with relatively high letter recognition, letters are wrong (W for b and E for i in mouse), many letters are flipped (P in Radmilla's signature, n in Darinka's signature), letters are missing (P, h, and n in potato), and word boundaries are missing (in the proverb).**

Table 2. *Subjects' Age and Performance (in %) on Experimental and Control Tasks*

Age	Letter recognition	Syllable counting	Phoneme counting	Phoneme deletion	PPVT	Tone counting	Forward digit span
			Poor letter recognition				
72	0	77	13	0	89	57	71[a]
61	0	72	25	0	90	60	57
74	25	67	47	0	87	74	86
68	30	0	66	0	84	65	43
70	50	67	77	0	86	52	43
69	40	77	0	0	89	50	43
60	40	80	47	0	91	75	86
			Medium letter recognition				
76	80	55	85	25	86	80	71
78	90	70	87	40	90	70	71
68	80	77	80	0	86	72	57
60	95	72	77	75	90	52	57
62	90	72	57	45	88	52	71
55	90	95	82	50	91	88	71
54	60	72	57	20	90	67	86
76	70	80	67	0	90	60	71
60	90	58	77	60	92	65	71
			Good letter recognition				
63	100	60	92	90	88	80	86
100	95	100	85	85	77	86	
58	100	82	100	100	99	70	100
58	100	92	92	85	92	52	86
64	100	85	92	70	97	94	86
60	100	85	95	85	92	67	71
65	100	70	97	80	93	90	71

[a]Forward digit span is expressed as a percentage of the maximum performance obtained with this population, which was seven items.

In order to assess whether such differences simply reflect general intellectual skills, a parallel 3 × 3 ANOVA for the control tasks (PPVT, tone counting, and digit span) was conducted (to be comparable to the other tasks, which were all percentages, forward digit span was scaled to the maximum performance in this population). The main effect of group, $F(2, 20) = 6.96$, $p < .01$, again indicates that skill at letter identification is somewhat associated with better overall performance on the tasks (GLR = 84%, MLR = 75%, PLR = 70%). The main effect of task, $F(2, 40) = 33.30$, $p < .0001$, indicates that subjects performed better on PPVT (90%) than on tone counting (68%) or forward digit span (72%). The interaction of group with task was not significant, $F(4, 40) = 1.72$, $p > .15$. Nonetheless, planned comparisons were carried out and revealed no significant group differences on PPVT or tone counting and only one difference (GLR-PLR) for forward digit span (Tukey, $p < .05$). It has been shown that verbal short-term memory limitations have a phonological basis (Baddeley, 1966) and are associated with differences in reading ability (Conrad, 1972; Shankweiler et al., 1979). The differences among letter recognition groups as a function of experimental and control tasks are shown in Figure 2.

A second analysis focused on the counting tasks in a 3 (good, medium, and poor letter identification) × 2 (syllable versus phoneme) × 4 (number of segments) ANOVA on the errors. The main effect of group, $F(2, 18) = 7.13$, $p < .01$, again indicates that better letter identification is associated with better performance in the form of fewer errors (PLR = 3.6, MLR = 2.6, GLR = 1.3). While there was a main effect of number of segments, $F(3, 54) = 6.11$ $p < .01$, there was no overall difference between tasks, $F(1, 18) < 1$. These effects are best seen in interactions. The Group × Task interaction, $F(2, 18) = 4.39$, $p < .03$, revealed that the group difference was attributable wholly to the phoneme counting task (Figure 3), with simple effects tests showing significance for phonemes, $F(2, 36) = 10.82$, $p < .0001$, but not for syllables, $F < 1$. The Group × Segments interaction, $F(6, 54) = 4.74$, $p < .001$, indicates that number of segments mattered for PLR (simple effects: $F(2, 54) = 10.20$, $p < .0001$) and MLR ($F(2, 54) = 4.30$, $p < .01$) but not for GLR ($F \approx 1$). The Task × Segments interaction, $F(3, 54) = 18.62$, $p < .0001$, found number of segments to matter for both tasks but in opposite directions: There were more errors with more phonemes ($F(3, 54) = 19.20$, $p < .0001$) but fewer errors with more syllables ($F(3, 54) = 6.4$, $p < .001$).

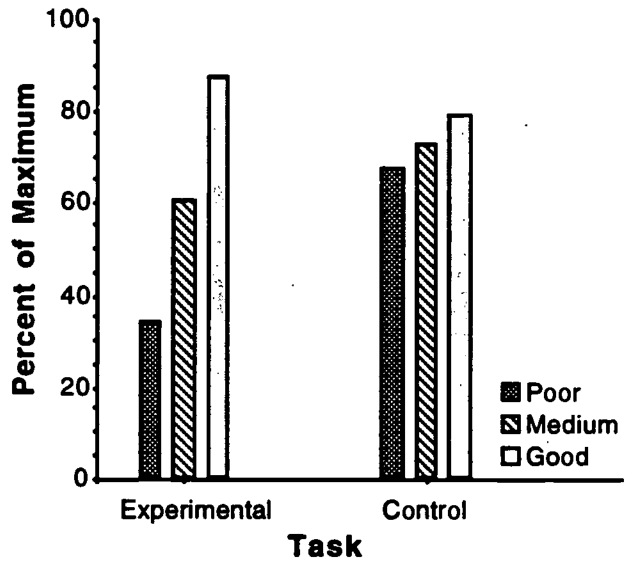


Figure 2. **Performance on experimental and control tasks (individual tasks of each type were combined; see Footnote 2) for the different letter recognition groups.**

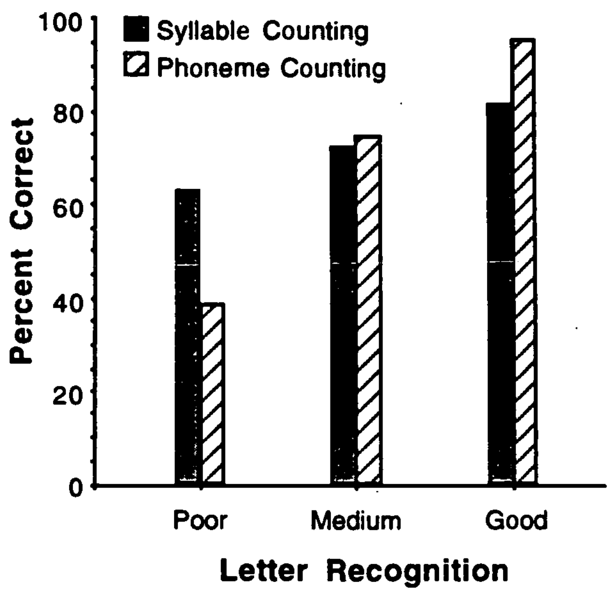


Figure 3. **Performance on syllable and phoneme counting tasks for the different letter recognition groups.**

The marginal Group × Task × Segments interaction, $F(6, 54) = 2.11, p < .07$, pinpoints this reversal and provides a dramatic contrast between the syllable and phoneme tasks. The flip is largely due to the MLR group who encountered special difficulty in counting the number of syllables in one-syllable words, averaging almost 7 errors out of a possible 10 (upper panel, Figure 4). The simple effect of number of syllables was not significant for either GLR or PLR. Systematic group differences are apparent in phoneme counting (lower panel, Figure 4), however: The number of segments increased errors slightly for GLR ($p < .01$), moderately for MLR ($p < .001$) and sharply for PLR ($p < .0003$).

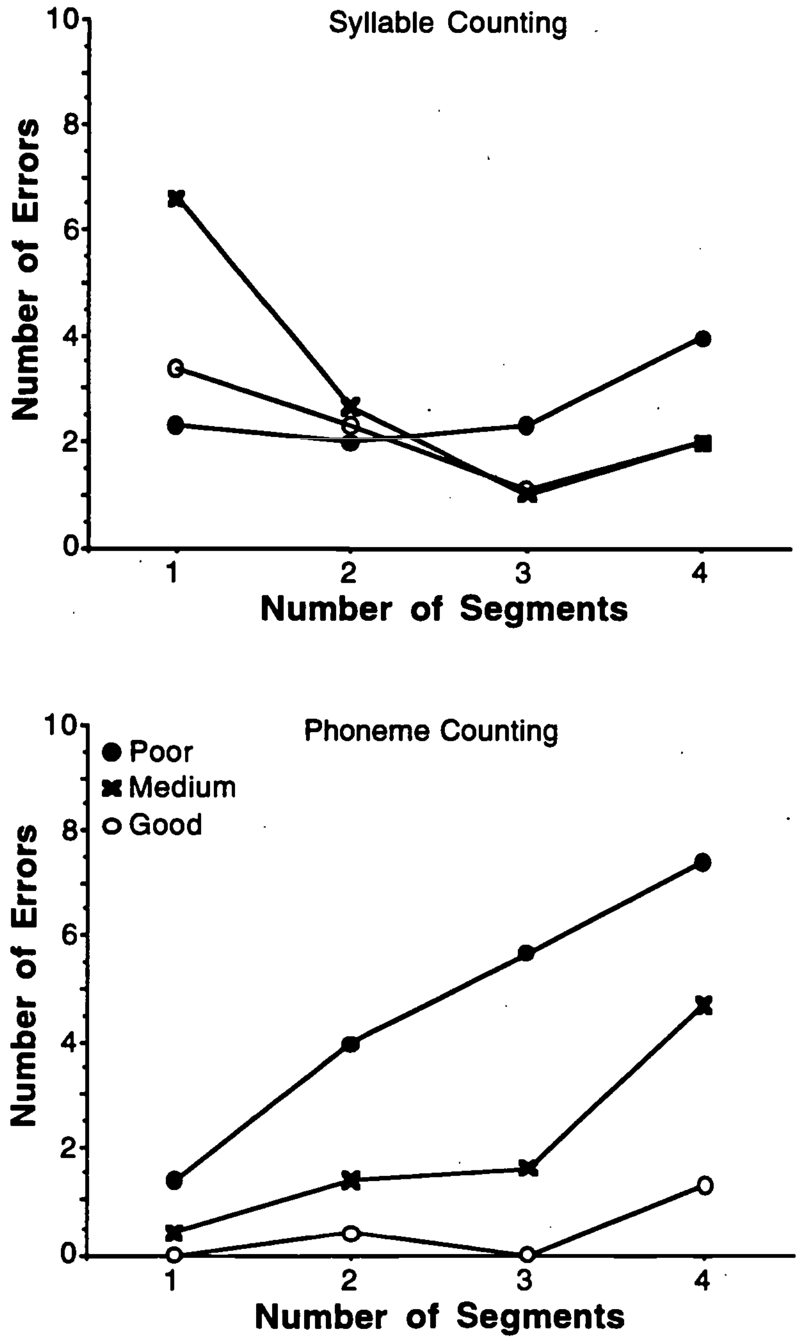


Figure 4. **Number of errors for syllable counting (top) and phoneme counting (bottom) as a function of number of segments and letter recognition group.**

These group differences were repeated in the analyses of the phoneme deletion scores. Letter recognition group was significant, $F(2, 20) = 43.75$, $p < .0001$, with average scores of PLR = 0 (s.d. = 0), MLR = 35 (s.d. = 25.9), and GLR = 85 (s.d. = 9.1). All paired comparisons were significant. In parallel with other studies testing illiterate adults (Morais et al., 1979; Read et al., 1984) and children (Stanovich et al., 1984), this task was the most difficult for our adult illiterate subjects.

Inter-relatedness of tasks

A correlation matrix of the eight measures makes apparent the interrelationships among tasks (Table 3). In particular, phoneme counting, phoneme deletion, letter recognition, and the literacy score are highly correlated with one another and less so with the other variables. Syllable counting is not correlated significantly with either of the phoneme tasks, nor is it correlated with the literacy score. A stepwise regression of the literacy score against the tasks enters only phoneme deletion and phoneme counting as significant, $r^2 = .92$, $p < .0001$. The pattern of results point to (1) a difference between the abilities to segment syllables and phonemes and (2) a relationship between literacy level and phonemic segmentation abilities.

The possibility of constitutional literacy deficits

We should address the possibility that our sample may have included individuals whose reading deficit is due to neurobiological factors, and that these may have contributed to the observed group differences. Such a possibility is suggested, for example, by the fact that some of the subjects had dropped out of reading instruction courses, although the reason given during the individual interviews pertained to socioeconomic factors rather than reading difficulties. Moreover, developmentally language-impaired individuals would be expected to have other language related difficulties, such as naming common pictorial objects (Katz, 1985; Wolf & Goodglass, 1986). However, PPVT scores did not differ across letter recognition groups and performance by all individuals was uniformly strong. Finally, a case for neurologic causation could be made for individuals who had good knowledge of the alphabet (which does not require sophisticated phonological skills) coupled with poor performance on phonologically taxing tasks.

Table 3. *Correlation Matrix for Experimental and Control Tasks*

	Letter recognition	Syllable counting	Phoneme counting	Phoneme deletion	PPVT	Tone counting	Digit span	Literacy score
Letter recognition	–							
Syllable counting	.32	–						
Phoneme counting	.85***	.10	–					
Phoneme deletion	.83***	.32	.73***	–				
PPVT	.36	.44*	.24	.52**	–			
Tone counting	.35	.09	.49*	.30	.24	–		
Digit span	.39	.45*	.39	.53**	.52**	.46*	–	
Literacy score	.85***	.34	.79***	.95***	.51**	.40	.45*	

*$p < .05$; **$p < .01$; ***$p < .0001$

But a reexamination of Table 2 reveals no instances of such a dissociation in the GLR group: Those who identified 100% of the letters of the alphabet performed well on both phoneme counting and phoneme deletion. One PLR individual and one MLR individual may be considered discrepant on phoneme counting (40% letters with 0% phoneme counting; 90% letters with 57% phoneme counting) and two different MLR individuals may be considered discrepant on phoneme deletion (both of whom scored 0% on phoneme deletion after identifying 70% and 80% of the letters). These two types of discrepancies do not identify the same individuals, as would be expected if their reading deficits were primarily constitutional in origin. Moreover, eliminating these individuals from the analyses (either collectively or as pairs) does not alter the pattern of significances, including the Group × Task interaction.

DISCUSSION

Native speakers of Serbo-Croatian who are adept at identifying the letters of their alphabet are also adept at performing tasks that tap phonological abilities, in particular, those that involve phoneme awareness. They are far superior to ostensibly comparable groups who identify letters less well. It should be emphasized that the women in this study have had no formal schooling. The only reading instruction they may have had lasted no longer than a month and occurred over 40 years prior to testing. Nonetheless those who succeeded, on the basis of such minimal exposure, in establishing a link between graphemes and phonemes achieved impressive phonemic awareness abilities. Indeed, alphabetic familiarity seems to be a useful index of the lower end of the literacy continuum in Serbo-Croatian.

Against this background, the results can be evaluated with respect to our predictions: namely, (1) minimal phonemic awareness among the truly illiterate, (2) better performance on the syllable task than on the phoneme tasks, and (3) a strong link between alphabetic familiarity and phonemic awareness in Serbo-Croatian. Consistent with (1), it was found that the truly illiterate adults—those in the PLR group—were unable to do the phoneme deletion task at all. This task has been found to be the most difficult for beginning readers as well (Stanovich et al., 1984). Table 4 allows a comparison of the phoneme deletion task used in the present study with that used by Morais et al. (1979) in their study of Portuguese illiterates. Note the gradation in performance as a function of alphabet familiarity by the Serbo-Croatian-speaking subjects. The best performed at ceiling, whereas the worst performed at floor. The third group was in between. This result suggests that the discreteness implied by the categories literate vs. illiterate is somewhat misleading. The superior performance of the Portuguese illiterates relative to the Yugoslav PLR group suggests that the Portuguese group probably included people with some alphabetic ability.

Table 4. *Percent Correct Phoneme Deletions by Portuguese and Serbo-Croatian Speakers on Targets Requiring Word or Nonword Responses.*

Language	Literacy	Response: Word	Response: Nonword
Portuguese[a]	Illiterates	26	19
	Literates	87	73
Serbo-Croatian	Poor letter recognition	0	0
	Medium letter recognition	62	51
	Good letter recognition	94	91

[a]Data are taken from Morais et al. (1979)

The poorer performance by the Portuguese recent literates relative to the Yugoslav GLR group, especially on nonwords, probably reflects the fact that this Portuguese group included people who were not firmly established in their literacy skills. In other words, the finer gradation apparent when subjects are divided on the alphabetic familiarity dimension reinforces the idea that literacy is a continuum.

The PLR group also had considerable difficulty with phoneme counting, averaging 39% correct. If we consider 25% to be chance performance (given that all of the counting tasks involved up to 4 entities), three people were at or below chance including the two who could recognize no letters of the alphabet. It should be noted, however, that two of the PLR women achieved modest success, scoring 66% and 77% correct. Nonetheless, it seems clear that phonemic awareness and literacy go hand in hand, a conclusion that is buttressed by developmental and cross-language data (e.g., Ball & Blachman, 1991; Bradley & Bryant, 1983; Byrne & Fielding-Barnsley, 1991; Cossu et al., 1988; Lundberg et al., 1980; Lundberg et al., 1988; Morais et al., 1986).

In support of (2), the PLR group was better at syllable counting than phoneme counting, but the opposite was true of the GLR group, however, although the GLR subjects did better than the MLR subjects on both tasks. This pattern is reminiscent of Italian children for whom the initial advantage of syllables over phonemes also reversed for first and second graders (Cossu et al., 1988). In contrast, for American children, the syllable advantage shrinks with age (nursery school, kindergarten, and first grade), but does not disappear (Liberman et al., 1974). The similar pattern for Italian children and Yugoslav adult illiterates is noteworthy given that, relative to English, both languages are phonologically shallow. Again, across ages and across orthographies, language users with minimal phonemic awareness have considerably less difficulty with syllables—that is, segments that are closer to the basic unit of articulation. Whether a syllable advantage persists (or, perhaps, reverses with literacy experience) seems to be language specific, however.

The corroboration of (3) provides what is, perhaps, the most striking feature of the results—how closely associated are letter knowledge and phonemic awareness for the unschooled Yugoslavs. The low literacy English-speaking adults of Liberman et al. (1985) and Read and Ruyter (1985) had considerably more schooling than our Yugoslav subjects—presumably their exposure to printed materials was far greater—and yet the MLR (74%) and GLR (95%) groups exceeded the phoneme counting scores of their English-speaking counterparts (39-58%). Interestingly, the syllable counting scores of the English-speakers (77% [Read & Ruyter, 1985]) were comparable to those of the MLR (72%) and GLR (81%) Yugoslavs. Amplifying this point, the lower panel of Figure 4 shows that with increasing ability to identify letters, subjects made fewer errors and were less bothered by multiple segments in counting phonemes. An obvious source of this cross-language difference is the shallow phonology of the Serbo-Croatian language, which lends itself to a correspondingly shallow orthography, with its set of one-to-one mappings between phonemes and graphemes. One consequence of the difference between the two orthographies, which may be important in understanding these results is the matter of the letter names. In Serbo-Croatian, letters are named by their sounds (/ah/, /buh/, /kuh/) rather than by non-corresponding names, as in the English ("ay," "bee," "see"). Reading instruction emphasizes this, so that knowing the letters is, in effect, knowing their phonemic correspondences. In English, some letter names refer to only one of the phonemic interpretations possible for that letter (e.g., "gee" identifies the soft consonant /dz/ but not the hard /g/), while the names of others are actually misleading (e.g., "aitch," "double u"). These differences help to explain why the link between alphabetic familiarity and phonemic awareness is language-specific.

Phonemic awareness and literacy achievement

The groups distinguished by letter recognition ability differed most clearly on phoneme counting and phoneme deletion tasks, but differed minimally, if at all, on syllable counting and the control tasks. Both syllable tasks and phoneme tasks have been studied as measures of phonological awareness (e.g., Lundberg et al., 1980, 1988; Mann & Liberman, 1984); they have been found to make unequal contributions in predicting subsequent reading success (e.g., Cossu et al., 1988; Liberman et al., 1974; Mann, 1984). Moreover, data sets with a sufficiently large sample size to permit a factor analysis have shown these tasks to load onto different underlying factors, whether for low literacy English-speaking adults (Read & Ruyter, 1985), Swedish 7-year old-kindergartners (Wagner & Torgesen's, 1987, reanalysis of data from Lundberg et al., 1980),[2] or Danish kindergarten children (Lundberg et al., 1988). Thus, across languages there is impressive consistency in the nature of the metaphonological measures that show the highest relations with literacy.

It is clear that whatever modest reading abilities adult illiterate Yugoslavs have attained covary dramatically with their degree of phonemic awareness. Indeed, the association between phonemic segmentation and literacy measures appears to be, if anything, more direct in shallow orthographies than in deeper ones. For English-speaking persons, we have reason to believe that phonological awareness and letter knowledge are each necessary, but not sufficient conditions for word recognition in reading. As Byrne and Fielding-Barnsley demonstrated (1989, 1991, 1995), these abilities do not account for all the variance in the reading scores of children studied during the acquisition phase. Moreover, as we noted, the association between individual differences in letter knowledge and phoneme segmentation skill, which is well-nigh total in the present study, appears to be far weaker for beginning readers of English and Danish. Training studies have shown how dissociable these abilities can be. In English, Ball and Blachman (1991) found that training in letter names and letter sounds alone did not significantly improve the segmentation skills or the reading skills of kindergarten children. Conversely, Lundberg and his colleagues showed that training Danish kindergartners to segment spoken words phonemically did not enhance their ability to identify letters beyond those of a control group. Certainly the present findings are very different from these. They fit with other indications (which are reviewed elsewhere, e.g. Carello, Turvey, & Lukatela, 1992; Lukatela & Turvey, 1990 a and b; Lukatela & Turvey, 1991) that the Serbo-Croatian orthography is highly phonologically penetrable.

Conclusion

Our research has found that, with Serbo-Croatian speakers at least, a little (letter) knowledge goes a long way. Some almost totally unschooled speakers of this language can penetrate remarkably far into the orthography, armed only with phonological awareness and alphabetic knowledge. We suggest that for a Serbo-Croatian speaker, knowing the letter units is the entry point into the alphabetic principle because letters and phonemes are related so straightforwardly. Knowing the letters is not as helpful for speakers of English, as discussed earlier. In all events, the present research lends support to the claim that literacy is a continuum. Just as differences in phonological awareness have been found between good and poor literates, so too have we found differences in phonological awareness between good and poor illiterates. How easily one moves along that continuum is dependent on language-specific features.

REFERENCES

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362-365.

Ball, E. W., & Blachman, B. A. (1991). Does phonological awareness raining in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*, 49-66.

Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301*, 419-521.

Bryant, P., MacLean, M., & Bradley, L. (1990). Rhyme, language, and children's reading. *Applied psycholinguistics, 11*, 237-252.

Bryant, P., MacLean, M., Bradley, L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26*, 429-438.

Byrne, B. (1990). Experimental analysis of the child's discovery of the alphabetic principle. In L. Rieben & C. A. Perfetti (Eds.), *Learning to read: Basic research and its implications*. Hillsdale, NJ: Erlbaum.

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology, 81*, 313-321.

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83*, 451-455.

Byrne, B., & Fielding-Barnsley, R. (in press). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up, and a new preschool trial. *Journal of Educational Psychology*.

Carello, C., Turvey, M. T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological? In R. Frost & L. Katz (Eds.), *Orthography, morphology, phonology, and meaning* (pp. 211-226). Amsterdam: Elsevier.

Conrad, R. (1972). Speech and reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading* (pp. 205-240). Cambridge, MA: MIT Press.

Cossu, G., Shankweiler, D., Liberman, I. Y., Katz, L., & Tola, G. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics, 9*, 1-16.

Cossu, G., Shankweiler, D., Liberman, I. Y., & Gugliotta, M. (in press). Visual and phonological determinants of misreadings in a transparent orthography. *Reading & Writing: An Interdisciplinary Journal*.

DeFrancis, J. (1989). *Visible speech*. Honolulu: University of Hawaii Press.

Ehri, L. C. (1984). How orthography alters spoken language competencies in children learning to read and spell. In J. Downing & R. Valtin (Eds.), *Language awareness and learning to read* (pp. 119-147). New York: Springer Verlag.

Feldman, L. B., Lukatela, G., & Turvey, M. T. (1985). Effects of phonological ambiguity on beginning readers of Serbo-Croatian. *Journal of Experimental Child Psychology, 39*, 492-510.

Katz, R. B. (1985). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. *Cognition, 22*, 225-257.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23*, 65-77.

Liberman, I. Y., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. Venezky (Eds.), *Orthography, reading, and dyslexia* (pp. 137-153). Baltimore, MD: University Park Press.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex, 18*, 367-375.

Liberman, I. Y., Rubin, H., Duquès, S., & Carlisle, J. (1985). Linguistic abilities and spelling proficiency in kindergartners and adult poor spellers. In J. Kavanagh & D. Gray (Eds.), *Biobehavioral measures of dyslexia* (pp. 163-176). Parkton, MD: York Press.

Liberman, I. Y., & Shankweiler, D. P. (1979). Speech, the alphabet and teaching to read. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading* (Vol. 2, pp. 109-134). Hillsdale, NJ: Erlbaum.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education, 6*, 8-17.

Liberman, I. Y., Shankweiler, D. P., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Liberman, I. Y., Shankweiler, D. P., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability* (pp. 1-33). Ann Arbor, MI: University of Michigan Press.

Lukatela, G., & Turvey, M. T. (1990a). Phonemic similarity effects and prelexical phonology. *Memory & Cognition, 18*, 128-152.

Lukatela, G., & Turvey, M. T. (1990b). Automatic and prelexical computation of phonology in visual word identification. European Journal of Cognitive Psychology, 2, 325-343.

Lukatela, G., & Turvey, M. T. (1991). Phonological access of the lexicon: Evidence from associative priming with pseudohomophones. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 951-966.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23*, 263-284.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Mann, V. A. (1984). Longitudinal prediction and prevention of early reading difficulty. *Annals of Dyslexia, 34*, 117-136.

Mann, V. A. (1986). Why some children encounter reading problems: The contribution of difficulties with language processing and phonological sophistication to early reading disability. In J. Torgesen & B. Wong (Eds.), *Psychological and educational perspectives on learning disabilities* (pp. 133-154). New York: Academic Press.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17*, 592-599.

Mann, V. A. Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory & Cognition, 8*, 329-335.

Mattingly, I. G. (1972). Reading, the linguistic process and linguistic awareness. In J. Kavanagh & I. Mattingly (Eds.), *Language by ear and by eye* (pp. 133-147).

Mattingly, I. G. (1980). Reading, linguistic awareness, and language acquisition. *Haskins Status Report on Speech Research, SR-61,* 135-150.

Mattingly, I. G. (1992). Linguistic awareness and orthographic form. In R. Frost & L. Katz (Eds.), *Orthography, morphology, phonology, and meaning* (pp. 11-26). Amsterdam: Elsevier.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24,* 45-64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7,* 323-331.

Morais, J., Content, A., Bertelson, P., Cary, L., & Kolinsky, R. (1988). Is there a critical period for the acquisition of segmental analysis? *Cognitive Neuropsychology, 5,* 347-352.

Read, C., & Ruyter, L. (1985). Reading and spelling skills in adults of low literacy. *Reading and Special Education, 6,* 43-52.

Read, C., Zhang, Y-F., Nie, H-Y., & Ding, B-Q. (1984). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24,* 31-44.

Rosner, J., & Simon, D. (1971). The auditory analysis test: An initial report. *Journal of Learning Disabilities, 4,* 384-392.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning & Memory, 5,* 531-545.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175-190.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

Wolf, M., & Goodglass, H. (1986). Dyslexia, dysnomia and lexical retrieval: A longitudinal investigation. *Brain and Language, 28,* 159-168.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 159-177.

FOOTNOTES

**Applied Psycholinguistics, 15,* 463-487.

†Also University of Connecticut, Storrs.

[1]Although Serbo-Croatian is phonologically precise—a given letter has a single pronunciation—there is a complication: It happens to be phonologically precise in two largely distinct but partially overlapping alphabets, Roman and Cyrillic (children are expected to be fluent in both by the second grade). The nature of the overlap is such that a few shared letter are pronounced the same way in the two alphabets (e.g., E, A, T are pronounced /e/, /a/, and /t/, respectively) and a few shared letters are pronounced differently, depending on which alphabet they are read in (e.g., B is /b/ in Roman but /v/ in Cyrillic; P is /p/ in Roman but /r/ in Cyrillic). Letter strings composed of a mix of these shared letters are phonologically ambiguous. For example, BETAP read in Cyrillic is /vetar/, the word for "wind." Read in Roman, it is /betap/, a nonword. Phonologically unique versions of these same letter strings can be constructed in one or the other alphabet (Feldman & Turvey, 1983). For BETAP, the phonologically unique control is VETAR, which can only be pronounced /vetar/, again the word for "wind." The standard result with adult readers (Lukatela, Feldman, et al., 1989; Lukatela, Turvey, et al., 1989) is that they take longer to name a phonologically ambiguous letter string than its phonologically unique counterpart.

[2]Wagner & Torgesen nonetheless concluded (on the basis of "confirmatory factor analyses") that tasks involving syllables and tasks involving phonemes tap a single latent ability.

Attention Factors Mediating Syntactic Deficiency in Reading-disabled Children*

Avital Deutsch† and Shlomo Bentin‡

Syntactic context effects on the identification of spoken words, and the involvement of attention in mediating these effects, were examined in seventh grade children with reading disabilities and children who were good readers. The subjects were asked to identify target words that were masked by white noise. All targets were final words embedded in unmasked sentences. Relative to a syntactically neutral context, the identification of targets whose morpho-syntactic structure was congruent with the context was facilitated and the identification of syntactically incongruent targets was inhibited. Reading-disabled children were less inhibited by syntactic incongruence than good readers. Presenting congruent and incongruent sentences in separate blocks reduced the amount of inhibition in good readers while having no effect on the reading-disabled. The percentage of correct identification of incongruent targets in the mixed presentation condition was larger for reading-disabled than for good readers, whereas in the blocked presentation condition the percentage of correct identification was equal across groups. The amount of facilitation was not affected by blocking the congruent and incongruent conditions, and was equal across reading groups. It is concluded that, in both reading groups, the syntactic structure of the context triggers a process of anticipation for particular syntactic categories which is based on a basic assumption that linguistic messages are syntactically coherent. Reading-disabled children are, however, less aware of this process and are therefore less affected when the syntactic expectations are not fulfilled.

INTRODUCTION

The existence of an impairment in the syntactic ability of children with severe reading difficulties is a matter of controversy. One view is that children with reading difficulties lack basic syntactic abilities due to delayed development of language skills (Byrne, 1981), or due to structural deficiencies in the language system (Stein, Cairns, & Zurif, 1984). This view is supported by ample evidence that reading-disabled children are inferior to good readers on various tests of syntactic ability (Bohannon, Warren-Leubecker, & Hepler, 1984; Bowey, 1986; Brittain, 1970; Byrne, 1981; Goldman, 1976; Guthrie, 1973; Flood & Menyuk, 1983; Siegel & Ryan, 1984; Stein, Cairns, & Zurif, 1984; Tunmer, Nesdale, & Wright, 1987; Vogel, 1974; Wigl, Semel, & Crouse, 1973; Willows & Ryan, 1986). An alternative view is that syntactic deficiency is not characteristic of reading disability. The proponents of this view point out, for example, that the speech of reading-disabled children is grammatically correct most of the time, and that they do not differ from their normally reading peers in using syntactic rules for generating sentences. Accordingly, those authors suggest that the deficient syntactic ability observed in reading-disabled children reflects a limitation of short-term memory caused by a basic difficulty in

This study was supported by a grant from the Israel Foundations Trustees to Shlomo Bentin and Partly by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. Avital Deutsch was supported by a stipend from the Levin Center for Child Development. We would like thank all of the schools for enabling us to carry out this study. We thank Dr. Oren Lam for allowing us to study the children in his clinic.

generating phonological codes (Fowler, 1988; Shankweiler & Crain, 1986; Shankweiler, Crain, Brady, & Macaruso, 1992; Shankweiler, Smith & Mann, 1984; Smith, Macaruso, Shankweiler, & Crain, 1989).

The two opponent views can be brought closer together by assuming that the apparent syntactic inferiority of children with reading difficulties does not reflect the absence of basic syntactic knowledge, but rather a poor ability to use this knowledge proficiently. Thus, this view agrees with those who assume that the syntactic knowledge of reading-disabled children is basically intact, yet it does not consider poor phonological skills to be the only cause of the poor syntactic performance observed in these children. Instead, a metalinguistic problem is suspected, which does not significantly affect the natural process of speech, but may interfere with less natural linguistic activities such as reading, or complex linguistic processes that may be required in specially designed experimental procedures.

Empirical support for a metalinguistic account of the impaired performance of the reading-disabled in tests of syntactic processing was provided in a study of syntactic-context effects on the identification of spoken words (Bentin, Deutsch, & Liberman, 1990). Previous studies showed that words are processed faster and more accurately when they are embedded in a congruent than in an incongruent syntactic context. This syntactic-context effect was demonstrated in fluently reading adults using visual word recognition (Carrello, Lukatela, & Turvey, 1988; Goodman, McClelland, & Gibbs, 1981; Gurjanov, Lukatela, Moskovljević, & Turvey, 1985; Lukatela, Kostić, Feldman, & Turvey, 1983; Lukatela & Moraco, Stojonov, Savić, Katz, & Turvey, 1982; Miller & Isard, 1963; Seidenberg, Waters, Sanders, & Langer, 1984; Tanenhaus, Leiman, & Seidenberg, 1979; West & Stanovich, 1986; Wright & Garrett, 1984) and in the identification of spoken words (Katz, Boyce, Goldstein, & Lukatela, 1987; Marslen-Wilson, 1987; Tyler & Wessels, 1983). In their study, Bentin et al. (1990) found that the effect of syntactic context on the identification of white-noise masked spoken words was lower in a group of reading-disabled children than in normally reading matched controls. More specifically, it was found that the reading-disabled identified syntactically incongruous targets better than controls, while being equal on the identification of congruent targets. Furthermore, although reading-disabled children performed worse than the good readers in judging the grammaticality of the sentences, as well as in correcting the ungrammatical sentences, the difference between the groups was significantly larger in the correction task.

Our interpretation of these results was based on the assumption that the interference with the identification of incongruent targets resulted from a mismatch between context-based general expectations regarding the grammatical form of the target and the incomplete phonetic information provided by the masked input. We suggested that this mismatch inhibited the identification process in the good readers more than in the reading-disabled, because good readers are more likely to take into account the available syntactic information in the process of target identification. Since inhibitory processes are assumed to reflect the operation of mechanisms that require attention resources (Becker, 1985; Neely, 1991; Posner & Snyder, 1975), we suggested that the deficient syntactic performance observed in reading-disabled children is related to inefficient or impaired application of attention-mediated strategies in processing syntactic structures.

The involvement of attention factors may also explain why the reading-disabled could judge the grammaticality of sentences better than they could correct them (see also Fowler, 1988). Although both tasks require explicit syntactic knowledge, they differ in their degree of complexity and in the amount of attention required to perform them. While both tasks require recognizing the syntactic structure and finding deviations from known syntactic rules, the sentence correction task requires, in addition, the utilization of syntactic knowledge in the creation of new syntactic structures. Therefore, correcting syntactically aberrant sentences is perceived to require more attention than judging their grammaticality (de Villiers & de Villiers, 1974; Fowler, 1988). On the basis of our interpretation of the Bentin et al. (1990) results, we proposed that a comparison between the syntactic ability of good and disabled readers should distinguish between (1) the ability to automatically apply basic syntactic rules and (2) the ability to strategically use this syntactic knowledge while processing sentences. Our hypothesis is that the difference between good and disabled readers in the syntactic domain is best characterized in terms of the second process.

The validation of the above hypothesis required two steps. The first was to demonstrate that the effect of syntactic context on the identification of spoken words is indeed composed of an inhibitory attention-based component and a facilitatory automatic component. The second was to show that the distinction between the syntactic ability of good and of disabled readers is determined by the attention-mediated component of the syntactic-context effect.

The first step was accomplished in a recent study in which the subjects were fluently reading adults (Deutsch & Bentin, 1994). Using a neutral condition, we first determined that congruent syntactic context facilitates the identification of white-noise-masked words, whereas incongruent context interferes with their identification. In comparison with a randomly mixed presentation of syntactically congruent and incongruent sentences, isolating the presentation of congruent and incongruent sentences in separate blocks reduced the interference of syntactic incongruence but had no effect on the identification of congruent targets. Also, increasing the ISI between context and target enhanced the interference of the incongruent context without affecting the identification of congruent targets. Together, these results demonstrated that the syntactic context effect includes a source of inhibition which is sensitive to the manipulation of attention. The present study was designed to examine whether, conforming to our hypothesis, the syntactic context effect is less sensitive to the manipulation of attention in reading-disabled children than in good readers. If, relative to a neutral baseline, syntactic incongruity interferes less with the identification of white-noise-masked words in reading-disabled than in good readers, and if blocking congruent and incongruent sentences reduces this interference more in the latter then in the former reading group, than our hypothesis would be supported.

METHODOLOGICAL CONSIDERATIONS

As in our previous studies (Bentin et al., 1990; Deutsch & Bentin, 1994), the subjects' task in the present experiments was to identify white-noise-masked spoken words embedded in unmasked sentences. This task was used because studies of visual word identification have suggested that the degradation of stimulus intelligibility magnifies context effects (Becker & Killion, 1977; Meyer, Schvaneveldt, & Ruddy, 1975; Neely, 1991; Stanovich & West, 1983). The auditory modality was used in order to avoid confounding genuine syntactic processing difficulties, which may distinguish disabled from good readers, with poor performance which may result from the basic difference between the two groups in their ability to decipher written words.

The addition of noise may alter the normal process of word identification (for example by over-emphasizing contextual influences), and therefore may hamper the generalization of the present results beyond the particular experimental circumstances. However, it should not interfere with our use of this method to examine the nature of syntactic contextual processes whenever the linguistic system sets them in motion and utilizes them for word identification.

In the present study, we manipulated the Hebrew agreement rule between subject and predicate for gender and number, and the rule of conjunction of the pronoun and the preposition. The essential role of agreement rules in Hebrew, which has no effect on semantic processing, is to specify the syntactic relation between the constituents of a sentence. For example, the predicate agrees with the subject in person, gender and number but, because specification of the gender and number is already available in the subject, violation of one or more of these types of agreements does not affect the meaning of the sentence (Shanon, 1973). Moreover, since the agreement rule is at the level of inflectional morphology, its violation does not cause changes in word class (changes that may have semantic implications; Carello, 1988). Take, for example, the sentence "A nice boy is writing" which translates into Hebrew as "Yeled (sub.) yafeh (attrib.) kotev (pred.)." The morphological unit "yeled" (boy) contains information about gender (masculine) and number (singular). The same root (y.l.d) with different affixes is used to form the feminine "yaldah" (girl) or the plural "yeladim" (boys). The agreement rules require that attributes and predicates agree with the subject in gender and number: "yafeh" (nice) is a singular masculine form, as is "kotev" (is writing). The sentence "Yaldah yafah kotev" contains a syntactic violation because the predicate is in the masculine form while both the subject and attribute are in the feminine form. The conjunction rule provides that a pronoun and a preposition appear in a conjoined form. For example, the preposition "el" (to) and the pronoun "ata" (you) are composed into one form "alecha." The decomposition of this form into "el" + "ata" is illegal. (See additional examples and details in the "Test and Materials" section.)

These rules have a common ground in that they are based on the use of an inflectional system. Both rules are formal, language-dependent, and based on convention. Both subject-predicate agreement, and preposition-pronoun conjunction are simple and essential in Hebrew grammar and are acquired at a very early age. Moreover, the productive use of these rules and the analysis of their contribution to the verbal message do not involve complex processing or excessive memory load. (To ensure that possible inter-group differences in memory capacity do not influence the results, all the sentences in this study were syntactically simple and short [three or four words]).

They were chosen for two reasons. First, while our hypothesis was that the difference between the good and disabled readers may be related to a difference in their respective ability to use basic syntactic knowledge strategically during sentence processing, we aimed at disentangling this ability from the basic syntactic competence as it is demonstrated by the spontaneous use in every-day speech. These features ensured that the observed performance would reflect the ability to use syntactic knowledge rather than the mere existence of such knowledge. The second reason was that we were interested in isolating the effect of the syntactic context from the effect of the semantic context. Although the agreement rule that we chose operates between subject and predicate, we were not constrained to present these two sentential elements continuously. We could therefore avoid lexical priming effects that may operate between adjacent words and focus on processes related to the syntactic structure of the sentence. Furthermore, in order to avoid lexical priming based on semantic relationships, none of the targets was semantically related to preceding words in the context or could have been predicted on the basis of the semantic context of the sentence. Moreover, although agreement rules are applied mainly by adding or changing suffixes, they usually also involve phonological modifications in the structure of the whole word, as required by phonetic rules. In the above example, for instance, the addition of the plural suffix "im" to the singular/masculine form "kotev" changes the morphological form "kotev" into "kotvim" rather than *"kotevim." It should also be noted that there are several suffixes that are used to mark gender and number, some of which are shared by nouns and verbs. For example, the feminine form of the noun "zamar" (singer) is "zameret" while the feminine form of "rakdan" (dancer) is "rakdanit" and the feminine form of "yeled" (boy) is "yaldah." Similarly, in the verb system, the feminine form of "roked" [(he) dances] is "rokedet" and of "yashen" [(he) sleeps] is "yeshenah." Thus, while the subject and the predicate agree in gender and number, they do not have to end with the same specific suffixes. Consequently, although the morphological form of the predicate can be predicted by the morphological form of the subject, its specific morphophonological form is not unequivocal. In summary, targets could neither be activated by semantic network connections nor could they be predicted or easily guessed on the basis of the sentential context.

The children in both reading groups were sampled from the seventh and the eighth grades in junior high schools. This age was selected in order to avoid the possibility that slow maturation of language functions might account for either the reading disorders or the syntactic inferiority of the reading-disabled children. A reading level control group was not examined in the present study for the following reason. A major justification for including control groups matched on reading level, is to dissociate performance differences between good and poor readers that are related directly to reading disability from differences that may be accounted for merely by differences in reading experience. This dissociation was irrelevant in the present study in which we have examined the processing of the most basic syntactic rules in Hebrew. These skills are mastered by all native speakers well before they learn to read and, therefore, there is no theoretical basis for the assumption that reading experience per se could have influenced performance in the word identification task (see also Shankweiler et al., 1992). Furthermore, using an auditory rather than visual word presentation we have technically minimized the possibility that reading-experience biased the children's performance in any way.

EXPERIMENT 1

In our recent study with fluently reading adults (Deutsch & Bentin, 1994) we demonstrated that the syntactic context effect on the identification of target words, masked by white noise, reflects both the facilitation of syntactically congruent targets and the inhibition of syntactically incongruent targets. The existence of these two separate processes provided the necessary empirical ground for our previous interpretation that the difference between good and disabled readers, which was restricted to incongruent targets, reflected a difference in the inhibitory process (Bentin et al., 1990).

The purpose of the present experiment was to extend these previous results, examining how the inhibitory and facilitatory components of the syntactic context effect interact with reading ability in children. To achieve this goal we compared the pattern of the syntactic context effect relative to a neutral base-line, in children with reading disorders and in good readers. On the basis of our previous results (Bentin et. al., 1990), we predicted that, while the percentage of correct identification of targets in the neutral condition should be equal in the two groups, an incongruent syntactic context should inhibit the identification in good readers more than in disabled readers, whereas the facilitatory effect of a congruent context should not interact with reading ability.

Method

Subjects. The "good readers' group": The good readers were 24 children (10 girls and 14 boys), selected from 39 children in ordinary junior high school classes. Their mean age was 13 years and 6 months (s.d = 10 months). Their mean IQ (based on Raven - see below) score was 113.4 (s.d = 9.9). In order for a child to be included in the good readers' group he or she had to read at least 20 pseudowords correctly (with at most 16% errors) and at least 32 sentences (with no more than 11.11% errors). For a detailed description of the reading tests, see section: "Tests and Materials." *The "disabled readers' group"*: The reading-disabled were 24 children (4 girls and 20 boys), selected from a population of 107 children attending "compensatory learning settings" (special classes for learning-disabled children attending regular schools). Their mean age was 13 years and 5 months (s.d. = 13 months). Their mean performance IQ score was 96 (s.d = 8.6). In order to be included in the disabled readers' group a child had to make at least 8 errors (33.3%) in reading pseudowords and at least 8 errors (22%) in reading sentences. In addition his or her mean reading time in both tests had to be at least twice as long as the mean reading time of the good readers' group.

The mean performance of the two groups in each of the reading tests is presented in Table 1.

Tests and Materials

A. Reading tests:

1. Test of phonemic deciphering ability: This test contained a set of 24 meaningless three or four letter strings (pseudowords) presented with vowel marks. All Hebrew consonant-letters and vowel-marks were used in constructing 24 pseudowords which were structured to comply with the Hebrew morpho-phonemic rules. They were presented one at a time on a computer screen subtending a mean visual angle of 2.85 degrees. Each trial consisted of the following events:

Table 1. *Reading performance of the two reading groups tested in Experiment 1.*

READING ABILITY	READING POINTED NONWORDS		READING UNPOINTED SENTENCES	
	Percentage of errors	Time per item (sec)	Percentage of errors	Time per item (sec)
Good readers	2.5%	1.8	2.0%	2.5
Reading-disabled	12.7%	4.2	11.4%	8.6

First a fixation mark was presented at the center of the screen for 500 ms simultaneously with a warning beep. 600 ms from the offset of the fixation mark, a pseudoword substituted for the fixation mark and remained on the screen until a response was made. The subjects were instructed to read each pseudoword aloud exactly as it was written. Reading accuracy was recorded, as well as reading time from stimulus onset. Self-corrections of initially wrongly read stimuli were recorded but not included in the count of correct responses. Reading time was always measured to the first response. Responses with latencies longer than 2 SD from the subject mean were counted as errors. Each subject was assigned two scores: the percentage of correct responses and the average reading time.

2. Reading unpointed words in context: This test contained 36 four-word sentences presented without the vowel-marks. (With the exception of prayer books, poetry and children's books Hebrew is generally printed without vowels. By the end of the fourth grade, children are expected to read "unpointed" print fluently). The last word in each sentence was always a noun, designated as the target. There were two target categories: In one, the targets were heterophonic homographs, i.e., consonant clusters each of which could be combined with several different vowel patterns to form several different words. In the absence of vowels, the correct reading of heterophonic homographs in a particular sentence context can be determined only by apprehending the meaning of the sentence. The targets in the second category were unambiguous words, i.e., consonant clusters each of which could take only one vowel pattern. Thus, even in the absence of vowel-marks each target in this category could be meaningfully read in only one manner. Among the 36 sentences, 24 ended with a heterophonic homographic target and 12 with a phonologically unambiguous target. Each of the homographic targets was presented in two sentences. In one, the phonological alternative implied by the context had a low word-frequency value, while in the other the context implied the reading of a high-frequency word. The 36 sentences were presented in quasi-random order, in which sentences containing alternatives of one heterophonic homograph target were separated by at least two other sentences.

As in the pseudoword reading test, each trial began with the presentation of a fixation mark and a warning tone; these were followed by a sentence centered around the fixation point. The subjects were instructed to read each sentence aloud. Reading time was measured for each sentence from its onset to the end of reading. Subjects' responses were first coded as correct or incorrect. Incorrect responses were further categorized according to four error types, but a detailed description of reading errors is beyond the scope of the present report. (The sub-types of the incorrect responses were: Type 1: The subject substituted one or more words so as to form a syntactically correct meaningful sentence. Type 2: The subject made errors while reading one or more words in the sentence, but read the target word correctly. Type 3: The subject substituted an incorrect phonological alternative of a heterophonic homograph for the one implied by the context. Type 4: The subject was unable to read the sentence). As in the pseudowords reading test, responses with latencies longer than 2 SD from the subject mean were counted as errors.

B. Intelligence tests. An IQ score for each child was obtained either using the performance score of the WISC (whenever those data were available) or by converting to IQ the subject score on the Raven's Progressive Matrices.

C. The syntactic context effect test. Syntactic context effects were assessed using an auditory word identification test similar to that described in our previous studies (Bentin et al., 1990; Deutsch & Bentin, 1994). The test contained 60 sentences. Each sentence included a clearly presented syntactically congruent context phrase followed by a target masked by white noise. The syntactic congruity between the target and the context was manipulated to form three congruity conditions: 1. "Congruent," in which the target word fit the syntactic structure of the sentence. 2. "Incongruent," in which the target word did not fit the syntactic structure of the sentence, that is, caused a violation of a syntactic rule. 3. "Neutral," in which the context was "The next word will be...."

The syntactic violations were constructed by changing the congruent sentences in one of the following ways.

Type 1: Violation of the gender agreement between subject and predicate. For example, the Hebrew sentence "Hasachkan harazeh yashen" (the skinny actor is sleeping) includes a noun, "hasachkan" (preceded by the definite article "Ha"), as the subject, an adjective, "harazeh" (also preceded by the definite article), as the attribute, and a verb, "yashen," as the predicate. In the congruent condition (the above sentence) both the subject and the predicate are in the masculine singular. In the incongruent condition, the same masculine predicate form was presented in a sentence in which the subject and the attribute were in the feminine: "Hasachkanit haraza yashen." (Note that according to another agreement rule in Hebrew the attribute agrees with the subject in gender, number and definite article. In all our examples syntactic structure was kept intact until the last target word.) This category included 12 target words repeated across the three context conditions, forming a total of 36 sentences. In the incongruent condition a masculine subject was presented with a feminine predicate (in 6 of the sentences) or a feminine subject was presented with a masculine predicate (in the other 6 sentences).

Type 2: Violation of the agreement in number between subject and predicate. For example, in the sentence "Hamechonit hayafa yekara" (The nice car is expensive) the feminine, singular predicate form, "yekara," agrees with the feminine singular subject, "mechonit." A violation of the number agreement might be "Hamechonyot hayafot yekara," where the same target is presented with a feminine plural subject (and attribute). Twelve target words (different from those in Type 1) were repeated across the three conditions, forming 36 sentences. In the incongruent condition a singular predicate followed a subject in the plural form (in 6 of the sentences), or vice versa (in the other 6 sentences).

Type 3: Violation of both gender and number agreement between subject and predicate. For example, in the congruent sentence "Harakdan hamefursam mitragesh" (The famous dancer is anxious) the masculine singular predicate, "mitragesh," is in agreement with the masculine singular subject, "harakdan," whereas in the incongruent sentence "Harakdaniyot hamefursamot mitragesh," the same masculine singular predicate relates to a feminine plural form, "harakdaniyot." This category also included 12 target words (different from those in types 1 and 2), which were repeated across conditions to form 36 sentences. In the incongruent condition the gender and number compatibility between subject and predicate was altered in each sentence. For example, a masculine singular subject might be followed by a feminine plural predicate. (We constructed all 4 possible combinations, with 3 sentences for each.)

Type 4: Decomposition of the conjunctive form of preposition and pronoun. This category included 8 target pronouns, each of which was combined with a different preposition, forming 24 sentences. In Hebrew, when a preposition precedes a pronoun, the two are always in a conjunctive form. Thus, in the incongruent condition, the conjunctive form was decomposed into its two elements. For example the conjunctive form "alecha" ("on you") was presented as two separate words: "al" (the preposition "on") and "ata" (the pronoun "you"). In the neutral condition the targets were presented as normal conjunctions.

The sentences of types 1 to 3 consisted of three words in the following order: subject, attribute, predicate. The masked target was always the predicate. All the words used to construct these sentences were basic in childrens' vocabulary. The predicate was either a verb or an adjective (participle form in nominal clauses). Type 4 sentences consisted of a subject, a predicate and a verbal completion (the conjunctive pronoun). The masked targets were the verbal completion in their normal conjunctive form (congruent and neutral conditions) or decomposed (the incongruent condition).

The sentences were organized into 3 lists of 60 sentences, 20 in each congruity condition. Each group of 20 included 4 sentences of each Types 1 to 3, and 8 sentences of Type 4. The targets in the sentences of Types 1 to 3 were rotated so that each subject saw each target only once but, across subjects, each target appeared in each congruity condition. Because the number of pronouns is small, the rotation of pronouns between congruity conditions was within subjects, so that each appeared 3 times in a list (once in the decomposed form). In order to avoid as far as possible the effect of repeating the context, a different sentence was used in each condition. Moreover, the contexts were counterbalanced across the three lists.

All the sentences were recorded on tape by a female, native in Hebrew and professional radio speaker. The tapes were digitized at 20 Khz and edited as follows. The duration of the mask was equal in all sentences, determined by the duration of longest target (750 ms). The white noise was digitally added to the target, starting slightly before onset with a signal-to-noise ratio of 0.35. This ratio was chosen in our previous study with adult fluent readers (Deutsch & Bentin, 1994) on the basis of pilot tests, so that the expected correct target identification level was about 50%.

The sentences in each list were randomized and output to tape at a 2 second intersentence interval.

Procedure. The children were tested individually in two sessions which were run consecutively. In the first session the reading and intelligence tests were administered. Only children who met the selection criteria were tested in the second session for syntactic context effects.

In the syntactic context effects test each child was randomly assigned to one of the three test lists. In each reading group, each list was used to test 8 children. The experimenter and the subject listened to the stimuli simultaneously, using two sets of interconnected earphones (HD420). The children were instructed to listen to the sentence and to repeat the last (masked) word during the silent interval at the end of each sentence. No time constraints were imposed; in a few instances, when the subject's response was delayed past the intersentence interval, the experimenter stopped the tape recorder. The responses were recorded verbatim by the experimenter and no feedback was provided. The experimental session began with 12 practice trials (4 sentences in each condition), followed by the test list.

Results

Subjects' responses were initially coded as correct responses (accurate identification of the inflected word) or errors. The errors made in the incongruent condition were further categorized into four types: 1) "Spontaneous syntactic correction" (making a correction of the syntactic violation) 2) "Logical substitution" (reporting a different word, yet forming a semantically and syntactical congruent sentence); 3) "Nonsense" (replacing the target with a word or nonword which yielded a meaningless sentence); 4) "No response" ("I don't know"). In the neutral and congruent conditions only the last three categories were possible.

Informal inspection of the percentage of correct identification in the different congruity conditions revealed that syntactic congruity had a very similar effect with all four types of violation. Moreover, formal statistical analyses of ten different types of syntactic violations (in a previous study) also revealed that these four violations were equally affected by syntactic context (Bentin et al., 1990). Therefore, identification performance was collapsed across sentence types.

The percentage of correct identification of target words in the congruent, neutral and incongruent grammatical conditions, across subjects and stimuli, for the good and reading-disabled groups are presented in Table 2.

A two-factor analysis of variance with subjects (F1) and stimuli (F2) as random variables revealed a significant interaction between syntactic congruence and reading group [$F1(2,92) = 8.48$, MSe = 74, $p < .0004$, $F2(2,118) = 3.26$, MSe = 261, $p < .0418$]. This interaction reflected the smaller effect of syntactic context in the disabled than in the good readers' group.

Table 2. *The percentage of correctly identified target (SEm) in each congruity condition in the two reading groups.*

CONGRUITY CONDITION	GOOD READERS	READING-DISABLED
Congruent	68.96% (10.2)	52.92% (14.3)
Neutral	39.17% (11.8)	29.17% (12.6)
Incongruent	15.00% (10.6)	13.33% (8.8)

The main effects of syntactic congruence and reading group were also statistically significant. Across groups, the percentage of correct target identification was highest in the congruent condition (61%), second in the neutral condition (34%) and lowest in the incongruent condition (14%) [$F1(2,92) = 358.55$, MSe = 74, $p < 0.0001$, $F2(2,118) = 71.78$, MSe = 926, $p < 0.0001$]. Across congruence conditions, the percentage of correct identification was higher for the good readers (41%) than for the reading-disabled (32%) [$F1(1,46) = 12.16$, MSe = 252, $p < .0011$, $F2(1,59) = 22.8$, MSe = 313, $p < .0001$)]. Post-hoc comparisons (Tukey-A) revealed, however, that the difference between the two groups was statistically reliable only in the congruent and neutral conditions (16.04% and 10%, respectively); in the syntactically incongruent condition the difference between the reading-disabled and good readers in percentage of correctly identified targets (1.67%) was not statistically reliable (HSD = 7.04).

Because the two reading groups differed considerably in IQ level, we examined the possibility that our results were tainted by differences in general intelligence level. First we matched selected groups of good and disabled readers on IQ level and calculated the performance of these subgroups separately. Second, within each reading group children with relatively high and relatively low IQ levels were selected and their performance on the word identification task was calculated separately. As is evident in Table 3, none of these manipulations changed the general pattern of the differences. Moreover, the influence of IQ on the absolute levels of performance was far from dramatic suggesting that, indeed, the identification performance was independent of IQ.

The distribution of errors in the reading disabled and control groups is presented in Table 4.

Table 3. *Percentage of correct identification in each congruity condition for each reading group, split by IQ level.*

	READING DISABLED				GOOD READERS			
	IQ	Congruent	Neutral	Incongruent	IQ	Congruent	Neutral	Incongruent
Whole Group	96	53%	29%	13%	113	69%	39%	15%
IQ Matched	104	48%	25%	13%	106	64%	38%	14%
Low IQ	92	58%	31%	14%	106	64%	38%	14%
High IQ	104	48%	25%	13%	123	71%	41%	14%

Table 4. *The distribution of errors among the different types in each congruity condition for good readers and reading-disabled (Mean percentage and SEm).*

CONGRUITY CONDITION	READING ABILITY	ERROR TYPE			
		Spontaneous correction	Logical substitution	Nonsense	No response
CONGRUENT	Good Readers	-	44.5% (28.4)	1.3% (4.4)	54.3% (29.8)
	Reading-disabled	-	55.4% (24.8)	12.9% (13.6)	31.6% (28.2)
NEUTRAL	Good Readers	-	60.4% (24.0)	0.0% (0.0)	39.6% (24.0)
	Reading-disabled	-	60.0% (20.9	4.4%)(6.6)	31.9% (18.2)
INCONGRUENT	Good Readers	12.5% (7.4)	19.8% (13.8)	1.6% (3.6)	66.0% (19.2)
	Reading-disabled	9.8% (7.8)	34.5% (15.9)	11.1% (10.5)	44.5% (24.4)

Because spontaneous corrections cannot exist in the syntactically congruent and neutral conditions, the statistical evaluation of the distributions was based on a two factor analysis of variance with factor, error type and reading group, within each congruity condition separately. These analyses showed a significant effect of error type in all congruity conditions [$F(2,92) = 30.63$, MSe = 827, $p < .0001$, $F(2,92) = 85.23$, MSe = 478, $p < .0001$, and $F(3,138) = 85.07$, MSe = 275, $p < .0001$ for the congruent, neutral and incongruent conditions, respectively]. A significant interaction was found between the effects of error type and reading group in the congruent condition [$F(2,92) = 5.56$, MSe = 827, $p < .0052$] and in the incongruent condition [$F(3,138) = 11.35$, MSe = 275, $p < .0001$] but not in the neutral condition [$F(2,92) = 0.93$, MSe = 478, $p > .3997$].

As is evident in Table 4, the percentage of "no response" errors was higher for the good readers than for the reading-disabled. This pattern was found in the syntactically congruent condition (54% and 32% for good and reading-disabled, respectively) as well as in the incongruent condition (66% and 45% for good and reading-disabled, respectively). Post hoc analyses of the interactions revealed that these differences were statistically significant in the incongruent condition (HSD = 18.42) but not in the congruent condition (HSD = 24.23).

Although a formal analysis of the Error type × Reading group × Congruency condition interaction could not be made, it is worth noting that the two-way interaction between error type and reading group is caused by opposite trends in the congruent and incongruent conditions. In the congruent condition disabled readers made fewer "no response" than "logical substitution" errors, while good readers showed an inverse tendency. In the incongruent condition, on the other hand, both groups made more "no response" than "logical substitution" errors, while the difference was considerably larger for good readers than for reading-disabled.

Discussion

The results of the present experiment showed that, for children as for adults (Deutsch & Bentin, 1994), the syntactic context effect on the identification of auditory masked words reflects two processes, facilitation and inhibition. Both processes were effective in disabled and in good readers. However, reading ability influenced the size of the relative contribution of each of these two processes to the global syntactic context effect.

A post-hoc analysis of the interaction of reading ability and syntactic context revealed that for good readers syntactic incongruence reduced the percentage of correct target identification (relative to the neutral condition) almost as much as syntactic congruence elevated this percentage (24% and 29%, respectively). In contrast, for disabled readers the relative contribution of the inhibitory process to the global syntactic context effect (15%) was significantly smaller than that of the facilitatory process (24%). Although disabled readers identified fewer targets than good readers in the syntactically congruent and neutral conditions, the performance of the two reading groups was similar in the incongruent condition. This pattern is in agreement with our previous findings, which suggested that incongruent syntactic context interferes less with the performance of reading-disabled than with that of good readers (Bentin et al., 1990).

In contrast to our previous findings (Bentin et al., 1990), in the present experiment the correct identification rate across syntactic congruity conditions was lower for reading-disabled (32%) than for good readers (41%). Note, however, that the overall identification performance of good readers was also poorer than that of adults (51%; Deutsch & Bentin, 1994). The reduction in identification performance occurred even though an identical masking intensity, identical stimuli and experimental procedures were used in both studies. It is possible that the masking conditions, which had been calibrated for adults, were too difficult for children, and that this difficulty was more conspicuous for disabled readers possibly because they may have had a phonological disability as well (Brady et al., 1983). It is possible therefore that the similarity between disabled and good readers in the identification of syntactically congruent targets reported in Bentin et al. (1990) reflected a reduced level of masking which was not very sensitive to phonological impairments. However, if the relatively inferior phonological ability of the reading disabled children had been the only factor accounting for the differences between the identification performance of the two groups, no interaction between reading level and syntactic congruency condition should have been observed. We had the opportunity to test this interpretation in Experiment 2, where the noise level was reduced.

The analysis of the error distribution yielded additional insights. Consider first the relatively small percentage of "spontaneous syntactic corrections" which was observed in both reading groups. Because only verbatim accurate responses were considered correct, the pattern of facilitation and inhibition might simply reflect that the subjects, facing uncertainty, used some partial phonological information extracted from the noise and the contextual information, to guess the target word. If this interpretation were correct, the difference in the percentage of correct identifications of inflected targets in the congruent and the incongruent conditions would reflect the correspondence or disagreement between the subject's intuition about how the identified word should have been inflected and what was actually presented. Such a strategy, however, would result in a high percentage of "spontaneous syntactic correction" errors in the incongruent condition. But this did not occur: The relatively small percentage of "spontaneous syntactic corrections" in both groups indicates that the pattern of masked-word identification in our subjects did not simply reflect an intelligent guessing strategy based on partial input.

The difference between the distributions of the different types of errors in the two reading groups provided additional support for the view that the reading-disabled children were generally less affected by the syntactic context, and, in particular that their performance was less impaired by syntactic incongruence than that of the good readers. In both the congruent and the incongruent conditions good readers produced more "no response" errors than reading-disabled, and disabled-readers produced more "logical substitution" errors than good readers. Within each group, the percentage of "no response" errors was larger in the incongruent than in the congruent condition. On the other hand, the percentage of logical substitutions was smaller in the incongruent than in the congruent condition. The abundance of "no response" errors for good readers, especially in the incongruent condition, suggests that children may have chosen to abstain from responding when facing uncertainty. This strategy was more appropriate in the incongruent than in the congruent condition because in the former condition the uncertainty caused by masking was increased by the mismatch between the partial information provided by the phonetic input and general expectations raised by the syntactic structure of the sentence context. The fact that reading-disabled produced fewer "no response" and (context-unrelated) substitution errors than good readers also supports our suggestion that the disabled readers' identification performance was less affected by the syntactic context than that of good readers.

In summary, the results of the present experiment showed that, for both good readers and reading-disabled, the identification of auditory masked targets presented in sentences is affected by the syntactic structure of the context. However, this syntactic context effect is reduced in reading-disabled, primarily because their performance is less impaired by syntactic incongruity. In our previous study (Deutsch & Bentin, 1994) we suggested that the inhibitory component of the syntactic context effect is mediated by attention. Consequently it is possible that the difference between the reading-disabled and the good readers reflects a deficient attention-mediated syntactic process in reading-disabled. Experiment 2 was designed to test this hypothesis.

EXPERIMENT 2

In the present experiment we tested the interaction between reading ability and the effect of attention-related mechanisms which mediate the inhibitory component of the syntactic context effect (Deutsch & Bentin, 1994). Specifically, we compared the effect of presenting the congruent and incongruent conditions in separate blocks, as opposed to random mixed presentation, for both disabled and good readers.

In our previous study (Deutsch & Bentin, 1994) we showed that when the congruent and incongruent conditions are presented in separate blocks the inhibitory component of the syntactic context effect was attenuated while the facilitatory component was not affected. In line with previous interpretations of similar effects on semantic priming (e.g., Fischler & Bloom, 1985; Stanovich & West, 1983; Tweedy, Lapinsky & Schvaneveldt, 1977), we suggested that the blocking manipulation primarily affects an attention-based mechanism which may be reflected more in the inhibitory than in the facilitatory component of the syntactic context effect. Blocking syntactically incongruent targets may, for example, discourage the elaboration of (automatically) generated syntactic expectations based on the structure of the context. Thus we used the blocking manipulation to disentangle the attention-related factors involved in the syntactic context effect from the more automatic factors, and to examine how each of these factors interacts with reading ability.

Our hypothesis was that the observed difference in syntactic performance between reading-disabled children and good readers reflects the malfunctioning of an attention-based mechanism for processing syntax. On the basis of this hypothesis we predicted that the effect of blocking the congruency conditions would be stronger in good readers than in disabled readers. Specifically, we predicted that discouraging the use of expectations by presenting the ungrammatical sentences in one block would decrease the amount of inhibition for the good readers, while reading-disabled would be significantly less affected by this manipulation.

In the present experiment we also tested our assumption that the inferior identification performance of the reading-disabled relative to good readers in the congruent condition of Experiment 1 was caused by a too high level of masking. To this end, the percentage of correct identification in the mixed presentation condition was compared between reading groups, using a higher signal to noise ratio. If our interpretation is correct, then the results reported by Bentin et al. (1990) should be replicated, i.e., the two groups should identify a similar percentage of congruent targets, while the disabled readers should identify more incongruent targets than the good readers.

Method

Subjects. The subjects were 120 children who had not taken part in the first experiment. They included 60 good readers (27 girls and 33 boys), and 60 disabled readers (13 girls and 47 boys), selected from seventh and eighth graders attending regular classes, using the same selection criteria as in Experiment 1 (Table 5). The mean age of the good readers was 13 years, with a mean IQ score of 116 (s.d = 10.6). The mean age of the reading-disabled children was 13 years and 2 months, with a mean IQ score of 105 (s.d = 11.5).

Test and Materials. The sentences were those used in Experiment 1, with the exception of the neutral stimuli. Thus each stimulus list consisted of 40 sentences, 20 syntactically congruent and 20 syntactically incongruent. In the "mixed" presentation the 40 sentences were randomized and presented in one set of stimuli. In the "blocked" presentation congruent and incongruent sentences were clustered separately in two blocks of 20 sentences each. The sentences were randomized within each of the two blocks. Each target appeared only once in each list (with the exception of sentences of Type 4). Across lists, each target appeared equally often in the congruent and incongruent conditions.

The intensity of masking was lowered from that of Experiment 1 by increasing the signal-to-noise ratio from 0.35 to 0.40.

Procedure. Different groups of 30 good readers and 30 disabled readers were tested with each presentation condition. The assignment of subjects to experimental conditions was random. The procedure used for the mixed presentation was identical to that used in Experiment 1.

Table 5. *Reading performance of the two reading groups tested in Experiment 2.*

READING ABILITY	READING POINTED NONWORDS		READING UNPOINTED SENTENCES	
	Percentage of errors	Time per item (sec)	Percentage of errors	Time per item (sec)
Good readers	2.4%	2.1	2.4%	2.5
Reading-disabled	12.0%	4.2	10.2%	7.6

In the blocked presentation, 15 subjects began with the block of syntactically congruent sentences, and the other 15 with the block of syntactically incongruent sentences. Each block was preceded by 8 practice sentences in the respective congruity condition. No special instructions were given before the "incongruent" block, but the ungrammatical structure of the sentences was not denied in reply to occasional queries raised by the subjects following practice with ungrammatical sentences (as was true for the mixed condition as well).

Results

The percentage of correct identification was averaged for each subject and target in each congruity condition. The reading groups did not differ in the percentage of correct identification of congruent targets either in the mixed or in the blocked presentation (about 60% correct). In contrast, the disabled readers differed from the good readers in the percentage of correct identification of target words in the incongruent condition. This difference, however, was influenced by the mode of presentation. The percentage of correct target identification in the incongruous condition was higher for reading-disabled than for good readers. This difference was particularly conspicuous in the mixed condition. Disabled readers identified 20% of the incongruent targets regardless of whether the congruency conditions were mixed or blocked. In contrast, good readers identified twice as many targets when incongruent targets were blocked (16%) than when incongruent and congruent targets were mixed (8%) (Figure 1).

The statistical significance of these differences was examined by a mixed model three factor analysis of variance, with subjects (F1) and stimuli (F2) as random variables. The between-subjects factors were reading ability and presentation mode, while the within factor was congruity condition. The influence of reading ability on the interaction between the syntactic congruity condition and presentation mode was demonstrated by a significant second-order interaction between congruity condition, presentation mode, and reading ability [$F1(1,116)=4.80$, $MSe = 125$, $p<.0305$, $F2(1,118)=5.59$, $MSe = 177$, $p<.0196$].

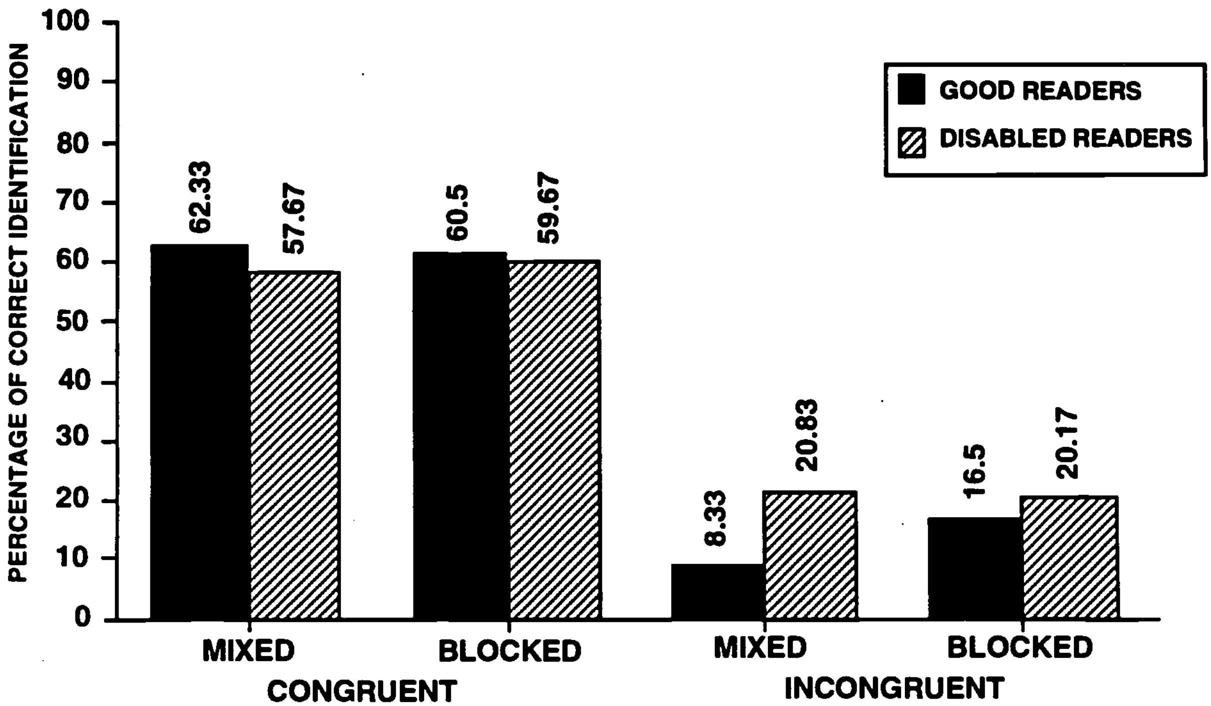


Figure 1. Percentage of correct identification of syntactically congruent and incongruent targets in the blocked and mixed presentation modes.

The syntactic congruity effect in the mixed condition was tested separately by a two factor analysis of variance. This analysis was performed in order to test our hypothesis that the unexpected difference between the two groups in overall identification performance in Experiment 1 was caused by excessive masking. The analysis revealed a significant interaction between reading level and congruity condition [$F(1,58) = 18.93$, $MSe = 112$, $p < .0001$]. Post hoc comparisons showed that this interaction was caused by a significantly higher percentage of correct identification of incongruent targets by the reading-disabled than by the good readers (a difference of 12%), in contrast to the statistically insignificant difference between the groups (less than 5% difference) in the percentage of correct identification of congruent targets ($MSe = 112$, $q(4,58) = 3.75$, $HSD = 5.134$).

The possible influence of IQ level on the revealed pattern of results was examined as in Experiment 1. As is evident in Tables 6 and 7 the pattern of results was apparently independent of IQ level.

The error distribution in the good and disabled readers is presented in Table 8.

A three-way analysis of variance, with error type, reading level and presentation condition as main factors, was performed separately within each congruity condition. This analysis revealed a significant main effect of error type in both conditions ($F(2,232) = 97.49$, $MSe = 550$, $p < .0001$ and $F2(3,348) = 170.07$, $MSe = 255$, $p < .0001$ for the congruent and incongruent conditions, respectively). The second-order interaction of error type, reading group and presentation condition (blocked or mixed) was significant for the incongruent targets [$F(3,348) = 8.47$, $MSe = 255$, $p < .0001$] but not for the congruent targets [$F(2,232) = 1.43$, $MSe = 550$, $p > .2420$].

Examination of the distribution of error types in the incongruent condition across the two presentation conditions indicated that the percentage of substitution of another word for the target word ("logical substitution" or "nonsense") was much higher for the reading-disabled children than for the good readers. In contrast, the percentage of "no response" is much higher for the good readers than for the reading-disabled children. As was previously found with fluently reading adults, the percentage of "no response" was lower in the blocked condition than in the mixed condition for the good readers, while an opposite trend was observed the reading-disabled.

Table 6. *Percentage of correct identification in each congruity condition for each reading group, split by IQ level, within the mixed presentation mode.*

	READING DISABLED			GOOD READERS		
	IQ	Congruent	Incongruent	IQ	Congruent	Incongruent
Whole Group	105	58%	20%	116	62%	10%
IQ Matched	106	57%	23%	103	62%	8%
Low IQ	92	57%	22%	103	62%	8%
High IQ	116	58%	21%	126	61%	6%

Table 7. *Percentage of correct identification in each congruity condition for each reading group, split by IQ level. within the blocked presentation mode.*

	READING DISABLED			GOOD READERS		
	IQ	Congruent	Incongruent	IQ	Congruent	Incongruent
Whole Group	105	60%	20%	116	61%	17%
IQ Matched	109	61%	23%	113	68%	21%
Low IQ	99	58%	26%	113	68%	21%
High IQ	109	61%	23%	128	59%	19%

Table 8. *The distribution of errors among the different types in each congruity condition in the mixed and blocked presentation mode for good readers and reading-disabled.*

CONGRUITY CONDITION	READING ABILITY	MODE OF PRESENTATION	ERROR TYPE			
			Spontaneous correction	Logical substitution	Nonsense	No response
CONGRUENT	Good Readers	Mixed	-	39.6%	3.2%	57.2%
		Blocked	-	51.6%	14.3%	33.5%
	Reading-disabled	Mixed	-	45.8%	2.4%	52.8%
		Blocked	-	46.1%	15.7%	37.1%
INCONGRUENT	Good Readers	Mixed	11.4%	17.7%	1.8%	69.1%
		Blocked	12.7%	25.1%	7.0%	55.0%
	Reading-disabled	Mixed	11.7%	31.2%	18.9%	35.3%
		Blocked	11.3%	27.4%	14.7%	46.7%

Discussion

The most interesting result in Experiment 2 was that the effect of presenting the syntactically incongruent sentences in a separate block as opposed to mixing them with congruent sentences was different in good and disabled readers. Whereas for the good readers there was less interference of syntactic incongruity with target identification in the blocked presentation mode than in the mixed presentation mode, disabled readers were not affected by this manipulation. Moreover, the amount of inhibition in the mixed presentation mode was smaller for the reading-disabled than for the good readers. In contrast to incongruent targets, syntactically congruent targets were equally well identified by both reading groups, and this performance was not affected by the mode of presentation (blocked vs. mixed).

The smaller inhibitory effect in the blocked than in the mixed presentation mode, and the absence of any effect of presentation mode on the facilitatory component of the syntactic context effect—both of which were observed among good readers—replicates a similar pattern found in fluently reading adults (Deutsch & Bentin, 1994). This pattern supports the hypothesis that the inhibitory component of the syntactic context effect is controlled by attention-mediated mechanisms while the facilitatory component of the syntactic context effect is more automatic.

The absence of any influence of presentation mode on the identification performance of disabled readers suggests that they were either less sensitive than good readers to the syntactic structure of the sentence (and therefore less disturbed by syntactic incongruity), or that they did not use that information to generate a performance strategy. The similar amount of facilitation in the syntactically congruent condition observed in the performance of disabled and good readers indicates that the second interpretation is more plausible than the first. We will elaborate this interpretation in the general discussion.

The equally good performance of disabled and good readers with syntactically congruent targets is noteworthy also because it supports our account for the unexpectedly poorer performance of disabled relative to good readers found in Experiment 1 across all congruity conditions. We assumed that this inferiority was caused by too intense masking. Indeed, when the signal-to-noise ratio was increased in the present experiment in comparison to that in Experiment 1 (i.e., the amount of masking was reduced), the general pattern of differences between the two reading groups replicated the pattern found by Bentin et al. (1990): Disabled readers identified as many targets as good readers in the congruent condition, and more targets than good readers in the incongruent condition. Thus, assuming that intense auditory masking affected disabled readers more than good readers, the smaller difference between the two groups in the incongruent condition than in the congruent condition which was observed in Experiment 1 (despite the high-intensity masking) may have reflected the same underlying mechanism suggested by the results of Experiment 2, i.e., that the identification of target words is less inhibited by syntactic incongruity in disabled than in the good readers.

Additional support for our account of the differential effect of presentation mode on good and disabled readers was provided by the pattern of errors. As we found in the analysis of incorrect responses, the two-way interaction between the type of errors that subjects made, presentation mode and reading ability was significant only in the incongruent condition. These results revealed that although good readers produced significantly more "no response" responses than poor readers, this trend was influenced in both groups by the mode of presentation. However, the manipulation had a different effect in each reading group. In the blocked presentation condition good readers produced "no response" errors less often than in the mixed presentation, but they had an increased tendency to commit substitution errors of various kinds. In contrast to the case of good readers, mode of presentation had no effect on disabled readers. The types of errors made by disabled readers in both conditions resembled the pattern found in good readers in the blocked presentation mode, i.e., there was a relatively high percentage of substitutions and a relatively low percentage of "no response" errors.

Recall that, as discussed in Experiment 1, "no response" may reflect the inhibition caused by the conflict between context based expectations and phonological input. Therefore, the fact that good readers produced a smaller proportion of "no response" errors in the blocked presentation condition than in the mixed presentation might have been the result of a strategic decision that the context is not very helpful in the target identification process, in the block of incongruent sentences and can therefore be ignored. Using the same logic, the finding that reading-disabled children produced fewer "no response" among their errors than good readers may reflect the fact that they did not rely as much on context-based expectations in either presentation mode. Although less informative, the pattern of substitution errors can also be integrated into the above interpretation. It is possible that when conflict is reduced (in the blocked presentation), subjects can more easily adopt a less conservative strategy and release intuitive responses.

In conclusion, the results of the present experiment supported our previous findings that syntactic incongruity disturbs reading-disabled children less than good readers, and revealed that one source of this difference is a reduction in the efficiency of an attention-based inhibitory component in the syntactic context effect. We will elaborate this mechanism and its possible implications for understanding reading disability in the next section.

GENERAL DISCUSSION

The present study was aimed at further investigating the basis of the difference in the ability of reading-disabled and good readers to use the syntactic information conveyed by a sentence context in word identification (Bentin et al. 1990). More specifically we examined how attention-related mechanisms may be a source of this difference. To achieve this goal, we have tested the interaction between reading ability and attentional mechanisms that mediate the syntactic-context effect on word identification. Auditory masked target words were embedded in an unmasked sentential context, and were either congruent or incongruent with the syntactic structure of the sentence. The results revealed that, as compared to a neutral condition, word identification was facilitated by syntactic congruence and inhibited by syntactic incongruence in both good readers and the reading-disabled children and that, as was predicted, the effect of inhibition was smaller in the latter than in the former group.

Before discussing the possible interpretations of these results it is worth mentioning that the absence of any relationship between general intelligence level and the word identification performance in any of the reading groups studied, supports previous claims that reading skills and intelligence are not closely related (Baddeley, Logie & Allis, 1988; Brady, Shankweiler & Mann, 1983; Fowler, 1988: Shankweiler, Crain, Brady & Macaruso, 1992; Siegel, 1988; Stanovich, 1991; Stanovich, Cunningham & Feeman, 1984). Discussing this issue Stanovich and his colleagues also showed that the correlation between reading and intelligence is limited to reading comprehension. On the basis of extensive research they concluded that intelligence scores may account for the performance of garden-variety type of poor readers, but it is less informative in accounting for the inferior linguistic performance of children with severe reading disorders which are based on specific phonological handicap.

The manipulation of attention related strategies, by presenting congruent and incongruent sentences in separate blocks, had a differential effect on the two reading groups. For good readers, this manipulation affected the magnitude of the inhibition but had no effect on facilitation. These results replicated our previous findings with fluent adult readers (Deutsch & Bentin, 1994), suggesting that attention mechanisms mediate mainly the inhibitory component of the syntactic context effect. In contrast to the performance of good readers, the performance of the reading-disabled children was not affected by the blocked/mixed manipulation in the identification of either congruent or incongruent targets. Moreover, the percentage of correct identification of incongruent targets was relatively high in the mixed as well in the blocked condition, resembling the performance of good readers in the blocked condition.

Our interpretation of the differences in performance between good readers and reading-disabled is based on a conceptualization of the mechanism of the syntactic-context effect that we elaborated in a previous study (Deutsch & Bentin, 1994). In analogy to a commonly held account of attention mediated factors in semantic priming (Fischler, 1977; Fischler & Bloom, 1979; Neely, 1977; Stanovich & West, 1981; 1983), we suggested that the involvement of attention in the syntactic context effect is related to a process of elaborating context based expectations. In an attempt to explain the nature of these expectations, we borrowed a concept put forward to account for the role of attention in the construction of expectations in the semantic domain, and extended it to the syntactic domain. This concept is the assumption of coherence formulated by de Groot, Thomassen & Hudson (1982). According to the coherence assumption, the reader (or the listener) covertly assumes that every linguistic message must be coherent. Consequently, he/she expects each word in the linguistic message to be congruent with the context in which it is embedded. This expectation induces a process of verification that this coherence indeed exists. According to de Groot et al.'s (1982) analysis, this verification is performed at a postlexical level, after a phonological unit has been provisionally identified. This process delays the final identification of the word until the coherence has been verified, and thus its effect on word identification is always inhibitory. However, the effect should be particularly strong when the coherence assumption is not satisfied.

In light of the fact that the syntactic system is more constrained than the semantic system, and supported by the residual inhibition observed in the incongruent block despite its obvious structure, we suggested that the same covert assumption of coherence, which was used in the semantic domain to account only for the inhibition process, may underlie both the inhibition and the facilitation processes in the syntactic domain (for a detailed elaboration of this claim see Deutsch & Bentin, 1994). It was shown that in incongruent contexts the process of inhibition is unavoidable (de Groot et al., 1982). Therefore the mere tendency to generate grammatical expectations and the triggering of the coherence verification cannot easily be controlled. Hence, we suggested that at the sentence level these expectations are probably generated by a veiled controlled (quasi-automatic) process which uses only minimal attention resources (Schneider & Shiffrin, 1977). The term "veiled controlled processes" is used to describe an intermediate stage of attentional processing which is carried out like an automatic process: It is characterized by low demands on attentional resources and is generally carried out without intention (Shiffrin & Schneider, 1977). Congruent targets are facilitated in comparison with a neutral condition because their morphological structure may have been previously activated while generating the expectations, and/or because they may be integrated more easily into a previously activated syntactic structure. On the other hand, when the same expectations are violated by incoherent input, attention is mobilized to control an additional process of reevaluating the basis of the syntactic expectations and/or re-examining the phonological input. The reevaluation may be the attention-mediated factor in the process of inhibition. This account for the syntactic-context effect accommodates the finding that strategic changes induced by the blocking manipulation influenced the magnitude of the inhibition effect without affecting the quasi-automatic facilitation. Moreover, because the generation of expectations is not under strategic control, residual inhibition may also exist when experimental circumstances discourage the initiation of the reevaluation process. This mechanism may account for the inhibition found in the blocked presentation mode.

In the present study we found that the performance of the reading-disabled was facilitated by syntactic congruence as much as that of as good readers. This finding precludes the possibility that the reading-disabled children were simply insensitive to the syntactic structures. Alternatively, this finding may suggest that the coherence assumption, and the uncontrolled generation of context-based expectations, function in reading-disabled children as well as in good readers.

In contrast to the equal facilitation, the inhibition process was different in the two groups. Good readers were inhibited more than the reading-disabled when the congruence conditions were mixed, whereas in the blocked presentation word identification was similarly inhibited in the two groups. The equal amount of inhibition found in the reading-disabled group across presentation conditions suggests that the natural manifestation of the inhibition process in reading-disabled children is similar to its manifestation in good readers when influenced by the artificial conditions created by blocking the incongruent sentences, i.e., when the validity of the coherence assumption was reduced. In our interpretation this residual inhibition is accounted for by the uncontrolled generation of syntactic expectations, while the attention demanding process of reevaluating the syntactic structure is reduced in the reading-disabled.

In accordance with the above analysis, we propose that reading-disabled children are as competent as good readers in assembling appropriate syntactic structures and generating morpho-syntactic expectations while identifying words in context. They are, however, inferior to good readers in using this system for the attention-based process of re-evaluating the generated structure and/or the phonologic input when their expectations are not fulfilled. Note that similar suggestions have been made in the semantic domain (Gernsbacher, 1993; Gernsbacher & Faust, 1991). These authors found that, while poor readers do not differ from good readers in generating context based expectations, they are inferior in suppression inappropriate contextual information.

The distinction between the attention-based and quasi-automatic mechanisms of the syntactic-context effect is similar to the distinction between the ability to comprehend and produce language, on the one hand, and the ability to reflect on this linguistic knowledge and use it intentionally, on the other hand. The latter skills comprise the concept of "linguistic awareness" (Hakes, 1980). Using this concept to interpret our present results, we suggest that, at least in regard to morpho-syntactic rules as manipulated in the present study, the syntactic knowledge of reading-disabled children may be intact, but that they are less competent than good readers in using this knowledge intentionally. Assuming that linguistic awareness develops on the basis of the accumulated knowledge in a particular linguistic domain and on the organization of this knowledge (cf. Hakes, 1980), the absence of syntactic awareness in these children may reflect deficiencies in the nature, quality and organization of the syntactic knowledge they possess. This deficiency should not impair the automatic processes of understanding and producing language, but it is evident in more complex linguistic tasks, such as reading, or in artificial tasks (such as those in the present study). The more complex tasks require the intentional activation of attention-based operations on the basis of this knowledge (for a similar conceptualization see Cupples & Holmes, 1992).

The relatively low percentage of spontaneous syntactic corrections of the incongruent targets in good readers, and the fact that the percentage of these error-types was similar across groups require additional consideration. Because corrections may draw on attentional resources (de Villiers & de Villiers, 1974; Fowler, 1988), this pattern of errors could be interpreted as evidence against our suggestion that good readers are more aware of their syntactic knowledge than poor readers. However, a different interpretation of correction errors is possible in the present context - nammely, that spontaneous corrections may not necessarily reflect syntactic awareness but may be sporadic responses characterizing children's performance in many linguistic tasks even at a developmental stage when linguistic awareness is still absent (Hakes 1980). This pattern is particularly relevant to the present experiment, in which the task was not to correct the sentence grammar (cf. Fowler, 1988) but rather to identify the target words. Thus, the low percentage of "spontaneous correction" errors which was similar in all reading groups and in both the mixed and the blocked presentation modes, may reflect a unintentional process which differs fundamentally from the processes we are interested in which is the ability to intentionally allocate attention to the syntactic structure.

The deficiency in the attention mechanism which according to our interpretation, underlies the impaired use of syntactic context by reading-disabled children is not necessarily a general attention deficit. In our present conceptualization we are considering attention only in its limited role as a vehicle for linguistic awareness. Hakes (1980) described the development of linguistic awareness on the basis of practice and repeated experience with existing knowledge. According to this model, linguistic awareness enables the perceiver of language to gain control of processes which may be triggered automatically. For example, the coherence assumption posits that the generation of the syntactic structure and of the expectations which are based on this analysis are automatic. However, without linguistic awareness and without the ability to direct a minimal amount of attention resources to these expectations, the perceiver would not be aware of their content and would not be able to respond to their violation. As we suggested above, this activity is the source of the attention-based interference caused by syntactic incongruence. This is not to say, however, that the perceiver is necessarily conscious of the process. The control may be veiled, as suggested by Shiffrin and Schneider (1977). For example, while using the syntactic context normally in the perception of congruent sentences, the subject may not be conscious of the expectations developed during the process of word recognition. Furthermore, attention-based control is not an all-or-none ability, and the amount of attention resources it requires may vary according with the complexity of the linguistic message and the subject linguistic sophistication. For example, Jou and Harris (1992) showed that, while incorrectly inflected verbs embedded in a text interfered with reading speed, the performance was improved by increasing the inflection error rate to 100%. This pattern suggests that, although the process of syntactic analysis is motivated by a covert automatized involuntary process based on prior knowledge, this process can be controlled and modulated by attention in novel situations (see also MacLeod & Dunbar, 1988).

The full development of attention-based elaboration of context-based expectations, as well as an accomplished linguistic awareness, probably requires the integrity of the linguistic system. Consequently, the reduced syntactic awareness in the reading-disabled may suggest an impairment in the consolidation and integrity of syntactic knowledge in this population. In the final section of this discussion we consider the role that syntactic awareness might have in reading.

Most reading-related studies of syntactic ability difficulties in the syntactic domain were related to advanced levels of reading comprehension (Bowey, 1986; Willows & Ryan, 1986). On the basis of the present results and our interpretation of them, we suggest that difficulties in syntactic awareness may also be related to primary stages of the reading process, namely the process of word identification (for a similar argumentation see Tunmer, Herriman & Nesdale, 1988). In the process of word identification, whether spoken or written, syntactic context may play a particularly role in resolving categorical ambiguity (e.g., deciding whether RUN is a noun or a verb). Syntactic information may be more important in reading than in speech because of the lack of prosodic cues which facilitate disambiguation in spoken but not in written language, and because there is additional ambiguity in print in the case of heterophonic homographs such as "WIND."

In Hebrew, the language in which the present study was carried out, there is a large incidence of heterophonic ambiguity. In Hebrew orthography most vowels are represented by diacritical marks which are usually omitted from print. This aspect of the orthography creates a situation in which the same sequence of consonants may be read in many ways, each with a different vowel-pattern. Consequently, heterophonic homographs and categorical ambiguity are more frequent in Hebrew than in any in most other languages. In fact even words that have only one meaningful phonological structure cannot be read entirely through the assembled phonology. (For a detailed description of Hebrew orthography and its consequences for reading see Frost & Bentin, 1992.) Given this specific orthographic structure, it is conceivable that the importance of context-based expectations in word identification is amplified in Hebrew. Nonetheless, we believe that our findings are not language-specific. This claim is based on existing evidence which was reviewed in the introduction, but undoubtedly the particular relationship between attention-based mechanisms of syntactic processing and reading disability require additional research.

REFERENCES

Baddeley, A. D., Logie, R. H., & Ellis, N. C. (1988). Characteristics of developmental dyslexia. *Cognition, 29,* 197-228.

Bentin, S., Deutsch, A., & Liberman, I. Y. (1990). Syntactic competence and reading ability in children. *Journal of Experimental Child Psychology, 48,* 147-172.

Bohannon, J. N., Warren-Leubecker, A., & Hepler, N. (1984). Word order awareness and early reading. *Child Development, 55,* 1541-1548.

Bowey, J. A. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology, 41,* 282-299.

Brady, S., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory code in relation to reading ability. *Journal of Experimental Child Psychology, 35,* 345-348.

Brittain, M. M. (1970). Inflectional performance and early reading achievement. *Reading Research Quarterly, 6,* 34-48.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? *Applied Psycholinguistics, 2,* 201-212.

Carello, C., Lukatela, G., & Turvey, M. T. (1988). Rapid naming is affected by association but not by syntax. *Memory & Cognition, 16,* 187-195.

Cupples, L., & Holmes, V. M. (1992). Evidence for a difference in syntactic knowledge between skilled and less skilled adult readers. *Journal of Psycholinguistic Research, 28,* 249-274.

Deutsch, A., & Bentin, S. (1994). Attention mechanisms mediate the syntactic priming effect in auditory word identification. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 395-607.

de Groot, A. M. B., Thomassen, A. J. W. M., & Hudson, P. T. W. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition, 10,* 358-370.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition, 5,* 335-339.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional process in the effects of sentence context on word recognition. *Journal of Verbal Learning and Verbal Behavior, 18,* 1-20.

Fischler, I., & Bloom, P. A. (1985). Effects on constraint and validity of sentence context on lexical decisions. *Memory & Cognition, 13,* 128-139.

Flood, J., & Menyuk, P. (1983). The development of metalinguistic awareness and its relation to reading achievement. *Journal of Applied Developmental Psychology, 4,* 65-80.

Fowler, A. E. (1988). Grammaticality judgment and reading skill in grade 2. *Annals of Dyslexia, 38,* 73-94.

Gernsbacher, M. A., & Faust, M. E. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science, 4,* 294-298.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 245-262.

Goldman, S. R. (1976). Reading skill and the minimum distance principle. A comparison of listening and reading comprehension. *Journal of Experimental Child Psychology, 22,* 123-142.

Goodman, G. O., McClelland, J. L., & Gibbs, R. W. (1981). The role of syntactic context in word recognition. *Memory & Cognition, 9,* 580-586.

Gurjanov, M., Lukatela, G., Moskovljević, J. Savić, M., & Turvey, M. T. (1985) Grammatical priming in inflected nouns by inflected adjectives. *Cognition, 19,* 55-71.

Guthrie, J. T. (1973). Reading comprehension and syntactic response in good and poor readers. *Journal of Educational Psychology, 65,* 294-299.

Hakes, D. T. (1980). *The development of metalinguistic abilities in children.* Berlin: Springler-Verlag.

Jou, J. (1992) Reading inflectionally incongruent texts: the automatic and strategic processes in language performance. *Journal of Psycholinguistic Research, 21,* 365-382.

Katz, L., Boyce, S., Goldstein, L., & Lukatela, G. (1987) Grammatical information effects in auditory word recognition. *Cognition, 25,* 235-263.

Lukatela, G., Kostić, A., Feldman, L. B., & Turvey, M. T. (1983). Grammatical priming of inflected nouns. *Memory & Cognition, 11,* 59-63.

Lukatela, G., Moraca, J., Stojnov, D., Savić, M. D., Katz, L., & Turvey, M. T. (1982). Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research, 44,* 297-311.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71-102.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic process in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84,* 1-66.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre and postlexical loci of contextual effects on word recognition. *Memory & Cognition, 12,* 315-328.

Shankweiler, D., & Crain, S. (1986). Language mechanism and reading disorder: A modular approach. *Cognition, 24,* 139-168.

Shankweiler, D., Crain, S., Brady, S., & Macaruso, P. (1992). Identifying the cause of reading disability. In P. B. Gough and L. C. Ehri (Eds.), *Reading acquisition* (pp. 275-305). Hillsdale, NJ: Erlbaum.

Shankweiler, D., Smith, S. T., & Mann, V. A. (1984). Repetition and comprehension of spoken sentences by reading-disabled children. *Brain and Language, 23,* 241-257.

Shanon, B. (1973). Interpretation of ungrammatical sentences. *Journal of Verbal Learning and Verbal Behaviour, 12,* 389-400.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Siegel, L. S., & Ryan, E. B. (1984). Reading disability as a language disorder. *Remedial and Special Education, 5,* 28-33.

Smith, S. T., Macaruso, P., Shankweiler, D., & Crain, S. (1989). Syntactic comprehension in young poor readers. *Haskins Laboratories Status Report on Speech Research, SR-97/98,* 147-169.

Stanovich, K. E. (1991). Discrepancy definitions of reading ability. *Reading Research Quarterly, 26,* 8-29.

Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984). Relation between early reading acquisition and word decoding with or without context: A longitudinal study of first-grade children. *Journal of Educational Psychology, 76,* 668-677.

Stanovich, K. E., & West, R. F. (1981). The effects of sentence context on ongoing word recognition: Tests of two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 658-672.

Stanovich , K. E., & West, R. F. (1983). On priming by sentence context. *Journal of Experimental Psychology: General, 112,* 1-36.

Stein, C. L., Cairns, H. S., & Zurif, E. B. (1984). Sentence comprehension limitation to syntactic deficits in reading-disabled children. *Applied Psycholinguistics, 5,* 305-322.

Tanenhaus, L. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic context. *Journal of Verbal Learning and Verbal Behavior, 18,* 427-440.

Tunmer, W. E., Nesdale, A. R., & Wright, A. D. (1987). Syntactic awareness and reading acquisition. *British Journal of Developmental Psychology, 5,* 25-34.

Tweedy, J. R., Lapinsky, R. H., & Schvaneveldt, R. W. (1977). Semantic context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition, 5,* 84-89.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics, 34,* 409-420.

Vogel, S. A. (1974) Syntactic abilities in normal and dyslexic children. *Journal of Learning Disabilities, 7,* 103-109.

West, R. F., & Stanovich, K. E. (1986). Robust effect of syntactic structure on visual word recognition. *Memory & Cognition, 14,* 104-112.

Wigl, E. H., Semel, M. S., & Crouse, M. B. (1973). The use of English morphology by high-risk and learning disabled children. *Journal of Learning Disabilities, 6,* 457-465.

Willows, D. M., & Ryan, E. B. (1986). The development of grammatical sensitivity and its relation to early reading achievement. *Reading Research Quarterly, 21,* 253-266.

Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition, 12,* 31-45.

FOOTNOTES

[*] *Journal of Experimental Child Psychology,* in press.

[†] Department of Psychology and School of Education, Hebrew University, Jerusalem.

[‡] Department of Psychology and School of Education, and The Center of Neural Computation Hebrew University, Jerusalem

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 81-93

Cognitive Profiles of Reading-disabled Children: Comparison of Language Skills in Phonology, Morphology and Syntax*

D. Shankweiler,† S. Crain,† L. Katz,† A. E. Fowler,‡ A. M. Liberman, S. A. Brady,§ R. Thornton,‡‡ E. Lundquist,‡‡ L. Dreyer,[1] J. M. Fletcher,[2] K. K Stuebing,[2] S. E. Shaywitz,[3] and B. A. Shaywitz[3]

A comprehensive cognitive appraisal of elementary school children with learning disabilities showed that within the language sphere, deficits associated with reading disability are selective: Phonological deficits consistently accompany reading problems whether they occur in relatively pure form or in the presence of coexisting attention deficit or arithmetic disability. Although reading-disabled children were also deficient in production of morphologically related forms, it was shown that this difficulty stems in large part from the same weakness in the phonological component that underlies reading disability. In contrast, tests of syntactic knowledge did not distinguish reading-disabled children from those with other cognitive disabilities, nor from normal children after covarying for intelligence.

The insight that underlies reading in an alphabetic system is, of course, that letters of the printed word correspond approximately to phonologic segments of the spoken word. Phoneme awareness, the ability to analyze words into consonant and vowel segments, is necessary for mastery of an alphabetic writing system. It is not surprising, then, that measures of phonological awareness constitute the strongest single correlate of reading success, far superior to measures of general intelligence in distinguishing dyslexic readers from normals (See Fletcher, Shaywitz, Shankweiler, Katz, Liberman, Stuebing et al., 1994; Goswami & Bryant, 1990; Share, Jorm, Maclean, & Matthews, 1984; Stanovich & Siegel, 1994 for reviews).

There is some evidence that special difficulties in achieving phonological awareness, and hence learning to read, reflect a general weakness in the child's phonological specialization for language (Liberman, Shankweiler, & Liberman, 1989; Olson, Wise, Connors, & Rack, 1990; Stanovich, 1988; Vellutino & Scanlon, 1991; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993). One would therefore expect to find in poor readers other manifestations of a phonological deficiency. Research has borne out this expectation. Reading-disabled children and adults are characterized by poor retention of phonological information in verbal working memory (Brady, 1991; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979), difficulties in retrieving the phonological shapes of words on object naming tasks (Katz, 1986; Wolf, 1991), and difficulties on

This research was supported chiefly by a Program Project Grant "Psycholinguistic and Biological Mechanisms in Dyslexia" (IP50HD21888) from the National Institute of Child Health and Human Development to Yale University Medical School and Haskins Laboratories, New Haven, CT. Partial support of preparation of the manuscript was provided by NICHD Grant HD-01994 to Haskins Laboratories. Assistance from John Holahan and Abigail Shneider is gratefully acknowledged. We also thank Carol Fowler, Ram Frost, and Michael Studdert-Kennedy for their comments on earlier drafts. Finally, we must acknowledge the invaluable contributions of our late colleague, Isabelle Y. Liberman, who did not live to see the completion of this project, but for whose guidance we are all indebted.

phonetically taxing speech production and perception tasks (Brady, Poggie, & Rapala, 1989; Brady, Shankweiler, & Mann, 1983).

Although these results appear to implicate a specific phonological deficit, poor readers also frequently fail on language tasks that are not ostensibly phonological. They sometimes fail in understanding the morphological relations among forms derived from a common root (Carlisle, 1988; Elbro, 1990; Vogel, 1975), and in comprehending some spoken sentences (Byrne, 1981; Mann, Shankweiler, & Smith, 1984).[4] It has been proposed that these difficulties reflect a deficit in morphosyntactic development, over and above the phonological deficit (Stein, Cairns, & Zurif, 1984; Vogel, 1975). Alternatively, these difficulties may be further symptoms of an underlying phonological weakness. If so, there would be a unitary explanation for the entire symptom complex (see Shankweiler & Crain, 1986).

There is some evidence that the morphologically-related problems associated with reading disability are at least in part phonological in origin. In a study preliminary to this one, Fowler and Liberman (1995) found that poor readers have particular difficulty in the production of morphological forms that involve a phonological change within the base morpheme (as in *courage/courageous*); they do not have as much difficulty when the phonology of the base does not change in the derived form (as in *danger/dangerous*). It appears that a weakness in the phonologic component may make some kinds of morphological relationships particularly difficult to learn.

Syntactic difficulties may require a different explanation. Among the syntactic structures that reading-disabled children find difficult to process are passives, relative clauses and sentences containing adjectives with exceptional control properties. In our research, however, we have found that reading-disabled children can succeed nearly as well as normal children with these structures when comprehension is tested by a task that minimizes demands on working memory (Bar-Shalom, Crain, & Shankweiler, 1993; Fowler, 1988; Macaruso, Shankweiler, Byrne & Crain, 1993; Smith, Macaruso, Shankweiler, & Crain, 1989). We have proposed, therefore, that poor readers have the relevant structures in their grammars, but may frequently perform less well than good readers because of phonologically-based limitations of their working memory.

In the present study we wished to make a more stringent test of the hypothesis that reading-disabled children can perform as well as non-disabled children in comprehending complex sentence structures when demands on working memory are minimized. For example, earlier work from our laboratory showed that when sentence structure is held constant, reducing the number of animate noun phrases and making the test sentences conform to presuppositional constraints allows reading-disabled children to perform nearly as well as normal readers (Smith, Macaruso, Shankweiler, & Crain, 1989). Here, also, we sought to make the response required of the child as simple as possible. Yet, at the same time, we sought to make the sentences challenging to interpret. First, we included a greater variety of complex structures than had previously been tested in an experiment involving the same children. Secondly, we included a large proportion of *a priori* implausible sentences. So, for example, we had mice chasing cats. The rationale was if children can succeed on these, it must be the syntax that is driving the analysis rather than *a priori* plausibility. Finally, we included syntactically ambiguous sentences that have two grammatical interpretations in order to find out whether both interpretations would be available to the children. Would we find differences between poor readers' and normals' ability to access the less preferred interpretation?

One purpose of this study was to explore the possibility that poor readers in the early elementary grades have difficulties in the morphological and syntactic domains that cannot be explained by a unitary phonological deficit. A second purpose was to compare reading-disabled children not only to normal readers, but also to children with arithmetic disability or attention deficits. To date, most studies supporting the phonological deficit hypothesis do not go beyond a comparison of poor readers with normal readers. Obtained differences in such studies are therefore ambiguous. One cannot rule out the possibility that other learning-disabled children might show similar cognitive profiles, even in the absence of reading disability. Here, we asked whether reading-disabled children, throughout the normal range of IQ, would display a specific pattern of language abilities that distinguishes them not only from normal children but also from children with attention deficits or arithmetic disability. In addition, we wished to learn whether phonological abilities remain the best predictors of reading skill when other accompanying deficits are present.

Method

Subjects. Comprehensive testing of linguistic and nonlinguistic abilities was carried out on 353 children, aged 7.5 to 9.5, recruited for disabilities in reading, arithmetic and attention, and representing a wide range of intelligence levels.[5] The data analysis included all the children recruited for the study who met either the criteria for reading disability, arithmetic disability, or attention deficit disorder. Other children who met none of the criteria, but who passed screening tests for vision and hearing, were included as normal controls. Classification was based on IQ, achievement in reading and arithmetic, and standard criteria for attention disorder (APA, 1980).

The criteria for reading/arithmetic disability were based on 1) a regression discrepancy of 1.5 standard errors between achievement (word and nonword reading and/or math (arithmetic) subtests of the Woodcock-Johnson Psycho-Educational Test Battery) and IQ (WISC Full Scale), *or* 2) low achievement (a score below the 25th percentile on one or both of the reading subtests or the math subtest). To meet either criterion, a subject must also have attained an IQ of 80 or above (see note, Table 1). A comparison of discrepancy (1) and low achievement (2) subgroups showed that the profiles of cognitive abilities were basically similar (See Fletcher et al., 1994; Stanovich & Siegel, 1994). Accordingly, in the present study, we combined these subgroups in the data analysis.

Tasks. The experimental language tasks tapped phonological, morphological and syntactic components. The phonological measures included tests of phonological awareness and verbal short-term memory. Measures of these skills have proven strongly diagnostic of reading disability in earlier studies with more limited samples. The phoneme awareness measure (PD, "phoneme deletion") tested the ability to segment spoken words by phoneme (Rosner & Simon, 1971). The child was asked to repeat a spoken word (e.g., *smile*), and then to repeat it again omitting a specified segment (e.g., "Can you say smile without the sss"?). Tests of short-term memory measured ability for immediate recall of three classes of linguistic materials: *random word sequences*, consisting of 20 sequences of five monosyllabic, high-frequency words, as in Mann, Liberman, and Shankweiler (1980); *digit sequences*, tested by standard procedures for the Digit Span Test, as specified in the WISC-R manual; and *sentence repetition*, based on a subset of the sentences from the Syntax test, which was administered on a later day. A composite short-term memory score (STM) made up from these three tests was used in the analysis.

The test of morphological awareness, adapted by Fowler and Liberman (1995) from Carlisle (1988), included two parts. In one, the experimenter articulated a word (a base form), then used it in a sentence designed to elicit the appropriate derived form. For example, "Four: My brother's team placed ______," and "Five: This prize would be her_____ ." In the other condition, the child's task was to extract the base from the derived form, using the same cloze procedure to prompt production of the target word. In half the items the target derived form results in no phonological change in the base (as in first example, *four/fourth*). In the other half, the base must have undergone phonological change in creating the derived word (as in the second example, *five/fifth*).

In designing the tests of syntax, we sought out structures that are considered to be mastered late in the course of language acquisition, and that our previous research had found to be difficult for children of this age range (Crain, Shankweiler, Macaruso, & Bar-Shalom, 1990). Testing was by a sentence-picture matching procedure. Tape-recorded versions of syntactically unambiguous or ambiguous sentences were presented over a loud speaker and their corresponding pictures were presented synchronously by computer. For unambiguous sentences, a "correct" match between sentence and picture reflected the proper interpretation of the sentence in the adult grammar. The mismatching pictures illustrated an incorrect interpretation. An example sentence with a mismatching picture is given in Figure 1. Subjects indicated by a key press whether or not single the picture displayed on the monitor was a good match for the sentence.

Table 1. *Means and standard deviations of assessment measures for selection and classification of subjects.*

				WISC		WOODCOCK-JOHNSON		WOODCOCK-JOHNSON		WOODCOCK-JOHNSON		WOODCOCK-JOHNSON	
GROUP	N	AGE		FULL-SCALE IQ		REAL WORDS		NONNWORDS		COMPREHENSION		MATH	
		MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
NORM	50	8.44	0.71	121.54	12.17	114.60	13.99	110.14	11.03	116.80	15.09	110.78	11.00
R	56	8.40	0.70	105.98	11.73	85.86	7.04	82.41	10.44	89.16	11.40	98.25	6.40
M	85	8.40	0.59	107.13	13.86	102.14	9.58	99.42	7.97	101.74	12.41	82.73	9.29
R + M	108	8.32	0.60	100.51	14.43	81.65	10.08	80.73	10.95	80.76	12.47	78.14	11.69
ADD	54	8.40	0.68	107.93	11.13	103.57	9.21	101.72	9.10	102.59	11.66	100.41	7.71

Note. IQ was assessed by the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974). Reading and arithmetic calculation were assessed by subtests from the Woodcock-Johnson Psycho-Educational Battery (Woodcock & Johnson, 1978): word reading (WJ-13), nonword reading (WJ-14), math (WJ-16), and paragraph comprehension (WJ-15). The latter subtest was not used in classification.

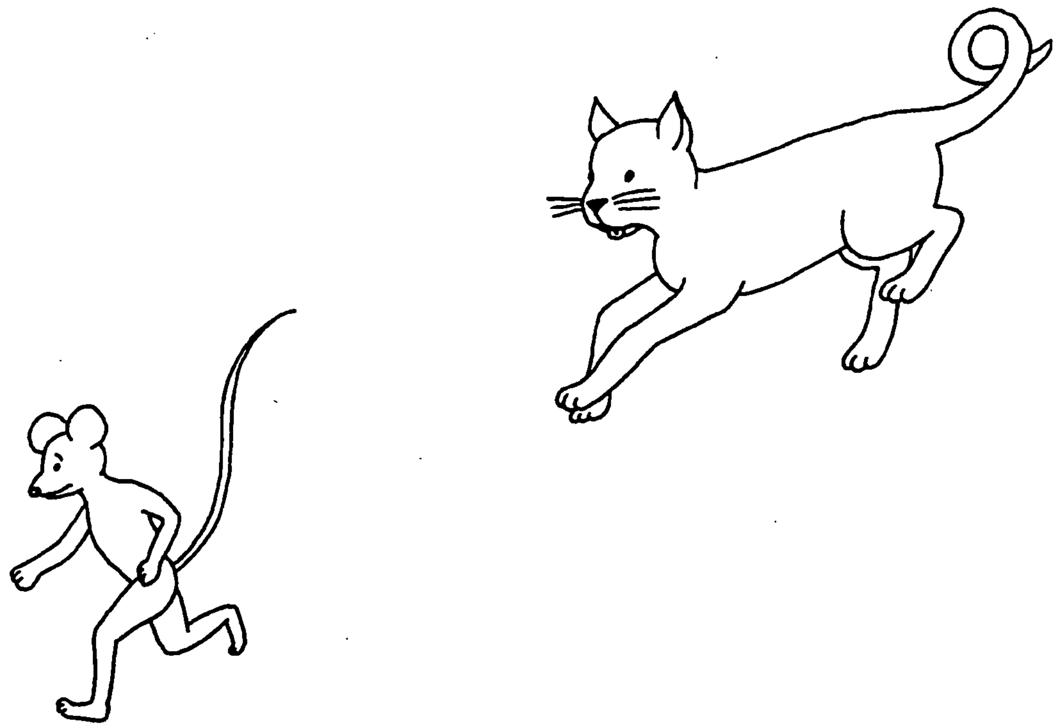


Figure 1. **Syntax Test. "The cat with a curly tail is being chased by the mouse." The figure depicts an incorrect interpretation in which agent and patient roles are reversed.**

Syntactic complexities turned on the following: relative clauses, passives, control properties of adjectives, and pronoun coreference.[6] Data from matching sentences and pictures, i.e., "yes" trials and mismatching , i.e., "no" trials, were entered as separate factors in the analysis. For the ambiguous sentences test, the same kinds of constructions were employed, but for each sentence there always existed two legitimate syntactic interpretations. Each sentence was accompanied by a picture corresponding to one or the other interpretation. Errors consisted of failures to recognize a match between sentence and picture. The individual syntax tests were thus of three types: Unambiguous sentences with correctly matching pictures (SYNyes), and with nonmatching pictures (SYNno), and ambiguous sentences (SYNamb) in which the picture matched one interpretation.

In addition to analytic language measures described above, the tasks included a test of listening comprehension at the discourse level and tests measuring three reading abilities: words (RWORD) and nonwords (RNWORD), each presented in list form, and comprehension of text (RCOMP). Each of the three reading measures was a composite made up of at least two independent tests of that ability (note, Table 2, gives the individual reading tests).

Results

Table 1 gives the results of the classification procedure, which partitioned the 353 children into the following five groups: Reading disability (R); math (e.g., arithmetic calculation) disability (M); reading and math disability (R+M); attention deficit disorder (ADD), and normal (NORM).[7] Each group was treated as a separate block in the data analysis.

It is notable that the three different types of reading measures—words, nonwords, and paragraph comprehension—were highly intercorrelated (see Table 2). Word reading and pseudoword reading correlated .92 with each other, and each correlated .89 and .79, respectively, with reading comprehension. Thus, comprehension scores at this age are largely determined by ability in decoding (see Perfetti, 1985; Shankweiler, 1989). The phonological measures (PD and STM) and the morphological measures (MORPH) are substantially correlated with each of the measures of reading ability (RWORD, RNWORD, RCOMP). Syntax measures, in contrast, are only weakly correlated with the other measures.

Table 2. *Correlations among assessment measures and experimental language tasks.*

	MATH	RWORD	RNWORD	RCOMP	STM	PD	MORPH	SYNamb	SYNyes
RWORD	.56								
RNWORD	.52	.92							
RCOMP	.56	.89	.79						
STM	.40	.54	.55	.52					
PD	.44	.73	.79	.66	.54				
MORPH	.49	.72	.68	.71	.56	.67			
SYNamb	.06	.05	.03	.06	.12	.04	.08		
SYNyes	.07	.12	.13	.14	.15	.09	.18	.50	
SYNno	.22	.22	.21	.21	.17	.19	.26	-.13	.19

Note. Math is represented by the arithmetic subtest (WJ-16) of the Woodcock-Johnson Psycho-Educational Battery (Woodcock 7 Johnson, 1978). Reading is represented by three composite measures derived from tests of reading words, reading nonwords, and text comprehension. For words (RWORD), the measures were Woodcock-Johnson, WJ-13; Wide Range Achievement Test-Revised (Jastak & Wilkinson, 1984); and Decoding Skills Test, Words (Richardson & DiBenedetto, 1986). For nonwords (RNWORD), the measures were Woodcock-Johnson, WJ-14, and Decoding Skills Test, Nonwords. For text comprehension (RCOMP), the measures were Woodcock-Johnson, WJ-15; Gray Oral Reading Test, Paragraphs (Gray, 1967); and Formal Reading Inventory, Form B (Wiederholt, 1986). (A parallel form of the Formal Reading Inventory, Form C, was administered in spoken form as the listening comprehension control.) STM = composite measure of short-term memory; PD = measure of phoneme awareness; MORPH = measure of morphological awareness; SYNamb = performance on syntax test with ambiguous sentences; SYNyes = performance on matching-pictures trials of syntax test with unambiguous sentences; SYNno = performance on nonmatching-pictures trials of syntax test with unambiguous sentences.

Each of the language measures was adjusted by covariance analysis for differences due to age, listening comprehension, and general intelligence and then the measures were standardized (as z-scores) to place them all on the same scale. This procedure leaves residual differences between groups that are specifically associated with differences in reading ability. IQ and listening comprehension are moderately correlated with each other ($r = .55$), and each is correlated at about the same magnitude with each measure of reading ability. It is appropriate to remove the contribution of these measures of general comprehension to isolate the specific contribution to reading of the experimental language measures that are the focus of interest (see Stanovich, 1991).

Means of the adjusted, standardized language measures for phonology, morphology and syntax are plotted for each subject group in Figure 2, yielding profiles of language abilities. A MANOVA yielded a significant effect of groups; $F(24,1191) = 6.73$, $p < .0001$. Subsequent univariate analyses showed significant effects of groups with $p < .0001$ for all of the following: STM, $F(4,346) = 11.71$, PD, $F(4,346) = 29.46$, and MORPH, $F(4,346) = 18.99$. In contrast, none of the tests for the syntax measures (i.e., SYNamb, SYNyes, and SYNno) approached significance.[8]

Factors specifically associated with reading disability are revealed by post hoc Fisher Least Significant Difference comparisons of differences among the individual subject groups. Each of the two groups of reading-disabled children, the pure reading disability group (R) and the reading and arithmetic group (R+M), differed significantly from normal children on each of the phonologically-driven tasks, PD and STM, and on the morphology test, MORPH (for all comparisons, $p < .0001$). The most telling comparison is between learning-disabled children with reading deficits alone and those with arithmetic deficits alone. Because these groups were well-matched in IQ (see Table 1), tasks that distinguished them (even without statistical adjustment) are likely to reflect essential aspects of the reading process. On each of the phonological tasks and the morphological task the R group's performance was significantly inferior to that of the M group (for all comparisons, $p < .0001$). (The R+M group resembled the R group, but showed a lower level of performance on the phonological tasks).

Decomposition of the morphology test

As noted, the test of morphological awareness (MORPH) yields an overall score made up from different types of items. On half of the items the task was to generate a derived form given the base (e.g., *fifth* from *five*), and for the remainder the task was the reverse (*five* from *fifth*). As it happened, the former was only slightly more difficult than the latter; therefore we collapsed across this factor in Figure 3. A more relevant difference among the items turned on whether the base undergoes phonological change in the course of generating the derived form (e.g., *five-fifth*) or whether the derivation involves addition of a suffix to the base without changing it (e.g., *four-fourth*). As seen in Figure 3, the phonological change condition was harder than the no-change condition, $F(1,348) = 401.14$; $p < .0001$. Notably, the interaction of Group × Condition is significant ($F(4, 348) = 7.86$; $p < .0001$), indicating that the phonological change condition resulted in greater differences among the groups. It is apparent that the two groups of poor readers were most affected by phonological change: Differences between the R and R+M groups and the normal group were greatest for these items.

Regression analyses in which word reading (RWORD) was the dependent measure indicate that variance attributable to MORPH is largely, but not completely overlapping with the variance attributable to PD. After residualizing for age and IQ, we varied the order of entry of PD and MORPH in a hierarchical regression. When PD is entered last, it accounts for slightly more of the variance in RWORD than MORPH does when it is entered last (.109 vs. .051 increment to R-square).

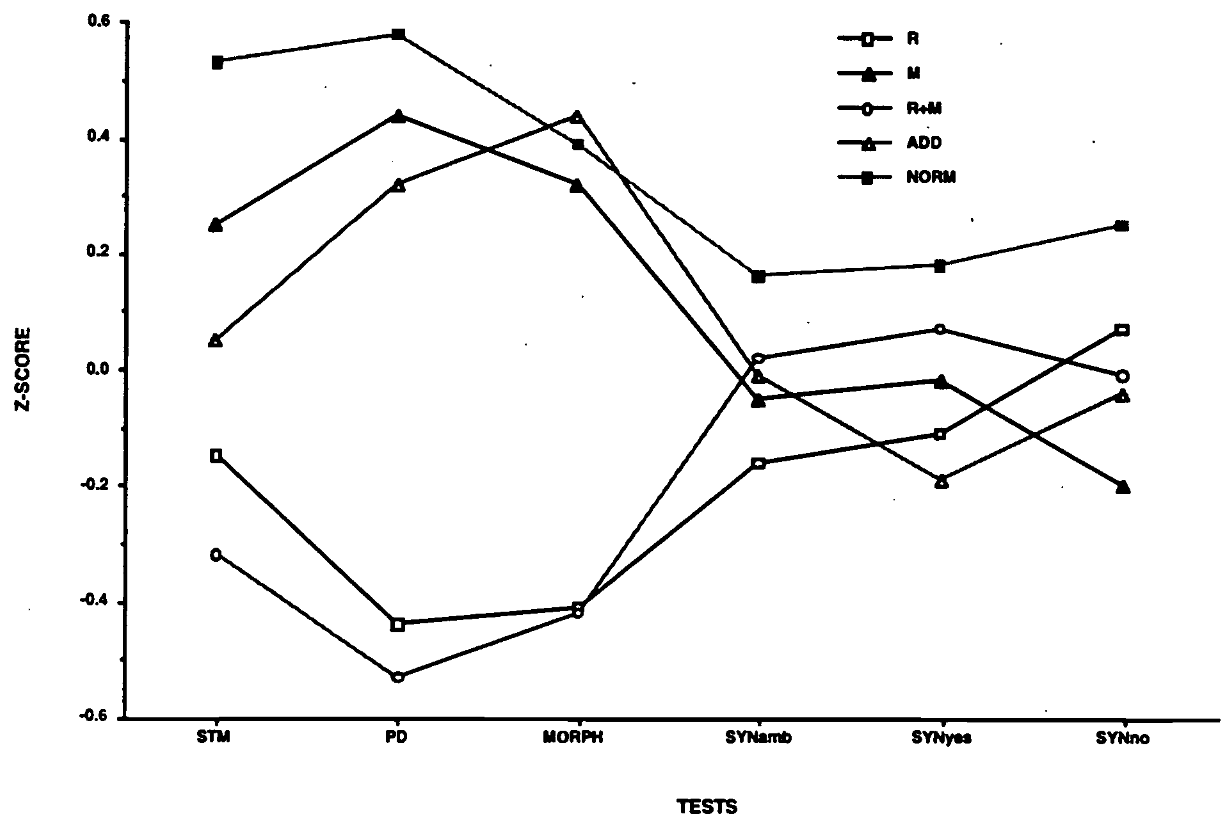


Figure 2. Means of the language measures (z scores). The plot shows scores after adjustment for age, listening comprehension, and Full-Scale IQ (Wechsler, 1974). Tests are as follows: STM, short-term memory composite measure: PD = phoneme awareness; MORPH = morphological awareness; SYNamb = structurally ambiguous sentences; SYNyes = unambiguous sentences that are correct matches for their accompanying pictures; SYNno = unambiguous sentences that are incorrect matches for their accompanying pictures. The groups are as follows: R = reading disability; M = math disability; R+M = reading and math disability; ADD = attention deficit (without reading or math disability); NORM = normal control subjects.

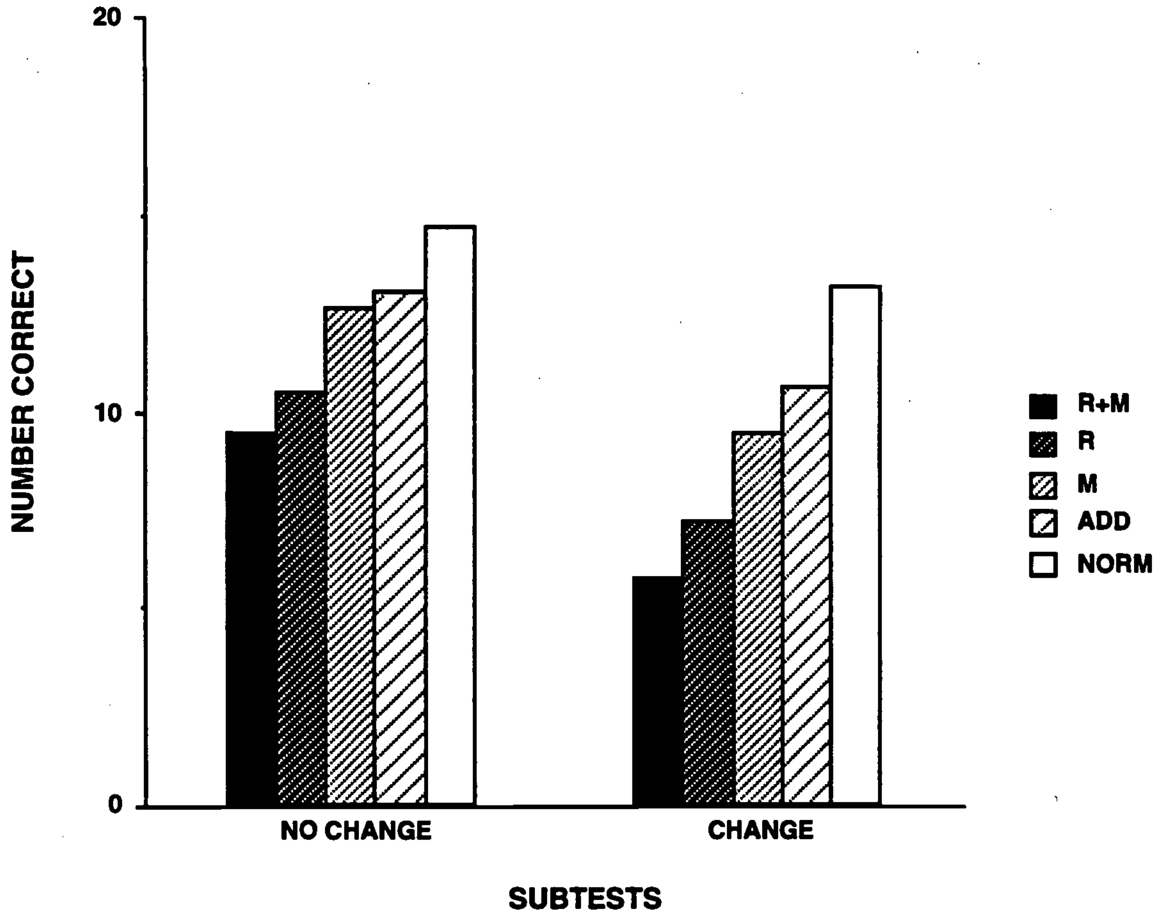


Figure 3. **Results from the test of morphological awareness. The bars on the left graph performance by group on trials on which the base does not change phonologically in the derived form; the bars on the right graph data for trials on which the base undergoes phonological change. See Figure 2 for an explanation of the abbreviations designating the five subject groups.**

Analysis of the syntax measures

Further analyses examined correct responses on the syntax tests (unadjusted for IQ and listening comprehension) by type of syntactic construction (see Figure 4). The results for ambiguous sentences are not included in the figure and are not discussed further since the findings closely resembled those with unambiguous sentences. For unambiguous sentences, incorrect matches (SYNno) and correct matches (SYNyes) are plotted separately. "No" trials proved more difficult than "yes" trials, as expected. In order to reject a picture as a depiction of a sentence, one must have detected a specific feature in which sentence and picture fail to match, whereas acceptance merely implies that a mismatch was not detected. There were significant effects of construction and group both for the "no" trials and the "yes" trials (for SYNno: construction, $F(3,348) = 43.42$, $p < .0001$; groups; $F(4,348) = 5.22$, $p < .0005$; for SYNyes; for construction, $F(3,348) = 4.58$, $p < .004$); groups; $F(4,348) = 2.47$, $p < .05$). There were no significant interaction effects between type of construction and group. The slight, across-the-board superiority of the normal group is wholly accounted for by higher IQ and listening comprehension scores. However, even without removing the influence of IQ and listening comprehension, we find no significant differences between the critical pair of groups, R and M, for any of the syntax tests.

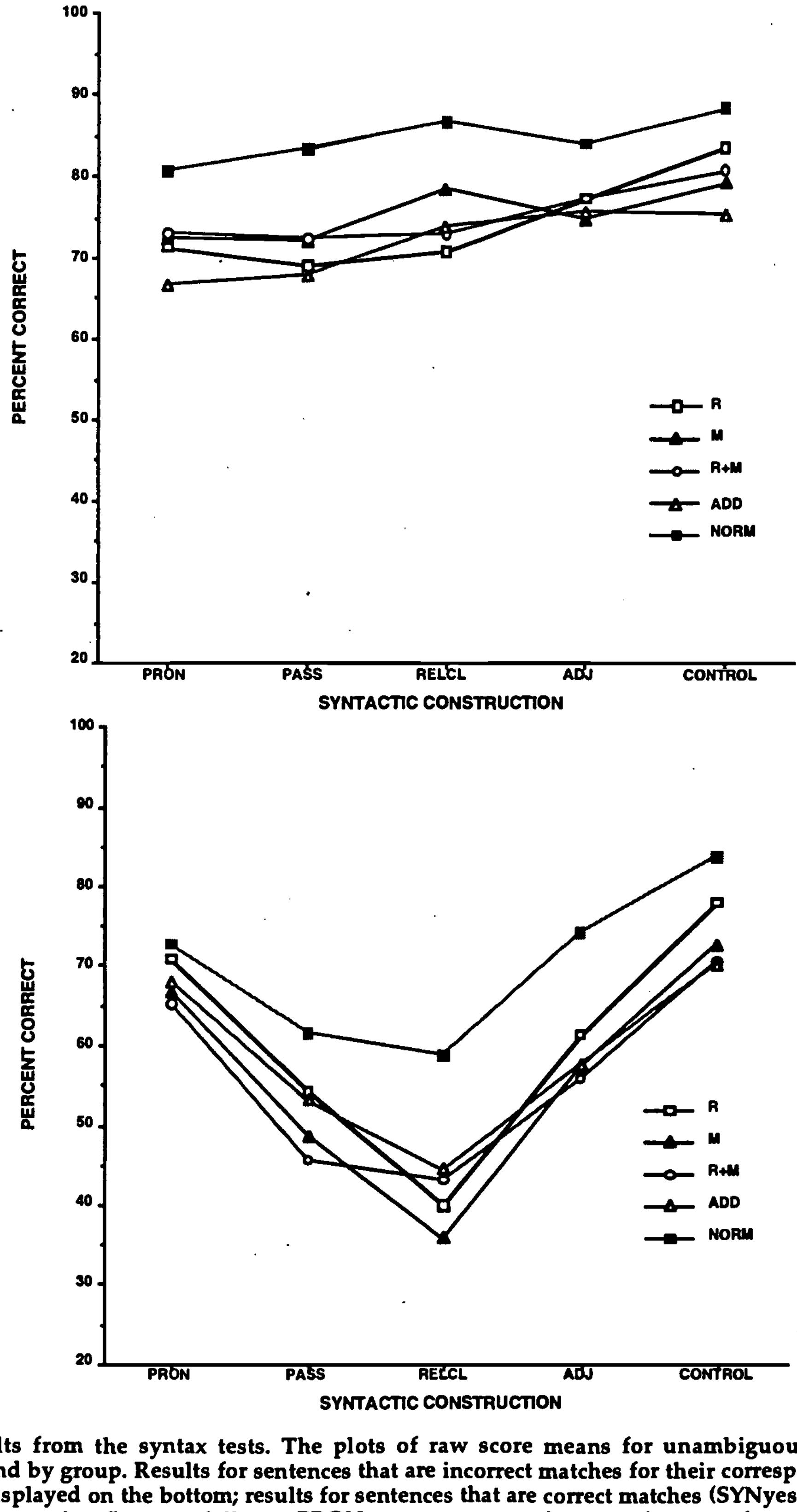


Figure 4. **Results from the syntax tests. The plots of raw score means for unambiguous sentences by construction and by group. Results for sentences that are incorrect matches for their corresponding pictures (SYNno) are displayed on the bottom; results for sentences that are correct matches (SYNyes) are on the top. "Syntactic constructions" are as follows: PRON = sentences testing co-reference of pronouns; PASS = passives; RELCL = sentences containing a relative clause or sentential complement; ADJ = sentences containing adjectives with exceptional control properties; CONTROL = syntactically simple sentences. See Figure 2 for an explanation of the abbreviations designating the five subject groups.**

DISCUSSION

The findings affirm that deficits on phonologically-driven tasks (PD, STM) are a common denominator in children with reading disability. Not only did these tasks distinguish reading-disabled children from normal children, they also distinguished children with reading problems from those with other cognitive disabilities. The results are in agreement with earlier research, including recent twin studies that identify phonological skills as the most likely mediators of genetically-based differences in reading ability (e.g., DeFries, Olson, Pennington & Smith, 1991). In addition, the results support earlier indications that morphology-dependent skills are deficient in poor readers. The phonological and morphological deficiencies were confirmed by their strong residual associations with measures of reading ability after the contributions of intelligence and listening comprehension had been removed. The Syntax test, on the other hand, did not discriminate either in numbers of errors or in distribution of errors across the various syntactic constructions.

It is notable that performances on phonological and morphological tasks were highly intercorrelated whereas neither was correlated more than weakly with the syntax task. The strong association between phonological awareness (PD) and morphological awareness (MORPH) tells us that both tests converge on a common ability. To interpret this association, it is important to note that the subset of the test words on MORPH that undergo phonological change were the most discriminating in separating poor readers from children with other disabilities and from normal children. Together these facts suggest that the difficulty that the poor readers experienced in generating appropriate derived forms is at least in part an expression of a phonological limitation. Consistent with this is the added fact that more of the unique variance in reading performance is associated with phonological awareness (PD) than with morphological awareness (MORPH). In sum, the findings with PD and MORPH largely reflect a common source of difficulty. It seems likely that what they have in common is their reliance on the phonological component.

On the syntax test, performance levels tended to be rather low, indicating that we achieved our goal in making the sentences difficult. The task has two components: an interpretative component and an execution component. The interpretative component is intrinsically difficult because the sentences are presented without supporting contexts. (In ordinary circumstances sentences occur in contexts that make them true, so the question of truth value does not arise). The test sentences are therefore not experienced as children would experience them in real life. In addition, as we noted, many of the sentences depicted events that would be unexpected in the real world. The response component, on the other hand, which required only a go/no-go response, should have posed few additional difficulties of its own.

Examining the results by sentence type, shows marked differences in difficulty of the various structures (for SYNno items). Relative clauses and passives were more difficult than pronoun coreference and adjective control. Each of these was, in turn, more difficult than simple-structure control sentences. Although we succeeded in making the sentences difficult, interactions between sentence type and group, with poor readers doing relatively worse than normals on the most difficult structures, did not materialize. The failure to find differences among the learning disabled groups indicates that the children's problems with this sentence comprehension task are not related to the (chiefly phonological) difficulties that distinguish good and poor readers. Moreover, the difficulty of the task is not one that could be expected to stress working memory. If, instead, the response required by the task had been more complex, as, for example, if the children had been required to choose between pictures or perform an act-out task, as is commonly done, reading group differences might well have emerged, as in previous studies.[9]

In sum, the data speak clearly on the point that syntactic abilities *per se* did not distinguish poor readers from normals after factoring out IQ, nor did they distinguish reading disabled children from other children with learning problems. The cause of comprehension difficulties in reading and spoken discourse must therefore lie outside of the syntax itself. In contrast, poor readers were distinguished from children with specific deficits in calculation and attention in the phonological and morphological domain. Thus, their weaknesses within the language system are selective.

Poor readers' phonological limitations, particularly as they are expressed in difficulties in parsing words phonemically, handicap them in acquiring the alphabetic principle and in acquiring good word recognition skills. The tightly correlated difficulties in reading comprehension must stem in large part from word recognition skills that are insufficiently accurate and rapid to enable the reader to pass smoothly from the lower-level to the higher-level structures of language. Early intervention is critical for children whose phonological limitations would otherwise predispose them to reading failure.[10]

REFERENCES

Adams, M. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

American Psychiatric Association. (1980). *Diagnostic and Statistical Manual of Mental Disorders* (3rd. ed.). Washington, DC.

Ball, E. W., & Blachman, B. A. (1991). Does phoneme segmentation training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26,* 49-66.

Bar-Shalom, E., Crain, S., & Shankweiler, D. (1993). A comparison of comprehension and production abilities of good and poor readers. *Applied Psycholinguistics, 14,* 197-227.

Bradley, L., & Bryant, P. E. (1983).Categorizing sounds and learning to read—a causal connection. *Nature, 301,* 419-421 .

Brady, S. A. (1991). The role of working memory in reading disability. In S. A. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 129-151). Hillsdale, NJ: Erlbaum.

Brady, S., Poggie, E., & Rapala, M. M. (1989). Speech repetition abilities in children who differ in reading skill. *Language and Speech, 32,* 109-122

Brady, S., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35,* 345-367.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? *Applied Psycholinguistics, 2,* 201-212.

Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 85,* 1-5.

Carlisle, J. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth and eighth graders. *Applied Psycholinguistics, 9,* 247-266.

Crain, S., Shankweiler, D., Macaruso, P., Bar-Shalom, E. (1990). Working memory and sentence comprehension: Investigations of children with reading disorder. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 477-508). Cambridge, England: Cambridge University Press.

DeFries, J. C., Olson, R. K., Pennington, B. F., & Smith, S. D. (1991). The Colorado reading project: An update. In D. D. Duane & D. B. Gray (Eds.), *The reading brain: The biological basis of dyslexia* (pp. 53-87). Parkton, MD: York Press.

Elbro, C. (1990). *Differences in dyslexia.* Munksgaard, Copenhagen.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L. Liberman, I. Y., Stuebing, K. K., Francis, D. J., Fowler, A. E., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86,* 1-23.

Fowler, A. (1988). Grammaticality judgments and reading skill in Grade 2. *Annals of Dyslexia, 38,* 73-94.

Fowler, A., & Liberman, I. Y. (1995). Morphological awareness as related to early reading and spelling ability. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 157-188). Hillsdale, NJ: Erlbaum.

Goswami, U., & Bryant, P. E. (1990). *Phonological skills and learning to read.* Hillsdale, NJ: Erlbaum.

Gray, W. S. (1967). *Gray Oral Reading Test.* New York: Bobbs-Merrill.

Jastak, S., & Wilkinson, G. S. (1984). *The Wide Range Achievement Test-Revised.* Wilmington, DE: Jastak Associates.

Katz, R. (1986). Phonological deficiencies in children with reading disability: Evidence from an object naming task. *Cognition, 22,* 225-257.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1-33). Ann Arbor, MI: University of Michigan Press.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263-284.

Macaruso, P., Shankweiler, D., Byrne, B., & Crain, S. (1993). Poor readers are not easy to fool: Comprehension of adjectives with exceptional control properties. *Applied Psycholinguistics, 14,* 285-298.

Mann, V. A., Liberman, I. Y., Shankweiler, D., (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory and Cognition, 8*, 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11*, 627-643.

Olson, R. K., Wise, B., Connors, F., & Rack, J. (1990). Organization, heritability, and remediation of component word recognition and language skills in disabled readers. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skill approaches* (pp. 261-322). New York: Academic Press.

Perfetti, C.A. (1985). *Reading ability*. New York: Oxford University Press.

Richardson, E., & DiBenedetto, B., (1986). *Decoding Skills Test*. Parkton, MD: York Press.

Rosner, J., & Simon, D. P. (1971). The auditory analysis test: An initial report. *Journal of Learning Disability, 4*, 384-392.

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development, 61*, 1728-1743.

Shankweiler, D. (1989). How problems of comprehension are related to difficulties in decoding. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 35-67). Ann Arbor, MI: University of Michigan Press.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition, 24*, 139-168.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning & Memory, 5*, 531-545.

Share, D., Jorm, A., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76*, 1309-1324.

Shaywitz, B. A., Shaywitz, S. E., Liberman, I. Y., Fletcher, J. M., Shankweiler, D. P., Duncan, J., Katz, L., Liberman, A., Francis, D., Dreyer, L., Crain, S., Brady, S. Fowler, A., Kier, L. Rosenfield, N., Gore, J., & Makuch, R. (1991). Neurolinguistic and Biological Mechanisms in Dyslexia. In D. D. Duane & D. B. Gray (Eds.), *The reading brain: The biological basis of dyslexia* (pp. 27-52). Parkton, MD: York Press.

Smith, S. T., Macaruso, P., Shankweiler, D., & Crain, S, (1989). Syntactic comprehension in young poor readers. *Applied Psycholinguistics, 10*, 429-454.

Stanovich, K .E., (1991). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly, 26*, 1-29.

Stanovich, K. E., (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21*, 590-612.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24-53.

Stein, C. L., Cairns, H. S., Zurif, E. B., (1984). Sentence comprehension limitations related to syntactic deficits in reading-disabled children. *Applied Psycholinguistics, 5*, 305-322.

Vellutino, F. R., & Scanlon, D. M. (1991). The Preeminence of phonologically based skills in learning to read. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 237-252). Hillsdale, NJ: Erlbaum.

Vogel, S. A. (1975). *Syntactic abilities in normal and dyslexic children*. Baltimore, MD: University Park Press.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K. & Rashotte, C.A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85*, 83-103.

Wechsler, D., (1974). *Wechsler Intelligence Scale for Children-Revised*. New York: The Psychological Corporation.

Wiederholt, J. L. (1986). *Formal Reading Inventory*. Austin, TX: Pro-Ed.

Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly, 26*, 123-141.

Woodcock, R., & Johnson, M. B. (1978). *Woodcock-Johnson Psycho-Educational Battery*. Allen Park, TX: DLM Teaching Resources.

FOOTNOTES

**Psychological Science*, (1995), *6*, 149-156.

† Also University of Connecticut.

‡ Also Wesleyan University.

‡‡ University of Connecticut, Storrs.

§ Also University of Rhode Island.

[1] Southern Connecticut State University.

[2] University of Texas Medical School at Houston.

[3] Yale University School of Medicine.

[4]Because English orthography represents both phonology and morphology, one could reasonably expect that an explicit awareness of both types of language structures should be required for mastery of the spelling system. To a large extent, English is written morpho-phonologically. So, for example, the word *health* is written thus (not as **helth*) to make transparent its relation with the root morpheme, *heal*. Moreover, there is much psycholinguistic evidence that morphemically complex words are in fact treated in an analytic fashion by skilled readers (See Fowler & Liberman, 1995, for a review).

[5]See Shaywitz et al. (1991) for a description of the plan of the project. See also Fletcher et al., 1994, for related findings.

[6]In the case of relative clauses, an interpretation that children have been reported to assign is the "conjoined clause" analysis. Thus, for the test sentence "The man is riding the horse that is wearing a hat," the picture for the incorrect sentence/picture match represented "The man riding the horse and wearing a hat." For targets testing control properties of adjectives, such as "The kangaroo with a ribbon around its neck is easy to reach," the picture depicted the kangaroo reaching an apple. For the sentence testing coreference of pronouns, the incorrect picture depicted illegitimate coreference. For example, "While the magician is sitting down, the prince is tickling him with a feather duster" was accompanied by a picture in which the prince is tickling himself with the feather duster.

[7]The R, M and R+M groups contain many children who also met criteria for attention deficit disorder. The ADD group includes only those children with attention deficit who do not meet criteria for reading or arithmetic disability).

[8]If it seems strange to remove the measure of listening comprehension when evaluating syntax, it should be noted that the listening test employed here consisted of content questions about discourse that assess the listener's inferencing skills and ability to follow a narrative, but make only modest syntactic demands. In fact, this measure proved only weakly correlated (.2 or less) with each of the syntax measures.

[9]It is conceivable that if we had tested children as young as three, differences in syntactic knowledge might have emerged as predictors of later reading success. Such a claim has been made by Scarborough (1990).

[10]There is evidence that children who are at risk for dyslexia on account of their weak phonological processing abilities can be successfully taught to decode using methods that foster phonological awareness (Adams, 1990; Ball & Blachman, 1991; Bradley & Bryant, 1983; Byrne & Fielding-Barnsley, 1993; Lundberg, Frost & Petersen, 1988).

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 95-115

Syntactic Processing in Agrammatic Aphasia by Speakers of a Slavic Language*

Katarina Lukatela,† Donald Shankweiler,† and Stephen Crain†

It is widely believed that agrammatic aphasics have lost the ability to assign complete syntactic representations. This view stems from indications that agrammatics often fail to comprehend complex syntactic structures, as for example, some types of relative clauses. The present study presents an alternative account. Comprehension by Serbo-Croatian-speaking agrammatic aphasics was tested on four types of relative clause structures and on conjoined clauses. The relative clauses varied in type of embedding (embedded vs. nonembedded) and in the location of the gap (subject position vs. object position). There were two control groups: Wernicke-type aphasics and normal subjects. The findings from a sentence-picture matching task indicated that agrammatic aphasics were able to process complex syntactic structures, as evidenced by their well-above chance performances. The success rate varied across different types of relative clauses, with object-gap relatives yielding more errors than subject-gap relatives in all groups. Each group showed the same pattern of errors: agrammatic subjects were distinguished from Wernicke subjects and normal subjects only in quantity of errors. These findings are incompatible with the view that the agrammatics are missing portions of the syntax. Instead, their comprehension deficits reflect varying degrees of processing impairment in the context of spared syntactic knowledge.

Explaining Comprehension Difficulties in Agrammatism

The view that the "agrammatism" of Broca's aphasia represents a disorder involving loss of some structural component of the language apparatus has enjoyed considerable influence in the last two decades (Berndt & Caramazza, 1980; Schwartz et al., 1980; Zurif, 1984). The appeal to missing structural knowledge rested in part on the promise it seemed to hold for explaining parallel deficits in language production and comprehension. Just as agrammatic aphasics produce syntactically deficient speech largely as a result of a tendency to omit function words and to distort inflections, so, too, it might be supposed that they understand sentences by inferring meaning without recourse to normal syntactic operations, using non-syntactic, lexically-based strategies instead. Several specific proposals have been offered, each seeking to ground the difficulties involving the closed-class vocabulary and the inflectional system on one or another level of linguistic representation: phonological (Kean, 1977), lexical (Bradley, Garrett, & Zurif, 1980), morphological (Lapointe, 1983), or syntactic (Caramazza & Zurif, 1976; Caplan & Futter, 1986; Grodzinsky, 1986, 1990; Hickok, Zurif, & Canseco-Gonzales, 1993; Mauner, Fromkin, & Cornell, 1993). Collectively, we call these proposals the Structural Deficit Hypothesis.

Whatever the plausibility of these proposals, there is mounting evidence that calls into question any form of the Structural Deficit Hypothesis. First, several case studies have reported patients who fail to show parallel deficits in production and perception. Some patients present agrammatic symptoms in production, but not in comprehension (Miceli, Mazzucchi, Mann, &

Partial support of preparation of this manuscript was provided by NICHD Grant HD-01994 to Haskins Laboratories.

Goodglass, 1983). Additionally, there are reports of patients who show agrammatic symptoms in comprehension despite fluent production of well-formed sentences (Caramazza, Basili, Koller, & Berndt, 1981; Smith & Bates, 1987). These findings suggest that expressive and receptive agrammatism may represent different deficits, though they often occur together.

The finding that agrammatic aphasics retain the ability to make metalinguistic judgments of grammatical acceptability presents a further challenge to the Structural Deficit Hypothesis. Retained ability to detect syntactic violations has been demonstrated even in patients who were severely agrammatic in both production and comprehension (Linebarger, Schwartz, & Saffran, 1983). Preserved sensitivity to syntactic structure in doubly agrammatic patients cannot readily be explained by a syntactic account of agrammatism. Spared ability to judge the grammaticality of complex syntactic structures has been confirmed in additional studies of English-speaking agrammatics (Shankweiler, Crain, Gorrell, & Tuller, 1989; Wulfeck, 1988). Sensitivity to violations of the inflectional morphology has also been demonstrated in Italian, German, and Serbo-Croatian agrammatics (Lukatela, Crain, & Shankweiler, 1988; Kolk & van Grunsven, 1985; Bates, Friederici, & Wulfeck, 1987, Friederici, Wessels, Emmorey, & Bellugi, 1992).

Central to the Structural Deficit Hypothesis is the assumption that the comprehension deficit in agrammatism is syndrome-specific. This assumption, too, is challenged by findings with other language impaired populations and with normal subjects. For example, sentence comprehension in children with reading problems shows the same ordering of difficulty across syntactic structures as is displayed by agrammatic aphasics (Smith, Macaruso, Shankweiler, & Crain, 1989). Moreover, normal adults working under time pressure have been found to conform to the same pattern (Milekic, 1993; Ni, 1988). Such consistencies that cut across diagnostic groups and normal subjects point to a common source of variation that would implicate a processing explanation, not a structural explanation.

Spurred by findings that are unfavorable to the Structural Deficit Hypothesis, an alternative has begun to crystallize. We call it the Processing Limitation Hypothesis.[1] This hypothesis appeals to the distinction between structural and processing components of the language apparatus. The structural components include the lexicon and the different levels of linguistic representation: phonology, syntax, and semantics. According to the Processing Limitation Hypothesis, impaired comprehension need not reflect loss of critical linguistic structures. Language processing involves not only the assignment of structural representations, it also requires a series of operations for storing and retrieving linguistic information and for coordinating the transfer of information between levels of linguistic representation. The Processing Limitation Hypothesis directs us to consider linguistic *processing* as a possible source of the comprehension deficits that are characteristic of aphasia.

In addition to giving direction to the quest for the source of sentence comprehension difficulties in agrammatism, the Structural Deficit Hypothesis and the Processing Limitation Hypothesis have implications for accounts of normal sentence processing. If an obtained pattern of preserved and impaired comprehension can be accounted for on the basis of the disruption of a particular component of syntactic representation postulated in one theory but not in others, then the data would provide support for that theory. However, if the pattern of performance can be accounted for on the basis of a limitation in processing capacity, then the data could not decide among linguistic theories, but would require a model of sentence processing that incorporates the appropriate processing components.

The intent of the present study was to compare a structural deficit versus a processing limitation account of syntactic comprehension difficulties in agrammatic aphasia. We proceed by examining a structure often implicated in agrammatism, *the relative clause*. We then present the rationale for a study of comprehension of relative clauses in agrammatic subjects who are speakers of the Slavic language, Serbo-Croatian. The highly inflected morphology of the Serbo-Croatian language is exploited to provide the appropriate experimental conditions for distinguishing between the two accounts and for testing specific proposals regarding difficulties in processing relative clauses.

Some initial comments about Serbo-Croatian are in order. The closed-class morphology, consisting of grammatical words and inflections, plays a somewhat different role in syntactic operations in a free word-order language, like Serbo-Croatian, than in a fixed word-order language, like English. In order to construct a grammatically correct sentence in Serbo-Croatian, words must match in gender, number, person, and noun case. This is accomplished by an appropriate suffix (an inflectional morpheme) added to the word root. In English, word order is used to indicate, for example, agent/object relations, both semantically and syntactically (e.g., "The girl pushed the boy"). Case is generally conveyed either by word order or by free-standing prepositions or pronouns in English. Case is conveyed by noun-inflections in Serbo-Croatian, however. In the absence of a consistent word order pattern, a Serbo-Croatian listener must rely on case markers and other agreement markers (subject-verb agreement, modifier-noun agreement, agreement between pronouns and their referents, etc.). Consequently, the English sentence from the example above can be translated into two Serbo-Croatian sentences having the same meaning but different word orders (e.g., "Devojčica$_{(nom)}$ je gurnula dečka$_{(accus)}$" and "Dečka$_{(nom)}$ je gurnula devojčica$_{(accus)}$"). The present study was designed to exploit this cross-language difference in the use of inflectional morphology to evaluate difficulties agrammatics experience in comprehending relative clauses. We proceed by examining relevant findings that have been reported in the literature.

Evidence from studies with relative clauses

Among the earliest evidence of a specific sentence processing deficit in agrammatism is the finding by Caramazza and Zurif (1976) of difficulties in comprehension of semantically reversible relative clause sentences. It is presently well-established that agrammatic aphasics often fail to understand correctly certain sentences with relative clauses if they are presented without the support of semantic content and/or pragmatic context. All types of relative clauses have not proven equally difficult, however. There appear to be selective difficulties on sentences with object-gap relatives, as compared to subject-gap relatives. For example, Caplan and Futter (1986) report such a pattern, based on a study of an agrammatic subject using an object manipulation test. Object-gap relatives contain a superficially empty noun phrase in object position (e.g., "The monkey that the rabbit grabbed _ shook the goat"). Caplan and Futter's subject performed more accurately with subject-gap relative clauses, i.e., where the empty noun phrase is in subject position, (e.g., "The sheep that _ pushed the cat jumped over the cow"). The authors suggest that the subject had lost the ability to interpret sentences using the rules of normal English syntax. On their view, the subject attempted to map thematic roles (agent, patient, theme, etc.) directly to linear sequences of words. This strategy could sometimes result in the correct linguistic interpretation even for subjects who lacked the relevant grammatical knowledge. This would happen with structures that conform to the canonical word order of the language in question. Canonical word order provides the right results in sentences of English that contain subject-gap relative clauses. This strategy would lead to consistent misinterpretation of sentences that depart from canonical S-V-O form of English sentences, however. One example is object-gap relatives.

The distinction between object-gap and subject-gap relatives has received a specific structural interpretation by Grodzinsky (1986, 1989). Grodzinsky explains agrammatics' comprehension difficulties within the framework of Chomsky's theory of Generative Grammar known as Government and Binding theory. One aspect of this theory is the postulation of a "trace" whenever a constituent is moved by a transformational rule from one level of representation, D-structure, to another level, S-structure. What is missing in the representations of agrammatics, according to this view, is the trace left behind by the transformation. Therefore the affected individuals are unable to maintain the crucial grammatical link between the "trace" and the moved constituent. Although Grodzinsky discussed several structures that involve constituent movement, we will be concerned here specifically with his discussion of relative clauses. One of the assumptions of Government and Binding theory is that traces are the bearers and transmitters of thematic roles. From this assumption it follows that the thematic role of a moved NP inside a relative clause will be unspecified in the absence of the trace. Accordingly, Grodzinsky proposes that agrammatics must resort to a default strategy for heuristically assigning thematic roles to disenfranchised NPs in relative clauses. In an SVO word-order language like English, the heuristic strategy assigns roles according to word order conventions: the initial NP would receive the role of agent. This strategy gives the right interpretation for sentences with subject-gap relatives such as, "The boy that kissed the girl is tall." In such a sentence, the transformation preserves the original NP order; therefore comprehension is preserved in spite of loss of traces in the S-structure representations. However, in object-gap relatives, as a result of trace deletion, the S-structure representation has two NPs preceding the verb (e.g., "The boy that the girl kissed was tall"). Because there are two possible agent candidates, the assignment of thematic roles can not be determined. Therefore, agrammatic patients should perform at chance in responding to object-gap relatives. According to Grodzinsky's theory, then, agrammatics generate complete syntactic representations except in the case of constructions that involve movement transformations, such as relative clauses and verbal passives.

A test of this conceptualization of the comprehension deficit in agrammatism is presented by Grodzinsky (1989). In this study, agrammatic subjects were tested using a sentence-picture matching task for comprehension of four types of relative clauses (embedding vs. nonembedding and subject- vs. object-gap). The results are interpreted in favor of the trace deletion account. We question whether the results do constitute unequivocal support for this hypothesis, however. For one thing, Grodzinsky's analysis is based on averaging across sentence types. Each sentence type should be considered separately, in our view. By pooling the results of two types of subject-gap relatives and comparing them with two types of object-gap relatives, and by comparing two types of embedded structures with two types of nonembedded structures, one is liable to lose sight of relevant variability. In addition, there were marked individual differences among the subjects. For example, the performance of the four subjects varied from 20-80% error in response to nonembedded object relatives. These differences cannot be explained on Grodzinsky's account.

We have presented two structurally-based accounts of agrammatic comprehension difficulties, indicating how each applies to sentences containing relative clauses. The accounts differ in their diagnosis of where within the structural apparatus the problem lies, but each assumes that critical syntactic information for the assignment of thematic roles is not available to agrammatics. We now consider how the two accounts might be differentiated empirically—Grodzinsky's specific trace deletion hypothesis and Caplan and Futter's more general syntactic simplification account—and how each, in turn, may be distinguished from the Processing Limitation Hypothesis.

Testing between the two hypotheses

Though differing in their assignment of the specific source of comprehension difficulty, each version of the Structural Deficit Hypothesis leads to specific predictions concerning the comprehension performance of an agrammatic subject. It is important to spell out the expectations in detail. (1) The affected individual would perform poorly on all sentences in which the correct interpretation depends on a full syntactic analysis that would bring into play the damaged component. Thus, if there is loss of syntactic knowledge there should be no significant variation across any sentence type that conforms to a specific syntactic pattern (this prediction would apply only if the putative syntactic loss was complete). If agrammatics construct incomplete syntactic representations, as on Grodzinsky's theory, they lack the means to determine the thematic role played by the moved NP. Therefore, they must apply a guessing strategy which should be reflected in chance performance on sentences with object-gap relatives. On the other hand, if agrammatics fail to construct hierarchical syntactic representations, but rely on simplified structures that are governed by word-order, as Caplan and Futter supposed, then, similarly, they should consistently err in responding to object-gap relatives. (2) There should be no significant variability in performance level across patients on a given sentence type. If in order to understand a particular construction, it is necessary to apply the syntactic rule that is assumed to be missing (for example, a rule for assigning thematic roles in relative clauses), all agrammatic subjects would be expected to perform deficiently (i.e., at chance, if syntactic roles are randomly assigned, or below chance, if some specific non-syntactic strategy is used). (3) If the syntactic deficit is structural, one could expect it to be syndrome-specific. Thus, a given pattern of results would characterize agrammatism but not other syndromes which differ in the underlying deficit. Agrammatic patients would be forced to rely on non-syntactic comprehension strategies and to assign a syntactic structure that deviates from that assigned by the normal population, or by another aphasic group whose syntactic problems, if any, are not identified with those of Broca-type aphasics (for example, Wernicke-type aphasics). In consequence, the pattern of performance on any structures that tax the damaged component (for example, relative clauses) should be qualitatively different in agrammatic subjects than in other aphasics, or in normal subjects. (4) If a missing structure is part of Universal Grammar, agrammatics in any language would be expected to erroneously process the critical structures. Thus, if one tests the critical syntactic structure across different languages, agrammatics from a non English-speaking population would be expected to fail in processing the missing structure just as their English-speaking counterparts.

The Processing Limitation Hypothesis makes different predictions about the comprehension difficulties associated with agrammatism. On this view, particular sentences place greater processing demands upon the language apparatus, and particular tasks further augment the difficulties imposed by these sentences. The processing limitation account makes specific predictions about the performance of agrammatic subjects. (1) Variability in performance levels across different sentence types is expected because processing difficulties are on a continuum. Agrammatic subjects are predicted to demonstrate more difficulty with syntactic structures that impose heavy demands on the processing system (e.g. object-gap relatives) as compared with structures that do not (e.g. subject-gap relatives). (2) Variability in performance levels across individuals on a given sentence type is expected, but each agrammatic subject should display the same rank order of sentence difficulty. The level of performance should vary according to the severity of each individual's processing impairment. Thus, we would expect a continuous distribution of scores across subjects, but with a consistent ordering of sentence types. (3) The relative difficulty of each syntactic structure should be the same in both aphasic subjects and normal subjects; the sentences that are most difficult for normal subjects should also be most difficult for agrammatic aphasics. Although, the pattern of performance across different syntactic structures should be the same for agrammatic aphasics and normals, the level of performance may well differ. If difficulties in comprehension are caused by a processing limitation, then we would expect that when normal subjects are pressed (e.g., by artificially speeded speech or text) they would show the same pattern of performance as agrammatic aphasics. (4) Variability in performance across languages is expected because languages use different means to accomplish the same syntactic ends. These may vary in their costs to the processing system.

The present study was designed to take advantage of the manner in which inflectional morphology is used syntactically in a free word-order language, Serbo-Croatian. Four types of relative clauses varied in their place of attachment (embedded vs. nonembedded), and in the grammatical role of the missing NP inside the relative clause (subject- vs. object-gap). These sentence types are abbreviated as SS, SO, OO, OS. The abbreviations use the first letter (S or O) to indicate the place of attachment (S = embedded, O = nonembedded). The second letter indicates the role of the missing NP (S = subject, O = object). In some relative clauses in Serbo-Croatian, as in subject-gap relatives (SS, OS) and nonembedded object-gap relatives (OO), the thematic role of an NP is determined by a noun-inflection marking the moved NP. Examples with underlined case-inflections are given in 1-3 :

(1) SS: Žena(nom) koja(nom) ljubi čoveka(accus) drži kišobran (nom, accus).
The lady who is kissing the man is holding an umbrella.

(2) OS: Žena(nom) ljubi čoveka(accus) koji(nom) drži kišobran (nom, accus).
The lady is kissing the man who is holding an umbrella.

(3) OO: Žena(nom) ljubi čoveka(accus) koga (accus) štiti kišobran (nom, accus).
The man is kissing the lady that the umbrella is covering.

Thus, in these sentence types the inflectional morphology aids in coindexation of the moved constituent and trace. However, in embedded object-gap relatives, as in (4), both NPs have the same nominative-case inflection and, therefore, thematic roles cannot be assigned by processing the noun-case inflection only.

(4) SO: Čovek_(nom) koga_(accus) žena(nom) ljubi drži kišobran_(nom, accus).
The man that the lady is kissing is holding an umbrella.

The relative pronoun of the relative clause is invariably marked by the thematically appropriate case inflection. Thus, although in SO sentences, the NPs cannot be thematically differentiated by processing only NP inflections, the thematic roles can nonetheless be differentiated by processing the relative pronoun-case inflection.

The fact that moved constituents are marked not only by traces but also by case inflections provides an additional cue for Serbo-Croatian users (which is unavailable to English users) when assessing thematic roles. It is this feature that enables us to test Grodzinsky's trace-deletion hypothesis. The trace-deletion account predicts that agrammatics will perform successfully on subject-gap relative clause sentences (OS, SS), but will be at chance on object-gap sentences (SO, OO). However, if, agrammatics have retained the inflectional morphology and are missing only traces of movement in their syntactic representations, as this hypothesis proposes, then Serbo-Croatian agrammatics are expected to have an advantage over English-speaking agrammatics because their preserved inflectional morphology would be sufficient for determining the thematic role. According to this account, therefore, Serbo-Croatian agrammatics are expected to perform at chance only on SO sentences but equally well on the other three types of relative clauses. Alternatively, if Serbo-Croatian agrammatics have intact inflectional morphology, their performance can be expected to be equally successful on all four types of relative clauses, since the relative pronoun in the SO sentences is marked by the thematically appropriate case inflection.

On the other hand, if Serbo-Croatian agrammatics are unable to make syntactic use of inflections, as an account of parsing deficiency resulting in incomplete, simplified syntactic representations would state, then noun and pronoun case inflections could not aid their comprehension of relative clauses. On this account Serbo-Croatian agrammatics are expected to err on all sentences that depart from canonical word order (SO, OO), systematically choosing the conjoined-clause interpretation instead. In consequence, Serbo-Croatian agrammatics should demonstrate a similar degree of difficulty as their English-speaking counterparts in comprehension of object-gap relatives.

In addition to testing agrammatics' ability to assign thematic roles in relative clauses we also asked whether they tend to simplify the syntactic structure of a relative clause. One possibility is that object-gap relative clauses might be treated as though they consisted of two conjoined clauses. This would be expected on the suggestion that agrammatics lose the ability to construct complete syntactic representations, regressing by default to simpler structures (e.g., Caplan & Futter, 1986). Thus, conjoined-clause sentences can be used to test the proposal that agrammatics fail in comprehension of relative clauses because they tend to simplify complex syntactic structures and employ heuristic, non-syntactic strategies (e.g., a canonical word-order strategy) when interpreting some relative clauses. Studies previously cited have focused on testing for ability to assign correct agent/patient (thematic) relations. The present study offers the first test of the possibility that agrammatics tend to simplify the complex syntax of relative clauses in certain sentences by construing them as though they contained two conjoined clauses (CC). A conjoined clause simplification of, for example, an SO sentence (4) would be:

(5) CC: Čovek ljubi ženu i drži kišobran.
The man is kissing the woman and holding an umbrella.

A test of this possibility was made by requiring subjects to choose between two pictures, one of which depicted the conjoined-clause analysis and the other the relative clause analysis. This technique was used successfully in previous research examining comprehension of relative clauses by agrammatic aphasics (Zurif & Caramazza, 1976, Wulfeck, 1988; Grodzinsky, 1989).

The present study makes use of the forgoing features of the Serbo-Croatian language to investigate comprehension of relative clauses by Serbo-Croatian speaking agrammatics. The fact that inflectional morphology plays such an important role in Serbo-Croatian syntax makes it an ideal language to contrast with English for testing theoretical claims about the basis of comprehension deficiencies in agrammatism. The experiment was designed to distinguish between versions of the Structural Deficit Hypothesis as well as between either version and the Processing Limitation Hypothesis. The study therefore addressed the following questions:

1. Are there systematic variations in performance among agrammatic subjects across different types of reversible relative clauses and conjoined clauses. Do these variations form a graded continuum or are they all-or-none?
2. Are there cross-language differences in comprehension of relative clause sentences between agrammatic speakers of a highly inflected language (Serbo-Croatian) and a fixed word-order language (English)?
3. Are there systematic differences between subject groups? Will Broca-type aphasics, Wernicke-type aphasics and normal subjects each show a distinctive pattern of errors? If a hierarchy of difficulty of sentence types is found, will it differ for the three subject groups, or will it be the same.

Method

Subjects. The aphasic subjects were seven non-fluent Broca-type aphasics (three females and four males) and five fluent Wernicke-type aphasics (one female and four males). All were outpatients of the Neurological Clinic or the Institute for Psychophysiology and Speech Pathology, in Belgrade, Yugoslavia. All were native speakers of Serbo-Croatian. The age range was 44-62 for Broca-type aphasics and 48-60 for Wernicke-type aphasics. All subjects had at least a secondary education and all were right-handed. Further details are given in Table 1.

Table 1. *Aphasic Subjects: Background Data.*

Aphasic Subjects	Sex	Age	Educ.	Etiology	Lesion
Broca subjects					
S.P.	M	53	16	CVA (1981)	L inf. frontal at the depth of the ventricle
D.R.	M	62	16	CVA (1983)	Large subcortical Broca's area, L motor strip, parietal area, patchy Wernicke's area
V.P.	M	46	12	CVA (1986)	Cortical and subcortical Broca's area
D.T.	F	52	16	CVA (1984)	L basal ganglia and int. capsule
A.T.	M	46	10	CVA (1985)	L frontal, lower motor cortex
V.M.	F	44	14	CVA (1986)	L inf. fronto-temporal cortex
M.J.	F	47	10	CVA (1985)	L inf. frontal
Wernicke subjects					
M.C.	M	57	14	CVA (1983)	L subcortical tempo-parietal, supramarginal and angular gyri
A.B.	M	59	14	CVA (1985)	L fronto-parietal cortex and basal ganglia
V.V.	M	48	16	CVA (1987)	L fronto-parietal cortex
M.B.	M	59	16	CVA (1981)	L temporo-parietal cortex
D.D.	F	60	12	CVA (1980)	L temporo-parietal cortex

The control group comprised seven neurologically normal subjects (four females and three males), roughly matched to the aphasic group in age and years of education.

All patients were categorized according to a neurological examination, results of a CT-scan, and the results of tests of language function based on the Serbo-Croatian version of the Boston Diagnostic Aphasia Examination (BDAE, Goodglass, & Kaplan, 1972). The etiology in all cases was a single cerebrovascular accident confined to the left cerebral hemisphere. Time since onset of the symptoms varied from six months to seven years. There was no history of drug abuse, and no significant disabilities in vision or hearing among either the patients or the control subjects.

All Broca-type patients showed the characteristic non-fluent speech and all displayed some degree of agrammatism (see Table 2 below). Their sentences were short with impoverished syntactic structure, consisting mainly of nouns and verbs with frequent omission of free-standing functors and occasional substitution of bound morphemes. A common error was to use the nominative case, in place of the appropriate noun case. All Broca-type subjects had measurable losses in language comprehension when tested with the BDAE. (Results by individual subjects on the comprehension subtests of the BDAE are given in Appendix A).

The Wernicke-type aphasics had fluent speech with an apparently normal melodic line. Their sentences were rife with semantic and phonetic paraphasias and paragrammatically inappropriate grammatical forms. Their comprehension was markedly impaired as measured with the BDAE. (Results by individual subjects on the comprehension subtests of the BDAE are give in Appendix A).

Materials. In designing semantically-reversible sentences containing relative clause, steps were taken to minimize possible difficulties in pragmatic interpretation that these sentences might induce: (a) to this end only two animate noun phrases were allowed in each sentence (in contrast, for example, to Caplan and Futter, 1986); (b) semantic relations among noun phrases were always plausible; (c) the third noun phrase in each test sentence was inanimate and conveyed descriptive information. The last restriction was imposed because findings with young children have shown that performance on an act-out task improved when the number of animate noun phrases in relative clause sentences was reduced from three to two (Goodluck & Tavakolian, 1982).

Experimental sentences

Four types of semantically reversible relative clause sentences were created and recorded on audiotape. The relative clauses varied in their place of attachment (embedded vs. nonembedded), and in the role of the missing noun phrase inside the relative clause (subject- vs. object-gap). See 1 - 4 above.

Control sentences

In addition to relative clause sentences, conjoined-clause (CC) sentences were included in the test materials. As noted, CC sentences have structures that are hypothesized to be syntactically less complex than relative clause sentences and are considered to be mastered earlier in development (Tavakolian, 1981). The CC sentences were derived from OS sentences. Each contained one empty noun phrase in the second clause, which is coreferential with the subject of the first clause, as illustrated below.

(6) CC: Žena drži kišobran i ljubi čoveka.
The lady is holding the umbrella and kissing the man.

Additional sentences, were added as controls to ascertain that the subjects were attending to the entire sentence. These control sentences were of the same form as three of the sentence types (SS, OS, CC), but their respective foils differed. The picture foils for all sentence types are described later.

Picture materials

Given that the task is a forced choice among alternative pictures, the design and choice of picture materials is critical. Steps were taken to create pictures depicting possible nonreversible situations. The so-called "felicity conditions" (Hamburger & Crain, 1982) were met by providing a natural context for the relative clause. This was accomplished by depicting more than one character corresponding to the head NP. These felicity conditions were not met in the sentences used to test comprehension in previous studies of aphasia.

A two-choice picture task was adapted from materials constructed by Smith et al. (1989). Both picture choices depicted plausible events. Since a relation between two animate noun phrases was depicted in each picture, the location of agents (left or right side of the picture) was randomized within sentence sets. In half of the arrays, the correct picture was in the top position, and in the other half the correct picture was in the bottom position. (Sample test materials for an experimental sentence are displayed in Appendix B).

Picture foils

The conjoined-clause analysis was used as the picture-foil, that is, the correct interpretation of SO, OS, and OO sentences was contrasted with foils depicting the conjoined-clause analysis interpretation. This misanalysis was chosen for the reasons indicated above.

The following examples are descriptions of correct target pictures and the incorrect foils that were used for stimulus sentences. :

(7) SO:[2] The man that the lady is kissing is holding an umbrella.
Target picture: a man holding an umbrella while a lady is kissing him.
Foil picture: a man holding an umbrella and kissing a lady

(8) OS: The lady is kissing the man who is holding an umbrella.
Target picture: a lady kissing a man while this man is holding an umbrella
Foil picture: a lady kissing a man and holding the umbrella

(9) OO[3]: The man is kissing the lady that the umbrella is covering.
Target picture: a man is kissing a lady while she is protected by an umbrella
Foil picture: a man is kissing a lady and he is protected by an umbrella

For the SS sentences a conjoined-clause analysis would yield the same result as interpretation of the relative clause. Therefore, a foil depicting a main clause only interpretation was used for the SS sentences (10).

(10) SS: The lady who is kissing the man is holding an umbrella.
Target picture: a lady while holding an umbrella is kissing a man
Foil picture: a lady is holding an umbrella

For the CC sentences, however, the foil depicted an erroneous minimum-distance principle interpretation (11).

(11) Stimulus sentence (CC): The man is kissing the lady and holding an umbrella.
Target picture: a man kissing a lady and holding the umbrella
Foil picture: a man kissing a lady and the lady holding an umbrella

For the control SS and OS sentences a relative-clause only interpretation was depicted in the foil. Finally, a first-clause-only interpretation was used for the control CC sentences.

Test design. The test contained 65 sentences: 10 sentences in each set (OO, SO, SS, OS, CC), plus 5 sentences in each set of foil-control sentences (SS, OS, CC). Two test orders were prepared, with the control sentences interspersed randomly. Practice trials consisting of four sentences and their picture sets were used to familiarize subjects with the procedure.

Procedure. When performing a sentence-picture matching task, the subject is asked to listen to each sentence and then to decide which picture, among simultaneously present alternatives, depicts the meaning of the sentence correctly. The dependent variable is error rate since performance on this task is not timed. Subjects were tested individually in a single, one-hour session. Before each sentence was presented, the picture array was exposed. A practice session was administered to familiarize subjects with the materials and the procedure. Subjects were instructed to listen carefully to the entire sentence, to look carefully at both pictures in the array, and then to point to the picture that matched the meaning of the sentence.

Results

There is clear separation between the subject groups on overall accuracy. The Broca subjects averaged 22% errors, (range 10-34%), Wernicke subjects averaged 37% errors, (range 28-54%), and normal control subjects averaged 6% errors (range 2-10%). Thus the Wernicke subjects were more impaired in sentence-picture matching of relative clause sentences than Broca subjects or normals.

Since the task consisted of two-picture choices, chance performance would be 50%. We define chance performance conservatively: as an error rate between 40 and 60%. Error rates less than 40% were considered to be above chance, whereas error rates above 60% were considered to reflect systematic application of a nonlinguistic strategy.

Table 2 displays the mean number of errors by individual subjects. Although all subjects demonstrated better comprehension of subject-gap sentences than of object-gap sentences, there is much individual variability in error rates, with Wernicke subjects performing overall worse than Broca subjects. All Broca subjects exhibited overall above-chance ability to match the correct picture to the experimental sentence. On the SO sentences the Broca subjects manifested performance that ranged from highly above chance (20% error) to chance (60% error). Four of the seven Broca subjects performed with an above chance success rate on this sentence type.

Table 2. *Percent of Errors on each Sentence Type for Broca and Wernicke Subjects.*

	Sentence type					
Individual Subjects	OO	OS	SO	SS	CC	Mean
Broca Aphasics						
S.P.	40*	10	30	10	30	24
D.R.	30	30	60*	20	30	34
V.P.	10	10	30	20	20	18
D.T.	60*	40*	60*	0	0	32
A.T.	30	20	40*	10	10	22
V.M.	20	10	20	0	0	10
M.J.	10	20	30	0	10	15
Mean	29	20	40*	9	14	
Wernicke Aphasics						
M.D.	70	40*	60*	50*	50*	54*
A.B.	50*	50*	50*	10	50*	42*
V.V.	40*	30	60*	20	30	33
M.Dj.	50*	20	80	0	0	30
D.D	40*	20	60*	10	10	28
Mean	50*	32	62*	18	28	

*chance performance

However, subject D.T. performed at chance on all object-gap relatives (OO, SO), and, in addition, on some of the subject-gap sentences (OS). All five Wernicke aphasics performed at chance level on the SO sentences. Moreover, one subject (M. Dj.) chose the conjoined clause option very frequently (80% error). Another Wernicke subject (M.D.) performed at chance on all sentence types, and a third (A.B.) performed at chance on all sentence types except the SS sentences. The task was evidently too difficult for these latter subjects, so that they judged sentences in a random manner. For the OO sentences the mean error rate was smaller but there was high variability. The range of errors for Broca patients was 10 - 60% and for the Wernicke patients 40 - 70%. Only two Broca patients performed at chance, whereas all of the Wernicke patients did so. The pattern of performance within each aphasic group (with exception of two Wernicke patients who performed equally poorly on all sentence types) shows the same hierarchy of sentence difficulty.

Factorial analyses of variance were performed separately on the experimental and conjoined-clause control sentences, and on the foil-control sentences. Since there was no effect of test order on the accuracy score, the data from both orders were combined for analysis. The error scores were analyzed by an ANOVA which compared the factors of Group (Broca, Wernicke, Control) and Sentence type (OO, OS, SO, SS, CC). Both main effects were significant. The main effect of Group ($F(2,16) = 26.35$, $p < .001$) indicates that there were differences between types of aphasia and the normal control group. The significant effect of Sentence type ($F(4, 64) = 21.83$, $p < .001$) indicates that all sentence types were not equally difficult. The interaction between Group and Sentence type was also significant ($F(8, 64) = 2.39$, $p < .02$). Its interpretation will be considered presently.

A post hoc Tukey test ($p = .01$) indicated that each subject group was significantly different from the others with the normal control group exhibiting the fewest errors and the Wernicke group exhibiting the most.

The rank order of difficulty for the sentence types was similar in both aphasic groups: The SO sentences were the most difficult. Three of the seven agrammatic subjects performed at chance level on these sentences, and four performed with above-chance success. The SO sentences were the most difficult for all the subjects including the control subjects, although the Sentence-type effect did not reach significance in this group because performance was at the ceiling level. A Post hoc Tukey test ($p = .01$) indicated that there were significantly more errors on SO sentences than on all others, with the exception of the OO type, from which they differed only at the $p = .05$ level. More errors occurred on the OO type than on either SS or CC sentences. The latter were not significantly different from each other. OS sentences gave rise to more errors than SS sentences but did not differ from OO or CC types. As was expected, the control CC sentences and the SS sentences were the easiest for all three groups of subjects.

The mean percent error per sentence type for the three groups of subjects is displayed in Figure 1. The figure shows the same pattern of performance across sentence types in all three subject groups. When the Broca group and the Wernicke group were compared, there was a significant effect of aphasia type ($F(1,11)= 9.19$, $p < .01$), but no aphasia-type by sentence-type interaction. The most difficult sentence type, the SO sentences, produced the only significant difference between aphasic groups ($p < .02$). Group differences on OO and OS sentence types were in the same direction, but failed to reach significance (each with $p < .09$). There was no significant difference between the two patient groups in the number of errors on control CC sentences and SS sentences.

Given the absence of interaction of type of aphasia and sentence type, we should ask why an interaction with subject group was obtained in the analysis that included all subjects. Figure 1 shows that few errors were made by control subjects; for them the plot of errors against sentence type is relatively flat. Thus, the presence of an interaction in the composite analysis is clearly attributable to the ceiling level performance of the control group.

Foil-control sentences

On the foil-control sentences both aphasic groups performed at a high level of accuracy. Broca's aphasics averaged 4% errors (range 0-13%); Wernicke's aphasics averaged 7% errors (range 0-20%). The difference was not significant ($F(1,10)= .46$, $p < .51$). The control group performed with 100% accuracy on these sentences.

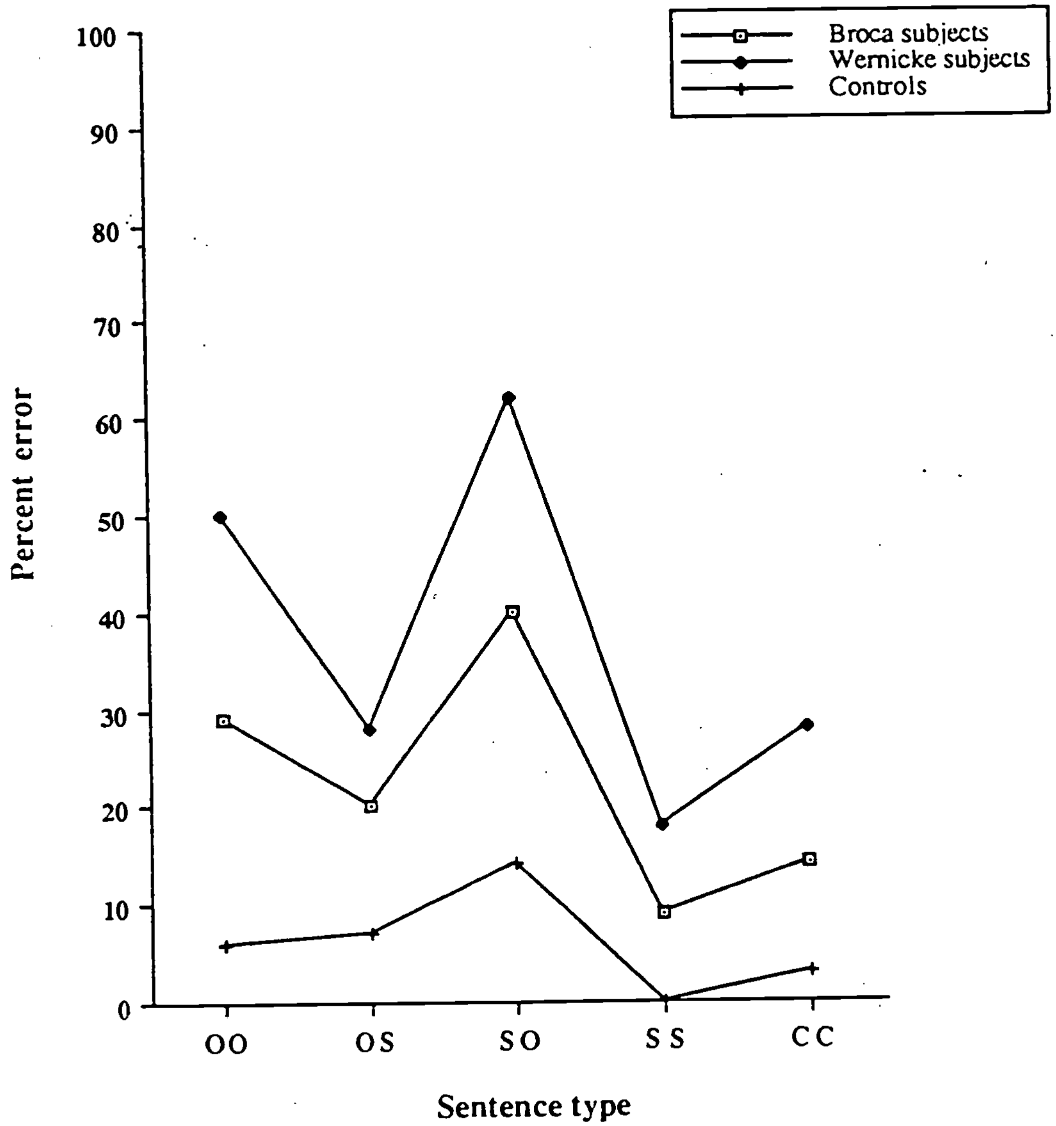


Figure 1. **Percentage of Errors by Sentence Type.**

DISCUSSION

Our purpose was to obtain evidence that could distinguish between the explanatory adequacy of two accounts of comprehension impairment in agrammatism. One explanation, the Structural Deficit Hypotheses, states that syntactic structures critical for sentence interpretation are lost or are unavailable. The other explanation appeals to a processing deficiency. To distinguish the two hypotheses we have studied one complex structure intensively, the relative clause. This structure is well suited also to our additional goal of bringing data from different languages bear on the problem.

This study is the first to present data from an inflected language on comprehension of a complete set of relative clause structures by Broca's and Wernicke's aphasics. The Structural Deficit Hypothesis predicts that agrammatics would fail to assign correct syntactic representations to relative clauses. If the requisite structures are lost, agrammatics would have no recourse but to apply nonsyntactic strategies. In the case of object-gap relatives, they might be expected, to assign thematic roles randomly (Grodzinsky, 1989) or to apply a canonical word-order strategy indiscriminately (Caplan & Futter, 1986). We therefore asked whether agrammatics do, in fact, lack the necessary syntactic structures to analyze object-gap relative clauses, and, if so, whether they tend to simplify these structures by treating them as conjoined clauses. The first question was addressed by comparing the comprehension of relative clauses that differed in place of attachment (i.e., embedded and nonembedded relatives) and in the location of the gap (i.e., subject- and object-gap relatives). This was accomplished by exploiting particular features of the grammar of Serbo-Croatian, taking advantage of the fact that Serbo-Croatian marks thematic roles in relative clauses by case inflections. This characteristic enabled us to tease apart two possible sources of syntactic deficiency: syntactic simplification amounting to loss of hierarchic structure and deletion of the traces of movement. The second question was addressed by using conjoined-clause sentences as controls, a syntactic structure that could plausibly result from simplification of a relative clause. Accordingly, four types of reversible relative clauses and conjoined clauses provided the critical materials for testing between the two hypotheses.

Notably, the agrammatic aphasics found the different relative clause structures to be unequal in difficulty. Object-gap relatives yielded the highest error rates, in keeping with the earlier findings with English-speaking agrammatics (Caramazza & Zurif, 1976; Grodzinsky, 1989; Caplan & Futter, 1986), and in agreement with both Grodzinsky's and Caplan and Futter's predictions concerning the expected order of difficulty. If one were to draw conclusions about agrammatic subjects' competence only by taking average performance into account, one might be led to conclude that the subjects of the present study had lost a portion of their grammatical knowledge and, consequently, were forced to rely on nonlinguistic strategies. However, the prediction of the Structural Deficit Hypothesis that agrammatics would perform at chance on object-gap sentences was not met, at least for a majority of subjects. Four out of seven of the agrammatic subjects performed well above chance on these structures. Thus, this result offers, at best, only partial support for the theoretical conceptions of Caplan & Futter and Grodzinsky.

It could be expected that trace deletion in object-gap relatives should impair Serbo-Croatian speaking subjects less than English speaking subjects, since critical information about the subject/object distinction can be extracted from noun case inflections regardless of word order. But these subjects, like their English-speaking counterparts would be expected to be at chance on SO sentences if they lacked traces, even if they were able to rely on noun-phrase inflectional morphology. That is because, as we explained, these sentences have ambiguous inflectional markings because both noun phrases are marked for the nominative case. On the other hand, performance on OO sentences should not differ from subject-gap relatives because on these sentences the inflections indicate thematic roles unambiguously.[4]

However, although in all subjects object-gap relatives gave rise to significantly more errors than subject-gap relatives neither of these expectations based on Grodzinsky's hypothesis (1989) finds support in the data. Concerning the first prediction, although SO sentences were more difficult than OO sentences, performance was at chance only for three of the seven agrammatic subjects. One of those (D.T.) who performed at chance on SO sentences also performed at chance on OO and OS sentences, which indicates that this subject was not taking advantage of the inflectional morphology. Concerning the second prediction, the OO sentences were comprehended with more errors than the subject-gap sentences. Additionally, the comprehension difficulties in these agrammatics cannot be explained on the trace deletion account, since difficulties in comprehension were also present to a lesser degree on other structures (subject-gap relatives) which would be expected on the trace deletion account to be analyzed normally.[5]

On the other hand, diffuse difficulties are expected when there is a special limitation in processing. Only on the processing limitation account is it expected that the comprehension difficulties in agrammatics would be most severe on specific syntactic structures, but also present to a lesser degree with other structures.

The results of the present study lend no support to Caplan and Futter's (1986) conjecture that the syntactic apparatus of agrammatics has undergone simplification that would necessitate use of a word-order strategy in the absence of syntactic parsing. On this proposal, agrammatics should choose the erroneous conjoined-clause interpretation on all relative clauses that fail to preserve canonical word order (OO, SO). That clearly did not happen. The possibilities for varying word order that are permitted by the grammar of Serbo-Croatian enabled us to hold word-order constant and to construct OS and OO sentences with the same sequencing of NPs and VPs. Keeping the same word order in OO and OS sentences would induce agrammatics to be incorrect on the OO sentences as often as they are correct on the OS sentences if they relied solely on a word-order strategy. Although, in fact, the agrammatics produced more errors on OO sentences than on OS sentences, their performance on OO sentences was above chance. They were successful in distinguishing different syntactic structures even though these structures had the same noun - verb sequences. Above-chance interpretation of the OO sentences is incompatible with the hypothesis that these agrammatics were using a linear word-order strategy.

A further test between the differing accounts of comprehension disorder in agrammatic aphasia was made by comparing the performances of agrammatic subjects with those of Wernicke's aphasics, and with the neurologically-normal control group. The Processing-Limitation Hypothesis predicts that the pattern of performance across sentence types should be consistent across subject groups. The present results showed that, in fact, both agrammatic Broca- and Wernicke-subjects experienced difficulties with interpretation of semantically-reversible relative clauses. The significant group differences were differences of degree, not qualitative differences in pattern. Both types of aphasic subjects were similarly affected by the variations in syntactic structure that were introduced by inclusion of different types of relative clauses; the same rank order of difficulty among sentence types was found for each group with the object-relatives (SO, OO) being the more difficult structures. This order of difficulty was also observed in the normal control group. The agrammatic group performed better overall than the Wernicke group, creating the significant between-group interactions. Finally, a consistent pattern of performance across sentence types was observed in the individual subject data.

In this connection it is relevant to note that the ordering of difficulty on the four relative clause structures obtained in this study is consistent with that which has been found in several research studies that explored acquisition of relative clauses in young children (deVilliers et al. 1979; Tavakolian, 1981). In addition, this pattern has also been demonstrated in dyslexic children (Smith et al. 1989; Crain et al. 1990). These similarities are unlikely to be mere coincidences. The existence of parallel findings across such diverse groups fits with the failure of the present study to find evidence for a syndrome-specific comprehension deficit in agrammatism.[6]

Taken together, these findings give us reason to prefer an account of agrammatism that appeals to damaged processors in favor of an account that evokes loss of grammatical structures. The processing limitation account is to be preferred because it shows itself capable of tying together a wider variety of findings. An adequate theory of "agrammatic" comprehension would account for syntax-related variations in performance among agrammatics and for parallels across different populations.

It remains to give positive underpinnings to the proposal that the source of the comprehension deficit in aphasia is in processing limitations. The question becomes: What is the origin of the difficulties in sentence processing that cause comprehension failures? Elsewhere, we have developed a proposal that appeals to an extended version of Fodor's (1983) modularity hypothesis (Crain & Shankweiler, 1990). On this view, language processing is carried out in discrete stages, organized in a hierarchical, bottom-up fashion (Forster, 1979). Language comprehension involves a series of translations between levels of representation: phonological, syntactic and semantic. A processing limitation view of aphasics' comprehension difficulties assumes at least two distinct ways in which syntactic processing could be impaired. On one possibility, although syntactic knowledge is preserved, its access and utilization are restricted during the process of parsing in which syntactic structures are assigned to the incoming string of lexical categories. One consequence of impaired parsing capacity may be that agrammatics, contrary to normals, do not access syntactic information in an automatic fashion but via a slow and controlled process (Kolk & vanGrunsven, 1985; Friederici et al., 1992; Zurif et al., 1993). Under this assumption, although the input to the syntactic parser is normal, processing at the syntactic level is disrupted.[7] On the other possibility (which we have discussed at length elsewhere, Crain et al., 1990), difficulties in processing syntax ultimately derive from deficient phonological input and memory processes. The phonology is especially vulnerable because it is the first level at which the input engages the language apparatus. Under this assumption, although syntactic processing per se is intact, the input to the parser is deficient, thus resulting in comprehension failure.

This notion is consistent with our findings concerning sentence comprehension in young children, normal adults, and the reading impaired, we would like to suggest a possible source of comprehension deficits in agrammatism that could also account for parallel findings in other populations. One resource limitation that young children, dyslexics, and aphasics may have in common concerns the processing of phonological input. The processing account can provide a basis for the observed parallels in syntactic comprehension performance. If the hypothesis is correct, then any group that suffers from a bottleneck in phonological processing, for any reason, should display limitations at sentence level.

In a subsequent study we obtained confirmation of phonological deficiencies in some of our agrammatic subjects. The capacity of phonological short-term memory was tested in the same Broca-type aphasics that participated in the present study (Lukatela & Shankweiler, 1990). The aphasic and control subjects were compared on verbal and nonverbal retention of rhyming and nonrhyming word strings, and nonsense drawings, testing in each case, memory for serial order. The results indicated that these subjects had a material-specific deficit in short-term retention. They differed from the control group on word strings but not on nonsense drawings. The deficit was exacerbated when all the words were phonologically similar (rhyming). Arguably, a phonological processing deficiency of this nature could impair the working memory sufficiently to impede comprehension, at least for sentences that are likely to require re-analysis (McCarthy & Warrington, 1987). These findings, therefore, lend further substance to the speculations that our agrammatic subjects' comprehension difficulties may stem at least in part from deficiencies in phonological processing that curtail the efficient use of working memory in comprehension tasks.[8]

In sum, the comprehension difficulties encountered by the agrammatic aphasics are more consistent on several counts with a Processing Limitation Hypothesis than with a Syntactic Deficit Hypothesis involving trace deletion or failure to interpret grammatical inflections: 1) Comprehension accuracy for agrammatic speakers of Serbo-Croatian was overall above chance for all types of relative clauses. 2) Agrammatics' difficulties in comprehending these structures proved remarkably similar to those displayed by Wernicke's aphasics: the pattern of performance did not distinguish syndromes. 3) Comparisons from the literature based on dyslexic children and normals tested under stressful conditions reveals an order of difficulty of relative clause structures consistent with that displayed by the aphasics. 4) Agrammatics can succeed in detecting syntactic violations of these and other sentence types with a high degree of accuracy,[9] and they are sensitive to syntactic priming. 5) There is independent evidence of phonological impairments in the agrammatic subjects of this study.

REFERENCES

Bates, E., Friederici, A., & Wulfeck, B (1987). Comprehension in aphasia: A cross-linguistic study. *Brain and Language*, *32*, 19-67.

Berndt, R. S., & Caramazza, A. (1980). A redefinition of Broca's aphasia: Implications for a neuropsychological model of language. *Applied Psycholinguistics*, *1*, 225-278.

Bradley, D. Garrett, M., & Zurif, E. (1980). Syntactic deficits in Broca's aphasia. In D. Caplan (Ed), *Biological studies of mental processes*. Cambridge, MA: MIT Press.

Caplan, D., & Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain and Language*, *27*, 117-134.

Caplan, D., & Waters, G. (1990). Short-term memory and language comprehension: A critical review of neuropsychological literature. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. (pp. 337-389). Cambridge: Cambridge University Press, .

Caramazza, A., & Zurif, E. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*, 572-582.

Caramazza, A., Basili, A. G., Koller, J., & Berndt, R. S. (1981). An investigation of repetition and language processing in a case of conduction aphasia. *Brain and Language*, *14*, 235-271.

Crain, S., Shankweiler, D., Macaruso, P., & Bar-Shalom, E. (1990). Working memory and sentence comprehension: Investigations of children with reading disorder. In G.Vallar & T. Shallice (Eds.), *Impairments of short-term memory*. Cambridge: Cambridge University Press.

Crain, S., & Shankweiler, D. (1990). Explaining failures in spoken language comprehension by children with reading disabilities. In D. Balota, F. d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading*, Lawrence Erlbaum Publishers.

deVilliers, J. G., Tager-Flusberg, H. B., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *8*, 499-518.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Friederici, D. A., Wessels, M.J., Emmorey, K., & Bellugi, U. (1992). Sensitivity to inflectional morphology in aphasia: A real-time processing perspective. *Brain and Language*, *43*, 747-763.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper & E. C. T. walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia: Lea & Febiger.

Goodluck, H., & Tavakolian, S. (1982). Competence and processing in children's grammar of relative clauses. *Cognition*, *11*, 1-27.

Grodzinsky, Y. (1986). Language deficits and the theory of syntax. *Brain and Language*, *27*, 135-159.

Grodzinsky, Y. (1990). *Theoretical perspectives on language deficits*. Cambridge, MA: MIT Press.

Grodzinsky, Y. (1989). Agrammatic comprehension of relative clauses. *Brain and Language*, *37*, 480-499.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj, II (Ed.), *Language development, Volume 1: Syntax and semantics*. Hillsdale, NJ: Erlbaum.

Hickok, G., Zurif, E., & Canseco-Gonzales, E. (1993). Structural description of agrammatic comprehension. *Brain and Language*, *45*, 371-395.

Kean, M. L. (1977). The linguistic interpretation of aphasic syndromes. *Cognition*, *5*, 9-46.

Kolk, H. H. J., & vanGrunsven, M. M. F. (1985). Agrammatism as a variable phenomenon. *Cognitive Neuropsychology*, *2*, 347-384.

Lapointe, S. G. (1983). Some issues in the description of agrammatism. *Cognition*, *14*, 1-39.

Linebarger, M., Schwartz, M., & Saffran, E. (1983). Sensitivity to grammatical structures in so-called agrammatic aphasics. *Cognition*, *13*, 361-392.

Lukatela, K., Crain, S., & Shankweiler, D. (1988). Sensitivity to inflectional morphology in agrammatism: Investigation of highly inflected language. *Brain and Language*, 33, 1-15.

Lukatela, K., & Shankweiler, D. (1990). Working memory, phonological decoding and agrammatic aphasia. Paper presented at the Academy of Aphasia, Baltimore.

Lukatela, K., Ocic, G., & Shankweiler, D. (1991). Syntactic priming of inflected nouns in agrammatic aphasia. Poster presented at the Academy of Aphasia, Rome.

Lukatela, K., Shankweiler, D., & Crain, S. (1988). Sentence comprehension in fluent and non-fluent aphasia. Paper presented at the Academy of Aphasia, Montreal.

Martin C. R., & Feher, E. (1990). The consequences of reduced memory span for the comprehension of semantic versus syntactic information. *Brain and Language*, *38*, 1-20.

Mauner, G., Fromkin, A. V., & Cornell, T. L. (1993). Comprehension and acceptability judgments in agrammatism: Disruptions in the syntax of referential dependency. *Brain and Language, 45*, 340-370.
McCarthy, R. A., & Warrington, E. K. (1987). Understanding: A function of short-term memory? *Brain, 110*, 1565-1578.
Miceli, G., Mazzucchi, A., Menn, L., & Goodglass, H. (1983). Contrasting cases of Italian agrammatic aphasia without comprehension disorder. *Brain and Language, 19*, 65-97.
Milekic, S. (1993). *Distance and position effects in sentence processing by agrammatic aphasics.* Unpublished doctoral dissertation, University of Connecticut, Storrs.
Ni, W. (1988). An investigation of grammatical knowledge and processing mechanisms using RSVP. *University of Connecticut Working Papers in Linguistics*, 69-81.
Schwartz, M. F., Saffran, E. M., and Marin, O. S. M. (1980). The word order problem in agrammatism. I Comprehension. *Brain and Language, 10*, 249-262.
Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorders: A modular approach. *Cognition, 24*, 139-168.
Shankweiler, D., Crain, S., Gorell, P., & Tuller, B. (1989). Reception of language in Broca's aphasia. *Language and Cognitive Processes, 4*, 1-33
Smith, S., & Bates, E. (1987). Accessibility of case and gender contrasts for assignment of agent-object relations in Broca's aphasics and fluent anomics. *Brain and Language, 30*, 49-60.
Smith, S., Macaruso, P., Shankweiler, D., & Crain, D. (1989). Syntactic comprehension in young poor readers. *Applied Psycholinguistics, 10*, 429-454.
Tavakolian, S. L. (1981). The conjoined-clause analysis of relative clauses. In S. L. Tavakolian (Ed.), *Language acquisition and linguistic theory*. Cambridge, MA: MIT Press.
Tzeng, O., Chen, S., & Hung, D. (1991). The classifier problem in Chinese aphasia. *Brain and Language, 41*, 187-202.
Wulfeck, B. B. (1988). Grammaticality judgments and sentence comprehension in agrammatic aphasia. *Journal of Speech and Hearing Research, 31*, 72-81.
Zurif, E. (1984). Psycholinguistic interpretation of the aphasias, In D. Caplan, A. R. Lecours, & A. Smith (Eds.), *Biological perspectives on language*. Cambridge, MA: MIT Press.
Zurif, E., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. *Brain and Language, 45*, 448-464.

FOOTNOTES

[*] *Brain and Language*, (1995), *49*, 50-76.
[†] Also University of Connecticut, Storrs.
[1] The term "processing impairment" has been used differently by different authors. Tzeng, Chen, and Hung (1991) use the term much as we do to refer to an account of language breakdown based on deficits in the processes by which a preserved knowledge base is accessed and deployed.
[2] For the SO sentences there are two possible erroneous conjoined-clause analyses One of these was the most commonly observed conjoined-clause response in studies with children (Tavakolian, 1981). Therefore this response type was selected to be the foil for this sentence type.
[3] The OO sentences, like the SO, offer two conjoined-clause analyses. Again, young children choose one conjoined clause response more often than the other (Tavakolian, 1981) and that is why it was used as a foil.
[4] Grodzinsky's description of agrammatic *production* proposes a more inclusive deficit at the level of S-structure than the putative deficit underlying comprehension disorder (Grodzinsky, 1990). On this account nonlexical terminals are deleted, including case markers and other aspects of inflectional morphology as well as traces of movement. If this description of the agrammatic deficit were extended to comprehension, additional difficulties in interpretation of the inflectional morphology would be anticipated.
[5] A recent paper by Hickok, Zurif, and Canseco-Gonzales (1993) also reports that agrammatics can experience difficulties with subject relatives. In addition to their failure on object-gap relatives, Hickok et al.'s subjects showed a deficit in comprehension of matrix sentences in subject relatives. The latter finding, the authors note, is incompatible with the trace-deletion hypothesis as framed by Grodzinsky (1990).
[6] The processing account gains further support from the finding that the performance gap between agrammatics and normal subjects can be eliminated when normal speakers are tested in a way that places them under a heavy processing load. Word-by-word reading, in which previous words disappear as new ones come into view, is a technique that was employed by Ni (1988) in studies of sentence processing in normal adult subjects. The task was to detect an anomalous word which occurred at the beginning, middle or end of the test sentence. Comparing the results of Ni's study with the results obtained by Shankweiler et al. (1989) with agrammatic aphasics who were tested (in a listening test) with the same set of sentences, we find a similar pattern of latencies and errors across structures. The strong similarities between the normal subjects under

pressure and the aphasics led Ni to infer that each group used syntactic mechanisms in the same way. The same conclusion was reached by Milekic (1993) in a comparison between performance patterns of Serbo-Croatian speaking agrammatics and normals in a word-by-word reading test. Milekic's findings yield further evidence that sentence processing in normal subjects can be profoundly affected by variation in processing demands, and that the relative difficulty of different structures mirrors the pattern exhibited by agrammatic aphasics.

[7]Recently, we have begun to use other paradigms to assess specific syntactic abilities in agrammatic subjects (Lukatela, Ocic, & Shankweiler, 1991). In an on-line study that used the syntactic priming paradigm, the same agrammatic subjects that participated in the present study, demonstrated sensitivity to syntactic priming when case was primed by a preposition. These results, however, do not necessarily indicate intact on-line sentence processing ability given that priming was demonstrated within a "minimal" syntactic context; a test trial consisted of a sentence fragment.

[8]The research literature presents a confusing picture of the relationship between sentence comprehension and working memory. Researchers have often noted that some aphasic patients with a severely restricted phonological short-term store are sometimes capable of sentence comprehension at a level far exceeding what would be expected on the basis of their span limitations (Martin & Feher, 1990; Caplan & Waters, 1990; McCarthy & Warrington, 1987). However, it is important to note that memory in these studies was assessed by measuring patients' span for unorganized material. In our theoretical model of working memory we have assumed that there are two components: a storage buffer and a mechanism whose primary task is to relay the results of lower-level analyses of linguistic input upward through the language apparatus (Shankweiler & Crain, 1986; Crain et al., 1990). In the studies cited above the patients may have suffered impairment of only one of the proposed memory components, the storage buffer, and have preserved a relay mechanism which maintained the ability to synchronize information flow.

[9]The subjects of the present study were tested on another occasion for detection of syntactic violations in relative clause sentences that were structurally identical to those of the sentence picture matching study (Lukatela, Shankweiler, & Crain, 1988). Although the results of the grammaticality judgment test cannot be compared directly with those of the sentence picture matching test, it is instructive to note that performance on judgments was appreciably more accurate averaging 90.6% correct for subject-gap relatives and 85.5% for object-gap sentences.

APPENDIX A

Aphasic Subjects: Comprehension Data

Broca Subjects	BDAE Comprehension A (15)	B (12)	C (10)	Speech Production Description of "Cookie theft" picture
S.P.	14	6	7	Kujna. Mama pere...ovaj tanjir. A ovaj decak i devojcica. Ova je voda pri..pri..E, voda je pr-li-la. Kitchen. Mama is washing...this plate. Boy and girl. This water..is li..li..Water is li-king.
D.R.	6	3	2	Seka...kolač...Majka...Casě...Vodu Girl...cookie...Mother...Glasses..Water
V.P.	10	9	6	Mama...pe-re. Sestra i brat. Ne mogu da kazem. Vidi ovde... Mama...wa-shing. Sister and brother. I can not say. Look here....
D.T.	11	9	7	Uzimaju kolače. Bo..bori se da ne..Tu..mama..pere. Taking the cookies. Is try-trying not to... Here..mama..washing.
A.T.	10	11	9	Ovde žena pere sudove a klin-ci se igraju. Jedan pao sa sto-lice. To..pere sudove. Here a woman washing dishes, kids are playing. One is fallen from the chair. This.. washing dishes.
V.M.	15	10	10	Deca uzimaju kolače. Devojčica gleda. Decăk je pao. Majka pere a voda curi. Children are taking cookies. The girl is watching. The boy is fallen. Mother is washing and water is liking.
M.J.	13	10	9	Majka pere sudje. Deca se igraju. Uzi-maju keks. Stolica..stolica..seka i braca..Pašće sa stolice. Mother is washing dishes. Children are playing. They are taking cookies. The chair..chair..brother and sister..They are going to fall from the chair.

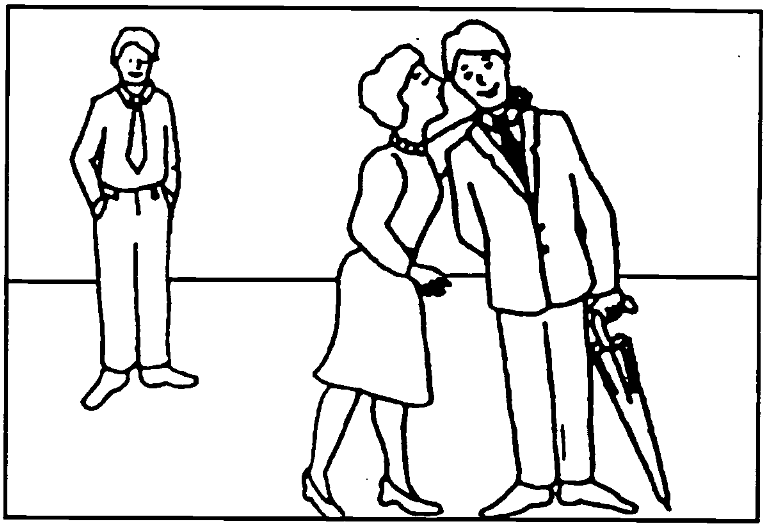
A - Commands (max. score 15)
B - Complex Ideational Materials (max. score 12)
C - Reading Sentences and Paragraph (max score 10)

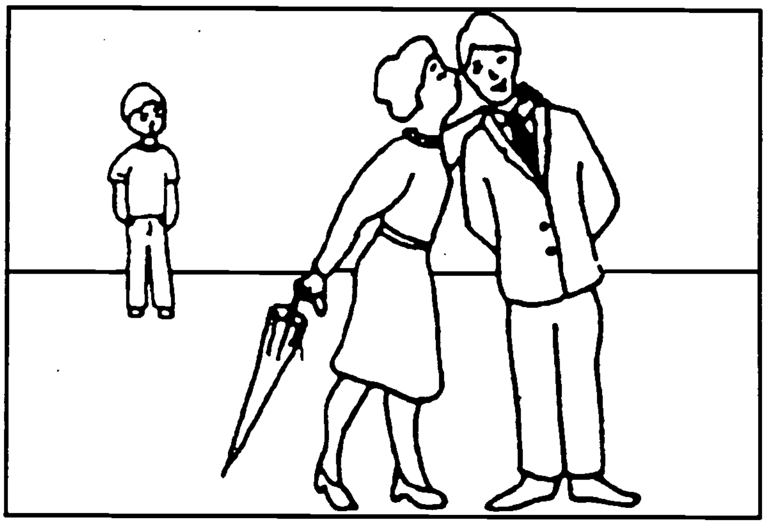
Wernicke Subjects	BDAE Comprehension A (15)	B (12)	C (10)	Speech Production Description of "Cookie theft" picture
M.C.	8	3	3	Majka radi, je bila je radila jedva za hranu, sa jedne strane to su radnici. A deca su uzela da jedu čokoladu. The mother is working, she was, she was working barely to support them, from one point this are workers. And children are eating the chocolate.
A.B.	2	0	0	Ovaj...dete je ustalo da pojede pekmez a ova žena je prosula vodu što je htela da pere pa je sve oprala. Well..the child stood up to eat the jelly and this woman has spilled the water, because she want to wash, she washed everything.
V.V.	6	1	1	Devojčica... ova stolica se valjda slomila. Ukrali su..ne mogu da se setim. Drugarica je donela kolače i sada deca kradu kolače. The girl...this chair, looks like it has broken. They are stealing...I can't say. The woman brought the cookies and the children are stealing cookies.
M.B.	4	0	0	Vidite ovde decu, devojcica, vidite ovaj stolnjak. Stolnjak nije u redu, a majka to ne vidi. U..kako se to zove. You see, here are children, the girl, you see this tablecloth. The tablecloth is not OK., and mother doesn't see that. Well, what's the name for this.
D.D.	5	3	4	Sta je ovo? Neka deca ovde su se popela, hoće da uzmu kolače. Zena pere...šta je ovo...tanjir. Ona stalno gleda kroz prozor. What's this? Some children have climbed here, they wont to take cookies. The woman is washing...what is this...the plate. She is looking through the window.

A - Commands (max. score 15)
B - Complex Ideational Materials (max. score 12)
C - Reading Sentences and Paragraph (max score 10)

APPENDIX B

Sample Picture Array for Sentence-Picture Matching Task





OS: The lady is kissing the man who is holding an umbrella.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 117-138

Tasks and Timing in the Perception of Linguistic Anomaly*

Janet Dean Fodor,† Weijia Ni, Stephen Crain,‡ and Donald Shankweiler‡‡

Three experiments were conducted to investigate the relative timing of syntactic and pragmatic anomaly detection during sentence processing. Experiment 1 was an eye movement study. Experiment 2 employed a dual-task paradigm with compressed speech input, to put the processing routines under time pressure. Experiment 3 used compressed speech input in an anomaly monitoring task. The outcomes of these experiments suggest that there is little or no delay in pragmatic processing relative to syntactic processing in the comprehension of unambiguous sentences. This narrows the possible explanations for any delays that are observed in the use of pragmatic information for ambiguity resolution.

1. INTRODUCTION

Suppose there were experimental results which showed, without any shadow of a doubt, that during ambiguity resolution there is a brief interval in which only grammatical principles (syntax and semantics) are operative, with effects of discourse, plausibility and world knowledge not occurring until later. If this could be demonstrated, would it entail that the language processing mechanism is a module? Or that, within the language system, syntax is autonomous? Some discussions imply that it would.

For example, Mitchell, Corley, and Garnham (1992) presented data showing the absence of a rapid contextual influence in resolving the temporary ambiguity in sentences such as (1), and then concluded (p. 85, our emphases): "The findings show that discourse information is **ignored** at first, even though this information becomes **available** (in the context paragraph) well before the point at which it could usefully have made a contribution to the process of ambiguity resolution."

(1) *The politician told the woman that had been meeting him that he was going to see the minister.*

We agree completely with the logic of this conclusion: To prove modularity of linguistic processing one would show that outside-the-module information (a) is available but (b) is ignored, at least temporarily. But the meaning of the word "available" is crucial here. It has to mean "available to the parsing routines," not just "present in the situation and potentially accessible to

The research reported in this paper was supported by Program Project Grant HD-01994 to Haskins Laboratories from the National Institutes of Child and Human Development, and in part by a grant to Stephen Crain and Donald Shankweiler from the University of Connecticut Research Foundation. We thank Jacques Mehler and Gerry Altmann for helpful information about speech compression algorithms. We are especially grateful to Ignatius Mattingly for his generous expert advice on speech compression and design of the experimental materials, and also for recording all the speech stimuli. Julie Boland, Brian McElree, Janet Nicol and Martin Pickering provided very helpful advice on an earlier draft.

any device with the ability to calculate it." The information has to be available inside the perceiver, at the time in question.

There is no guarantee that this will be so. One of the motivating arguments presented by J. A. Fodor (1983) in favor of a mental module for language was that thinking (the kind of central inferencing that draws on knowledge of the world) is open-ended and is therefore apt to be slow. Syntactic processes (and semantics in the strict sense) are narrower in scope, and therefore can proceed much more quickly—which is why it is good design for them to be unshackled from central processes and allowed to go at their own pace. But notice how the reasoning here turns around on itself. If pragmatic processing is **inherently** slower than syntactic processing, then there is no need to invoke the modularity thesis to explain an experimental finding that pragmatic processing lags behind syntactic processing. To put it more strongly: It would be improper to claim such a time lag as evidence for a modular structure of the mind.

A note on terminology is needed here. For brevity, we will be contrasting what we call syntactic processing and pragmatic processing. But these terms have been used with various meanings, and the choice of interpretation makes a difference to the empirical content of the modularity thesis for language. In what follows we will use the word **syntax** as a shorthand to denote syntax and formal compositional semantics and those aspects of reference, etc. that fall within the grammar. We will use the term **pragmatics** for matters of plausibility and all inferencing based on general knowledge and information provided by the discourse.

The substantive point we wish to make is that, if pragmatic inferences are just slow to be computed, there will be a delay before the use of pragmatic information whether or not the mental architecture is modular. This is what we call **de facto modularity**: Syntax precedes pragmatics as a matter of fact, for whatever reason. Given a finding of de facto modularity, further work is needed to establish its source. In order to argue that the use of pragmatic information in parsing is held up on principled grounds, it would be necessary to show that, whatever delay there might or might not be in computing the relevant pragmatic properties, there is an even greater delay in the use of them in making parsing decisions.

Mitchell et al.'s (1992) experiment did not include a check on whether the potentially disambiguating pragmatic information in the discourse had been processed by subjects in time to be useful, but the point can be illustrated with an experiment by Ferreira and Clifton (1986). They tested sentences as in (2).

(2) a. *The defendant examined by the lawyer turned out to be unreliable.*
b. *The evidence examined by the lawyer turned out to be unreliable.*

Of interest here is the word *examined*, when it follows *defendant* and when it follows *evidence*. Following *defendant* it is sensible (on both structural analyses), and the reading time data showed that it was easy to process. Following *evidence*, *examined* is anomalous (if they are parsed as subject and predicate), and reading times showed that difficulty was high here. This reading-time difference on *examined* shows that the incongruity of the subject-verb analysis for *evidence examined* has already been detected at the word *examined*. With this established, it becomes interesting that the *evidence examined* sentence is still difficult at the *by*-phrase, which syntactically blocks the subject-verb analysis and forces a shift to the reduced relative analysis. At the *by*-phrase, Ferreira and Clifton's data indicated that the *evidence examined* sentence and the *defendant examined* sentence were both significantly harder to process than their unambiguous controls. Thus apparently the anomaly of *evidence examined* is recognized, but this information is not used right away to shift attention to the other analysis. This is the classic syntactic autonomy demonstration.

Later versions of this experiment have had different outcomes (cf. MacDonald, 1994; Trueswell, Tanenhaus, & Garnsey, 1994) and we will not dwell on the data here. It is also true that some methodological problems remain, problems which are difficult to avoid in experiments on modularity. For instance, a positive result in the availability test [at *examined* in (2b)] creates a disturbance of processing; therefore a finding of difficulty at the disambiguation point [the *by*-phrase in (2b)] might be due merely to persistence of this disturbance over the next word or two before it declined. In that case it would not show that the garden-path continued until the *by*-phrase, and so it would not show non-use of available pragmatic information prior to that point.[1] This account of the data could be eliminated by delaying the disambiguation point for several more words, until the upset due to *examined* has demonstrably declined. But by a psycholinguistic Catch-22, a delayed test point might be too late to detect a genuine garden-path if one were present, since all models allow pragmatic information to have an influence at some point. Thus some delicate adjustments of materials or presentation might be needed in order to provide proof that processing difficulty first rises (at anomaly detection), then falls to baseline, and then rises again (at garden path detection). Also, Martin Pickering (personal communication) points out the importance of ensuring that this double sensitivity occurs within trials; it is not sufficient that for some subjects or on some trials the anomaly is detected, and for other subjects or trials the garden path occurs. To satisfy all these practical demands is not easy. Nevertheless, the basic logic of Ferreira and Clifton's (1986) experimental design is surely just right. We believe that this logic should be applied to all purported modularity demonstrations: Unless the availability criterion for pragmatic information is shown to be satisfied at the relevant point in the sentence, that test point may reveal de facto modularity but it cannot provide evidence for a modular architecture for language.

2. AVAILABILITY OF PRAGMATIC INFORMATION: EXPERIMENTAL EVIDENCE

2.1. RATIONALE

Despite years of experimental research (see Tanenhaus & Trueswell, 1994, for a recent review), there is still considerable disagreement about whether the data do in general support a modular or non-modular organization of the language faculty. We take no stand here on which interpretation of the experimental findings is correct; our concern in this paper is solely with the methodological issues involved in testing modularity or autonomy hypotheses. It seemed possible to us that neglect of the availability factor might be responsible for some of the apparent inconsistencies, especially if the time course of pragmatic processing is not constant but varies with the complexity of the inference, the speed of sentence presentation, and so forth. It could be useful, therefore, to establish how rapidly pragmatic implausibilities are computed, independently of the issue of how rapidly they are put to work in resolving syntactic ambiguities.

Eye movement studies like Ferreira and Clifton's certainly seem to show sensitivity to pragmatic problems within as little as 300 to 400 ms. And in ERP (Event-related Brain Potentials) studies the negative shift at approximately 400 ms. in response to pragmatic anomalies suggests much the same time frame. But current evidence is incomplete in several respects, and particularly so with respect to the speed of pragmatic processing **relative** to syntactic processing.[2] Few studies have compared pragmatic and syntactic processing in situations other than ambiguity resolution. Recent exceptions are the work of McElree and Griffith (1995) and Boland (submitted). McElree and Griffith have used an anomaly judgment task in which they force the pace by training subjects to respond promptly at a signaled time following the sentence. This results in a speed-accuracy trade-off (SAT) whose time course can be tracked. Using this method, McElree and Griffith found a delay of pragmatic judgments relative to syntactic judgments; their results "indicated that thematic role violations began to be detected 50-100 ms. later than either constituent structure or subcategorization violations" (p. 152). The speed-accuracy trade-off procedure is an interesting innovation in experimental methods for investigating modularity, and it permits estimates of very small temporal differences. However, for reasons discussed below it is not entirely certain that the observed lag in pragmatic judgment reflects a lag in pragmatic anomaly detection *per se* (see section 2.4 for discussion). Boland (submitted) used a cross-modal naming task with auditory sentence fragments (at normal speed) including examples such as (a) *Which necklace did Nancy describe* and (b) *Which salad did Jenny toss*, followed by visual presentation of a name such as *Bill*. If the name is integrated into the sentence, it creates a syntactic anomaly with (a) and a pragmatic anomaly with (b). The data for these sentence versions appear to show a stronger and earlier influence on naming RT of the syntactic anomaly; the pragmatic anomaly version shows a significant influence on naming only for targets presented 300 ms. later than the offset of the verb. (Note: the auditory input stopped at the end of the verb.) However, Boland's own interpretation of her data, including a comparison of naming and lexical decision on the visual stimulus, is that the naming task is not sensitive at all to pragmatic anomaly for reasons other than its temporal properties.

In the research reported here, our strategy for detecting any small timing differences that may exist in the availability of information during sentence comprehension was to put the processing mechanism under severe pressure by using compressed speech input. This approach was proposed some years ago by Chodorow (1979), but his suggestion has not previously been followed up. Chodorow's rationale for using compressed speech was as follows (p. 88): "If component processes in comprehension have different performance characteristics (e.g., resource requirements or maximal rates) then increasing the overall rate of input ought to affect them differentially. In this way, we might reasonably expect to be able to pull apart otherwise entangled components." Chodorow compared lexical and syntactic processing, not syntactic and pragmatic processing. For syntactic processing his experiments showed that when speech input is speeded up to twice the normal rate, the processor cannot keep up but builds up a backlog of processing to be done after the sentence is over.[3] His results showed that sentence recall declined for compressed input relative to normal input but, at least for unambiguous sentences, the provision of extra time between the compressed input and the recall task allowed performance to return to normal levels. The extent of the backlog was bracketed between 200 ms. (since this provided insufficient time for recovery) and 750 ms. (which was shown to be sufficient). Based on this finding for syntax, we speculated that forcing the pace of processing by using compressed speech would exaggerate any difference in the timing of syntactic and pragmatic processes, so that if such a difference exists, it would be measurable however slight it normally is. Having set up this situation where a pragmatic delay is most likely to occur, and shown that our methods could detect it, it would then be possible to vary the task demands in various ways to see under what circumstances the delay would diminish or disappear. At issue would be whether it fluctuates widely enough to account for some of the apparently contrary outcomes of ambiguity resolution experiments. In fact, we have not completed this broad program of investigation, and for reasons that will become clear below, it has somewhat changed its character as a result of the data reported here.

The experimental sentences were the same for all experiments described below, except for minor differences of constituent length as necessitated by different tasks. They were not ambiguous, either globally or temporarily. They contained syntactic anomalies or pragmatic anomalies of a kind similar to those commonly used in experiments on ambiguity resolution to disqualify one of the potential analyses, as illustrated in (3).

(3) *It seems that the cats from across the road...*

SYNTACTIC ANOMALY *won't eating*

PRAGMATIC ANOMALY *won't bake*

BASELINE (NO ANOMALY) *won't eat*

...the food that Mary puts out on the porch every morning as soon as she gets up.

The syntactic anomaly always involved a modal verb followed by an *-ing* form. The pragmatic anomalies involved unsuitable pairings of agents and actions, such as *cats - bake*, compared with the acceptable *cats - eat*. The unsuitability was often but not always due to a mis-match of animacy. The anomalously paired noun and verb were also not just low associates, compared with high associates in the acceptable baseline sentence. For example, the anomalous pairing *kangaroo-swear* contrasted with the baseline pairing *kangaroo-sit*, though the latter are not high associates; also, the anomalous *songs-learn* contrasted with the baseline *songs-tend*, which is not a high associate pair. The anomaly did not depend on words following the verb: for example, cats baking something is anomalous regardless of what is being baked. No pre-test was run to validate that the anomalous versions were indeed perceived by subjects as anomalous; the results of the main experiments sufficiently confirmed that they were. Thirty matched triples of experimental sentences like (3) were used in all three experimental paradigms reported here. The verbs were matched for mean frequency across sentence versions (pragmatic anomaly version 89.46; baseline version 79.96, Francis & Kučera, 1982). In each experiment, 30 experimental sentences were presented to a subject, 10 of each of the three versions (syntactic anomaly, pragmatic anomaly, baseline). They were dispersed semi-randomly among 76 filler sentences, with at least one filler between neighboring experimental sentences. The filler sentences were of various syntactic types; some were similar and most were dissimilar to the experimental sentences; 58 were acceptable and 18 were anomalous in various ways (additional or missing words, noun/verb substitution, etc.). Half the subjects received the sentences in one order, and the other subjects received them with the first and second halves of the list interchanged. (Order of presentation was not a significant factor in any of the outcomes and will not be discussed further.) A practice session preceded each experiment.

Experiment 1 was an eye movement study designed to establish that our materials gave results similar in temporal profile to those standardly reported in the literature. Experiment 2 used compressed speech input to put the processing routines under time pressure; the time of detection of the anomaly was established by measuring its effect on a concurrent lexical decision task. Experiment 3 used the same compressed speech input but required subjects to monitor for anomalies, as in several recent studies with visually presented materials. (See for example, Boland, Tanenhaus, and Garnsey (1990) and Boland, Tanenhaus, Garnsey, and Carlson (in press) for discussion of the "stop-making-sense" task.) We would emphasize that none of these experiments was designed to resolve the modularity issue for language processing. Their aim was merely to clear away one possible source of uncertainty so that modularity questions can be asked and answered more clearly.

2.2. EXPERIMENT 1: EYE MOVEMENTS

Due to space limitations we present Experiment 1 in outline only. Further details and all statistical analyses are reported in Ni, Fodor, Crain and Shankweiler (in prep.). Sentences as in (3), adjusted to a maximum length of 76 characters, were presented visually on a single line and eye movements were recorded. Subjects were 24 college students who were paid for their participation. For purposes of data analysis the sentences were divided into regions as shown in (4).

(4) *It seems that / the cats / won't usually / VERB the / food we / put on the porch.*
1 2 3 4 5 6

Region 1 was the beginning of the sentence prior to region 2; it was 0 to 3 words long. Region 2 consisted of the last two words of the subject NP of the main verb; this was always the head noun plus a preceding word (determiner or adjective). Region 3 was the modal verb plus a following adverb (inserted for purposes of this experiment to make it possible to measure regressions from the main verb to the modal). Region 4 was the main verb plus the next word, regardless of category. Region 5 was the next two words, regardless of category. Region 6 was the remainder of the sentence, 0 - 4 words.

First pass residual reading times[4] and percent of regressions (i.e., the percentage of all first pass fixations which resulted in regression to a prior region) are shown in Figures 1 and 2.

At the verb (region 4), there was no effect of either the syntactic or the pragmatic anomaly on reading times. However, there was an increase in regressions compared with baseline in this region for both anomaly types, though only approaching significance for the pragmatic anomaly. The difference between the syntactic and pragmatic effects is not significant.

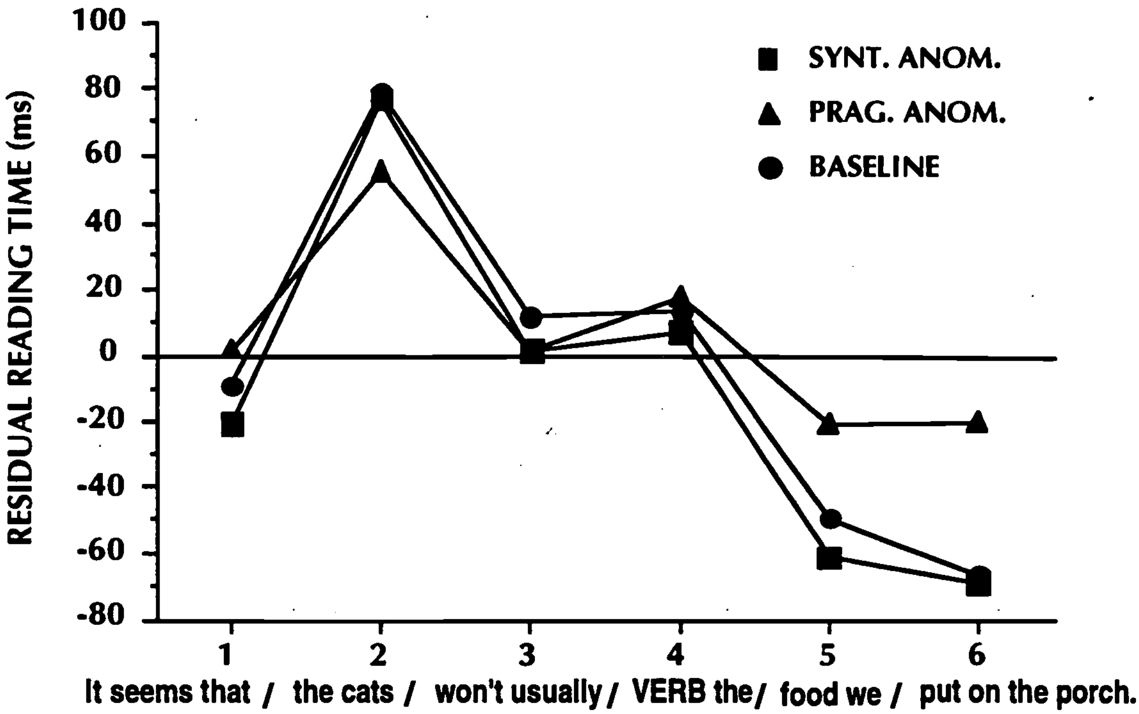


Figure 1. **Experiment 1: Mean first pass residual reading times**

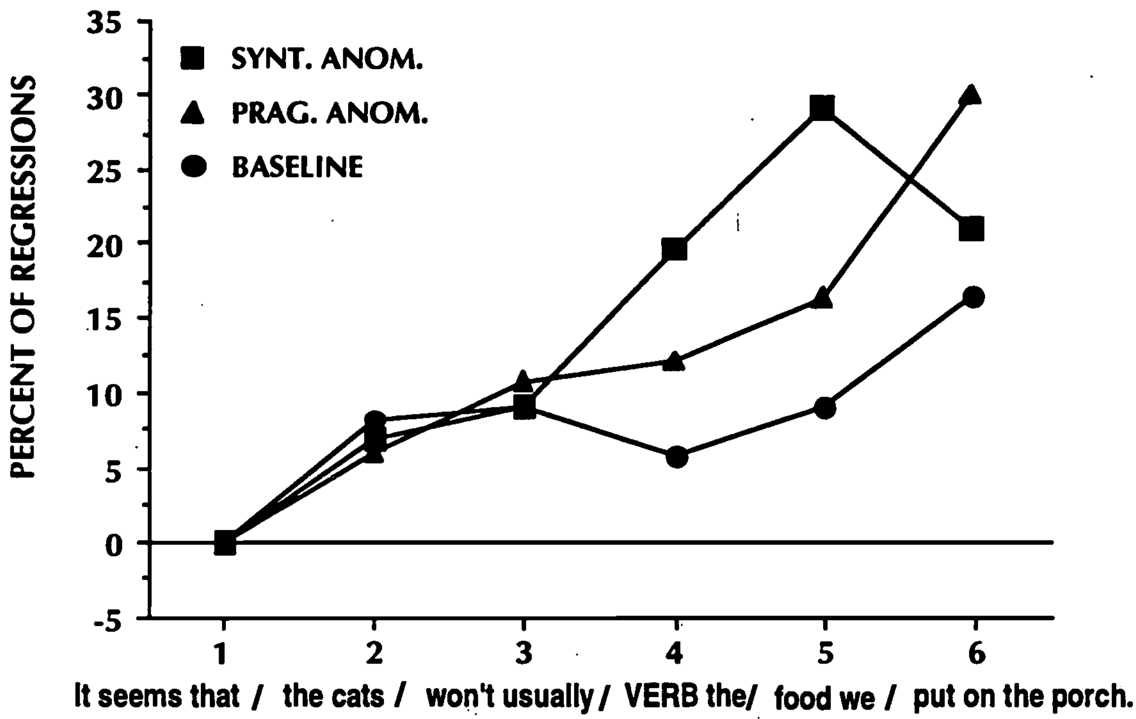


Figure 2. **Experiment 1: Percent of regressive eye movements**

Thus, eye movement responsiveness to both types of anomaly was essentially immediate and simultaneous, within the limits of measurement imposed by the paradigm; that is, at least within the time taken to read region 4 (mean 401.55 ms). To obtain finer information about timing it would be necessary to divide the sentence into even smaller regions, but since readers don't normally fixate on every word, results would become erratic and would be complicated by uncertainty about the extent of parafoveal effects.

For the remainder of the sentence (regions 5 and 6), responsiveness to the two types of anomaly is differently distributed between reading time and regressions. For the syntactic anomaly, reading times do not depart significantly from baseline at any point. For the pragmatic anomaly, by contrast, reading times rise relative to baseline at region 5 and remain high at region 6. The difference falls just short of significance for each region separately, but for both regions combined it is significant. For syntactic processing, regressions increased, relative to baseline, at region 5, but dropped to baseline level by region 6. For pragmatic processing, by contrast, regressions continue to increase relative to baseline throughout the sentence.

In summary: The syntactic anomaly has little effect on first pass fixation durations; it causes an immediate increase in regressions, but this is short-term. The effects of the pragmatic anomaly are divided between reading times and regressions, and both effects become progressively stronger as the sentence continues. It seems reasonable to attribute these qualitatively different patterns to different strategic responses by the parsing mechanism to on-line problems. (We don't mean to imply by this that eye movements are controlled by deliberate decisions on the part of the reader, but merely that some part of the mental mechanism involved in eye movement planning and control is responsive to the outcomes of higher level linguistic processing.) The syntactic anomaly is easily detectable. It triggers immediate regression to check the source of the problem, and then, since the error does little damage to sentence comprehension, there can be a quick return to normal processing. For the pragmatic anomaly, we suppose that the parser has some uncertainty about the source of the problem, or some hope that it might be resolved by later words in the sentence. So though it did check back to confirm that all was not well, its main strategy was to keep pressing forward, but slowly, hoping that matters would eventually resolve themselves. At the end of the sentence, when it became clear that no resolution was forthcoming, regressions took an upward turn. Thus in general, the results of Experiment 1 show rapid sensitivity to both types of anomaly in our experimental materials, with qualitatively different profiles for syntax and pragmatics that are, on reasonable assumptions, in keeping with previous findings in the literature on eye movement patterns in response to anomalous sentences (for example, Ni, Crain, & Shankweiler, in press; Pearlmutter, Garnsey, & Bock, 1995).

Eye movements provide generous quantities of on-line information and show very rapid response to sentence properties of interest. However, they are not ideal for the purpose of establishing relative timing relations between syntactic and pragmatic processes, for several reasons. As noted above, it is not easy to narrow down the intervals within which events occur. The division of effects between forward reading and regressions differs across anomaly types and may be susceptible to properties of the materials that are not of central interest. Because the reading is self-paced, the comprehension system can take as much time as it needs to complete all levels of processing, so minor differences in timing could go overlooked. There is some hint of a pragmatic processing lag in Experiment 1, but it is far from clear. To sharpen up the evidence we turn to compressed speech input.

2.3. EXPERIMENT 2: CROSS-MODAL LEXICAL DECISION, COMPRESSED SPEECH

A cross-modal dual-task paradigm (cf. Shapiro, Zurif, & Grimshaw, 1987) was used to establish the time course of anomaly detection under conditions of rapid processing. The sentences were produced by a male speaker at a normal rate (average 330 ms. per word) and recorded using DigiDesign, Inc.'s Sound Designer II, an audio editing application for the Macintosh computer. Sound Designer II was then used to produce versions of these utterances compressed to approximately half of their original duration (54 % on average, range 53% to 55%, the variability due to the fact that the program safeguards quality of the signal by refraining from compression at certain points). We checked the output of the program by making spectrograms of the stimuli both before and after compression, and observed no untoward changes. The timing was uniformly scaled, and the pitch contour and formant frequency contours were essentially unchanged. On the intelligibility of compressed speech, Altmann and Young (1993) report: "We have found that there is virtually no loss of intelligibility, or subjective 'quality', when sentences are compressed to, for instance, 50% of their original duration." Gerry Altmann (p.c.) informs us that at 50% compression, intelligibility (after a brief period of adaptation) is around 85-90% for plausible sentences, where intelligibility is established by the percentage of words correctly reported by subjects after hearing the sentence. (The sentences were 8.5 words long on average, shorter than ours. The compression program differed from the one we used but, as far as we know, not in any consequential way.)

Sixty college students were paid for their participation in the experiment. All reported normal hearing. The compressed speech was played to subjects through headphones. The experimental sentences were as in (3). A minimum of 10 words followed the anomaly in all cases, to ensure that the concurrent task (lexical decision) did not overlap with sentence-final wrap-up effects. The first 14 sentences after the practice session were fillers, to provide subjects with the chance to adapt to the rapid speech. Data from three other subjects were excluded from the analysis due to poor performance on the comprehension task (described below).

A visual lexical decision target was presented at five different time points, both before and after the critical verb (*eat, eating, bake* in (3) above) was heard. Previous work (Ni, Fodor, Crain, Shankweiler, & Mattingly, 1993) had shown the importance of tracing the rise and fall of sensitivity to the anomaly. A single test point is incapable of revealing timing differences between anomaly types if it happens that there is overlap between their time-envelopes. The target word appeared at -81 ms., 0 ms., 81 ms., 162 ms., or 243 ms. relative to the offset of the verb. (These test points were 150 ms. apart before speech compression.) For filler sentences the target words appeared at a wide variety of sentence positions, randomly determined. The lexical decision targets for experimental sentences were all low-frequency words (mean 7.82 per million, range 0 to 36; Francis and Kučera, 1982) of medium length (mean 6.33 letters, range 3 to 10). Targets for the 76 filler sentences were 23 words, and 53 non-words created by changing one letter in a word. Each target word appeared centrally on a computer screen until the subject responded or after 700 ms. All target words were unrelated in meaning to the sentences during which they appeared. All 3 versions of an experimental sentence were associated with the same target word; each subject heard only one version. There was a complete rotation of sentence versions and test points: for each of the 5 test points, a subject heard 6 experimental sentences (2 syntactically anomalous, 2 pragmatically anomalous, 2 baseline); thus each subject was presented with 2 tokens of each of the 15 presentation conditions. There was also an end-of-sentence task to ensure that subjects would take the trouble to listen to the sentences under these very demanding conditions; we will discuss the nature of this task below.

On the basis of Chodorow's (1979) findings, and our own impressions from listening to the compressed materials, we expected that the computation of content would fall behind the processing of structure, so that the lexical decision task would show interference from the syntactic anomaly at an earlier point than it would show interference from the pragmatic anomaly. As will become clear, this prediction was not confirmed.

The results are given in Figure 3, where lexical decision time is shown for the anomalous sentence versions relative to lexical decision time for the baseline version.

Note that Figure 3 and all statistics associated with it derive from an analysis not of absolute RTs but of z-scores (= RTs expressed in terms of standard deviations from each subject's mean RT across all items). This is because the data were pooled from subjects divided into two groups whose mean RT differed as discussed in detail below. (Mean RTs in ms. for all three sentence versions are given by group in footnote 5.) In all analyses, only RTs for correct lexical decision responses were included, and lexical decision RTs were trimmed by changing those that were more than 2 standard deviations above or below the subject's mean RT to exactly 2 standard deviations above or below the mean respectively.

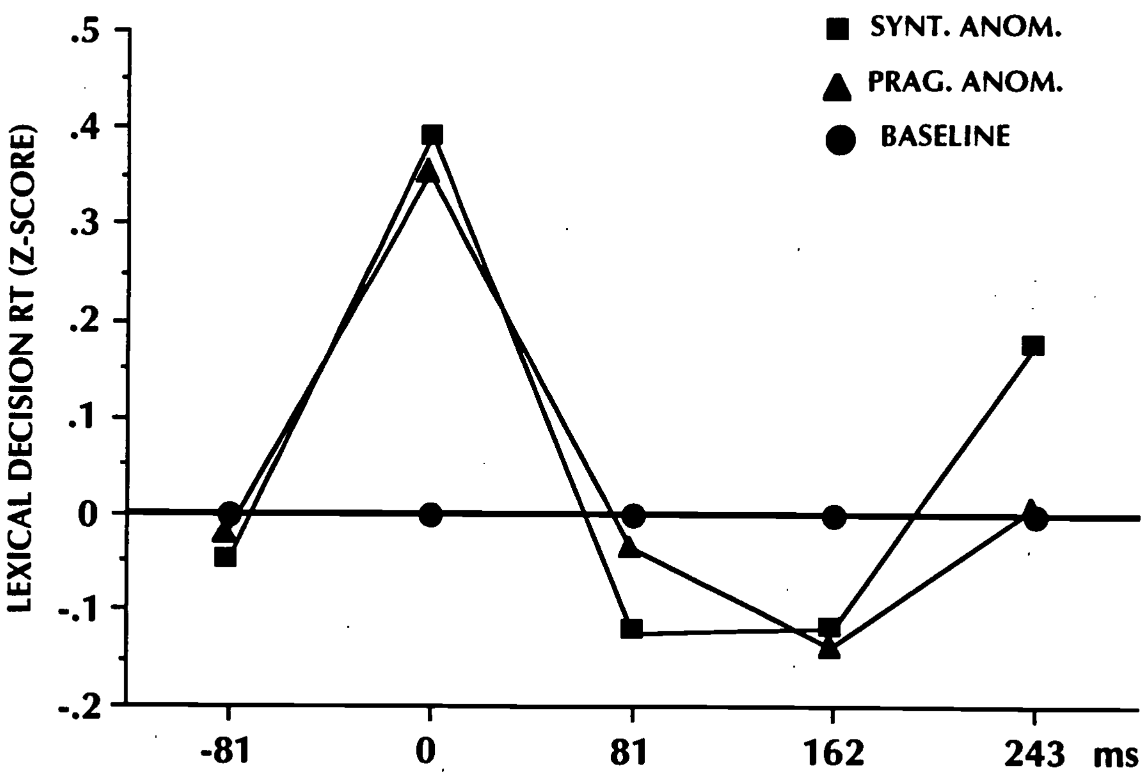


Figure 3. **Experiment 2: Lexical decision RT (z-score) for anomalous sentence versions relative to baseline (all subjects)**

Analysis of variance showed that the only position at which the anomalous sentences differed from the baseline was at the 0-ms test point. For the syntactic anomaly the mean difference in RT at this point was 56 ms. ($F_1(1,59) = 11.46, p = .0013$; $F_2(1,29) = 5.54, p = .0256$). For the pragmatic anomaly the mean difference in RT at this point was 52 ms. ($F_1(1,59) = 8.07, p = .0062$; $F_2(1,29) = 7.610, p = .0100$). There was no difference between the two anomaly versions at any other test point. There was an apparent rise in RT for the syntactic anomaly version at the 243 ms. test point but it was not significant relative to either the baseline ($p > .1$) or the pragmatic anomaly ($p > .1$). Thus the result seems very clear. The peaks in lexical decision RT indicate the increase in sentence processing load due to the anomaly. And the peaks for syntactic and pragmatic processing lie on top of each other. There is no sign here of any delay in pragmatic processing relative to syntactic processing, at least within the grain provided by the 81 ms. intervals between the test points.

The data in Figure 3 are from all 60 subjects. These numbers combine results from subgroups of subjects who performed different post-sentential comprehension tasks. Group A did an oral paraphrase task: after 20% of the filler sentences a bell sounded and the word "PARAPHRASE" appeared on the screen; the subject then had to speak into a microphone, giving the meaning of the preceding sentence in his or her own words. For Group B subjects, following the same sentences, a bell sounded and a simple comprehension question appeared on the screen; the subject answered it with the same "yes" and "no" keys as for the lexical decision task. Performance on these ancillary tasks was recorded but not analyzed except for purposes of screening out inattentive subjects. However, when the lexical decision RT data are analyzed for the two groups separately, it appears that the comprehension tasks made an interesting (unanticipated) difference to lexical decision performance. It should be noted that this difference between groups is not statistically significant, but it exhibits a trend which is of sufficient theoretical interest to be worth discussing, even though conclusions must necessarily be tentative. Figures 4 and 5 show the results for the two groups separately.[5]

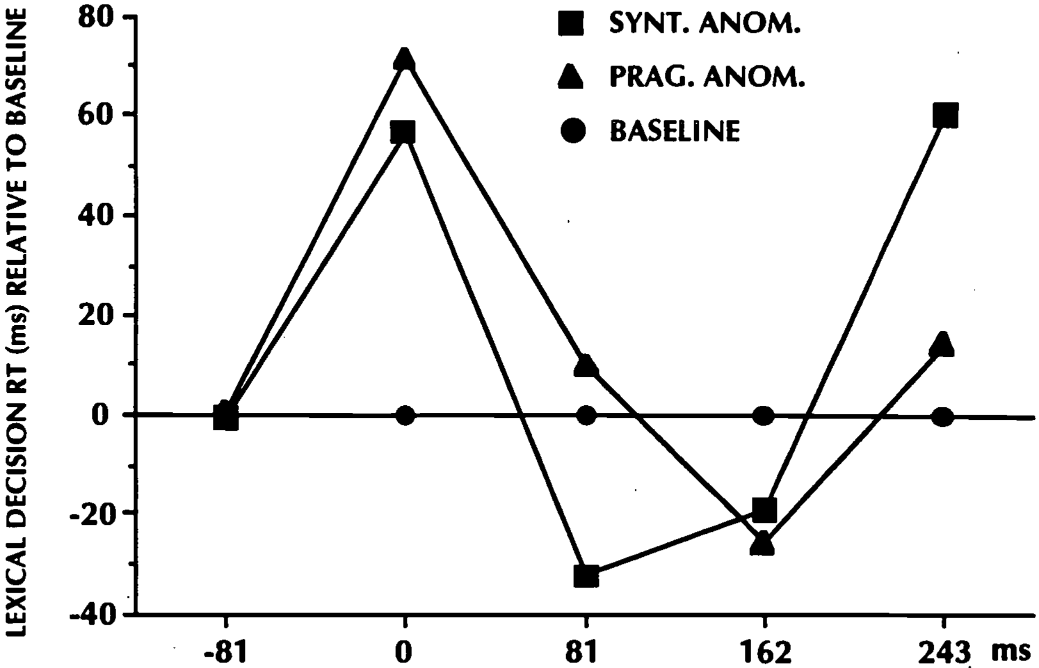


Figure 4. Experiment 2: Lexical decision RT (ms) for anomalous sentence versions relative to baseline: Group A (paraphrase task)

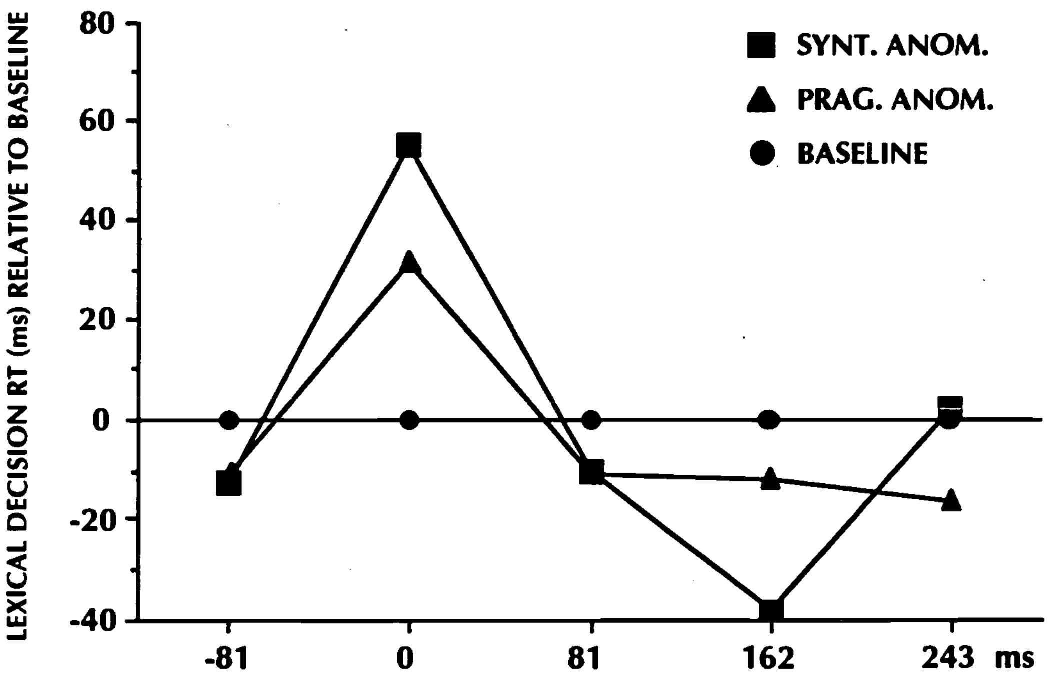


Figure 5. Experiment 2: Lexical decision RT (ms) for anomalous sentence versions relative to baseline: Group B (comprehension question)

The timing of the peaks for syntax and pragmatics is still identical, in each group. But the height of the peaks differs across groups, indicating a difference in the degree of sensitivity to the two types of anomaly. For the paraphrase subjects (Group A), sensitivity to the pragmatic anomaly was greater than sensitivity to the syntactic anomaly (Figure 4). For the comprehension question subjects (Group B), the opposite was the case (Figure 5). (Neither of these differences was significant; $p > .1$ in both cases.) In fact for Group B the pragmatic peak was not reliably a peak at all; the difference from the baseline dropped to 32 ms. and was not significant ($p > .1$). The other three peaks remain significantly different from baseline. (For Group A, syntax vs. baseline, $F_1(1, 29) = 4.61, p = .0402$; $F_2(1,29) = 4.64, p = .0397$; pragmatics vs. baseline, $F_1(1, 29) = 4.83, p = .0361$; $F_2(1,29) = 6.16, p = .0191$. For Group B, syntax vs. baseline, $F_1(1, 29) = 4.82, p = .0363$; $F_2(1,29) = 3.88, p =.0584$.).

What is of interest is the difference across groups in the sensitivity to the pragmatic anomaly. For pragmatics, the mean RT difference from baseline for Group A = 71 ms, but for Group B = 32 ms. The difference between these (i.e., the interaction between subject groups and sentence type, for pragmatic anomaly vs. baseline) approached significance. This contrasted with sensitivity to the syntactic anomaly, which is almost identical across groups. For syntax, the mean RT difference from baseline for Group A = 57 ms., for Group B = 55 ms. Plausibly, this apparent difference in pragmatic sensitivity between the subject groups is no accident. The paraphrase task was more demanding than the question answering task. It required more careful attention to the content of the sentence. This is reflected in the fact that for the paraphrase group the overall mean RT for lexical decision was 61 ms. higher than for the yes/no question group. (Mean RT for Group A = 792 ms., for Group B = 731 ms.; $F_2(1, 29) = 52.30$; $p =.0001$.) Sensitivity to the pragmatic anomaly rose and fell as the task demands did. The syntactic effect, by contrast, stayed constant in magnitude across tasks. This suggests that when processing resources are limited, as in the compressed speech situation, the processor concentrates efforts on syntax. It seems that syntactic processing is mandatory, but pragmatic processing is not; it may be sacrificed when time is short. This is one interesting sense in which it seems that syntax does take priority over pragmatic processing, and it is worth noting that mandatoriness is another of the properties that J. A. Fodor (1983) proposed as indicative of a module at work. However, this is the **only** evidence we found in this experiment for the priority of syntax over pragmatics. The temporal delay of pragmatics that might have been expected with speeded speech was not evident at all.

To sum up: It can be concluded that differences in time of availability of syntactic and pragmatic information are **not** the source of timing differences in their use for ambiguity resolution, at least for materials such as these where detecting the pragmatic anomaly requires no complex inference. Of course, the data reported here don't exclude the possibility that pragmatics runs some milliseconds behind syntax, as might be expected on any model in which the pragmatic analysis is fed by the syntax. (Pragmatic analysis not fed by the syntax occurs in "strong interaction" models; see McClelland, St. John & Taraban, 1989, and Bates & MacWhinney, 1989.) Our experimental results do not discriminate between a very short delay and none at all. But they do appear to exclude a delay on the scale of those standardly cited in favor of syntactic autonomy in ambiguity resolution. Such effects are not always precisely timed but seem to be on the order of several hundred ms. The classic paper by Rayner, Carlson, & Frazier (1983; p. 371) concludes "probabilistic semantic and pragmatic information does not influence the processor's initial choice of a syntactic analysis...[but] semantic and pragmatic plausibility information does influence the ultimately preferred analysis of a sentence;" however, the data as presented do not permit computation of specific temporal relations between the initial and the ultimate analyses. Mitchell et al. (1992) looked for an influence of discourse on ambiguity resolution at two points in their material: at the beginning and end of the ambiguously attached *that*-clause (see example (1) above). As noted in Section 1, they found no discourse influence at the earlier disambiguation point immediately after *that*, but they did find a marginally significant effect of discourse at the later point (e.g., at the second *that*-clause in *The politician told the woman that he had been meeting that he was going to see the minister*, which disambiguates the first *that*-clause as a relative. The data as presented did not afford an exact calculation of when this second disambiguation occurred, but it appeared to have been at least 1200 ms. later than the early test point. It was not possible to tell whether discourse information had been used for ambiguity resolution any earlier than that. A recent study by Urbach, Pickering, Branigan, and Myler (1995) was designed to provide more precise timing information. Urbach et al. tested syntactic and pragmatic cues for disambiguation of the familiar main clause / reduced relative clause ambiguity (e.g., *The cook helped / helping was busy, The teacher / language taught was Spanish*). ERP patterns indicated that both cues were helpful in staving off a garden-path at the word *was* when sentences were presented visually one word at a time at 550 ms. per word, but only the syntactic anomaly averted the garden-path when the presentation rate was 400 ms. per word. This could suggest a pragmatic lag in ambiguity resolution of somewhere between zero and 550 ms. However, some caution is necessary here since no data are presented for the "anomalous" word (e.g., *taught* following *language*), and without this it is not possible to estimate whether the pragmatic anomaly had been detected prior to *was* at the faster presentation rate, i.e., whether these materials meet the availability criterion. We return to the relationship between ambiguity resolution and unambiguous processing in our general discussion of the findings.

With appropriate provisos, we conclude from Experiment 2 that even at twice-normal input rate, pragmatic computations can keep pace with syntactic computations. Perhaps we should have guessed that this might be so, on the basis of results reported by Young, Altmann, Cutler, and Norris (1993). They were looking for prosodic effects on intelligibility of compressed speech, but they concluded: "Overall, the only consistent predictor of intelligibility across the two experiments was plausibility. The more plausible a sentence, the better recognized are the component words." Obviously, plausibility could not have influenced word recognition unless plausibility were being calculated rapidly. However, the plausibility effect was not necessarily occurring in the word recognition stage; it could have been due instead to response bias, since the task was to write down the sentence after hearing it. Young et al. note: "...the present results do not allow us to decide whether this is because a word can be better predicted, and hence better recognized, on the basis of the preceding words as the sentence is heard, or because during the subsequent transcription of the sentence it is easier to reconstruct words which had not been recognized originally. The latter explanation, invoking reporting biases based on listeners' experience, is certainly consistent with the finding that the effect of plausibility remains constant across compression rates." It also, of course, does not imply rapid pragmatic processing. However, the results of Experiment 2 make it more likely that Young et al.'s plausibility effect was not (just) a response bias.

The outcome of Experiment 2 does not conform to our expectation, based on Chodorow's (1979) findings, that syntactic and pragmatic processing would be "pulled apart" by time compression of the input. But here too, hindsight offers explanations. Our paradigm differed substantially from Chodorow's. Chodorow estimated processing load post-sententially. With compressed speech it seems especially likely that the processor would want to re-survey the whole sentence during "wrap-up" operations; this could create a post-sentential lag at some level of processing, even if none occurred on-line. Thus there is no contradiction between Chodorow's results and our own. Note also that Chodorow compared the extent of the post-sentential spill-over of processing when speech was compressed with when speech was at normal speed, and found that (for difficult sentences) the former was greater than the latter. Our data do not speak to this. They show that pragmatics was not **differentially** slowed by the speech compression, but they do not exclude the possibility that all levels of processing were retarded (by an equal amount).

It is important that other varieties of pragmatic anomaly be tested to establish whether it is generally the case that pragmatic processing is as rapid as syntactic processing, regardless of how intricate the pragmatic reasoning involved. The anomalies we tested resemble the sentence-internal subject-predicate incompatibilities investigated in many studies, including Ferreira and Clifton's (1986) experiment described above. But other mis-matches, such as between a sentence and the prior discourse, might be established more slowly. What complicates matters is that it is often unclear what initiates the relevant computations. For instance, in example (1) above from

Mitchell et al.'s (1992) experiment, the singular noun *woman*, if it had no modifying phrase, would be referentially anomalous following a discourse which established two women as potential referents. As Mitchell et al. noted, the existence of two women was established in the context several words before the noun *woman* in the test sentence. However, there is no guarantee that the processor would make use of this time to deduce that if an unmodified singular noun *woman* were to occur it would be infelicitous. Perhaps such an inference is drawn only when it becomes relevant to some later decision. In that case the computation would not start until the word *woman* was received. Even this presupposes an active processor that anticipates what may follow the current input. (To what extent the human parser is anticipatory has never been very clearly established, but see Gorrell, 1989, for discussion.) A merely passive parser would process structure and meaning in step with the input and would not react to problems until they have arisen. In that case, the discourse information about two women would not initiate relevant inferencing until the following word, *that*, was received—but this is the very word whose attachment needs to be disambiguated.

The only safe course would be to test all materials designed to be used for experiments on modularity. The question then arises of what experimental paradigms are suitable for this purpose. Working with compressed speech cannot be commended on grounds of convenience. It is considerably more demanding in terms of both labor and technical resources than the use of normal spoken or visual stimuli. Moreover, a dual-task paradigm (at least, if the secondary task provides just a single RT, not a continuous response measure) is not economical as a way of monitoring changes in processing load over an extended stretch of a sentence. It is of practical importance, therefore, to establish whether an adequate test of information availability could be based instead on a direct response to a sentential event such as an anomaly, and at normal presentation rates. Arguably, the latter is insufficient because of the importance of catching small timing differences in availability that could have a powerful effect in a winner-takes-all system for ambiguity resolution (see discussion in Section 3). Unless more precise time-sensitive response measures are used than in most current experiments, there may be no way to achieve this other than by forcing the pace of the processing routines, as in Experiment 2 or in the manner of McElree and Griffith (1995), so that small timing differences cannot be absorbed and become measurable. It is possible, though, that the measurement could be made by means of other on-line tasks than the cross-modal lexical decision task of Experiment 2. To check this, we investigated the value of a simple anomaly monitoring task. (See also Osterhout, Nicol, McKinnon, Ni, Fodor, & Crain, 1993, for an ERP study of sentences like those in the experiments reported here.)

2.4. EXPERIMENT 3: ANOMALY MONITORING, COMPRESSED SPEECH

The same compressed speech materials as in Experiment 2 were presented to 9 subjects in a monitoring task. They were instructed to "push the NO button as fast as you can, if you hear a mistake." A practice session gave examples of "mistakes," both syntactic and pragmatic. All sentences continued to completion whether the button was pressed or not. Table 1 shows RT and percentage of sentences judged anomalous for each sentence version.

Table 1. *Experiment 3: Sentences judged anomalous.*

	Percent	RT (ms)
SYNT. ANOM.	83.3	1011
PRAG. ANOM.	75.3	1535
BASELINE	9.9	2186

The pragmatic anomalies were detected almost as often as the syntactic ones; the difference in accuracy was not significant ($p > .1$). We may conclude that subjects' judgments were roughly equally secure for both kinds of example. However, subjects' reaction to the pragmatic anomalies was very much slower than to the syntactic anomalies; the difference of 524 ms. is highly significant ($F_1(1, 8) = 13.33, p = .0001$; $F_2(1, 29) = 12.62, p = .0001$). However, in view of the striking lack of difference that we found in Experiment 2, where an explicit judgment on the sentences was not called for, we believe that this delay of response to the pragmatic anomaly when the subjects' task was monitoring for anomalies must be attributed to judgment processes rather than to architectural constraints on information flow between processing components. Indeed, the monitoring task would be expected to show a similar sort of strategic effect as we saw for eye movements in Experiment 1, perhaps in heightened form because the button push in Experiment 3 depends on a more deliberate decision than the control of eye movements. The sentence processing routine, not in doubt about when a syntactic error has ruined a sentence, would respond promptly in the judgment experiment. But in the case of a pragmatic anomaly, the processor might anticipate that the apparent problem would fade away, or that it would be rescued somehow or other by later parts of the sentence; what at first seemed absurd might turn out to make sense when the message was eventually grasped as a whole. So the processor would wait for further words, to see if things improved. If so, then the button-pressing *response* to the pragmatic anomaly would be delayed even if there were no delay in pragmatic processing at all.

Other research shows that when the processor is not allowed to wait for more words before making a pragmatic anomaly decision, processing is slowed overall. Tanenhaus, Garnsey and Boland (1990) note that in the "stop making sense" paradigm, self-paced reading time increased by about 200 ms. per word (for "makes sense" judgments) relative to normal self-paced reading. Response times for negative ("stop making sense") responses are not reported, but they are presumably at least as slow. Therefore, this paradigm, like the anomaly monitoring paradigm of Experiment 3, is not optimal for purposes of establishing precise timing of events in normal sentence processing. As noted in Section 2.1, an explicit judgment methodology for studying anomaly detection has also been employed by McElree and Griffith (1995), who observed a significant delay in pragmatic processing relative to syntactic processing. They note (p. 152): "For the active conditions, the average estimate of when thematic role processing began was 279 ms. as compared with 233 ms. for syntactic processing. In the passive conditions, the estimate was 289 ms. as compared with 172 ms." McElree and Griffith's experiment differs in several ways from both Experiment 2 and Experiment 3. The sentences were only 4 - 6 words long, and the anomalous word was always the final one. Therefore subjects would know that the anomaly could not be reprieved by anything later in the sentence. Also, the judgments would be made in the region of "wrap-up" processing, which may differ from on-line processing. Sentences were presented visually at 200 ms. per word (somewhat slower than our compressed speech materials). The test points (at which a bell signaled subjects to respond, whether ready or not) were at 14, 157, 300, 557, 800, 1500, and 3000 ms. after the onset of the final word. Thus, the early test points were 143 ms. apart, and this is where sensitivity to the anomalies was estimated to have begun. The estimates of when, within one such interval, the different types of anomaly first became discriminable were made on the basis of curve fitting that included the later test points as well.

Since the goals of this project and our own were similar, it is of interest to consider their findings together, although comparison across different paradigms could only be suggestive. McElree & Griffith's thematic role violations are similar to our pragmatic anomalies (e.g., *Some people alarm books*), and their category and subcategorization violations are not unlike our syntactic anomalies (e.g., *Some people rarely books, Some people agree books*). Tentatively, therefore, the small but significant estimated difference in onset of sensitivity to the thematic versus (sub)category anomalies in McElree and Griffith's experiment may be compared with the lack of an observed pragmatic/syntactic difference in our Experiment 2 and the large difference in our Experiment 3. One way to reconcile these findings is to assume, as suggested by Brian McElree (p.c.), that Experiment 2 would also have shown a pragmatic lag if the intervals between test points had been shorter. As noted earlier, this is certainly possible. It is clear that in future experiments, the test points should be even closer together; we know now that if there is a lag it is very brief. However, we believe that it is at least as likely that the pragmatic lag that McElree & Griffith report was induced by the anomaly judgment task, under the stress of the speed-accuracy trade-off test procedure; that is, that it represents a delay in the **judgment**, but not necessarily a delay in **detection** of the anomaly by the processing routines.

The reason for using anomalous sentences for purposes of comparing syntactic and pragmatic processing is that anomaly is a property that both levels of processing can exhibit regardless of differences in other respects. But even so, there are possible disparities between the two types of anomaly that must be guarded against. Syntactic errors are often more sharply defined than pragmatic ones, and may be judged more confidently. This is a difference that is not easy to eliminate, so it is important to protect experimental designs against its influence, as far as is possible. However, requiring subjects to make an overt anomaly judgment response may magnify this difference between anomaly types. The subject is faced with what amounts to a double judgment: (i), Do I, the subject, find this sentence odd? (ii), Is its oddity such that the experimenter classifies it as a mistake? One judgment is about the sentence; the other is about the likelihood of inter-speaker agreement about the sentence. For the materials in our experiments and in McElree and Griffith's (1995), subjects are probably more certain that the experimenter agrees with them on syntactic judgments than on pragmatic judgments. We have no data on this matter, but it would not be difficult to test experimentally. If it is correct, it would explain why subjects might need a little more time before overtly classifying a sentence as containing an error if the error is pragmatic than if it is syntactic. The fact that the final judgment is the same in both cases (as the identical asymptote implies) does not mean that there was not more doubt along the way in one case than in the other.

The issues are complex, but one outcome that is clear is that the subjects' task makes a difference. Experiments 2 and 3 gave very different results for identical sentence materials. It seems clear that some experimental paradigms are less suitable than others for establishing timing relations among processing events, because the measurement task may itself introduce causes for differential delays in response. It is crucial to know what is controlling response time, to avoid misinterpreting any delayed response as evidence of delayed sensitivity to the input. The data reported here suggest that when explicit recognition of an anomaly is called for, as in the monitoring task of Experiment 3, the judgment of pragmatics is quite tardy, though there is no evidence of pragmatic delay when subjects listen to sentences for comprehension only, as in Experiment 2. Reading for comprehension, as in Experiment 1, falls somewhere between these poles. There is no overt judgment to be made and correspondingly no delay due to deliberation. But there are "decisions" to be made by a lower-level mechanism that controls the scanning of the text, and it does not treat pragmatic and syntactic problems alike. The methodological conclusion to be drawn would seem to be that for purposes of investigating autonomy of linguistic processing, measurement should be as indirect as possible. There must be some means of establishing when an anomaly has been detected, but this should not require subjects to indicate their evaluation of the linguistic status of the sentences. Dual-task paradigms are well-suited to this purpose. And they have the additional advantage of allowing the sentential input to be paced by the experimenter, so that the perceiver cannot slow processing down to a pace at which all levels of processing accommodated and no differences between them could emerge.

3. GENERAL DISCUSSION: AVAILABILITY AND MODULARITY

The data reported here suggest that there is virtually simultaneous (unconscious) processing of syntax and pragmatics, even under tough conditions where pragmatics might have been expected to fall behind. This is the most reasonable interpretation of the results of Experiment 2, and Experiments 1 and 3 do not in any way contradict it. Though perhaps unexpected, this is a welcome outcome for both sides of the modularity debate. Our results support the rapid availability of pragmatic facts. And this, as we will argue, is a pre-condition for using timing data to establish **either** the autonomy **or** the non-autonomy of syntactic processing.

For interactionists, the present findings have at least the virtue that they do not contradict empirical claims of immediate effects on parsing of all relevant forms of information. They validate the assumption that human language processing could be interactive not just architecturally but de facto as well: pragmatic information is accessible in time for it to have an influence on ambiguity resolution, if the structure of the system permits it to do so. (Whether or not it does permit it to do so—and what counts as pragmatic in the relevant sense—is the issue that has drawn most attention in recent years, but it is not addressed here.) Of course, we have availability data so far only on simple subject-predicate mis-matches like *the cats wouldn't bake the food*, and as noted above, more complex pragmatic inferences might become available at later times. We cannot now estimate at what point greater complexity of inference, or lack of sharp contrasts, gives rise to detectable delays.[6] But at least the current results imply that materials demanding fairly simple pragmatic processing can be used in experiments designed to reveal the interaction of all information sources on an equal footing, without danger that an availability delay will inadvertently mask the legitimacy of putting pragmatic information to work whenever and wherever it is useful.

The immediate availability of pragmatic facts is advantageous for the modularity hypothesis also. It protects it against objections of mere de facto autonomy compatible with an underlying interactive architecture. It implies that, whenever a delay of pragmatic influence on ambiguity resolution is successfully demonstrated (with a suitable task), it is most likely due to structural restrictions on how sub-processors are permitted to commune with one another. This saves proponents of modularity from having to demonstrate an even longer delay (a delay greater than the delay in availability of pragmatic information) to prove the same point. The possibility that pragmatic processing is slow for materials where the relevant inferences are unclear or complex must of course be borne in mind, as noted above. But at least it appears that simple pragmatic anomalies can confidently be used to look for signs of architectural restrictions on parsing.

Exactly how these restrictions operate needs to be explicated in a modular model. If there is essentially no pragmatic lag in normal processing of unambiguous strings, why and how should a detectable lag occur in the processing of ambiguous strings? We may distinguish three answers to this question. (For purposes of this discussion let us suppose that a temporal advantage of syntax over pragmatics in ambiguity resolution **has** been empirically established.)

(i) TIME DELAY. It might be that some fixed number of milliseconds is allotted to syntactic resolution strategies before pragmatics is consulted.

(ii) DOMAIN DELAY. The lag might be defined by some significant linguistic domain such as a clause or a theta-domain.

(iii) PRIORITY DELAY. There might be a priority ranking which requires non-syntactic factors to wait until syntax has done all that it is capable of doing with the current input item, however much or little that might be.

A priori, (ii) and (iii) seem more credible than alternative (i). A pure time delay just in the case of ambiguity resolution cannot be ruled out, but it has no obvious rationale. Note that it would have to be something other than a simple blockage or detour in the pipeline carrying input from the syntactic to the pragmatic processor that would delay the pragmatic processor equally in ambiguity resolution and in normal unambiguous processing. But even if (i) is implausible, either (ii) or (iii) could provide the modularity hypothesis with an adequate account of why a pragmatic lag shows up specifically in ambiguity processing.

Arguments for a domain delay have been given by Frazier and colleagues (e.g., Rayner et al., 1983; Frazier, 1990). The proposal is that syntactic and thematic/pragmatic analyses are conducted in parallel and that the two units compare notes at the end of each thematic domain. The thematic analysis therefore cannot be based on a full syntactic analysis but it could be fed by a crude recognition of lexical categories such as noun and verb. The thematic processor determines their optimal relations on the basis of plausibility, and reports its views to the syntactic processor. If they agree, the syntactic analysis proceeds without alteration; this would be the normal case for unambiguous non-anomalous sentences. For ambiguous or anomalous sentences the two processors may disagree, in which case the syntactic processor then looks for a different analysis that does not conflict with pragmatics. The delay before this reanalysis begins would be variable; it would be determined by the distance between the ambiguity and the end of the relevant thematic domain in the particular sentence under analysis. Note that no delay is predicted in the **registering** of a pragmatic anomaly (with impact on a concurrent task as in Experiment 2), but only in when it is **acted on** in establishing the structure for the sentence. This point is not explicitly discussed by Rayner et al. (1983) and Frazier (1990), and it may seem incompatible with their assumption that the thematic processor ignores syntactic constraints and arranges arguments and predicates in whatever way best suits itself. A thematic processor of this kind would not immediately spot a pragmatic anomaly consisting of a reversal of a plausible predicate-argument structure, such as *The icecream ate the children from the orphanage*; this would be detected only when the syntactic and thematic processors confer at the end of the clause. The model could be adjusted to predict anomaly detection in this case (e.g., by assuming the thematic processor employs heuristics such as an N-V-N strategy), but there is no need to do so to account for the present results, since the pragmatic anomalies in our materials were not curable by reversing them. Both a cat baking food and food baking a cat would offend the thematic processor. (Only one out of the 30 examples would be improved by reversing its arguments: *The pacifier we bought in Japan will drop the cranky baby....*)

Arguments for a priority delay model are given by Meltzer (1995) based on comparison of the processing of the empty categories PRO and *pro* of Government Binding Theory (Chomsky, 1981) in Spanish. Meltzer's results suggest that pragmatic selection of an antecedent for a dependent element may be either immediate or delayed depending on how much the grammar has to say about the choice of the antecedent. If syntactic principles do not determine the antecedent (as in the case of *pro*), pragmatic selection of the antecedent may begin immediately. But pragmatic selection must hold back if syntactic principles are relevant, e.g., if syntax entails that an obligatory antecedent will occur later in the sentence, as is the case for a controlled PRO in a clause that has been fronted. More generally, pragmatic delays would be expected to vary in length depending on whether or not the input word enters into a syntactic dependency with material to its right, and if so, how distant that material is. Such a dependency extends the domain in which syntax has information to contribute, and thereby postpones the point at which pragmatic guessing may begin. For model (iii), there would be essentially no pragmatic delay in cases where local information is fully sufficient to establish the correct analysis. Syntax tells pragmatics: *Wait until I'm through*; but where there is no local ambiguity, and no rightward dependency, syntax could be finished so quickly that there is no measurable waiting period for pragmatics at all. Why would there be delay in the case of ambiguity? It would result from the fact that syntax will have made its choice between the alternative analyses before pragmatics is given a chance to vote. When pragmatics enters on the scene, there is no decision left to be made: its only option is to accept the decision that syntax made, or else to override it. And overriding a decision once it is in place presumably takes more time and effort than initial decision making does. Therefore (depending on how difficult the revision process is[7]), effects of pragmatics on ambiguity resolution would be observed quite late.

This priority-delay variant of the modularity hypothesis raises a new methodological challenge. It reconciles the absence of a detectable pragmatic delay in normal processing with the possibility of a significant delay in ambiguity resolution, by the simple assumption that once syntax has started its work on an input item, it won't stop until it has done all that it can. As a result, even a very brief headstart for syntax can be magnified into a much greater one because of the difference between making a decision and overthrowing a decision. But now, this same kind of winner-takes-all priority system could be combined with a **non**-modular design in which **any** processor can compete for priority and the winner is determined by a race. Whichever gets there first with useful information to contribute gains the right to proceed and is permitted to complete its work before the other begins. The observable manifestations of such a system would resemble those of modular model (iii) if, de facto, syntax is usually the first to contribute relevant information—first by any tiny interval, and for whatever practical reason. In that case syntax would usually have the headstart and pragmatics could at best struggle against it and so would be delayed, though not as a matter of principle, and not because of fixed architectural barriers restricting information flow.

Once again, then, we see that the deductive link between overt timing facts and underlying design characteristics is very fragile. Unless this case can be excluded somehow on other grounds, it raises the stakes considerably on the investigation of timing relations in unambiguous sentence processing. It would no longer be sufficient, as a pre-condition on being able to demonstrate modularity, to establish the availability of pragmatics prior to the point at which pragmatics could have been useful but was not used. It would be necessary now to entirely eliminate the possibility of any systematic tendency for syntactic facts to become accessible slightly before pragmatic ones. This is hardly feasible with current methods. Yet anything less than simultaneity of availability could lend itself, as we have seen, to a winner-takes-all priority model compatible with either a modular or a non-modular design for the language processing system.

As observed above, this kind of indeterminacy weakens both sides of the debate. Interactionism might be true, and yet be obscured by delays overlaid on it by a headstart priority system. Syntactic autonomy might be true, and yet not demonstrable because its manifestations could always be non-autonomously accounted for. This is unfortunate. It points up the perilousness of the project that so much of psycholinguistics has occupied itself with in recent years: the project of deducing underlying mental organization from its one-dimensional projection onto chronometric relations between observable operations. The goal is to infer mental structure from facts about the timing of mental processes. This is always chancy, but it is particularly tricky when it is modal notions that are under test. We have to distinguish "these processes cannot interact" from "these processes could interact but do not." Unless we are prepared to give up on the quest for architectural conclusions, other sources of evidence may need to be found.

REFERENCES

Altmann, G. T. M. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes, 3*, 73-97.

Altmann, G. T. M., & Young, D. (1993). Factors affecting adaptation to time-compressed speech. *Proceedings of Eurospeech, 93*.

Bates, E., & MacWhinney, B. (1989). *Functionalism and the competition model*. Cambridge, UK: Cambridge University Press.

Boland, J. E. (submitted). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes*.

Boland, J. E., Tanenhaus, M. K., & Garnsey, S. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29*, 413-432.

Boland, J. E., Tanenhaus, M. K., Garnsey, S., & Carlson, G. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*, 774-806

Carlson, G. N., & Tanenhaus, M. K. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Syntax and semantics, Vol. 1: Thematic relations*. London: Academic Press.

Chodorow, M. S. (1979). Time-compressed speech and the study of lexical and syntactic processing. In: W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*. Cambridge: Cambridge University Press.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research, 23*, 407-434.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston MA: Houghton Mifflin Company.

Frazier, L. (1990). Exploring the architecture of the language-processing system. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press.

Friederici, A. D., Mecklinger, A., Steinhauer, K., & Hahne, A. (1995). Processing violations of syntactic structure versus violations of syntactic preference: Evidence from ERP studies. Paper presented at the 8th CUNY Conference on Human Sentence Processing, Tucson, AZ.

Gorrell, P. (1989). Establishing the loci of serial and parallel effects in sentence processing. *Journal of Psycholinguistic Research, 18,* 61-73.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 32,* 692-715.

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes, 4,* 287-336.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 134-157.

Meltzer, S. G. (1995). *Issues in the processing of null subjects.* Unpublished doctoral dissertation. City University of New York.

Mitchell, D. C., Corley, M. M. B., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory and cognition, 18,* 69-88.

Ni, W., Crain, S., & Shankweiler, D. (in press). Sidestepping garden paths: Assessing the contribution of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes.*

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (in preparation). Anomaly detection: Eye movement patterns

Ni, W., Fodor, J. D., Crain, S., Shankweiler, D., & Mattingly, I. G. (1993). Evidence of the autonomy of syntax and pragmatics. Poster presented at the 6th Annual CUNY Sentence Processing Conference, March, 1993. University of Massachusetts, Amherst, MA.

Osterhout, L., & Holcombe, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31,* 785-806.

Osterhout, L., Nicol, J., McKinnon, R., Ni, W., Fodor, J. D., & Crain, S. (1994). An event-related brain potential investigation of the temporal course of sentence comprehension. Poster presented at the 7th Annual CUNY Sentence Processing Conference. City University of New York, NY.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. J. (1995). Subject-verb agreement processes in sentence comprehension. Paper presented at the 8th CUNY Conference on Human Sentence Processing, Tucson, AZ.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction between syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22,* 358-374.

Shapiro, L., Zurif, E., & Grimshaw, J. (1987). Sentence processing and the mental representation of verbs. *Cognition, 27,* 219-246.

Swinney, D., & Osterhout, L. (1989). Inference generation during auditory language comprehension. In A. Graesser & G. H. Bower (Eds.) *The psychology of learning and motivation.* Academic Press.

Tanenhaus, M. K., Garnsey, S. M., & Boland, J. E. (1990). Combinatory lexical information and language comprehension. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives.* Cambridge, MA: MIT Press.

Tanenhaus, M. K., & Trueswell, J. C. (1994). Sentence comprehension. In J. Miller & E. Eimas (Eds.), *Handbook of perception and cognition: Volume 11.* Academic Press.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33,* 285-318.

Urbach, T., Pickering, M., Branigan, H., & Myler, L. A. (1995). Event-related potential effects and parsing. Poster presented at the 8th CUNY Conference on Human Sentence Processing, Tucson, AZ.

Young, D., Altmann, G. T. M., Cutler, A., & Norris, D. (1993). Metrical structure and the perception of time-compressed speech. *Proceedings of Eurospeech, 93.*

FOOTNOTES

**Journal of Psycholinguistic Research,* (1996), *25(1),* 25-57.

†Also CUNY Graduate Center.

‡Also University of Maryland.

‡‡Also University of Connecticut, Storrs.

[1] It might be countered that the recognition of anomaly (e.g., of *evidence examined*) at the ambiguous attachment site is itself evident for syntactic autonomy. The argument would be as follows: anomaly detection at *examined* indicates that the attachment ambiguity for *examined* was resolved on a syntactic basis and only then subjected to pragmatic evaluation; if the more plausible reduced relative analysis of (2b) had been selected immediately, there would have been no anomaly to detect. If correct, this would undercut our claim in this paper, that it is necessary to give independent evidence of timely access to pragmatic information. However, this counter-proposal presupposes that processing is slowed by an anomaly only if the parser has adopted the anomalous analysis. If, on the other hand, processing can be slowed by the anomaly of any analysis that the parser is contemplating as an option, then a difficulty at *examined* would not prove that a single analysis

had been selected (by syntax) at that point. Hence, difficulty at *examined* (ignoring data for the *by*-phrase) would be equally compatible with a "weakly interactive" parser, in which syntactic analyses are computed in parallel and pragmatic factors select between them (cf. Crain & Steedman, 1985). But then the data for *examined* would not bear straightforwardly on modularity, since a weakly interactive parser is also architecturally modular.

[2]In ERP studies, N400 responses to pragmatic anomaly contrast with late positivity (P600) in response to syntactic anomalies (Osterhout & Holcombe, 1992; Osterhout, Nicol, McKinnon, Ni, Fodor, & Crain, 1994). This probably should not be taken as evidence that pragmatic processing is faster than syntactic processing. Possibly the P600 represents a re-analysis process which follows breakdown of the first-pass parse, as suggested by Friederici, Mecklinger, Steinhauer, and Hahne (1995). If so, it does not illuminate current concerns.

[3]On the basis of the data previously available on sentence processing at approximately 50% compression, Chodorow observed (1979, p. 95): "Early investigations...of time-compressed speech revealed that intelligibility of individual words spoken in isolation remains quite high.... By contrast, **comprehension** of passage of connected text... is very poor at such compression rates.... These results mirror the subjective feeling often reported by listeners who experience a sensation of 'falling behind' the input when they are given time-compressed passages of text."

[4]See Trueswell et al. (1994) for the rationale for using this measure.

[5]Mean lexical decision RT (ms) for each sentence version and test point, by group, are as follows:

	GROUP A					GROUP B				
TEST POINT	-81	0	81	162	243	-81	0	81	162	243
SYNT ANOM	794	803	770	802	810	726	745	730	717	739
PRAG ANOM	795	817	812	796	763	728	722	729	743	720
BASELINE	794	746	802	822	749	738	690	740	755	736

[6]See Swinney and Osterhout (1989) on the speed of various types of inference involved in language comprehension. They distinguish "perceptual" inferences (rapid and mandatory, e.g., essentially immediate determination of antecedents for pronouns) from "conceptual" inferences (slow and "nonautomatic," e.g., metaphorical reasoning, presupposition, implicit instruments). Altmann (1988) presents data indicating a delay of approximately 300 ms due to the complexity of an inference relating a relative clause to its context.

[7]Fodor and Inoue (1994) have proposed, on the basis of informal judgments of processing difficulty in Japanese and English, that when the "symptom" of a garden path is a pragmatic anomaly (as in *They told the boy that the girl met the story*) the parser is less successful at finding the correct analysis than when the symptom is a syntactic anomaly (as in *They told the boy that the girl met not to go home*). If true, this suggests that pragmatic information cannot easily overthrow an established syntactic structure; only a syntactic anomaly has the power to face down a prior syntactic decision. This does not deny that pragmatic cues may be helpful in shaping the direction of a structural reanalysis (cf. Carlson & Tanenhaus, 1988) but it implies that they will be more effective if the need for reanalysis has been signaled syntactically.

APPENDIX: MATERIALS

Notes: For Experiment 1, the sentences as illustrated in (3) above were shortened by deletion of words at the beginning and end, but included the adverbs shown here in curly brackets. For Experiment 2, the words shown here in parentheses following the sentences were used as lexical decision targets. The alternative verb forms for all three experiments (syntactically anomalous, pragmatically anomalous, baseline) are given here in square brackets. Filler sentences are not included here.

1. It seems that the cats from across the road won't {usually} [eat / eating / bake] the food that Mary puts out on the porch every morning as soon as she gets up. (RINK)
2. Apparently the argument given by the astronomer might {even} [prove / proving / shout] that there are many canals on the moon though this has not been widely believed for over a decade. (MOTH)
3. In case of a break-in, the alarm system we just installed will {surely} [warn / warning / swear] that there is an intruder in the building and alert the local police department. (BISCUIT)
4. The new species of orchid that was discovered in Peru will {only} [grow / growing / sing] in tropical regions of South America or the islands around Madagascar. (HEXAGON)
5. This very expensive ointment from South East Asia will {supposedly} [cure / curing / loathe] all known forms of skin disease but only if it is used in accord with the instructions on the package. (METEOR)
6. This old electric blender that the bartender uses doesn't {really} [crush / crushing / own] icecubes any more but the management can't afford to spend money on a new one. (TRIBE)
7. This exotic spice from Aunt Ellen's kitchen may {possibly} [add / adding / seek] the subtle oriental flavor that John enjoys and is so difficult to find in this country. (HERMIT)
8. According to reports, the new fighter-plane that was tested in Nevada can {apparently} [fly / flying / walk] faster than anyone had expected it to when it was originally designed. (HUT)
9. The big wooden boxes in the attic may {still} [hold / holding / find] many old photographs and souvenirs from our trips abroad in the sixties when we hitchhiked all around Europe and India. (ELBOW)
10. Despite its other merits, this new test of mathematical reasoning might {occasionally} [fail / failing / hate] to discriminate between students of quite different abilities or aptitudes. (BLOUSE)
11. The inspector asserts that helicopters taking off from the roof may {repeatedly} [shake / shaking / paint] the walls and windows of the top floor and do considerable damage to the building. (TOOTH)
12. The plumber warned us that the leaking water he noticed yesterday might {slowly} [seep / seeping / speak] out from behind the refrigerator and ruin the linoleum tiles in the kitchen. (OUNCE)
13. They are confident that the fingerprints on the gun next to the body could {clearly} [prove / proving / judge] that the defendant is innocent though he had both motive and opportunity to commit the crime. (MATTRESS)
14. A family of beavers that lived in our duckpond would {sometimes} [chew / chewing / melt] the garden hose beside the shed so that we were unable to water the lawn. (SAUCER)
15. The fancy French clock that was selected by the mayor doesn't {always} [tell / telling / ask] the time during a power failure because the special batteries that it is designed to take are unavailable. (PEACH)
16. Critics say the latest rap songs that are played on MTV might {supposedly} [tend / tending / learn] to lead impressionable young people into immoral or indecent forms of behavior. (PETAL)
17. Those small red spiders with very long legs would {often} [spin / spinning / burn] beautiful webs in the rose bushes beneath the old maple trees near the barn. (SYNTHETIC)

18. I am sure that the pacifier we bought in Japan will {immediately} [soothe / soothing / drop] the cranky baby within a few minutes and then we will all be able to get some peaceful sleep at last. (SPLINTER)
19. People complained that the skyscraper being built by the city would {eventually} [block / blocking / seek] out the sunlight even in the middle of the day because it is positioned too close to the other buildings in the street. (BRIDESMAID)
20. Unfortunately, these grape vines from Southern France don't {usually} [grow / growing / jog] well in sandy regions where the topsoil is too loose to provide enough support for their long roots. (NICOTINE)
21. One elderly kangaroo in the San Diego Zoo would {just} [sit / sitting / swear] all day at the gate that the keeper usually enters through when he brings food and fresh water. (EXPLOSION)
22. The full-length portrait of great uncle Henry doesn't {really} [look / looking / talk] like him or like anybody ELSE in the family but it is extremely handsome. (NATIONALISM)
23. Hopefully the new heater in the maid's room should {quickly} [dry / drying / kick] the laundry that she hangs over the towel rack when the weather is too bad to use the out-door clothesline. (EPISODE)
24. Don't you think that the strawberry beds being planted by the gardener might {soon} [tempt / tempting / lift] rabbits and other animals into the backyard and create a serious problem of pest-control for the future? (IDIOTS)
25. The exquisite colors woven into these sweaters shouldn't {ever} [fade / fading / cry] when they are washed in hot soapy water but I think it is always safer to send things to the dry cleaner. (GLOBE)
26. A chemical additive now being tested may {also} [tend / tending / want] to lower the freezing point of sea water so that ships can be kept ice-free all winter long. (SANCTUARY)
27. The sleek black sea lions that inhabit the little bay can {happily} [bask / basking / read] on the beach all day long when the weather is fine and can sleep on the rocky ledge at night. (GHOST)
28. We are pleased to report that the security camera at the bank will {now} [take / taking / tear] photographs of everyone who uses the automatic cash machine or the overnight deposit box. (MEDAL)
29. Sam is scared because the bull that escaped could {easily} [smash / smashing / mend] the wooden fence around the meadow and get into the field where the sheep are grazing. (MAST)
30. It is clear that the lever on the basement wall does not {reliably} [shut / shutting / leap] off the air-conditioning unit or the power supply to the elevators. (GEOGRAPHY)

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 139-173

Sidestepping Garden Paths: Assessing the Contributions of Syntax, Semantics and Plausibility in Resolving Ambiguities*

Weijia Ni, Stephen Crain,† and Donald Shankweiler‡

A central issue in the study of sentence processing is the manner in which various sources of information are used in resolving structural ambiguities. According to one proposal, the garden path model (e.g., Frazier & Rayner, 1982), perceivers are initially guided by strategies based solely on the structural properties of sentences. Another class of models, constraint satisfaction models, emphasizes the influence of lexical properties in decisions among the alternative analyses of an ambiguous sentence fragment (e.g., Tanenhaus, Garnsey, & Boland, 1991). In this paper we explore the predictions of an alternative model, the referential theory (e.g., Crain & Steedman, 1985). The referential theory maintains that the relative complexity of discourse representations plays a key role in determining the perceiver's immediate parsing preferences. We present four experiments designed to weigh the influence of semantic/referential complexity and general world knowledge in the on-line resolution of two kinds of structurally ambiguous sentences. In each experiment, we examined pairs of sentences that were identical except for the alternation between the definite determiner THE and the focus operator ONLY. Two techniques were used to assess ambiguity resolution: Word-by-word reading and eye movement recording. The results indicate that semantic/referential principles are applied immediately in on-line ambiguity resolution, and that these principles pre-empt general world knowledge. The use of world knowledge was found to depend on working memory capacity, whereas the resolution of ambiguity by means of semantic/referential principles appeared to be independent of memory resources. Taken together, the findings are interpreted as support for the referential theory of ambiguity resolution.

1. INTRODUCTION

One of the central goals of psycholinguistic research is to provide a systematic account of how people interpret structurally ambiguous sentences. This paper presents the results of an interlocking set of experiments that shed further light on the bases of ambiguity resolution. Two experiments focus on the kind of structural ambiguity exhibited by Bever's (1970) well-known "garden path" sentence *The horse raced past the barn fell.* In this sentence, the verb *raced* is morphologically ambiguous: It can be analyzed as a simple past tense main verb, or as a past-participle. It is striking that most people find the sentence extremely difficult to process at the

The research reported in this paper was supported in part by a Project Program HD-01994 from the National Institute of Child and Human Development to Haskins Laboratories and in part from a grant to Stephen Crain and Donald Shankweiler from the University of Connecticut Research Foundation. Portions of this research were presented at meetings of the Annual CUNY Conference on Sentence Processing: Rochester, NY, 1991; Amherst, MA, 1993; and New York, NY, 1994. We would like to express our gratitude to Michael Tanenhaus and John Trueswell for providing us with operating software and instructions in conducting eye-movement research. Thanks are also due to Slavoljub Milekic for his assistance in setting up the eye-tracking laboratory, and to Meredith Daneman for her courtesy in providing the materials for the memory-span test. Finally, we are indebted to Gerry Altmann, Don Mitchell, and an anonymous reviewer for many insightful comments on earlier drafts of the paper. Errors that remain are our own.

verb *fell*, suggesting an overwhelming preference to analyze *raced* as a main verb: This decision leads to a "garden path effect." Two additional experiments investigate a more subtle garden path effect,[1] the momentary semantic anomaly that is manifested in sentences like *The spy saw the* cop *with a revolver*. As Rayner, Carlson and Frazier (1983) have demonstrated, the anomaly arises because people prefer to analyze the prepositional phrase *with*...to modify the verb *saw*, rather than the immediately preceding noun phrase *the* cop. Evidence of a preference for verbal attachment in constructing a structural representation was found by comparing eye movement patterns during reading in sentences such as *The spy saw the* cop *with a revolver* and sentences such as *The spy saw the* copy *with binoculars*. Apparently, an elevation in eye fixation durations occurs when people encounter the noun phrase *a revolver*. Because the prepositional phrase is attached to the verb, this noun phrase makes the sentence anomalous. By contrast, the noun phrase *binoculars* is a plausible continuation of the sentence fragment *The spy saw the* cop *with....*

The existence of garden path effects and semantic anomalies is evidence that the human sentence processing mechanism (the parser) makes rapid decisions about which alternative, grammatically well formed structural representation to adopt when the input is ambiguous. The factors that influence the decision-making of the parser and the manner in which the parser is influenced are matters of controversy, however. One account of ambiguity resolution is known as the *garden path model* (Frazier, 1979; Frazier & Rayner, 1982; Rayner et al., 1983; Ferreira & Clifton, 1986; Clifton & Ferreira, 1989). The garden path model maintains that the parser is a serial processing device. The model contends that the parser's initial analysis of ambiguity is based solely on structural properties of the linguistic input. One structurally-based parsing strategy, Minimal Attachment, instructs the parser to pursue the analysis that postulates the fewest non-terminal nodes in constructing the phrase structure representation of a sentence. The model predicts that in the sentence *The horse raced past the barn fell*, the word *raced* is initially analyzed as the main verb because this analysis is structurally simpler than the alternative analysis on which the verb *raced* is analyzed as a past participle. Consequently, the "real" main verb *fell*, which comes later, cannot be readily incorporated into the analysis. The main verb analysis thus leads the parser down a "garden path." Although the parser uses only structural information in making its initial decisions, according to the garden path model, other sources of information contribute to reanalysis if the initial analysis turns out to be incompatible with subsequent linguistic material. For example, Ferreira and Clifton (1986) propose that a second stage "thematic processor" operates on the output of the syntactic component of parsing. When triggered by "an error signal *in the disambiguating region*" (Ferreira & Clifton, 1986, p. 366; emphasis ours), the thematic processor supplies alternative thematic representations that may prove useful in structural reanalysis.

A recent version of the garden path model, by Mitchell, Corley, and Garnham (1992), maintains that the parser may begin to revise a misresolved ambiguity even before the point of disambiguation. According to this account, the effects of syntactic parsing strategies such as Minimal Attachment persist only briefly, perhaps for no longer than a word or two. This allows the parser sufficient time to begin processing semantic and discourse information, which may then be used in the reanalysis of the sentence, if that is needed. Mitchell et al. (1992) point out that most of the early research on the influence of discourse representations on ambiguity resolution test for syntactic commitments two or more words after the onset of ambiguity. Such test points arrive too late to reveal the effects of parsing strategies, according to Mitchell et al., because discourse factors may have had sufficient time to override the parser's initial syntactic commitment. It is therefore necessary to test for the parser's initial syntactic commitment at the "earliest feasible point" in the unfolding structural analysis assigned by the parser.[2] It is important to note that on all versions of the garden path model, non-structural sources of information can only serve as evidence confirming or disconfirming the parser's initial structurally-based decision. Recent research motivated by this theoretical framework has therefore been concerned with the costs incurred by reanalysis, and with the diagnostics used by the parser in recovering from ambiguities that have been misresolved (Frazier, 1994; Frazier & Clifton, in press).

While a number of studies provide support for one or another version of the garden path model (e.g., Britt, Perfetti, Garrod, & Rayner, 1992; Murray & Liversedge, 1994; Rayner, Garrod, & Perfetti, 1992), much recent work has led to alternative accounts of ambiguity resolution that explore the possibility that on-line decisions of the parser are affected by a range of non-structural factors. Adopting the terminology used in a recent review article by Tanenhaus and Trueswell (in press), we refer to one general line of research as the *constraint satisfaction model.* Proponents of this approach have investigated the on-line influence of lexically-based factors such as verb frequency, information from argument structure and conceptual-semantic information. Each of these factors is assumed to play a role in evaluating the alternative structural representations of ambiguous sentences (e.g., Boland, Tanenhaus, & Garnsey 1990; Juliano & Tanenhaus, 1993, 1994; MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994; Merlo, 1994; Spivey-Knowlton & Sedivy, in press; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994; Taraban & McClelland, 1990; Trueswell & Tanenhaus, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). Tanenhaus and Trueswell (in press) conclude, "When these [lexically-based] factors are quantified and combined...there is no need for either an initial category-based parsing stage or a separate revision stage." In a similar vein, other relevant factors have been identified, including intonation and prosody (e.g., Beach, 1991; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Nagel & Shapiro, 1994; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Speer, Crowder, & Thomas, 1993), and the memory costs associated with the processing of various syntactic constructions (e.g., Gibson, in press; Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992).

So far we have identified the garden path model and the constraint satisfaction model. In this paper, we pursue the predictions of a third model. On this model, yet another class of factors is viewed as essential in the resolution of structural ambiguities, namely, the referential properties of sentences. We call this model the *referential theory* (e.g., Crain & Steedman, 1985, Altmann & Steedman, 1988, Ni & Crain, 1990). According to the referential theory, the complexity of the alternative discourse representations (corresponding to the alternative structural analyses) is often crucial in the resolution of structural ambiguities. A wide variety of parsing preferences that have often been attributed to structural properties of sentences are viewed by the referential theory as consequences of the application of semantic/referential principles.

Both the constraint satisfaction model and the referential theory can be contrasted with the garden path model in certain respects. Each of the former typically maintains that the parser computes multiple (partial) structural analyses of an ambiguous phrase; the parser is regarded as a *parallel processing* mechanism. In addition, both models take the position that real-world knowledge is invoked to resolve some structural ambiguities. Crain and Steedman (1985) have explicitly proposed and supported the claim that parsing decisions are influenced by considerations of general knowledge of the world: "If a reading is more plausible in terms either of general knowledge about the world or of specific knowledge about the universe of discourse, then, other things being equal, it will be favored over one that is not." (p. 330) While the use of information about the *a priori* plausibility of the alternative readings of an ambiguous sentence fragment is acknowledged, the referential theory maintains that principles of discourse pre-empt *a priori* plausibility. As Crain and Steedman put it: "...in case of a conflict between general and specific knowledge, the latter must clearly take precedence." (op. cit.)

The research presented in this paper is designed to test these two specific tenets of the referential theory, i.e., the claim that there is immediate application of semantic/referential principles in the resolution of ambiguity, and the claim that these principles pre-empt general knowledge of the world. A further goal of the present study was to examine the manner in which these two different sources of information are used by subjects with varying working memory capacities to resolve structural ambiguities and to recover from garden paths. To accomplish these goals, two experiments investigated main-verb/reduced-relative-clause ambiguity, and two investigated ambiguities involving the attachment of prepositional phrases. The experiments were conducted using two experimental techniques: Self-paced word-by-word reading, and eye-movement recording.[3] In the two experiments on main-verb/reduced-relative-clause ambiguities, the structural content of the test sentences informed the subject whether or not he/she has been led down a garden path. In contrast, in the two experiments testing the preferences for attachment of prepositional phrases, the subject was informed by *a priori* plausibility when an incorrect analysis had been pursued. The results of the latter experiments were analyzed by dividing subjects into groups according to individual differences in memory span (Daneman & Carpenter, 1980). Between-group comparisons enabled us to distinguish properties of sentence processing that are relatively undemanding of memory resources from properties that are highly sensitive to limitations in memory. The view has been put forward that people with high memory spans are better able to maintain parallel alternative structural analyses of an ambiguous sentence, and are better able to use a variety of sources of information in resolving ambiguities (MacDonald et al., 1992; Pearlmutter & MacDonald, 1995). The findings of our research are consistent with these claims. All our subjects appeared to be capable of applying semantic/referential principles on-line to resolve local ambiguities, but those with greater memory capacity also proved able to rapidly access real-world knowledge to recover from a misanalysis, whereas subjects with less memory capacity typically re-read portions of ambiguous sentences when the use of real-world knowledge was required to recover from a misanalysis.

The findings bear directly on the issues of how and when ambiguities are resolved and the costs that are incurred. As we discussed, the referential theory leads to the expectation that decisions concerning specific discourse representations should be made earlier than decisions that are based on general world knowledge. In a resource-limited system, such as the human sentence processing mechanism, disruptions should occur most often at later stages of processing; if principles used to construct discourse representations are accessed earlier than information about the real world, then the former should be less likely than the latter to exhaust memory resources. We appeal to this assumption of the referential theory to explain why individuals with higher memory capacity appear to be better able to access real-world knowledge on-line in order to recover from a misanalysis.

Since this research was motivated by the referential theory, it is appropriate to spell out its operating principles in greater detail (Section 2). Following that, in Section 3, we describe the semantic properties of the focus operator ONLY, and explain the rationale for alternating ONLY and THE as pre-nominal modifiers in our experimental manipulations. Section 4 contains the experimental findings and, finally, Section 5 is a general discussion the findings.

2. THE REFERENTIAL THEORY

The referential theory contends that primary responsibility for resolving structural ambiguities rests with the immediate, word-by-word evaluation of alternative structural analyses by the semantic/discourse processor. On this view, no particular structural configurations are intrinsically prone to elicit garden paths but, instead, certain discourse contexts either promote or deter garden path effects. The theory assumes a "weak" interaction between components of the language processing system. The syntactic processor putatively computes multiple (partial) structural analyses when it encounters an ambiguous sentence fragment. The alternative analyses are shunted to the semantic/discourse processor, which chooses among them. Here is a list of the basic tenets of the theory:

A. All permissible structural analyses of an ambiguous sentence are computed in parallel by the syntax. They are presented to the semantic/discourse processor for adjudication.
B. Semantic evaluation is carried out incrementally, more or less word by word.
C. The semantic/discourse processor evaluates and chooses among the alternative syntactic analyses on the basis of their fit to the conversational context.
D. If no decision is rendered by the semantic/discourse processor, then factors such as general knowledge of the world may be used to decide on the analysis to pursue.

Ambiguity resolution is ordinarily achieved within some discourse context. According to the referential theory, decisions by the parser follow what Crain and Steedman have called *the principle of referential success*: "If there is a reading that succeeds in referring to entities already established in the perceiver's mental model of the domain of discourse, then it is favored over one that does not." (Crain & Steedman, 1985, p. 331). Another version of this principle was formulated by Altmann and Steedman (1988), as *the principle of referential support*: "An NP analysis which is referentially supported will be favored over one that is not."

The referential theory also explains how ambiguities are resolved in absence of context. When processing a sentence in the so-called null context, the perceiver actively attempts to construct a mental representation of a situation that is consistent with the sentence. In addition to the characters and events *asserted* in a sentence, the construction of a mental model of the situation sometimes requires the perceiver to represent information that a sentence *presupposes*. The process of augmenting one's mental model to represent the presuppositional content of sentences has been called "the accommodation of presuppositional failure" by Lewis (1979), "extending the context" by Stalnaker (1974) and Karttunen (1974), and the "addition of presuppositions to the conversational context of an utterance" by Soames (1982).

On the referential theory, the accommodation of presuppositional failure plays a critical role in explaining how ambiguous sentences are interpreted outside of context. The parser attempts to construct all permissible discourse representations of a sentence but, due to limited computational resources, it settles on the analysis that requires the fewest modifications in establishing a coherent representation. Crain and Steedman (1985, p. 333) call this *the principle of parsimony*: "If there is a reading that carries fewer unsatisfied but consistent presuppositions than any other, then that reading will be adopted and the presuppositions in question will be incorporated in the perceiver's mental model."

The principle of parsimony can explain why the sentence *The horse raced past the barn fell* produces a garden path effect in the absence of context: At the onset of the ambiguous phrase, i.e., at the verb *raced*, the parser must actively create a mental model of a discourse in which the sentence could felicitously occur. According to the principle of parsimony, when there is a choice between alternative analyses, the analysis that requires the fewest extensions to the mental model is favored. The noun phrase *the horse* leads the parser to assume that a particular horse is in the domain of discourse. To make felicitous the alternative reduced relative clause analysis, the parser would have to establish a representation in which there is more than a single horse, with one of the horses being raced by someone. Because nothing in the fragment *The horse raced*...demands such additions to the mental model of the discourse, the main verb analysis of *raced* is favored, and the parser is led down the garden path.

Empirical support for the referential theory has come chiefly from studies that show the on-line influence of linguistic context on the resolution of structural ambiguities (Altmann & Steedman, 1988; Altmann, Garnham, & Dennis, 1992; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Spivey-Knowlton & Tanenhaus, 1994).[4] However, much debate has centered on the *immediacy* of such influences, i.e., whether or not contextual information is available early enough to be effective in resolving local ambiguities (e.g., Altmann, Garnham, & Henstra, 1994; Ferreira & Clifton, 1986; Mitchell et al., 1992; Crain & Steedman, 1985; Fodor, Ni, Crain & Shankweiler, in press). It is apparent that the information represented in the mental model of an extended discourse may not be accessed immediately, and that the effectiveness of context on parsing may be differentially affected by characteristics of the contextual manipulations. A more rigorous comparison between the different accounts of ambiguity resolution can be made, therefore, if garden path effects are manipulated without providing an explicit discourse context. The experiments we report in this paper follow this research strategy. The test sentences differ only in a single respect: One version of each sentence contains the focus operator ONLY, and another version contains the definite determiner THE (e.g., *Only horses raced past the barn fell.* vs. *The horses raced past the barn fell*). Because the experiments involve *minimal pairs* of sentences, the influence of the referential properties of sentences is investigated in isolation from other factors. The experimental manipulations vary only the referential content of the test materials, while holding constant the effects of verb frequency, argument structure, semantic/conceptual information, and so on.

To sum up, by manipulating referential properties sentence-internally, the present experiments circumvent problems that have sometimes plagued studies attempting to investigate referential effects by manipulating discourse context. Because the referential contributions of the focus operator ONLY, and of the definite determiner THE, are essential to incremental semantic interpretation, the influence of these pre-nominal modifiers in the resolution of ambiguity is likely to be felt immediately. Indeed, we will demonstrate that garden path effects can either be instigated or deterred by substituting one pre-nominal modifier for another. Thus, with all other factors held constant, the present studies demonstrate robust effects of changes in referential content (which arise by substituting ONLY for THE in the experimental materials) on the decisions made by the parser when an ambiguity is encountered. These findings are predicted by the referential theory. In the following section, the rationale for using the focus operator ONLY in the experimental manipulations is explained more fully.

3. THE FOCUS OPERATOR ONLY

The semantic function of the focus operator ONLY is to signal that the denotation of a linguistic constituent, which we call the focus element, is being contrasted with a set of alternatives. Consider (1):

(1) In New Haven, only Willoughby's coffee is really good.

It is appropriate to use the sentence in (1) only if coffee from Willoughby's is being compared to coffee from other shops in New Haven. If the speaker had sampled coffee from Willoughby's, but nowhere else, it would be infelicitous to utter (1). Notice, however, that the use of sentence (1) does not *assert* that a comparison among coffee shops has been made; rather, this comparison is *presupposed* to have occurred prior to the utterance of (1). The presupposition that coffee from other shops has been sampled is triggered by the focus operator ONLY.

The semantic representation of sentences with the focus operator ONLY can be partitioned into three parts. One part represents *background information,* a second represents *the element in focus,* and the third represents *a contrast set*—the alternatives to the focus element. The contrast set is not mentioned explicitly in the sentence; instead, it is presupposed to exist. Two conditions must be met for sentences with ONLY to be true. First, the information in the background must apply to the element in focus. Second, the background information must *not* apply to any members of the contrast set. That is, the background must apply *uniquely* to the focus element.[5]

Based on semantic properties associated with ONLY, the referential theory predicts that sentences like (2) will not evoke garden path effects, but that ones like (3) will. These differences are expected despite the fact that (2) and (3) are identical following the initial noun phrase; in particular, the sentences are identical at the point of disambiguation and thereafter.

(2) *Only businessmen loaned money at low interest were told to record their expenses.*

(3) *The businessmen loaned money at low interest were told to record their expenses.*

According to the referential theory, the subject NP *only businessmen* in (2) causes the parser to establish a discourse representation (a mental model) of the conversational context in which a set of businessmen is represented. The pre-nominal modifier ONLY in the initial NP prompts the parser to search for a contrast set. If a contrast set has not been previously established in the discourse, the parser has two options. First, it could attempt to construct a contrast set "from scratch." That is, the parser could conjure up some set of individuals to be contrasted with businessmen. There is a second option, however. Since the verb *loaned* is ambiguous, the parser could choose to satisfy the presupposition associated with ONLY by adopting the reduced relative clause analysis of the verb phrase. Pursuing this second option requires the parser to partition the set of businessmen already admitted into the mental model, rather than adding new entities. According to the principle of parsimony, the second option should therefore be preferred.

If a decision is made to analyze the ambiguous fragment as a reduced relative clause, no garden path effect will occur when the main verb (*were told*) is encountered. On the referential theory, then, sentences like (2), which begin with the focus operator ONLY, should tend to pattern like sentences with an unambiguous verb, such as (4).

(4) *The vans stolen from the parking lot were found in a back alley.*

There is a further prediction of the referential theory: If a contrast set is established *before* the ambiguity is encountered, then garden path effects should tend to emerge in sentences with ONLY. This is illustrated in (5).

(5) *Only wealthy businessmen loaned money at low interest were told to record their expenses.*

The phrase *wealthy businessmen* in (5) satisfies the requirement of setting up a contrast set: the set of businessmen who are not wealthy. Having established the contrast set in advance of the ambiguity, the main verb analysis is more highly favored (by the principle of parsimony, as discussed earlier). Adopting the main verb analysis results in a garden path effect, however, when the real main verb, *were told,* is encountered.

In the following sections, we report the results of four experimental studies that test the predictions of the referential theory, using both the self-paced word-by-word reading paradigm and the technique of monitoring subjects' eye movements during reading. The former technique is used to give continuity with past research and the latter to gain greater naturalness and finer temporal resolution. Experiments 1 and 2 test sentences with main-verb/reduced-relative-clause ambiguities. Experiment 3 and 4 test sentences with ambiguous attachment sites for prepositional phrases.

EXPERIMENT 1

The purpose of Experiment 1 was to examine whether the parser's on-line decisions are affected by manipulations of the referential properties in sentences containing main-verb/reduced-relative-clause ambiguities. A single substitution of either the focus operator ONLY or the definite determiner THE was made within the ambiguous test sentences. This substitution does not alter the syntactic structure at the point of ambiguity, i.e., at the ambiguous verb, but it does alter the referential content of the initial NP. According to the referential theory, sentences in which the word ONLY precedes a noun that is followed by an ambiguous verb should not induce garden path effects at the main verb, in contrast to their counterparts that substitute the word THE. In addition, the ambiguous test sentences were manipulated by including or excluding an adjective in the noun phrase that contained either THE or ONLY. Garden path effects are expected to occur for both these sentences, following the referential theory.

Method

Subjects. Thirty-two undergraduate students participated in the experiment. All were native speakers of English, and were naive about the purpose of the experiment.

Materials. Thirty-two ambiguous test sentences and sixteen unambiguous controls were constructed for the experiment.[6] There were four versions of each of the test and control sentences. One version of the test and control sentences contained the definite determiner THE in the initial noun phrase ("The-amb" and "The-unamb"); one version of each contained the word ONLY in the initial noun phrase ("Only-amb" and "Only-unamb"); one version of each contained THE and an adjective in the initial NP ("The-adj-amb" and "The-adj-unamb"); and, finally, one version of the test and control sentences contained the word ONLY followed by an adjective in the initial NP ("Only-adj-amb" and "Only-adj-unamb"). A full list of test and control sentences can be found in Appendix A. The following is an example of one complete set of test and control sentences.

<u>Ambiguous Test Sentences</u>:

The-amb	*The businessmen loaned money at low interest were told to record their expenses.*
Only-amb	*Only businessmen loaned money at low interest were told to record their expenses.*
The-adj-amb	*The wealthy businessmen loaned money at low interest were told to record their expenses.*

Only-adj-amb	*Only wealthy businessmen loaned money at low interest were told to record their expenses.*

Unambiguous Control Sentences:

The-unamb	*The vans stolen from the parking lot were found in a back alley.*
Only-unamb	*Only vans stolen from the parking lot were found in a back alley.*
The-adj-unamb	*The new vans stolen from the parking lot were found in a back alley.*
Only-adj-unamb	*Only new vans stolen from the parking lot were found in a back alley.*

The test and control sentences were intermixed in the experiment. A counterbalanced design yielded four lists of stimuli, such that no more than a single version of any particular test or control sentence was present in each list. Each of the four lists contained 32 ambiguous test sentences (with 8 tokens of each of the 4 versions) and 16 unambiguous control sentences (with 4 tokens of each of the 4 versions). These sentences were interspersed among 92 filler sentences. The fillers included a variety of structures, half of which were grammatically ill-formed. Eight subjects were randomly assigned to be tested on one of the four stimulus lists. All versions of test and control sentences were shown to each subject; therefore, the substitutions between THE and ONLY, and the presence and absence of an adjective in the initial noun phrase are within-subject (and within-item) variables, while ambiguity is a within-subject but between-item variable.

Procedure. The experiment used a grammaticality judgment task embedded in the self-paced word-by-word reading paradigm (Ford, 1983; Kennedy & Murray, 1984). Each consecutive word in a test sentence appeared on a CRT, from left to right, at the request of the subject by way of a key press. The words accumulated on the screen. The subject's task was to press a key marked "YES" if each newly-appearing word was a grammatical continuation of the previous material. Subjects were instructed to press a key marked "NO" whenever a sentence fragment stopped being grammatical. The "NO" key was used by subjects thereafter to finish displaying the remainder of the sentence. The computer recorded the time elapsed, in milliseconds, between the onset of each new word and the next key press. Subjects' responses ("YES" or "NO" to each word) were also recorded by computer. A short pretest was conducted to familiarize the subjects with the task.

Results and Discussion

Two-way ANOVAs were performed separately for the ambiguous test sentences and for the unambiguous control sentences. These ANOVAs examined the two types of pre-nominal word (THE/ONLY) and the presence or absence of an adjective (ADJ/NOADJ) intervening between the word THE or ONLY and the head noun. Planned comparisons by subject were performed between ambiguous test sentences and unambiguous controls. The dependent variables were mean reaction times and percent of errors. The reaction time data included the time subjects took at each word to correctly judge that the word was a grammatical continuation of the ongoing sentence fragment. A "NO" response to any word in a test or control sentence was counted as an error, and reaction times on any sentence in which an error occurred were excluded from the reaction time analyses.

For comparisons of reaction times and error rates, test and control sentences were divided into six regions. Region 1 contained the subject NP *(The / Only (wealthy) businessmen).* Region 2 included the first verb *(loaned)*, which is morphologically ambiguous in the test sentences. Region 3, *(money at low),* contained the remainder of the first verb phrase excepting the last word. The sole content of Region 4 was the last word in the first verb phrase *(interest).* Region 5 was the region of focal interest in this experiment. It contained two words: either an auxiliary verb and the main verb, or the main verb and the following word *(were told).* These words either confirmed a correct analysis or corrected a misanalysis. The final region included the remainder of the sentence minus the terminating word *(to record their).* The final word was excluded from the analysis to avoid the distorting influence of end-of-sentence wrap-up effects (see Just & Carpenter, 1980).[7]

Figure 1 depicts the mean reaction time per word at each of the six regions for the ambiguous test sentences (left) and for the unambiguous control sentences (right).

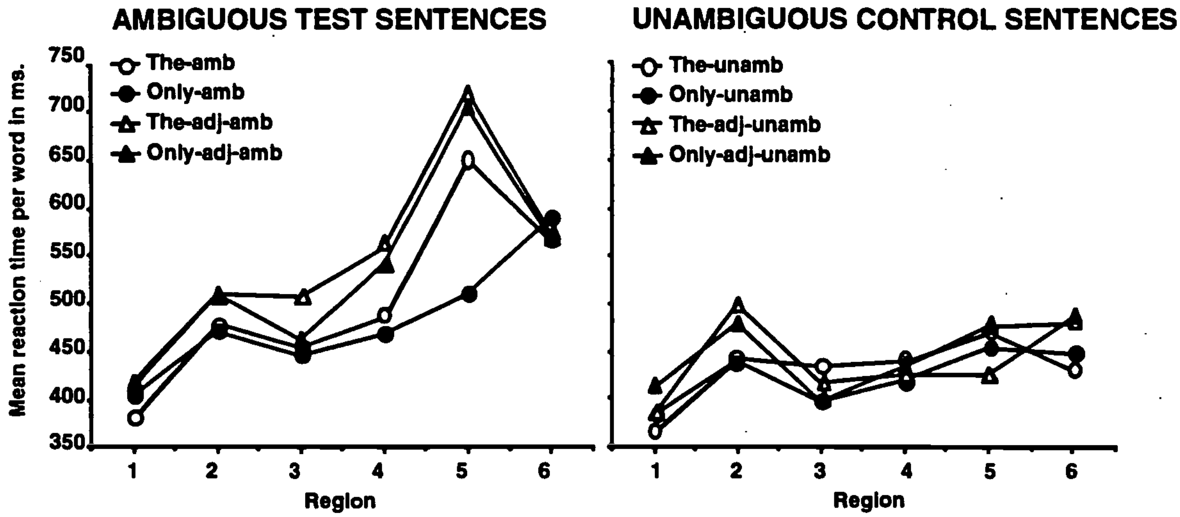


Figure 1. **Experiment 1—Mean reaction time per word at each region for ambiguous test sentences (left) and unambiguous control sentences (right)**

Ambiguous sentences and unambiguous controls were analyzed separately. At Regions 1 through 4, there was no significant effect of the substitution between THE and ONLY (hereafter THE/ONLY), or between the presence and absence of an adjective in the initial noun phrase (hereafter ADJ/NOADJ). However, average reaction time were greater for ambiguous sentences than for unambiguous controls at Region 2 ($F(1,31) = 4.83, p < .04$), Region 3 ($F(1,31) = 18.21, p < .01$), and Region 4 ($F(1,31) = 17.89, p < .01$).[8]

At Region 5, a two-way analysis of variance revealed a main effect of THE/ONLY for the ambiguous test sentences ($F_1(1,31) = 10.31$, $p < .01$; $F_2(1,31) = 3.55$, $p < .07$): ambiguous sentences containing THE induced longer reaction times than those containing ONLY, although the effect only approached significance in the analysis by items. There was also a main effect of ADJ/NOADJ ($F_1(1,31) = 3.54, p < .07; F_2(1,31) = 8.14, p < .01$): longer reaction times were found in ambiguous sentences with an intervening adjective than in those without one, although in this case, the effect by subjects only approached significance. There was no interaction between these factors, however. The lack of interaction is presumably due to the high variance associated with the sentences that evoked longer reaction times. Of the four versions of test sentences, three—"The-amb," "The-adj-amb," "Only-adj-amb"—were expected to produce garden path effects, but "Only-amb" sentences should not induce a garden path effect, according to the referential theory. This interpretation is supported by a comparison of standard deviations ("Only-amb" = 123.23, "The-amb" = 335.61, "The-adj-amb" = 356.78, "Only-adj-amb" = 336.91). Indeed, a planned contrast between the reaction times of the "Only-amb" sentences and an average of the reaction times of the other three versions revealed that the former (mean = 502.11 ms.) was significantly shorter than the latter (mean = 695.76 ms.) ($t(1) = 18.94; p < .01$).

While there were no significant main effects or interactions among the four versions of the unambiguous control sentences at Region 5, a planned comparison between ambiguous and unambiguous sentences (hereafter AMB/UNAMB) revealed a significant main effect ($F(1,31) = 45.49$, $p < .01$), with longer reaction times associated with the ambiguous sentences. Sentences with THE were also found to induce longer reaction times than those with ONLY ($F(1,31) = 5.90$, $p < .03$). The difference between sentences with or without an adjective was non-significant ($F(1,31) = 3.22$, $p < .09$). There was a significant interaction between the factors THE/ONLY and AMB/UNAMB ($F(1,31) = 8.70$, $p < .01$): whereas sentences with THE induced longer reaction times in the ambiguous cases, this was not true of the unambiguous cases. The interaction between THE/ONLY and ADJ/NOADJ approached significance ($F(1,31) = 3.84$, $p < .06$): an intervening adjective induced longer reaction times in the ambiguous sentences but not in the unambiguous ones. There was not a significant three-way interaction (i.e., AMB/UNAMB by THE/ONLY by ADJ/NOADJ).

Turning to the analyses of the error data, we conducted separate two-way (THE/ONLY by ADJ/NOADJ) analyses of variance for the ambiguous test sentences and for the unambiguous controls. The profiles by region are depicted in Figure 2.

As Figure 2 indicates, there were no significant differences between versions in the first three regions. At region 4, the word before the disambiguating region, more errors were found for ambiguous sentences with an adjective than ones without ($F(1,31) = 4.29$, $p < .05$). For the unambiguous sentences, the error rate was higher for sentences containing THE than for sentences containing ONLY ($F(1,31) = 8.75$, $p < .01$).

At Region 5, for ambiguous sentences, there was a main effect of THE/ONLY ($F_1(1,31) = 36.78$, $p < .01$; $F_2(1,31) = 34.55$, $p < .01$), and a main effect of ADJ/NOADJ ($F_1(1,31) = 27.35$, $p < .01$; $F_2(1,31) = 23.37$, $p < .01$). In addition, there was a significant interaction of THE/ONLY vs. ADJ/NOADJ ($F_1(1,31) = 5.20$; $p < .03$; F_2 $(1,31) = 8.49$, $p < .01$). More errors were made on ambiguous sentences with THE than on ones with ONLY, and for those with an adjective than those without one. For the unambiguous controls, no such effects existed. There was a significant main effect of AMB/UNAMB at Region 5 ($F(1,31) = 253.47$, $p < .01$), with ambiguous sentences inducing more errors than unambiguous sentences. There was no significant effect of interaction on a 3-way ANOVA (AMB/UNAMB by THE/ONLY by ADJ/NOADJ).

Finally, we carried out two planned comparisons on the reaction time and error data between the "Only-amb" test sentences and an average based on the four versions of unambiguous control sentences. With mean reaction time as the dependent measure, the difference between the "Only-amb" sentences and the unambiguous controls approached significance (t $(1) = 4.04$; $p < .06$). However, there was a robust difference between them when error rate was the dependent measure (t $(1) = 30.88$; $p < .01$). This result brings out the fact that while subjects responded with greater accuracy and read the "Only-amb" sentences faster than other versions of the ambiguous sentences, they nevertheless responded less accurately and read these sentences somewhat slower than the unambiguous control sentences. This pattern is predicted by the referential theory, as discussed below.

To summarize, the results of Experiment 1 lend support to the referential theory. Most importantly, there was a significant decrease in reaction times and errors on the "Only-amb" sentences at the disambiguating region 5, as compared to the "The-amb" sentences.

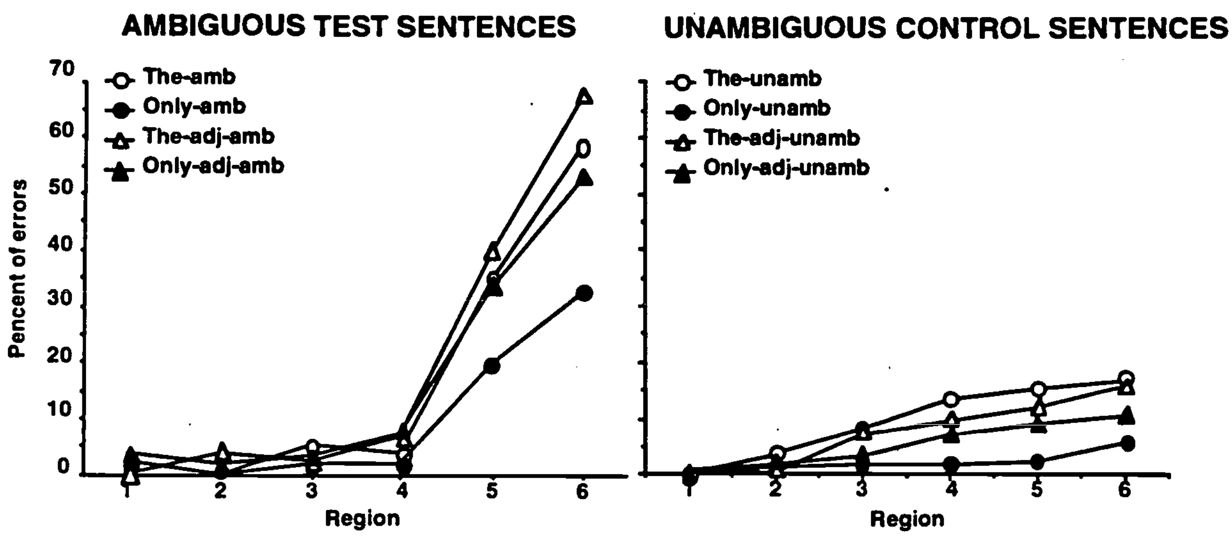


Figure 2. Experiment 1—Percent of errors at each region for ambiguous test sentences (left) and unambiguous control sentences (right)

The requirement posed by ONLY for a contrast set apparently directed the parser to opt for the reduced relative clause alternative within the ambiguous region, hence the main verb came as no surprise—a potential garden path effect was averted. By contrast, in "The-amb" sentences, the main verb analysis was generally chosen in the ambiguous region, resulting in a severe garden path effect at the disambiguating region. There was also a marked difference between "Only-amb" sentences and "Only-adj-amb" sentences, with the latter, but not the former, showing a garden path effect at the disambiguating region. This suggests that the appearance of an adjective satisfied the requirement for a contrast set and, accordingly, the parser preferred to analyze the first verb phrase as the main verb, leading to the garden path effect.

It remains to comment on the comparison between the "Only-amb" sentences and the unambiguous control sentences. As we saw, compared with unambiguous controls, there was a slight elevation in reaction times for the "Only-amb" sentences at the disambiguating region; these sentences also produced significantly more errors. This, we suggest, is chiefly a consequence of parallel processing. When an input string is ambiguous, albeit briefly, more than one analysis is entertained. Although the principles of the referential theory direct the parser in its decisions, even a dispreferred interpretation will be accepted by some proportion of the subjects. The roughly 20% increase in errors for the "Only-amb" sentences, as compared to the unambiguous control sentences, suggests that while the presence of ONLY sufficiently promoted the reduced relative clause analysis in the majority of cases, the main clause reading was not ruled out on every occasion. This resulted in some elevation in reaction times and errors, as compared to the unambiguous controls.

The findings of the present experiment would not be expected on a serial processing model. If parsing strategies such as Minimal Attachment was always applied first, then a lexical substitution in the initial noun phrase would not affect the structural analysis at the onset of an ambiguity, because at that point, each of the ambiguous sentences has the same structure. As noted earlier, however, on the model proposed by Mitchell et al. (1992), relevant semantic information could be used very rapidly to override a brief initial misanalysis within the ambiguous region. This could explain the decrease in reaction times on the "Only-amb" sentences at Region 5 (the so-called disambiguating region). Note that if reanalysis is performed, it would occur following a garden path effect, however brief it might be. Although the findings of the present experiment do not indicate the existence of a garden path effect in the ambiguous region of the "Only-amb" sentences, in contrast to the "The-amb" sentences, it is conceivable that non-structural information was used within the ambiguous region to override a brief and mild garden path effect, but that the word-by-word reading measure used in this experiment was insufficiently sensitive to detect such an effect.

By its nature, the self-paced word-by-word reading technique has an inherent limitation. Because a decision is called for at every word, reading speed is slowed to levels far below normal reading rates. Therefore, measures of word-by-word reading may include the cumulative influences of a number of factors, and some of these factors may be used by the parser earlier than others. It has been suggested repeatedly, on the grounds such as these, that measures of word-by-word reading may be insensitive to the exact timing of the availability and application of different sources of information at the potentially most informative points in sentence processing, and that measures of eye movements may afford greater precision[9] (e.g., Rayner, 1993; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989; Rayner & Morris, 1991; but see Ferreira and Clifton, 1986, for evidence that eye movement results often confirm findings from word-by-word experiments). Based on these considerations, Experiment 2 repeated Experiment 1, using the technique of recording subjects' eye movements during reading.

EXPERIMENT 2

As noted in the Introduction, the influence of non-syntactic information in sentence interpretation is not in dispute. At issue, however, is the exact time course of the application of these sources of information in on-line sentence processing. A central question is whether non-syntactic information is available and used immediately to resolve ambiguities, or whether structurally-based strategies suffice to explain the parser's early decision-making. To answer this question, a more time-sensitive measure of parsing is required than that afforded by the word-by-word reading method. Accordingly, in Experiment 2, we adopted the technique of eye movement recording. Tracking subjects' eye movements while they are reading arguably affords the required precision. First, this technique permits normal, uninterrupted reading. Secondly, it permits the experimenter to identify specific fixation locations in the line of print. When reading materials are analyzed by region, we can ascertain not only how long the subject's eyes remain in a region when that region is read for the first time, but also how often a regressive eye movement is initiated from that region and where it lands.

The capability to examine both first pass reading and the incidence of regressive eye movements is important in addressing the question of *when* different sources of information are used by the parser. It is our working assumption that first pass fixations are most indicative of on-line processes; that is, they indicate the influence of information that is immediately accessed by the reader. We will assume that regressions are not indicative of on-line processing because they occur only sporadically in normal reading. The disparate patterns of regression on different types of sentences are therefore informative. Frequent regressions may signal difficulties that lead the parser to reprocess earlier material. Exploiting the reciprocity between reading times and regressions, we can use eye-movement tracking as an aid to infer which sources of linguistic information are used immediately, and which sources are used somewhat later in processing.

The object of Experiment 2 was to find out whether the information contributed by the focus operator ONLY is used in first pass reading. In this experiment, subjects' eye movements were monitored while they read test materials similar to those used in Experiment 1. The predictions by the referential theory are straightforward: 1), the parser should follow semantic/referential principles in constructing discourse representations, and 2), garden path effects should be modulated by the presence or absence of the focus operator ONLY.

Method

Subjects. Twenty-two undergraduate students participated in the experiment. All were native speakers of English. All reported normal vision or normal vision with soft contact lenses.

Materials. Twenty-four test sentences and sixteen controls were randomly selected from materials used in Experiment 1. Each sentence had two versions; the sole difference between them was the alternation between THE and ONLY in the initial NP. Some minor revisions were made to the test materials so that none of the sentences exceeded 76 characters. This enabled us to present each sentence on a single line beginning at the left margin of center screen. A sample set of test sentences and their corresponding controls is as follows:

The-amb	*The businessmen loaned money at low interest were told to record expenses.*
Only-amb	*Only businessmen loaned money at low interest were told to record expenses.*
The-unamb	*The vans stolen from the parking lot were found in a back alley.*
Only-unamb	*Only vans stolen from the parking lot were found in a back alley.*

A counterbalanced design was used, with test and control sentences evenly distributed in each of the two experimental conditions, which contained either a pre-nominal THE or ONLY. No subject read the same sentence with THE and with ONLY, and all test and control sentences occurred in either condition over the two stimulus lists. Eleven subjects were tested on each list in which test and control sentences were intermixed among 60 fillers in a pseudo-random fashion, such that each test or control sentence was followed by at least one filler. Each stimulus list was divided into two halves containing an equal number of test and control sentences, which were presented in separate sessions, preceded by 10 warm-up trials.

Equipment. Subjects' eye movements were recorded using the IRIS infrared-light eye-movement system, (SKALAR model 6500). The IRIS system uses a differential reflection method of eye-movement recording. In this technique, infrared-emitting diodes and infrared sensitive detectors are positioned in front of the eye so that their receptive fields match the iris-sclera boundary, both on the nasal side and on the temporal side. Upon horizontal rotation of the eye, the nasally positioned detector measures an increase in scleral infrared reflection, while the detector on the temporal side measures a decrease in infrared reflection. Subtraction of the signals from nasal and temporal detectors gives eye position relative to head position. Eye position is sampled every millisecond by a computer equipped with an analog-to-digital conversion board. Each eye fixation is represented by an x and y screen coordinate, a starting time, and an ending time. Eye movements are recorded from the right eye only, but viewing is binocular. Stimuli are displayed on a 13-inch High Resolution RGB monitor set 64 centimeters from the subject's eyes. In our test materials (mixed case in Courier 14-point font) the visual angle of each character was slightly greater than 12 min of arc, permitting a resolution of less than one character width.

Procedure. Subjects were given written instructions that contained a brief description of the eye-movement monitoring technique. Since eye positions are determined relative to the head position, head movements were kept to a minimum. Head stabilization was achieved by a bite-bar and a forehead rest. After the eye tracker was calibrated, the subject was told to begin reading sentences from the starting point at the center left of the screen, where a fixation cross was presented before each sentence. The experimenter emphasized that the sentence should be read at the subject's normal rate. When the subject finished reading the sentence, he/she pressed on the mouse and the sentence was erased. On one third of the filler trials, a comprehension question appeared on the screen after a sentence disappeared. The subject answered the question by moving a mouse-controlled arrow to a "YES" box or "NO" box on the screen and clicking on it. Feedback was given by the computer, informing the subject whether or not the answer was correct. After each trial, there was a calibration-checking routine, at which time the subject fixated consecutively on five coordinates on the screen. Adjustments of signal strength were occasionally performed, but recalibration was rarely needed. A pretest containing six sentences was conducted.

Results and Discussion

We present analyses based on the recorded eye fixation durations and percent of regressions. An eye fixation was recorded if a subject's eye dwelled upon a character for 8 milliseconds or longer. Total fixation duration in milliseconds was computed for each scoring region of the test and control sentences. Incidence of regressions from these regions was also recorded. For the purpose of data analysis, test and control sentences were divided into the same six regions as those described in Experiment 1: Region 1 contained the subject NP. Region 2 included the first verb, which was morphologically ambiguous for the test sentences. Region 3 contained the remainder of the first verb phrase except the last word, which was the sole content of Region 4. Region 5 contained either an auxiliary verb and the main verb, or the main verb and the following word. The final region consisted of the remainder of the sentence.

Planned factorial comparisons at each region were carried out by means of two-way ANOVAs, testing the two types of pre-nominal word (THE/ONLY) and the two types of initial verb phrase (AMB/UNAMB). The measure of first pass reading time included fixation durations by subjects who read a certain region of the sentence for the first time, provided that they had not read beyond that region. The dependent variable used in the analyses was residual reading times (RRT).[10] RRT was calculated by conducting a regression analysis for each subject, using the length of each region (the number of letters and spaces) as the independent variable and the total duration of all eye fixations at each region as the dependent variable. This measure statistically removes the length of a region as a factor. The incidence of regressive eye movements was based on the percent of subjects' first pass readings of a region that ended in a leftward regression to a portion of the sentence that had either been visited before or skipped.

First pass reading times: Figure 3 presents a profile of mean first pass residual reading times (RRT) at each region for the test sentences and controls. A significant main effect of THE/ONLY was found at Region 1, where reading times on sentences beginning with ONLY were significantly greater than those beginning with THE ($F(1,21) = 6.72$ $p < .02$). No significant main effect of AMB/UNAMB was found, nor was the interaction significant between THE/ONLY and AMB/UNAMB. It is clear that the focus operator ONLY had an impact on the parser such that it induced longer reading times than did the definite determiner THE.

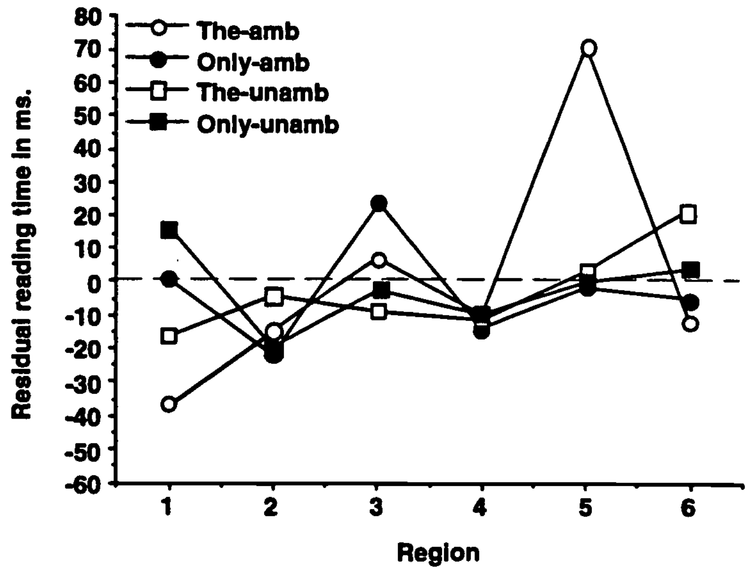


Figure 3. **Experiment 2—Mean first pass residual reading time (RRT) at each region.**

The only significant effect in the ambiguous region (Regions 2-4) was a main effect of AMB/UNAMB at Region 3, where reading times on the ambiguous test sentences were longer than those on the unambiguous controls ($F(1,21) = 6.17$ $p < .02$). Ambiguous sentences apparently presented more difficulties than unambiguous sentences within the ambiguous region. The convergence of reading times on all versions of sentences at Region 4 suggests that resolution of the ambiguity was reached by the end of the ambiguous region.

At Region 5, where the main verb appeared, there was a main effect of THE/ONLY: reading times for sentences beginning with THE were significantly longer than for those beginning with ONLY: ($F(1,21) = 23.14$, $p < .01$). There was also a significant main effect of AMB/UNAMB: ambiguous test sentences took significantly longer to read than unambiguous controls ($F(1,21) = 9.91$, $p < .01$). Moreover, there was a significant interaction of THE/ONLY by AMB/UNAMB ($F(1,21) = 7.05$, $p < .02$). A pairwise comparison showed that reading times for test sentences with THE ("The-amb") were significantly greater than those for the "Only-amb" sentences ($F_1(1,21) = 21.83$ $p < .01$; $F_2(1,23) = 13.65$, $p < .01$). On the other hand, no significant difference was present between the two versions of the unambiguous controls ($p > .1$). Reading times did not differ between the "Only-amb" version and either of the two versions of the controls ($p > .1$).

Reading times at Region 5 suggest that a garden path effect occurred for test sentences beginning with THE: Long fixation durations in that region indicated that subjects were apparently surprised to see that the material (the main verb) could not be incorporated into the analysis they had adopted, and as a result, they were forced to pause. For sentences that began with ONLY, on the other hand, no significant rise in fixation durations was recorded, suggesting that the main verb was expected, in keeping with the cases of the unambiguous controls. This pattern leads us to suppose that the reduced relative clause analysis was adopted in the ambiguous test sentences beginning with ONLY. Considering the fact that these are first pass fixation times, the information carried by ONLY must have been used extremely rapidly. However, in order to conclude that the processing of "Only-amb" sentences is genuinely different from that of "The-amb" sentences, subjects' eye-regression patterns must also be considered.[11]

Incidence of regression: Two-way ANOVAs were performed on the percent of first pass readings that resulted in regressive eye movements. Inspection of the percent of regressions at each region revealed a significant main effect of AMB/UNAMB at Region 5 ($F(1,21) = 10.05, p < .01$) and at Region 6 ($F(1,21) = 8.73$, $p < .01$). In each case, ambiguous test sentences induced more regressions than unambiguous controls. There was no main effect of THE/ONLY, nor was there an interaction at any region. Figure 4 displays the pattern of regressions for each of the four versions of sentences at each region.

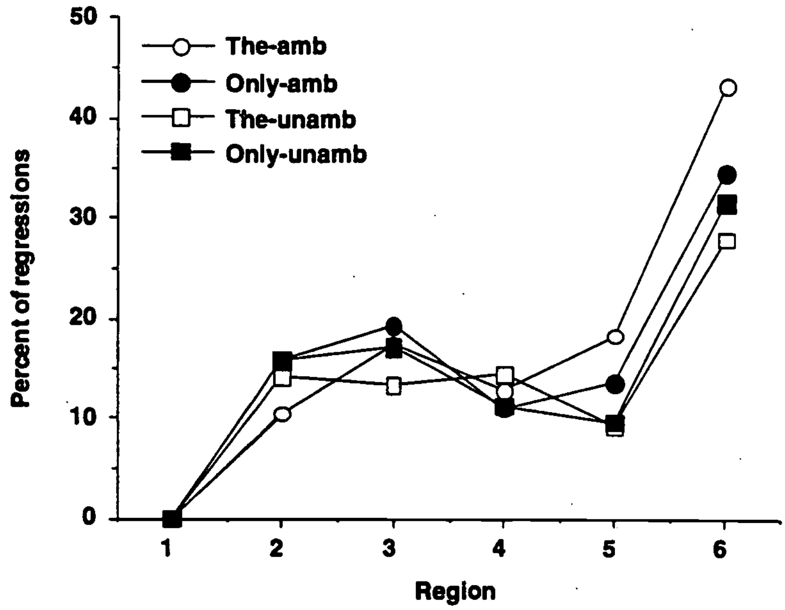


Figure 4. **Experiment 2—Percent of regressions from each region.**

The ambiguous test sentences were found to induce more regressions than their unambiguous controls, despite the fact that first pass reading times were significantly different between the two types of ambiguous test sentences at Region 5. This is reminiscent of the results of Experiment 1, where reaction times showed a significant difference between "The-amb" and "Only-amb" sentences at the point of disambiguation, but both versions showed a higher error rate than the unambiguous controls.

In light of the fact that ambiguous sentences induced more regressions than unambiguous ones, we also analyzed the data from first pass reading times by creating a new dataset that excluded any trial on which there was a regression. That is, we conducted a regression-contingent analysis of first pass reading times at Region 5 (See Altmann et al., 1992, for a detailed discussion of this method). The purpose of this analysis was to determine whether a difference in reading times persisted at Region 5 between "The-amb" sentences and "Only-amb" sentences using an uncontaminated measure of first pass reading. Shorter reading times in a region sometimes result when there are many regressions from that region. In the present analysis, we wanted to ensure that this was not the source of the relatively faster reading times at Region 5 on the "Only-amb" sentences. We expect, on the referential theory, that the difference in processing ambiguous sentences beginning with THE vs. those beginning with ONLY should be present *in the absence of a regression*.

The results of the regression-contingent analysis closely resembled the reading time patterns reported before: There was a main effect of THE/ONLY ($F(1,21) = 15.39$, $p < .01$), as well as AMB/UNAMB ($F(1,21) = 7.76$, $p < .02$). The effect of THE/ONLY by AMB/UNAMB interaction approached significance ($F(1,21) = 3.55$, $p < .08$). A comparison between sentence versions showed that reading times on "The-amb" sentences were significantly longer than those on the "Only-amb" sentences ($F_1(1,21) = 12.15$, $p < .01$; $F_2(1,23) = 7.54$, $p < .02$). On the other hand, there was no difference between the two kinds of controls ($p > .1$). No difference existed between test sentences beginning with ONLY and either of the two kinds of unambiguous controls ($p > .1$).

In sum, the results from Experiment 2 confirm the main findings of Experiment 1: they support the contention of the referential theory that referential effects occur on-line in ambiguity resolution. The referential information carried by the pre-nominal focus operator ONLY strongly influenced the parser's initial analysis of the ambiguity, as attested by the eye-movement patterns, especially the significant interaction between THE/ONLY and AMB/UNAMB at the disambiguating region on the reading time measure. First pass reading times in the absence of regressions showed the same result. These findings are as predicted by the referential theory. The theory also allows that, because more than one analysis is entertained for ambiguous sentences, these sentences prove to be more difficult than their unambiguous counterparts; this is shown by a significantly higher rate of regressions in the former than in the latter.

In discussing findings of Experiment 1, we noted the difficulty in distinguishing predictions of the referential theory from predictions of the recently revised structurally-based garden path model (Mitchell et al., 1992). According to this account, as we saw, reanalyses can be implemented early, using semantic and/or discourse information: If non-syntactic sources of information are delayed only briefly, then a Minimal Attachment analysis could be rapidly overridden within the ambiguous region in cases like the "Only-amb" sentences of the present experiment. As a result, no garden path effect would be expected to occur when the structural disambiguating material is encountered later. Because Experiment 1 used a word-by-word measure of reading, it may have lacked sufficient sensitivity to evaluate this possibility. With the measurement of eye movements, however, we are in a better position to reconstruct the pattern of processing within the ambiguous region. In fact, there was an apparent elevation of reading times for ambiguous sentences in this region.[12] The referential theory contends that this elevation is a consequence of parallel processing (see discussion of Experiment 1). The effect within the ambiguous region could be reconciled with the model proposed by Mitchell et al. (1992), but only with the addition of two assumptions. One assumption is that reanalysis can be triggered and completed within the ambiguous region. The second assumption is that there are two distinct kinds of garden path effects: One kind that is sensitive to structural information and is costly of processing resources, and another kind that is responsive to non-syntactic sources of information and is cost free for reanalysis. Although such a distinction is possible in principle, it would have to be motivated on independent grounds. In addition, some empirical way of distinguishing a costly garden path effect from a cost-free reanalysis would be required (cf. Fodor & Inoue, 1994; Frazier, 1994).

The two experiments we have presented so far converge on the same conclusion: garden path effects can be modulated by referential factors within the test sentences. The results from eye-movement recording closely parallel those found with word-by-word self-paced reading. As noted earlier, we employed within-sentence manipulations in the present study to enable us to examine referential effects in resolving sentences containing a main-verb/reduced-relative-clause ambiguity, a structure that has proven to be relatively impervious to extra-sentential context (Ferreira & Clifton, 1986; Murray & Liversedge, 1994; but also see Spivey-Knowlton & Tanenhaus, 1994, for evidence of the influence of referential context for this construction). We focused on the referential contributions of pre-nominal modifiers (ONLY versus THE) within a test sentence while holding constant factors such as the frequency of a verb (with an "*-ed*" ending) being used as a past participle. Therefore, although the findings are entirely consistent with the constraint satisfaction model, it is important to note that the results were predicted by the principles of the referential theory.

In Experiments 3 and 4, we pursue a related issue: the use of general world knowledge (plausibility) in on-line sentence processing. The focus of these experiments is on the locus of attachment of prepositional phrases in structurally ambiguous sentences. In Experiment 3, we ask whether plausibility is used to resolve structural ambiguities. Experiment 4 investigates the time course of the availability and application of this source of information, using the eye-movement recording technique.

EXPERIMENT 3

The parser's preference in resolving local ambiguities involving the site of attachment of prepositional phrases has received much discussion. As the examples below illustrate, one option in the sentences under consideration is to attach a prepositional phrase (PP) to the preceding verb phrase. This will be referred to as the VP-attachment analysis. A second option is to attach the prepositional phrase to the immediately preceding noun phrase. This is referred to as the NP-attachment analysis. The following examples show that while the PP *with new brushes* in (A) can only be attached to the verb to make the sentence felicitous, the PP *with large cracks* in (B) can only be used to modify the immediately preceding noun.

(A) *The man [painted the doors [with new brushes]] before the festival.* VP-attachment
(B) *The man [painted [the doors with large cracks]] before the festival.* NP-attachment

It has been demonstrated repeatedly that subjects find sentences like (B) more difficult than those like (A) (e.g., Altmann & Steedman, 1988; Britt et al., 1992; Ferreira & Clifton, 1986; Perfetti, 1991; Rayner et al., 1983, 1992).[13] Various explanations have been advanced to explain this difference. According to the referential theory, both VP-attachment and NP-attachment analyses are computed by the syntax at the onset of the preposition *with*, and both are evaluated by the semantic processor in a word by word fashion. The parser pursues the analysis that best fits the context (e.g., Altmann & Steedman, 1988). In the absence of prior linguistic context, however, a mental model of the discourse is set up which includes, for examples (A) and (B) above, a set of doors as required by the definite noun phrase, *the doors*. Having augmented the mental model with a set of doors, the VP-attachment analysis of the following PP is pursued. The alternative NP-attachment analysis requires the parser to further modify the mental model by distinguishing a subset of doors with specifications from other doors. The principle of parsimony therefore predicts that this analysis should be dispreferred because it requires more extensions to the mental model than the VP-attachment analysis. The pursuance for the VP-attachment analysis results in a temporary anomaly in sentences like (B), since *cracks* are not things that one can use to *paint* with, and reanalysis is instigated, leading to increased reading difficulty.

However, if the noun *doors* is preceded by ONLY, such as in (C) and (D), then the referential theory predicts a reversal in parsing preferences, namely, the NP-attachment analysis will be preferred, not the VP-attachment one:

(C) *The man [painted only doors [with new brushes]] before the festival.* VP-attachment
(D) *The man [painted [only doors with large cracks]] before the festival.* NP-attachment

On the referential theory, the presence of the pre-nominal focus operator ONLY invites the parser to assume the existence of a set of entities that contrasts with those referred to by the noun. The most parsimonious way to construct a contrast set is to divide an existing set into subsets. Pursuing this option, an NP-attachment analysis provides the needed information for a contrast set, namely, a specific set of doors (with cracks). As a consequence, sentences like (D) will be easy to process, but those like (C) will induce a temporary anomaly because it is infelicitous to modify *doors* with *brushes*, and reanalysis is required. Experiment 3 was designed to test these predictions, using a word-by-word reading paradigm.

Method

Subjects. Forty-four undergraduate students participated in the experiment. All were native speakers of English and were naive as to the purposes of the experiment. These subjects were randomly assigned to four groups.

Materials. The experiment included 20 sets of test sentences in each of four versions: VP-attachment sentences with THE ("The-VP") or with ONLY ("Only-VP") and NP-attachment sentences with THE ("The-NP") or with ONLY ("Only-NP"), as shown in the examples below. A full list of test sentences can be found in Appendix B. Forty filler sentences were interspersed among the test sentences. Four lists of stimuli were constructed. VP-attachment sentences were rotated through two lists, and NP-attachment sentences were rotated through the other two lists.[14] Each list was tested on a different group of 11 subjects.

The-VP	*The man painted the doors with new brushes before the festival.*
The-NP	*The man painted the doors with large cracks before the festival.*
Only-VP	*The man painted only doors with new brushes before the festival.*
Only-NP	*The man painted only doors with large cracks before the festival.*

Procedure. Subjects read sentences displayed on a CRT one word at a time, and the words remained on the screen until the sentence ended. The instructions given to the subjects were similar to those given in Experiment 1, except for one change. Since both VP-attachment and NP-attachment were grammatical, subjects were asked to decide whether or not the sentence continued "to make sense" as each consecutive word appeared (See Boland et al, 1990 for a more detailed discussion of this "stop making sense" task). The computer recorded the duration (in milliseconds) between the onset of each new word and the subject's key press. Subjects' responses ("YES" or "NO") were also recorded for each word. A pretest was conducted that contained eight example sentences.

Results and Discussion

Two-way ANOVAs were carried out, testing the effects of the two types of pre-nominal modifiers (THE/ONLY) and the two sites of attachment (VP/NP). The dependent variables were mean reaction time (RT) per word and percent of errors. Mean RT included the time subjects took to correctly accept each newly presented word to be a sensible continuation of the ongoing sentence fragment. A "NO" response to any word in a test sentence was interpreted as indicating that the subject erroneously deemed the sentence nonsensical. Reaction times on any sentence in which an error occurred were excluded from the analyses. Error analyses included all the responses from the subjects.

For the purpose of conducting statistical analyses, the test sentences were divided into five regions.[15] Region 1 contained the subject NP *The man*. Region 2 contained the main verb *painted*. Region 3 contained the object noun phrase that was preceded by either the definite determiner THE or the focus operator ONLY, and followed by the preposition *the/only doors with*. Region 4 contained the object NP of the prepositional phrase, the content of which either confirmed or disconfirmed a particular attachment *new brushes/large cracks*. Region 5 contained the remainder of sentence minus the last word *before the*. We report results from all five regions, focusing on Region 4: At this region, the referential theory predicts an interaction of THE/ONLY by VP/NP.

Analyses of reaction times revealed no significant main effect or interaction at either Region 1 or Region 2, as expected, since all the test sentences were identical in these regions. At Region 3, where the sentences diverged as to whether the definite determiner THE or the focus operator ONLY preceded the object noun, there was a significant main effect of THE/ONLY ($F_1(1,43) = 10.76$, $p < .01$; $F_2\ (1,19) = 5.64$, $p < .03$): Reaction times on sentences with ONLY were longer than those with THE. At Region 4, there was a significant interaction between THE/ONLY and VP/NP ($F(1,19) = 13.47, p < .01$). A similar pattern existed at Region 5, where the THE/ONLY by VP/NP interaction was significant ($F(1,19) = 18.99$, $p < .01$). Because the design of this experiment used VP/NP as a between-subjects variable, the effects of VP/NP (and THE/ONLY by VP/NP interactions) are calculated as analysis by items only. Figure 5 depicts the mean reaction times at each region.

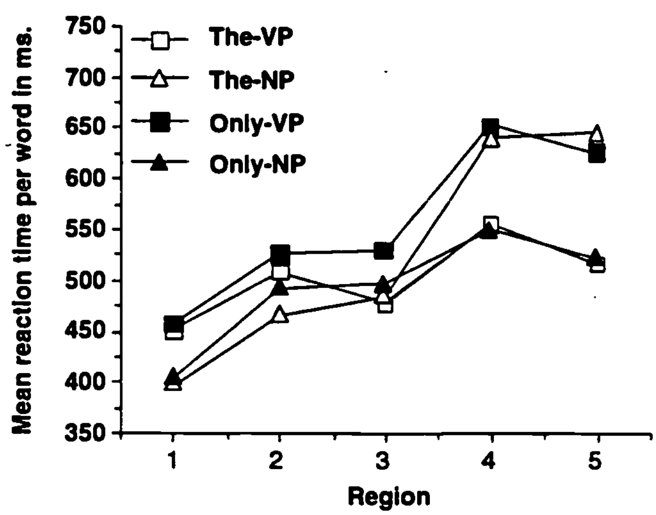


Figure 5. Experiment 3—Mean reaction time per word at each region.

Region 4 merits further analysis, because it contains the pragmatic information that either confirmed or disconfirmed the subjects' earlier parsing decision. A pairwise comparison between sentence versions revealed that, as expected, reaction times on "The-NP" sentences were longer than those on "The-VP" sentences, although the effect only approached significance ($F(1,19) = 3.71$, $p < .07$). A reversal occurred between the two versions with ONLY: Reaction times to the "Only-NP" sentences were much shorter than those to the "Only-VP" sentences ($F(1,19) = 9.56$, $p < .01$). Reaction times to the "Only-VP" version were longer than those to the "The-VP" version ($F_1(1,21) = 8.27$, $p < .01$; $F_2(1,19) = 8.85$, $p < .01$); Reaction times to the "Only-NP" version, on the other hand, were shorter than those to the "The-NP" version ($F_1(1,21) = 6.82$, $p < .02$; $F_2(1,19) = 4.36$, $p < .05$).

Figure 6 depicts the mean error rate at each region. The error count across all regions revealed a highly significant interaction of THE/ONLY by VP/NP ($F(1,19) = 126.47$, $p < .01$). The effect occurred mainly at regions 4 and 5. ANOVAs carried out on combined scores at these regions revealed a significant main effect of THE/ONLY ($F_1(1,43) = 39.56$, $p < .01$; $F_2(1,19) = 38.32$, $p < .01$). The effect of VP/NP approached significance ($F(1,19) = 3.42$ $p < .08$). The interaction between the two factors was significant ($F(1,19) = 39.87$, $p < .01$). A pairwise comparison showed that "Only-VP" sentences induced significantly more errors than did "Only-NP" sentences ($F(1,19) = 18.33$, $p < .01$). The difference between "The-VP" and "The-NP" sentences approached significance ($F_2(1,19) = 4.02$, $p < .06$), with "The-NP" sentences inducing more errors. "Only-VP" sentences induced significantly more errors than "The-VP" sentences ($F_1(1,21) = 27.69$, $p < .01$; $F_2(1,19) = 69.79$, $p < .01$). There was no difference between the "The-NP" and the "Only-NP" versions.

The results of reaction times and error rates confirmed the predictions of the referential theory. According to the referential theory, both VP-attachment and NP-attachment analyses are computed when the prepositional phrase is encountered. The focus operator ONLY makes the parser anticipate a modification of the noun phrase, opting for the NP-attachment analysis. The anticipation is met when the NP *large cracks* is encountered, because the prepositional phrase is a plausible modifier of the object NP *doors*. On the other hand, the anticipation of an NP modifier in "Only-VP" sentences leads the parser to encounter the noun phrase *new brushes*, which makes the NP-attachment analysis of the PP anomalous. Reanalysis is therefore instigated, resulting in a significant elevation of reaction times, as well as increased erroneous rejections of the sentence.

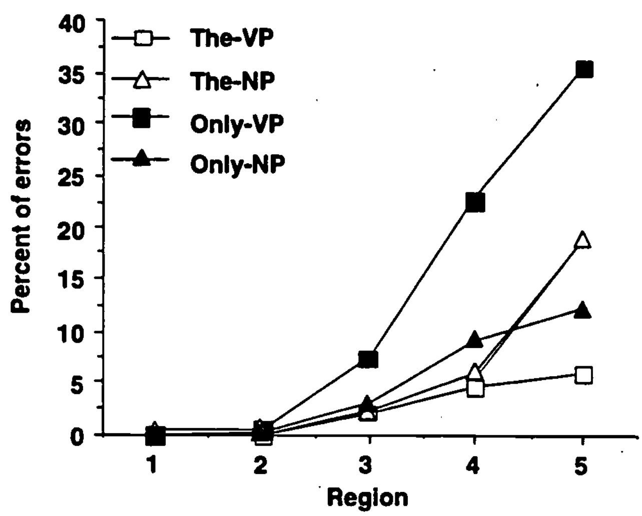


Figure 6. Experiment 3—Percent of errors at each region.

The findings of Experiment 3 indicate that a simple substitution between THE and ONLY in a noun phrase followed by a prepositional phrase can change the parser's immediate decision as to how a prepositional phrase should be attached. It was found in Experiments 1 and 2 that the semantic information carried by the focus operator ONLY is used on-line, establishing an immediate preference for an analysis that sets up a contrast set. The obtained effects in this experiment hinge on the added assumption that subjects use *a priori* plausibility, elicited by the noun phrase of the prepositional phrase, as their criterion for PP attachment. The disambiguating factor here is real-world knowledge. It is clear from Experiment 3 that information governing the *a priori* plausibility of the alternative representations of an ambiguous phrase is used quickly in reaching decisions about PP attachment. The manner in which such information is used by the parser is assessed with greater precision in Experiment 4, which uses the eye-movement recording technique.

EXPERIMENT 4

The eye-movement recording methodology offers a gain in measurement precision. With this technique, we are in better position to detect subtle garden path effects, or to detect the speedy recovery from a misanalysis involving semantic anomaly. This methodology also has the advantage that it can reveal the time course of the availability and use of different sources of information during on-line sentence processing. The issue of timing is at the crux of current research on sentence processing. If some sources of information are used earlier than other sources, this circumstance could be used to decide between competing models of ambiguity resolution (cf. Fodor, et al., in press). As we saw, it is a basic tenet of the referential theory that information about specific conversational context takes precedence over the use of real world knowledge. In this experiment, we ask whether the query of the store of world knowledge can keep pace with the assimilation of semantic information contributed by the focus operator ONLY.

A third issue we address in this experiment concerns the origin of individual differences in profiles of sentence processing. Previous research by MacDonald et al. (1992) has established that individual differences in working memory capacity constrain the ability of a subject to process ambiguous sentences. The authors interpret this finding as evidence that subjects with high memory spans can maintain multiple syntactic representations for ambiguous sentences, whereas subjects whose memory capacities are more limited can maintain only the representation that is most frequently used. Pearlmutter and MacDonald (1995) also found that high span subjects were more sensitive to the relative plausibility of alternative representations of an ambiguous phrase. On the basis of these findings, we are invited to infer that in ambiguity resolution, subjects with higher memory capacity may be more efficient in using diverse sources of information than those with more limited memory capacity. Arguably, rapid decision-making is facilitated by maintaining alternative representations in memory, based on whatever information the parser has at its disposal. Therefore, individual differences in working memory may be correlated with differences in the time at which various sources of information in a sentence are used. A subject who finds it difficult to maintain alternative representations of an ambiguous sentence will have difficulties in resolving the ambiguity, especially if late-arriving information is critical to the decision. This subject will be expected to "look back" more frequently in reading. As noted earlier, it can be inferred from the principles of the referential theory that persons with limited memory capacity will have greater difficulty than those with higher-memory capacity in resolving ambiguities that require information about the relative plausibility of the alternative meanings of a sentence.

The test materials used in this experiment were the same as those used in Experiment 3. An example set is repeated here:

The-VP	*The man painted the doors with new brushes before the festival.*
The-NP	*The man painted the doors with large cracks before the festival.*
Only-VP	*The man painted* only *doors with new brushes before the festival.*
Only-NP	*The man painted* only *doors with large cracks before the festival.*

The predictions by the referential theory remains the same as those made for Experiment 3: While "The-VP" sentences will not cause processing difficulty and "The-NP" sentences will, a reversal effect should occur for sentences with ONLY. "Only-VP" sentences will prove difficult to read, because the focus operator ONLY will lead the parser to favor an NP-attachment analysis, resulting in an anomaly at the noun phrase of the prepositional phrase. On the other hand, the NP-attachment analysis of the PP is anticipated for "Only-NP" sentences, and no effect of anomaly should occur.

Method

Subjects. Thirty-two undergraduate students participate in the experiment. All were native speakers of English. They were not informed of the purpose of the experiment. Subjects were selected who had uncorrected vision or wore soft contact lenses.

Materials. The same 20 sets of test sentences and 60 fillers used in Experiment 3 were used in this experiment. Four stimulus lists were generated. The four versions of each set of test sentences (The-VP, The-NP, Only-VP, Only-NP) were rotated through the four lists. Each list was tested on a different group of eight subjects. This experiment used a fully factorial repeated measure design, with both THE/ONLY and VP/NP as within-subjects and within-items factors. Eight warm-up sentences preceded each stimulus list.

Equipment. The equipment and data analysis programs used in this experiment were the same as those employed in Experiment 2.

Procedure. The procedure was as in Experiment 2. In addition, a memory span test was administered following the reading test.[16] In the memory span test, subjects listened to recorded materials over headphones. Their task was to report in order the last word of each sentence in a set of spoken sentences when prompted by a nonverbal signal. They were also required to judge, after each sentence within a set, whether it was "True" or "False." The number of sentences in each set was gradually increased from two to five. Subjects were encouraged to say as many terminal words as they could, even if they were unsure about their order of occurrence. A total of 42 words was solicited, and subjects were ranked according to the number of words they correctly reported. Based on span length, subjects were divided into two groups of 16 by a median split. High Span versus Low Span was treated as a separate factor in the statistical analysis.

Results and Discussion

The analyses were based on recorded eye fixations at five regions of the test sentences, as defined in Experiment 3: Region 1 was the subject noun phrase; Region 2 contained the main verb; Region 3 contained the object noun phrase (preceded either by THE or ONLY) and the following preposition; Region 4 contained the object noun phrase of the prepositional phrase. Region 5 contained the rest of the sentence. Region 4 was the focus of interest, because attachment preferences are contingent upon the processing of the semantic content of this region.

Two-way ANOVAs were performed, testing the two types of pre-nominal word (THE/ONLY) and the two types of attachments (VP/NP), with first pass residual reading times (RRT) and percent of regressions as the dependent variables. As was discussed in Experiment 2, RRT provides a more adequate measure than total fixation durations, since reading time comparisons are made between sentences that contain different words of varied lengths (especially at Region 4).

First pass reading times. Collapsing across regions, ANOVAs performed on first pass residual reading times revealed a significant main effect of THE/ONLY: Sentences with ONLY yielded significantly longer reading times than those with THE ($F_1(1,31) = 5.51, p < .03$; $F_2(1,19) = 6.33, p < .01$). There was no main effect of VP/NP ($p > .1$), nor was there a significant interaction of THE/ONLY by VP/NP ($p > .1$). Separate ANOVAs were performed on reading times at each region. A profile by region is shown in Figure 7.

No significant effects were found at either Region 1 or Region 2. At Region 3, there was a significant main effect of THE/ONLY ($F_1(1,31) = 10.72, p < .01$; $F_2(1,19) = 10.97, p < .01$).

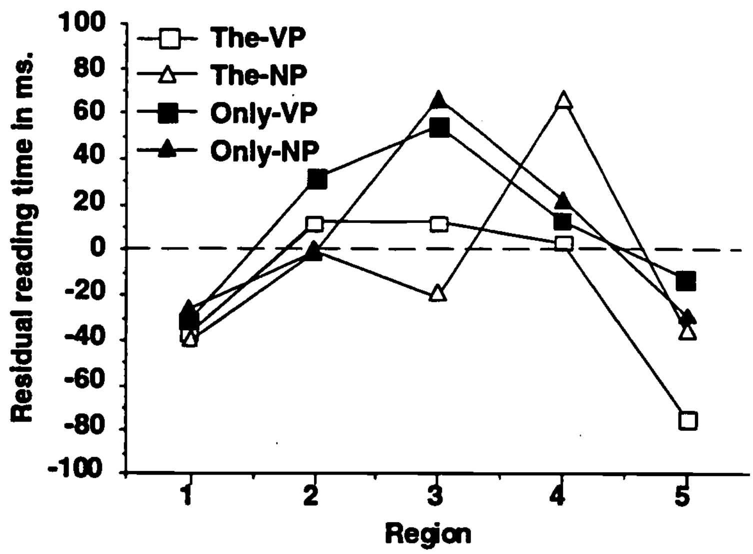


Figure 7. **Experiment 4—Mean first pass residual reading time (RRT) at each region..**

Reading times were significantly longer in sentences with ONLY. At Region 4, the effect of VP/NP approached significance ($F_1(1,31) = 4.09, p < .06; F_2(1,19) = 3.59, p < .08$), and there was no effect of THE/ONLY. The interaction of THE/ONLY by VP/NP was significant in the analysis by subjects ($F_1(1,31) = 5.09, p < .04$), but not in the analysis by items ($p > .1$). A planned comparison between sentence versions at Region 4 revealed that reading times for "The-NP" sentences were significantly longer than those for "The-VP" sentences ($F_1(1,31) = 10.01, p < .01; F_2(1,19) = 7.39, p < .02$). There was no significant difference between the two versions with ONLY ($p > .1$). Region 5 revealed no significant effects.

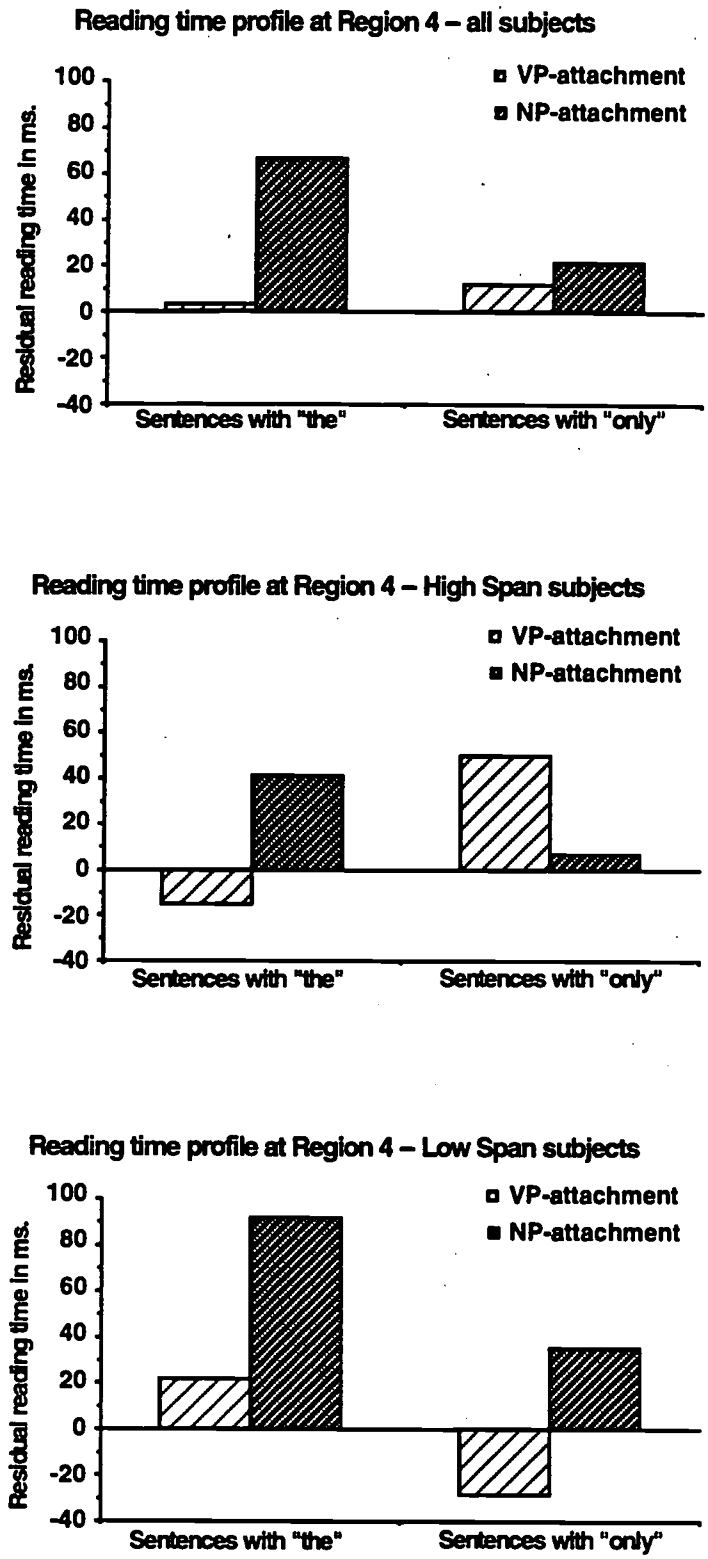
In sum, first pass reading times at Region 4, which is the region of interest, does not seem to meet the specific predictions of the referential theory on sentences with ONLY. Reading times on "Only-NP" sentences were *not* shorter than those on "Only-VP" sentences, as predicted by the referential theory. But neither were reading times for "Only-NP" sentences longer than those for "Only-VP" sentences, which would go against the referential theory.

Since the critical information needed to recover from a misanalysis of PP attachment turns on *a priori* plausibility, its use is expected to be sensitive to individual differences in memory capacity. We therefore turned to the analyses which partitioned subjects according to their scores on the memory span test. As illustrated in Figure 8, there is an apparent discrepancy between the two subject groups on the reading-time profiles at Region 4, compared with the pattern with all subjects combined (A): High Span subjects (B) acted in accordance with the predictions of the referential theory, while Low Span subjects (C) did not.

For High Span subjects, there was a significant interaction of THE/ONLY by VP/NP in the analysis by subjects ($F_1(1,15) = 11.78, p < .01$); the analysis by items approached significance ($F_2(1,19) = 3.78, p < .07$). As expected, a significant delay at Region 4 existed in reading "The-NP" sentences as compared to the reading of "The-VP" sentences ($F_1(1,15) = 6.32, p < .02; F_2(1,19) = 4.32, p < .05$). However, reading times on "Only-VP" sentences were longer than those on "Only-NP" sentences, producing an effect that approached significance in the analysis by subjects ($F_1(1,15) = 3.65, p < .08$), though not in the analysis by items ($p > .1$).

For Low Span subjects, there was a main effect of VP/NP ($F_1(1,15) = 4.61, p < .05; F_2(1,19) = 7.01, p < .02$): NP attachment sentences were read more slowly than VP attachment ones at Region 4. The main effect of THE/ONLY was significant in the analysis by items ($F_2(1,19) = 5.31, p < .04$), and approached significance in the analysis by subjects ($F_1(1,15) = 3.92, p < .07$). The interaction between THE/ONLY and VP/NP was not significant ($p > .1$). Notice, however, that the reading times on the "Only-NP" sentences were about the same as those on the "The-VP" sentences, and both were markedly different from reading times on the "The-NP" sentences.

A



B

C

Figure 8. Experiment 4—Mean first pass residual reading time (RRT) at Region 4 by all subjects (A), High Span subjects (B) and Low Span subjects (C)

The discrepancy between High Span and Low Span subjects on first pass reading time profiles may be attributable to the way these two groups use general world knowledge in reading Region 4. As we indicated, appeal to *a priori* plausibility is required in order to make and revise attachment decisions. It is useful, therefore, to look at incidence of regressive eye movements from Region 4, which may indicate reading difficulties.

Incidence of regressions. As with the profiles of reading times, the pattern of regressions did not yield any systematic effect in the undifferentiated data set, as shown in the region-by-region profile (Figure 9).

However, distinctive patterns of regression did occur for High- and Low-span groups at Region 4, as depicted in Figure 10. For High Span subjects (B), there was no systematic effect: ($p > .1$ in all analyses). These subjects seemed to have recovered from misanalysis in the course of their first pass reading, and presumably because of this, they showed little difference in regression patterns among the different versions of the test sentences. By contrast, Low Span subjects (C) displayed dissociated patterns. There was a main effect of VP/NP ($F_1(1,15) = 9.28, p < .01; F_2(1,19) = 8.41, p < .01$): VP-attachment sentences induced significantly more regressions than NP-attachment ones. There was no main effect of THE/ONLY ($p > .1$), but the interaction between VP/NP and THE/ONLY was significant ($F_1(1,15) = 7.38, p < .02; F_2(1,19) = 5.43, p < .03$). A pairwise comparison revealed that Low Span subjects made significantly more regressions on the "Only-VP" sentences than on the "Only-NP" sentences ($F_1(1,15) = 26.42, p < .01; F_2(1,19) = 11.34, p < .01$).

Here is our explanation of the combined results of first pass reading times and the incidence of regressions. We maintain that the divergent processing patterns by the subject groups resulted from discrepancies in the time course of application of plausibility information associated with individual differences in memory capacity. Let us look first at the sentences with the definite determiner, THE. As expected on the referential theory, it turned out to be relatively easy for both High and Low Span subjects to construct a mental representation for these sentences. Having pursued the referentially simple VP-attachment analysis of both "The-VP" and "The-NP" sentences, both groups of subjects detected the implausibility of the noun phrase *large cracks* as a modifier of the verb in "The-NP" sentences. Reanalysis was therefore initiated, triggering the long first pass reading times. That the relatively simple recovery process was generally accomplished on-line is also attested by the relatively low incidence of regressions by either High Span or Low Span subjects.

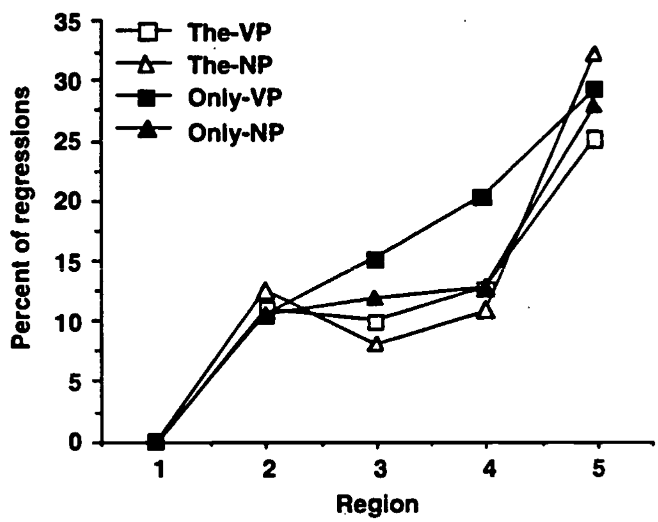


Figure 9. Experiment 4—Percent of regressions at each region.

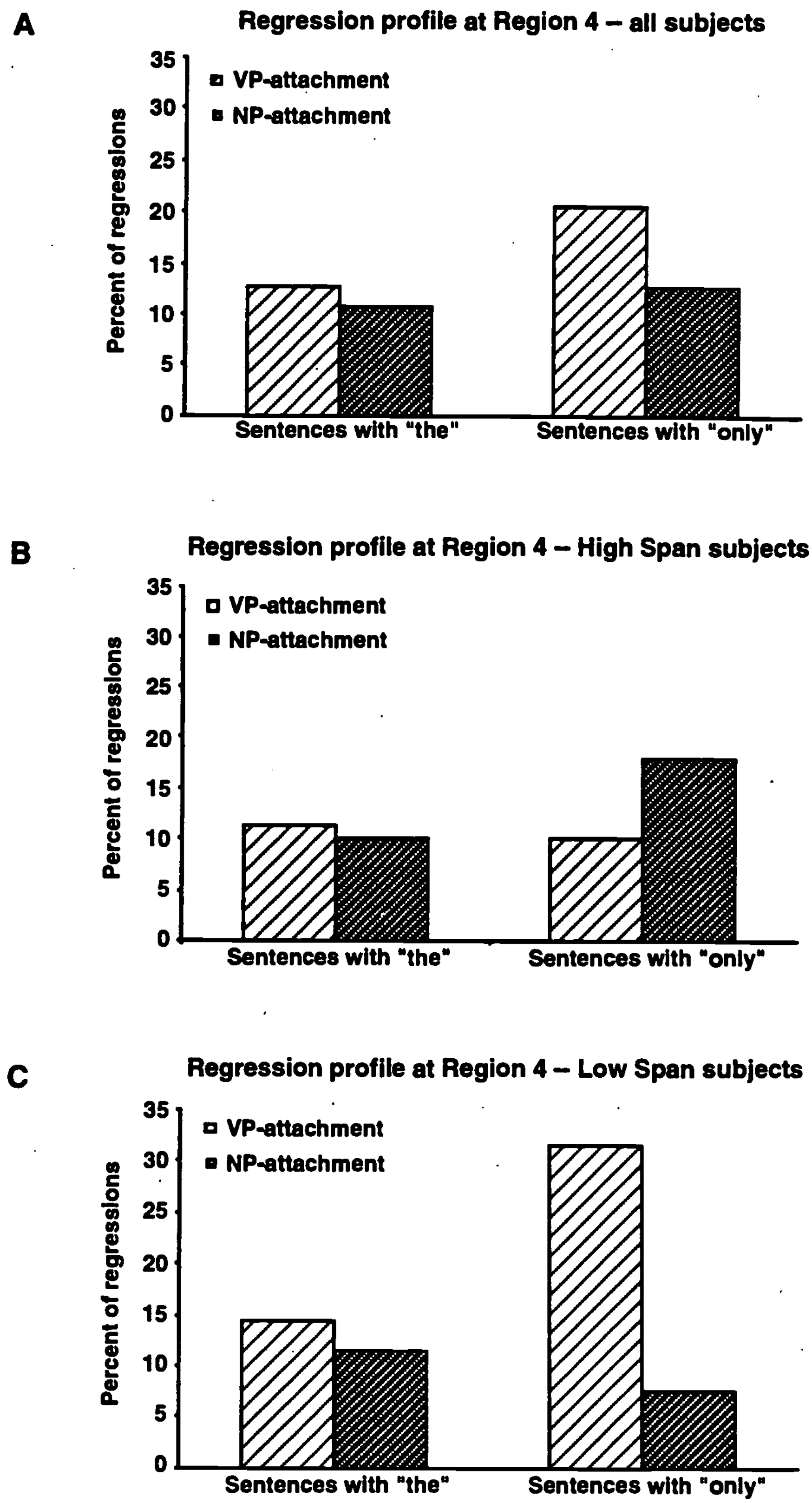


Figure 10. **Experiment 4—Percent of regressions at Region 4 by all subjects (A), High Span subjects (B) and Low Span subjects (C)**

Differences in the pattern of responses of High and Low Span subjects were observed, however, in processing sentences with ONLY. High Span subjects were apparently able to recover on-line from the anomaly in "Only-VP" sentences, where the PP *with new brushes* was not a plausible modifier for the NP *doors*. The long first pass reading times at Region 4 on "Only-VP" sentences by these subjects, coupled with the absence of appreciable regressions from that region, indicated that this group successfully reanalyzed the sentence without looking back to earlier regions. Low Span subjects had relatively fast reading times at Region 4, but these subjects made a greater number of regressive eye movements than the High Span group. This pattern of results suggests that these subjects had difficulties recovering on-line from an initial misanalysis; hence, they were compelled to review material that they had read earlier.

These findings related to memory span can be readily accommodated within the referential theory. Recall that in order to interpret the sentences with ONLY, a contrast set must be constructed. Failing to locate the contrast set within the current mental model of the conversational context, where it is expected to be found, the perceiver's next option is to examine the incoming string of words to see if a contrast set can be motivated on this basis. (Failing that, the alternative is to conjure up a contrast set. However, as we saw, this would require unmotivated additions to the mental model, and therefore, this option is dispreferred, according the referential theory.) Attempts to locate the contrast set in the incoming string require the perceiver to attach the ensuing PP as a modifier of the preceding NP containing ONLY. Having made the decision to attach the PP to the preceding NP, the parser continues its search for the contrast set within the PP. It turns out, however, that when the noun phrase of the PP is encountered in the "Only-VP" version of the test sentences in Experiment 4, the contrast set that presents itself is semantically anomalous. To achieve a semantically coherent interpretation, not only must the parser revise the structure of the earlier portion of the sentence, but it must also pursue its last-resort option for constructing a contrast set by making one up from scratch. Not surprisingly, the combined effort in making these computations proves highly demanding of memory resources and, therefore, pushes apart groups of subjects who differ in memory span.

The rapid first pass reading by Low Span subjects at Region 4 on "Only-VP" sentences suggests that, although they detected the pragmatic incompatibility of the noun *brushes* with their initial analysis, their memory resources had already been exhausted, thus triggering repeated regressive eye movements. Frequent resort to looking back would seem to imply that Low Span subjects are unable to use real-world knowledge effectively in on-line recovery from the misanalysis.[17] The referential theory gives a parsimonious explanation for this complementary pattern of eye fixations and regressions: The use of semantic information (carried by the focus operator ONLY) makes relatively light demands on memory resources. This explains why there are no significant differences between subject groups on the "Only-NP" sentences and no differences in any region preceding the anomaly in the "Only-VP" sentences. It is apparent that in the presence of ONLY all subjects pursued the NP-attachment analysis, following the principle of parsimony. It was the specific requirement that plausibility information be invoked to disconfirm the initial parsing decision in the "Only-VP" sentences that distinguished subjects with different memory capacities. (See Crain, Shankweiler, Macaruso and Bar-Shalom, 1990, for discussion of other related effects of working memory differences for sentence processing.)

In keeping with the preceding experiments, Experiment 4 also provides support for the referential theory. What is new is that the findings indicate that while the semantic information carried by ONLY is used on-line in resolving ambiguities, the use of plausibility information may not be. If and when plausibility is used depends upon the memory capacity of the individual. For High Span subjects, plausibility information seems to be used rapidly to recover from a misanalysis, whereas its use appears to be delayed for Low Span subjects. How quickly it is used depends on two factors that can be identified. One factor is the memory capacity of the reader. An additional factor is the point within the sentence at which plausibility information is available for use. If information pertaining to plausibility is encountered before the point of ambiguity, it can be effective in resolving local ambiguities that are encountered subsequently (Trueswell et al., 1994). On the referential theory, however, plausibility is used only to adjudicate among competing alternative partial structural analyses. Plausibility does not compete with more specific information about the conversational context if the latter is operative. The present findings conform to this expectation. Apparently, if information about plausibility is encountered *after* the point of ambiguity, then even though this information may be available to the parser, its implementation in decision-making may be delayed. In contrast, semantic (focus) information is rapidly used to adjudicate among competing structural analyses. The distinction between availability and use of plausibility information is consistent with the Modularity Hypothesis (see Fodor et al., in press).

5. GENERAL DISCUSSION

Findings were presented from four experiments that were designed to test predictions of the referential theory, and to assess its explanatory scope in the context of the current debate on the relative timing of the use of a variety of non-syntactic information in on-line sentence processing. The first experiment confirmed a major prediction of the referential theory, namely, that referential complexity is critical for the parser to resolve ambiguities involving main-verb/reduced-relative-clause analyses. A word-by-word reading test revealed that a simple substitution of ONLY for THE in the subject noun phrase substantially reduced garden path effects. The findings provide circumstantial evidence for the contention that ambiguity resolution is influenced by properties of discourse representations that are assigned to the alternative analyses of an ambiguous phrase.

Experiment 2 confirmed the results of Experiment 1, using records of subjects' eye movements. First pass reading times on garden path sentences containing the focus operator ONLY did not differ to a significant degree from those of their unambiguous control sentences in the disambiguating region. There was a significant difference at this region, however, when garden path sentences beginning with the definite determiner THE were compared with appropriate controls. The results were interpreted as evidence for the rapid on-line use of the semantic contribution carried by the focus operator ONLY.

It should be noted that in Experiment 1 all ambiguous sentences, with THE or with ONLY, produced significantly more errors than their respective unambiguous controls. These sentences also produced more regressions in Experiment 2. In our view, this effect is the result of parallel processing. Because more than one representation is computed within the ambiguous region, subjects occasionally select the one that is inconsistent with the discourse context. However, the presence of ONLY was sufficient to promote the reduced relative clause analysis on the majority of trials. Compared to the unambiguous control sentences, there was about a 25% increase in overall errors (in Experiment 1), and a 4% increase in regressions (at Region 5 in Experiment 2) for ambiguous sentences with ONLY. In contrast, for ambiguous sentences with THE, there was a 46% increase in overall errors in Experiment 1, and a 9% increase in regressions in Experiment 2.

Taken together, the findings of Experiments 1 and 2 provide evidence for the influence of referential content of noun phrases in the resolution of garden path sentences such as the infamous *The horse raced past the barn fell.* The findings of Experiments 1 and 2 support the joint predictions of the referential theory that semantic (focus) information is used to decide among competing syntactic analyses (Experiments 1 and 2) and that this information is used on-line (Experiment 2). These findings would not be anticipated, and are difficult to explain on any account of ambiguity resolution that ignores the referential properties of sentences in the initial decisions made by the parser. Although some researchers have shown themselves willing to acknowledge the involvement of referential factors in resolving ambiguities involving prepositional phrase attachments, it is probably generally believed that the effects of structurally-based strategies such as Minimal Attachment predominate in sentences with a main-verb/reduced-relative-clause ambiguity (see, e.g., Tanenhaus & Trueswell, in press). To the contrary, the findings of the present research show referential effects to be as strong, if not stronger, in the resolution of the main-verb/reduced-relative-clause ambiguity, than in resolution of ambiguities in the attachment of prepositional phrases, which was the subject in Experiments 3 and 4.

Returning to the method of word-by-word reading in Experiment 3, we found that subjects used information about *a priori* plausibility of alternative representations of an ambiguous phrase in arriving at their preferred attachment of a prepositional phrase. A simple substitution of ONLY for THE in a noun phrase followed by a prepositional phrase changed the parser's decisions as to where to attach the prepositional phrase within an ambiguous sentence. The subsequent effects were seen to depend on subjects' use of plausibility information contributed by the head noun of the prepositional phrase. The results were interpreted as further confirmation for the referential theory, which, as we noted, maintains the view that the parser bases its initial decisions on semantic/referential principles.

Experiment 4 was designed to estimate the time at which plausibility information is used by the sentence processing system. Adopting the standpoint of the referential theory, we anticipated that the use of this source of information would be pre-empted by referential factors contributed by the focus operator ONLY. This expectation is based on the principle of *a priori* plausibility, which maintains that plausibility information is invoked only if the semantic/referential content of a sentence does not offer a sufficient basis for selecting among the competing interpretations. The technique of eye movement recording was used to test this prediction using the same sentences as in Experiment 3. In analyzing the results of Experiment 4, subjects were grouped according to working memory span, to ask under what conditions the resolution of ambiguities would show individual differences in memory. The results suggested that the time course of the use of plausibility information did in fact co-vary with memory span. Such information was used rapidly by individuals who had relatively high memory spans, but its use was delayed and was probably less effectively used by low-span individuals. These results therefore support the view that the ability to take advantage of *a priori* plausibility is highly resource-dependent. Moreover, if the relevant information is encountered after the onset of ambiguity, as is the case with ambiguities of prepositional phrase attachment, its value for ambiguity resolution is more limited than if it is encountered prior to the onset of ambiguity. We take this result, too, as support for the referential theory, according to which semantic (focus) information is used on-line in the construction of discourse representations, with plausibility information exerting its influence only with ambiguities that remain unresolved after semantic principles have been applied.

Let us now take stock. There is a degree of consensus among researchers concerned with sentence processing that some kinds of non-syntactic information are very rapidly assimilated by the parser, as when it goes about the business of ambiguity resolution. There are simply too many empirical facts in the literature that point in this direction to make a denial plausible. But, as we have emphasized, non-syntactic influences are not all of the same kind, and researchers who adhere to the referential theory characteristically differ from adherents to the garden path model and the constraint satisfaction model, in how and when various kinds of non-syntactic information are incorporated by the parser. The referential theory makes a case for the primacy of discourse considerations. An integrated set of experiments was designed to find out how well the referential theory could predict and explain findings from experimental paradigms that permit comparison of the times at which discourse principles and factors governing real-world knowledge become operative. Our findings consistently confirmed that discourse principles are operative on-line in resolution of two kinds of structural ambiguities, and that they take precedence over plausibility. We further clarified the costs associated with use of these two kinds of information, showing that the parser's appeals to real-world knowledge, but probably not its application of discourse principles, are heavily resource consuming and are dependent on processing resources that vary greatly among individuals. We therefore maintain that the findings form a coherent package that can adequately be explained by the referential theory. Can the alternative accounts of sentence processing deal as well with findings such as these? We leave it for future research to decide.

REFERENCES

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191-238.

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31,* 685-712.

Altmann, G. T. M., Garnham, A., & Henstra, J. A. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 209-216.

Beach, C. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30,* 644-663.

Bever, T. G. (1970). The cognitive basis for linguistic structure. In J. R. Hayes (Ed.), *Cognition and the Development of Language.* New York: John Wiley.

Boland, J. E., Tanenhaus, M. K., & Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29,* 413-432.

Britt, M. A. (1991). *The Role of Referential Uniqueness and Argument Structure in Parsing Prepositional Phrases.* Ph.D. Dissertation, University of Pittsburgh, PA.

Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing and discourse: Context effects and their limits. *Journal of Memory and Language, 31,* 293-314.

Clifton, C. & Ferreira, F. (1989). Ambiguity in context. Language and Cognitive Processes, 4, 77-103.

Crain, S., & Ni, W. (1991). Parsermony. In Proceedings of Texas Conference on Cognitive Science, Vol. X. . University of Texas at Austin, TX.

Crain, S., Shankweiler, D., Macaruso, P., & Bar-Shalom, E. (1990). Working memory and comprehension of spoken sentences: Investigations of children with reading disorder. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory.* Cambridge: Cambridge University Press.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives.* Cambridge: Cambridge University Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348-368.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research, 23,* 407-434.

Fodor, J. D., Ni, W., Crain, S. & Shankweiler, D. (in press). Tasks and timing in the perception of linguistic anomaly. *Journal of Psycholinguistic Research.*

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior, 22,* 203-218.

Frazier, L. (1979). *On comprehending sentences Syntactic parsing strategies.* Unpublished doctoral dissertation, University of Connecticut, Distributed by the Indiana University Linguistics Club, Bloomington.

Frazier, L. (1994). Sentences (re)analysis. Paper presented at the 7th CUNY Conference on Human Sentence Processing, CUNY, New York.

Frazier, L., & Clifton, C. (in press). *Construal.* Cambridge, MA: MIT Press

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178-210.

Gibson, E. (in press). Memory Limitations and Sentence Processing Breakdown. Cambridge, MA: MIT Press.

Jackendoff, R. (1972). *Semantic interpretation of generative grammar.* Cambridge, MA: MIT Press.

Juliano, C., & Tanenhaus, M. K. (1993). Contingent frequency effects in syntactic ambiguity resolution. In *Proceedings of the 15th Annual Conference of the Cognitive Society.* (pp. 593-598). Hillsdale, NJ: Erlbaum.

Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research, 23,* 459-471.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Karttunen, L. (1974). Presupposition and linguistic context. *Theoretical Linguistics, 1,* 181-194.

Kennedy, A., & Murray, W. S. (1984). Inspection times for words in syntactically ambiguous sentences under three presentation conditions. *Journal of Experimental Psychology: Human Perception and Performance, 10.6,* 833-849.

Krifka, M. (1991). A compositional semantics for multiple focus constructions. *Linguistische Berichte, Sonderheft, 4.* Also in *Proceedings of Semantics and Linguistic Theory (SALT) I, Cornell Working Papers II.*

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8,* 339-359.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 32,* 692-715.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24,* 56-98.

MacDonald, M. C., Pearlmutter, N., & Seidenberg, M. S. (1994). Syntactic ambiguity resolution as lexical ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. (pp. 123-153), Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology, 45A*, 73-87.

Merlo, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research, 23*, 435-457.

Mitchell, D. C., Corley, M. M. B., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory and cognition, 18*, 69-88.

Murray, W., & Liversedge, S. (1994). Referential context and syntactic processing. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 359-388). Hillsdale, NJ: Erlbaum.

Nagel, H. N., & Shapiro, L. P. (1994). Prosody and the processing of filler-gap sentences. *Journal of Psycholinguistic Research, 23*, 473-485.

Ni, W., & Crain, S. (1990). How to resolve structural ambiguities. In *Proceedings of the North Eastern Linguistic Society, 20/2*, (pp. 414-427), University of Massachusetts, Amherst.

Pearlmutter, N., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language, 34*, 521-542.

Perfetti, C. A. (1991). The cooperative language processors: Semantic influence in autonomous syntax. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 205-230). Hillsdale, NJ: Erlbaum.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel,. S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956-2970.

Rayner, K. (1993). Eye movements in reading: Recent developments. *Current Directions in Psychological Science*. 81-85.

Rayner, K., & Morris, R. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G. B. Simpson (Ed.), *Understanding word and sentence*. North Holland: Elsevier Science Publishers B.V.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction between syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*, 358-374.

Rayner, K., Garrod, S., & Perfetti, C.A. (1992). Discourse influences during parsing is delayed. *Cognition, 45*, 109-139.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 3/4, SI*, 21-49.

Rooth, M. (1985). *Association with focus*. Unpublished doctoral dissertation, University of Massachusetts at Amherst.

Sedivy, J. C., & Spivey-Knowlton, M. J. (1994). The use of structural, lexical, and pragmatic information in parsing attachment ambiguities. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 389-413). Hillsdale, NJ: Erlbaum.

Soames, S. (1982). How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry, 13*, 483-545.

Speer, S. R., Crowder, R. G., & Thomas, L. M. (1993). Prosodic structure and sentence recognition. *Journal of Memory and Language, 32*, 336-358.

Spivey-Knowlton, M. J., & Sedivy, J. C. (in press). Resolving attachment ambiguities with multiple constraints. *Cognition*.

Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. (pp. 415-439). Hillsdale, NJ: Erlbaum.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental psychology, 47*, 276-309.

Stalnaker, R. (1974). Pragmatic presuppositions. In M. K. Munitz & P. K. Unger (Eds.), *Semantics and Philosophy*. New York: New York University Press.

Tabossi, P., Spivey-Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umilta & M. Moscovitch (Eds.) *Attention and Performance XV*. Hillsdale, NJ: Erlbaum.

Tanenhaus, M. K., Garnsey, S. M., & Boland, J. E. (1991). Combinatory lexical information and language comprehension. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge, MA: MIT Press.

Tanenhaus, M. K. & Trueswell, J. C. (in press). Sentence comprehension. In J. Miller & E. Eimas (Eds.), *Handbook of Perception and Cognition: Volume 11*. Academic Press.

Taraban, R., & McClelland, J. L. (1990). Parsing and comprehension: A multiple-constraint view. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension Processes in Reading*. Hillsdale, NJ: Erlbaum.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a constraint-based lexicalist approach to syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on Sentence Processing*. (pp. 155-179), Hillsdale, NJ: Erlbaum.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285-318.

FOOTNOTES

[*] *Language and Cognitive Processes*, in press.

[†] Also University of Maryland

[‡] Also University of Connecticut

[1] We are using the term "garden path effect" quite broadly, to refer to cases in which the attachment of a linguistic items in an on-going structural analysis results in a structure that is incompatible with later input, and also, as a description of cases in which an attachment requires *re*interpretation.

[2] The empirical basis for Mitchell et al.'s claims has been challenged by Altmann, Garnham and Henstra (1994), who question both the effectiveness of the test materials and the proper interpretation of the data. Regardless of the outcome of this debate, however, it is important to heed the general point made by Mitchell et al. These researchers emphasize the importance of testing for syntactic biases at the earliest point possible, so as to detect subtle, possibly unconscious garden path effects, if these exist. One way to accomplish this is to obtain a precise record of the parser's on-line operations, using measures that are sufficiently sensitive. We return to this point in the discussion of Experiment 2.

[3] Although the main findings of the two techniques are almost entirely complementary, the results from eye-movement recording provide a more fine-grained view of processing difficulties within the ambiguous region of a "garden path" sentence.

[4] Crain and Ni (1991) show how the referential theory applies to purely semantic ambiguities; that is, ambiguities that could not, in principle, be explained by structurally based criteria. Based on the observation that the principles of discourse and reference are independently motivated, and cover a range of phenomena not handled by structurally-based models, Crain and Ni contend that the referential theory has an edge on these other models to the extent that the referential theory also can also explain the garden path effects that occur in structural ambiguities of the sort discussed in the literature.

[5] Formally, the semantic value of the focus operator ONLY is captured by the following rule (adapted from Krifka, 1991; also see Jackendoff, 1972, Rooth, 1985):

MEANING RULE FOR ONLY:

$$B(F) \,\&\, \forall X[\{ X \in CON(F) \,\&\, B(X) \} \rightarrow X = F]$$

Where X is a variable of type F, and CON(F) is a set of contextually

determined alternatives to F.

The first conjunct of the meaning rule, B(F), states that the background must apply to the focus element. The second conjunct is the statement of uniqueness: $\forall X[\{ X \in CON(F) \,\&\, B(X) \} \rightarrow X = F]$. Here, the universal quantifier ranges over a *meta*variable, X. By replacing the metavariable X with actual variables of different types, different interpretations may be derived, depending on the nature of the entities that are being contrasted with the focus element. This provides the flexibility to cope with alternative interpretations for sentences with ONLY. If the element in focus is an *individual*, then the contrast set contains individuals. In this case, the metavariable is replaced by an individual variable: x, y, and so on. By contrast, if the focus element is a *property* of individuals, then the contrast set consists of sets of properties of individuals, rather than individuals themselves. In such cases, the metavariable is replaced by a variable of this type: P, Q, and so on. The meaning rule ends by guaranteeing the *uniqueness* of the focus element—for each member of the contrast set, if the background applies to it, then that member is the focus element itself.

[6] We used fewer control sentences than experimental sentences in an effort to minimize subject fatigue.

[7] Just and Carpenter (1980) observe that there are extra processing requirements associated with the end of a sentence. Subjects tend to pause longer at the word or phrase that terminates a sentence for several reasons. First, they may be searching for references that have not been assigned; Second, they may be constructing interclause relations (contextual integration); and finally, they must handle any inconsistencies that could not

be resolved within the sentence. In a word-by-word reading task, the maintenance of verbatim information in working memory may also contribute to the sentence wrap-up effect.

[8]To simplify the exposition, we will use the following conventions for expressing p values in ordinary language: $p < .01$ = highly significant; $p < .05$ = significant; $p < .1$ but $> .05$ = approaching significance; $p > .1$ = not significant. Also, because experiments 1, 2 and 3 used a mixed design, analysis by subjects (F_1) and analysis by items (F_2) were not both carried out in all cases. The symbol (F) will be used if only one analysis is possible, either by subjects or by items.

[9]Eye-movement patterns may reflect precognitive processes, however, as pointed out by Altmann and Steedman (1988, p. 217).

[10]Trueswell et al. (1994) maintain that per-character reading time, a widely used measure, may provide a misleading basis for comparison across regions of different length. Per-character reading time progressively distorts the accuracy of the analyses as the length of a region decreases. If the regions being compared are identical in length and/or in material, then the measure of uncorrected total reading time suffices. If the portions being compared are different (especially in length), then residual reading time provides a more accurate measure.

[11]Closely related to regression patterns are second pass reading times, which are calculated from the fixation durations in the portions of sentences that have either been read earlier or skipped entirely. Our data show that, overall, ambiguous test sentences required longer reading times and were reread more often than their unambiguous controls. There was no indication that the two types of test sentences were reread differently, at least in cases where rereading did occur. However, in this paper, we report regression patterns rather than second pass reading times. As pointed out by the editor and an anonymous reviewer, it is often difficult to interpret second pass reading time data, because what counts as a second pass fixation depends on how a region is defined. For instance, a particular fixation Y that follows a fixation X but lands to the left of X may be counted as a first pass fixation if the landing site is within the boundary of a particular region. On the other hand, if the region boundary is redefined so that it falls between fixations X and Y, then Y is counted as a second pass fixation.

[12]A finer-grained, word-by-word analysis of first pass eye fixation durations within the ambiguous region was conducted to look for signs of any effect of reanalysis for the "Only-amb" sentences. The most we could find was a non-significant elevation that occurred in the middle of the ambiguous region for "Only-amb" sentences relative to "The-amb" ones. This was probably because ambiguous sentences with ONLY required the construction and evaluation of a contrast set in addition to the difficulties imposed by parallel processing. There was also an effect of AMB/UNAMB that approached significance ($F(1,21) = 4.20$, $p < .06$) in the middle of the ambiguous region, with both ambiguous sentences taking longer to read than their unambiguous controls. By the end the ambiguous region, however, reading times for all of the versions had converged.

[13]A distinction has been made in recent studies (e.g., Britt, 1991; Britt et al, 1992; etc.) between attachment preferences due to an argument and ones due to an adjunct of a PP: A PP that follows a verb which requires a Goal argument (e.g., *put*) prefers to attach to the verb, while one that follows a verb like *throw*, which does not require an argument, may be more responsive to contextual manipulations. Other factors such as definiteness of NPs also affect parsing decisions (Crain & Steedman, 1985; Sedivy & Spivey-Knowlton, 1994). The present research does not consider these factors, which are held constant in the experimental manipulations.

[14]This experiment used a mixed design in which half of the subjects read only VP-attachment sentences (with THE and ONLY), and another half read only NP-attachment sentences (with THE and ONLY). As a result, THE/ONLY was a within-subject variable but VP/NP was a between-subject variable. Most of the data analyses, therefore, were by-item analyses (F_2). The mixed design was used because a pilot test showed that subjects were confused when both VP- and NP-attachment sentences were present, and there was a spill-over effect. This was probably caused by the task that asked subjects to judge whether or not the sentence continued to make sense at every word. In Experiment 4, which used the same material in an eye movement monitoring study, a fully crossed design was used.

[15]The division of regions was based on that used by Rayner et al. (1983).

[16]Daneman and Carpenter (1980) report correlation between subjects' performance on measures of language comprehension and memory span .

[17]The alternative is to suppose that Low Span subjects followed the Minimal Attachment strategy and did not initially pursue the NP-attachment analysis of prepositional phrases in sentences with ONLY. On this account of the findings, the significant number of regressions by these subjects to "Only-VP" sentences is unexplained. Mitchell (personal communication) suggests that the garden path model can account for this kind of result if it stipulates that early corrections occur when resources are available. However, evidence is yet to come that a revision was attempted after a brief garden path effect within the prepositional phrase.

APPENDIX A

TEST MATERIALS FOR EXPERIMENTS 1 & 2

(Experiment 2 used a subset of these sentences with some modifications)

Test Sentences

(Each sentence has four types: The-amb, Only-amb, The-adj-amb, Only-adj-amb)

1 The/Only (smart) people taught new math will pass the test.
2 The/Only (wealthy) businessmen loaned money at low interest were told to keep record of their expenses.
3 The/Only (brave) soldiers killed in the line of duty were mourned.
4 The/Only (frequent) visitors issued passes used them to leave on weekends.
5 The/Only (protesting) generals presented copies of the report blamed the government for cutting defense spending.
6 The/Only (greedy) wives left during the first month of marriage demanded alimony at the hearing.
7 The/Only (bold) students pushed into the flow of traffic got badly hurt.
8 The/Only (inexperienced) boxers punched hard in the early rounds were unable to finish the bout.
9 The/Only (poor) shopkeepers charged for repairs thought they were being cheated.
10 The/Only (old) union workers warned about possible layoffs picketed the company's main office.
11 The/Only (trained) paratroopers dropped into the dense jungle were captured by guerrilla forces.
12 The/Only (gourmet) chefs asked to have food ready refused to do so.
13 The/Only (cranky) children rocked to sleep soon woke up.
14 The/Only (new) homeowners hurt because of the increase in taxes decided to join forces against the administration.
15 The/Only (junior) pilots delivered the warning notice went out on strike.
16 The/Only (senior) doctors stopped while driving to work were not fined by the police.
17 The/Only (skinny) clowns tripped during the skit remained on the ground until the end of the performance.
18 The/Only (new) owners offered tempting food gulped it down.
19 The/Only (dishonest) students furnished answers before the exam received high marks.
20 The/Only (adventurous) swimmers drowned in the icy lake were not found until the spring.
21 The/Only (social) organizations donated emergency supplies helped to provide shelters to the earthquake victims.
22 The/Only (irate) listeners called during prime time programs didn't answer their phones.
23 The/Only (fishing) ships salvaged during the hurricane returned to the dock.
24 The/Only (fresh) turkeys roasted for under three hours were ready in time for the banquet.
25 The/Only (trained) social workers lectured about the dangers of smoking tried to help their own friends quit.
26 The/Only (chocolate) cookies baked in the brick ovens were sold at the carnival.
27 The/Only (big) boulders rolled down the mountain stopped the approaching trucks.
28 The/Only (crooked) dealers sold forgeries went straight to the police.
29 The/Only (heavy) boats floated down many rivers failed to get over the rapids.
30 The/Only (senior) senators elected to hold dinners for fundraisers were allowed to missed the vote.
31 The/Only (famous) actors paid for the entertainment performed in an outdoor theater.
32 The/Only (retired) men delivered junk mail threw it in the trash.

Control Sentences

(Each sentence has four types: The-unamb, Only-unamb, The-adj-unamb, Only-adj-unamb)

1 The/Only (young) hunters bitten by ticks worried about getting lime disease.
2 The/Only (dangerous) criminals taken into custody at the riot were not released the next day.
3 The/Only (homeless) people shaken by the earthquake feared an aftershock.
4 The/Only (misery) jewelers given huge diamonds cut them into small stones.
5 The/Only (pretty) models drawn by the illustrator were used for a magazine cover.
6 The/Only (candidate) managers chosen by the company answered every question at the interview.
7 The/Only (white) strangers seen at the time of the robbery had scars.
8 The/Only (indecent) scientists proven to be incorrect faked their data.
9 The/Only (short) hymns sung with great emotion were worth listening to.
10 The/Only (long) speeches written by the candidate were hard to understand.
11 The/Only (blues) vans stolen from the parking lot were found in a back alley.
12 The/Only (small) crops grown by farmers were damaged by the frost.
13 The/Only (strong) horses ridden past the finish line were given the prizes.
14 The/Only (sports) cars driven at high speeds were found to be defective.
15 The/Only (dance) shoes worn by the famous actress were put on display.
16 The/Only (fried) poultry eaten at the fair gave people an upset stomach.

APPENDIX B

TEST MATERIALS FOR EXPERIMENTS 3 & 4

(Each sentence has four types: The-VP, The-NP, Only-VP, Only-NP)

1. The burglar blew open (the/only) safes with (high quality dynamite/high quality diamond) and fled.
2. The historian studied (the/only the) maps with (a magnifying glass/large print) carefully.
3. The cleaners wiped (the/only the) windows with (a heavy durable cloth/a heavy covering of dirt) every day.
4. The workman opened (the/only the) valves without (much effort/much rust) during the flood.
5. The monkey tried to eat (the/only the) bananas with (silverware/bruises) during the show.
6. The dressmaker cut (the/only) material with (unusual scissors/unusual patterns) for the wedding.
7. The little girl cut (the/only the) oranges with (paring knives/thin skins) before dinner.
8. The drunk smashed (the/only the) windows with (an empty bottle/stained glass) last night.
9. The vet tranquilized (the/only the) tigers with (a dart gun/bad tempers) before treating them.
10. The tribesmen killed (the/only) lions with (poison arrows/sharp teeth) as part of the ritual.
11. The company demolished (the/only) buildings with (huge bulldozers/cement foundations) last weekend.
12. The craftsman stripped (the/only) cabinets with (paint stripper/brass hinges) for the building.
13. The doctor examined (the/only) women with (a cold stethoscope/high fever) before the surgery.
14. The secretary typed up (the/only) reports with (an IBM typewriter/few diagrams) after lunch.
15. The thief opened (the/only the) doors with (a credit card/faulty locks) during the robbery.
16. The woman repaired (the/only the) socks with (some thread/large holes) during the TV show.
17. The fireman broke (the/only) windows with (the ax/rusty hinges) during the fire.
18. The gardener cut down (the/only the) trees with (the chainsaw/the disease) last fall.
19. The detective was watching (the/only) women with (binoculars/straw hats) at the station.
20. The man decided to paint (the/only) doors with (new brushes/large cracks) for the festival.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 175-183

Why is Speech so Much Easier than Reading and Writing?*

Alvin M. Liberman

About 25 years ago, some of my colleagues posed the question that was, in their view, basic to an understanding of the reading process and the ills that so frequently attend it: what must the would-be reader know that mastery of speech will not have taught him? Drawing on a combination of common sense, old knowledge about language, and new knowledge about speech, they arrived at the hypothesis that a missing and necessary condition was what has come to be called phonological awareness—that is, a conscious understanding that words come apart into consonants and vowels. Research then demonstrated that such awareness is not normally present in preliterate children or illiterate adults; that measures of awareness provide, perhaps, the best single predictor of reading achievement; and that training designed to develop awareness has generally happy consequences for those who receive it.

But the pioneers of phonological awareness rather neglected the flip side of their inquiry: why is phonological awareness not necessary for speech? My aim is to repair that omission. To that end, I will seek reasons, in addition to those my colleagues found, why the phonologic structures that are common to speaker and hearer are nevertheless not noticed by either. Beyond further rationalizing the hypothesis—now a fact—that phonological awareness is not a normal by-product of learning to speak, those reasons should lay bare the critical difference between speech and reading/writing, and so let us see why the one is so much easier than the other. Moreover, when taken together with considerations having to do with the operation of the phonological facility, they may enlarge our understanding of certain deficiencies that poor readers have, apart from the process of reading itself (Liberman, Shankweiler, & Liberman, 1989).

I begin, however, not with notions about why speech does not require awareness, but rather with some speculations about why that aspect of the issue was initially scanted. That is surely a chancy and presumptuous thing for me to do, for I cannot expect researchers to have written about the questions they never raised, so I cannot know whether my colleagues did not think to ask them, or did not think them fit to ask. I will therefore rely on what I know of the awareness issue as it developed in the mind of Isabelle Liberman, one of the pioneers of the awareness enterprise. Because Isabelle habitually used me as a sounding board, I was privy to the intellectual trial and error that led to the insights behind her signal contributions. I remember the hits and the misses, the turns, both right and wrong, and, of particular interest for the purposes of this essay, the turns not made at all because they lay on roads not taken.

It all began for Isabelle when, in order to accommodate her career to the constraints of the University's nepotism rule, she was assigned to teach teachers how children learn to read, and why some don't. She had not, at that point, done research in the field, and was unacquainted with the literature, but she was nonetheless determined that her teaching be grounded in reasonably solid science. She therefore undertook a two-part program. First, she took stock of what she, and presumably every other educated person, knew about speech and language that might be relevant, carefully selecting only those facts and generalizations that were firmly established. Then, given those pieces of secure and presumably pertinent knowledge, she measured their implications against the received wisdom about reading as it appeared in textbooks and the research literature.

What Isabelle (and, as she thought, everybody else) knew that seemed relevant fell into two categories: the difference between language and all other forms of natural communication; and the difference between speech and reading/writing.

As for the relevance of the difference between language and other natural modes of communication, Isabelle knew that by far the most important property of language is that it is generative, in contrast to all nonhuman, but equally natural, modes of communication, which are not. This is to say that language can communicate an indefinitely large and various set of messages, including many that are entirely novel. Human beings communicate in this wonderfully productive way as easily and naturally as they walk, but only because they have ready access to the two generative devices—phonology and syntax—that their language faculty provides. Lacking both a phonology and a syntax, nonhuman animals have only that apparatus which is necessary to connect a few signals to an equal or smaller number of unchangeable messages. Isabelle wondered, therefore, why some reading specialists should nevertheless suppose, as they seemed to, that skilled readers go directly from print to meaning, presumably by-passing their phonological and syntactic processes altogether (Smith, 1971). Surely, that would be to do in reading what is not normally done in speech, where phonologic and syntactic processing are not a matter of choice, but mandatory, and so to trade generativity for the severe constraints that characterize all other natural forms of communication. She reckoned that a writing system, as well as the manner of its use, must preserve generativity at all costs, and that our alphabetic system does it passing well, but only when properly used. To her, proper use required that the reader attach the artifacts of the alphabet to the natural structures of his language, taking care to make the connection at the earliest possible stage. That done, the reader gets all the rest of the complex processing for free, courtesy of the biological specialization for language that he owns simply by virtue of his membership in the human race.

Thus, at the level of the word, the reader who has read it right can deploy his powerful phonologic resources, with the result that generativity is preserved, and he is not reduced to treating words within the narrow limits that the nonhuman, nonphonologic modes allow. In this connection, it seemed to be little appreciated in the reading community that the phonology a reader can exploit is not merely a list of sounds—or letter-sound correspondences—but rather a marvelous combinatorial scheme, unique to speech, that comprehends all the words the reader already knows, as well as those he has forgotten, and those he has yet to learn.

As for the sentence, no intellectual exertions are necessary once the reader has made proper contact with his natural language faculty, for then even the most complex sentences will be handled as easily as they are in speech, which is, more often than not, easy enough. Isabelle could not understand, therefore, why so few among reading specialists were concerned to know where or how the contact might be made, but rather seemed to assume the existence of a 'visual' language, where readers not only perceive the print visually, as they must, but also represent the words that way, as Isabelle thought they must not. For surely, the natural way of understanding the sentence would be entirely beyond the reach of such a 'visual' reader, if only because the syntactic component of the biological specialization for language cannot have evolved to deal with anything but phonologic representations (together, of course, with their grammatical appendages); there is no reason to think it would know what to do with the outputs of the visual system. So if the representations were exclusively visual, readers would be required to develop a wholly new mechanism, and one for which they had no natural bent, simply to do that which the old system, given a more propitious input, is adapted to do automatically. Since the most sophisticated linguists and psycholinguists had been unable to figure out how the old system works, it seemed most unlikely that a 'visual' reader could succeed where they had failed, and so invent a new system just to meet the unnatural demands of the visual way he had unwisely chosen to read. Most generally, in this connection, Isabelle came to think it seriously misleading to suppose that language can be 'visual' or 'auditory' when, in fact, it can be neither. For reasons I will develop later, she was in process of conceiving that language is a biologically coherent modality in its own right, possessed of its own uniquely linguistic structures and processes. Why, then, might it be hard for someone who is perfectly at home in that modality to enter it by way of a seemingly simple transcription of its modality-specific structures? Just about everything Isabelle wanted to know about beginning reading was brought into focus by that question.

Turning to the difference between speech and reading-writing, Isabelle, and presumably everybody else, knew that the former is a species-typical product of biological evolution, arguably the most apparent and defining of our genetically determined characteristics, in contrast to the latter, which is an intellectual achievement of an apparently difficult sort. There is, then, a strong presumption that we've been talking ever since we emerged as a species (or, according to some, as a genus), which was from 200,000 to several million years ago (depending on whether you think speech began with the one class of creatures or the other); but it was less than 4000 years ago that some of our fellow humans discovered the alphabetic principle, and put it to practical use. What was truly unique about this discovery was not the idea that drawings can be used to represent speech instead of objects or ideas. That was of critical importance, to be sure, but it had been exemplified in the rebus, the first true (i.e. generative) writing system, and then elaborated several times independently in the syllabic or morphosyllabic scripts of, for example, Sumerian, Mayan, and Chinese. The unique discovery underlying the alphabet was neither more nor less than what I have already identified as segmental phonology, the part of grammar that generates all words by variously combining and permuting a small number of consonants and vowels. Seen that way, the alphabet was a triumph of applied linguistics. But why has it to be reckoned a triumph? Why has the discovery been made only once, all applications having been borrowed, in effect, from that first, seminal event? In short, why was it hard 4000 years ago for all pre-alphabetic humans, and why is it hard now for the pre-alphabetic child?

Surely, reading teachers must have asked those questions, for they could hardly know what it is they have to teach, or how to teach it, without knowing what it is the student must learn, and why the learning might not be easy. But when Isabelle searched the textbooks and the research literature for the answers, she could not even find the questions. Nevertheless, ideas about reading were thick on the ground, and all did at least imply answers to her questions, answers that ranged, in her view, from the improbable to the impossible. Way off at the impossible end of the continuum was a notion, now in full flower as a basic assumption of Whole Language, that foreclosed almost all questions about the reading process by asserting that learning to read is just as easy and natural as learning to speak, or would be, if only we taught reading the way we teach speech (Goodman & Goodman, 1979). To the end of her days, Isabelle found it shocking that this proposition—so obviously false in light of absolutely everything we know about language and its biology—should be taken seriously by so many, and should, indeed, have become the cornerstone of what is currently the most widely accepted theory of reading and method of instruction. Whenever I asked her how a proposition like that could possibly have prospered, she would either offer one of the mordant comments for which she was justly famous, or else say, resignedly, "Go figure."

A little closer to being merely improbable was the claim—made, incidentally, by a guiding spirit of Whole Language—that reading is a 'psycholinguistic guessing game' (Goodman, 1976). But could it be that the great event underlying the development of the alphabet was that some human being had discovered, at long last, that people can guess at language, and, accordingly, that guessing is the critical skill the reader has to acquire? Not likely, it seemed. Anyway, it rather offended Isabelle's sense of the rightness of things that readers would want to guess what the word might be when the actual word was right there in plain sight.

But by far the most numerous theories located the difficulty somewhere in the eyes. Now surely, a person cannot be expected to read if he cannot see the print. But, just as surely, it could hardly have been pandemic visual deficiencies that had so effectively blocked the development of an alphabet; hence, it can hardly be true, except in special cases, that rectifying such deficiencies is now the critical step in the development of the ability to use one,

Of course, some did make what seemed an appropriate obeisance to phonology by supposing that children had to learn the so-called letter-sound correspondences—the heart of 'phonics' instruction. But Isabelle thought that a trivially easy task, hence not likely to be the core of the child's problem. Indeed, she had already begun to see that emphasis on those correspondences rested uneasily on the false assumption that speech is an acoustic alphabet, for it is only on such an assumption that a child might have been able to 'synthesize' a word out of its letter-sound constituents—that is, 'sound it out.' For reasons to be developed later, speech cannot be, and is not, an acoustic alphabet, so attempts to sound out a word from the 'sounds' of its letters will typically produce an utterance—as often as not, a nonword—that has as many syllables as the target word has letters. Still, Isabelle was ever willing to grant that learning the sounds of the letters might be of some help in moving the child to the right insight. At the least, it was better than leaving the child to his own devices, or trying to mislead him into believing that the printed word is a picture, a meaning, or an idea, when, in fact, it is a piece of language—actually, a phonologic structure—to which certain meanings may, or may not, be attached. But sounding out was not the only way to get the child to see what the game is all about, nor did it seem, given what Isabelle was beginning to learn about speech, necessarily the best way.

All the foregoing is by way of telling where Isabelle wanted to go, and how it was that she could find in the ideas of the reading specialists only that which would have taken her somewhere else. So she, together with the other early members of the Haskins reading group, including, especially, Donald Shankweiler and Ignatius Mattingly, turned away from those ideas, and put their attention, instead, on speech. That seemed the thing to do, since speech and an alphabetic writing system have the same primary function, which is to convey the internal structure of words. Therefore, one might hope to find in speech the key to understanding why that structure was so hard to get from the printed page.

Happily for the reading group, other Haskins colleagues had for some time been doing research on speech, and had uncovered a few characteristics that might be relevant to the reading problem. The one that seemed most likely was that speech is not the acoustic alphabet so many had assumed it to be, so it cannot be mapped directly onto the optical alphabet a reader must learn to use (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The reason is easily understood once one thinks about the requirements of phonologic communication. The most obvious of these is imposed by the generative combinatorial strategy that phonology exploits. For if words are to be formed by combining and permuting a set of meaningless units, then the units must be commutable, which is to say discrete, invariant, and categorical. Just like the letters of the alphabet, indeed. But if those units were sounds, and the sounds had to have those characteristics, then the sounds could only be produced by articulatory maneuvers that had them, too. In that case, as Isabelle was later to emphasize repeatedly, a speaker could not say [bag], but only [bə] [æ] [gə], and to say [bə] [æ] [gə] is not to speak, but to spell. Communication by spelling would, of course, be painfully slow, and the listener would presumably find it nearly impossible to organize the phonologic segments into the larger units of words and sentences, It is also relevant to the rate problem that even if it could somehow be solved in production, the result would defeat the ear. Speech delivers phonologic information at rates of 10 to 20 consonants and vowels per second. But if each consonant and vowel were a unit sound, as it would be in an acoustic alphabet, rates that high would strain the temporal resolving power of the ear and overreach its ability to keep sequential order straight. So, an acoustic alphabet is impossible, if people are to speak and listen as fast as they must.

Some of Isabelle's speech-research colleagues believed—correctly, as I think—that nature had solved the rate problem by defining the phonologic units, not as sounds, but as abstract motor structures that control the articulatory movements by which those sounds are made (Liberman & Mattingly, 1985). The critical advantage for the speaker is that unit gestures corresponding to the discrete, invariant, and categorical units of the phonology can be coarticulated—that is overlapped and merged—with the further result that speakers can run them off at the high rates that characterize speech and make language possible. For the listener, coarticulation efficiently packs information about several phonological segments into the same piece of sound, thus loosening the constraints that are imposed by the temporal resolving power of the ear. As for the difficulty that auditory perception has with sequential order, coarticulation produces context-conditioned variations in the acoustic signal that mark order by the shape of the signal, not by some temporal sequence of its presumably discrete pieces. Thus, to perceive that [b] comes first in [ba] and second in [ab], the listener relies on the fact that the acoustic cues for the consonants are mirror images, reflecting the coarticulated gestures from consonant closure to vowel opening, in the one case, and the reverse progression in the other. If such syllables are of relatively short duration, the acoustic signals will carry information about both consonant and vowel throughout their lengths, so acoustic shape can be the only basis for determining the order of the phonetic units. Given a linguistic specialization that recovers the gestures, of which more later, the acoustically different signals will nevertheless evoke the same phonetic percept, accurately marked for its position in the sequence.

The general consequence of coarticulation is that almost any piece of sound, no matter how short, carries information about, not one, but several units in the phonologic string. Accordingly. the sounds of speech are not a substitution cipher on the phonologic structure, but a complex and specifically linguistic code in which there is simply no correspondence in segmentation between any delimitable acoustic segments and the segments of the phonology. This is not to say that the underlying phonology is not alphabetically segmented, only that the segmentation is not apparent at the acoustic surface.

It was, then, the foregoing facts and speculations that initially suggested to the reading group that a child who had mastered speech might nevertheless be unaware of the discrete segments it conveys. In listening to a word, he would not have heard a succession of discrete, segmented sounds. Moreover, as previously noted, it would have been hard to develop awareness in the child simply by showing him how to divide a word into, or synthesize it from, alphabet-size pieces of speech sound, because, apart from vowels, such sounds do not exist.

As I earlier implied, our reading-research colleagues did not have all the reasons why awareness is lacking, but they had at least one, and that was enough to head their enterprise in the right direction. They could take satisfaction in the fact that their reason was well grounded in the background science (Mattingly, 1971; Shankweiler & I. Y. Liberman, 1971). Moreover, the hypothesis it led them to held up under empirical test, and bent agreeably to necessary elaborations, and amendments. But most important was the practical application, which lay in the assumption that, to get the child up and running, someone should teach him how words come apart, for speech had not revealed to him that they do, yet that was exactly what he needed to understand if he was properly to appreciate and apply the alphabetic principle. Once the child had that principle, he would know what to look for, so further refinements in his understanding of the exact relation between the alphabetic script and the language could come with experience in reading, as the phonemic and morphophonemic regularities of the writing system revealed themselves. This would happen the more readily, of course, if the teacher provided the right help by contriving exercises designed to make the regularities most apparent. As for the irregularities, enlightened instruction would introduce them gradually, while also showing that many were not so wholly irregular as they seemed.

So, given what the reading group had yet to do in the further development and testing of their fertile hypothesis, there was no compelling reason for them to wonder why awareness was not essential in learning to speak. Indeed, in the matter of speech, they had gone about as far as they could go, for even their speech-research colleagues did not, at that time, have a firm grip on the rest of the story. We may not have that even now, but, as I dare to say below, we may at least have it in hand.

To understand why speech is different in the matter of awareness, and so to close the circle, I observe again that, in spite of the complexly encoded nature of the speech signal, phonological structures are, in fact, contained within it. Those stuctures *must* be produced and received by the speaker and listener, whether they know it or not, for if they were not, language as it has come to be would not exist. Moreover, it *is* possible to become aware of those structures, for, if it were not, alphabetic reading and writing as they have come to be would not exist. No matter, then, that the speech process itself is fully automatic, hence unavailable to consciousness as process; for the listener, that process must nevertheless produce phonologic representations of which the listener can be conscious. In that connection, Mattingly has noted that the automatic process that derives meaning from speech need not, in principle, ever make available to consciousness the phonologic structures that are intermediate to its goal, and, alone among students of language, he has speculated about the possible function that such representations might serve (Mattingly, 1990). For our purposes, however, it is enough to know that the representations are there and are available. The previously noted bad fit in segmentation between those representations and the acoustic signal is but one reason, and not necessarily the most compelling one, why there is, nevertheless, no awareness; in any case, it can hardly account for the fact that, for speech, in contrast to reading and writing, awareness is not necessary. What, then, does account for that fact?

The easy answer is that speech is a kind of instinct, and therefore a thought-free process, while reading/writing is, as I said earlier, an intellectual achievement of sorts. But that is just to restate the question. I want, therefore, to try a possibly more satisfying answer. Unfortunately for that purpose, such an answer arises from an unconventional theory of speech that is probably not well known to students of the reading process. Worse yet, a proper account of the theory would require that I describe several experiments, and that would require, in turn, a quick tour through the most technical details of acoustic phonetics. To avoid that thicket, and stay within the limits of space and time our host has set, I will try to describe and support the theory on grounds of plausibility only, and, within those bounds, keep reading ever in view. However, even that over-simplified approach requires an initial detour.

Consider, then, how reading/writing and speech meet a requirement that is imposed on both. Indeed, this requirement is imposed on every kind of communication, easily qualifying as the most fundamental requirement there is. It is odd, then, that it appears not to enter into the calculations of researchers in speech or reading, the more so because it is easy to see and easy to understand. The requirement is simply that what counts for the sender must count for the receiver. For that requirement to be met, two closely related conditions must be fulfilled: out of all possible signals, a certain few must be recognized as relevant to language; and the representations in the minds of sender and receiver must, at some point, be the same. On the assusmption that a thing is more likely to be noticed if it has a name, Mattingly and I have called this the requirement for 'parity,' and have challenged theorists of any stripe to say how it was established and maintained for whatever kind of communication they study (Liberman & Mattingly, 1989).

In the case of alphabetic reading and writing, it is easy to see exactly what the parity requirement is, and how it is met. Suppose, for example, that I write 'P' and also '/.' Every user of the Roman alphabet knows that the first character counts for language, but the second does not. Moreover, the writer of 'P' is in league with the reader of 'P,' because they have a common understanding of what linguistic unit the 'P' counts for: the bilabial voiceless stop consonant [p] that introduces the syllable [pat]. So parity exists, and the system will run. But we see immediately that parity, though real, is arbitrary; a user of the Cyrillic alphabet would agree that the chracter 'P' counts, but insist that it counts, not for a stop consonant, but for the continuant [r] , as in [rat]. Of course, parity has got to be arbitrary in writing systems, because it was established by agreement. Those who developed the Roman alphabet arrived at a compact that bound them to certain arbitrary decisions about which optical shapes would index which phonological units of the language. All who use that alphabet must become parties to that compact; adhering to its terms, they can communicate; otherwise, not.

But what of speech? How is it that [p] bcame relevant to language and a snort did not? Surely, not by agreement, for nobody invented speech, or somehow derived it as a secondary cipher on some more basic mode of communication; hence, nobody got everybody else to speak according to arbitrary decisions about which percepts count, and what they count for. To assume the contrary would be hardly less absurd than to assume divine intervention, as if there had been some extra commandments that Moses dropped on his way down from Sinai, one of which said, "Thou shalt not commit the phoneme [p], except as it is thy intention to communicate." No more is it plausible to suppose that speaker and listener have a common representation just because they both subscribe to an agreement that [p] is the name of some otherwise ordinary sound they both hear. As Studdert-Kennedy remarked some time ago, the thing about phonemes is that they "name themselves." My aim is to tease out the theoretical implications of that piece of wisdom.

But first, I would describe the conventional view of speech, just to show how the system it envisions fails to meet Studdert-Kennedy's requirement—how, in other words, it falls short in the matter of parity, and therefore cannot enlighten us about the most fundamental difference between speech and reading/writing. To simplify the matter, let us, for the moment, consider only speech perception. The view held explicitly by almost all researchers in speech—and, I should think, at least tacitly by people concerned about reading—goes as follows (for discussion, see Liberman (1992) and Liberman and Mattingly (1985)). The elements of speech are sounds, and their perception is as it would be for sounds of any other kind. All depend on the general processes of auditory perception, so all produce percepts of a generally auditory sort. Thus, the percepts evoked by a stop consonant and a squeaking door can differ only in the mix of auditory primitives—pitch, loudness, and timbre, for example—out of which they are presumably formed. They are made of the same perceptual stuff, as it were, which is to say that the percept evoked by the speech sound can be no more phonetic than the percept evoked by the squeaking door. In its most general form, this comes down to the assumption that language simply appropriated for its uses the most general processes and representations of the auditory modality. On that conventional assumption, however, the nonlinguistic auditory percepts would have somehow to be connected to language if they were to enjoy the very particular communicative privileges that linguistic status confers. According to the conventional theory, the necessary link is made at a cognitive stage, beyond perception, where the auditory percepts are associated with the phonetic units of language, and so, in effect, given phonetic names.

Of course, the percepts evoked in reading are just ordinarily visual in fact, as the percepts of speech are ordinarily auditory in theory, so it is a matter of fact, not theory, that the visual percepts have to be given phonetic names if they are to be used for lingistic purposes. No mystery there, however, since those percepts have only to be named after the perceivable phonetic units that are evoked by the sounds of speech, independently of the alphabetic shapes that were arbitrarily selected to stand in their stead. But what could the presumably auditory percepts of speech be named after? Surely, not themselves. The specifically phonetic names the conventional theorist would attach to them are neither primary acts nor percepts, so they must be in the nature of ideas, presumably innate, that are peculiar to language. For the theorist who has a taste for such innate ideas, assuming their existence does not settle the parity issue, for it still remains to explain how particular auditory percepts came to be connected to them. Presumably, many were called, but few were chosen. How, then, were the choices made, who made them, and what guaranteed that all would choose the same way? The seemingly inescapable conclusion is that, on the conventional view, parity in speech must have been established by agreement. But that's no way to run a natural communication system. Parity by agreement is acceptable, even necessary, for a biologically secondary process like reading, but not for the primary processes of speech. To say otherwise is to claim that speech is an artifact, like the alphabet, and that does violence to the facts.

It has simply got to be, I think, that the percepts evoked by the sounds of speech, in contrast to those evoked by the alphabet, are phonetic, not by virtue of having been given phonetic names, but *ab initio,* by their very nature. That is, they cannot be commonly auditory, as the conventional view would have it, but must rather belong to a phonetic modality that is as different from auditory as auditory is from visual. The primary perceptual response to speech would then be recognized as phonetic in the same way that a percept is recognized as belonging to any distinct modality; there is no need for some cognitive process to endow it with phonetic significance by giving it a phonetic name. On this unconventional view, what evolved was a communicative modality. Unlike the communicative modalities that evolved in other animals, this one was linguistic in nature, and therefore had a phonetic component. Given the functional requirements of phonological communication, the constituents of the phonetic component were, as I've already said, not sounds, but gestures. That these gestures are specifically, distinctly, and exclusively phonetic is not just a matter of plausibility, but of fact: they are a distinct set, different from those we make with the same organs when we swallow, move food around in the mouth, or lick our lips; having evolved to have a phonetic function, they serve no other. As for parity, it is built into the very bones of the phonetic modality, for the speaker produces specifically phonetic gestures, and the listener perceives them. Thus, the acts and percepts of speaker and listener do not have to be arbitrarily connected to language or to each other. The gestures provide a common phonetic currency, good for all linguistic transactions.

Also specialized for phonetic communication is the manner in which the gestures are controlled, since in the elaborate overlapping and merging that I called coarticulation, they are required, as other kinds of action are not, to preserve and transmit information about the discrete string of (phonetic) control structures that are their distal sources. Mattingly and I have proposed that all of this is managed by what we have called a 'phonetic module,' a biological specialization that, like all such specializations, has its own domain, its own mode of automatic signal processing, and its own primitives. The consequence for the purposes of this essay is that a speaker does not have to know how to spell a word in order to say it. Indeed, he does not even have to know that it has a spelling. He has only to think of the word, whatever that means; the phonetic module then spells it for him, automatically selecting and coordinating the gestures that form the phonologic structure. Small wonder the speaker doesn't notice that the structure is spelled, or how.

The listener is in similar case. Presented with the speech signal, he need not puzzle out the complex relation between it and the segmented phonetic structure it conveys. That, too, can be left to the phonetic module, for its complementary perceiving face is specifically adapted to parsing the signal so as to represent phonetic structure by recovering the underlying gestures that are its elements.

It is also relevant here to say again that the phonetic module is but one component of the larger specialization for language, for then we see that syntax, the other major component, must have evolved to work in close harmony with its phonetic partner; theirs would have been a marriage made in heaven. We should expect, then, that the representations produced as the output of the phonetic component would be grist for the syntactic mill, precisely adapted to what syntax wants and needs to do its job. Hence, those representations would pass through to syntax exactly as they came out of the phonetic module. They would not call attention to themselves because they would not require attention, and they would not require attention because they would not have to be made into something that they originally were not.

I must not suppose that all I've said here is what Studdert-Kennedy meant when he said that phonemes "name themselves," but it's the only story I am able to tell, and until I think of a better one, I am happy to be stuck with it. I like it, even as it is, because, aside from its fit to experimental results I've not presented here, it enables me to see why phonological awareness is neither a result nor a condition of learning to speak. More generally, it tells me how speech differs from reading and writing, why it is easier, and, not least, what's wrong with the conventional story about how it works.

Perhaps this story also redeems the claim I made at the outset about the various deficiencies that poor readers might have. The point is that the phonetic module cannot be expected to work equally well for everybody, and that a faulty module might have a variety of consequences, including some that are not directly reflected in reading performance. But, given that the module works only in the phonetic domain, the consequences, whatever they are, should be found there and nowhere else (Liberman, Shankweiler, & Liberman, 1989).

As a kind of coda, I venture that if Isabelle had been able to read all that I have just written, she would have offered a characteristically incisive comment. "I see," she probably would have said, "the point is that speech *is* language, but the alphabet only *refers* to it." Then she would have informed me that, reduced to that simple statement, my view would not have passed the grandmother test. By that she would have meant that her succinct characterization would predictably have elicited from Grandma the query, "So, what else is new?" or, perhaps, "You mean, I sent you to college for eight years to learn *that* ?" In my defense, I might then have observed that it is one thing to say the obvious, quite another to explain it. That would likely have wrung from Grandma the concession that "explanation can't hurt," but that would have been the extent of her praise for the notions I have advanced here.

REFERENCES

Goodman, K. S. (1976). Reading: A psycholinguistic guessing game. In H. Singer & R. Ruddell (Eds.), *Theoretical models and processes of reading* (2nd ed.). Newark, DE: International Reading Association.

Goodman, K. E., & Goodman, Y. M. (1979). Learning to read is natural. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading* (Vol. 1, pp. 137-154). Hillsdale, NJ: Lawrence Erlbaum Associates.

Liberman, A. M. (1992). The relation of speech to reading and writing. In L. Katz & R. Frost (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 167-178). Amsterdam: Elsevier Science Publishers.

Liberman, A. M., Cooper, F. S., Shankweiler, D. D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243*, 489-494.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1-33). Ann Arbor, MI: University of Michigan Press.

Mattingly, I. G. (1971). Reading, the linguistic process, and linguistic awareness. *Haskins Laboratories Status Report on Speech Research, SR27*, 23-24.

Mattingly, I. G. (1990). Reading and the biological function of linguistic representation. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shankweiler, D. P., & Liberman, I. Y. (1971). Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Reading by ear and by eye: The relationships between speech and reading* (pp. 293-317). Cambridge, MA: MIT Press.

Smith, F. (171). *Understanding reading* (1st ed.). New York: Holt, Rinehart, & Winston.

FOOTNOTE

*To appear in M. Joshi & C. Hulme (Eds.), *Reading and spelling: Development and disorder*. Hillsdale, NJ: Erlbaum Associates.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 185-195

How Theories of Speech Affect Research in Reading and Writing*

Alvin M. Liberman

When we remember that words are sounds merely, we shall conclude that the idea of representing those sounds by marks, so that whoever should, at any time after, see the marks would understand what sounds they meant, was a bold and ingenious conception, not likely to occur to one man of a million in the run of a thousand years...That it was difficult of conception and execution is apparent, as well by the foregoing reflections as by the fact that so many tribes of men have come down from Adam's time to ours without ever having possessed it. —Abraham Lincoln

My aim is to promote two notions about the relation between reading/writing and speech: the right theory of speech is essential to a coherent account of reading/writing; and the conventional theory of speech is the wrong theory. To say, as I will, that the conventional theory has therefore made it hard for researchers to see how reading/writing differ from speech is not to deny progress in the field; indeed, I believe, to the contrary, that research of the last few decades has brought insights that are both new and important. I would only suggest that the researchers who are responsible for those insights have either worked from the right theory or else managed somehow to ignore the wrong one. Naturally, I believe about the 'right' theory, not that it is perfectly and forever true, only that it is, by comparison with its more conventional competitor, more nearly right, and more likely, therefore, to head the reading/writing researcher in the right direction.

I should say that the conventional theory I consider wrong and the unconventional theory I consider right are only about speech in the narrow sense, by which I mean the component of the broader language faculty that comprises the production and perception of consonants and vowels. Though one of the virtues of the unconventional theory is that it makes speech an organic part of language, instead of the biologically arbitrary appendage that the conventional theory portrays, it is nonetheless possible for our purposes to deal with speech in isolation. However, I reserve the right to suggest at a later point that the biologically based fit of speech to the other components of language is an important reason why appreciation of the alphabetic principle is hard to come by.

To advance the two notions that are the point of this paper, I will rely almost entirely on facts that are in plain sight, requiring only to be thought about if their implications for speech and reading are to be seen. Because I had not myself thought about those facts, I would therefore emphasize I was long ago taken in by the conventional view. Specifically, I was led by it to suppose that speech is an acoustic alphabet, with segments of sound as discrete as the letters that convey them, and that I could, therefore, contrive an acoustic alternative for use in a reading machine for the blind (Liberman, 1996). Only after I had failed miserably to produce that alternative, and had then done a lot of research to find out why, did I begin to see that I might have known better before I started had I simply gone beyond surface appearances to take account of the somewhat deeper, if still visible, considerations I will invite you to think about. (A reading machine for the blind is now a reality, but it produces speech, not an acoustic alphabet.) I will claim that the conventional view fails the reading/writing researcher for much the same reason it failed me. If that failure has gone largely unnoticed, it is not because the conventional speech researchers have been unable or unwilling to understand what might seem plain, but only because they have not been concerned with the relevance of their theory to research on reading/writing or reading machines, and have therefore not had occasion to measure its implications against the hard realities of those enterprises.

In addition to the facts about language that are apparent to everyone, I will refer to just a few that have come from research on speech, but those are easily understood without a technical background in acoustic phonetics; moreover, they are not in dispute. All this is to say that the matter is not difficult at all, except, perhaps, in the telling.

The relevance of a theory of speech. As for the first notion—that a proper theory of speech is essential to an understanding of how people read—the most relevant consideration arises out of the deep biological gulf that separates the two processes. Speech, on the one side, is a product of biological evolution, standing as the most obvious, and arguably the most important, of our species typical behaviors. Reading/writing, on the other, did not evolve biologically, but rather developed (in some cultures) as a secondary response to that which evolution had already produced. A consequence is that we are biologically destined to speak, not to read or write. Accordingly, we are all good at speech, but disabled as readers and writers; the difference among us in reading/writing is simply that some are fairly easy to cure and some are not.

Viewing the matter from a slightly different angle, we see that, being at least as old as our species, speech has been around for 200,000 years or more, while the idea that it could be rendered alphabetically was born no more than 4000 years ago. Subtracting the latter number from the former, we conclude that it took our ancestors at least 196,000 years just to discover how to describe what it was they did when they spoke. Why did it take so long? Why was it so hard for our prealphabetic ancestors to make the momentous discovery, and why is it so hard for our preliterate children to understand it? Why has an alphabet been developed only once in all of human history? Surely, questions like those cry out for a theory of speech that explains in the same breath why an alphabetic description of speech is not immediately apparent to everyone, and why it should be almost wholly beyond the reach of some, Nothing less will do if we are to know how to teach children who are somehow ready to cope, while also helping those who are not.

Contrasting views of the biology of speech. There is a question that goes to the heart of the difference between the conventional and unconventional views of speech: does the specialization for language extend to the motor and perceptual processes underlying the consonants and vowels that speech and reading/writing use in common? The guiding assumption of the conventional view is that it does not (Crowder, 1983; Diehl & Kluender, 1986; Fujisaki & Kawashima, 1970; Kuhl, 1981; Lindblom, 1991; Massaro, 1987; Stevens & Blumstein, 1978; Sussman, 1989; Sussman, 1991; Warren, 1993). On that view, language simply appropriated modes of motor control and auditory perception that had already evolved in connection with nonlinguistic functions. Having been adopted by language for its purposes, those plain vanilla processes are now seen on the conventional view to work horizontally, serving linguistic and nonlinguistic behaviors alike.

According to the unconventional view, the specialization for language extends much farther, embracing even the very low level where the primary motor and perceptual representations of speech are to be found (Liberman & Mattingly, 1985; Liberman & Mattingly, 1989; Mattingly & Liberman, 1990; Mann & Liberman, 1983; Remez, Rubin, Berns, Pardo, & Lang, 1994; Whalen & Liberman, 1987). In other words, there are distinctly linguistic representations, not in the higher reaches of the cognitive machinery, but down among the structures of action and perception. Thus, language is seen as a vertically organized system in which linguistically specialized structures (and processes) are as central to phonetics as they are to syntax.

Among the more particular assumptions of the two views, perhaps the most fundamental concerns the nature of the ultimate constituents of speech (and, for that matter, language). On the conventional view, they are sounds. Just about everybody (including Lincoln, as the otherwise insightful epigraph makes clear) simply takes that for granted. And just about everybody holds that the sounds of speech are serviceable as consonants and vowels to the extent that they evoke distinctive auditory percepts.

The unconventional theorist, on the other hand, takes the ultimate constituents to be, not sounds, but articulatory gestures. Thus, the consonant we write as 'b' is a closure of the vocal tract at the lips, 'd' a closure at the alveolar ridge, and so forth. This notion came originally from research on speech that revealed vast context-conditioned variations in the sound as a result of the coarticulation of seemingly invariant gestures. Among the unconventional investigators who take such gestures to be the primitives of the phonetic system, there is some question about exactly how they should be defined, but little of what I mean to say here turns on the answer. What is most important for our purposes are just two considerations: one is that the gesture-as-primitive view permits us to see a system in which the defining gestures can be overlapped and merged (i.e., coarticulated) so as to produce phonetic strings at the high rates that are, in fact, achieved; the other, that the phonetically relevant gestures were presumably selected (and refined) in evolution because they lent themselves to just those articulatory and coarticulatory maneuvers that were appropriate to their specifically phonetic function. Accordingly, they form a natural class, a phonetic modality, as it were, that has a linguistic purpose and no other.

As for speech production, the conventional view is that it is controlled by mechanisms of a general motor sort, mechanisms that are constrained to produce exactly the sounds that define the consonants and vowels. According to the unconventional view, on the other hand, the mechanisms of articulation and coarticulation are not instances of some more general mechanism of motor control, but rather the workings of a biological specialization—a phonetic module—that is no less distinctly linguistic than the specialized gestures it manages. The aim of its specialized gestures is not to achieve particular acoustic targets, but to represent consonants and vowels invariantly in rapidly produced strings, allowing the resulting sounds to go wherever the acoustically complex effects of coarticulation happen to take them. That the articulation of consonants and vowels is, in fact, a biological specialization is plainly shown by the inability of nonhuman primates to learn to produce even the simplest syllables. They can't do it, not because they are not smart enough, or because they lack the appropriate pieces of anatomy, but because, being other than human, they are not endowed with a phonetic module.

Turning now to perception, we see, on the conventional view, only the most general processes of the auditory modality, which is to say that perception of consonants and vowels is supposed to be no different from perception of other sounds. All use the same mode of signal analysis, and evoke in the same perceptual register the same set of auditory primitives. Thus, the difference in perception between a consonant and some nonspeech sound is only in the particular mix of auditory primitives they comprise. They are made of the same perceptual stuff.

According to the unconventional theorist, on the other hand, the phonetic gestures are recovered in perception by the specialized phonetic module that controlled their production. Such a specialized process is necessary in order that proper account be taken of the specifically phonetic complications that coarticulation introduces into the relation between acoustic signal and the gestural message it conveys. Given that the message and the process that recovers it are both specific to phonetic communication, the resulting representation is specific to that kind of communication, too, which is to say that its modality is distinctly phonetic, not auditory.

There is one more important assumption of the conventional view, this one made necessary by the prior assumption that speech rests on motor and perceptual representations of some general sort. For it falls to anyone who holds that speech is supported in that way to explain how its initially nonphonetic representations are invested with phonetic significance, and so made appropriate for linguistic communication. The conventional explanation is that this is done at a cognitive stage, beyond action and perception, where the very ordinary motor and auditory representations are translated into units of a linguistic sort. There are various notions among the conventional theorists about exactly how that is done, but those seem to be distinctions without a difference, for they all come, necessarily, to the same thing: speaker and listener must, in effect, attach phonetic labels to their respective nonphonetic acts and percepts; neither party can experience phonetic representations at the level of action and perception, because phonetic representations are not supposed to exist there.

The unconventional theory needs no such assumption as the one just described, because it takes the primary representations of speaker and listener to be immediately phonetic; they are precognitive acts and percepts, not cognitive afterthoughts.

The implications for reading and writing. From what has so far been said about the two theories, it is clear that they see the relation of speech to reading/writing in drastically different ways. They must, of course, agree that reading/writing are not supported by a biological specialization at the level of act or percept—that is, in production or perception of the letters of the alphabet—and, accordingly, that the letters can take on linguistic significance only by virtue of being named after the consonants and vowels to which they have been arbitrarily assigned. Given that area of agreement between the theories, the critical difference hinges, then, on the clear implication of the more conventional theory that what is true of the link between signal and language in reading/writing must be true in speech, too: the primary acts and percepts of speech can be no more linguistic than those of reading/writing, and no less arbitrarily connected to language. Thus, the conventional view reduces the difference between speech and reading/writing to a matter of making or hearing sounds, in the one case, and drawing or seeing print, in the other.

On the unconventional theory, however, the difference between speech and reading/writing is profound. In contrast to the letters of the alphabet, the gestural representations that are the input to the phonetic module in production and its output in perception are, by their very nature, pieces of language, not arbitrary stand-ins. Accordingly, speaker and listener are immediately engaged in the language business in a way that writer and reader are not; the difference between making or hearing sounds, on the one hand, and drawing or seeing print, on the other—the only difference the conventional view allows—has precious little to do with the matter.

Consider, now, the implications of the contrasting views of speech for the questions we should answer if we would understand the reading process and the difficulties that some have with it.

Are writing/reading hard? The answer given by the conventional view of speech is: not really; no harder, certainly, than speech. That is, of course, exactly what the avatars of Whole Language take as their most fundamental premise (Goodman, 1986). Indeed, it may very well be the conventional theory of speech that initially emboldened them to promote a proposition that is so at odds with the most obvious facts, and so harmful to an understanding of how we ought to teach children to read. But then if they had really thought hard about the implications of the conventional theory, they would have been led to the even worse conclusion, if that were possible, that reading and writing must be, not just as easy as speech, but significantly easier. For if, as the conventional view would have it, the difference is only that between sound-tongue-ear for speech and print-finger-eye for reading/writing, then reading/writing has the advantage on all counts.

Taking, first, the nature of the signal, one quickly sees the superiority of print. Typically, the printed characters are crisp and clear; the signal-to-noise ratio could hardly be better. The speech signal, on the other hand, leaves much to be desired from a physical point of view, if only because much of the acoustic information that is most important for phonologic purposes is least prominent acoustically. As for the effectors—fingers versus tongue—the fingers win, and by a wide margin. For the moving finger writes, and having written, moves on to play Bach's Goldberg Variations or do brain surgery; in contrast, the moving tongue speaks, and having spoken, lapses into inactivity, except as it is occasionally called on to lick the lips or help in swallowing. Imagine fixing a stylus to your tongue and trying then to write your name. Turning finally to the receptors—the ear vs. the eye—I simply note that, as a channel for transmission of information, the eye is better than the ear by several orders of magnitude. How, then, are we to understand why it is that speech is, by every conceivable measure, the easier. Indeed, if linguistic communication were as the conventional view of speech says it is, then our concerns would be the exact reverse of what they are: having taken it for granted that reading and writing are dead easy, the members of this conference would be exchanging ideas about how to teach would-be speakers to overcome the difficulties caused by the evident limitations of tongue and ear, and what to do for those who can't manage. The unconventional view does not blink those shortcomings, but rather shows how speech, in a triumph of evolution over engineering, found ways around them. Special exertions by speaker and listener are not called for. What that means will become clearer below, where we consider the requirements of phonological communication and how they are met.

What is hard about writing and reading? Surely, we can't know how to teach a child to read or write except as we understand what he has to learn and why the learning might not be easy. But, as we have seen, the conventional view of speech tells us that reading/writing should be even easier than speech, which we already know to be quite easy, so the conventional view is not likely to be helpful at this very earliest stage of our inquiry. Let us, however, overlook that most unfortunate implication of the conventional view, and put our attention instead on what it reveals about the well-documented difficulty of grasping the alphabetic principle. For that purpose, we must digress a bit to consider the nature of phonological communication and the requirements it imposes.

Everyone understands that the function of the phonologic mode of communication is to generate an uncountably large number of words by variously combining and permuting the small number of meaningless segments we call consonants and vowels. That is the combinatorial principle that allows to language its property of openness or generativity, a property that is unique among natural communication systems; not surprisingly, then, it is the design feature that characterizes language at all levels, But if the principle is to work at the level of phonology, two requirements must be met, The more obvious is that the segments be commutable, which is to say discrete, invariant, and categorical. The possibly less obvious requirement derives from the fact that, if all utterances are to be formed by a small number of segments, then, inevitably, those segments will run to long strings, so it becomes essential that their production and perception be expeditious.

Now, on the conventional view, it is sounds and the ordinary auditory percepts they evoke that must have those critical properties, from which it follows that speech could only be an acoustic alphabet, offering a discrete, invariant, and categorical sound (and auditory percept) for each phonetic segment. Of course, the sounds would presumably be smoothed and connected at the places where they join, much as the shapes of cursive writing are, but, one way or another, there would have to be, for each segment, a commutable piece of sound. To produce such sounds, a speaker would necessarily make a discrete articulatory gesture for each one, in which case he could not produce a syllable like 'bag,' but only the three syllables, 'buh ah guh.' If speech had to be like that, it would come nowhere near meeting the requirements for commutability and speed, so a communication system that was generative at the level of word formation would not be possible.

Nor would things be much better if a means could somehow be found to deliver the alphabetic sounds more rapidly, for that would surely defeat the ear. The point is that speakers normally produce phonetic segments at rates that average about 10 or 12 per second and, for short periods, run up to 20 or 25. Now if each of those segments were represented by a discrete sound, as the conventional view says it must be, then rates that high would strain the temporal resolving power of the ear, and also overreach its ability to keep the order of the segments straight (Warren, 1993).

But even if we put all of the foregoing considerations aside, and assume that speech as portrayed by the conventional view could somehow be made to work, we are still not at all enlightened about why the alphabetic principle should not have been almost immediately apparent to those who lived before it was discovered, and why it is not equally apparent now to every normal child. All would have mastered a language that was conveyed, presumably, by an acoustic alphabet. Why, then, would they not already understand the alphabetic principle, and quickly learn to apply it in the visual modality simply by substituting the alphabetic letters for the correspondingly alphabetic sounds?

But if the conventional view leaves us puzzled about why it is hard to be aware that words come apart, it does suggest why some teachers might misunderstand how they are put together.

The misunderstanding manifests itself, and also begins to take its toll, when, having taught a child the 'spelling-to-sound rules,' the teacher urges him to 'blend'—that is, to form the alphabetic sounds, 'buh, ah, and guh,' for example, into the proper word 'bag.' I cannot presume to know what is in the mind of the teacher who tries to get the child to do that, but I suspect it is a resolution of the apparent conflict between what she believes about speech and what her ears tell her about how it sounds. With encouragement from the conventional view, she presumably believes that there are three sounds in 'bag,' and that these are represented by the letters 'b,' 'a,' and 'g.' However, I should think she would find it unsettling that she can't really hear three sounds, but rather something that is, from a purely auditory point of view, all of a piece. Perhaps, then, she supposes that the auditory appearances are deceiving, that the three sounds have been so thoroughly blended as to hide their individual identities. If so, then she is using the word 'blend' in its correct sense to mean a combination in which the constituent parts are indistinguishable; but she is imagining a most unfortunate contingency, for blending would cause language to lose its vital phonologic core, with the result that the combinatorial principle would no longer be available to produce vocabularies that are large and expandable. In any case, it is physically and physiologically impossible to produce a word by blending, or otherwise combining, the discrete sounds that are taken to be its individual phonetic constituents. So while 'blend' is the right word, it is the wrong idea.

I do not mean to suggest that implying to a child that 'bag' is a blend of three sounds is necessarily to court failure in reading and writing. It is rather to tell a white lie, and is better by far than characterizing the printed word as a picture, or advising the child to guess what the print says. Learning letter-to-sound correspondences and trying, on that basis, to 'sound out' words is likely, at least, to help bring the child to the correct understanding that words come apart, and that the alphabet has something to do with the parts. The error is in the belief that the parts are sounds. Most, but obviously not all, children who are taught that error manage somehow to rise above it, and so learn to read and write. Still, things would almost certainly go better if they were acquainted with the true state of affairs.

In contrast to the conventional theory, the unconventional account of these matters shows how phonological communication is possible, and, by the same token, why the alphabetic principle is hard to grasp. Remember that speakers are able to produce strings of phonetic segments at high rates, but only because the segments are gestures that are efficiently overlapped and merged. By that means, the speaker succeeds in producing phonologic structures that effectively 'spell' the words they convey. But there are several reasons why the illiterate or preliterate speaker nevertheless does not know how to spell, or even that words have a spelling. Perhaps the most obvious is that the phonetic module spells for the speaker. Once he has thought of the word, whatever that means, the phonetic module takes over, automatically selecting and coordinating the appropriate gestures. The speaker cannot know how the module did what it did, because it is true of all biological modules that their processes are not available to conscious inspection. On the other hand, the speaker <u>can</u> be aware of the representations the phonetic module deals with; but there is no reason he should be, because being inherently phonetic, the motor structures that are represented do not require translation, so they do not invite attention. And, finally, it is probably relevant that the mechanisms of articulation and coarticulation produce smoothed and context-sensitive movements at the surface, and so obscure the exact nature of the distal motor structures that are the actual phonetic units.

The relevant considerations are much the same in perception. There, coarticulation has allowed information about several successive segments to be conveyed simultaneously in the acoustic signal, and so relaxed the constraint on rate of perception imposed by the temporal resolving power of the ear. The listener can, therefore, keep up with the speaker, but only because his phonetic module is specialized to process the acoustic signal so as to extract the coarticulated gestures that produced its uniquely phonetic complications. A consequence is that the listener is likely to lack phonologic awareness for much the same reasons that keep a speaker in the dark, Though the signal is, in fact, parsed into its phonetic constituents, the listener is none the wiser, because the module runs on automatic in perception just as it does in production.

Deliberate, cognitive procedures are never necessary to do the job. Indeed, the job cannot be done cognitively, because the complexities of the speech code are apparently too great, too special to language, and too deep in our biology; certainly, no one has succeeded yet, though, given the intense and long-continued efforts to build an automatic speech recognizer, we know it is not for want of trying.

As for the representations that are the result of the module's efforts, they are already phonetic, as I've said so many times, hence perfectly appropriate for all further linguistic processing. Therefore, the listener does not have to give them the attention they would require if, like the letters of the alphabet, they had to be translated into pieces of language. Finally, coarticulation has destroyed anything that remotely resembles a straightforward relation between the segments of the phonetic message and such segments as can be found in the acoustic signal. A consequence is that the consonants, at least, have never in the listener's experience been isolated and pointed to, as words, for example, commonly are. Surely, that is one reason why preliterate children are more likely to be aware of words than of the phonologic segments that form them. None of this is to say that listeners cannot be aware of the phonologic constituents of words—indeed, if they could not, the use of alphabetic transcriptions would be impossible—only that the unconventional view shows us why such awareness does not come for free with mastery of speech.

What links writing to reading and speaking to listening? In linguistic communication, where every sender is a receiver and every receiver a sender, the processes of production and perception must somehow be linked. Mattingly and I have called this the 'requirement for parity,' and wondered how it is met (Liberman, 1996; Liberman & Mattingly, 1989).

In reading/writing, parity cannot be said to rest on a primary biological base, but must rather have been established by agreement. Somehow, those who developed an alphabet arrived at a compact that specified which optical shapes were relevant to language, and which piece of language each was relevant to. A result is that learning to read and write is largely a matter of mastering the arbitrary terms of that compact, and, for all the reasons the unconventional view has revealed, that is rather hard, and commonly requires help from a tutor.

In the matter of parity, as in all things, the conventional view of speech implies that as it is in reading/writing, so must it be in speech, in which case learning speech has got to be just like learning to read and write. So here again we find in the conventional view of speech full justification for one of the fundamental, and fundamentally wrong, assumptions of Whole Language—namely, that children should learn to read as they learned to speak, which is to say that the educational process should be geared to provide conditions just like those under which speech was acquired; children need not, and should not, be taught to analyze the language as linguists do (Goodman, 1986).

The unconventional view, on the other hand, claims that parity in speech does not derive from a compact of some kind, but rather reflects a fundamental aspect of our biology. For parity is exactly what evolved: the necessary link between production and perception is given immediately by the genetically determined phonetic module, which provides that the specifically phonetic motor structure in the mind of the speaker is reproduced in the mind of the listener; there is no need for the two parties to connect grossly dissimilar but equally nonphonetic acts and percepts that were, like the letters of the alphabet, selected by earlier generations and then arbitrarily assigned to phonetic categories. Thus, the phonetic module makes for a deep and immediate intimacy between speaker and listener, an intimacy as necessary for linguistic communication as that which sex affords is necessary for reproduction. An important difference, of course, is that the one proceeds from parity, while it is disparity that lies at the root of the other. But an equally important similarity is that both kinds of intimacy are the products of co-evolution, since the two sides of the connection had, in each case, to develop in step, change for change, else neither system could ever have become functional.

Though parity in speech is part of its underlying biology, it does not follow that speech is not learned, only that it need not be taught. For surely, the necessary and sufficient conditions for learning speech are but two: membership in the human race, and exposure to a mother tongue.

To get an idea of the nature of that kind of learning, it is helpful, I think, to see the phonetic module as one of a class of modules that have certain characteristics in common (Liberman, & Mattingly, 1989). One of those is plasticity over periods of time during which the module is shaped by environmental conditions. An example is the module for sound localization, which responds to interaural differences of time and intensity, using them as a basis for computing and then representing location in azimuth. Of course, those interaural differences change considerably as the head grows, and the distance between the ears increases, so the module must be continuously recalibrated. One might reasonably suppose that, in a somewhat similar way, the biologically coherent phonetic module is calibrated over several years by the phonetic environment in which it finds itself. In that case, the obvious effect of experience on speech would be to shape or hone a genetically determined system, not, as in the case of reading/writing, to provide the basis for acquiring arbitrary connections by processes of a cognitive sort. Given a normal environment, speech 'emerges,' in the terminology of Whole Language, but reading and writing most certainly do not. Thus, the unconventional view permits us to see that learning to speak and learning to read or write are fundamentally different processes.

Implications for reading disability. Accepting the conventional view of speech, the reading/writer researchers must believe, as I earlier said, that the phonologic segments are plainly displayed on the auditory surface, there for all to hear. Accordingly, such researchers have no way to see why it should be hard to be aware of those segments, and so to grasp the alphabetic principle. They can hardly be expected, then, to look to that difficulty for the causes of reading/writing disability, and indeed, they do not. Rather, they look where the conventional view most directly tells them to, which is at some aspect of the visual system. That system is the seemingly most promising target, because the substitution of eye for ear is virtually the only important difference between speech and reading/writing that the conventional view allows. So if reading proves to be hard, then it must be that some aspect of vision is at fault. Small wonder, then, that one or another aspect of visual function is the place where many of the theories of disability locate the problem (Geiger & Lettvin, 1988; Orton, 1937; Pavlides, 1985; Stein, 1988).

At the same time, the conventional view permits, if it does not actually encourage, the belief that the problem might be with the ear. That belief begins with the conventional assumption that speech is a string of brief sounds that follow each other in rapid succession. The problem, then, is that the auditory system of some children can't keep up. As a consequence, they have language problems, from which reading problems follow (Tallal, 1980). Now if speech were a string of acoustic segments, one for each phonologic segment, then it would be true that the relevant sounds would, indeed, be very brief, and would follow each other in rapid succession—so brief and in such rapid succession, that, as I earlier pointed out, sound segments would come along at rates between 10 and 25 per second. It is also true, as I said in the same context, that rates that high would strain the ability of everybody's auditory system, not just those of some unfortunate children. Fortunately, speech does not require people to do what their ears do poorly, which takes us now to the unconventional view and its radically different implications for how we might see disability in reading/writing.

Let us consider first the theory about disability just alluded to. On the unconventional view of speech, the known limitation on ability to perceive the order of brief sounds presented rapidly and in series is irrelevant to speech perception, because phonetic segments are not sounds, and speech is not a string of them. The unconventional view tells us that the true phonetic elements are gestures, and that their coarticulation smears the information for each one over a considerable stretch of the acoustic signal, overlapping it grossly with information for other segments. One important consequence is that ordinal position is marked, not by the temporal order of the sounds, but by their acoustic shapes. Thus, the syllables ba and ab, when pronounced briefly, have acoustic patterns in which information about consonant and vowel are completely overlapped. Accordingly, there are not two acoustic segments—one for each phoneme -- hence no way to perceive which came first by paying attention to the way sounds succeed each other in time. Nevertheless, the listener infallibly knows which came first and which second because the acoustic shapes of the two syllables are very different. In fact, in that case, they are exact mirror images. Given the services of the phonetic module, which are always at the disposal of the listener, the one shape is perceived as an opening gesture (consonant first, vowel second), the other as a closing gesture (vowel first, consonant second).

Having just seen what the unconventional view says is not true of speech, and therefore of a theory that locates the cause of reading/writing disability in the ear, I turn now to what it says is true, and how that gives us an entirely different slant on the probable causes of failure. We earlier saw how the unconventional view shows that learning to speak, however fluently, will not be sufficient to produce awareness of phonologic structure. Acting on precisely that consideration, Isabelle Liberman, Donald Shankweiler, Ignatius Mattingly, and their colleagues began the line of thought that led them to find that phonologic awareness is, in fact, largely absent in preliterate children (Liberman, 1973; Mattingly, 1972; Liberman, Shankweiler, Fischer, & Carter, 1974). Subsequent research by them and others amply confirmed that finding, while also establishing that the extent to which awareness is present counts as one of the best predictors of success in reading/writing (for reviews, see Blachman, 1989; Routh & Fox, 1984), and that training in awareness has generally happy consequences for those who get it (Bradley & Bryant, 1983; Content et al., 1986; Ball & Blachman, 1988; Lundberg, Frost, & Peterson, 1988; Olofsson & Lundberg, 1983; Vellutino & Scanlon, 1987).

Proceeding further with the implications of the unconventional view, researchers picked up on its assumption that there is a distinct phonological faculty—I have here called it a phonetic module—that is independent of cognition and, indeed, of all other-than-linguistic modes of production and perception. They found it reasonable to suppose that if such a faculty exists—though the conventional view provides no place for it—then it might work more or less well among otherwise normal children, with the result that there would be differences in the ease with which they would learn to read and write (Liberman, I. Y., & Liberman, A. M., 1990; Liberman, A. M., 1992; Liberman, I. Y., Shankweiler, D., & Liberman, A. M., 1989; Brady, S. A., 1991). Most obviously, the effect would be on the general quality or clarity of the phonologic representations, which would, in turn, affect the child's ability to become aware of them, and so to comprehend and apply the alphabetic principle. One would expect, too, an effect on the phonologic basis of the working memory that is an integral part of syntactic processing, and therefore on the child's ability to comprehend at the level of the sentence. Indeed, everything about language or reading/writing that depends on phonologic structures and processes would presumably be affected in some way. Exactly how, and with what consequences, are questions that motivate the research our colleagues are now actively pursuing. What is reasonably clear at this point is only that the leads afforded by the unconventional view are promising, and that the relevant research on the role of specifically phonologic processes is nearer its beginning than its end. It is very hard, I think, to see how we should ever have arrived at that beginning if researchers had remained true to the conventional view of speech. On the other hand, the hypothesis that phonologic factors deserve careful attention is now common enough that researchers may have lost sight of what the unconventional view of speech had to do with it.

REFERENCES

Ball, E. W., & Blachman, B. A. (1988). Phoneme segmentation training: Effect on reading readiness. *Annals of Dyslexia, 38*, 208-225.

Blachman, B. A. (1989). Phonological awareness and word recognition: Assessment and intervention. In A. G. Kamhi and H. W. Catts (Eds.), *Reading disorder: A developmental language perspective*. San Diego, CA: College-Hill Press.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301*, 419-421.

Brady, S. A. (1991). The role of working memory in reading disability. In S. A. Brady & D. P. Shankweiler (Ed.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 129-151). Hillsdale, NJ: Lawrence Erlbaum Associates.

Content, A., Kolinsky, R., Morais, J., & Bertelson, P. (1986). Phonetic segmentation in pre-readers: Effects of corrective information. *Journal of Experimental Child Psychology, 42*, 49-72.

Crowder, R. G. (1983). The purity of auditory memory. *Phil Trans R Soc Lond, B302*, 251-265.

Diehl, R. L., & Kluender, K. R. (1989). On the categorization of speech sounds. In S. Harnad (Ed.), *Categorical perception* . Cambridge: Cambridge University Press.
Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121-144.
Diehl, R. L., & Kluender, K. R. (1989). Reply to Commentators. *Ecological Psychology, 1(2),* 195-225.
Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute.* Faculty of Engineering, University of Tokyo, Vol. 29, 207-214.
Geiger, G., & Lettvin, J. Y. (1988). Dyslexia and reading as examples of alternative visual strategies. In C. von Euler, I. Lundberg, & G. Lennerstrand (Ed.), *Brain and reading (Wenner-Gren Symposium Series 54)* . New York: Macmillan.
Goodman, K. S. (1986). *What's whole in whole language: A parent-teacher guide.* Portsmouth, NH: Heinemann.
Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70,* 340-349.
Liberman, A. M. (1996). *Speech: A special code.* Cambridge, MA: MIT Press.
Liberman, A. M. (1992). The relation of speech to reading and writing. In R. Frost & L. Katz (Ed.), *Orthography, phonology, morphology, and meaning* (pp. 167-178). Amsterdam: Elsevier Science Publishers B. V.
Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.
Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243,* 489-494.
Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Haskins Laboratories Status Report on Speech Research, SR-33,* 157-166.
Liberman, I. Y., Shankweiler, D., Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle.* IARLD Research Monograph Series. Ann Arbor: University of Michigan Press.
Liberman, I. Y., & Liberman, A. M. (1990). Whole language vs. code emphasis: Underlying assumptions and their implications for reading instruction. *Annals of Dyslexia, 40,* 51-76.
Liberman, I. Y., Shankweiler, D. P., Fisher, F. W., & Carter, B. J. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18,* 201-212.
Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1-33). Ann Arbor, MI: University of Michigan Press.
Lindblom, B. (1991). The status of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Ed.), *Modularity and the Motor Theory of speech perception* (pp. 7-24). Hillsdale, NJ: Lawrence Erlbaum Associates.
Lundberg, I., Frost, J., & Peterson, O.-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263-284.
Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14,* 211-235.
Massaro, D. W. (1987). Psychophysics versus specialized processes in speech perception: An alternative perspective. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 46-65). Boston: Nijhoff.
Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.
Mattingly, I. G. (1972). *Reading, the linguistic process, and linguistic awareness.* Cambridge, MA: MIT Press.
Mattingly, I. G. (1972). Speech cues and sign stimuli: An ethological view of speech perception and the origin of language. *American Scientist, 60(3),* 327-337.
Mattingly, I. G., & Liberman, A. M. (1990). Speech and other auditory modules. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Ed.), *Signal and sense: Local and global order in perceptual maps* (pp. 501-519). New York: John Wiley & Sons.
Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? *Scandinavian Journal of Psychology, 24,* 35-44.
Orton, S. T. (1937). Reading, writing and speech problems in children. New York: Norton.
Pavlides, G. T. (1985). Eye movement differences between dyslexics, normal, and retarded readers while sequentially fixating digits. *American Journal of Opto and Physiological Optics, 62,* 820-832.
Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129-156.
Routh, P. K., & Fox, B. (1984). M...m is a little bit of May: Phonemes, reading and spelling. In N. D. Gadow & P. Bralen (Ed.), *Advances in learning and behavioral disabilities* . Greenwich, CT: JAI Press.
Stein, J., Riddell, P., & Fowler, S. (1988). Disordered right hemisphere function in developmental dyslexia. In C. v. Euler, I. Lundberg, & G. Lennerstrand (Ed.), *Brain and Reading* . New York: Macmillan.
Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64,* 1358-1368.

Sussman, H. (1989). Neural coding of relational invariance in speech perception: Human language analogs to the barn owl. *Psychological Review, 96*, 631-42.

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language, 9*, 182-198.

Vellutino, F. R., & Scanlon, D. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33*, 321-363.

Warren, R. M. (1993). Perception of acoustic sequences: Global integration versus temporal resolution. In S. McAdams & E. Bigand (Eds.), *Thinking in sound*. Oxford: Oxford University Press.

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237*, 169-171.

FOOTNOTE

*To appear in B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention.* Hillsdale, NJ: Erlbaum Associates.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 197-219

New Evidence for Phonological Processing during Visual Word Recognition: The Case of Arabic*

Shlomo Bentin† and Raphiq Ibrahim‡

Lexical decision and naming performance were examined using visually presented words and pseudowords in literary Arabic as well as transliterations of words in a Palestinian spoken dialect which has no written form. Although the transliterations were completely unfamiliar visual stimuli, and in some cases their phonologic structure violated the phonology of literary Arabic (the only form of Arabic which can be legally written), they were not easily rejected in the lexical decision task and more slowly accepted in a phonologically based lexical decision task. Naming transliterations of spoken words was delayed relative to naming literary words and also delayed relative to pseudowords. This pattern suggests that phonological computation aimed at retrieving the phonological structure of the word is mandatory for lexical decision as well as for naming. A very sizeable frequency effect in lexical decision as well as in naming which was three times as big for literary words than for the transliterations suggest that addressed phonology is an option for very familiar orthographic patterns. Moreover, the frequency effect on processing transliterations indicated that lexical phonology is involved with pre-lexical phonological computation even if addressed phonology is not possible. These data support a combination between a cascade type process, in which partial products of the grapheme-to-phoneme translation activate phonologic unites in the lexicon, and an interactive model in which the activated lexical units feed-back shaping the pre-lexical phonological computation process.

Several models of visual word recognition have proposed that fluent readers do not use the phonological information conveyed by printed words, until after their meaning has been identified, (e.g., Banks, Oka, & Shugarman, 1981; Jared & Seidenberg, 1991; Paap, Newsome, McDonald, & Schvaneveldt, 1982; Saffran & Marin, 1977). Accordingly, the term "post-lexical" phonology has been used to denote the idea that the phonological lexicon is accessed via a top-down process initiated by the activation of a semantic node (Besner, Davis, & Daniels, 1981; Foss & Blank, 1980; Patterson & Coltheart, 1987). In their extreme forms, such models assume that, although orthographic units may automatically activate phonological units in parallel with the activation of meaning, lexical access and the recognition of printed words may be mediated exclusively by orthographic word-unit attractors in a parallel distributed network (if one takes a connectionist approach, e.g., Hinton & Shallice, 1991; Seidenberg & McClelland, 1989) or by a visual logogen system (if one prefers a more traditional view, e.g., Morton & Patterson, 1980).

Much of the empirical evidence supporting the orthographic-semantic models of word recognition comes from the neuropsychological laboratory. For example, patients with a form of acquired alexia labeled "deep dyslexia" apparently cannot use grapheme-to-phoneme translation, yet they are able to identify printed high-frequency words (Patterson, 1981). Furthermore, the

This study was supported by a grant from the Israel Foundations Trustees to Shlomo Bentin and partly by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. Raphiq Ibrahim was supported by a stipend from the Joseph Meyerhoff Foundation. We thank Drs. D. Besner, R. Frost, C. Perfetti, A. Pollatsek and one anonymous reviewer for very constructive comments on earlier drafts.

reading errors made by such patients are predominantly semantic paralexias and visual confusions (for a review, see Coltheart, 1980). These data were therefore interpreted as reflecting identification of printed words via their whole-word visual/orthographic (rather than phonologic) structure. The propriety of generalizing these data to normal reading is questionable, but additional support for the orthographic-semantic view can also be found in studies of normal word recognition. For example, in Hebrew (as in Arabic) letters represent mostly consonants while vowels may be represented in print by a set of diacritical marks (points). These points are frequently not printed and under these circumstances isolated words are phonologically and semantically ambiguous. Nevertheless, it has been found that, in both Hebrew (Bentin, Bargai & Katz, 1984) and Arabic (Roman & Pavard, 1987), the addition of phonological by disambiguating vowel points inhibits (rather than facilitates) lexical decision. On the basis of such results, it has been suggested that, at least in Hebrew, correct lexical decisions may be initiated on the basis of orthographic codes before a particular phonological unit has been accessed (Bentin & Frost, 1987). In English, a distinction has been made between frequent and infrequent words. Whereas it is usually accepted that phonological processing is required to identify infrequent words; frequent words are presumed to be identified on the basis of their familiar orthographic pattern (Seidenberg, 1985a).

Advocates of phonological mediation on the other hand, claim that access to semantic memory is necessarily mediated by phonology (e.g., Frost, 1995; Liberman, 1992; Liberman & Liberman, 1990). In a "weaker" form of the phonological-mediation view it is suggested that, although the phonologic structure may not necessarily be a vehicle for semantic access, it is automatically activated and integrated in the process of word recognition (Van Orden, 1991; Van Orden, Pennington, & Stone, 1991). Such models assume that phonological entries in the lexicon can either be accessed by assembling the phonological structures at a pre-lexical level, or addressed directly from print, using whole-word orthographic patterns. The problem of orthographic-phonemic irregularity is thus solved by acceptance of the concept of addressed phonology. Indeed, cross-language comparisons indicate that addressed phonology is the preferred strategy for naming printed words in deep orthographies (Frost, Katz, & Bentin, 1987, but see Frost, 1995).

Given that all of the above strategies are in principle possible, the focus of most contemporary studies of word recognition has shifted from attempting to determine which of the above theories is better supported by empirical evidence, to understanding how the different kinds of information provided by printed words interact during word recognition (e.g., Taraban & McClelland, 1987). Along these lines, one aim of the present study was to examine whether the reader has the option of ignoring the phonological information provided by printed stimuli when such information may interfere with efficient performance. To achieve this aim we took advantage of a specific property found in the Arabic language in which the spoken dialects are not used in print. A second aim of the present study was to examine word recognition processes in a language has some unique features and has not been extensively investigated. Comparisons of reading Arabic and French suggest that word recognition processes may be slightly different in these two languages, possibly because of the additional morphologic complexity of Arabic relative to French (Courrieu & Do, 1987; Farid & Grainger, in press).

The Arabic language has two major forms. One, literary Arabic, is universally used throughout the Arab world in all written texts, from the Koran to modern newspapers. Literary Arabic is not, however, used in mundane speech communication. For ordinary speech there are spoken dialects that differ across different Arab countries (and often across different regions within one country). These dialects are the mother tongue of the great majority of native speakers of Arabic, while the literary form is first learned in school. Although a subset of words are similarly pronounced and have the same meaning in both languages, literary and spoken Arabic are phonologically different. In addition to their having different lexica, there are phonological structures that may appear in only one of the two forms. For example, none of the literary words may start with a sequence of two consonants or with a consonant and a schwa (the neutral vowel), whereas many spoken words do. In addition, there are vowels that are pronounced differently in each language. For example, the vowels /o/ and /ɛ/ are used only in spoken Arabic; in literary Arabic they are pronounced /au/ or /u/ and /aɛ/ or /i/, depending on phonetic context.

The orthography of literary Arabic is visually complex. Consonants are represented by letters and frequently include diacritic marks. Vowels are usually represented by diacritic marks, although, as in other Semitic languages, some vowels are also represented by letters. Thus, like Hebrew, if all the diacritics are presented, Arabic orthography is phonologically transparent. However, if the vowel dots are missing, the print becomes phonologically opaque, at least to some extent. Printed material in Arabic usually includes all consonantal diacritic marks but only those vowels that were necessary for unequivocal reading as a meaningful word (see examples in Figure 1).

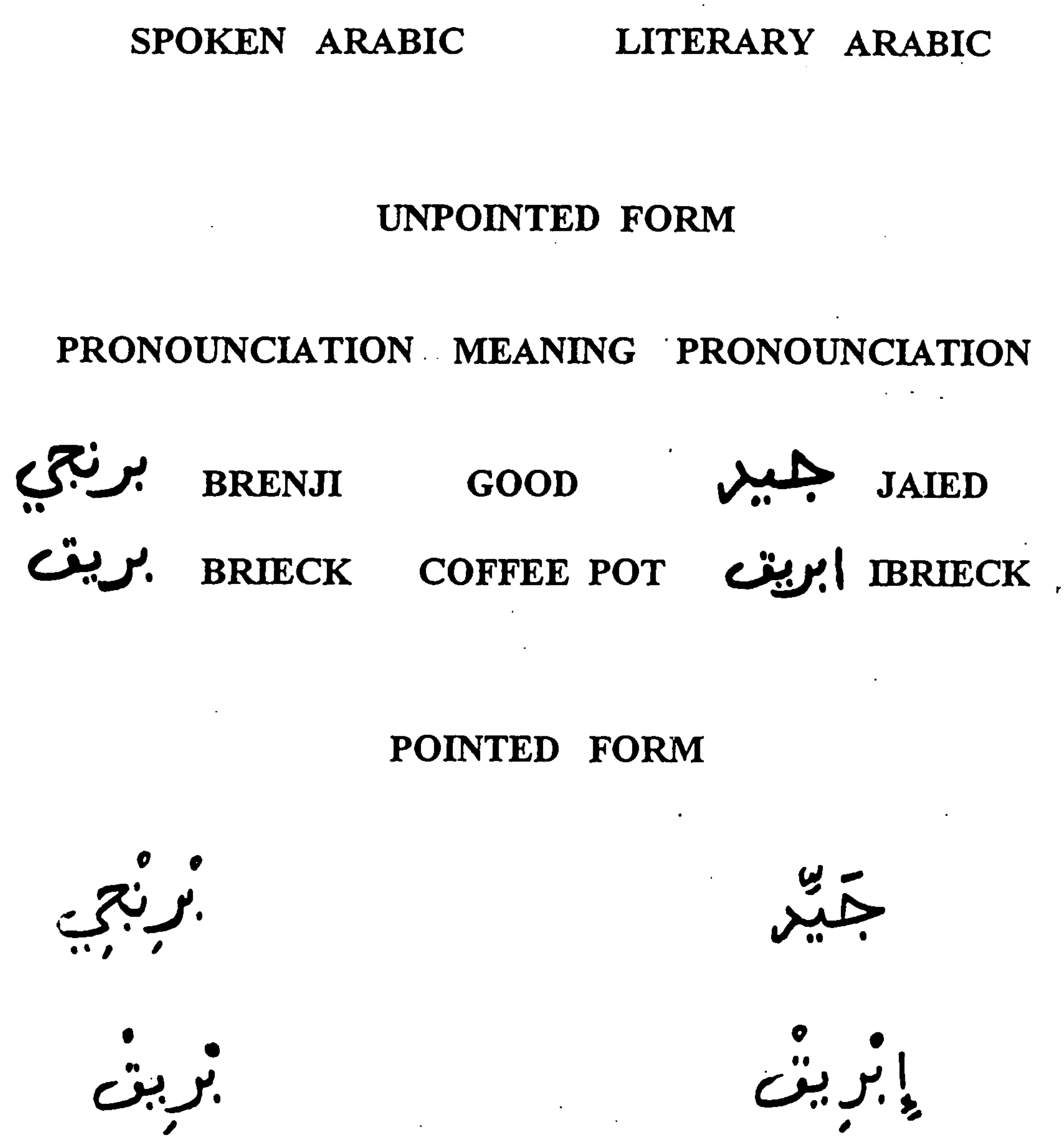


Figure 1. Examples of literary Arabic printed words and transliterations of words in spoken Arabic.

Because letter-to-phoneme translation is regular in Arabic orthography, it is possible to present spoken words in a printed form by using transliterations based on phoneme-to-letter transformations. Such an orthographic pattern would be very unfamiliar to readers of Arabic, but if they reverse the translation process, i.e. if they use grapheme-to-phoneme transformations, the resulting assembled phonological unit should match a phonological lexical entry. The effect of presenting such stimuli in a lexical decision task should, therefore, depend upon the nature of the word recognition process. If lexical decision may be based solely on the orthographic pattern, unless subjects are specifically instructed to accept all stimuli that sound like words, transliterated spoken words should be processed as very unfamiliar (or illegal) nonwords i.e., rejected very fast, faster than phonologically and orthographically legal nonwords (pseudowords). However, if in lexical decision subjects process the phonological information conveyed by the print, the transliterations should pose a particular problem. On the one hand they are unfamiliar orthographic patterns, but on the other hand they "sound" like real words, albeit in spoken, not in literary Arabic. Thus, these specific stimuli may have an effect similar to the pseudohomophone effect described in English—that is, they should be rejected more slowly than pseudowords (Coltheart, Davelaar, Jonasson, & Besner, 1977; Gough & Cosky, 1977; Rubenstein, Lewis, & Rubenstein, 1971; for similar effects in Hebrew see Bentin et al., 1984 - Experiment 2). Such a delay could be explained by assuming that the phonological information extracted from these letter strings activates a lexical entry and rejection is based on a post-lexical orthographic check (e.g., Dennis, Besner, & Davelaar, 1985), or by assuming that phonological and orthographic information are pooled in a pre-lexical logogen system and that the partial activation initiated by the matching phonology postpones "no" decisions (e.g., Coltheart et al., 1977).

Predictions about naming transliterations of spoken Arabic words were also theory dependent. There is ample evidence that words are named faster than nonwords. This difference has traditionally been explained by assuming that words may access the lexicon "directly" using whole-word orthographic codes, thereby immediately accessing whole-word phonological information. In contrast, the pronunciation of nonwords must be based on a longer and less efficient process of pre-lexical phonological assembling (e.g., Coltheart, Besner, Jonasson, & Davelaar, 1979; Frederiksen & Kroll, 1976; Seidenberg et al., 1984). Conforming to such a theory, because the orthographic pattern of the transliterations was (at least) as unfamiliar as the orthographic pattern of the nonwords, transliterations should have been named as rapidly as pseudowords, and both more slowly than literary words. On the other hand, more recent theories and data suggest that lexical information may support pre-lexical phonological assembly in naming (e.g., Besner, & Smith, 1992; Carello, Turvey, & Lukatela, 1992; 1994; Frost, 1995). For example, there is evidence that pseudohomophones are named faster than orthographically similar nonhomophonic nonwords (McCann & Besner, 1987). Accordingly, transliterations should be named faster than pseudowords.

In three experiments we examined the processing of words in literary Arabic, legal nonwords (pseudowords) produced by substituting letters in literary Arabic words, and orthographically presented spoken words (transliterations) formed by using Arabic letters to stand for their associated phonemes.

GENERAL METHOD

Subjects. The subjects were 60 high-school seniors (30 boys and 30 girls), all native speakers of Arabic (Palestinian dialect) attending a school in which Arabic is the official language. High-school pupils were chosen because many undergraduate students in Israeli universities read Hebrew and English more than Arabic. All subjects were volunteers.

Stimuli and Materials. All stimuli were hand-written by a skilled native speaker of Arabic and scanned for presentation by a Macintosh SE computer. All stimuli were strings of 3-6 characters and included the diacritical marks that were part of the consonants as well as some of the vowels. The included vowels were attached to the initial letters unequivocally specifying a meaningful reading (see Appendices). However, not all the vowels were included. There were four stimulus categories: 1) words used in both literary and spoken Arabic; 2) words existing only in literary Arabic; 3) phonetic transliterations of words that exist only in spoken Arabic; and 4) pseudowords, i.e. letter strings that were constructed replacing one or two letters in literary words. Hence, the pseudowords were phonologically and orthographically legal in both forms of Arabic but had no meaning in either of them. About 1/3 of the phonetic transliterations included structures that were phonologically illegal in literary Arabic (words beginning with two consonants or a consonant and a schwa).

The three word categories were further categorized as high- or low-frequency. In the absence of a computerized word-frequency count in Arabic, frequency was determined empirically by asking 50 high-school students (who did not participate in the present experiments) to rate the frequency of 480 letter strings. A scale of 1 (very infrequent) to 7 (very frequent) was used. The stimuli were presented for frequency rating in two lists. One included words that exist either only in literary Arabic or in both literary and spoken Arabic. The other list included words that exist only in the spoken dialect, and thus had no written form. Before rating the spoken-only words the subjects were instructed to use grapheme-to-phoneme translation and imagine the spoken word that was represented by the print. On the basis of this rating the high-frequency words selected for the three categories in this study had mean ratings of 6.37, 4.85, and 4.88, for the literary and spoken, literary-only, and spoken-only categories respectively, while the low frequency words had mean ratings of 2.95, 2.04, and 2.04. The mean length of the stimuli on the screen was 4 cm (ranging from 1.5 cm to 6 cm), seen from a distance of about 70 cm.

Procedure. Performance in both lexical decision and naming was examined in the first two experiments, whereas only lexical decision was examined in the third experiment. In the lexical decision task, subjects pressed one button with their right-hand index finger for positive answers and another button with their left-hand index finger for negative answers. Naming onset was measured from stimulus onset, using a voice key. The reaction times (RTs) were measured to the nearest ms by the computer. Only the RTs to correct responses were included in the analyses.

The experiments were conducted at the school in a relatively quiet classroom. After the instructions were given, ten practice trials, and a "ready" signal preceded each test list. Once the ready signal was on the screen, subjects could initiate the test list by pressing a button. The stimuli remained on the screen until a response had been given or for 2.5 seconds. The ISI between stimuli was 2.5 seconds. Errors were recorded by the computer in the lexical decision task and by the experimenter in the naming task. Because the same stimuli were used for both naming and lexical decision, different subjects were tested for each task. The same subjects were examined in Experiments 1 and 2. Each subject was randomly assigned either to lexical decision or to naming tasks. Half of the subjects began the session with Experiment 1 and the other half with Experiment 2.

EXPERIMENT 1

The words used in Experiment 1 were selected from the subset of words that are shared by spoken and literary Arabic. Thus, the subjects' performance in this experiment could be compared with that in most other languages in which lexical decision and naming have been investigated. On the basis of previous studies of lexical decision and naming performance using pointed and unpointed Hebrew words (Bentin & Frost, 1987; Frost, 1994), we predicted that both naming and lexical decision would be faster for high-frequency than for low-frequency words and slowest for pseudowords.

Method

Subjects. The subjects were 40 high-school seniors, 20 boys and 20 girls. Half of them were instructed to make lexical decisions for the stimuli, and the other half to name the same stimuli.

Stimuli. Ninety-six different stimuli were used, 48 words and 48 pseudowords. The words were from among the subset used in both spoken and literary Arabic. Among them, 24 were high-frequency and 24 were low-frequency. One high-frequency word, 4 low-frequency words and 3 pseudowords had a vowel at onset. The initial consonants in the high-frequency group were 10 stops, 12 fricatives, and a semivowel, and in the low-frequency group 14 stops, 4 fricatives, and 2 semivowels. The mean number of characters per word was not significantly different among stimulus groups (3.8, 4.0 and 3.7 for high-frequency words, low-frequency words and pseudowords, respectively), and the orthographic redundancy (i.e., the number of "neighbors," defined as the value "N" representing the numbers of different words that can be formed by changing only one letter in each stimulus, Coltheart et al., 1977; see also McClelland & Rumelhart, 1981) was similar across groups (1.45, 1.37 and 1.62 for high-frequency words, low-frequency words and pseudowords, respectively). The Arabic stimuli, their pronunciation, and their English translations are presented in Appendix A.

Results

Mean RTs for correct responses and percentage of errors were calculated for each subject, separately for high- and low-frequency words and for pseudowords. RTs that were above or below 2 standard deviations (SD) from the subject's mean in each condition were excluded and the mean was re-calculated. About 2% of the trials were excluded by this procedure. These data are presented in Table 1.

Because the method for collecting RT data was different for naming and lexical decision, these data were analyzed separately for each task. For each task we have analyzed the stimulus type effect within subjects (F1) and between stimulus-types (F2). Although the difference between the average number of letters per stimulus was similar across stimulus types, because low-frequency words had slightly more letters per word than high-frequency words and pseudowords, the stimulus analysis included the number of letters per stimulus as a covariate.

In the lexical decision task the stimulus type effect was significant, $F1(2,38) = 26.8$, $MS_e = 67029$, $p < .001$, $F2(2,92) = 76.3$, $MS_e = 30210$, $p < .001$. The influence of the stimulus length covariate on the main effect was not significant [$F2(1,92) = 2.0$ $p > .15$]. Post hoc (Tukey-A) comparisons revealed that while the decisions were significantly faster for high-frequency words than for the other two stimulus types, low frequency words were not significantly faster than pseudowords.

A similar pattern of effects was found for naming. The stimulus-type effect was highly significant $F2(2,38) = 51.3$, $MS_e = 5467$, $p < .001$, $F2(2,92) = 31.0$, $MS_e = 12760$, $p < .001$. The stimulus length covariate had no influence on the main effect [$F2(1,92) < 1.00$]. Post hoc (Tukey-A) comparisons revealed that naming high-frequency words was faster than naming both low-frequency words and pseudowords. The difference in the speed of naming low-frequency words and pseudowords was significant in the subject analysis ($p < .05$) but not in the stimulus analysis.

The percentages of errors in naming and lexical decision and in each stimulus category were compared using a two-way ANOVA. This analysis showed that more errors were made in the naming (6.4%) than in the lexical decision task (4.7%) [$F1(1,38) = 4.16$, $MS_e = 1.5$, $p < .05$], but a significant interaction between the task and the stimulus type effects [$F1(2,76 = 27.13$, $MS_e = 1.3$, $p < .0001$] and post hoc comparisons revealed that for high-frequency words there were more errors in the lexical decision than in the naming task, while for low-frequency words there were more errors in naming than in the lexical decision task. Finally, for pseudowords the percentage of errors in the two tasks was similar.

Table 1. *Mean reaction time (SEm) in milliseconds and percentage of errors for words that exist in both literary Arabic and the Palestinian spoken dialect and for pseudowords in the lexical decision and naming tasks.*

		WORDS		PSEUDOWORDS
WORD FREQUENCY		HIGH	LOW	
LEXICAL DECISION	RT	614 (12)	1078 (44)	1133 (27)
	ERRORS	5.2% (0.5)	4.8% (0.5)	4.2% (0.5)
NAMING	RT	634 (15)	815 (27)	856 (17)
	ERRORS	0% (0.0)	15.0% (1.7)	4.2% (0.7)

Discussion

The general trend of the results of the present experiment resembled that found in similar studies conducted in other languages but several interesting specificities were found. The most interesting aspect of these data was the unusually large word-frequency effect found in both tasks (464 ms for lexical decision and 181 ms for naming). This large frequency effect was not expected and therefore any explanation must necessarily be post hoc. A possible interpretation is suggested by the fact that, overall, the response times in both tasks were relatively longer than those reported in similar studies conducted in many other languages (particularly for the low-frequency words) and frequency effects might be a proportion of the overall response time. In addition, it is possible that the relative slowness of the Arabic native speakers in these visual word processing tests reflected a situation in which the subjects read a language which they do not usually use and have not mastered well. The previous frequency ratings obtained from other pupils from the same population and the relatively normal percentage of errors in lexical decision suggested that the subjects did recognize most of the words. It is possible however, that the experience that they had with reading the infrequent words was minimal, by far smaller than that typical in other studies. The statistical similarity between the performance with low-frequency words and pseudowords supports the latter interpretation.

The very large frequency effect in naming, albeit considerably smaller than in lexical decision, was also unprecedented. Such a large effect was particularly unexpected because, although not all the diacritics symbolizing vowels were attached to the consonants, the script included sufficient information for an unequivocal meaningful reading. Therefore this pattern contradicts previous reports in which, if the orthography was sufficiently shallow (i.e., the print provided sufficient information to enable pre-lexical assembling of the phonological structure) frequency effects in naming have been small or inexistent, (e.g., Frost, 1994; Frost et al., 1987; Katz & Feldman, 1983). In a nutshell, this sizeable word-frequency effect suggests that lexical phonological information was used to facilitate naming in literary Arabic. We will elaborate and discuss the implication of this suggestion in the discussion of Experiment 2 and in the General Discussion.

EXPERIMENT 2

The stimuli in the present experiment were: a) orthographic patterns that represent words in literary Arabic but do not exist in the spoken dialect; b) transliterations of words in spoken Arabic that do not exist in literary Arabic; and c) pseudowords i.e., orthographic patterns that were phonologically and orthographically legal in literary Arabic but had no meaning in either of the two forms of the language. The same stimuli were used in both the lexical decision and the naming tasks with different subjects assigned to each task.

In the lexical decision task the subjects were instructed to "accept" only words in literary Arabic and "reject" all other stimuli. Because spoken Arabic is never written in Israel, the transliteration of the spoken Arabic words formed orthographic patterns that were very unfamiliar. Moreover, about 1/3 of these patterns contained phonological combinations that are illegal in literary Arabic (see above). Hence, these stimuli may be considered analogous to phonologically illegal nonwords in English. Consequently, if the categorical decision between words and nonwords is based purely on the familiarity of the orthographic patterns, transliterations of spoken words should be rejected easily and at least as fast as pseudowords. On the other hand, if lexical decision in Arabic involves some phonological computation, transliterations of spoken words might access the phonological lexicon, thus inhibiting their rejection. Such an effect might, in fact, be expected given the similarity of this condition to pseudohomophones in visual lexical decision. As mentioned above, previous studies have reported that non-words that sound like words (e.g., brane) take more time to reject in lexical decision than orthographically similar nonwords that do not sound like words (e.g., brate) (Rubenstein et al., 1971). On the other hand, unlike the presently used transliterations, the pseudohomophones used in previous studies sounded like words in the same language in which the real words were presented. Therefore, a pseudohomophone effect could not be a priori predicted for these stimuli without some caution.

Naming performance for literary words and for pseudowords was expected to be similar to that observed in Experiment 1: high-frequency words should be named faster than low-frequency words and both faster than pseudowords. In addition, because the orthographic pattern of the transliterations was totally unfamiliar, literary words should be named faster than spoken words. According to the accumulating evidence supporting lexical involvement in prelexical assembly of phonological codes (Besner & Smith, 1992; McCann & Besner, 1987), transliterations should be named faster than pseudowords. On the other hand, because the transliterations represented words in a language that was different than the one in which the "real" words were presented, and their printed form was not only unfamiliar but also "strange" looking (including orthographic sequences that are totally inexistent in literary Arabic), it was possible that the transliterations of spoken words would be named as slowly as pseudowords, and the frequency of the spoken words should not affect naming performance.

Method

Subjects. The subjects were the same 40 pupils who were tested in Experiment 1. Subjects participated either in the lexical decision or naming task in both experiments.

Stimuli. The stimuli were 24 high-frequency and 24 low-frequency literary words, 24 transliterations of spoken words (12 high-frequency and 12 low-frequency words), and 24 pseudowords. From the subjects' point of view, however, there were only two equally represented stimulus categories (legally written, i.e., literary) WORDS, and NONWORDS (including the spoken words).

Among the words in literary Arabic, 2 high-frequency and 7 low-frequency words began with a vowel. The initial consonants in the other high-frequency literary words were 14 stops, 4 fricatives, and 4 semivowels. Among the low frequency literary words that did not began with a vowel 12 began with a stop consonant and 5 with a fricative. Among the high-frequency transliterations, 1 began with a vowel, 1 with a semivowel, and 10 with stop consonants. Among the low-frequency transliterations, 1 began with a vowel, 7 with stop consonants and 4 with fricatives. Out of the 24 transliterations, 3 high-frequency and 4 low-frequency began with letter combinations that, in the literary Arabic are not existent, i.e., were phonologically illegal. The mean word length was similar across groups, 4.5, 4.4, 4.0, 4.2, and 3.8 letters for high-frequency literary words, low-frequency literary words, high-frequency transliterations, low frequency transliterations and pseudowords, respectively. There was no significant difference in orthographic redundancy across groups (The mean "N" values were 0.96 and 1.04 for high- and low-frequency literary words, 0.92 and 0.88 for high- and low-frequency transliterations, and 1.08 for pseudowords, respectively). These stimuli are presented in Appendix B.

Procedure. The procedure was the same as in Experiment 1. In the lexical decision task the instructions indicated the possibility that some of the "nonwords" might have meaning in spoken Arabic but that these "odd" stimuli should be rejected. In the naming task, the nature of the stimuli was also explained, but the subject was instructed simply to read the pattern presented on the screen as fast as he or she could. Each task began with 10 practice trials that included stimuli of all the kinds.

Results

The RTs were averaged for each stimulus across subjects and for each subject according to five stimulus categories: high-frequency literary words, low-frequency literary words, high-frequency transliterations, low-frequency transliterations, and pseudowords. RTs that were above or below 2 SD from the subject or the stimulus mean in each category were excluded, and the mean was recalculated. Less than 3% of the stimuli were outliers, equally distributed among stimulus categories.

For both tasks, the RTs to transliterations of spoken words were slower than to pseudowords, while the RTs to literary words were the fastest. High-frequency words were processed faster than low-frequency words (Table 2).

Table 2. *Reaction time (SEm) in milliseconds and percentage of errors in the lexical decision task and in naming for words that exist only in literary Arabic, only in the spoken dialect, and pseudowords.*

RESPONSE TYPE		"YES" RESPONSES		"NO" RESPONSES		
STIMULUS TYPE		LITERARY ARABIC		SPOKEN DIALECT		
WORD FREQUENCY		HIGH	LOW	HIGH	LOW	PSEUDOWORDS
LEXICAL DECISION	RT	887 (43)	1411 (92)	1377 (99)	1517(113)	1140 (111)
	ERRORS	5.4 (0.6)	4.8 (0.5)	3.3 (0.9)	5.4 (0.9)	5.8 (0.7)
NAMING	RT	721 (28)	881 (44)	929 (49)	977 (56)	877 (49)
	ERRORS	2.1 (0.7)	16.2 (1.8)	13.7 (1.9)	41.7 (3.8)	7.9 (1.3)

The stimulus-type effect was analyzed separately for each task by a one-way ANOVA within subjects ($F1$), and one-way ANCOVA between stimulus-types (F2). The number of letters per stimulus was the covariate factor in the stimulus analysis. The stimulus-type effect was significant for both lexical decision and naming. [For the lexical decision task the ANOVA yielded $F1(4,76) = 21.44$, $MS_e = 59732$, $p < .001$ and $F2(4,90) = 28.0$, $MS_e = 41044$, $p < .001$. For the naming task the statistics were $F1(4,76) = 22.0$, $MS_e = 8423$, $p < .001$ and $F2(4,90) = 5.76$, $MS_e = 32858$, $p < .001$]. The influence of the number of letters per stimulus on the stimulus-type effect was marginal and statistically nonsignificant both for lexical decision ($p > .06$) and for naming ($p>.07$). Tukey-A post hoc comparisons revealed the following pattern: In the lexical decision task the rejection of both high- and low-frequency transliterations of spoken words was slower than the rejection of pseudowords. The frequency effect was significant for the acceptance of literary words, but not for the rejection of the transliterations. In the naming task, low- but not high-frequency transliterations were slower than pseudowords while high- but not low-frequency literary words were named faster than pseudowords. Within each frequency group, transliterations of spoken words were named more slowly than literary words. Because of the excessive number of errors in naming low-frequency transliterations (Table 2), we analyzed the naming data using only 12 subjects who made less than 50% errors in that condition. The RTs and the results of that analysis were similar to the above.

The effect of language on naming was also examined by a two-factor ANOVA with repeated measures. The factors were Language (literary, spoken), and Frequency (high, low). Literary words were named faster than spoken words [$F1(1,19) = 25.9$, $MS_e = 17810$, $p < .001$ and $F2(1,68) = 13.8$ $MS_e = 32788$, $p < .001$], and high-frequency words were named faster than low-frequency words [$F1(1,19) = 36.1$ $MS_e = 5992$, $p < .001$ and $F2(1,68) = 6.3$, $MS_e = 32788$, $p < .02$]. The interaction between the Language and the Frequency effects was significant in the subject analysis [$F1(1,19) = 18.0$, $MS_e = 3526$, $p < .001$] but not in the stimulus analysis [$F2(1,68 < 1.0]$.

More errors were made in the naming task (16.32%) than in the lexical decision task (4.94%) [$F1(1,38) = 52.76$, $MS_e = 2.5$, $p < .0001$, $F2(1,38) = 34.02$, $MS_e = 1.8$, $p < .001$]. The stimulus-type effect was significant across tasks [$F1(4,152) = 28.1$, $MS_e = 1.1$, $p < .0001$, $F2(4,180) = 31.45$, $MS_e = 0.9$, $p < .0001$], as was the interaction between the two factors [$F1(4,152) = 37.32$, $MS_e = 1.1$ $p<.0001$, $F2(4,180) = 34.87$, $MS_e = 0.9$, $p < .0001$]. The interaction was examined by a separate one-way ANOVA for each task. These analyses showed that errors in lexical decision were evenly distributed among the stimulus categories [$F1(4,76) = 1.66$, $MS_e = .011$, $p > .18$, $F2 < 1.00$]. In naming, on the other hand, the percentage of errors was different for the different types of stimuli [$F1(4,76) = 64.41$, $MS_e = 0.7$, $p < .0001$ $F2(4,90) = 48.35$, $MS_e = 0.9$, $p < .0001$]. Post hoc comparisons showed that for both literary and spoken words more errors were made with low- than with high-frequency words, and that pseudowords produced less errors than low-frequency words.

EXPERIMENT 2-A

Because spoken and literary Arabic differ in several important phonological aspects, it is possible that the frequency and the language effects on naming reflected difficulties at the production stage. To control for this possibility, a delayed naming task was also investigated (e.g., Besner & Hildebrandt, 1977). Sixteen new subjects from the same population were asked to read aloud all the stimuli used in Experiment 2. They were instructed, however, to delay reading onset until a signal was given. The signal was an asterisk which was presented 2.5 s after the onset of the stimulus. The exposure time of each word was 2s. The delayed naming times and the percentage of naming errors are presented in Table 3.

A within-subject ANOVA showed that delayed naming time was equal across stimulus types [$F1(4,60) = 1.19$, $MS_e = 8455$, $p > .30$ and $F2(4,91) = 1.77$, $MS_e = 7894$, $p > .15$]. Hence, under these circumstances, naming was equally fast for words in literary Arabic (363 ms), spoken Arabic (379 ms) and pseudowords (360 ms). Further more, although there was a tendency to name high-frequency words faster than low frequency words (355 ms vs. 387 ms, respectively), a Frequency × Language ANOVA showed that this difference was nonsignificant [$F1(1,15) = 2.43$, $MS_e = 7748$, $p > .14$ and $F2(1,68) = 1.82$, $MS_e = 9096$, $p < .18$] and that the two effects did not interact [$F1(1,15) = 1.64$, $MS_e = 9276$, $p > .22$, and $F2(1,68) = 1.72$, $MS_e = 9096$, $p < .19$]. The errors analysis, on the other hand, showed that even when naming was delayed, the distribution of errors was not even across the stimulus-types [$F1(4,60) = 18.7$, $MS_e = 41.9$, $p <$ 001, and $F2(4,91) = 13.5$, $MS_e = 51.1$, $p < .001$]. Post hoc comparisons revealed that more errors were made naming transliterations of low frequency words than naming any other type of stimulus, and fewer errors were made naming high-frequency literary words than high-frequency transliterations. Naming pseudowords was equally accurate as naming all other stimulus types except low-frequency transliterations.

Discussion

One of the most interesting results of Experiment 2 was that transliterations of spoken words were processed more slowly than pseudowords. Similar to the pseudohomphone effect in English or Hebrew, the rejection of transliterations of both high- and low-frequency words in the spoken dialect was delayed in lexical decision relative to the rejection of pseudowords derived from literary Arabic. Similarly, the naming of both high- and low-frequency transliterations was also delayed relative to pseudowords, although this difference was statistically significant only for the low-frequency stimuli. The word frequency effect was significant for literary words in both lexical decision and naming. On the other hand, the numerically faster RTs for high- than for low-frequency transliterations was nonsignificant in lexical decision (where these transliterations had to be rejected), while in naming it was significant only for the subject analysis. The direction of the differences in the error data was similar to that found for RTs, suggesting that the stimulus type effects on RTs were not caused by a speed-accuracy trade-off. When naming was delayed by 2.5 seconds, RTs were similar across conditions but naming the low-frequency spoken words was still highly inaccurate (19.7% errors), significantly less accurate than naming any other stimulus type.

Table 3. *Reaction time (SEm) in milliseconds and percentage of errors in the delayed naming task for words that exist only in literary Arabic, only in the spoken dialect, and pseudowords.*

STIMULUS TYPE		LITERARY ARABIC		SPOKEN DIALECT		
WORD FREQUENCY		HIGH	LOW	HIGH	LOW	PSEUDOWORDS
DELAYED	RT	331 (12)	395 (20)	379 (40)	380 (27)	359 (13)
NAMING	ERRORS	2.1 (0.61)	5.5 (1.08)	10.4 (3.1)	19.7 (3.8)	6.0 (1.09)
IMMEDIATE	RT	390 ms	486 ms	550 ms	597 ms	518 ms
MINUS						
DELAYED	ERRORS	0.1%	10.8%	4.4%	22.0%	2.1%

The general similarity of the pattern of results in naming and lexical decision suggests that post-lexical, decision-related factors cannot totally account for the stimulus-type effects found in this experiment. Furthermore, the fact that the RTs in the delayed naming experiment were similar across stimulus-type indicates that an important part of this effect, at least in naming, was related to stimulus encoding rather than production factors. However, the unusually large percentage of errors in naming low frequency transliterations which persisted even when naming was delayed suggests that these stimuli presented a particular problem to the subjects.

Unlike pseudohomophones, the transliterations of spoken words were not constructed by substituting allophones for one or two letters in real literary words. Therefore, orthographic similarity to items in the "word" category could not account for the delayed rejection of these stimuli relatively to pseudowords (cf. Taft, 1982). In fact, the subjects in the present experiment had no previous experience with the transliterations of spoken words. Informal comments made by most subjects while performing the tasks expressed their surprise that "spoken" words could also be written. Therefore, if the familiarity of the orthographic patterns had been a major factor in determining the speed of the lexical decision, the transliterations should have been rejected as fast as pseudowords—or even faster, because some of these strings included combinations of letters that are illegal in written Arabic (cf. Balota & Chumbley, 1984). Consequently, the fact that these stimuli took longer to reject than pseudowords can be more easily explained assuming that during the process of lexical decision the phonological representations of the transliterations (i.e., the phonological units representing words in spoken Arabic) had been activated.

Why should the lexical activation of words in spoken Arabic delay their rejection in a task in which only words in literary Arabic should be classified in the "positive" category? A possible explanation is that once the lexicon is accessed, and particularly because literary and spoken Arabic share a subset of words, lexical decisions required an additional classification, one between literary and spoken words. This additional classification was not necessary for pseudowords, because pseudowords do not fully activate lexical units. Note that this mechanism should have delayed lexical decisions for both literary and spoken words. Indeed, a comparison between lexical decision RTs in Experiment 1 and in Experiment 2 revealed that, while for pseudowords, the mean RT was almost identical in both experiments, the RTs to literary words were significantly longer in Experiment 2 than in Experiment 1. This difference was particularly conspicuous for low-frequency words. In fact, in Experiment 2, the time required to accept low-frequency literary words was longer than the time required to reject pseudowords.

Although the above interpretation suggests that the delay in lexical decision for both literary words and transliterations can partly be explained by decision-related processes, it is based on the assumption that the phonological representations in the lexicon are activated before the lexical decision is made. Furthermore, because the orthographic pattern of the transliterations could not have been used to address whole-word phonologic representations, the activation of these lexical units necessarily required some pre-lexical phonological computation. The stimulus-type effect on naming performance (which does not involve decision processes) supported this argument and helped elaborate the nature of the lexical involvement in the phonological processing of written Arabic words.

Naming transliterations of spoken words was slower and less accurate than naming literary words. These results are congruent with the relationship between naming orthographically familiar and unfamiliar words in Katakana (Besner & Hidebrandt, 1987) as well as with the relationship between naming pseudohomophones and nonwords in English (McCann & Besner, 1987; Taft & Russell, 1992). The frequency effects found in naming performance for both literary and spoken words suggest that lexical phonology assists phonological encoding not only when whole-word phonological units are addressed in the lexicon but also when phonology is pre-lexically assembled.

A caveat to this interpretation was introduced by the unpredicted result that naming transliterations was also slower (at least for the low-frequency spoken words) than pseudowords. This relationship is in sharp contrast with the results reported by Besner and Hildebrandt (1987) who used a fairly similar manipulation. In that study, native Japanese speakers were asked to read aloud words printed in Katakana (one of the two Japanese syllabic scripts). Some of these stimuli were words that are usually written in Kanji (a logographic script) hence they were orthographically unfamiliar to the subjects. Although these orthographically unfamiliar words were read more slowly than orthographically familiar words, in contrast to the present results, they were faster than orthographically matched nonwords. A possible (post hoc) explanation of the unexpected difference in naming transliterations and pseudowords as well as an account of the other results of this experiment is provided by a model recently proposed to account for naming in Hebrew (Frost, 1995).

According to the Frost (1995) model, generating phonology from print consists first of a computational stage during which a tentative phonological representation is formed, thereby converting letters and letter clusters into phonemes and phonemic clusters. According to our interpretation of Frost's model, the size of the orthographic unit used in that computation may vary from single letters (when the letters-string does not contain familiar orthographic structures) to whole words (when the letter string is very familiar). We assume that most of the time this computation is based on a combination of letters and sub-word orthographic structures. Partial results of this computational analysis are sufficient to feed forward and activate a set of whole-word lexical units (at which stage frequency effects may occur). The lexical units feedback and help shape the computational process, allowing a correct pronunciation. The process thereby combines a cascade-type process (e.g., McClelland, 1979) with an interactive process (Seidenberg & McClelland, 1989) during which different lexical units are activated to different extents (depending upon their respective compatibility with the partial ad hoc phonological output of the computational process). The feedback from the lexicon to the prelexical computational system might, in turn, determine the relative level of activation of these units.

According to this model, the naming of the transliterations was slowed down during the initial computational stage. This could have happened for several reasons. First, the stimuli looked sufficiently unfamiliar to prevent any attempt to address whole word units in the lexicon. Second, because the spoken Arabic dialect is never written and because they were randomly interspersed among twice as many normal (literary) words, the partial products of the computational phonological process might have been addressed to the literary lexicon. Consequently, the information available in the spoken-word lexicon might have been late to intervene and did not facilitate naming. Third, as described above, some of the transliterations contained letter sequences that were either completely illegal in normal print, or had a different pronunciation which did not fit the spoken lexicon. Pseudowords, on the other hand, were not inhibited by either the lexical process or orthographical irregularity. Therefore, reading the transliterations was more difficult than reading the pseudowords. Note that this situation is considerably different than that of the orthographically unfamiliar words in Katakana that addressed the same lexicon as the familiar words. Support for these assumptions and particularly for the particular difficulty in naming the illegal clusters was provided by the analysis of errors.

More errors were made in naming transliterations than literary words. The most striking aspect of the distribution of errors was the extremely high percentage of errors in naming low-frequency transliterations. Half of these errors, however, (21%) were made with the four transliterations that had a phonologically illegal onset. Similarly, among the errors made in naming the high-frequency transliterations, 10% were made with the three phonologically illegal clusters in this group. If we consider only the errors made in naming phonologically legal words, we are left with an expected error rate for high-frequency words but an unusually high error rate for low-frequency words. Moreover, some of these errors persisted even when naming was delayed. These errors consisted mostly of using the literary pronunciation of the letter clusters while reading the transliterations.

Our account of the lexical decision results in this experiment suggested a second stage of processing during which spoken words were distinguished from literary words. This second stage was necessary because literary and spoken words were classified in different response categories. To control for this problem and get a "cleaner" measure of the difficulty in processing the transliterations, we ran an additional experiment in which subjects were asked to make a phonological lexical decision, i.e., to "accept" any legal Arabic word and "reject" only the pseudowords.

EXPERIMENT 3

In the present experiment subjects were instructed to read the visually presented stimuli silently and make a phonologically based lexical decision (e.g., Taft, 1982). They were told to accept as "words" any letter string that sounded like a word in Arabic, regardless of whether the phonological product was a word in literary Arabic or in the spoken dialect.

If lexical decisions in the previous experiment were delayed (particularly for transliterations) mainly because a secondary classification between literary and spoken words was imposed by the task, then the difference between literary and spoken words in the present experiment should be minimal. On the other hand, if our model of word recognition is correct, the phonological computation should be more difficult for the transliterations than for literary words, regardless of the response category to which these stimuli must be assigned. Consequently, we predicted that phonological decisions will take longer for the transliterations of spoken words than for the literary words.

Method

Subjects. The subjects were 20 high-school pupils, 10 boys and 10 girls. They were naive about the purpose of the study and had not participated in any of the previous experiments.

Stimuli. The stimuli were 96 words and 96 nonwords. Among the words, 48 were the literary Arabic stimuli used in Experiment 2 and 48 were transliterations of spoken Arabic words; the transliterations were the 24 used in Experiment 2 and 24 new words, 12 high- and 12-low-frequency (see Appendix C). The pseudowords were the 48 used in Experiment 2 and 48 additional stimuli constructed by replacing one letter in literary Arabic words. All pseudowords were meaningless but phonologically legal.

Design. The RTs were grouped using a univariate five-level design, within subjects and between stimuli. The levels were high-frequency literary, low-frequency literary, high-frequency spoken, low-frequency spoken and pseudowords. As in the previous experiments, the stimulus analysis was based on ANCOVA controlling for the number of letters per stimulus (which was used as the covariate). In addition, the responses to literary and spoken words were compared using a Frequency × Language within-subject and between-stimulus design.

Procedure. The procedure was similar to that used in Experiment 2 except that the subjects were told that some of the letter strings would be transliterations of spoken words, and they were instructed to distinguish words (whether spoken or literary) from pseudowords.

Results

RTs and errors were averaged for each stimulus condition across subjects and stimuli. RTs above or below 2 SD from the subject or the stimulus mean in each condition were excluded. About 2% of the responses were outliers, equally distributed across conditions. RTs to spoken words were slower than to both literary words and pseudowords (Table 4).

Table 4. *Reaction time (SEm) in milliseconds and percentage of errors in the phonological lexical decision task for words that exist only in literary Arabic, only in the spoken dialect, and pseudowords.*

RESPONSE TYPE		"YES" RESPONSES				"NO" RESPONSES
STIMULUS TYPE		LITERARY ARABIC		SPOKEN DIALECT		
WORD FREQUENCY		HIGH	LOW	HIGH	LOW	PSEUDOWORDS
LEXICAL DECISION	RT	793 (33)	1196 (65)	1210 (42)	1422 (69)	1032 (86)
	ERRORS	5.4 (0.53)	3.7 (0.73)	5.6 (0.55)	4.8 (0.62)	4.8 (0.31)

The statistical analysis showed that the stimulus-type effect was significant [F1(4,76) = 29.04,MS_e = 37754, p<.001 and $F2(4,186) = 44.5$, $MS_e = 28859, p < .001$]. Stimulus length did not influence this main effect [$F2(1,186) = 1.29, p > .25$]. Post hoc comparisons showed that all the differences between any two single categories were significant, except for the difference between low-frequency literary words and pseudowords. The Frequency × Language ANOVA showed that responses to spoken words were slower than to literary words [$F1$(1,19 = 223.0, $MS_e = 972, p < .001$ and $F2(1,91) = 66.8$ $MS_e = 37862, p < .001$]. Responses to high-frequency words were faster than to low-frequency words [$F1(1,19) = 55.6$, $MS_e = 34047, p < .001$ and ($F2$1,91) = 41.9, MS_e = 37862, $p < .001$], but a significant interaction between the two factors revealed that the frequency effect was significantly larger for literary words (403 ms) than for spoken words (212 ms) [$F1(1,19) = 14.4$, $MS_e = 12917, p < .01$ and $F2(1,91) = 5.6$, $MS_e = 37862, p < .025$]. As in the one-way ANCOVA, the stimulus length did not influence these effects (p >.6).

In order to estimate the contribution of the decision-related factor to lexical decision performance in Experiment 2 we compared the RTs to literary words in the two experiments. Note that in both experiments these stimuli were accepted as real words. A mixed-model ANOVA was used in which Experiment was a between-subject factor and Word-Frequency a within-subject factor. This analysis revealed that literary words were accepted faster in Experiment 3 (994 ms) than in Experiment 2 (1149 ms) [$F(1,38) = 4.0$ $MS_e = 119538, p < .056$], high-frequency words were accepted faster than low-frequency words in both experiments [$F(1,38) = 119.8$ MS_e = 35870, $p < .001$], while the interaction between the two factors was not significant [F(1,38) = 2.046, MS_e = 35870, p >.16]. In contrast, the rejection of pseudowords was equally fast in both experiments $t(38) = 0.774$ p > .48.

An analysis of the errors indicated that the differences between stimulus categories was not significant within subjects [$F1(4,76) = 1.67$, $MS_e = 0.06, p > .16$]. However, the between stimulus-type analysis followed by Tukey-A post hoc comparisons, showed that more errors were made with low-frequency words (literary and spoken) than with high-frequency words or pseudowords [$F2(4,186) = 32.41$, $MS_e = 4.7, p < .0001$]

Discussion

The results of the present experiment demonstrated that the delay in processing transliterations of spoken words resulted from pre-lexical encoding difficulties as well as from decision-related factors. On the one hand, lexical decisions were slower for transliterations than for literary words even though the decision was phonologically based (i.e., both literary and spoken words were classified in the same response category). Hence, assuming that our subjects were at least as familiar with the phonological representations of spoken words as they were with those of literary words, the delay in the phonological lexical decision for the spoken relative to the literary words indicates that the process of generating the phonological code from the transliterations was slower (more difficult) than from literary words. On the other hand, literary words were accepted faster in the present experiment (in which phonological decisions were required) than in Experiment 2 (in which the decisions could, at least according to some theories, be based on the visual familiarity of the orthographic patterns). This outcome supports our assumption that in Experiment 2 as well as in the present experiment lexical decisions were based on the phonological structure of the visual stimuli, and that, given the nature of the task, a second distinction between spoken and literary words was necessary only there.

Assuming that lexical decision requires the recovery of the phonological structure of printed words, we suggest that this process has similar components in naming and in lexical decision. Thus, generating the phonological structure was faster for literary than for spoken words because: 1) The orthographic patterns of literary words were relatively more familiar and therefore some of these stimuli could directly address whole-word phonological units in the lexicon 2) Spoken words are not usually written and, therefore, when transliterations were processed the partial phonologic output of the pre-lexical computation might have been addressed (by rule) first to the literary lexicon and 3) The unusual combination of letters might have inhibited pre-lexical computation of the transliterations and limited the size of the orthographic structure used in the translation process to single phonemes.

GENERAL DISCUSSION

The present study was aimed at examining the role of phonology in lexical decision and naming by assessing the effects of the phonological structure of orthographic patterns representing words in spoken Arabic on lexical decision and naming performance. Because only literary Arabic is written, transliterations of words that are specific to the spoken Palestinian dialect were novel orthographic stimuli for all our subjects. Consequently, unless the phonologic structure of the orthographic pattern was processed by the reader while performing these tasks, transliterations should have been treated as unfamiliar nonwords.

The results of the three experiments can be summarized as follows. Both lexical decision and naming performance were inhibited while processing the transliterations in comparison to literary words and pseudowords derived from literary Arabic. Transliterations of spoken words were more difficult to reject than meaningless pseudowords (Experiment 2), but also were accepted more slowly than literary words in a phonologically based lexical decision task (Experiment 3). Lexical decisions were inhibited for both literary and spoken words when transliterations had to be rejected relatively to a condition when all the words were in literary Arabic (Experiment 1) or when the decision was phonologically based. Naming transliterations took longer and was less accurate than naming literary words. Unexpectedly, naming transliterations was also slower and less accurate than naming pseudowords, although the latter difference was significant only for low-frequency stimuli. Finally, an unusually large word frequency effect was found in naming as well as in making "positive" lexical decisions for both literary and spoken words.

These data are congruent with views suggesting that word recognition is always mediated by phonology. The subjects in the present study did not ignore the phonological structure of the transliterations even though ignoring it would have facilitated the lexical decision. Moreover, the longer RTs to literary Arabic words when transliterations had to be rejected (Experiment 2) than when phonological analysis was imposed by the task (Experiment 3), strongly suggests that even if given the option to use the orthographic pattern for lexical categorization (and in fact, this strategy would have been the most efficient), subjects could not ignore the phonological structure of the literary words. Hence, familiar as well as unfamiliar orthographic patterns are analysed phonologically during the course of lexical decision.

This is not to say, however, that phonology is always computed pre-lexically (cf. Frost, 1995). In that paper Frost distinguished between a strong and a weak version of the phonological mediation theory. According to the strong version, the initial process of recovering phonologic information from print necessarily involves the translation of graphemes into phonemes and does not make use of the notion of addressed phonology, i.e., accessing whole-word phonologic units using whole-word orthographic patterns (Carello et al., 1992; Lukatela & Turvey, 1990; Van Orden, Pennington & Stone, 1990). In contrast, the weak version of the phonological mediation theory, while still emphasizing that phonological encoding is obligatory and necessarily mediates word recognition, views the generation of phonology from print as a process that involves computations at the level of subword orthographic units, in addition to direct connections between whole-word orthographic units and whole-word phonologic units. Obviously, pre-lexical computation and addressed phonology are not mutually exclusive. In fact both processes may be attempted in parallel and, to some extent support each other. What determines the relative contribution of these two processes to the retrieving of the phonological structure of a printed word is the ease with which pre-lexical phonology can be achieved. Thus, when the orthographic patterns of the words are relatively unfamiliar, for example infrequent words (Seidenberg, 1985b), or very unfamiliar as in pseudowords, pre-lexical computations are dominant. On the other hand, when the orthographic pattern is very familiar and/or the sub-word orthographic units are phonologically ambiguous (for example phonologically irregular words) or when the print provides only incomplete phonological information (such as in unpointed Hebrew or Arabic) addressed phonology could have a more important role (Frost et al., 1987). Although the present data do not disprove the strong version of the phonological mediation theory, they are more easily accommodated by its weak version. Note that support for the strong version has been mainly indirect, in the form of demonstrations that all lexical effects on naming can be explained without assuming addressed phonology. Until recently, empirical evidence has been provided only for reading very shallow orthographies, such as Serbo-Croatian (e.g., Lukatela, Turvey, Feldman, Carello, & Katz, 1989). In his recent paper, however, Frost (1995) supported the strong version of the orthographic mediation theory by showing that for unpointed Hebrew words of equal length and frequency, naming, but not lexical decision time, is positively and monotonically related to the number of missing phonemic units. Moreover, word frequency effects in naming were found only when phonemic information was missing, but not when it was complete.

Contrary to Frost's results, we found that naming Arabic words was significantly influenced by word-frequency even though sufficient phonemic information was provided to enable unequivocal pronunciation, although not all the diacritical marks were added. Moreover, the huge word-frequency effect in both naming and lexical decision, which was more than three times as large for literary than for spoken words, suggests a qualitative difference in the processing of high- and low-frequency literary words. It is reasonable to assume that while a relatively large proportion of high-frequency printed literary words could rapidly retrieve their phonological structure via associative connections between whole-word orthographic patterns and whole word phonological units, addressed phonology was not an option for transliterations and low frequency literary words. This hypothesis is also supported by comparisons between the word-frequency effects on naming regular vs. exception words (Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987), on comparisons between word-frequency effects on naming in languages with deep and shallow orthographies (Frost et al., 1987), and on morphological or prosodic manipulations (Monsell, 1991; Monsell, Doyle & Haggard, 1989; Paap, McDonald, Schvaneveldt, & Noel, 1987).

Other studies, however, have raised doubts about the role of word-frequency in lexical access, and suggest that word-frequency effects in naming reside in the connection between visually accessed lexical entries and their articulatory output (McCann & Besner, 1987) and that in lexical decision it plays a role only at a post-lexical decision stage (Balota & Chumbley, 1984). The present results indicate that phonological encoding factors probably account for most of the word-frequency effects in naming and suggest a pre-lexical as well as a decision-related influence of word-frequency in lexical decision.

The word-frequency effect on phonological decisions for both literary words and transliterations of spoken words (which contradicts the results reported by McCann, Besner, & Davelaar, 1988) and particularly the large frequency effects in naming transliterations of spoken words (which contradicts the results reported by McCann & Besner, 1987) indicate that word frequency has a role in phonological encoding and lexical access (see also Taft & Russell, 1992). Additional support for this claim was provided by the delayed naming experiment where naming time was not affected by word frequency (see also Monsell et al., 1989). Moreover, as elaborated in the discussion of Experiment 2 the present data indicate that lexical phonology is directly involved in pre-lexical phonological computation even when addressed phonology is impossible.

With regard to lexical decision, Balota and Chumbley's two-stage model suggests that word frequency determines the value of the stimulus on a familiarity/ meaningfulness (FM) dimension. Since the same transliterations were used in Experiment 2 and Experiment 3, their FM values must have been the same in both experiments. Therefore, if word frequency had affected only decision strategies it should have had an opposite effect on opposite decisions. Yet, lexical decision for high-frequency words was faster than for low-frequency words, both when the subject accepted these stimuli as words and when they were rejected. Note however that the frequency effect was statistically significant only for the phonological lexical decisions.

In conclusion, the present study provides additional evidence that phonological information is automatically analyzed during visual word recognition (cf. Perfetti & Bell, 1991; Perfetti, Bell, & Delany, 1988) and that the phonological structure of printed words is used by the reader during word recognition. Unlike strong versions of the phonological mediation theory however, we assume that the orthographic pattern of very frequent words may be associated to whole-word phonological units in the lexicon and that these associations may be used to retrieve the word's phonological structure addressing by the lexicon directly. These data are not supportive of

models suggesting that word recognition in general and lexical decisions in particular can be based solely on visual familiarity and orthographic analysis (e.g., Besner & McCaan, 1987). The word-frequency effect on processing literary and spoken words supports Monsell's (1991) suggestion that the effect of frequency reflects lexical transcoding from orthography to phonology, and suggest that lexical phonology may contribute to this process by shaping pre-lexical phonological computation even when addressed phonology is not possible. A model accounting for this pattern may combine a cascade-type feed forward activation of lexical phonological units by partial output of pre-lexical phonological computation (e.g., McClelland, 1979) with feedback from the activated units which may shape pre-lexical computation (e.g., McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982).

REFERENCES

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340-357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production frequency? *Journal of Verbal Learning and Verbal Behavior, 24*, 89-106.

Banks, W. P., Oka, E., & Shugarman, S. (1981). Recording of printed words to internal speech: Does recording come before lexical access? In O. J. L. Tzeng & H. Singer (Eds.), *Perception of Print: Reading Research in Experimental Psychology* (pp. 137-170). Hillsdale, NJ: Erlbaun.

Bentin, S., Bargai, N., & Katz, L. (1984). Graphemic and phonemic coding for lexical access: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 353-368.

Bentin, S., & Frost, R. (1987). Processing lexical ambiguity and visual word recognition in a deep orthography. *Memory & Cognition, 15*, 13-23.

Besner, D., Davis, J., & Daniels, S. (1981). Reading for meaning: The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology, 33A*, 415-437.

Besner, D., & Hildebrandt, N. (1987). Orthographic and phonological codes in the oral reading of Japanese Kana. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 335-343.

Besner, D., & McCann, R. S. (1987). Word frequency and pattern distortion in visual word identification and production: An examination of four classes of models. In M. Coltheart (Ed.), *Attention and Performance XII* (pp. 201-219). Hove, E. Sussex: Erlbaum.

Besner, D., & Smith, M. C. (1992). Basic processes in reading: Is the orthographic depth hypothesis sinking? In R. Frost & L. Katz (Eds.), *Advances in psychology: Orthography, phonology, morphology, and meaning* (pp. 45-66). Amsterdam: Elsevier/North Holland.

Carello, C., Turvey, M. T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological? In R. Frost & L. Katz (Eds.), *Advances in psychology: Orthography, phonology, morphology, and meaning* (pp. 211-226). Amsterdam: Elsevier/North Holland.

Carello, C., Turvey, M. T., & Lukatela, G. (1994). Lexical involvement in naming does not contravene prelexical phonology: A reply to Sebastian Gallés (1991). *Journal of Experimental Psychology: Human Perception and Performance, 20*, 192-198.

Coltheart, M. (1980). Deep dyslexia: A review of the syndrome. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 22-47). London: Routledge & Kegan Paul.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological recording in lexical decision task. *Quarterly Journal of Experimental Psychology, 31*, 489-507.

Coltheart, M., Davelaar, E., Jonasson, J. T. & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*. London: Academic Press.

Courrieu, P., & Do, P. (1987). Perceptual analysis of words in Arabic. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 451-458). North-Holland: Elsevier.

Dennis, I., Besner, D., & Davelaar, E. (1985). Phonology in visual word recognition: Their is more two this than meats the I. In. D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 167-198). Academic Press.

Farid, M., & Grainger, J. (in press). How initial fixation influences visual word recognition: A comparison of French and Arabic. *Brain and Language*.

Foss, D. J. & Blank, M. A. (1980). Identifying the speech codes. *Cognitive Psychology, 12*, 1-31.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 361-379.

Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 116-129.

Frost, R. (1995). Phonological computation and missing vowels: Mapping lexical involvement in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 398-408.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies of visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance 13*, 104-115.

Gough, P. B., & Cosky, M. J. (1977). One second of reading again. In N. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2). Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*, 74-95.

Jarred, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120*, 358-394.

Katz, L., & Feldman, L. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning Memory and Cognition, 9*, 157-166.

Liberman, A. M. (1992). The relation of speech to reading and writing. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 193-210). North-Holland: Elsevier.

Liberman, I. Y., & Liberman, A. M. (1990). Whole word vs. code emphasis: Underlying assumptions and their implications for reading instruction. *Bulletin of the Orton Society, 40*, 51-76.

Lukatela, G., & Turvey, M. T. (1990). Automatic and prelexical computation of phonology in visual word identification. *European Journal of Cognitive Psychology, 2*, 325-343.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Context effects in bi-alphabetical word perception. *Journal of Memory and Language, 28*, 214-236.

McCaan, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models pronunciation assembly and the locus of word frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 13-24.

McCaan, R. S. Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance, 14*, 493-706.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287-307.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

Monsell, S. (1991). The nature and the locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148-197). Hillsdale, NJ: Erlbaum.

Monsell, S., Doyle, M. C. & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118*, 43-71.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76*, 340-354.

Morton, J. & Patterson, K. E. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. Patterson & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 91-118). London: Routledge & Kegan Paul.

Paap, K. R., McDonald, J. E., Schvaneveldt, W. R. & Noel, R. W. (1987). Frequency and pronounceability in visually presented naming and lexical decision tasks. In M. Coltheart (Ed.), *Attention and Performance XII* (pp. 221-243). Hove, E. Sussex: Erlbaum.

Paap, K. R., Newsome, S. L., McDonald, J. E. & Schvaneveldt, W. R. (1982). An activation-verification model for letter and word recognition. *Psychological Review, 89*, 573-594.

Patterson, K. E. (1981). Neuropsychological approaches to the study of reading. *British Journal of Psychology, 72*, 151-174.

Patterson, K. E., & Coltheart, V. (1987). Phonological processes in reading. In M. Coltheart (Ed.), *Attention and Performance XII* (pp. 421-447). Hove, E. Sussex: Erlbaum.

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backword masking and masked priming. *Journal of Memory and Language, 30*, 473-485.

Perfetti, C. A., Bell, L., & Delaney, S. (1988). Automatic phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27*, 59-70.

Roman, G., & Pavard, B. (1987). A comparative study: How we read Arabic and French. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 431-440). North-Holland: Elsevier.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recording in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10*, 645-657.

Rumelhart, D. E., & McClelland, J. E. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement and some tests and extensions of the model. *Psychological Review, 89*, 60-94.

Saffran, E. M., & Marin, O. S. M. (1977). Reading without phonology: Evidence from aphasia. *Quarterly Journal of Experimental Psychology, 29*, 515-525.

Seidenberg, M. S. (1985a). The time course of information activation and utilization in visual word recognition. In D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 199-252). New York: Academic Press.

Seidenberg, M. S. (1985b). Constraining models of visual word recognition. *Cognition, 20,* 169-190.

Seidenberg, M. S., & McClelland, J. L. (1989). The distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523- 568.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23,* 383-404.

Taft, M. (1982). An alternative to grapheme-phoneme conversion rules? *Memory & Cognition, 10,* 465-474.

Taft, M., & Russell, B. (1992). Pseudohomophone naming and the word-frequency effect. *The Quarterly Journal of Experimental Psychology, 45A,* 51-71.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 25,* 608-631.

Van Orden, G. C.. (1991). Phonological mediation is fundamental to reading. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 77-103). Hillsdale, NJ: Erlbaum.

Van Orden, G. C., Pennington, B. F. & Stone G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488-522.

FOOTNOTES

[*] *Journal of Experimental Psychology: Memory, Learning, and Cognition,* in press.

[†] Center for Neural Computation, Hebrew University, Jerusalem

[‡] Department of Psychology and School of Education, Hebrew University, Jerusalem

[1] According to our conceptualization, the lexicon is a sub-system in semantic memory which initially stores phonological information about words. With practice, orthographic information may be added to some of the lexical entries. This lexicon is the data-base used for both word recognition and word production.

[2] Alternative explanations of the pseudohomophone effect were based on the orthographic similarity between the pseudohomophone and the related real word (e.g., Taft, 1982). Such explanations, however, are irrelevant to the present study in which the transliterations were not pseudohomophones of words in literary Arabic and, therefore, did not bare any specific orthographic similarity to written words.

[3] Most of these theories assume the existence of a separate orthographic lexicon from which a phonological (output) lexicon may be addressed. For coherence reasons, while citing these theories, we chose to use our conceptualization of the unified lexicon in which each word entry contains both phonologic and orthographic information. For the present exposition we don't see radical differences between these two definitions of the lexicon.

[4] In Arabic there is no difference between print and handwriting. We decided to use calligraphic-written stimuli rather than computer fonts because of the poor quality of the latter.

[5] The illegal nature of these letter strings stems from the phonological differences between spoken and literary Arabic, and from the fact that spoken Arabic is usually not written. Hence, two consonants would never occur at the beginning of the word in writing (as the trigram "ZBL" does not exist in English written words).

[6] There were no significant hesitations or coughs in the naming task, therefore all the correct responses were included in the analysis.

[7] Note, however, that not all the diacritics were included. In principle, the subjects could have read the words as nonwords, assigning a meaningless pronunciation.

[8] In some Arab countries there is a tendency to introduce spoken words in newspapers and other popular reading material. Our high-school students, however, do not usually read this literature.

[9] Because the most interesting comparison in this experiment was between transliterations of spoken words and legal nonwords, and in order to maintain a 1:1 ratio between "yes" and "no" responses, the number of spoken words presented was equated to the number of nonwords rather than to the number of literary words.

[10] Note however that the difference between orthographically unfamiliar words and nonwords persisted also when naming was delayed. Although the authors suggested that this persistence was caused by an insufficient delay of naming (1 s), these results cannot be considered conclusive.

APPENDIX A

Words used in Experiment 1[1]

High-Frequency	Approximate Pronunciation	Meaning	Low-Frequency	Approximate Pronunciation	Meaning
قلم	kalam	pen	طريده	tarida	insanity
مدرسه	madrasah	school	طاغيه	tagia	gorge
كثير	kathir	much	وعكه	waakah	pain
سؤال	sual	question	قريحه	karixa	appetite
نار	nar	fire	مناقصه	munakasa	auction
شمعه	shamaah	candle	تأزم	taazam	got in trouble
دفتر	daftar	copybook	محصن	muxasan	fortified
شجره	shajara	tree	مرتزقه	murtazaka	mercenary
صوره	surah	picture	حاشيه	xashia	entourage
باب	bab	door	وهله	waxlah	sudden
شباك	shubak	window	مكتوم	maktum	hidden
أرض	ard	land	خنصر	xansar	little finger
بحر	baxr	sea	تقشف	takashuf	modesty
صيف	sayf	summer	طوفان	tufan	deluge
سطح	satx	roof	عرقله	arkalah	failure
شهر	shahr	month	قوميه	kaumiah	nationality
شارع	sharia	road	إبزيم	ybzym	buckle
فنجان	funjan	cup	لزوم	luzum	need
دار	dar	house	طلبيه	talabyah	petition
صحن	saxyn	plate	إنشق	ynshaka	cracked
شمس	shams	sun	تشتت	tashatat	spread
مدير	mudeer	director	تعمد	taamad	premeditated
معلم	mualem	teacher	حصانه	xasanah	immunity
لوح	laux	blackboard	فخفخه	faxfaxa	wealth

[1] All the words are used in both literary and spoken Arabic.

APPENDIX B

Words used in Experiment 2

Transliterations of spoken words*

High-Frequency	Approximate Pronunciation	Meaning	Low-Frequency	Approximate Pronunciation	Meaning
طرطوع	tartuah	young	بايز	bayez	broken
تزنطر	tzantar*	in rage	سكج	sakaj	managed
كاره	karah	dough-cloth	قرطوز	kartuz	servant
أزوك	yzwak	curved	مليفك	mlyfak*	cunning
بجم	bajam	stupid	مشختف	mshaxtaf*	miser
مزمّل	mzamel*	freezing	جلص	jyls	impertinent
عفارم	afarym	compliment	جلوق	jaluk	mouth
بخشيش	baxshish	tip	صنديحه	sandixa	forehead
قزّاز	kazaz	glassmaker	هشلم	xashlam	bush
مقحميش	mkaxmesh*	dry	فندله	fandaleh	show off
قرمز	karmaz	kneeled	مشعتل	mshatal*	distributed
تبّان	taban	silo	مرهدل	mraxdal*	untidy

Words used only in literary Arabic

High-Frequency	Approximate Pronunciation	Meaning	Low-Frequency	Approximate Pronunciation	Meaning
مقام	makam	holy grave	دبال	dubaal	volcanic soil
مجهر	myjxar	telescope	دجّج	dajaj	hunter
مضارع	mudarya	present	شذرات	shadarat	parts
درّاجه	darajah	bicycle	كهنوت	kaxanut	religious
نموذج	namuthaj	example	أطناب	atnaab	enlarged

* Phonologically illegal in literary Arabic

APPENDIX B - continued

High-Frequency	Approximate Pronunciation	Meaning	Low-Frequency	Approximate Pronunciation	Meaning
فهرس	faxras	content	جهر	jaxara	informed
نعت	naat	description	نبراس	nybras	flag
مثلثات	muthalathat	trigonometry	تقنين	taknyn	moderate
حاسوب	xasub	computer	حذافير	hadjafyr	detailed
همزه	xamzah	the vowel /a/	كابسوله	kabsulah	pill
متراميه	mutaramiah	raising	مضمار	mydmar	range
مفردات	mufradaat	concepts	أطروحة	uutruxa	thesis
دوره	dawrah	course	إختلج	yxatalaja	moved
نظرية	nadjariah	theory	إعتباط	aytybat	illogical
مبتدا	mubtadaa	subject	إبتذال	abtydal	sacrifice
كسره	kasra	the vowel /y/	شتان	shatana	gap
كائن	kaayn	exist	تصعيد	tasyyd	worsening
خلية	xalyah	cell	إحتذى	yxtadah	followed
سكون	sukun	schwa	دأب	daaba	interested
طاقة	takah	energy	إيماء	ymaah	indication
عشوائي	ashwaee	occasional	مثانة	mathana	bladder
مسدس	musadas	revolver	متضلع	mutadaleh	expert
عاري	aary	shrewd	سجالا	syjalan	mutual
أطلق	atlaka	fire (a gun)	تقنية	tikanya	technique

APPENDIX C

Additional transliterations used in Experiment 3*

High-Frequency	Approximate Pronunciation	Meaning	Low-Frequency	Approximate Pronunciation	Meaning
زوم	zum	cycle	زعير	zaeer	active
مكعبل	mkaabal*	spheric	متكه	matakeh	reason
مخرّق	mxarek*	duck (verb)	شلاتي	shlaty*	lazy bum
مطجن	mtajen*	clod	براره	brara*	rejected product
طايش	taysh	intrigued	تلاطش	tlatash*	struggled
متنك	mtanek*	ancient	سكنه	sakaneh	cigarette ash
فرش	farash	extended	بايخ	bayex	dull
فنس	fanas	disappoint	مترب	mtareb*	worn out
شوفه	shofeh	sight	خوش	xosh	close by
مطبش	mtabash*	broken	مكرسح	mkarsax*	destroyed
تبضع	tbadaa*	shopped	إنقلز	ynkalaz	died
مزتوت	maztut	thrown away	تعمس	tamas	insinuated

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 221-234

Lexical and Semantic Influences on Syntactic Processing

Avital Deutsch,† Shlomo Bentin,‡ and Leonard Katz†††

The present study addressed the issue of the independence of syntactic processing from lexical and semantic processing. Syntactic (inflectional) priming was manipulated by preceding verb, adjective, and pseudoword targets with noun phrases that either agreed or disagreed in gender and/or number with the target. In Experiment 1, similar syntactic priming effects were found whether the target was a word or a pseudoword. For both, subjects' decisions that targets were the same/different to a probe were faster for targets that were syntactically congruent with their sentential context than for incongruent targets. In Experiment 2, there was a congruency effect for gender: Naming a target that did not agree in gender with the preceding noun phrase was delayed relative to naming a congruent target, but only if the noun phrase's subject was animate. For inanimate targets, syntactic congruency had no statistically significant effect. We suggest that inflectional analysis may not require the full activation of a lexical entry initially; however, subsequent syntactic analysis does interact with a word's semantic information.

The issue of the autonomy of syntactic processing in language perception is controversial. Some authors, adopting a modular approach to the structure of the linguistic system, suggest that communication between syntactic and other cognitive levels of analysis is independent and takes place only at the output of the respective modules (Fodor, 1983; Forster, 1979). An alternative, interactive, view posits mutual influence between the different cognitive domains throughout the processing of the linguistic input (McClelland, 1987; Marslen-Wilson & Tyler, 1987; Tanenhaus, Spivey-Knowlton, Eberhard, Sedivy, 1995).

The autonomy of syntactic processes in sentence comprehension has been supported by studies using a variety of techniques such as self-paced word-by-word reading, or the examination of eye movements during the reading of sentences that were syntactically ambiguous (e.g., Ferreira & Henderson, 1990; Mitchel, 1987; Ferreira & Clifton, 1986; Rayner, Carllson, Frazier, 1983). Syntactic (but not semantic) ambiguity was formed in these studies by using, for example, a reduced relative clause (e.g., "The performer sent the flowers was very pleased") or using sentences in which there was an attached prepositional phrase (e.g., "The spy saw the cop with *a revolver*") (see Rayner, et al., 1983). These studies revealed that when the reader encounters the disambiguating part of the sentence, the pace of reading is reduced and the reader's gaze regresses (a *garden path effect*). For the present perspective, the important aspect of the garden path effect was that it was observed even when the semantic characteristics of the sentence were unambiguous (as in the examples above). Consequently, Rayner et al. (1983) suggested that sentence processing, is initially governed by syntactic parsing based on the minimal attachment principle (Frazier & Rayner, 1982), and that it is independent of semantic

This study was supported by a grant from the Israeli Foundations Trustees to Shlomo Bentin, and partly by a grant from NICHD # 01994 to Haskins Laboratories. This paper was written while S. Bentin was a visiting professor at the Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto Canada, supported by the Ben and Hilda Katz International Visiting Scholar Program. We are indebted to skillful assistance from Naomi Wexler and Omer Nirhod.

or pragmatic constraints.[1] Similar results were obtained manipulating the animacy of the first noun phrase, thus influencing the thematic role that it performs (Ferreira & Clifton, 1986), or using verb subcategorization to constrain the syntactic analysis of ambiguous sentences: for example, using verbs that 'prefer' a noun-phrase or a sentence complement (Ferreira & Henderson, 1990; Mitchell, 1987, 1989). These studies supported the syntactic autonomy by revealing a garden path effect in eye movements (hence indicating the application of the minimal attachment principle) regardless of the animacy of the noun or the type of the verb. Finally, additional support for a functional and possibly neuroanatomical dissociation between the syntactic and semantic systems was recently provided by event-related potential (ERP) studies of semantic and/or syntactic processing (Munte, Heinze, & Mangun, 1993; Rosler, Putz, Frederici, & Hahne, 1993). These studies demonstrated that violation of the syntactic integrity of sentences modulates ERPs that have a different scalp distribution and latency than the N400 potential which is modulated by semantic incongruence.

In contrast to the above evidence, other studies, suggested an interaction between syntactic and semantic processing during sentence comprehension. For example, Taraban and McClelland (1988) found that, contrary to the minimal attachment principle, the attachment of the prepositional phrase was initially influenced by content based expectations. In a self-paced reading task in sentences with ambiguity induced by prepositional attachment, they showed that the reading rate at various parts of the sentence was a function of the consistency between the reader's context-based expectations for a specific attachment (whether minimal or non-minimal), and the ultimate structure of the sentence, rather then being related to the specific syntactic structure of the sentence. Thus, by biasing the sentential context with pragmatic cues into minimal or nominal attachment, it is possible to eliminate the difficulties which may be occasionally observed in sentences with a prepositional attachment that is inconsistent with the minimal attachment principle. Similar conclusions were reached in additional studies. In an attempt to replicate the Ferreira and Clifton (1986) study, Trueswell, Tanenhaus, and Garsney (1994) found that the animacy of the noun significantly constrained the initial parsing of ambiguous sentences with a reduced relative. It is possible however, that the discrepancy between Ferreira & Clifton (1986) and Trueswell et al. (1994) results is accounted for by a difference between the materials used in these two studies. Unlike the first, Trueswell et. al (1994) avoided using inanimate nouns that could be the subject of active verbs (such as, for example, instruments). They also avoided using verbs with ergative meanings that could form acceptable predicates of inanimate nouns (such as, for example, "The trash smell...").

Also contradictory to the hypothesis of autonomous syntactic processing were further studies in which sub-categorization of verbs was found to guide syntactic parsing (Holmes, Stow, & Cupples, 1989; Stowe, 1989). For example, Stowe (1989) found that the sub-categorization preference of causative verbs can be influenced by the animacy of the subject. In that study the garden-path effect was eliminated in ambiguous sentences such as "While his mother was drying off the boy began to go in," by replacing the first noun in the subordinate clause ("mother" in the above example) with an inanimate noun (for example, "towel"). Hence, the noun's animacy biased the subcategorization of the verb from transitive to intransitive and, consequently eliminated the garden-path effect by eluding a subject-verb-object parsing of the sentence. Hence, it is possible that some of the contradictory findings about syntactic autonomy is explained by difficulties in determining the role that certain semantic manipulations may have on the construction of the thematic roles of syntactic units and, consequently, on the syntactic parsing of the sentence.

A different approach taken to study syntactic processing in general and the question of its independence from other linguistic processes, in particular, was the investigation of syntactic context effects on word recognition and lexical decision. Taking this approach, ample evidence have been provided showing that, regardless of semantic relationship, target words are processed faster and more accurately when their inflectional forms are congruent with the syntactic context in which they appear than when they are syntactically incongruent (Carello et al., 1988; Goodman, McClelland, & Gibbs, 1985; Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985; Katz, Boyce, Goldstein, & Lukatela, 1987; Marslen-Wilson, 1987; Seidenberg et al., 1984; Sereno, 1991; Tanenhaus, Leiman, & Seidenberg, 1979; West & Stanovich, 1986). In addition, the idea of

1991; Tanenhaus, Leiman, & Seidenberg, 1979; West & Stanovich, 1986). In addition, the idea of independence between syntactic and semantic processes has been supported by several studies that have shown that the syntactic priming effect has different characteristics than the well known effect of semantic priming (Carello, Lukatela, & Turvey, 1988; Goodman, McClelland, & Gibbs, 1981; Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985; Katz, Boyce, Goldstein, & Lukatela, 1987; Seidenberg, Waters, Sanders, & Langer, 1984; Tyler & Wessels, 1983; West & Stanovich, 1986). For example, whereas semantic priming effects are usually found in both naming and lexical decision, syntactic priming effects were not found in naming (Carello et al., 1988; Seidenberg et al., 1984), or were very small (Stanovitch & West, 1986). Moreover while semantic priming requires, by definition, the manipulation of lexical units such as base-word, several studies in Serbo-Croatian have shown that a lexical decision to a noun was faster if it was preceded by congruently inflected pseudoadjective (a nonword inflected like an adjective) than if the noun and pseudoadjective were syntactically incongruent (Gurjanov et al., 1985; Katz et al., 1987). Hence, in contrast to semantic priming, syntactic priming based on inflectional morphology can be obtained even if the stem of the prime is not included in the lexicon. To the best of our knowledge, there is no direct evidence for the interactive view which is based on studies of syntactic priming.

In the present study we used the syntactic priming effect to explore possible interactions between syntactic and lexical or semantic processes. To this end we focused our research on the effects of violating agreement rules that are anchored in Hebrew inflectional morphology.

Hebrew is a highly inflected language. Most nouns (with the exception of a few categories like collective nouns and proper names) and adjectives are inflected for gender and number. Similarly, all verbs are also inflected for gender and number (as well as for person, tense, aspect, etc., with the exception of the present tense which is not inflected for person) Inflection is formed by affixation (mostly by suffixes) of a base-form which itself is a combination of two morphemes: a consonantal root (usually a three-consonant sequence) and a word-pattern of vowels supperimposed on the consonants. For example, the consonantal sequence Y-L-D is combined with a vowel pattern to produce the masculine word YELED (boy). In order to transform this (unmarked) form into the feminine form (i.e., girl), a specific vowel pattern that includes the feminine suffix /A/ replaces the unmarked (masculine) form's pattern to produce YALDA. A further vowel pattern transformation produces the derived form YALDUT (childhood). Note that the tri-consonant sequence remains invariant while the vowel pattern changes. Occasionally, as in the last example, the transform includes additional consonants (see Frost & Bentin, 1992 for description of the structure of Hebrew words).

In Hebrew the masculine singular form constitutes the unmarked form. The feminine gender is marked usually by one of three possible suffixes /a/, /ɛt/, or /yt/ and the plural is marked by the suffix /ym/ (usually used for masculine) and /ot/ (usually used for the feminine). Frequently, the addition of inflectional affixes also changes the infixed word structure. For example, "yalda" (girl) is the feminine form of "yɛlɛd" (boy). Note that y-l-d is the root which is common of both words and that the addition of the suffix 'a', denoting the feminine, induced a change in word structure. On the other hand, the masculine form "shofɛt" (judge) is, in the feminine, "shofɛtɛt"; here, the infixed word structure of the masculine was not altered. The same suffixes are also used by the verbal system in the various tenses to denote the feminine gender, for example: The masculine verb- form "katav" (he wrote) becomes "katva" (she wrote) and the form "rokɛd" (he dances) becomes "rokɛdɛt" (she dances). For a more detailed description of Hebrew morphology and word structure see Bentin & Frost, 1994.

Agreement rules exist; they are based on inflectional matching between words that carry inflection. They are the most fundamental tool for specifying syntactic relations in Hebrew sentences. For example, the agreement rule according to which the subject and the predicate agree in gender and number (and also in person if the predicate is a verb in the past, future, or imperative forms) is nearly always an unequivocal cue for specifying the subject and the predicate in a sentence.[2] Thus, a sentence like: "The suspicious (male) judge fell down" which translates into Hebrew as: "Hashofɛt (article /Ha/ + subject; "the judge") haxashdan (article /ha/ +

We chose to investigate the interaction between syntactic and other linguistic processes (lexical and semantics) using agreement rules, for two main reasons: (1) Agreement rules are a purely syntactic tool for indicating a functional relation between certain words. Thus by manipulating agreement rules, particularly between basic elements in sentences such as subject and predicate, we undoubtedly tap processes that are predominantly syntactic. (2) While the agreement rules are based on inflectional morphology, their violation does not induce changes in word class which may entail semantic implications (Carello, et al., 1988). Furthermore, because the information regarding the subject's number and gender is already available in the subject's form, violation of the agreement, does not affect the basic meaning of the sentence. Consequently, contextual effects observed using Hebrew agreement rules in a syntactic priming paradigm, may safely be related to syntactic rather than semantic processes.

In Experiment 1, we examined the role of lexical factors in syntactic priming using a target-probe match/mismatch decision paradigm. In Experiment 2, we examined the interaction between syntactic and semantic factors via a target-naming paradigm.

EXPERIMENT 1

In the strict modular view of syntactic processing, a word is first parsed into its lexical morpheme (i.e., its unmarked consonantal sequence) and its syntactic morphemes (such as its inflectional pattern). After parsing, the syntactic process is conjectured to be independent of other lexical information such as the word's semantic information. The present experiment was aimed at addressing this question of syntactic processing autonomy.

Targets, which were inflected words or pseudowords, were presented following a noun phrase containing a subject and its attribute (i.e., the target's context). Each word-target was a predicate that completed a three-word sentence; it could not be predicted on the basis of the semantic content of the context. Half of the words and half of the pseudowords were inflected for gender and number in agreement with the context, while the other targets were syntactically incongruent with the preceding context. The experimental subjects were instructed to match these targets to probes presented just before the context phrase. In our previous studies (e.g., Deutsch & Bentin, 1994), violation of agreement inhibited word-identification. Consequently, we predicted that RT for matching syntactically congruent words would be faster than syntactically incongruent words. Assuming that pseudowords are not represented in the lexicon, similar syntactic priming effects for words and pseudowords should support a modular view in which syntactic analysis is dependent only on the syntactic information in the stimulus.

Method

Subjects. The subjects were 72 undergraduate students who participated in this experiment as a course requirement. All of the subjects were native speakers of Hebrew and had normal or corrected vision.

Stimuli and design. Each stimulus item was a three word sentence which consisted of a noun phrase followed by a predicate, the target. The targets were 72 words and 72 pseudowords; each constituted the predicate in a three word sentence. For all the targets used in this study, inflection for the feminine gender and for plural number required the addition of a suffix to the unmarked form.

The pseudowords were constructed by substituting some of the consonants of a real word's root morpheme without changing the vowels or consonants of the syntactic word-pattern and without changing the inflectional morphemes. Hence, the global morpho-phonological structure of the pseudowords was identical to that of the word targets, except that these inflected phonological structures had no root meaning. In order to avoid phonemic ambiguity, all the targets were presented with the vowel points (Frost, 1994; Frost & Bentin, 1992).

The noun phrase (the context) preceding each target consisted of a noun subject and an attribute (e.g., "The pretty girl...). The syntactic congruence between the target and its context was manipulated, forming two conditions. In the Congruent condition, the target agreed in gender and number with the syntactic structure of the context. In the Incongruent condition, gender and number with the syntactic structure of the context. In the Incongruent condition, either the gender or the number or both the gender and the number of the target were different than those of the context, thus violating the rules of agreement.

In producing congruent stimuli, we avoided physical identity between inflections that agree (i.e., the two inflections were not the same letters or phonemes). This was done to avoid the confounding of syntactic congruence with rhyming or orthographic repetition effects. In doing so, we took advantage of the fact that different words may take different inflectional morphemes (suffixes) to denote a given gender. Thus, the subjects and attributes that were selected to form the context phrase were inflected by a different inflection than the one used for the same purpose in the target. Take, for example, the syntactically congruent sentence "Hayalda hayafa rokɛdɛt" (The pretty girl is dancing). The subject "yalda" (girl) and the attribute "yafa" (pretty) use the suffix /a/ to denote the feminine form but the predicate "rokɛdɛt" (is dancing) uses the suffix /ɛt/ for the same purpose.

Half of the probes presented prior to the context were words and half were pseudowords. For each lexical category of the target (words or pseudowords), half of the trials were "match" trials, in which the target and the probe were identical, and half were "mismatch" trials in which the probe was different than the target. The probes used in the mismatch trials and their paired targets were different derivations of the same roots.

To summarize, there were 8 experimental conditions representing all combinations of three factors: a) lexicality (word, pseudoword), b) syntactic congruence (congruent, incongruent), and target-probe matching (match, mismatch). Each of the 72 target words and the 72 pseudowords were rotated across subjects to appear in each of the four possible combinations of syntactic congruence and matching. For example, in the following sentences the target word is *mitragɛsh* ("is anxious," masc., sing.):

1. Word-congruent-match: probe *mitragɛsh* - "Harakdan hamɛfursam mitragɛsh" (The famous dancer is anxious).
2. Word-incongruent-match: probe - *mitragɛsh*, sentence: "Harakdaniyot hamɛfursamot mitragɛsh" (The famous dancers [fem., pl.] is nervous).
3. Word-congruent-mismatch: probe - *ragish* (is sensitive, masc., sing.), sentence: "Harakdan hamɛfursam mitragɛsh" (The famous dancer is nervous).
4. Word-incongruent-mismatch: probe - *ragish*, sentence: "Harakdaniyot hamɛfursamot mitragɛsh" (The famous dancers is anxious).

The same context phrases were used for the target pseudoword *mitkatzɛsh*.

Hence, to complete this design, 576 sentences were needed, divided into 4 equal and balanced lists. A different group of 18 subjects was assigned to each list so that each subject was examined with 18 different targets in each condition and each target was presented in each condition to 18 subjects. Targets were rotated across subjects so that each target appeared with each context. Every experimental subject was presented with all conditions but received no repetitions of a given sentence.

Procedure. The experiment was conducted in a quiet and dimly lit room. The stimuli were presented on an MAC-SE computer using a standard Hebrew font. The same computer collected and timed key-press responses.

At the beginning of each trial a fixation mark was presented for 250 ms at the right end of the screen[4] placed vertically one line above the middle of the screen. The fixation point was replaced by the probe such that the first letter of the probe replaced the fixation point. The probe remained on for 350 ms, and was followed by a 200 ms blank ISI. The context phrase was presented next for 700 ms. It was located one line under the (now absent) probe. A second ISI of 200 ms blank period separated the context from the target which was placed on the screen as a natural continuation of the line. The target was on the screen until the subject answered or up to 2 s. Two additional seconds separated one trial from the other.

The subjects were instructed to silently read the probe, and read the context aloud while waiting for the target. Their task was to press one button if the target and the probe were identical and another button if they were not. They were instructed to answer as fast as they

Results

Means and standard deviations of reaction time for correct responses were calculated for each subject in each experimental condition. Reaction times that were slower or faster than two standard deviations from the subject's mean in each condition have been excluded fro the average. These outliers accounted for less than 5% of all responses.

For both match and mismatch trials, RTs were faster for syntactically congruent than for incongruent targets regardless of whether these were words or pseudowords. However, because the variance in the mismatch trials was high, we analyzed only trials on which the target matched the probe. These data are presented in Table 1.

The statistical significance of the apparent differences in RT was assessed by a CONGRUENCE (Congruent, Incongruent) × LEXICALITY (Words, Pseudowords) analysis of variance. A within-subjects design was used for the subjects analysis ($F1$) and a mixed-model design for items analysis ($F2$). These analyses showed that both main effects were significant. Correct match decisions were faster when the targets were syntactically congruent with the context (635 ms) than when they were incongruent (670 ms) [$F1(1,69) = 18.14$, MSe = 4959, $p < 0.0001$ and $F2(1,140) = 5.06$, MSe = 7723 $p < 0.05$], and word-targets (635 ms) were faster than pseudoword-targets (670 ms) [$F1(1,69) = 17.84$, MSe = 4783, $p < 0.0001$ and $F2(1,140) = 14.24$, MSe = 6574 $p < 0.0001$]. More importantly, the interaction between CONGRUENCE and LEXICALITY was not significant and its F-ratios were small [$F1\ (1,69) = 1.31$, MSe = 7114, $p > 0.2562$; and $F2(1,140) = 1.43$, MSe = 7696, $p > 0.2346$].

The percentages of pseudoword errors (8.9%) and word errors (8.3%) were similar, $F1(1,70) < 1.00$. For both words and pseudowords, more errors occurred in the syntactically incongruent condition (10.15%) than in the congruent condition (6.94%), $F1(1,70)= 11.45$, MSe = 64.58, $p < .001$. As was the case with RT, the interaction between CONGRUENCE and LEXICALITY was not significant, $F1(1,70) < 1.0$.

Discussion

A syntactic congruency effect was found in the present experiment even though the match/mismatch task could have been successfully accomplished without any reference to the target's sentential context. Because the context was not visible when the probe was processed and it was irrelevant to the task, the syntactic priming effect is not likely to have been related to the match/mismatch *decision*. More plausibly, it reflects the context's influence on encoding the target when it appeared. Because the target was a natural continuation of the sentence, its processing was presumably integrated into the context. Notice, however, that the target could not have been predicted on the basis of the semantic context of the preceding phrase. Moreover, the gender-agreement rule did not have any semantic consequences. This integration probably included an accounting of the congruence between the inflectional morpheme in the target and the syntactic expectancies evoked by the context (see Deutsch & Bentin, 1994 for an elaboration of this priming mechanism).

Table 1. Mean reaction time in ms, and percentage of errors for words and pseudowords in the syntactical congruent and incongruent conditions.

SYNTACTIC CONDITION	WORDS		PSEUDOWORDS	
	RT (SEm)	% ERRORS	RT (SEm)	% ERRORS
INCONGRUENT	659 (18.6)	9.98 (1.10)	682 (20.0)	10.33 (1.11)
CONGRUENT	612 (18.8)	6.69 (1.09)	658 (19.5)	7.16 (0 [illegible])

The most interesting result of the present experiment was, of course, the fact that significant syntactic priming was found for pseudowords as well as for words. Furthermore although the congruence effect was numerically bigger for words than for pseudowords, the interaction

The most interesting result of the present experiment was, of course, the fact that significant syntactic priming was found for pseudowords as well as for words. Furthermore although the congruence effect was numerically bigger for words than for pseudowords, the interaction between the two factors was not statistically significant. The syntactic congruence effect on processing the pseudowords, and the absence of the interaction between this effect and lexicality suggest that syntactic processes of inflectional agreement can proceed without a full analysis of the word's lexical representation and, therefore, without the intervention of lexical semantic information.

The syntactic priming effect on pseudowords is particularly interesting given that, in the pseudoword condition, both the probe (that preceded the context phrase) and the target were pseudowords. Hence, having the probe, the subject knew that the target was also going to be a pseudoword. Apparently, despite this knowledge, experimental subjects did not inhibit either their processing of the inflection or their formation of syntactically based expectations; they analyzed the inflection in the inflected pseudoword *vis-à-vis* the sentential syntactic structure. This outcome supports the hypothesis of an autonomous syntactic processor.

Although there was no significant interaction between lexicality and syntactic congruence, the numerical difference in the size of the syntactic priming effect between the word condition and the pseudoword condition weakens support for the autonomy hypothesis and suggests, instead, that the syntactic process may not be completely independent from lexical semantic information. This suggestion was followed up in Experiment 2.

EXPERIMENT 2

As Experiment 1 suggested, although an inflectional analysis of the stimulus occurred whether it did or did not have a lexical entry (i.e., a semantic sense), the process of analysis may, nevertheless, be modulated by lexical-semantic factors.

It is usually difficult to disentangle semantic from purely grammatical aspects of syntactic information. However, the gender-agreement rule in Hebrew provides one such approach. As previously mentioned, all the nouns in Hebrew (animate and inanimate) are marked for gender, either masculine or feminine. Within animate word pairs, the masculine is typically the unmarked form and the feminine gender the marked. The same inflectional markers used for animate feminine nouns (the suffixes described above) are used to mark the feminine for inanimate subjects. This consistency provides the experimenter with a way of distinguishing between the purely syntactic use of gender (i.e., for inanimate nouns which can have no sexual gender) and a syntactic-plus-semantic use of gender (i.e., for animate nouns).[5] Thus, because the same inflectional markers are used to denote both grammatical and real gender, manipulation of the congruency of gender inflections may affect both grammatical *and* semantic/pragmatic processes for animate nouns, but only grammatical processes for inanimate nouns. Consequently, differences in the effects of manipulating the subject-predicate gender-agreement rule for inanimate versus animate subjects may reflect the difference in processing syntax which has only formal syntactic consequences (e.g., agreement) versus processing syntax which has, in addition, semantic consequences (e.g., sexual gender). If the syntactic processes that entail the analysis of agreement are influenced by semantic factors, the interference caused by violation of the gender agreement may have a stronger effect for animate than for inanimate nouns.

Because in the present experiment all the target items were words, we were able to avoid certain complications caused by the decision task that was used Experiment 1. In Experiment 1, RT speed may have been affected by nuisance factors related to the decision itself (i.e., the decision same vs. different). In the present experiment we avoided that potential problem by using a naming task.

Although naming might involve only pre-lexical phonological computation, and therefore, in principle, need not be sensitive to higher-level linguistic processes, recent models and data on naming in shallow languages such as Serbo-Croatian (Carello, Turvey, & Lukatela, 1992; 1994) as well as in deep languages such as English (McCaan & Besner, 1987) and Hebrew (Frost, 1995) have demonstrated that lexical information shapes the pre-lexical computation in a top-down manner. Moreover, when the orthographic pattern of the word is very familiar or when the lexicon directly (for a recent discussion of this alternative option in naming, see Bentin & Ibrahim, in press). In order to encourage direct lexical access, all the stimuli in the present experiment were typed in unpointed Hebrew letters.

Method

Subjects. The subjects were 48 undergraduate students who did not participate in the first experiment. They were all native speakers of Hebrew, who took part in the experiment for course credit or for payment.

Stimuli and design. The critical stimuli in this experiment were 48 word-targets that were the predicates concluding three-word sentences. Each target was embedded in four different sentential contexts, one of the four different combinations of a 2 × 2 design: ANIMACY (Animate, Inanimate) × SYNTACTIC CONGRUENCY (Congruent, Incongruent). In animate contexts the subject of the sentence was an animate noun. In inanimate contexts, the subject of the sentence was an inanimate noun. Half of the noun subjects in each animacy condition were masculine gender and half were feminine. Each noun subject was followed by a congruently inflected attribute. The syntactic congruence between the target (predicate) and its sentential context was manipulated within each animacy condition, forming two levels of the SYNTACTIC CONGRUENCE factor. In the Congruent condition the target was inflected to agree in gender with the subject and attribute. In the Incongruent condition the inflection of the target did not agree in gender with the context. Take, for example, the target "*nafal*" (fell down). The four different trial types in which this target was used were:

1. Animate-congruent: "The suspicious judge fell down" which is, in Hebrew,: "Hashofɛt (sub., masc.) haxshdan (attrib., masc.) *nafal* (pred., masc.)."
2. Animate-incongruent: "Hashofɛt*ɛt* (fem.) haxshdan*it* (fem.) *nafal* (masc.)."
3. Inanimate-congruent: "The shiny fork fell down" which, in Hebrew, is: "Hamazlɛg (sub., masc.) hanozɛz (attrib., masc.) *nafal* (pred., masc.)."
4. Inanimate-incongruent: "The shiny spoon fell down": "Hakap*it* (sub., fem.) hanozɛz*ɛt* (attrib., fem.) *nafal* (pred., masc.)."

As in the previous experiment, phonetic priming was avoided by using subjects and attributes that take different inflectional morphemes to denote the feminine gender that taken by the target.

The resulting 192 sentences (48 targets presented in four context conditions) were divided into 4 stimuli lists. Each list included 48 different sentences, 12 in each of the four animacy/congruence conditions. Twelve different subjects were randomly assigned to each list. Thus, each subject read sentences in every experimental condition and each target was presented (across subjects) in all conditions. This rotation allowed a two-factorial within-subject ANOVA with both subjects and items as random factors.

Procedure. The sentential context appeared on the center of the screen for 2000 ms, followed by the target word which appeared as a continuation of the sentence. The target remained on the screen for 1000 ms. RTs were measured from the onset of the target until the onset of naming. Subjects were asked to read aloud the context as well as the targets, but speeded performance was required only for the targets. The experiment started with a practice session of 16 sentences.

Results

Naming RTs that were shorter than 150 ms accounted for less than 3% of the responses and were discarded. Means and standard deviations were calculated for each subject and separately for each target in each of the conditions. Outliers of more than two standard deviations accounted for less than 0.5% of the responses and were excluded from the recalculated averages.

For both animate and inanimate context, congruent predicates were named equally fast. Syntactic incongruence led to slower responses. Importantly, the congruency effect was three times as large in the animate than in the inanimate condition (Table 2).

Table 2. *Mean reaction time in ms and SEm, in parentheses, for targets in the syntactically congruent and incongruent context, for animate and inanimate conditions.*

CONGRUENCY CONDITION	ANIMACY CONDITION ANIMATE	INANIMATE
INCONGRUENT	560 (11.9)	538 (10.3)
CONGRUENT	528 (10.8)	528 (9.3)
CONGRUENCY EFFECT	**32 ms**	**10 ms**

ANOVA showed that the syntactic congruency effect was statistically reliable by both subjects [$F1(1,47)=16.33$, MSe = 1311, $p < .0001$] and items analyses [$F2(1,47) = 3,68$, MSe = 868, $p < .0001$]. The interaction was significant for subjects [$F1(1,47)=6.43$, MSe = 925, $p < .025$], but not for items [$F1(1,47) = 2.54$, MSe = 1994, $p < .12$]. Post-hoc comparisons on subject means revealed that while the syntactic congruency effect was significant for animate condition [$t(47)= 18.73$, $p < 0.001$), it was not reliable for the inanimate condition [$t(47)= 2.65$, $p > 0.11$). Moreover, a planned t-test revealed that the 32 ms congruency effect for the animate targets was significantly bigger than the 10 ms congruency effect for the inanimate targets, both for the subjects [$t(47)=3.48$, $p<0.001$] and items analyses [$t(47)=2.38$, $p < 0.025$]. The 10 ms congruency effect for inanimate nouns was not significant.

Discussion

The most important outcome of the present experiment was that the syntactic priming effect was significantly bigger when the subject of the sentence was animate than when the sentence had an inanimate subject. Because syntactic priming in the present experiment was related to the agreement or disagreement in *gender* between the predicate and the preceding noun-phrase denoting the subject of the sentence, it is evident that the interaction between the syntactic congruency effect and the animacy of the subject reflected the difference in the role of gender for animate and inanimate sentence-subjects. Apparently, readers are more disturbed by violation of gender-agreement when the gender has a semantic/pragmatic value than when it denotes an arbitrary, pure syntactic agreement. The sensitivity of the syntactic process to the semantic meaning may indicate that the inflectional processor is exposed to semantic information of the word, and not just to its grammatical characteristics.

Before concluding this section two caveats should be considered. One is the possibility that all the observed priming effect is explained by semantic rather than syntactic factors. Such a suspicion might be elicited, for example, by previous studies in which syntactic priming was found in lexical decision but not in naming (Carello et al., 1988; Seidenberg et al., 1984; Sereno, 1991) or in which, relative to lexical decision, syntactic priming in naming was significantly attenuated (West & Stanovich, 1986). However, it is unlikely that semantic factors solely accounted for the present priming effect. This claim is supported primarily by the fact that the same predicates (targets), were used with both animate and inanimate targets. Therefore the semantic relationships within the sentences in both conditions were very similar and should have produced equal effects. Moreover, none of the targets was semantically or associatively related to the preceding words in the context or could have be predicted on the basis of the sentence semantic context. In addition, it is logically necessary to process the inflectional incongruence between the subject and the predicate in order to experience any semantic incongruence.

The second caveat is that, relative to the animate group of nouns, the inflectional system for the group of inanimate nouns is less regular; in this group there are more exceptions in which a masculine noun takes a feminine plural suffix (and vice versa). It is possible, therefore, that the difference observed between the animate and inanimate conditions stemmed from an effect of ambiguity (i.e., predictability) due to their difference in inflectional consistency. Experiment 3 was designed to control for this possibility.

EXPERIMENT 3

The irregularity of the inflectional system associated with the gender of inanimate nouns is particularly conspicuous in plural form. As mentioned above, two suffixes in Hebrew are u·ʳd to denote a plural: one that is regularly used for masculine /im/ and one regularly used for feminine /ot/. However, whereas for words denoting animate concepts this rule is almost always true, for inanimate concepts there are irregular cases. Most of these cases are masculine nouns that take the /ot/ suffix to denote the plural form (and use masculine inflections for their predicates), but there are also a few in which feminine nouns accept the /im/ suffix to denote the plural form. Thus, although for inanimate nouns there is a correlation between the inflectional structure of a word and its grammatical gender, their inflectional structure coincides with their gender less regularly than it does for animate nouns. Accordingly, the native speaker of the language may be less disturbed by gender disagreement in inanimate nouns because the inflectional system is less regular than for animate nouns.

In order to examine the effect of inflectional regularity on the syntactic priming effect, in the present experiment we compared this effect for "regular" and "irregular" nouns. As in the previous experiments syntactic priming was induced by violating the gender subject-predicate agreement, while manipulating gender-regularity in inanimate nouns. If the difference between the syntactic priming effect for animate and inanimate subjects reflected mainly the difference between processing inflectional regular and irregular word categories, a smaller syntactic priming effect should be found for irregular than for regular forms.

Method

Subjects. The subjects were 48 undergraduates students, who did not participate in any of the two previous experiments. They were all native speakers of Hebrew, who took part in the experiment for course credit or payment.

Stimuli and design. In the present experiment we used 48 target words. Becau· :, as described in the introduction, most irregular nouns are masculine, we could not manipulate the agreement between the subject and the predicate by changing the noun phrase (context) while keeping the predicate (target) intact. Therefore, unlike in the previous experiments, the masculine form of the target was used in the congruent condition while its feminine form was used in the incongruent condition. The same target was used, however, within congruity conditions for both regular and irregular nouns. Take for example the target "fell down" which in the masculine form sounds "nafal" whereas in the feminine forms sounds "nafla." This target was used in conjunction with the regular masculine noun /yahalom/ (dimond) (in plural form yahalom*im*), and with the irregular masculine noun /mazleg/ (fork) (in plural form mazleg*ot*) to form the following 4 experimental conditions:

1. Regular-congruent: "Hayahalom (sub. masc.) hanotzεz (attrib. masc.) *nafal* (pred. masc)" - (The shining diamond fell down).
2. Regular-incongruent: "Hayahalom (masc.) hanotzεz (masc.)*nafala* (fem.)."
3. Irregular-congruent: "Hamazlεg (sub. masc.) hanotzεz (attrib. masc.) nafal (pred. masc.)" - (The shiny fork fell down).
4. Irregular-incongruent: "Hamazlεg (masc.) hanotzεz (masc.) nafala (feminine)."

Each subject was examined in all four conditions, using different targets in each condition. This design allowed a within subject (F1) and within item (F2) ANOVA design with REGULARITY (regular, irregular) and SYNTACTIC CONGRUITY (congruent, incongruent) as main effects .

Procedure. Experimental procedure of Experiment 3 was identical to this of Experiment 2.

Results

As in the previous experiment, responses that were shorter than 150 ms (less than 4% of the responses) and outliers of more than 2 Sds (less than 3%) were excluded. For sentences having an irregular noun subject as well as for sentences with a regular subject, syntactically congruent targets were named faster than syntactically incongruent targets (Table 3).

Table 3. *Mean reaction time in ms and SEm, in parentheses, for regular and irregular inanimate targets in the syntactically congruent and incongruent context.*

CONGRUENCY CONDITION	REGULARITY CONDITION REGULAR	IRREGULAR
INCONGRUENT	632 (12.1)	637 (12.8)
CONGRUENT	599 (13.0)	609 (11.9)
CONGRUENCY EFFECT	**33 ms**	**28 ms**

ANOVA corroborated these observations showing a significant syntactic congruity effect [$F1(1,47) = 22.94$, MSe = 2036, $p < 0.001$, $F2(1,47) = 18.06$, MSe = 2887, $p < 0.001$], and no significant effect of regularity [$F1(1,47) = 1.54$, MSe = 1709, $F2(1,47) < 1.0$]. Most importantly, as revealed by a non-significant interaction [$F1(1,47)<1.0$, $F2(1,47) < 1.0$], the syntactic congruity effect was similar for regular and irregular subject nouns.

Discussion

The similarity of the syntactic priming effect induced by the violation of subject-predicate agreement for regular and irregular nouns suggests that the syntactic priming effect is not influenced by the inflectional transparency of the noun gender classification. That is, the difference between animate and inanimate nouns in sensitivity to agreement in gender was not due to the existence of greater inflectional inconsistency in the inanimate group. Hence, these results help rule out the concern that the difference between the syntactic priming effects observed for animate and inanimate nouns was mediated by inflectional rather than semantic factors. At the same time, however, the magnitude of the syntactic priming effect for *inanimate* noun phrase contexts contrasts with the pattern of results in Experiment 2 and may be in conflict with their interpretation.

In Experiment 2 we have found a very small (and statistically unreliable) syntactic priming effects for inanimate nouns. In contrast, in the present experiment the magnitude of the syntactic priming effect was bigger, in fact, as big as the syntactic priming effect in for animate noun phrases in Experiment 2. There are several possible explanations for this difference. One stems from the fact that in Experiment 3, unlike in the previous experiments, all congruent targets were masculine (i.e., not inflected) while all incongruent targets were feminine (i.e,. inflected). Therefore, the syntactic congruence effect was confounded (al least in part) with simple inflectional, or even phonetic effects because the feminine-inflected targets were more complex and longer than the masculine targets. It is well known that naming time is positively correlated with the length of the word (Frederiksen & Kroll, 1976).

To assess the hypothesis that feminine nouns take longer to name because they are longer in length, we ran a separate group of subjects in a task of naming each stimulus in isolation using the specific stimuli used in Experiment 2. Although there was a difference in the expected direction (masculine nouns averaged 506 ms and feminine nouns 510 ms), the difference was small and not significant.

A second possible explanation is that the processing difference found between sentences with animate or inanimate subjects was strategic: induced, in Experiment 2, by mixing the two noun categories (every experimental subject received both kinds of stimuli). It is possible that this mixture sensitized the subject to the difference between these two categories thereby affecting sentence processing strategies. Additional research is necessary to examine these explanations.

GENERAL DISCUSSION

In the present study we examined the independence of the relationship between syntactic processes based on Hebrew inflectional morphology, and lexical and semantic factors. In three experiments, we manipulated the gender agreement between the noun phrase and the predicate in three-word sentences. This manipulation induced a syntactic priming effect reflected by a faster processing of syntactically congruent than of syntactically incongruent targets.

In Experiment 1 we found that the magnitude of this syntactic priming effect was similar for word and pseudoword targets. This pattern suggests that syntactic processes based on inflectional morphology are automatically applied to all phonologically legal structures, regardless of whether they are or they are not represented in the lexicon. In this respect, the present results are similar to those found when inflectional morphology of case agreement between adjective and noun was manipulated in Serbo-Croatian (Katz et al., 1987). Using a lexical decision task with spoken stimuli these authors found an equivalent syntactic priming effect when the prime was a meaningless pseudoadjective as well as when it was real adjective. However, the results of Experiment 2 suggested that syntactic processing is not indifferent to the semantic characteristics of the target. In that experiment we have found that disagreement in gender between the predicate target and the preceding noun phrase delayed naming the target significantly more if the subject of the sentence was an animal or a human being (i.e., a word whose grammatical gender correlated with one of its semantic/pragmatic values) than if the subject was inanimate, (i.e., a word whose grammatical gender had only grammatical value). The result that naming in the animate incongruent condition was slower than the other three conditions (which were all equal among themselves) suggests that the effect was inhibitory. Finally, the results of Experiment 3 showed that the difference in syntactic priming for sentences including animate and inanimate subjects was not accounted for by the relatively higher percentage of irregular inflection for gender in the inanimate than in the animate nouns.

The syntactic priming effect on processing inflected pseudowords suggests that the inflectional analysis of phonological stimuli (on which the syntactic processor could have operated), does not require full activation of a specific lexical entry. In other words, when the reader is exposed to an orthographic representation of an inflected phonological unit, inflectional analysis of the stimulus is initiated to identify its grammatical characteristics. The initiation (and probably the successful completion) of this process probably does not depend on the successful completion of lexical access or semantic identification. Such a description would be in accord with the inflectional decomposition conception of lexical organization, by which a connection is established between inflectional base unit and the various inflectional affixes with which it usually combines in the language (Marslen-Wilson, Tyler, Waksler, & Older, 1994).

This is not to say, however, that the syntactic process is completely independent of the lexical status of a stimulus (i.e., word or pseudoword) and, if the stimulus is a word, its semantics. The results of Experiment 1, although not significant, hinted at a stronger congruity effect for words than for pseudowords, a trend we have found repeated in several unpublished experiments in our laboratory. The animacy manipulation in Experiment 2 provided more direct support for an interaction between syntactic, lexical, and semantic cognitive information. Since the animacy value of a word is an fundamental part of a word's semantic characteristics, the influence of this factor on syntactic priming indicates that the syntactic processing of inflectional morphology is sensitive to lexical and semantic processes.

Semantic information may, for example, support the processing of a sentence in relatively late stages of sentence integration. This interpretation is supported by the asymmetrical effect of animacy on the congruent and incongruent conditions. If animacy would have had affected the process of identifying the grammatical characteristics of the word, the difference between the animate and the inanimate conditions would have been observed in both the congruent and the incongruent condition. However, naming of congruent predicates was equivalent for animate and inanimate noun subjects.

On the other hand, the effect of animacy can not be so late as to be irrelevant to word recognition because the interaction between animacy and syntactic congruence occurred for a process that is considered to be relatively shallow: naming. Naming is considered to require minimal contact with the lexicon as opposed to "deeper" tasks like lexical decision. The interaction result is consistent with previous studies in which syntactic congruence affected word identification, and supports our previous suggestion that syntactic priming and its interaction with semantic information occurs at a relatively low level of processing (Deutsch & Bentin, 1994). We conjecture that the process of word identification is supported in parallel by many levels of linguistic analysis, phonological, inflectional and semantic.

In summary, the above interpretation may fit into an interactive model of linguistic system where various processes associated with various aspects of the linguistic input may operate independently. However, possible mutual connection between these processes may facilitate or inhibit each of this processes or the operation of the whole system as a unit.

REFERENCES

Bentin, S., & Frost, R. (1994). Inflectional factors in visual word identification in Hebrew. In L. B. Feldman (Ed.), *Inflectional aspects of language processing*. (pp. 273-294). Hillsdale NJ: Erlbaum

Bentin, S., & Ibrahim, R. (in press). New evidence for phonological processing during visual word recognition and naming: The case of Arabic. *Journal of Experimental Psychology: Learning Memory, and Cognition*.

Carello, C., Lukatela, G., & Turvey, M. T. (1988). Rapid naming is affected by association but not by syntax. *Memory & Language, 16*, 187-195.

Carello, C., Turvey, M. T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 211-226). Amsterdam: Elsevier.

Deutsch, A., & Bentin, S. (1994). Attention mechanisms mediate the syntactic priming effect in auditory word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 595-607.

Ferreira, F., & Clifton, C. J. R. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Ferreira, F., & Henderson, J. M. (1990). The use of verb information in syntactic parsing: A comparison of evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 555-568.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing* (pp. 25-86). Hillsdale, NJ: Erlbaum.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 361-379.

Frost, R. (1994). Prelexical and postlexical strategies in reading: evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 116-129.

Frost, R. (1995). Phonological computation and missing vowels: Mapping lexical involvement in reading. *Journal of Experimental Psychology: Learning Memory, & Cognition, 21*, 1-11.

Frost, R. & Bentin, S. (1992). Reading consonants and guessing vowels: Visual word recognition in Hebrew orthography. In R. Frost & L. Katz (Eds.), Orthography, phonology, morphology, and meaning (pp. 27-44). Amsterdam: Elsevier.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human, Perception and Performance, 13*, 105-115.

Holmes, W. M., Stowes, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language, 28*, 668-689.

Goodman, G. O., McClelland, J. L., & Gibbs, R. W. J. R. (1981). The role of syntactic context in word recognition. *Memory & Cognition, 9*, 580-586.

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by inflected adjectives. *Cognition, 19*, 55-71.

Katz, L., Boyce, S., Goldstein, L., & Lukatela, G. (1987). Grammatical information effects in auditory word recognition. *Cognition, 25*, 235-263.

Marslen-Wilson, W. D., & Tyler, L. K. (1987). Against modularity. In J. Garfield (Ed.), *Modularity in knowledge representations and natural language understanding* (pp. 37-62). Cambridge, MA: MIT Press.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*, 3-33.

McCaan, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models pronunciation assembly and the locus of word frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 13-24.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 3-36). Hillsdale, NJ: Erlbaum.

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp. 601-618). Hillsdale, NJ: Erlbaum.

Mitchell, D. C. (1989). Verb guidance and other lexical effect in parsing. *Language and Cognitive Processes, 4*, 123-154.

Munte, T. H., Heinze, H. J., & Mangun, G. R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. Journal of *Cognitive Neuroscience, 5*, 335-344.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentence. *Journal of Verbal Learning and Verbal Behavior, 22*, 358-374.

Rosler, F., Putz, P., Friederici, A., & Hahne, A. (1993). Event-related brain potential while encountering semantic and syntactic constraint violation. *Journal of Cognitive Neuroscience, 5*, 345-362.

Seidenberg, M., S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre-and postlexical loci of contextual effects on word recognition. *Memory & Cognition, 12*, 315-328.

Sereno, J. A. (1991). Graphemic, associative, and syntactic priming effects at a brief stimulus onset asynchrony in lexical decision and naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 459-477.

Stowe, L. A. (1989). Thematic structures and sentence comprehension. In: G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistics structure in language processing* (pp. 319-358). Dordrecht: Kluwer Academic.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic context. *Journal of Verbal Learning and Verbal Behavior, 18*, 427-440.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Taraban, R., & McClelland, J. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language, 27*, 597-632.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285-314.

Tyler, L. K. & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics, 34*, 409-420.

West, R. F., & Stanovich, K. E. (1986). Robust effect of syntactic structure on visual word recognition. *Memory & Cognition, 14*, 104-112.

FOOTNOTES

[†]Hebrew University, Jerusalem

[‡]Also Hebrew University, Jerusalem.

[†††]Also University of Connecticut, Storrs.

[1]The minimal attachment principle postulates that the initially preferred syntactic parsing is the one which entails the minimal number of syntactic nodes. Accordingly, the initial parsing of a sentence that includes a prepositional attachment ambiguity will be of a simple active sentence in which the prepositional phrase will be attached to the main verb phrase. For example in the sentence "The spy saw the cop with a revolver." the prepositional phrase "with a revolver" will be initially attached to the main verb phrase "saw" rather than to the preceding noun phrase "the cop."

[2]Except for few cases of specific nominal sentences.

[3]Another agreement rule requires that the subject and attribute will agree in gender, number and definity. Accordingly, in the above example, the form of the attribute ("haxashdan") has also changed ("haxashdanit") when the masculine subject noun phrase had been replaced by a feminine noun.

[4]Hebrew is written from right to left.

[5]For example, the inanimate noun "knisa" (an entrance) uses the suffix "a" to denote its feminine gender like it is used to change the masculine "yɛlɛd" (a boy) into the feminine "yalda."

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 235-254

Decomposing Words into their Constituent Morphemes: Evidence from English and Hebrew*

Laurie Beth Feldman,† Ram Frost,‡ and Tamar Pnini†††

Evidence for the psychological reality of the morpheme was observed in a segment shifting task. This task, modeled after the reordering of segments that occurs in spontaneous speech (speech errors), requires that subjects segment and shift a sequence of letters from a source word to a target word and then name the product aloud. Morphemic and nonmorphemic letter sequences from morphologically complex words such as HARDEN (morphemic) and their morphologically simple (nonmorphemic) controls such as GARDEN were phonemically matched. In a series of four experiments, naming latencies were faster for morphemic sequences than their nonmorphemic controls in both English, in which the morphemic status of the shifted sequence was varied and sequences were appended after the base morpheme (linearly concatenated), and in Hebrew, in which morphological transparency of the root (base morpheme) was varied and one morpheme was infixed inside the other (nonconcatenative) so that the phonological and orthographic integrity of the morphemic constituents was disrupted. Moreover, the likelihood with which both affixes and bases combine to form words influenced segment shifting times. In conclusion, skilled readers in both languages are sensitive to the morphological components of words whether or not they form contiguous orthographic or phonological units.

Speakers and readers possess knowledge about what words mean and how they sound and this knowledge constitutes the mental lexicon. A major theme for theories of language processing is how lexical knowledge is organized. One crucial line of inquiry focuses on units and contrasts word- and morpheme-based accounts (e.g., Caramazza, Laudanna, & Romani, 1988). Related to the morphological units position is the issue of how regular and irregular morphological formations are represented in the lexicon (e.g., Fowler, Napps, & Feldman, 1985; Stemberger & MacWhinney, 1986, 1988; Pinker, 1991). A second line of investigation distinguishes between units for access and units for representation. Proponents of the full word position (e.g., Butterworth, 1983) claim that words composed of several morphemes (morphologically complex) are represented in the lexicon as are morphologically simple words, without regard to morphological structure. In fact, some theorists (e.g., Seidenberg, 1987; Seidenberg & McClelland, 1989; but see Rapp, 1992) have claimed that morphological effects can be accounted for in terms of similar form and/or similar meaning without invoking morphological units at all. By this account, morphological effects reflect the covariation of forms and meaning that characterize a language. By contrast, morpheme-based accounts (e.g., Feldman, 1994; Laudanna, Badecker, & Caramazza, 1992; MacKay, 1978) focus on the facilitatory or inhibitory interaction among morphemic units in the lexicon. Issues of lexical representation may or may not be logically dissociable from issues of units for lexical access (Henderson, 1985). Nevertheless, it is

We wish to thank Avital Deutsch, Carol Fowler, Albrecht Inhoff, Jim Neely, Philip Smith and several anonymous reviewers for their suggestions. The research reported here was conducted at The University of Delaware, SUNY at Albany and The Hebrew University and was supported by funds from National Institute Of Child Health and Development Grant HD-01994 to Haskins Laboratories. Portions were presented to the November 1987 meeting of the Psychonomic Society.

sometimes claimed that prefixes are stripped from complex words and that decomposition to a base morpheme is necessary before access to the lexicon can occur (Taft & Forster, 1975; Taft, 1979). Alternatively, it is sometimes claimed that constituent structure does not reliably influence access (e.g., Henderson, Wallis, & Knight, 1984; Manelis & Tharp, 1977; Rubin, Becker, & Freeman, 1979). In summary, in part because different theorists are searching for different types of morphological effects and in part because other relevant variables may have been confounded with morphology (see Henderson, 1989; Smith, 1988), the status of morphemes as psychologically relevant structures in word recognition is not unambiguous.

Across languages, the structure and prevalence of morphologically complex words varies. In *isolating* languages such as Chinese or Vietnamese, words tend to be monomorphemic and cannot be analyzed systematically into smaller meaningful components. By definition, morphemes in isolating languages tend to be represented as physically distinct units. In *agglutinating* languages such as Turkish, morphological elements are appended to a base form although the particular form of the affix may be phonologically and orthographically modified by properties of the base morpheme. In *inflecting* or fusional languages such as English and Hebrew, words are sometimes composed of multiple morphemes but the boundary between morphemes is not always straightforward. Stated generally, across languages, words differ with respect to the phonological and orthographic variability of their morphemes and this influences the salience of constituent morphemes (Comrie, 1981). This structural variation may have implications for how the components of a word are processed.

Morphological processing in concatenative languages

The most common type of morphological formation in inflecting languages consists of affixation of an element to a base morpheme (Matthews, 1974). Affixation encompasses three processes, defined by the position in which the affixation occurs. These include prefixation, suffixation, and infixation, respectively, in positions initial, final and internal to the base morpheme. By definition, prefixation and suffixation entail the linear concatenation of elements, whereas infixation is nonconcatenative insofar as the integrity of the base morpheme is disrupted. The tendency in English and other Indo-European languages is to concatenate (prefixes and) suffixes to the base (and to retain the base morpheme intact). Hebrew, by contrast, relies on the intertwining of two morphemes; a skeleton of consonants (i.e., root) and a phonological pattern of vowels[1] (i.e., the word pattern) (McCarthy, 1981). When a word pattern is infixed within the root, the integrity of the root morpheme is necessarily compromised relative to concatenated combinations. English and Hebrew contrast, therefore, in the principle by which morphological units are combined and this may have implications for how lexical access for words composed of multiple morphemes occurs in the two languages.

Alternatively, it is possible that morphological effects reflect representation in the lexicon such that contrasts between the affixing inflectional morphology of English and the infixing (and affixing) inflectional morphology of Hebrew pose no special problem. The appeal to abstractness of lexical representations across types of inflecting languages is intended to parallel the argument across modality of presentation whereby access representations but not lexical representations for visual and auditory forms are distinct (Marslen-Wilson, Tyler, Waksler, & Older, 1994). The experimental evidence in concatenative languages for lexical representation of morphology is based primarily but not exclusively on visual word recognition methodologies and how the pattern of decision latencies is influenced by (a) the particular combination of morphological components, (b) the frequency or productivity of constituent morphemes, or (c) repetition, across successive trials, of a morphological component. Some theorists capture knowledge about morphology in terms of lexical representations whose component structure is morphologically decomposed (e.g., Caramazza, Laudanna, & Romani, 1988). Other theorists capture knowledge about morphology in terms of a principle of lexical organization among full forms that are morphological relatives (e.g., Lukatela, Gligorijević, Kostić, & Turvey, 1980; Lukatela, Carello, & Turvey, 1987). For example, in Italian, rejection latencies in a lexical decision task for nonwords composed of illegal combinations of real morphemes (viz., verbal stems and affixes) vary as a function of the type of violation between stem and affix (Caramazza et al., 1988). Similarly, in a phoneme monitoring task with Dutch materials, the identification of words that include stems and prefixes is easier than that of words composed of stems without a prefix (Schriefers, Zwitserlood, & Roelofs, 1991). Moreover, rejection latencies for Dutch pseudowords are sensitive to the productivity of their morphemic components (Schreuder & Baayen, 1994). These studies provide evidence of the psychological reality of the morpheme as a sub-word unit that cannot be purely prelexical in locus because latency is sensitive to the combination of morphemes.

It is also the case that decision times for morphologically-complex forms in English are influenced by the cumulative frequency of all forms that include its base as well as by the surface frequency for that particular form (Taft, 1979). This is true both when the shared base morpheme differs in spelling (Kelliher & Henderson, 1990) and when spelling is preserved (Katz, Rexer, & Lukatela, 1991; Nagy, Anderson, Schommer, Scott, & Stallman, 1989; Taft, 1979). Similar effects have also been observed in Italian (Burani, Salmaso, & Caramazza, 1984). Evidently, morphological knowledge can be represented in a manner that is sufficiently abstract to tolerate changes in surface form. Moreover, the morphological mechanism is sensitive to the frequency with which a base morpheme appears across words and to the number of formations into which an affix enters.

Finally, in the repetition priming task, decision latencies to English words (e.g., CAR) preceded earlier in the list by a morphological relative (e.g., CARS) were faster than to targets (a) presented for the first time (Fowler, Napps, & Feldman, 1985), (b) preceded by an unrelated word (e.g., CARD) that was orthographically similar and (c) preceded by a semantically-related word (Napps, 1989; Napps & Fowler, 1987). Similarly, in a double lexical decision task, decision latencies to unrelated prime-target pairs formed from homographic base morphemes (e.g., PORTE "doors" and PORTARE "to carry") were slowed relative to orthographic controls (Laudanna, Badecker, & Caramazza, 1989). Moreover, for pairs formed from the same base morpheme, the effect depended on the type of morphological relation (viz., inflection, derivation) that exists between members of the pair (Laudanna, Badecker, & Caramazza, 1992; Marslen-Wilson et al., 1994). Finally, for derivational relatives presented in a cross-modal paradigm, the outcome depended on whether affixes preceded or followed the base morpheme (Marslen-Wilson et al., 1994). These findings suggest that a morphological principle of organization among either base morphemes or whole forms is present in the lexicon. Either activation spreads among whole word forms that share a base morpheme or morphological relatives are represented as compositional variations of the same base morpheme. Stated generally, there is experimental evidence from a variety of word recognition tasks in a variety of concatenative languages, in particular English, Italian and Dutch, that morphemes are psychologically real and that morphological effects reflect more than simply similar meanings or similar forms.

Morphological processing in Hebrew, a nonconcatenative language

Most words in Hebrew are composed of two morphemes, a root and a phonological word pattern. Although both are morphemes (Berman, 1978), the semantic contribution of each morpheme is not equal. The semantic information contributed by the root is usually more salient than that of the word pattern. The root conveys the core meaning of the words formed around it. The word pattern, by contrast, may in some cases carry nothing more than word class information. It is possible, therefore, that morphological processing in Hebrew is dominated by the semantic analysis of a word's root.[2] Consistent with this claim are findings in a repetition priming study (Bentin & Feldman, 1990) showing that effects of morphological relatedness among words that share the same root were evident even at long lags whereas effects of semantic association were evident only at short lags. This outcome suggests that different mechanisms underlie morphological processing and the appreciation of semantic association even though both rely on semantic analysis of the root.

Vowels and consonants in Hebrew are not represented in print in the same manner, and this may have implications for the processing of morphological structure as well. Consonants are represented by letters. Vowels are generally depicted by diacritical marks (points and dashes) presented beneath (or sometimes above) the consonant letters, although some vowels can also be conveyed by letters. Roots are composed of consonants. Word patterns are composed mainly (but not exclusively) of vowels. It is the convention that diacritical marks are omitted from most reading material although pointed text can be found in poetry, children's literature, and religious scripts. Stated generally, the morphological information conveyed by the Hebrew orthography is more salient and unambiguous for roots which are composed of consonants than for word patterns which are composed predominantly of vowels (see Frost & Bentin, 1992; Bentin & Frost, in press).

In addition to the difference between roots and word patterns, morphological relatives in Hebrew and English contrast because of the process of infixation. That is, in Hebrew, the entire root does not necessarily appear as an uninterrupted phonological unit. Rather, it may be distributed across multiple syllables of the word. In its printed form, however, because vowels tend not to be represented, the root often forms an orthographic unit. For example, in printed text, ZEMER ("a song") and ZAMAR ("a singer") will be printed in the same (right-to left) manner (i.e., זמר) although they will be pronounced differently. Because, in Hebrew, vowels tend not to be printed and because some morphologically-related forms differ primarily with respect to the vowels that are infixed among the consonants that compose the root, related forms will tend to differ more with respect to their pronunciation than to their visual form. As regards the repetition priming task in Hebrew, in which successive visual presentations of words formed from the same root reduce target decision latencies, the contributions of visual and morphological factors are not easily differentiated (but see Feldman & Bentin, 1994). Materials are typically constructed so that the repeated component is the root and target facilitation is observed when that component is presented in the prime and then in the target.

The repetition priming paradigm has been used to examine morphological processing of disrupted and continuous Hebrew roots but, as noted above, the role of word patterns in that task may be minimal. The present study uses a relatively new methodology that entails manipulation of the word pattern. An examination of how word patterns and roots are processed is important because the semantic character of the word pattern is relatively indistinct as compared with that of the root. Moreover, word patterns are morphemes that are never realized as continuous orthographic or phonological sequences appended to a root.

Segment Shifting Task

Laboratory-induced errors of the type that occur in spontaneous speech (Dell, 1987; Fromkin, 1973; Stemberger, 1984) have recently been examined by means of the segment shifting task (Feldman, 1991; Feldman & Fowler, 1987). In this task, subjects are instructed to segment and shift a designated segment from a source word onto a target word and to name the new result aloud as rapidly as possible. If the morphological structure of source words is analyzed and decomposed in the course of segmenting the designated segment and building up the utterance to be named, then morphological effects are anticipated in this task.

The experimental manipulation exploits the fact that the same sequence of letters (e.g., EN) can function morphemically in one source word (e.g., HARDEN) and nonmorphemically in another (e.g., GARDEN). The manipulation is designed to determine whether naming latency for a word formed from a target and a shifted segment (e.g., BRIGHTEN formed from the target BRIGHT and the segment EN) is faster if the segment comes from a source word in which it was a morpheme. It is assumed that a match between morphological components specified in the lexical representation and components designated for shifting in the present task facilitates performance in a manner in which an arbitrary (i.e., nonmorphological) segmentation could not and that the task encourages subjects to draw on morphological knowledge if it is available. Comparisons between morphological and nonmorphological segments have been reported with materials and native speakers of Serbo-Croatian which is a highly-inflected language (Feldman, 1991; Feldman, 1994). In all cases investigated previously, the output was a morphologically complex and real word (e.g., equivalents of BRIGHTEN) and shifting latencies to targets formed from morphemic letter sequences were faster than to those formed from nonmorphemic sequences. In the present study, we replicate the effect of morphological status of a final letter sequence in the segment shifting task with English materials and then extend those results to Hebrew morphology by exploiting its unique characteristics.

As described above, morphologically-complex words in both English and Hebrew are composed of two or more morphemes although they are constructed according to two different linguistic principles. In English as in Serbo-Croatian, discrete morphemic constituents are linked linearly. There is a base morpheme to which other elements are appended so as to form a sequence. This principle defines a concatenative morphology. In Hebrew, morphemic word patterns are infixed between the consonants of the root. This defines a nonconcatenative morphology. The aim of the present series of experiments was to manipulate the morphological properties of units in the segment shifting task in languages with concatenative (Experiments 1 and 2) and with nonconcatenative (Experiments 3 and 4) morphological systems in order to obtain evidence for or against the psychology reality of linguistically-defined morphological units that are not necessarily coextensive with either phonological (i.e., syllabic) or orthographic units. Experiments 1 and 2 use English materials and native speakers of English and the experimental manipulation focuses on the morphological status of a letter sequence. Experiment 1 uses real words as source and target items and it also encompasses an index of affix reliability, the reliability with which a particular letter sequence functions as a morpheme. Experiment 2 uses the same real words as source items but uses pseudowords as targets so as to eliminate interpretations based on a special relationship between source words and targets. Experiments 3 and 4 use Hebrew materials and native speakers of Hebrew and the experimental manipulation is cast in terms of root transparency of the source word; that is, the tendency for some but not all roots to combine with many word patterns. (The word pattern is always morphemic therefore the manipulation of morphological status is not entirely analogous to that introduced in English.) As noted above, in principle, morphological structure could become available either prelexically or be part of the lexical representation. Because morphemes are combined according to different principles in English and in Hebrew, similar outcomes across experiments and languages would provide evidence of morphological processing that is not easily described in terms of orthographic or phonological access units.

EXPERIMENT 1

The segment shifting manipulation is based on the observation that the morphemic composition of many words is not independent of lexical information. In the absence of a word context, the morphemic status of many sequences of letters is indeterminate. In the presence of a word context, some sequences may, but need not, be morphemic. Consider the status of final sequence EN in words such as GARDEN and HARDEN. The former is morphologically simple. The letter sequence is part of the unitary morpheme. The latter is morphologically complex in that the sequence EN forms a morpheme which is affixed to the base morpheme HARD. In short, whether or not EN is a morpheme depends on the particular word to which it is affixed. In Experiment 1 ambiguity of morphological status for sequences of letters was exploited in order to probe morphological processing. In particular, if language users have access to the morphological structure of words while performing the segment shifting task, then segmenting and affixing morphological segments may be easier than segmenting and affixing morphologically arbitrary but phonemically equivalent segments. Moreover, if the effect reflects affixation procedures, then compatibility between word class of source word and target word may be relevant. Alternatively, if the effect is dominated by segmentation (or decomposition) procedures on the source word, then properties of the affix or base morpheme (root) may be relevant insofar as they help to reveal constituent structure.

Recent work by Burani and Laudanna in Italian (Burani & Laudanna, 1992) and by Schreuder and Baayen in Dutch (1994; see also Baayen 1992) has suggested that the reliability with which a letter sequence functions morphemically is also important in determining how morphologically complex words are processed. Although a variety of measures have been proposed (e.g., some are type-based, others are token-based), most consider both the number of words in which a letter sequence functions morphemically and the total number of occurrences of that sequence in some corpus. As a first approximation of morphological reliability for the English materials, the *Capricorn Rhyming Dictionary* was used to estimate the ratio of the number of words which end in a particular morpheme to the total number of words (without proper nouns and archaic terms) that end in that letter sequence.

Methods

Subjects. Forty-six American college students from an Introductory Psychology course at the University of Delaware participated in Experiment 1 in partial fulfillment of course requirements.

Stimulus materials. Forty source pairs of English materials were constructed. Each source pair included a morphologically complex word composed of a base morpheme and a morphological suffix and a morphologically simple control word composed of only one morpheme. The control word ended with the same sequence of letters that functioned morphemically in its counterpart complex word. Morphemic and nonmorphemic endings were controlled for phonemic and syllabic overlap. Nineteen of the simple and complex source words were matched for length, although overall, the average length of the simple words was slightly shorter (6 letters) than for complex words (6.4 letters). The average surface frequency for morphologically complex source words was 19 (S.D. 42). The average frequency for morphologically simple source words was 69 (S.D. 107).[3] Both derivational and inflectional forms well as morphemically complex forms that could be either inflections or derivations were included. For example, morphologically complex source word pairs consisted of inflected words such as WINNING and their phonemically matched morphologically simple counterpart such as INNING and derived words such as HARDEN and their phonemically matched morphologically simple counterpart such as GARDEN.

Members of each source pair with morphological affixes included eleven pairs with ER including both inflectional (e.g., COLDER) and derivational (e.g., SWIMMER) functions; six pairs with EN which also included inflectional (e.g., DRIVEN) and derivational functions (e.g., SOFTEN); six pairs each with ING which is ambiguous as to morphological type[4] (e.g., WRITING), and six each with EST (e.g., NEATEST), and Y (e.g., LACY). In addition, there were two derivational pairs with AL (e.g., RENTAL), and one each with IC (viz., SCENIC), STER (viz., MOBSTER), and OR (viz., SCULPTOR). Measures of morpheme reliability for these affixes are summarized in Table 1 and are based on listings in the *Capricorn Rhyming Dictionary* with proper nouns, hyphenations and archaic terms deleted. They include the total number of morphemically complex words and the ratio of morphemic to total entries.

Target words were selected so that when the morphological or nonmorphological ending was added to it no spelling change or pronunciation change to its base morpheme was required. It was the case, however, that segmenting the affix (e.g., EN) from the morphologically complex source word (e.g., DRIVEN) sometimes required orthographic or phonological adjustments to its base (e.g., DRIVE). Source and target words were phonologically and semantically unrelated. The morphologically complex source words together with their morphologically simple controls and the target words to which the designated segment was shifted constituted a triad.

Table 1. *Number of morphemic entries, total number of entries and ratio of morphemic to total number of entries for the affixes used in Experiment 1.*

AFFIX	MORPHEME	TOTAL COUNT	RATIO
AL	42	80	.53
EN	63	104	.61
ER	177	311	.57
EST	25	61	.41
IC	101	132	.77
ING	184	238	.77
OR	39	99	.39
STER	10	17	.59
Y	176	250	.70

Procedure. Subjects were tested individually in a dimly lit room. They sat approximately 70 cm from an Atari computer and screen and stimuli subtended a visual angle of approximately 4 degrees. Following the procedure developed for Serbo-Croatian (Feldman, 1994), subjects viewed a fixation point for 200 ms then source words such as HARDEN or GARDEN appeared. After 750 ms, the EN was highlighted, the target word BRIGHT appeared below the source word and a clock started. The source and target words remained visible for 1500 ms and then a blank screen returned. The motivation for choosing a relatively long duration for the source word was to ensure that access to lexical knowledge was possible. It should be noted that the present durations work against a prelexical account of morphological processing by giving virtually unconstrained processing time as lexical access is generally available before 750 ms have transpired.

Subjects were instructed to segment and shift the designated segment from the source word to the target word and to name the new result aloud as rapidly as possible. For example, subjects were instructed to shift the EN of the source word GARDEN (or HARDEN) to the target word BRIGHT that appeared below it and to say the new form aloud. In each case, subjects said BRIGHTEN and onset to vocalization was measured and errors were recorded. The segment shifting procedure used in Experiment 1 is depicted in Figure 1.

Design. Two lists of items were created. Simple and complex members of a source word pair were counterbalanced across experimental lists so that half of the items in each list had each type of structure. Stated differently, a target that was preceded by a morphologically complex word in one list was preceded by a morphologically simple source word in the other and both lists included equal numbers of simple and complex source words. Each subject saw only one list and one member of a source pair. During the course of the experimental session, therefore, each subject saw both morphologically complex and simple source words.

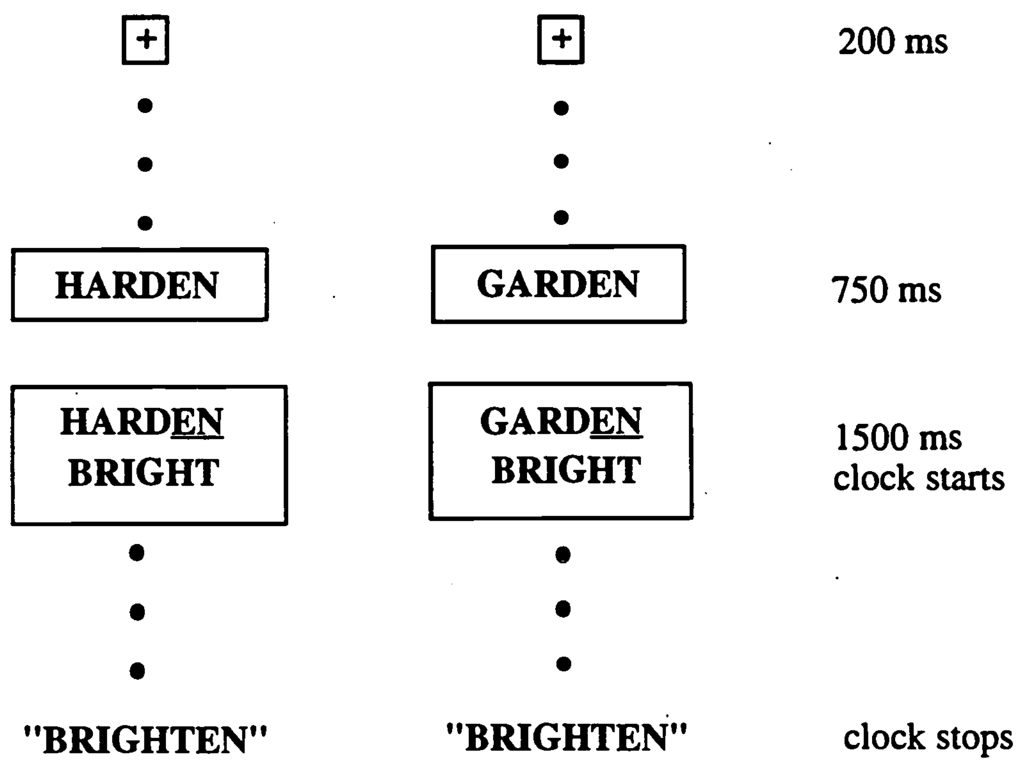


Figure 1.

Results and Discussion

Means and standard deviations of latencies for correct shifting of patterns from source words to target words were calculated for each subject in the morphologically complex and morphologically simple experimental conditions. Reaction times that were more extreme than 2.5 standard deviations from the mean of each subject in each condition were replaced by the overall mean of that condition for the analysis of variance for Experiment 1. Outliers and errors accounted for fewer than 9% of all responses. Means in the complex and simple conditions were 583 ms and 598 ms respectively. Shifting latencies were 15 ms faster to targets formed with a morphemic ending than to targets formed with a nonmorphemic ending. The statistical significance of that difference was assessed by F tests across subjects (F_1) and across stimuli (F_2) [F_1 (1,43) = 8.48, MSe = 4496, $p < .005$; F_2 (1,39) = 10.04, MSe = 4947, $p < .003$]. Due to the severe constraints on creating materials in the inflectionally impoverished language of English, no comparison between types of morphological formations was possible. Error rates were 8% for both the morphologically complex and simple conditions, therefore no analyses were performed.

The results revealed that it was easier for subjects to segment and shift letter sequences from a source word to a target word when the sequence constituted a morphological unit than when it did not. This outcome indicates that skilled readers are sensitive to a word's constituent morphological structure such that lower-frequency (morphologically complex) source words produced faster shifting latencies than higher-frequency (morphologically simple) source words. Paired item means were used to calculate the difference in segment shifting latencies in the complex and simple conditions. The difference was not significantly correlated with the surface frequency of either the morphologically complex or the morphologically simple source word nor with the frequency of the word from which the complex source word was formed.[5] Finally, the magnitude of the difference in shifting latencies was correlated with a crude measure of morphological reliability for each affix. The correlation of shifting latencies with proportion of morphemic entries was significant ($r = .35$; $p < .05$).

It was always the case that a morphologically complex source word contained another word, specifically its base. Morphologically simple words, by contrast, were variable. For example, HUNGER and ARMY encompass other unrelated words (viz., HUNG and ARM, respectively) whereas THUNDER and BABY do not. In a post hoc analysis, we asked whether this difference had an effect on shifting latencies. Morphologically simple source words were sorted as to whether or not they contained a word internally. The magnitude of the segment shifting effect was computed separately from item means for simple source words with and without internal words. The difference in shifting latencies between simple and complex forms was 16 ms [(F_1, 17) = 5.24; $p < .04$] for THUNDER type source words and 19 ms [(F_1, 18)= 4.50; $p < .05$] for HUNGER type source words.[6] Evidently the necessary presence of a word internal to morphologically complex source words and the optional presence of a word internal to morphologically simple source words cannot account for the difference in shifting latencies for morphemic and nonmorphemic letter sequences.

Output was equated over the morphologically complex and simple experimental conditions. Therefore, results in the segment shifting task results cannot simply reflect the phonological relationship between target word (e.g., BRIGHT) and what subjects produced (e.g., BRIGHTEN). Because source words were selected so that several letters close to the shifted portion were always phonologically very similar and that the shifted portion itself was identical, it is unlikely that this outcome reflects differences in phonological sequences or sequential probabilities of letters between simple and complex source words. It is possible, nevertheless, that lexical properties of the target (Dell, 1990) or a lexical editor of sorts (Dell, 1987) influences the outcome in the segment shifting task. In essence, subjects may be attempting to generate a response by applying a morphological rule appropriate for a particular target.

EXPERIMENT 2

One methodological concern with the segment shifting task is the inherent similarity between the base morpheme of the morphologically complex source word and the target word to which the segment is shifted. Typically, for example, source and target word belong to the same word class in the morphologically complex condition but not in the morphologically simple condition. The aim of the second experiment was to determine whether word class compatibility underlies the effect observed in the segment shifting task. Note that the issue of word class compatibility is relevant to the locus of facilitation in the present task. One possibility is that it is easier to detach a letter sequence from a source word if it is a morpheme than if it is not. Alternatively, it may be easier to append the letter sequence from a source word onto a target word of the same class than onto a target of a different word. In Experiments 2, 3, and 4, subjects were required to segment and shift sequences onto target pseudowords. In this case, no privileged relation between target word and source word was present.

Methods

Subjects. Forty-four American college students from an Introductory Psychology course at the University at Albany participated in Experiment 2 in partial fulfillment of course requirements.

Stimulus materials. The forty source pairs from Experiment 1 were combined with pseudoword targets. For example, morphologically complex source words consisted of inflected words such as WINNING and their phonemically matched morphologically simple pair such as INNING and derived words such as HARDEN and their phonemically matched morphologically simple pair such as GARDEN. Segments were shifted onto orthographically legal pseudowords such as REEN or EAP.

Procedure. The procedure was identical to that of Experiment 1.

Results and Discussion

Outliers were handled as in Experiment 1 and accounted for approximately 6% of the responses. Means in the morphologically complex and morphologically simple conditions were 813 ms and 833 ms respectively. Shifting latencies were 20 ms faster to targets formed with a morphemic ending than to targets formed with a nonmorphemic ending. The difference was significant across subjects and items [F_1 (1,43) = 7.97, MSe = 10516, $p<.007$; $F_2(1,39) = 5.83$, MSe = 7940, $p<.02$] in the analysis of latencies. No effects were significant in the analysis of errors but responses for complex source words tended to be more accurate than those for simple source words. Error rates were 11% and 13% for complex and simple conditions, respectively.

As in Experiment 1, paired item means in the complex and simple conditions were used to compute the difference in shifting latencies for each item pair. The correlation between the magnitude of the segment shifting effect and morphological reliability was not significant although it was in the same direction as in the previous experiment. The correlation of shifting latencies with proportion of morphemic entries was $r= .22$. Following the analyses described in Experiment 1, shifting latencies were examined separately for THUNDER and HUNGER type source words. Morphological effects were not statistically different for simple source words whose internal structure did and did not contain another lexical item.

In effect, the same pattern of results was obtained when segments were shifted to real word targets and when they were shifted to pseudoword targets. This finding eliminates accounts based on word class compatibility between source and target word in the morphologically complex condition but not in the morphologically simple condition.[7] The plausibility of a compatibility argument is further weakened by the finding that, in Serbo-Croatian, homographic morphemes (an English example is agentive and comparative ER) were shifted to appropriate targets no more quickly than to inappropriate targets (Feldman, 1991; Feldman, 1994). That is, in Serbo-Croatian, shifting the nominative plural "I" to another noun is no faster than shifting the third person plural "I" from a verb to a noun.

EXPERIMENT 3

In Experiments 1 and 2, the morphological status of a letter sequence influenced the time subjects require to produce a new form. Although it was the case that a variety of morphological patterns was examined in English, all morphological sequences and their controls took the (approximate) form of final syllables. The subsequent experiments in Hebrew were designed to probe segment shifting in linguistic environments with differing morphological characteristics and a description of that structure will be elaborated below. The relevance of investigating the processes that underlie segment shifting in a language with a nonconcatenative structure such as Hebrew was threefold. First, the experimental manipulation can be extended beyond a comparison of the morphological status of *final* syllables. Second, the experimental manipulation is extended to a more general characterization of morphology. Specifically, the transparency of the root as contrasted with the affix is manipulated. Third, because word patterns are distributed throughout the root, it is possible to contrast two types of to-be-shifted segments: (1) word patterns that are written exclusively as diacritics and disrupt only the phonological integrity of the root and (2) word patterns that are written with a combination of diacritics and letters and disrupt both the phonological and orthographic integrity of the root. This manipulation is not possible in English.

In Hebrew, roots and word patterns are abstract in that only by virtue of their joint combination (together with the application of phonological and phonetic rules) are specific words formed. Although word patterns carry morphosyntactic and some semantic information, their meaning is often obscure and typically changes for each root-pattern combination (see Berman, 1978). Moreover, there are no unequivocal rules for combining roots and word patterns to produce specific word meanings. For example, the word KATAVA ("a newspaper article") is composed of the root KTV, and the word pattern /-a--a-a/ (the dashes indicate the position of the root consonants and in the present example, the second consonant is doubled). The root KTV (כתב) refers to the concept of writing, whereas the word pattern /-a--a-/ is often (but not always) used to form nouns that are the product of the action specified by the root. Other word patterns may combine with the same root to form different words with different meanings that may be closely, but may be very remotely, related to writing. For example, the word KATAV ("press correspondent") is formed by combining the root KTV with the word pattern /-a--a/. The /-a--a/ pattern carries the morphosyntactic information that the word is a noun which signifies a profession.[8] Whereas some word patterns consist exclusively of vowels, others consist of a prefixed or suffixed consonant as well as vowels. The word KTOVET ("address") is formed by combining the KTV root with a word pattern that includes a final consonant. The /--o-et/ pattern carries the morphosyntactic information that the word is a feminine noun. When the same phonological pattern is applied to other roots, it forms different verbs or nouns, each of which is related to its respective root action. For example, KTORET ("incense") is formed from the KTR root together with the /--o-et/ pattern. In summary, it is the combination of root together with word pattern that specifies the meaning of a particular word.

In Experiment 3 the morphological processing of printed Hebrew words consisting of roots and infixed word patterns was investigated. Following the pointed source words, the subjects were presented with an unpointed pseudoroot target. They were required to detach the vowels from the previously presented source word and to read aloud the pseudoroot target using those vowels. They were instructed to proceed as rapidly as possible without making errors. The purpose of using pseudoroot rather than real roots as targets was to eliminate target-specific lexical effects from responses (i.e., naming the target word by using lexical knowledge and ignoring the source word) by forcing subjects to combine the segmented word pattern and the target.[9]

The aim of Experiment 3 was to investigate the processing of word patterns that are conveyed in unpointed Hebrew exclusively by diacritic marks. By analogy with previous studies in English and Serbo-Croatian that employed the segment shifting task, subjects were presented with pointed source words that varied along a morphological dimension. Here, the morphological source words were polygamous in that they had three-consonant roots that appeared in other Hebrew words and the consonants were pointed to convey a specific word pattern. That is, the root was fully transparent. The morphologically opaque words were monogamous in that they consisted of three consonant roots that did not occur in any other Hebrew word and were pointed with the same vowel configuration as the morphologically transparent words. These are analogous with the English control words in Experiments 1 & 2. The word patterns were composed exclusively of vowels represented by diacritics and, like the English experiment, the to-be-shifted material did not interrupt the orthographic integrity of the target words. Because

Hebrew word patterns are all considered to be morphemic, in necessary contrast to Experiments 1 & 2 in which a morphologically complex-simple comparison was possible, in Experiments 3 and 4 the experimental manipulation treated morphology in a continuous manner, linking it to the root and its transparency.

If morphological processes in Hebrew entail a segmentation or decomposition similar to that of Serbo-Croatian and English, then word patterns from transparent roots should be segmented (or affixed) faster than the same pattern from an opaque root. A morphological outcome evidenced by an effect of root transparency in Hebrew would indicate sensitivity to morphemes in a nonconcatenative language and provide additional support for the lexical representation of morphological structure.

Methods

Subjects. Thirty six students at the Hebrew University, all native speakers of Hebrew participated in the experiment for course credit or for payment.

Stimuli and Design. Forty triads of stimuli were constructed. Each set consisted of a word composed of a morphologically transparent root and a word pattern, a word composed of an opaque root with the same word pattern, and a pseudoroot target. The pseudoroots consisted of a sequence of three consonants that did not form a meaningful word in Hebrew by any possible vowel combination. The roots were all three consonants in length and the word patterns always consisted of two vowels.

Within a set, the word patterns for source words with transparent and opaque roots were identical so that shifting a word pattern to a target resulted in identical phonological structures (CVCVC) in the transparent and opaque conditions. For example, a triad could consist of (1) a source noun like DEVEK ("glue") that contains the transparent root DVK ("the action of sticking") and the word pattern /-e-e-/. (2) a source noun like GEFEN ("vine") that contains the three consonants GFN which do not form a transparent root, and the same infixed word pattern /-e-e-/. (3) a meaningless target consonant string like ZTM that does not convey meaning when combined with any pattern of vowels. All source nouns, both morphologically transparent and opaque were presented in their pointed form. The pseudoword target roots were presented in an unpointed form.

Two lists of words were constructed. A target pseudoroot that was paired with a morphologically transparent root and a particular word pattern in one list was paired with a morphologically opaque noun and the same word pattern in the other list, and vice versa. Each subject was tested on one 40 item list, and was thus presented with 20 source words with morphologically transparent roots and with 20 source words with morphologically opaque roots. Eighteen subjects were randomly assigned to each of the two lists.

Procedure and apparatus. Subjects were tested individually in a dimly lit room. They sat approximately 70 cm from a Macintosh II computer screen so that the stimuli subtended a horizontal visual angle of approximately 4 degrees. Due to the orthographic characteristics of the Hebrew materials, the presentation conditions were modified slightly from the English study. Following the appearance of a fixation point for 500 ms at the center of the screen, a pointed word appeared for 600 ms. Immediately afterwards, the target, which was an unpointed and meaningless sequence of consonants, was presented and the source word disappeared. The target was visible for 1500 ms. A blank field followed the display and lasted for 2000 ms.

Subjects were instructed to segment and shift the word pattern from the source word onto the target letter string and to name the new pseudoword aloud as rapidly as possible. For example, subjects were instructed to shift the vowels of the source word GEFEN (or DEVEK) to the target string ZTM in order to produce ZETEM. Onset to vocalization was measured from the presentation of the target consonant string, and errors were recorded. The experimental session started with 16 practice trials that were followed immediately by the 40 experimental trials presented in one block.

Results and Discussion

Outliers were determined as in Experiments 1 and 2. Outliers together with errors accounted for fewer than 5% of all responses. Shifting latencies for targets formed from words with morphologically transparent and morphologically opaque roots were 607 ms and 619 ms respectively. The shifting of vowels from words containing a transparent root was 12 ms faster than the shifting of vowels from words that contained an opaque root.

The analyses revealed an effect of morphological transparency that was marginally significant in the subject analysis F_1 (1, 35) = 2.6, MSe = 591, $p<.10$ but was significant in the stimulus analysis F_2 (1, 39) = 4.1, MSe = 659, $p<.04$. No effects were significant with the error measure and the error rates averaged less than 4%.

The results seem to indicate that, when phonologically matched, word patterns can be shifted from source words to target consonant strings more rapidly if the source word includes a transparent root than if the root is not transparent. This outcome, although weak, especially across subjects, is consistent with results reported previously in Serbo-Croatian and with results of Experiments 1 and 2 in English, both of which are concatenative languages. It suggests, moreover, that the variation in presentation conditions had no effect. If replicable, this outcome indicates that the component structure of morphologically complex words is represented in the Hebrew lexicon. Finally, by using pseudoroots as targets as in Experiment 2, interpretations based on similarity between the stationary portion of the source word and the target string in the transparent condition but not in the opaque condition become irrelevant.

The results of Experiment 2 with Hebrew materials in which word patterns were infixed between the consonants of a root (more precisely, pseudoroot) potentially extend the results of previous segment shifting studies. Here, the morphological manipulation is based on the transparency of the root of the source word because all shifted segments were, in fact, linguistically morphemic (i.e., there is no Hebrew equivalent of a "nonmorphological affix," more precisely, infix). This finding demonstrates a morphological effect in the segment shifting task even when morphemes are not phonologically coherent units because the root and word pattern are intertwined. It is not clear in this experiment whether the locus of the effect should be assigned to the word pattern or the root. The logical requirements of the task direct subjects to shift the word pattern from the source word to the target pseudoroot, which implies that they should focus on the word pattern. Nevertheless, it is possible that the classification of the word pattern also hinges on the transparency of the root with which it combines. Accordingly it is important to ascertain that, along nonmorphological dimensions, source words were equated across experimental conditions.

One possible criticism of the results of Experiment 3 is that transparency and surface frequency of the source words are confounded and that segment shifting times basically reflect recognition latencies. Accordingly, faster shifting latencies for transparent root patterns occur because those words appear more frequently in print than do the words with opaque roots. Although the English results are not consistent with this account because complex source words had a lower average frequency, one alternative interpretation of the outcome of Experiment 3 is that the greater facilitation in the segment shifting task for transparent roots relative to opaque roots merely reflects enhanced processing of more familiar words in Hebrew. In order to explore this possibility, the subjective frequency of each word was assessed. The pointed source words employed in Experiment 3 were presented to undergraduate students from the Hebrew University who rated their frequency on a 7-point scale, from very infrequent (1) to very frequent (7). The average frequencies across 50 judges were computed. Based on the frequency ratings, pairs of transparent and opaque source words whose frequencies were matched or with a higher opaque frequency were selected for further analysis. With this control for surface frequencies (20 pairs), shifting latency was again faster for words containing transparent patterns (604 ms) than for words containing opaque patterns (621 ms) [$t(19) = 2.9$, $p<.01$]. If anything, the effect increased in magnitude when the surface frequency of transparent words was equal to or lower than that of its opaque pair.

The same materials were presented to 40 new subjects from the Hebrew University in a lexical decision task. Latencies were also faster for words in the transparent condition (577 ms) than for words in the opaque condition (602 ms). However, the difference in shifting latencies for word patterns from transparent and opaque source words was not significantly correlated with lexical decision latencies for either word type (for transparent source words r=.10; for opaque source words r=-.19). Moreover, there was a suggestion that the difference between shifting latencies for the two conditions was negatively correlated with the difference in lexical decision latencies for transparent and opaque source words (r=-.22).[10] Consistent with the results obtained in English, the ease in shifting Hebrew word patterns from words with morphologically transparent roots relative to words with morphologically opaque roots apparently cannot be explained by frequency of forms in print. In addition, it cannot be explained by factors that produce differences in recognition (viz., lexical decision) latencies to those forms.

EXPERIMENT 4

The instructions to subjects in the previous experiment were deliberately vague. Subjects were simply instructed to shift the vowels from the source word onto the target pseudoroot in a manner that formed a possible word of Hebrew. Remarkably, most subjects were able to perform the task with only minimal practice and the error rates indicate that they were quite accurate. In Experiment 3, the word pattern that subjects were required to segment and infix consisted exclusively of vowels represented as diacritics. Thus, as with the segment shifting materials for analogous studies in English or Serbo-Croatian, the orthographic integrity of the target was never disrupted by shifting a pattern of vowels onto it. Stated differently, the "segment" to be shifted was always orthographically specified below the consonant letters of the source word root. Consequently, the orthographic form of the target was not disrupted by the addition of the word pattern.

In Experiment 4 , the morpheme manipulation was made graphemically more complex in an attempt to strengthen the outcome of the previous experiment. There exist some word patterns in Hebrew that disrupt the orthographic integrity of the root because the pattern consists of vowels that are represented by a combination of diacritics and letters. Other word patterns consist of a combination of both vowels and consonant letters appended either before or after the root. Two types of word patterns that disrupt the orthographic integrity of the root were examined in Experiment 4. One type included exclusively vowels represented by a combination of letters and diacritics. The other included a combination of pointed vowels and consonant letters. In neither case could the word patterns that were shifted in Experiment 4 be spatially defined below the root as was the pattern of diacritics.

Methods

Subjects. Thirty-six students at the Hebrew University, all native speakers of Hebrew, participated in the experiment for course credit or for payment. None of the subjects had participated in Experiment 3.

Stimulus materials. As in Experiment 3, 40 triads that consisted of a word with a morphologically transparent root, a word with a morphologically opaque root, and a pseudoroot target were assembled. However, in contrast to Experiment 3, all source words contained more than 3 letters, and their decomposition into (three consonant) roots and word patterns was, therefore, more complex. There were two types of word patterns. The first type consisted of vowels only, where these vowels were conveyed in print by a combination of diacritical marks under the letter and additional letters that were affixed in the medial and in the final position of the word. For example, the root PRS (פרס, "the action of slicing"), when infixed with the word pattern /--u-ah/, specifies PRUSAH ("a slice"), and is written פרוסה. The vowel /u/ is represented in this case by the letter "ו," the third letter of the word, and the final vowel /a/ is represented according to Hebrew spelling rules by the letter "ה" in word final position. (Recall that Hebrew is read from right to left.)

The second type of word pattern consisted of vowels and consonants that were affixed before and after the root. For example, the root ZKR (זכר), meaning "the action of remembering," when combined with the word pattern /ma--e-et/, forms MAZKERET ("a souvenir"), and is written מזכרת. In this case, the word pattern consists of a prefix /ma/, a vowel /e/ marked by a diacritic and a final suffix /et/. It is important to note that both the medial-final word pattern /--u-ah/ and the initial-final word pattern /ma-e-et/ can participate in the formation of many Hebrew words.

The design of the segment shifting procedure requires that the control source words never included a transparent root but always contained the same word pattern as the words with transparent roots. For example, the words PLUMAH ("feathers") and MAKHZELET ("mat") are the opaque controls for PRUSAH and MAZKERET. These words have the same word pattern as their transparent root pairs, but, because no other word is formed around the PL-M or the KHZ-L sequence, these consonant sequences are not considered to be transparent roots. As in Experiment 3, the pseudoroot targets in the present experiment were composed of a sequence of three consonants that did not represent a meaningful word in Hebrew with any vowel combination.

Subjects were instructed to read the source word and to create from it and the target a new word that has a similar word pattern. For example, the pseudoroot BNZ should have been read BNUZAH by applying the /--u-ah/ pattern, and the pseudoroot KSZ should be read "MAKSEZET" by applying the /ma--e-et/ pattern.

Procedure. The procedure and apparatus were identical to those of Experiment 3. Subjects were instructed to segment and shift the word pattern from the source word to the target string and to name the result aloud as rapidly as possible. For example, subjects saw MAZKERET and KSZ and then they produced MAKSEZET.

Results and Discussion

Two triads of stimuli were removed from the analysis because more than 50% of the subjects did not apply the word patterns to their respective targets appropriately. Means and standard deviations of latencies to name pseudoword targets with the word patterns of the source word were calculated for each subject in the transparent and opaque experimental conditions according to the same criterion as that of Experiment 3. They are summarized in Table 2. Means for the transparent and opaque conditions were 781 ms and 823 ms, respectively. Based on an examination of segment shifting latencies, the word patterns included in Experiment 4 were more difficult than those of Experiment 3 and it is likely that this reflects either the graphemic or phonological complexity of the word pattern to be shifted.[11] Nevertheless, as in the previous experiment, shifting of word patterns from words containing a transparent root was faster than shifting that same word pattern from words that contained opaque roots.

Table 2. *Reaction times (standard deviations) and errors of to shift word patterns consisting of diacritic plus letter vowels and diacritic plus letter consonants from transparent and opaque Hebrew source words in Experiment 4.*

	Source Word	
	Transparent	Opaque
Diacritic plus letter vowels		
reaction time	790	842
	(92)	(90)
errors	3.9%	4.4%
Diacritic plus letter consonants		
reaction time	773	805
	(58)	(85)
errors	8.9%	8.9%

The statistical significance of those differences was assessed by a two-way ANOVA with the factors of morphological transparency of source word (transparent, opaque) and of word pattern (vowels only, vowels and consonants). The analysis revealed a significant effect of morphological transparency F_1 (1,35) = 4.8, MSe = 7410, p<.03; F_2 (1,36) = 5.1, MSe = 6809, p<.03. The main effect of word pattern was significant in the analysis by subjects F_1 (1,35) = 4.8, MSe = 5806, p<.03 but missed significance in the analysis by stimuli F_2(1,36) = 1.96, MSe = 6809, p<.17. The two-way interaction was not significant (F<1.0). With the error measure, means for the transparent and opaque conditions were 6.4% and 6.6% respectively. Neither the main effect of word pattern nor the interaction of morphological status by word pattern approached significance with errors as the dependent measure.

In order to assess the contribution of surface frequency to the outcome, as in Experiment 3, 50 subjects from the Hebrew University were asked to rate the frequency of all items on a 7 point scale. The effect of root transparency was examined for pairs that were matched in frequency or where the frequency of the opaque source word was higher. The analysis, based on the 19 pairs that met these constraints, again revealed that word patterns were shifted more rapidly from source words with transparent roots (777 ms) than from source words with opaque roots (823 ms) [t(18) = 2.04, p<.05]. Thus, in replication of Experiment 3, the effect of root transparency in the segment shifting task could not be attributed to frequency differences among transparent and opaque source words.

As in the previous experiment, the materials from Experiment 4 were also presented to 40 students from the Hebrew University in a lexical decision task. Latencies were again faster for pointed words in the transparent condition (561 ms) than in the opaque condition (595 ms) and differences in shifting latencies for word patterns from transparent and opaque source words were not significantly correlated with mean lexical decision latencies for source word pairs with transparent roots (r=.14) or source words with opaque roots (r=.05). Moreover, the correlation between the differences for transparent and opaque source words in the lexical decision and segment shifting tasks was not significant (r=.22). Across experiments, lexical decision latencies were faster in Experiment 4 than in Experiment 3 and yet the magnitude of the segment shifting effect was enhanced.

In summary, the results of Experiment 4 are consistent with the segment shifting results of the previous experiments. Segment shifting latencies were slower overall but morphological effects were statistically significant where previous comparisons in Experiment 3 were marginal. With Hebrew materials, word patterns from transparent roots were shifted faster than from opaque roots and orthographic familiarity of the surface forms was not relevant.

The effect of root transparency on segment shifting was significant for both the diacritic plus letter vowel morphological patterns as well as for the diacritic plus letter consonant patterns and did not differ statistically between the two. The vowel only (i.e., diacritic plus letter) word pattern entailed the addition of a letter in word final position but it also required the infixation of a vowel letter between the consonants of the root. This letter disrupted the orthographic coherence of the consonant sequence that defined the root. Nevertheless, a distinction was observed between transparent and opaque roots with vowel diacritic plus vowel letter patterns. This finding suggests that orthographic integrity of the root does not play a role in the present task.

Significant effects for the letter consonant pattern are particularly informative. This pattern always included a consonant and vowel in initial position so that, in principle, articulation of the initial syllable could have commenced before the morphological status of the pattern had been determined. That is, the source words as well as the targets for /ma-e-et/ patterns all had /ma/ as the initial syllable. Nevertheless, the word patterns were shifted to a new target string more rapidly when the embedded roots were transparent than when they were opaque. This outcome suggests that in this task, the morphological word pattern is treated as one entity despite its distribution throughout the word. That is, the segment shifting measure is not sensitive to the sequential characteristics of the morphological pattern that must be moved.

GENERAL DISCUSSION

The results of the present study replicate those reported previously in Serbo-Croatian (Feldman, 1991; 1994). That is, morphological affixes were shifted faster than their nonmorphological but phonologically-matched controls. The facility with which morphological segments relative to their controls can be manipulated is consistent with the misorderings of morphological segments that occur in spontaneous speech productions (Fromkin, 1973). Specifically, morphological segments are more vulnerable to misorderings than are phonologically-matched but nonmorphemic segments (Garrett, 1980; 1982; Stemberger, 1984). In Experiments 1 and 2, letter sequences that were morphemic in status were shifted more rapidly to real word or to pseudoword targets than were their controls. In Experiments 3 and 4, the experimental manipulation was one of morphological transparency and all segments were shifted to pseudoroot targets. Word patterns from source words with transparent roots were shifted faster than patterns from opaque roots. Despite methodological differences, the same pattern of results was replicated in four experiments. The fact that similar effects were observed for pseudoword targets and for real word targets is consistent with the absence of word class compatibility results between source and target words in Serbo-Croatian described above and is relevant to the locus of the effect (Feldman, 1991; 1994). It suggests that the outcome in the segment shifting task does not depend on morphological compatibility (in the morphologically complex condition but not in the morphologically simple condition) between source word and target. By implication, the effect cannot reflect a special relation between what subjects produced and the morphologically complex source word they viewed (e.g., BRIGHTEN and HARDEN), otherwise again, it should have been absent for pseudoword targets. Because both meaningful and meaningless utterances in English showed a similar effect of morphology in the present task, the effect is unlikely to reflect lexical editing at output or other late processes. Finally, because the effect cannot be linked to the lexicality of the output or to the relation between source and target items, accounts that emphasize the semantic characteristics of the morphologically complex items or of morphemes in general are invalidated.

Although the presentation format varied slightly across languages so that comparisons must be interpreted with caution, it was the case that similar results were observed with materials constrained by the concatenative morphology of English as well as by the nonconcatenative morphology of Hebrew. In concatenative morphologies such as English and Serbo-Croatian, a morphological affix typically appears in word final (or sometimes, word initial) position[12] and the morphological status of a letter sequence typically depends on the word in which it occurs. The experimental manipulation of morphology with English materials focused on shifting letter sequences which varied with respect to morphological status. Importantly, the effect occurred whether or not a word was embedded inside the source word. Moreover, its magnitude was linked with a preliminary measure of affix reliability. In the nonconcatenative language of Hebrew, a word is composed of a sequence of (usually) three consonants that define a root and a word pattern that is phonologically and sometimes orthographically intercalated between those consonants. Word patterns disrupt the phonological integrity of the root whereas the effect on orthographic integrity depends on whether the vowels are written with optional diacritics or with letters (and on whether the pattern also includes consonants). When segmentation and shifting of the Hebrew word pattern disrupted the orthographic pattern by removing letters as well as diacritics from the source word and adding them to the target (Experiment 4), results were enhanced relative to when the orthographic pattern of the root and the target was preserved (Experiment 3). By most accounts, word patterns are always morphemic. Therefore, the exact manipulation used in the segment shifting studies with English and Serbo-Croatian materials was not possible in Hebrew. Instead, the morphological manipulation was defined in relation to the root with which the word pattern combined. Specifically, if the root appeared with other word patterns (i.e., was polygamous) then the root was morphologically transparent. If the root appeared exclusively with a particular word pattern (i.e., was monogamous) then it was morphologically opaque. That is, the morphological manipulation focused on the transparency of the root. An analog of transparent and opaque base morphemes to which English speakers can easily relate occurs in the BUDS-SUDS pair. Note that the former is transparent in that both BUD and S appear in other formations whereas SUD, the "root" of the later, appears only in this form. SUDS type constructions are typical of the control words in the Hebrew experiments. (The anticipated outcome of the English analog would be faster shifting latencies for S from BUDS than from SUDS.)

The segment shifting result appears not to be sensitive to the familiarity of a source word's orthographic pattern. In English morphologically complex source words tended to be lower in surface frequency than morphologically simple source words. In Hebrew, morphologically transparent source words tended to be higher in frequency (as assessed by both subjective frequency and lexical decision times) than opaque source words. Post hoc analyses within each language indicated that results with the segment shifting task could not be linked directly to surface frequency of the source word. Also, it is important to reiterate that the outcome in the segment shifting task was not linked to either the position of the shifted letter sequence in a word or its integrity as a unit. Similar results were observed when the shifted segment preserved the orthographic integrity of the target root (as is typical of morphological processes in English) and when it disrupted it (as can occur in Hebrew). Collectively, results across all four experiments indicate that the segment shifting task is sensitive to the morphological components of the source word and their tendency to combine with other morphemes. It is not sensitive to source word (surface or cumulative) frequency nor to lexical properties of the target.

Accounts based on sequential probabilities between letters are not plausible even in English because the composition of morphemic and nonmorphemic sequences was well matched in this study (see also Rapp, 1992). Segment shifting effects in nonconcatenative languages such as Hebrew are also not amenable to a sequential account of either letter or morpheme units because, as elaborated above, two morphological patterns are mixed within the word and thus they do not form units. If units are defined phonologically, morphological units lack coherence because the vowels of the word pattern are infixed between the consonants of the root. If units are defined orthographically, morphological units are coherent in Experiment 3 but lack coherence in Experiment 4 where the word pattern also included letters (vowels or consonants) that disrupted the sequence of consonant letters that formed the root. In essence, morphemes need not be orthographically defined units in order to produce effects in the segment shifting task. Because morphemes do not form orthographic segments in a nonconcatenative morphology, it is implausible that the observed effects arose independently of lexical knowledge.

In summary, the segment shifting task demands segmentation of source words. A match between the (experimentally-induced) components of the source word and the components of that word specified lexically appears to facilitate performance. For morphologically simple (and monogamous) source words, segmentation and affixation of the final sequence of letters is difficult because it is linguistically arbitrary. For morphologically-complex (and polygamous) source words, by contrast, segmentation and affixation is relatively easy because it is principled and may depend on units made more salient by their tendency to combine to form many different words.

Although it is evident that lexical knowledge is required in order to make the morphological structure of a word available in Hebrew, two accounts are plausible. One option is that priority is granted to the root and that the word pattern can only be determined once the root has been identified. Alternatively, it is possible that first the word pattern must be extracted from the source word. Admittedly, the two options described comprise a very fine distinction and assume serial processes and the present results cannot unequivocally specify how the component morphological structure becomes available in Hebrew. Nevertheless, several studies have suggested that the root serves as a lexical unit in Hebrew so it seems plausible that root extraction figures prominently in linguistic processing for the Hebrew speaker (see Frost & Bentin, 1992; Bentin & Frost, in press for a review). By this account, properties of the root should influence the pattern of results and this hypothesis deserves further investigation.

Reliability effects of the affix in English and transparency effects of the root in Hebrew are both compatible with an account that emphasizes lexical representations that specify morphemic components and rules for their combination. As generally defined, some affixes tend to be broader in scope so that many new words can be formed by combining them with other morphemes. Inflectional affixes tend to be more broad in scope than derivational affixes and it is interesting that in Serbo-Croatian, inflectional affixes produced more stable results in this task than did derivational affixes (Feldman, 1994). The affixes in Experiments 1 and 2 showed a significant relation between reliability of the segment as a morpheme and segment shifting latency. The observation from Hebrew that morphemic word patterns could be shifted from transparent roots more rapidly than from opaque roots extends this result by showing that transparency of the root is also relevant.

Smith (1988) proposed that morphological transparency influences how morphologically complex words are processed so that the structure of words whose base morphemes are words (e.g., TEST in PROTEST) is more accessible than that of words whose base morphemes are never words (e.g., LUSION in ILLUSION). In addition to semantic compositionality (e.g., Marslen-Wilson et al., 1994), a factor that may contribute to transparency is the frequency with which a particular morpheme enters into combinations with other morphemes or, in Hebrew, with other word patterns. This is the core of the transparency manipulation used in the third and fourth experiments and is similar in spirit to several recent proposals (e.g., Taft & Zhu, in press). In effect, segmentation as required to perform the segment shifting task may be facilitated when the morphemic components appear as morphemes in many different words.

Studies of word recognition are generally based on languages with concatenative structures where morphemes are typically coextensive with syllabic units and those units are linearly concatenated. It is frequently assumed that the morphemes of a word are processed in a serial fashion (Hudson, 1990) or that the morpheme can be represented in terms of orthographic structure (Chialant & Caramazza, in press). The disruption manipulation in Hebrew as well as analogous outcomes with concatenative and nonconcatenative materials generally show that the specialness of the morpheme is based on neither its orthographic nor its phonological integrity. Comparable results when segmenting final letter sequences from stationary portions that varied in morphological status in English and when segmenting word patterns from transparent or opaque roots in Hebrew indicate that whether morphemic structure is concatenative or nonconcatenative shifted segments are not treated in isolation from the lexicon and that lexical knowledge must include a word's morphological structure.

REFERENCES

Baayen, H. (1992). Quantitative aspects of morphological productivity. In G. E. Booij & J. van Marle (Eds.), *Yearbook of Morphology* (pp. 109-149). Dordrecht: Kluwer.

Bentin, S., & Feldman, L. B. (1990). The contribution of morphological and semantic relatedness to the repetition effect at long and short lags: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology, 42A*, 693-711.

Bentin, S., & Frost, R. (in press). Morphological factors in visual word recognition in Hebrew. In L. B. Feldman (Ed.), *Morphological aspects of language processing*. Hillsdale: Lawrence Erlbaum Associates.

Berman, R. A. (1978). *Modern Hebrew structure*. Tel Aviv: University Publishing Projects.

Burani, C., & Laudanna, A. (1992). Units of representation in the lexicon. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (pp. 361-376). Amsterdam: Elsevier.

Burani, C., Salmaso, D., & Caramazza, A. (1987). Morphological structure and lexical access. *Visible Language, XVIII*, 4, 342-352.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production*(Vol. 1, pp. 257-294). London and San Diego: Academic Press.

Caramazza, A , Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28*, 287-332.

Chialant, D. & Caramazza, A. (in press). In L. B. Feldman, (Ed.), *Morphological aspects of language processing*. Hillsdale: Lawrence Erlbaum Associates.

Comrie, B. (1981). *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.

Dell, G. S. (1987). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93, 285-321.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*, 313-349.

Feldman, L. B. (1991). The contribution of morphology to word recognition. *Psychological Research, 53*, 33-41.

Feldman, L. B. (1994). Beyond orthography and phonology: Differences between inflections and derivations *Journal of Memory & Language, 33*, 442-470.

Feldman, L. B. & Bentin, S. (1994). Morphological analysis of disrupted morphemes: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology, 47A*, 407-435.

Feldman, L. B., & Fowler, C. A. (1987). Morphemic segments shift faster than nonmorphemic controls. Paper presented to the Psychonomic Society, Seattle, WA.

Fowler, C. A., Napps, S. E., & Feldman, L. B. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13*, 241-255.

Fromkin, V. A. (1973). *Speech errors as linguistic evidence*. Mouton: The Hague.

Frost, R., & Bentin, S. (1992). Reading consonants and guessing vowels: Visual word recognition in Hebrew orthography. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 27-44). Amsterdam: North Holland.

Garrett, M. F. (1980). A perspective on research in language production. In J. C. Mehler, E. Walker, & M. F. Garrett (Eds.), *Perspectives on mental representation* (pp. 179-220). Hillsdale, NJ: Lawrence Erlbaum.

Garrett, M. F. (1982). Production of speech. Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19-76). London: Academic Press.

Henderson, L. (1985). Toward a psychology of morphemes. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 15-72). London: Erlbaum.

Henderson, L., Wallis, J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X*. (pp. 211-224). Hillsdale, NJ: Erlbaum.

Hudson, P. T. W. (1990). What's in a word? Levels of representation and word recognition. In D. A. Balota, G. B. Flores d'Arcais & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 187-201). Hillsdale: Erlbaum.

Katz, L., Rexer, K., & Lukatela, G. (1991). The processing of inflected words. *Psychological Research, 53*, 25-31.

Kelliher, S., & Henderson, L. (1990). Morphologically based frequency effects in the recognition of irregularly inflected verbs. *British Journal of Psychology, 81*, 527-539.

Laudanna, A., Badecker, W., & Caramazza, A. (1989). Priming homographic stems. *Journal of Memory & Language, 28*, 531-546.

Laudanna, A., Badecker, W., & Caramazza, A. (1992). Processing inflectional and derivational morphology. *Journal of Memory & Language, 31*, 333-348.

Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M .T. 1980). Representation of inflected nouns in the internal lexicon. *Memory & Cognition, 8*, 415-423.

Lukatela, G., Carello, C., & Turvey, M. T. (1987). Lexical representation of regular and irregular inflected nouns. *Language and Cognitive Processes, 2*, 1-17.

MacKay, D. (1978). On the retrieval and lexical structure of verbs. *Journal of Verbal Learning and Verbal Behavior, 15*, 169-182.

Manelis, L., & Tharp, D. A. (1977). The processing of affixed words. *Memory & Cognition, 5*, 690-695.

Marslen-Wilson, W., Tyler, L. K. Waksler, R., & Older, L. (1994). Morphology and meaning in the English lexicon. *Psychological Review, 101*, 3-33.

Matthews, P. H. (1974). *Morphology*. Cambridge: Cambridge University Press

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry, 12*, 373-346.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly, 24*, 263-282.

Napps, S. E. (1989). Morphemic relationships in the lexicon: Are they distinct from semantic and formal relationships? *Memory & Cognition, 17*, 729-739.

Napps, S. E., & Fowler, C. A. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research, 16*, 257-272.

Pinker, S. (1991). Rules of language. *Science, 253*, 530-535.

Rapp, B. C. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory and Language, 31*, 33-53.

Redfield, B. G. (1965). *Capricorn Rhyming Dictionary*. New York: Capricorn Books.

Rubin, G. S., Becker, C. A. & Freeman, R. H. (1979) Morphological structure and its effect on visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 18*, 757-767.

Schreuder, H., & Baayen, H. (1994). Prefix-stripping revisited. Journal *of Memory & Language, 33*, 357-375.

Schriefers, H., Zwitserlood, P., & Roelofs, A. (1991). The identification of morphologically complex words: Continuous processing or decomposition? *Journal of Memory & Language, 30*, 26-47.

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart (Ed.), *Attention and performance XII: Reading* (pp. 245-263). Hillsdale: Lawrence Erlbaum Associates.

Seidenberg, M., & McClelland, J. L. (1989). A distributed, developmental model of visual word recognition and naming. *Psychological Review, 96*, 523-568.

Smith, P. T. (1988). How to conduct experiments with morphologically complex words. *Linguistics, 26*, 699-714.

Stemberger, J. P. (1984).Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology, 1*, 281-313.

Stemberger, J. P., & MacWhinney, B., (1986). Frequency and lexical storage of regularly inflected words. *Memory and Cognition*, 14, 17-26.

Stemberger, J. P., & MacWhinney, B., (1988). Are inflected forms stored in the lexicon? In M. Hammond & M. Noonan (Eds.), *Theoretical morphology*. San Diego, CA: Academic Press.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition, 7*, 263-272.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*, 638-647.

Taft, M., & Zhu, X. (in press). Processing bound morphemes in English and Chinese. In L. B. Feldman, (Ed.) *Morphological aspects of language processing*. Hillsdale: Lawrence Erlbaum Assoc.

FOOTNOTES

[*] *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1-14 (1995).

[†] Also University at Albany

[‡] Also The Hebrew University, Jerusalem

[†††] The Hebrew University, Jerusalem

[1] The word pattern may include a syllabic prefix or suffix in addition to the vowels that are infixed into the root.

[2] There are cases, however, where the word pattern has a more explicit semantic content. In these cases it is not easy to determine which of the two morphemes dominates the semantic character of the word.

[3] Zero was entered for items not listed in the frequency count. The difference in frequency between simple and complex source words may be overestimated by using surface frequency values. In fact, the average frequency of the base form from which the morphological source word was formed was 79 (S.D. 102).

[4] For example, if WRITING is a noun then ING forms a derivational affix. If WRITING is a verb then ING is an inflectional affix.

[5] Nor was the difference significantly correlated with cumulative frequency or log cumulative frequency of words formed from the base of the complex source word.

[6] Three words were omitted from the analysis because they included letters in addition to an unrelated word and a potential affix. Specifically, CANDY includes D in addition to CAN and Y; CAPITAL includes IT in addition to CAP and AL; STARLING includes L in addition to STAR and ING.

[7] The pseudoword finding also reduces the similarity between the present outcome and that of Manelis and Tharp (1977). In that study, faster decision latencies were observed for pairs of words with the same morphological structure (i.e. simple or complex) than for pairs with differing structures (i.e., one simple and one complex). The effects of structure were interpreted in terms of similarity of meaning between word pairs. In the present study, analogous effects for word and pseudoword targets argue against interpretations based on similarity between source and target items in the morphologically complex condition but not the morphologically simple condition.

[8] The doubling of the middle consonants in the present example is a morphological marker that distinguishes between the word pattern that specifies a profession and a similar word pattern /-a-a/, common in adjectives, that signifies attributes.

[9] The similar results of Experiments 1 and 2 suggest that the introduction of pseudoword targets is appropriate because it demonstrated morphological processing of word targets as required by the segment shifting task generalized to pseudoword targets.

[10] The value for 38 df and $p<.05$ is .32.

[11] The word patterns were not spatially defined as they included letters in addition to diacritics positioned under letters. They also included more syllables than in Experiment 3.

[12] Because several affixes can be appended to the same base in English, an affix can also occur in positions that are not word final.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 255-266

An Articulatory View of Historical S-aspiration in Spanish*

Joaquín Romero†

The historical process of s-aspiration is common to many dialects of Spanish. This phonological process can be characterized as assimilation or loss of syllable-final /s/. The exact origins of aspiration are controversial and a variety of explanations have been proposed to account for the conditions that triggered the change. In this paper, a dialect comparison approach is taken in order to provide some experimental phonetic data on the phenomenon. It is suggested here that the origin of aspiration can perhaps be found in the evolution of Middle Spanish sibilants, which gave rise to a difference in the articulatory characteristics of /s/ in two dialects: Castilian, with an apical /s/, and Andalusian, with a laminal /s/. It is further suggested that it is precisely the laminal nature of Andalusian /s/ that might have given rise to aspiration, through gestural reduction, in this dialect but not in Castilian. In order to investigate this hypothesis, articulatory data were obtained from the aspirating dialect—Andalusian—and the non-aspirating dialect—Castilian—through the use of an electromagnetic tracking device (EMMA). The experiments confirmed the presence of a lamino-predorsal /s/ in Andalusian and an apical one in Castilian. Further, they revealed substantial differences in articulatory and dynamic characteristics of the two /s/s, which are taken as support for the gestural reduction hypothesis.

0. INTRODUCTION

The purpose of this paper is to investigate some phonetic factors related to the historical development of aspiration[1] of syllable-final /s/ in certain dialects of Spanish. Following the assumption that sound change takes place first at the phonetic level and that synchronic dialect comparison is a useful tool in explaining diachronic processes, it is suggested here that the origin of /s/-aspiration can be found in the evolution of sibilants in Middle Spanish and, more precisely, in the consequences that this evolution had for the articulatory characteristics of Spanish /s/ across dialects. Articulatory data will be presented from two different dialects of the language: Andalusian, with /s/-aspiration, and Castilian, without /s/-aspiration. A comparison of the data supports the notion that the origin of /s/-aspiration in Andalusian might be related to the presence of a lamino-predorsal /s/ in this dialect. It is also claimed here that the subsequent evolution of syllable-final /s/ in Andalusian can be attributed to a process of weakening or reduction of the gestural magnitude associated with this consonant.

1. Aspiration in Spanish. Description and origins

Aspiration of syllable-final /s/ is a common phenomenon in many dialects of Spanish, both in Spain and in the Americas. In Spain, although it is also observed elsewhere, it is most prominent in Andalusian, the dialect spoken in the southern part of the peninsula. Within Andalusian, two main dialect subareas are often recognized, among other things, on the basis of the effects of the

This work was supported by NIH Grants DC-00121 and HD-01994 to Haskins Laboratories. I thank Catherine Browman, Alice Faber and Robert Davis for their comments on earlier versions of this paper.

aspiration rule: the eastern variety, where aspiration of word-final /s/ has given rise to new vowel categories and a redistribution of the vowel space (Salvador, 1977), and the western variety, where the effects of the rule are context dependent and the vowel system is not affected (Alvar, 1955).

The realization of the aspiration rule in the western variety of Andalusian has occasionally been described as a simple substitution of [h] for syllable-final /s/ (Goldsmith, 1981), as in /espéra/ → [ehpéra] 'wait.' More detailed accounts (Zamora Vicente, 1969; Carbonero Cano, 1982) have remarked a wide range of variation in the phonetic outcome of the rule. In absolute final position /s/ is generally lost: /lúnes/ → [lúne] 'Monday'; in word-final position preceding a vowel, [h] is commonly heard: /lúnes aburído/ → [lúneh aβuriðo] 'boring Monday'; in syllable-final position (word-internal or final) preceding a consonant, the underlying /s/ assimilates in different ways to the consonant: complete assimilation (gemination) is often heard in contact with nasals, laterals and other fricatives, as in /ásma/ → [ámma] 'asthma,' /eslábo/ → [elláβo] 'slavic' and /esfwérso/ → [effwérso] 'effort,' respectively; phonetic pre- or post-aspiration (and occasionally gemination) in contact with voiceless stops, as in /estádo/ → [e^{h}taðo] or [et^{h}áðo] 'state'; in front of voiced stops aspiration interacts with spirantization (a process by which voiced stops surface as continuants in certain environments) and voiced fricatives are often the result: /rjésgo/ → [rjéɣo] 'risk.' As can be seen, the effects of aspiration of implosive /s/ in western Andalusian are in reality far more complicated than it has sometimes been reported. Work on other dialects that undergo the historical phenomenon of aspiration (Marrero, 1990) suggests that the effect of the rule is not the simple /s/ → [h] change in those dialects, either, which could be an indication of the general inadequacy of the common representation of the rule for all dialects.

The historical origins of aspiration are controversial both in terms of its date of appearance and the conditions that caused it. Estimated dates of appearance of the phenomenon vary considerably. The most conservative place it as a relatively new phenomenon, non-existent before the 17th or the 18th centuries (e.g., Torreblanca, 1989), while the most adventurous claim evidence of aspiration even in late Latin inscriptions (Frago, 1983). The most widely acknowledged estimates (Lapesa, 1981) place it, as expected, somewhere in between. It is probably safe to assume that aspiration of /s/ was not considered an unusual phenomenon in Andalusia, particularly in the Seville area, toward the first half of the 16th C. As for the causes of the change (not just in Andalusian, but also in related Romance languages such as French and Occitan), several hypotheses have been advanced (see Marrero, 1990 for a more detailed account). Grammont (1946) suggests the possibility that the previous vowel caused the opening of the constriction for /s/. Martinet (1955) mentions a general tendency in languages toward open syllables. Straka (1964) raises the possibility that the weakening of the tongue movement might be associated with the presence of a predorsal /s/ in syllable-final position. Finally, Méndez Dosuna (1987) mentions syllabic principles as the possible cause for /s/-aspiration.

2. An articulatory perspective on historical sound change

Explanation of historical sound change has for some time been constrained and biased by the theoretical assumptions about the nature of phonology implicit in the generative model (Chomsky & Halle, 1968; King, 1969). By concentrating on phonological patterns and distinctiveness, theorists assumed that sound change operates at the same formal level as synchronic contrast, i.e. the phonemic level. Therefore, explaining a diachronic change consists in developing the rules that can account for such a change. The disadvantages of such an approach to diachronic linguistics have been pointed out extensively in recent years (Harris-Northall, 1990; Faber, 1992). It has also been pointed out that sound change is hardly ever categorical, as a rule-based approach would imply, but rather the effects of a particular mutation in a language's sound system are most generally gradual and progressive (Labov, Yaeger, & Steiner, 1972). Change spreads through the system both syntagmatically and paradigmatically. Such a vision of sound change suggests that change takes place, at least in its earlier stages, at the phonetic level (Ohala, 1974; Faber, 1992) and, most likely, within a restricted lexical group. It also implies that several stages of a sound change could be present at any given time within a particular language community. It should, therefore, be possible to obtain important information regarding sound change from comparative phonetic studies across dialects of a given language (Labov, 1974; Terrell, 1981) or even by looking at similar processes in different languages.

Phonetic approaches to sound change have suggested the possibility that certain diachronic changes are conditioned by acoustic/auditory similarity (Ohala, 1981). In short, a sound might be substituted for another because the two share some acoustic properties that, under certain conditions, might make them hard for the listener to distinguish. Ohala (1974) explains the Norwegian s → ʃ change in these terms. The well-known change from /x/ → /f/ in English, as in Middle English 'rough' [roux] → Modern English [rʌf], has been ascribed to the acoustic similarities between labials and velars (Jonasson, 1971). An explanation in these terms for the aspiration of /s/ in Spanish is proposed in Widdison (1991, 1992, 1995), which suggests that, under certain speech conditions such as fast rate, unstressed environment or syllable coda position, the [h]-like portion of the transition between a vowel and an /s/ might have been identified as the /s/ itself.

Some important theoretical and practical shortcomings of an acoustic similarity approach to sound change have been pointed out recently by Mowrey and Pagliuca (1987, 1995) and Pagliuca (1982), who advocate, instead, a theory of sound change based on articulatory principles. They suggest the possibility that most if not all sound changes are articulatorily based and can be viewed as weakening processes, where articulatory gestures overlap and blend over time and are reduced in their magnitude. However, they do not provide much articulatory evidence for the kind of processes they propose (but see Mowrey & Pagliuca, 1995, for some relevant data). It is logical to assume, however, that diachronic and synchronic processes are governed by the same basic phonetic/phonological principles. Thus, processes that are part of a language's synchronic phonology are likely to be the same as or similar to certain diachronic processes in a different language: one could speculate that the present variation in the outcome of /s/ aspiration in Andalusian is likely to be similar to the situation in French in the 12th or 13th C. (Straka, 1964.)

Given that similarity, we can conclude that a theory of phonology that attempts to explain synchronic phenomena in terms of articulatory organization ought to be able to offer equally valuable insights in the case of diachronic processes. The theory of articulatory phonology (Browman & Goldstein, 1989, 1992) does precisely this in that it regards articulatory gestures as the basic units of phonological organization. Experimental articulatory data support the notion that at least some common phonological processes can indeed be explained in terms of overlap and blending of gestures or in terms of reduction of gestural magnitude. Browman and Goldstein (1991) present an explanation (from Pagliuca, 1982) of the above mentioned /x/ to /f/ change in English in terms that are in accordance with the general linguistic and evolutionary principles and predictions stated in Mowrey and Pagliuca (1987, 1995).

In this paper, an attempt will be made to explore, from an articulatory perspective, some possible factors behind aspiration of /s/ in Andalusian. For that purpose, we will compare experimental data from two dialects of Spanish: western Andalusian (from now on referred to simply as Andalusian), an aspirating dialect described above, and Castilian, a non-aspirating dialect. First, however, we will take a look at the historical developments that may have led to the emergence of aspiration in one dialect but not in the other.

3. The possible articulatory basis of aspiration. Comparison of aspirating and non-aspirating dialects

The two factors that characteristically differentiate Castilian from Andalusian are, on the one hand, the number and nature of voiceless fricatives and, on the other hand, the presence or absence of aspiration of implosive /s/. Castilian has a contrast between voiceless fricatives at the dento-alveolar point of articulation: /s/, usually described as apicoalveolar, and /θ/, a laminodental or interdental non-strident fricative, while Andalusian has only one category in that region, which we will assume to be underlyingly /s/, and that can be realized, in broad terms, either as [θ] or as [s], depending on the region (Zamora Vicente, 1969; Vaz de Soto, 1981). Because of the effects of aspiration described above, this Andalusian fricative only surfaces as [s]/[θ] in syllable initial position, whereas the two Castilian segments /θ/ and /s/ can occur either in syllable-initial or syllable-final positions.

While in Castilian the distinction between /θ/ and /s/ is stable, the [s]/[θ] variation in Andalusian is not. Traditionally Andalusian has been thought to realize its dentoalveolar fricative either as [θ] or as [s], but it is likely that such a view is largely influenced by the existing categories in Castilian (the prestige dialect in Andalusia) and by the characteristics of the spelling system, which, following the Castilian model, has different symbols for /θ/ and for /s/: 'z' and 'c' for /θ/, but 's' for /s/. In reality, however, the situation is more complex. Accounts of the articulatory nature of /s/ in Andalusian (Zamora Vicente, 1969; Narbona Jiménez, & Morillo-Velarde, 1987) reveal that there is much variation in its phonetic realization: basically it can be realized as a laminal, coronal or predorsal fricative with a constriction location situated anywhere between the teeth and the alveolar ridge.

The historical origin of the differences between the two dialects and of the variation in Andalusian can be found in the so-called 'sibilant turmoil' of Middle Spanish (approximately from mid 15th to mid 17th C.) (Kiddle, 1977; Lapesa, 1981). Medieval Spanish presented a rather large set of fricatives and affricates in the dentoalveolar and alveolopalatal region. These consonants arose from a large number of palatalization processes in Protoromance (Lapesa, 1981; Väänänen, 1963), many of which are common to other varieties of Western Romance. Toward the end of the 13th C. Spanish had the following set of sibilants: the dental affricates /t̪s̪/ and /d̪z̪/, the alveolar fricatives /s/ and /z/, the alveolopalatal fricatives /ʃ/ and /ʒ/, and /tʃ/, a voiceless alveolopalatal affricate – as well as [dʒ], an affricated variety of /ʒ/ that most likely occurred in initial position.

A simplification of this system took place in all dialects of the language, but the outcome was not the same in the North (Castile) as in the South (Andalusia) (Alarcos Llorach, 1961; Alonso, 1969). In all areas the occlusive element of the dental affricates /t̪s̪/ and /d̪z̪/ was lost, so that /t̪s̪/ became /s̪/ and /d̪z̪/ became /z̪/. Further, the voicing distinctions were generally lost, so that / and /z̪/ merged into /s̪/, /s/ and /z/ merged into /s/, and /ʃ/ and /ʒ/ merged into /ʃ/. In the northern dialects of the language (Castilian) the three-way contrast /s̪, s, ʃ/ was maintained by polarizing the distinctions, thus making them more salient: dental /s̪/ was fronted to /θ/, alveolar /s/ was retracted slightly to become an apicoalveolar /s/, and alveolopalatal /ʃ/ was pushed backwards in the mouth to become velar /x/. In the south (Andalusian) the distinction between /s̪/ and /s/ was lost and the two merged into a single segment that, as we saw, varies from laminodental to predorsoalveolar, while /ʃ/ presumably became /x/ as in Castilian before being weakened further to /h/.

The rearrangement of the fricative/affricate system of the language can be seen as responsible for the articulatory differences in the realization of /s/ in the two dialects (Zamora Vicente, 1969). In Castilian, because of the polarization between the dental and alveolar points of articulation, the tip of the tongue was retracted to acquire the characteristic apical – bordering on retroflex – position of modern /s/ (Joos, 1952). In Andalusian, on the other hand, the blending of the fronted laminodental position of the tongue for /s̪/ with the alveolar constriction location of /s/ resulted most likely in a constriction location in between the two, with a characteristic laminal tongue shape but a variety of constriction locations.

Other changes aside, the different solutions to the rearrangement of sibilants represented a clear reduction in the homogeneity of the language. As we mentioned above, the treatment of dentoalveolar sibilants is one of the characteristics that distinguish Castilian from Andalusian (and the American dialects). It is conceivable that the other distinguishing characteristic, aspiration of syllable-final /s/, is somehow related to the development of sibilants, since it has not been proven that aspiration existed before the sibilant merger and it seems to have appeared at a similar time or shortly thereafter. That hypothesis might have been implicit in early accounts of the articulatory characteristics of /s/ in Andalusian (as a means to investigate the possible evolution of the same process in Old French) in work by Chlumsky (1929), Malmberg (1950) and Straka (1964). Even though none of the mentioned accounts actually link the origin of /s/ aspiration to the rearrangement of sibilants, some do see a connection between the existence of a predorso-dentoalveolar realization of /s/ and aspiration, either as its direct cause (Malmberg, 1950) or as the appropriate kind of environment for weakening processes to take place (Straka, 1964.)

In any case, whether directly caused by the merging of dentoalveolar fricatives or not, it seems that the origin of aspiration might have to do with a weakening of the tongue-tip gesture related to the characteristics of the tongue during the production of /s/.[2] It is, of course, not possible to look at how weakening affects syllable-final /s/ in Andalusian, since, in that position, phonetic [s] does not occur. If, however, the weakening process has to do with the predorso-dentoalveolar nature of /s/, then it should be possible to obtain some information concerning the relationship between that type of consonant and weakening of the tongue-tip gestures by looking at the /s/ that does occur in the dialect, that is, in syllable-initial position. For that purpose, we shall try to compare the articulatory characteristics of the predorso-dentoalveolar /s/ of Andalusian and its apicoalveolar counterpart in Castilian. Experimental articulatory data for /s/ in the two languages and an examination of some of the results are reported in the next section.

4. Experimental data. Methods and results

The data reported here were collected in two separate experiments: one with a native speaker of Castilian from the Barcelona region and one with a native speaker of Andalusian from Seville. Tongue movement data for the two subjects were collected using an electromagnetic midsagittal articulometer – EMMA or magnetometer –(Perkell et al., 1992). The magnetometer consists of two main parts: a) a head mount with three magnetic transmitters that generate a magnetic field covering the entire area of articulation of the subject. b): a set of small transducer coils that can be attached to numerous places in the midsagittal plane of the subject's vocal tract. As the articulators, such as the tongue, move inside the vocal tract during speech, the transducer coils create distortions in the magnetic field which result in a set of voltages. The voltages thus created can be converted, through software manipulation, to distance. In the present experiments, coils were placed on the upper and lower lips, tongue tip (TT), tongue blade or lamina (TBL) and tongue body or dorsum (TB), as well as at the lower incisors for an estimate of jaw movement and the bridge of the nose and upper incisors for head movement correction.

The experimental designs included stops and fricatives at three different points of articulation: labial, dental and velar in a variety of syllabic positions and in different vowel contexts. The subjects were presented with a list of words embedded in the carrier sentence "Diga _____ cada vez"('say _____ each time') for the Castilian speaker and in the sentence "Diga _____ muchas veces" ('say _____ many times') for the Andalusian speaker. The Castilian speaker read each utterance five times, while the Andalusian speaker read each utterance three times. Reported here are results for the utterances in Table 1.

Figure 1 illustrates the spatial position of the tongue-tip (TT) and tongue-blade (TBL) movements associated with Castilian and Andalusian /s/ — the trajectories are averages of five tokens in Castilian and three tokens in Andalusian.

Table 1.

CASTILIAN		ANDALUSIAN	
CASABA	[kasáβa]	PASABA	[pasáβa]
TASATA	[tasáta]	TASATA	[tasáta]
PASBAPA	[pasβápa]	PASAPA	[pasápa]
PESBEPE	[pesβépe]	PESEPE	[pesépe]

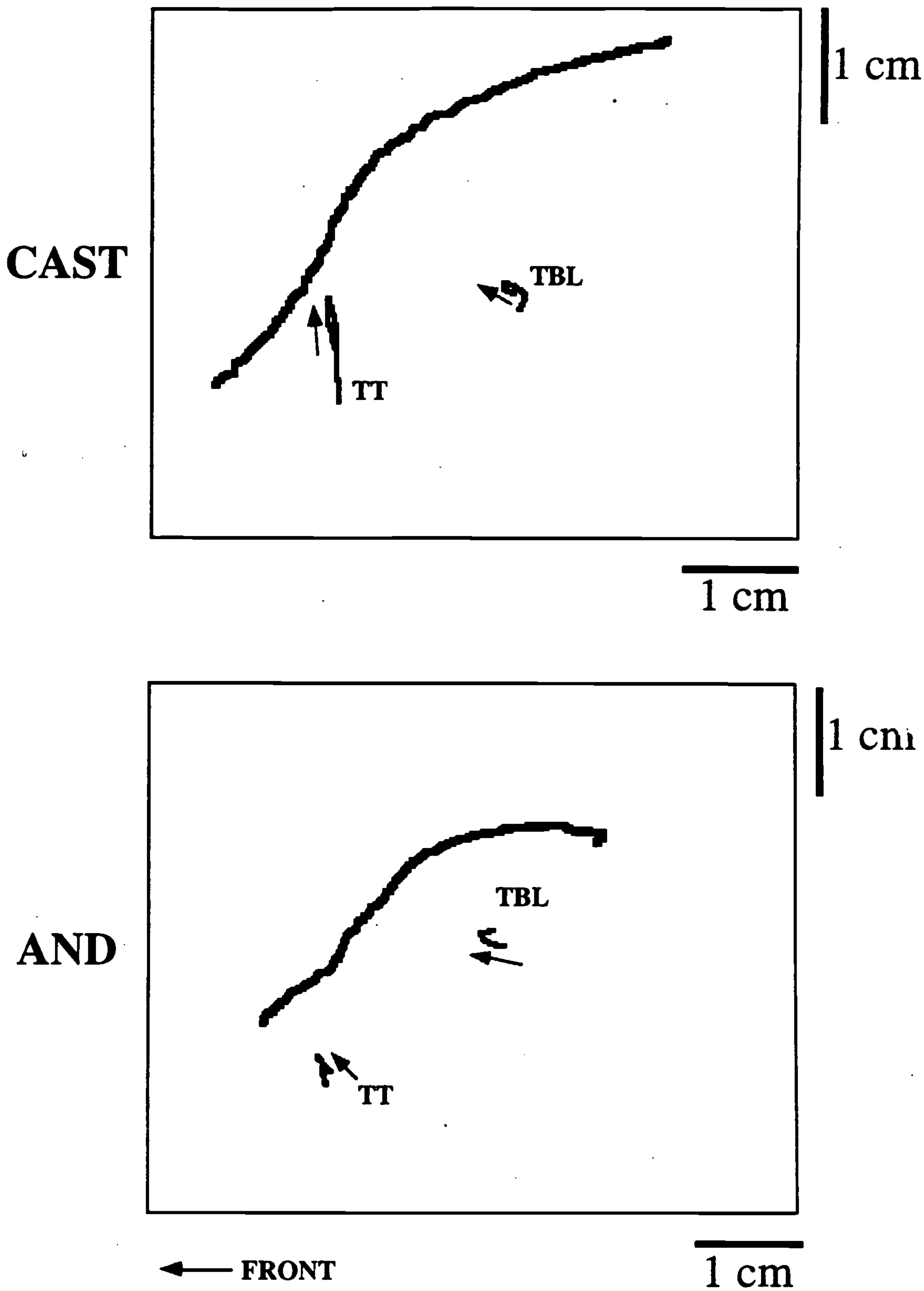


Figure 1. Spatial representation in XY space of tongue coil movement trajectories for the VCV portion of Castilian pAS(b)Apa (top) and Andalusian pASApa (bottom). A trace of the palate for each subject is displayed over the coil trajectories as visual help, although the position of the palates with respect to the trajectories is only approximate and should not be taken as an indicator of constriction degree. The arrows indicate the direction of the movement for each coil.

It can be seen very clearly from the displays in this figure that the tongue behaves very differently in the two dialects. In Castilian the tongue tip moves mostly vertically from the low position for the vowel /a/ toward the alveolar region. At the point of achievement of the target for the /s/ tongue tip and tongue blade are at nearly the same height, which, given that the tongue blade does not move a great deal from its original position for the preceding vowel, may be interpreted as a confirmation of the apical character of /s/ in Castilian. The picture in Andalusian, on the other hand, is rather different. The tongue tip moves slightly forward and upward from its position for the vowel /a/ toward the dental region, but even at its maximum position or target, the tongue tip is still at a much lower height than the tongue blade, which, as in Castilian, does not move much from its position for /a/ to the /s/ target.

We have seen, thus, that the tongue behavior in Castilian and Andalusian /s/ are clearly different and that the articulatory trajectories obtained in this experiment are in accordance with previous descriptions of the apicoalveolar versus predorso-dentoalveolar opposition. In order to show indications of gestural reduction, however, we need to look at the dynamic properties of the gestures involved in the achievement of the particular articulatory configurations. For that purpose, the data were analyzed as follows.

The movement of the TT and TBL coils were separated into X (horizontal, along the front/back dimension) and Y (vertical, along the high/low dimension) channels. These channels, over time, were displayed simultaneously with the corresponding speech signal. First derivatives were obtained for each channel in order to get an estimate of the movement velocity profiles. For each token, the VSV portion of the target word was analyzed and three different sets of X and Y measurements were obtained: displacement, or difference in cm between the tongue position for the preceding V and the target for the S; peak velocity of the articulator during the closing gesture from V to S; target interval, identified as the time interval between acoustic onset for the S and the maximum point of articulator movement or target. Figure 2 provides an example of how the three events were identified.

Articulator displacement is an indication of the extent of the movement from the previous vowel to the target associated with the consonant, which can provide an estimate of the change in tongue position associated with /s/. Peak velocity is the point of maximum velocity in the movement from the vowel to the consonantal target and it can be seen as an indication of the articulator speed during the formation of a constriction. The target interval has to do with the phasing between the two gestures involved in the production of /s/: the separation of the vocal folds in the larynx, which allows heavy airflow through the mouth, and the tongue-tip gesture in the oral cavity, which creates the narrow constriction associated with fricatives.

Everything else being equal, we should expect the tongue-tip displacement in Castilian /s/, because of its apicoalveolar nature, to be larger than in Andalusian in the vertical dimension, whereas Andalusian /s/, by virtue of being dentoalveolar (that is, a more fronted position), should show a larger horizontal displacement than in Castilian. Correspondingly, the velocities would be expected to be larger for Castilian than for Andalusian in the Y dimension, but the other way around in the X dimension. As for the target interval, the phasing between the lingual and the laryngeal gestures should perhaps be more synchronous in the Y dimension for Castilian and in the X dimension for Andalusian.

If, however, there is a reduction in the magnitude of the tongue-tip gesture in Andalusian, then we might expect small displacements in this dialect, even in the horizontal movement. Accordingly, peak velocities should also be lower in Andalusian than in Castilian, even in the X dimension. Finally, a reduction in magnitude in the tongue-tip gesture of Andalusian /s/ could be associated with a longer target interval in this dialect, which could indicate a delay in the phasing of the lingual and the laryngeal gestures.

Table 2 displays articulatory data corresponding to the TT coils during /s/ in the utterances in Table 1. For the Castilian data, each figure represents an average of five tokens, while in the Andalusian data each figure is an average of three tokens.

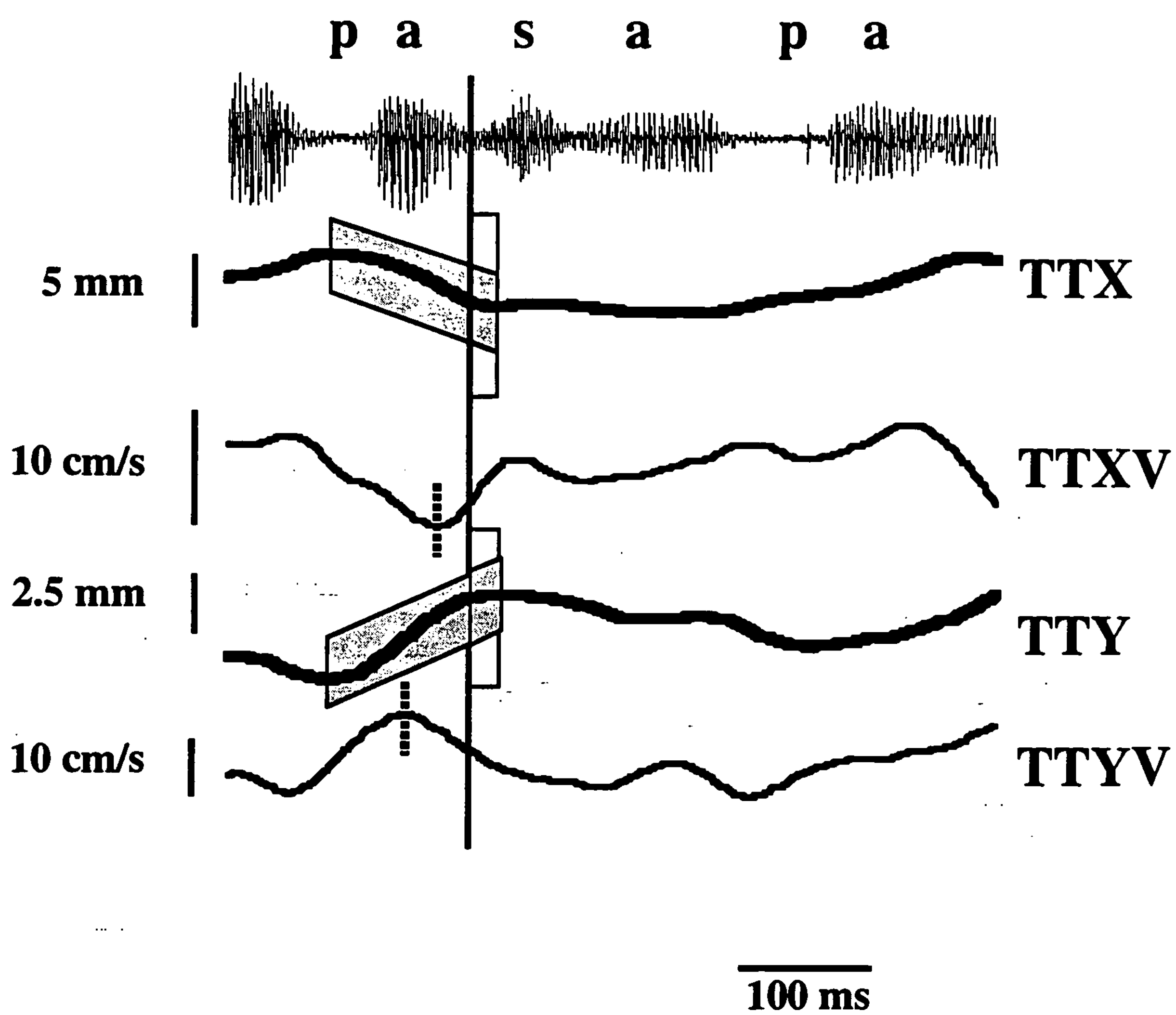


Figure 2. Display over time of acoustic and physiological signals for Andalusian 'pasapa.' The thicker traces represent the movement of the tongue-tip coil in the horizontal dimension (TTX) and in the vertical dimension (TTY). In the TTX trace, a downward movement represents a forward movement of the tongue and an upwards movement stands for a backward tongue movement (for ex. the tongue is at a relatively back position for [a] and moves forward (down in the signal) to make the /s/ constriction). The thinner traces show the corresponding velocities (TTXV and TTYV). The vertical line that crosses all signals indicates the acoustic onset for /s/. The filled boxes correspond to the displacements, the clear boxes to the target interval and the smaller striped lines in the velocity traces to the peak velocity values in the closing gestures.

Table 2.

		DISPLACEMENT (cm)		VELOCITY (cm/s)		TARGET INTERVAL (ms)	
		X	Y	X	Y	X	Y
CAS	CASABA	0.47	1.2	5.03	14.28	25.41	13.20
	TASATA	0.05	0.54	1.11	8.93	55.71	5.02
	PASBAPA	0.36	1.48	7.57	20.10	2.99	9.46
	PESBEPE	0.07	0.91	-1.4	12.32	12.41	9.05
AND	PASABA	0.51	0.5	6.56	5.36	21.1	15.46
	TASATA	0.17	0.18	3.25	3.16	31.19	17.46
	PASAPA	0.38	0.38	5.33	4.62	-3.3	16.86
	PESEPE	0.13	0.26	2.97	4.26	29.21	18.05

The data were analyzed using analyses of variance with two factors: dialect and utterance. A separate ANOVA was performed for each measure: displacement X and Y, peak velocity X and Y, and target interval X and Y. For those analyses that showed significant interactions, individual t-test were performed for each utterance pair using a Bonferroni correction (alpha = .0125). Even though the results in Table 2 cannot be considered definitive because of the small number of tokens per variable, some general trends can be observed with respect to the predictions made above. The utterances will be compared in a pairwise fashion, as illustrated in Table 1.

The results for displacement show differences in the direction of the prediction. In general, displacements in the X movement are larger for Andalusian than for Castilian, whereas the reverse is the case in the Y dimension: both X and Y show significant main effects for dialect ($F(1, 23) = 24.41, p < .01$ for X and $F(1, 23) = 776.31, p < .01$ for Y) as well as significant interactions. Looking at the values in detail, however, we see that the differences in vertical displacement between the two dialects are large. T-tests show that they are all statistically significant: $t(6)= -18.02, p <.01$ for *casaba/pasaba*, $t(6)= -15.61, p <.01$ for *pasbapa/pasapa*, $t(5)= -15.21, p <.01$ for *pesbepe/pesepe* and $t(6)= -9.62, p <.01$ for *tasata/tasata*. Differences in horizontal displacement, on the other hand, are very small: the pairs *casaba/pasaba* and *pasbapa/pasapa* are statistically non-significant, the *tasata/tasata* pair is barely significant at $t(6)= -3.53, p < .0123$, while the *pesbepe/pesepe* pair is significant at $t(5)= -5.86, p < .01$. It seems as if, in spite of its supposedly dentoalveolar character, /s/ in Andalusian is being produced with very little movement of the tongue-tip, whether in the vertical or in the horizontal dimension. In fact, the actual values of Andalusian vertical and horizontal displacements are rather similar, which seems to indicate just a slight upward and forward movement of the tongue from the position for the vowel toward the dental region.

In terms of peak velocity, the results are highly correlated with the results for displacement: here also both X and Y show significant main effects for dialect ($F(1, 23) = 23.42, p < .01$ for X and $F(1, 23) = 457.67, p < .01$ for Y), as well as significant interactions. Again we see that, in most cases, the peak velocity of the tongue tip in its movement from the vowel to the /s/ target is higher in Andalusian than in Castilian in the horizontal dimension, but higher for Castilian than for Andalusian in the vertical dimension. But also again a similar discrepancy between X and Y can be observed as in the displacements. The differences between Castilian and Andalusian in the vertical dimension are always statistically significant: $t(6)= -12.5, p <.01$ for *casaba/pasaba*, $t(6)= -14.52, p <.01$ for *pasbapa/pasapa*, $t(5)= -7.17, p<.01$ for *pesbepe/pesepe* and $t(6)= -9.22, p <.01$ for *tasata/tasata*. The results for the horizontal dimension, on the other hand, are more variable: the *casaba/pasaba* pair is non significant, the *pasbapa/pasapa* pair is barely significant at $t(6)= 3.57, p < .0119$, while the *tasata/tasata* and *pesbepe/pesepe* pairs show significant differences at $t(6)= -3.89, p < .01$ and $t(5)= -7.5, p < .01$, respectively. It seems to be the case, then, that the tongue tip in Andalusian is not only moving very little, but also rather slowly.

Finally, differences in target interval or phasing between the two dialects also seem to indicate a trend in the predicted direction. Generally, the lag between acoustic onset of /s/ and achievement of the tongue-tip target is longer in Castilian than in Andalusian for the horizontal dimension, even though neither the main effect for dialect nor the interaction were significant; the utterance main effect, however, was significant ($F(3, 23) = 9.01, p < .01$). Individual utterances differ considerably in both dialects, which suggests a lack of consistency in the achievement of a target in the horizontal dimension. In the vertical dimension, on the other hand, the figures are much more consistent across utterances. Here, Andalusian shows a longer interval than Castilian in all cases. From a statistical point of view, the ANOVA showed a significant main effect for dialect ($F(1, 23) = 6.44, p < .018$) but no significant interaction between dialect and utterance.

5. Interpretation of results in terms of reduction of gestural magnitude

As we said, the results presented above cannot be considered conclusive as to the reduced nature of the tongue-tip gesture of Andalusian /s/. A larger corpus of data is required to confirm or refute the reduction hypothesis with certainty. However, the data do show that, compared with Castilian /s/, Andalusian /s/ is produced with very little movement of the tongue-tip, in the horizontal as well as in the vertical dimensions. They also show that whatever movement there is, is slower than in the other dialect, as demonstrated by the very low values for peak velocity in Andalusian Y movement and the almost negligible values for the X dimension. On top of that, there is some indication that there is a lag in the phasing between the oral and the laryngeal gestures associated with Andalusian /s/, at least in the vertical dimension. All these factors might point to a rather wide constriction degree for /s/ in this dialect.

Nevertheless, it is possible that, because of the predorso-dentoalveolar nature of Andalusian /s/, the tip is not the most appropriate part of the tongue for measuring constriction degree for this consonant. Perhaps we should be looking at a more posterior part of the tongue in order to appreciate the actual constriction degree for /s/. Looking at Figure 1 we see that the tongue-blade coil moves just about as little as the tongue-tip coil in Andalusian. It is also not very different from the overall extent of the movement of the tongue-blade coil in Castilian. Still, the constriction might conceivably be realized at some point on the tongue between the positions of the TT and TBL coils, in which case neither TT nor TBL would be giving us the best estimate of /s/-related tongue movement. It seems unlikely, however, that a large movement of the tongue at the predorsal level would have such little effect on either the tongue tip or the tongue blade, especially since there seems to be no evidence that the predorsum functions as a separate articulator in the dialect. Unfortunately, the issue cannot be pursued directly with the available data.

It has been observed, however, that, at least in English and Castilian Spanish, /s/ is generally associated with a strong movement of the jaw (Keating, Lindblom, Lubker, & Kreiman, 1990; Romero, in preparation). Apparently, such jaw movement, both in the vertical and the horizontal dimensions, contributes to the achievement of the particular tongue configuration necessary for the production of /s/. Thus, if Andalusian /s/ did not show signs of reduction, but rather our coils did not capture the relevant part of the tongue that makes the constriction, then we might expect a similarly strong jaw movement to be associated with Andalusian /s/ as with the reported English and Castilian cases. In order to test that possibility, estimates of jaw displacement and velocity were obtained for the two dialects and compared. Subject differences aside, the results show that the displacement of the jaw, both in the vertical and the horizontal dimensions, are always significantly larger for the Castilian speaker than for the Andalusian speaker. Also, the X and Y velocities are always significantly higher for Castilian than for Andalusian. Thus, to the extent that jaw movement is an indicator of tongue activity for /s/, also here Andalusian shows less of it than Castilian.

Again, such characteristics may or may not be a sign of a reduced tongue-tip gesture in Andalusian /s/. We have to keep in mind, however, that the Andalusian /s/ we have been looking at appears always in stressed syllable-initial position. Of all possible positions, stressed syllable-initial is the one where we would expect the least amount of reduction to take place. One would expect that the observed characteristics of Andalusian syllable-initial /s/ would be affected by possible universal weakening effects in syllable-final or unstressed positions. Given the nature of Andalusian /s/ in syllable-initial position, it is not hard to imagine how a reduction in its magnitude in certain weak environments would lead to a nearly complete disappearance of the tongue-tip gesture: even less tongue-tip displacement and at a slower speed than we have observed in syllable-initial position would most likely result in almost no movement at all in syllable-final or other weak positions. Moreover, if we consider the fact that it is not uncommon to hear [h] as a substitute even for syllable-initial /s/ in certain areas of the dialect and in very fast, careless speech, we can see how weaker syllabic positions would easily lead to generalized loss of the lingual component of /s/. The differences in gestural phasing or target interval also seem to be pointing in that direction.

As outlined in section 1, many different realizations of the same aspiration process can be found in Andalusian, depending on the area or on other factors such as style, rate, etc. However, most of the variation can be reasonably explained if we start from the premise that /s/ aspiration is indeed the result of a process of reduction in the magnitude of the lingual gesture of the /s/. Whether such reduction is caused by the articulatory nature of this sound in Andalusian (as suggested in Malmberg 1950 for another aspirating dialect) cannot be proved here, but if the articulatory characteristics of Andalusian syllable-initial /s/ that we have seen are to be considered normal (that is, not reduced), then perhaps one could speculate that a predorso-dentoalveolar /s/ is not, in evolutionary terms, what could be considered a 'good' /s/ and is, consequently, likely to be weakened.

If such a speculation is at all correct, then we would expect to find other common phonological processes, whether synchronic or diachronic, that could be explained in similar terms, as suggested in Mowrey and Pagliuca (1987, 1995). In any case, we believe that it is essential, as stated in Faber (1992), to be able to provide experimental data that can, if nothing else, hint at the correctness of the speculation.

REFERENCES

Alarcos Llorach, E. (1961). *Fonología Española*. Madrid: Gredos.

Alonso, A. (1969). *De la Pronunciación Medieval a la Moderna en Español*. Madrid: Gredos.

Alvar, M. (1955). Las hablas meridionales de España y su interés para la lingüística comparada. *Revista de Filología Española, 39*, 284-313.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201-251.

Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 341-376). London: Cambridge University Press.

Browman, C., & Goldstein, L. (1991). Gestural structures: Distinctiveness, phonological processes and historical change. In I. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech production* (pp. 313-38). Hillsdale, NJ: Lawrence Erlbaum Associates.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155-180.

Carbonero Cano, P. (1982). *El habla de Sevilla*. Sevilla: Biblioteca de Temas Sevillanos, Servicio de Publicaciones del Ayuntamiento de Sevilla.

Chlumsky, J. (1929). L's andalouse et le sort de l's indo-européene en slave. *Slavia, 7*, 750-753.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Faber, A. (1992). Articulatory variability, categorical perception, and the inevitability of sound change. In Farry W. Davis & Gregory K. Iverson (Eds.), *Explanation in historical linguistics* (pp. 59-75). Philadelphia: John Benjamins Publishing Co.

Frago García, J. A. (1983). Materiales para la historia de la aspiración de la /s/ implosiva en las hablas andaluzas. *Lingüística Española Actual, 5*, 153-171.

Goldsmith, J. (1981). Subsegmentals in Spanish phonology: an autosegmental approach. In W. Cressey & D. J. Napoli (Eds.), *Linguistic symposium on romance languages: 9* (pp. 1-16). Washington: Georgetown University Press.

Grammont, M. (1946). *Traité de phonétique*. Paris: Librairie Delagrave.

Harris-Northall, R. (1990). *Weakening processes in the history of Spanish consonants*. New York: Routledge.

Jonasson, J. (1971). Perceptual similarity and articulatory reinterpretation as a source of phonological innovation. *STL-QPSR, I*, 30-42.

Joos, M. (1952). The medieval sibilants. *Language, 28*, 222-231.

Keating, P. A., Lindblom, B., Lubker, J. & Kreiman, J. (1990). Jaw position in English and Swedish VCVs. *UCLA Working Papers in Phonetics, 74*, 77-95.

Kiddle, L. B. (1977). Sibilant turmoil in middle Spanish (1450-1650). *Hispanic Review, 45*, 327-336.

King, R. D. (1969). *Historical linguistics and generative grammar*. Englewood Cliffs, NJ: Prentice Hall.

Lapesa, R. (1981). *Historia de la lengua española*. Madrid: Gredos.

Labov, W. (1974). On the use of the present to explain the past. *Proceedings of the Eleventh International Congress of Linguists* (Vol. 2, pp. 825-851). Bologna: Società Editrice Il Mulino.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.

Malmberg, B. (1950). Études sur la phonétique de l'espagnol parlé a l'Argentine. *Études Romanes de Lund X*.

Marrero, V. (1990). Estudio acústico de la aspiración en Español. *Revista de Filología Española, 70*, 345-397.

Martinet, A. (1955). *Économie des changements phonétiques*. Berne: Éditions A. Francke.

Méndez Dosuna, J. (1987). La aspiración de s como proceso condicionado por el contacto de sílabas. *Revista de la Sociedad Española de Lingüística, 17*, 15-35.

Mowrey, R., & Pagliuca, W. (1987). *The reductive character of phonetic evolution*. Unpublished manuscript. State University of New York at Buffalo.

Mowrey, R. & Pagliuca, W. (19??). The reductive character of phonetic evolution. *Rivista di Linguistica*.

Narbona Jiménez, A. & Morillo-Velarde Pérez, R. (1987). *Las hablas andaluzas*. Córdoba: Publicaciones del Monte de Piedad y Caja de Ahorros de Córdoba.

Ohala, J. (1974). Phonetic explanation in phonology. In A. Bruck, R. Fox & M. LaGaly (Eds.), *Papers from the Parasession on Natural Phonology* (pp. 251-274). Chicago: Chicago Linguistic Society.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178-203). Chicago: Chicago Linguistic Society.

Pagliuca, W. (1982). *Prolegomena to a theory of articulatory evolution*. Unpublished doctoral dissertation, State University of New York in Buffalo.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992). Electro-magnetic midsagittal articulometer systems for transducting speech articulatory movements. *Journal of the Acoustical Society of America, 92*, 3078-3096.

Romero, J. (in preparation). *Gestural organization in Spanish: an experimental analysis of spirantization and aspiration.* Unpublished doctoral dissertation, University of Connecticut.

Salvador, G. (1977). Unidades fonológicas vocálicas en andaluz oriental. *Revista Española de Lingüística, 7*, 1-23.

Straka, G. (1964). Remarques sur la "désarticulation" et l'amuïssement de l's implosive. In J. Renson (Ed.), *Mélanges de linguistique romane et de philologie médievale, offerts à M. Maurice Delbouille* (pp. 606-628). Gembloux: Éditions J. Duculot.

Terrell, T. (1981). Diachronic reconstruction by dialect comparison of variable constraints: s-aspiration and deletion in Spanish. In D. Sankoff & H. Cedergren (Eds.), *Variation Omnibus* (pp. 115-124.) Carbondale, IL: Linguistic Research Inc.

Torreblanca, M. (1989). La /s/ implosiva en español: Sobre las fechas de su aparición. *Theasurus, 44*, 281-303.

Väänänen, V. (1963). *Introduction au latin vulgaire*. Paris: Klincksieck.

Vaz de Soto, J. M. (1981). *Defensa del habla andaluza*. Sevilla: Cuadernos de Cultura Popular, Consejería de Cultura de la Junta de Andalucía / Edisur.

Widdison, K. A. (1991). *The phonetic basis for /s/ aspiration in Spanish*. Unpublished doctoral dissertation, University of California at Berkeley.

Widdison, K. A. (1992). Hacia los orígenes de la /s/ aspirada en español. *Estudios de fonética experin...ta V.* Barcelona: Promociones Publicaciones Universitarias.

Widdison, K. A, (19??). An acoustic and perceptual study of the Spanish sound change s > h. *Rivista de Linguistica.*

Zamora Vicente, A. (1969). *Dialectología Española* (2ª edición). Madrid: Gredos.

FOOTNOTES

*Appears in *Rivista di Linguistica, 7.1*, 113-120 (1995).

†Also University of Connecticut.

[1]Throughout this paper, the terms aspiration and /s/-aspiration will refer to the historical process by which syllable-final /s/ is lost and, in broad terms (see below) replaced by [h]. Phonetic aspiration, on the other hand, will be used to refer to the phonetic phenomenon characterized by the presence of a puff of air following the release of a stop consonant, as in English 'ten' [tʰɛn] or 'cat' [kʰæt].

[2]In this respect, it is interesting to note that another laminodental segment, the spirantized dental [ð] is also commonly reduced in certain environments, both in Andalusian and in Castilian, especially in final position as in "verdad" /berdád/ → [berdá⁽ð⁾] 'truth' but also in intervocalic position in, for example, certain past-participle forms, e.g. "cantado" /kantádo/ → [kantá⁽ð⁾o] 'sung'(Alarcos Llorach, 1961.)

Manfred Clynes, Pianist*

Bruno H. Repp

First encounter, battle, and retreat

I first met Manfred Clynes at the 1985 Workshop on Physical and Neuropsychological Foundations of Music in Ossiach, Austria. At the time he was head of the music research center at the New South Wales State Conservatorium of Music in Sydney, Australia. I was a researcher in speech perception with a strong interest in music perception and performance. With two colleagues, Mary Lou Serafine and Robert Crowder, I had done some experiments on memory for songs, which had been my only foray into music-related research so far but enabled me to present a paper in Ossiach and thus to attend my first music conference. Incidentally, it was also my first conference in my native country, which I happasen to share with Clynes. At the time, I had not heard of his work, but I was struck immediately by its originality and its relevance to my musical interests.

I was also very skeptical. After reading as many of Clynes's publications as I could lay my hands on, I decided to conduct a perceptual test of his theory of "composers' personal pulses" (Clynes, 1977, 1983). He very kindly assisted me by synthesizing the musical materials for that study (Repp, 1989) in his laboratory, as I did not have the necessary equipment and experience then. He also provided much advice which later turned into criticism when I deviated from the original design of the study. I subsequently acquired a digital piano and MIDI software and conducted a second perceptual study with my own materials (Repp, 1990b), as well as an analysis of recorded piano performances in search of the "Beethoven pulse" (Repp, 1990c). Both studies elicited strong critiques from Clynes (1990, 1994), followed by desperate defenses and counterattacks on my part (Repp, 1990a, 1994b). I did not emerge unscathed from this battle. Clearly, my studies had some shortcomings, for which I was duly reprimanded. They were not totally worthless, however: Having appeared in mainstream journals, they attracted attention to Clynes's important ideas, and they stimulated him to conduct a perceptual study of his own which provided impressive support for his theory (Clynes, 1995). I accept it as the last word on the issue, for the time being.

Thus I entered the world of music research on a rocky path and with bruised knees, but I did not turn back. My initial experiments had been done on the side, as it were, but I soon began to phase out my speech perception research and decided that music research was what I wanted to do henceforth. This decision was facilitated by the liberal atmosphere and generosity of Haskins Laboratories, whose support (together with a 3-month research fellowship from the Institute for Perception Research in Eindhoven) tided me over a few unstable years, until I obtained the grant from the National Institute of Mental Health that, at the time of writing, is holding my chin above water. In these years I carved out a small niche for myself in the sparsely populated research areas of objective performance analysis, perception of expressive microstructure, and experimental aesthetics of music performance. Although every study I conduct reveals how much

During preparation of this paper, I was supported by NIH grant MH-51230. I am grateful to Janet Hander-Powers and Lisa Robinson for helpful comments on an earlier draft. Most of the commercial recordings of Bach's Goldberg Variations were loaned from the Yale Music Library.

more I still have to learn, I have never regretted my decision to change fields and am enjoying my research greatly. I am deeply grateful to Manfred Clynes for providing the initial stimulus and for remaining a source of inspiration.

A second, more peaceful encounter

My purpose here is not to dwell on the past—all wounds have healed by now—nor to comment further on Clynes's scientific theories. In my more recent work I have not been directly concerned with them, although they are often on my mind. He, meanwhile, has made spectacular progress in developing a performance synthesis system that provides audible proof of the power of his ideas (and of their limits). I would like to focus here on another kind of audible proof of his fertile mind and musical imagination that has had a profound effect on me.

Over the years, Clynes has been kind enough to send me copies of several tapes of his performances as a pianist, recorded during some of his now very infrequent public appearances. Most outstanding among these recordings is his deeply moving interpretation of Bach's Goldberg Variations, a towering masterpiece of the keyboard literature.[1] Indeed, it was the expressive range and transcendent beauty of Clynes's music-making, more than any of his somewhat idiosyncratic scientific writings, that gave me confidence in his work, without necessarily removing all my skepticism. Research on music generally tends to be limited by the researcher's level of musical feeling and thought. For Clynes, however, there is no such limit. After hearing a few samples of his playing, I knew that he had the ability of penetrating to the profoundest musical truths.

Clynes and the Goldberg Variations

In the remainder of this paper, I would like to present a few glimpses of Clynes's extraordinary art in the form of graphic analyses of a few excerpts from his performance of the Goldberg Variations. This performance was recorded live in a concert given in Sydney on September 12, 1978, and was issued on cassette tape by the American Sentic Association.[2]

In order to confirm and better appreciate the uniqueness of Clynes's performance, I listened to a number of commercial recordings of the Goldberg Variations: the piano versions by Glenn Gould (CBS Masterworks MK 37779 [1981]), Charles Rosen (Sony Classical SBK 48173), Rosalyn Tureck (VAI Audio VAIA 1029), and Xiao-Mei Zhu (AVACCA 02-2); and the harpsichord versions by Maggie Cole (Virgin Classics VC 7 91444-2), Kenneth Gilbert (Harmonia Mundi HMC 901240), Wanda Landowska (EMI CDH 7610082), and Gustav Leonhardt (Teldec 8.43632).[3] Each of these interpretations has its merits, with Landowska's lively and colorful rendition deserving special mention. But, to my ears, only Gould's is on the same exalted level as Clynes's. Gould's performance is an extraordinary artistic achievement, as has been recognized by critics and music lovers worldwide. However, his approach is fundamentally different from Clynes's. Gould treats the work essentially as a giant Chaconne: He takes hardly any repeats, connects most variations without breaks, and observes strict tempo proportionality, which results in some unusual tempo choices for individual variations. His Aria is probably the slowest on record. His slow variations are serene and unbelievably focused, whereas the faster ones are lively and sharply articulated with the characteristic Gould touch. His playing emphasizes the structural aspects of the composition rather than its emotional content; it is fascinating and occasionally mesmerizing. And, of course, it is technically perfect, as Gould was not only one of the most accurate pianists but also a dedicated editor in the recording studio.

Clynes's live performance is not technically perfect, but it does not matter. He takes all the repeats and emphasizes the diversity and individual character of the variations. His interpretation is intensely emotional, especially in the slower variations, and he applies a degree of rubato and a dynamic range that one rarely encounters in Bach. However, his approach is vindicated by its convincing and powerful effect. Where others play just chains of notes, he finds (or rather introduces) expressive shapes that evoke deep resonances in the listener, very much as predicted by his theory of sentics (Clynes, 1977). Almost certainly, his theoretical ideas have influenced his performance style, and vice versa. In his hands, the variations become a colorful procession of character pieces and dances that alternately move the listener's soul and body, while the structural intricacy of the variations fades into the background. Musical motion and emotion occupy center stage, like living flesh surrounding the structural skeleton. Gould's performance, by comparison, is abstract and otherworldly.

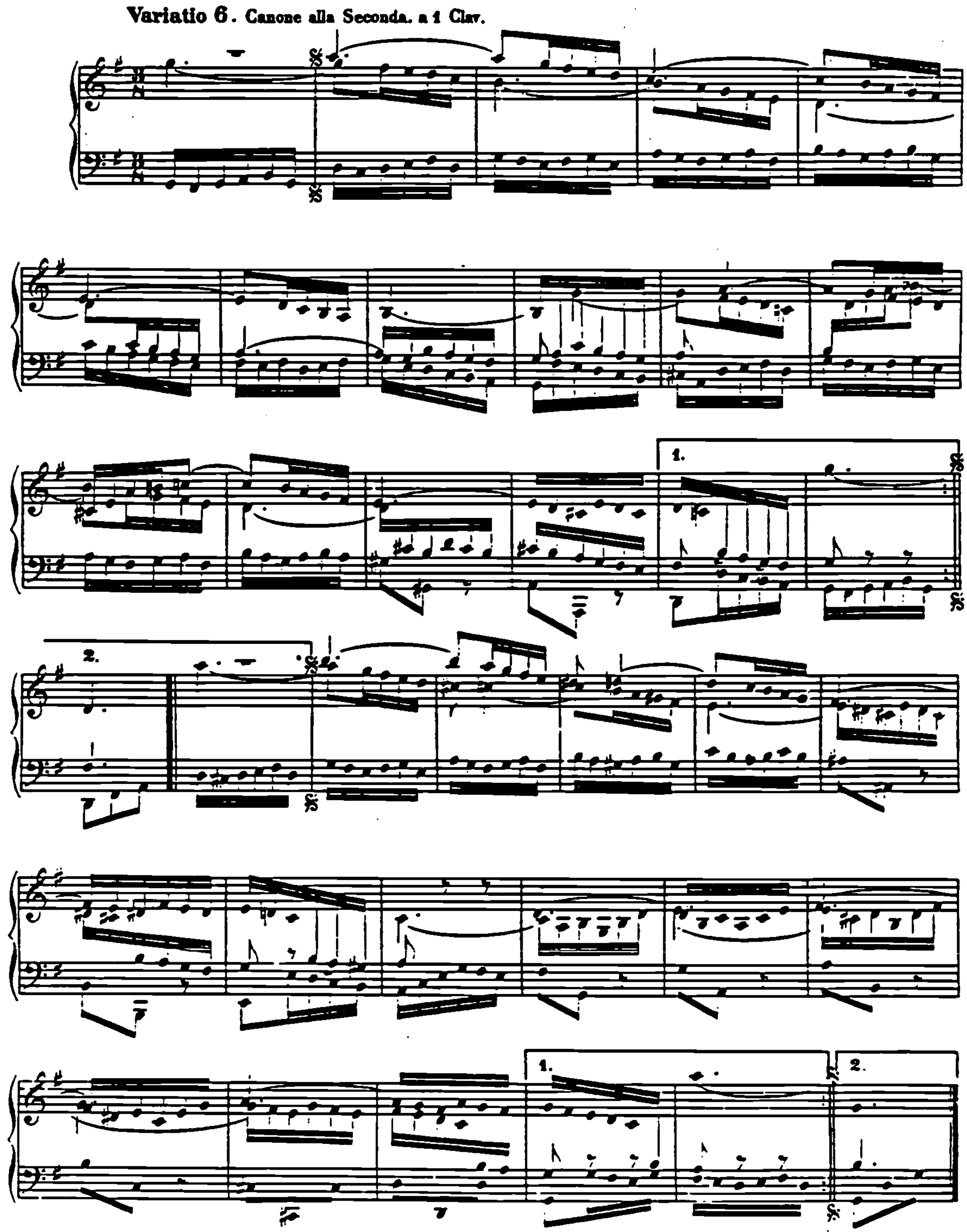
The verbal characterization of performance qualities is a difficult undertaking that always remains subjective and vague, compared to the qualitative precision of the auditory impression. It is not easy to tell by ear and to describe accurately what an artist has done to achieve a certain perceived quality. Objective performance analysis (Seashore, 1936) provides a means of capturing expressive variation quantitatively and portraying it graphically, so that the expressive shape of a performance lasting several minutes can be surveyed in a glance. While it cannot be a substitute for listening, it can reveal the agogic and dynamic devices an artist uses to achieve certain effects. It is unfortunate that dynamic variation is very difficult to measure accurately from an acoustic recording of polyphonic music. Clynes makes very effective use of the full dynamic range of the piano, and there is absolutely no attempt on his part to imitate the dynamically restricted harpsichord sound. However, the present measurements were limited to expressive timing. The relevant excerpts were digitized at 22.255 kHz, and the onsets of successive tones were measured in a waveform display with auditory feedback, using SOUNDEDIT16 software on a Macintosh Quadra 660AV computer.[4] Three particularly instructive excerpts will be considered, and in each case Clynes's very special agogics will be contrasted with that of one other pianist.

Variation 6. This variation is the *Canone alla Seconda* in G major (Example 1).[5] It is an ingenious canon in which both melodic voices are played by the right hand while the left hand provides a figurative or punctate accompaniment. The two voices are out of phase by one measure and differ in pitch by a major second. There is stepwise pitch motion on the accented beats from bar to bar, with a cadence every 8 bars. The variation is divided into two 16-bar sections, each with a repeat. The meter is 3/8, and there is continuous sixteenth-note motion throughout, provided either by the melody voices or by the accompaniment, or by both. This made it easy to examine expressive timing: The temporal distance from one sixteenth-note onset to the next was measured and plotted as a function of score position. An expressionless performance would appear as a straight line in this graph.

Figure 1 compares the expressive timing profiles of the performances by Manfred Clynes and Xiao-Mei Zhu.[6] Zhu's performance of this variation, while fluent and articulate, comes close to being devoid of expression. She takes a rather fast tempo (about 230 ms per sixteenth note, which translates into 130 eighth-note or 43 dotted quarter-note beats per minute) and omits the repeat of the second section. Her deviations from strict timing are, with a few exceptions, small and irregular. Some of this variation may be just random "motor noise" and some may be systematic but due to fingering. There is a pronounced *ritardando* at the end of each section, and a smaller one in bar 24 (a cadence). In bar 30, expressive lengthening occurs on the downbeat, the last dissonance before the final cadence. Phrase-initial lengthening (bars 1 and 17) may also be observed.[7] There is not much else to said about this plain rendition.

Contrast this with Clynes's grandly sculpted timing profile. First of all, his tempo is much slower, somewhere around 400 ms per sixteenth note (75 eighth-note or 25 dotted quarter-note beats per minute). This is the slowest tempo I have heard in this variation. Clynes needs this tempo, however, to obtain the desired expression for the principal motive, a descending sequence of five notes which recurs many times and always ends on a downbeat. While Zhu and others consider it merely as a descending scale fragment or treat the four sixteenth notes as an extended upbeat to the final long note, for Clynes it is an emotional gesture signifying (to my sensibilities) something akin to benevolence or the offering of comfort. To be effective, the gesture needs a slight *crescendo* as well as a pronounced *ritardando*, which is what we see in Clynes's timing profile. However, there is great variety in his execution of this expressive shape, and the degree of *ritardando* varies from bar to bar. Some of this variability may be due to "motor noise" or fingering patterns, as in Zhu's case, but much of it probably is intentional.

Between bars 9 and 14 a steady slowing of the tempo may be observed, especially in Clynes's first traversal, which culminates in a very large *ritardando* in bar 13. These bars have a denser texture than other bars because the two canonic voices overlap and cross each other in simultaneous sixteenth-note motion. A melodic and harmonic peak is reached in bar 12, whereupon the 5-note descending motive is stated once more in a single voice, leading to a final dissonance on the downbeat of bar 14 that then resolves into the final cadence. It is this final statement of the motive that Clynes builds up to and that forms the expressive climax of the whole variation, a particularly poignant moment not found in any other performance I have heard. Finally, it should be noted that Clynes intensifies his expressive maneuvers in the repeats: Many of the *ritardandi* are larger and start earlier in his second traversal of the music. The emotional impact on the listener is magnified correspondingly.



Example 1

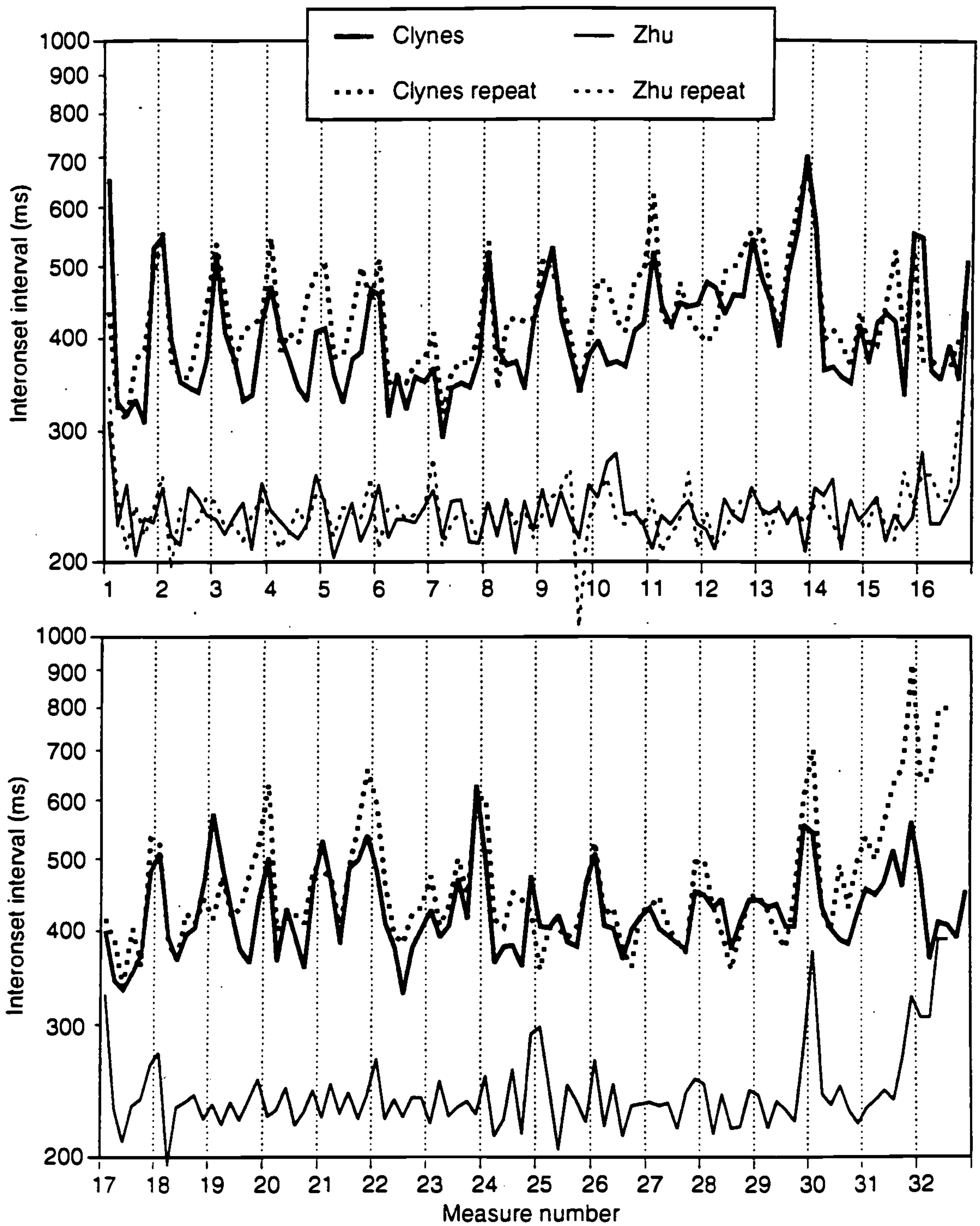


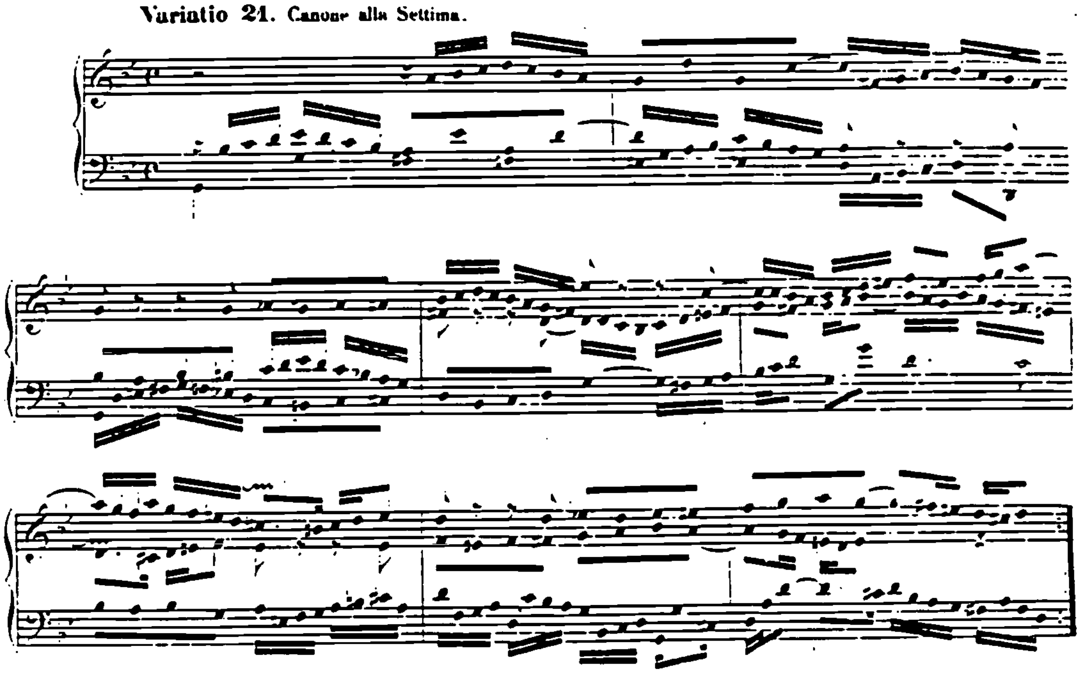
Figure 1. **Expressive timing profiles of Variation 6, as played by Manfred Clynes and Xiao-Mei Zhu (AVACCA 02-2).**

Variation 21. Another variation in which Clynes achieves extraordinary powers of expression, especially in comparison to other artists, is the *Canone alla Settima* in G minor. This is a somber and chromatic piece of great beauty, surely one of the finest variations in the set. It is in common time and is divided into two 8-bar sections, with repeats. It will suffice to consider the first 8 bars only (Example 2). As in Variation 6, the music is in continuous sixteenth-note motion. The timing data are shown in Figure 2.

As the comparison performance I have chosen the one by Charles Rosen here. His performance is rigorous and scholarly; it captures the serious tone of the variation well but shows little flexibility. This is confirmed by his timing profile. His tempo is much faster than Clynes's, approximately 280 ms per sixteenth note or about 54 quarter-note beats per minute. There is little pronounced agogic variation; even the *ritardando* at the end of the section is small. The repeat is rather similar to the first rendition. In bars 3 and 7–8, regular oscillations can be seen. In these bars, one or two voices move chromatically in eighth notes, and Rosen displaces the onsets of the intervening sixteenth notes in the third voice towards the following eighth notes, which he plays with much dynamic emphasis.

The tempo of Clynes's performance is much slower, again about 400 ms per sixteenth note or 38 beats per minute. It shows pronounced initial lengthening (bar 1) as well as an extended final two-stage *ritardando* (bar 8). Significant *ritardandi* also occur halfway through bars 2 and 4. The salient melodic motive in this variation consists of an 8-note sequence which first ascends by a fourth and then descends by a fifth in stepwise motion, ending on a strong beat. It is stated four times in bars 1–2. The first three statements are superimposed on descending chromatic steps in the bass which reach the dominant on the third beat of bar 2 and then resolve to the tonic. The fourth statement thus has a different emotional character: Whereas the first three seem to convey weariness or fatigue, the fourth seems lighter and relieved, as if a heavy weight had been deposited on the third beat of bar 2. In bar 4, something else occurs: A statement of a modified version of the 8-note motive leads to a striking unresolved dissonance, after which the modified motive (now with an extended prefix) recurs in inverted form. Clynes emphasizes the dissonance, especially in his repeat. Even more than in Variation 6, he slows down in the repeat and increases the expressive modulations during bars 1–4.

Variatio 21. Canone alla Settima.



Example 2

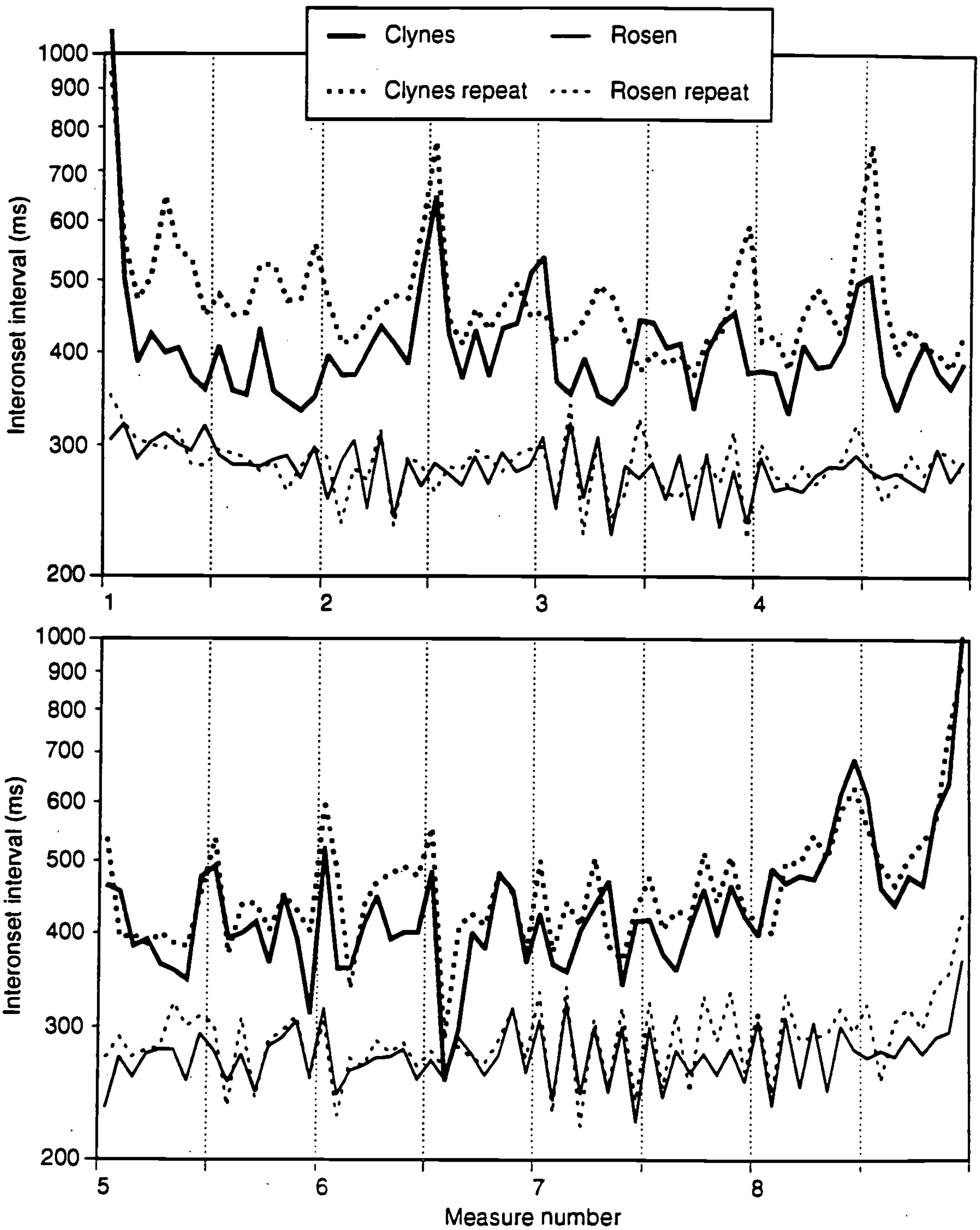


Figure 2. **Expressive timing profiles of Variation 21 (first half), as played by Manfred Clynes and Charles Rosen (Sony Classical SBK 48173).**

The difference between the two renditions is less pronounced in bars 5–8; in fact, they are very similar. A curious local phenomenon here is the very short second interonset interval in the third beat of bar 6, which occurs in the left hand, following a short trill in the right hand, perhaps to compensate for the lengthening associated with the trill. The local lengthening on the first beat of bar 6 is also caused by a trill but is not followed by a compensational maneuver. The final two-stage *ritardando* is explained by the fact that the alto voice ends on the third beat of bar 8, whereas the soprano voice, being out of phase by two beats, goes on to resolve to the dominant (the local tonic) and also changes the mode from minor to major, supported by the bass voice. All these agogic variations are of course supported by—or, rather, serve to pace—Clynes's exquisite dynamic shaping, which cannot be conveyed here graphically.

Aria. Finally, I turn to the Aria in G major as the third excerpt to be considered. Even though it opens the work, I saved its discussion for the end because of its greater rhythmic complexity.[8] It is in 3/4 meter and is divided into two 16-bar sections with repeats; again, I will examine only the first section here (Example 3). The richly ornamented melody contains a number of thirty-second notes, grace notes, and *appoggiature*, which were ignored in the present analysis, unless they were played metrically as sixteenth or eighth notes. Timing was measured at the sixteenth-note level. Intervals longer than a sixteenth note were normalized (i.e., divided by the number of sixteenth notes they contain) and graphed as plateaus extending over their nominal duration along the x-axis. For comparison with Clynes's performance, that of Glenn Gould [1981] was selected. The data are shown in Figure 3.

Gould's performance is very slow and relatively unmodulated. He does not take the repeat. The first three bars seem to be at a somewhat faster tempo than the remainder, which moves in the vicinity of 500 ms per sixteenth note, or 30 quarter-note beats per second. On closer inspection, there is a systematic pattern to the agogic variation: Temporal shapes comprising a brief *accelerando* followed by a longer *ritardando* occupy bars 1–2, 3–4, 9–10, 11–12, 13–14 (in part), 14–15, and 16. Each of these segments corresponds to half a phrase, bar 16 to an extension of the final cadence. Only bars 5–8 are relatively rigid, but with a *ritardando* at the end of bar 7. Gould's timing thus can be seen to follow the phrase structure very closely, which is consistent with the structure-oriented impression that his performance makes on the listener.

Example 3

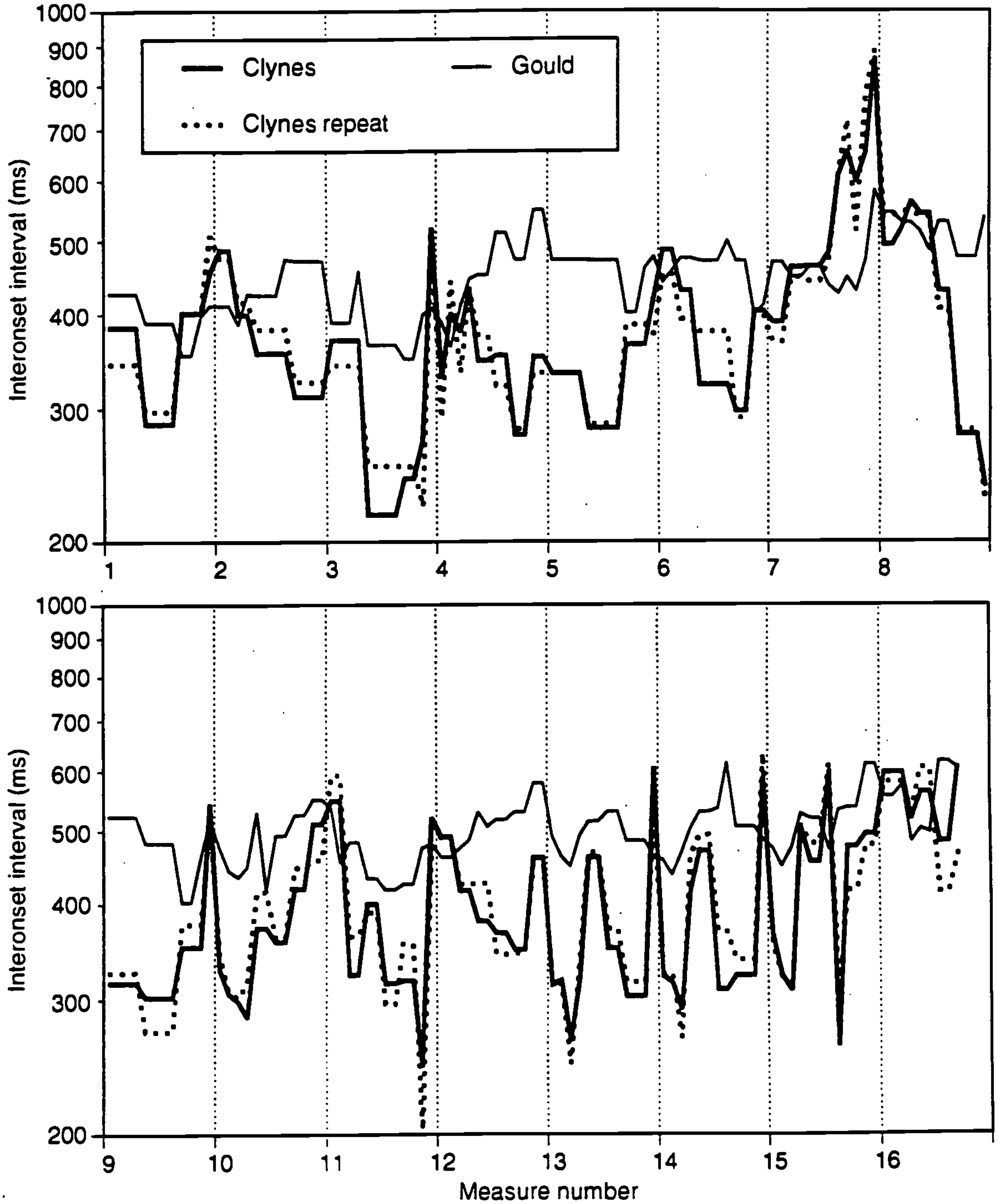


Figure 3. **Expressive timing profiles of the opening Aria (first half), as played by Manfred Clynes and Glenn Gould (CBS Masterworks MK 37779 [1981]).**

Clynes's performance, by contrast, is extremely modulated, so much so that it is difficult to assign any basic tempo to it. My best guess would be that it is somewhere around 300 ms per sixteenth note, or 50 beats per minute, on the assumption that most expressive deviations are lengthenings. Clynes takes the repeat and is amazingly consistent here; the two renditions are very nearly identical. This demonstrates that his very complex timing pattern is governed by a carefully worked out plan.[9] Rather than giving half-phrases a simple shape, Clynes tends to break them up, or rather pivots them on an expressive lengthening of the central sixteenth-note anacrusis to the following downbeat. Sharp "spikes" associated with this anacrusis can be seen at the ends of bars 3, 7, 9, 13, and 14, where it precedes another sixteenth note, while narrow peaks including the downbeat (here, an eighth note) occur at the onsets of bars 2, 6, and 12. This salient expressive device and the resulting local *ritardandi* and *accelerandi* account for a substantial part of the timing variation in Clynes's performance.

Other noteworthy features are the following: In bars 1, 5, and 9, two successive quarter notes of the same pitch occur phrase-initially; Clynes always shortens the second note relative to the first. This tendency is magnified in bar 3, where the second note, ornamented with a trill, is shortened dramatically, together with the following two notes. In bar 7, there is an enormous *ritardando* which brings the musical motion almost to a standstill. This is followed by an equally dramatic acceleration in bar 8, which leads into the next phrase. The emotional atmosphere I sense throughout is one of love, perhaps even devotion.[10] In bar 10, a pronounced *ritardando* leads to the arpeggiated chord at the beginning of bar 11, which is executed with great tenderness. In bars 13–16, each half-bar motive is set off from the next one by final lengthening. There is no *ritardando* at the end of the section, though the local tempo is slow (equal to Gould's here).

Conclusion

All three excerpts discussed illustrate the extraordinary sensitivity and flexibility of Clynes's performances, whose emotional impact is further enhanced by a masterful use of dynamics that unfortunately cannot be conveyed here. The other pianists' performances, by comparison, seem relatively rigid and unimaginative in their timing. Of course, their dynamics and timbres must also be taken into consideration, and in Gould's case the rigidity is clearly intentional, as is also evident in his carefully measured ornaments.[11] Surely, there will be some who will shake their head and say that rubato of the extent seen in Clynes's performance is inappropriate for Bach, not in style, Romantic, or inauthentic. Here Richard Taruskin, the leading critic of the notion of historical "authenticity" may be quoted, who has argued strongly that true authenticity is "founded to an unprecedented degree on personal conviction and on individual response to individual pieces" (Taruskin, 1995, p. 77). From this perspective, with which I wholeheartedly agree, Clynes is one of the most authentic musicians alive. His performances have emotional power and conviction, and a listener with an open heart and mind is carried along by them as if by a strong current. In today's world of technically flawless but often emotionally impoverished performances his art stands like a beacon, reminding us of what music can yield when it is tended with love and care.

REFERENCES

Clynes, M. (1977). *Sentics: The touch of the emotions.* New York: Doubleday. (Reprinted by Prism Press [Bridport, Dorset, UK] in 1989.)

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (pp. 76–186). Stockholm: Royal Swedish Academy of Music.

Clynes, M. (1990). Some guidelines for the synthesis and testing of pulse microstructure in relation to musical meaning. *Music Perception, 7,* 403–421.

Clynes, M. (1994). Comments on "Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists" [J. Acoust. Soc. Am. 88, 622-641 (1990)]. *Journal of the Acoustical Society of America, 96,* 1174–1178.

Clynes, M. (1995). Microstructural musical linguistics: Composers' pulses are liked most by the best musicians. *Cognition, 55,* 269–310.

Repp, B. H. (1989). Expressive microstructure in music: A preliminary perceptual assessment of four composers' "pulses". *Music Perception, 6*, 243-274.

Repp, B. H. (1990a). Composers' pulses: Science or art? *Music Perception, 7*, 423-434.

Repp, B. H. (1990b). Further perceptual evaluations of pulse microstructure in computer performances of classical piano music. *Music Perception, 8*, 1-33.

Repp, B. H. (1990c). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America, 88*, 622-641.

Repp, B. H. (1994a). Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research, 56*, 269-284.

Repp, B. H. (1994b). Response to "Comments on 'Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists'" [J. Acoust. Soc. Am. 96, 1174–1178 (1994)]. *Journal of the Acoustical Society of America, 96*, 1179–1181.

Seashore, C. E. (1936) (ed.). *Objective analysis of musical performance.* Iowa City, IA: The University Press.

Taruskin, R. (1995). *Text and act.* New York: Oxford University Press.

FOOTNOTES

*To appear in M. A. Mills (Ed.), *Festschrift for Manfred Clynes* (in preparation).

[1]Clynes was active as a concert pianist in his younger years and received high acclaim from critics and the general public, particularly for his performances of the Goldberg Variations.

[2]I do not know whether this association still exists and whether the cassette is available from it. Anyone interested should probably contact Clynes himself.

[3]Zhu's recording is a French CD that I received as a gift; it may not be commercially available in the United States. In addition to the recordings named, I am familiar with Gould's 1955 recording and with Ralph Kirkpatrick's harpsichord version, though I have not listened to them recently.

[4]In the case of asynchronous onsets of nominally simultaneous tones, the melodically most important tone was measured.

[5]All examples are reproduced from the Bach-Gesellschaft Edition (Leipzig, 1853/63), as reprinted by Dover Publications (New York, 1970).

[6]The ordinate is scaled logarithmically in order to make expressive deviations at different tempos comparable, on the assumption that they are roughly proportional to the basic tempo (see Repp, 1994a), and also to reduce the graphic excursion of large *ritardandi*. Note that a slowing of tempo corresponds to an upward excursion in the graph.

[7]The unusually short interval in bar 9 may reflect a slip of the finger or possibly a bad splice on the CD.

[8]Of course, the Aria also returns at the end of the work. However, my measurements were made on the opening Aria.

[9]I do not mean to imply that every deviation is consciously planned. Rather, the timing profile represents the replicable interaction between a musical structure and an exquisitely sensitive organism.

[10]It seems apt, though it can hardly have been Clynes's intention, that the timing profile of bars 1–8 resembles the silhouette of a medieval town with several gabled houses and two Gothic churches, one at a river (bars 3–4) and the other one on a mountain (bars 7–8). Gould's profile provides an appropriate counterpoint in the third dimension, lending depth to the illusion.

[11]Gould plays grace notes and trills metrically, whereas Clynes usually shortens grace notes and plays trills more freely.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 279-281

The Difficulty of Measuring Musical Quality (and Quantity)*
Commentary on R. W. Weisberg (1994). Genius and madness? A quasi-experimental test of the hypothesis that manic-depression increases creativity (*Psychological Science, 5,* 361-367)

Bruno H. Repp

In a recent article in *Psychological Science*, Weisberg (1994) examined the hypothesis that creative individuals suffering from manic-depressive disease are not only more productive during hypomanic phases (which they commonly are) but also produce works of *higher quality* than during normal or depressed periods. As his test case he took the composer Robert Schumann (1810-1856) who suffered from a bipolar affective disorder (Slater & Meyer, 1959; Ostwald, 1985; Jamison, 1993) and who left extensive records of his mood swings in letters and diaries. A plot of the number of Schumann's compositions by year of completion (Weisberg's Figure 1) reveals two periods of particularly intense activity: the years 1840 and 1849-1851, especially 1849. According to Slater and Meyer (1959), the years 1840, 1849, and 1851 coincide with hypomanic periods in Schumann's life. Years classified as mostly depressive periods, by contrast, show very low productivity.

This correlation between mood state and productivity seems robust and plausible, despite the crudity of the analysis. As Weisberg acknowledges, neither Schumann's periods of increased or reduced productivity nor his corresponding mood states necessarily extended over full calendar years, and there were other factors that impinged on his productivity, such as editorial work and periods of travel (see Ostwald, 1985). The measure of compositional quantity is also problematic. Weisberg's definition of "composition" (not stated, but following Slater & Meyer, 1959) is music published with a single opus number. However, works with different opus numbers may vary widely in size and complexity. In Schumann's case, they range from single short pieces such as the *Toccata*, op. 7, to extended works such as the *Three String Quartets*, op. 41. Alternative measures of compositional quantity worth considering would be the number of movements of comparable length and complexity, or even the number of distinct musical themes ("ideas").[1] Defining musical quantity is not a trivial problem.

There are also dangers in assuming that a composer's published works fully represent his creative output. Composers, like scientists, generally publish only the works that meet a certain self-imposed criterion of quality. However, that criterion may vary with the cycles of manic-depressive disease. If Schumann lowered his criterion for what he deemed publishable during his hypomanic periods, then the data would be biased against Weisberg's hypothesis that average compositional quality increases during these periods.

The most serious flaw of Weisberg's research, however, is his measure of quality. To assess the relative quality of Schumann's works in different years, Weisberg counted the number of recordings listed in two popular record guides. He placed the responsibility for this choice on an earlier author, in whose view these numbers represent "the combined judgments of musicians, recording companies, and the record-buying public" (p. 363). However, what the numbers really reflect is *popularity,* not artistic quality. There may well be a positive correlation between these two attributes, but since the quality of a work of art is extremely difficult to quantify, the strength of the correlation is unknown.[2] In most other domains (such as literature, films, or food), the items that appeal to the largest number of people are not those of the highest quality. The same principle may well apply within the select group of classical music listeners. Beethoven's Fifth Symphony is surely his most popular work, but is it his best? Schumann's *Carnaval,* op. 9, is far more popular than his *Humoreske,* op. 20, but is it of higher quality? Even those who know these works most intimately, namely musicians and musicologists, may have no simple answers to these questions. It seems naive to assume that indices of commercial value can substitute for experts' opinions.

Weisberg defined the relative quality of Schumann's compositions in a given year as the average number of recordings per work or, alternatively, as the proportion of works listed at least once in the guide. These measures showed no significant difference between hypomanic and depressive years, which led to the main conclusion that "Schumann's mood affected the quantity of his work, but not its quality" (p. 366). This may well be so, but it does not follow from the data, even if "(future) popularity" is substituted for "quality," because of additional complicating factors.

Consider the relationship between *genre* and popularity. Up to 1840, Schumann composed virtually only piano music and only a few works per year. In his first exceptionally productive year, 1840, he suddenly turned to song. According to Weisberg's calculations, that year was relatively undistinguished in terms of relative quality. However, German *Lieder* are definitely less popular with the record-buying public than are solo piano music, symphonies, or chamber music. *Lieder* are intimate and require knowledge of the language to be appreciated fully, and there is only a small number of outstanding performers in this special repertoire. Some of Schumann's songs are for two or more voices, which further reduces their popular appeal. As a result, fewer recordings are made of these works. Those song cycles that have been recorded relatively often, such as *Liederkreis*, op. 39, *Frauenliebe und Leben,* op. 42, and *Dichterliebe*, op. 48, are generally considered among the finest German songs ever written. But who is to say that the many other songs of that year are of lesser quality? One biographer of Schumann writes: "In originality, in beauty—in everything, indeed, that makes for his greatness as a composer—Schumann had reached his peak by the 'Liederjahr' [year of songs] of 1840" (Taylor, 1982, p. 191).

Changes in Schumann's compositional style over time represent another complicating factor. Taylor continues: "What followed, *pace* the occasional return to the heights in moments such as the Piano Concerto, was a slow decline, the companion of the irreversible deterioration of his physical and psychological condition" (p. 191). This opinion, which was widespread until recently, has been challenged by some musicologists who argue that the stylistic characteristics of Schumann's later works were the result of artistic choice, not deterioration (e.g., Struck, 1984). Whatever the cause may be, the later works are rarely performed. Weisberg's statement that "the proportion of high-quality compositions was essentially constant over the years of Schumann's career" (p. 365), whose literal accuracy remains uncertain, becomes false when "popularity" is appropriately substituted for "quality." The later compositions include such unwieldy works as the oratorio *Der Rose Pilgerfahrt*, op. 112, the dramatic poem *Manfred*, op. 115, songs for chorus, pieces for wind instruments, and piano duets, all of which are genres for which there is little demand in record shops and concert halls. The relative paucity of recordings of Schumann's later compositions is thus explained not only by their stylistic properties but also by their unusual form and instrumentation, neither of which has a direct bearing on quality.

Finally, the origin of Weisberg's hypothesis is unclear. He attributes it to Kraepelin (1921) who observed that "mania can produce qualitative changes in thinking, that is, changes in the kinds of ideas that the person produces" (Weisberg, 1994, p. 361). However, a qualitative change can be either an improvement or a deterioration—or neither. Kraepelin (p. 17) only talks about a "certain furtherance" of artistic activity. Jamison (1993, pp. 54-55) cites the same passage to exemplify the view that *increased artistic productivity* is linked to manic-depressive illness. Weisberg calls Jamison a "recent advocate of Kraepelin's view" (p. 361), but she too only talks about changes in quantity and quality—"increased fluency and frequency of thoughts," "speed increase," "unique ideas and associations" (Jamison, 1995, p. 66)—not about improvements. In fact, there is a high likelihood of a deterioration in quality during hypomania: "...the real capacity for work invariably suffers a considerable loss. The patient no longer has any perseverance, leaves what he begins half finished, is slovenly and careless in the execution of anything..." (Kraepelin, 1921, pp. 57-58). The idea of "increased creativity" that Weisberg derived from Kraepelin and Jamison may merely denote an increase in a creative person's productivity. To make the hypothesis of increased relative quality plausible, it would be necessary to consider in detail how the cognitive demands of musical composition might be furthered by the specific changes in cognitive functioning associated with hypomania. For example, it is quite possible that song composition benefits from increased fluency of thought whereas composition of large-scale works does not.

In summary, while Weisberg's study raises many interesting questions, it provides no answers because of serious empirical and theoretical inadequacies. The issues he means to address are very complex, and a simplistic approach will not do.

REFERENCES

Hofmann, K., & Keil, S. (1982). *Robert Schumann. Thematisches Verzeichnis sämtlicher im Druck erschienenen musikalischen Werke mit Angabe des Jahres ihres Entstehens und Erscheinens*. Hamburg: J. Schuberth & Co.

Jamison, K. R. (1993). *Touched with fire: Manic-depressive illness and the artistic temperament*. New York: Free Press.

Jamison, K. R. (1995). Manic-depressive illness and creativity. *Scientific American*, February, 62-67.

Kraepelin, E. (1921). *Manic-depressive insanity* (transl. R. M. Barclay). Edinburgh: E. & S. Livingstone.

Ostwald, P. (1985). *Schumann: The inner voices of a musical genius*. Boston, MA: Northeastern University Press.

Simonton, D. K. (1987). Musical aesthetics and creativity in Beethoven: A computer analysis of 105 compositions. *Empirical Studies of the Arts*, 5, 187–104.

Slater, E., & Meyer, A. (1959). Contributions to a pathography of the musicians: 1. Robert Schumann. *Confinia psychiatrica*, 2, 65-94.1

Struck, M. (1984). *Die umstrittenen späten Instrumentalwerke Schumanns: Untersuchungen zur Entstehung, Struktur und Rezeption*. Hamburg: Karl Dieter Wagner.

Taylor, R. (1982). *Robert Schumann: His life and work*. London: Granada.

Weisberg, R. W. (1994). Genius and madness? A quasi-experimental test of the hypothesis that manic-depression increases creativity. *Psychological Science*, 5, 361-367.

FOOTNOTES

[*] *Psychological Science*, 7, 123-124 (1996).

[1] The author has carried out such an alternative count based on a catalogue of Schumann's works (Hofmann & Keil 1982), using the sonata movement as the basic unit and assigning fractional weights to smaller works, such as short piano pieces or songs. While Schumann's productivity histogram changed in certain details, it still showed a correlation between mood state and productivity across calendar years.

[2] Simonton (1987) reports a correlation of 0.66 between "compositional popularity" (measured by frequency of mention in record-buying guides, anthologies, etc.) and "aesthetic significance" of Beethoven's works, as rated by one musicologist. This illustrates the correct approach, but the reliability and validity of the ratings are unknown, and the correlation cannot be generalized to a different composer.

Haskins Laboratories Status Report on Speech Research
1994-1995, SR-119/120, 283-291

A Review of David Epstein's *Shaping Time: Music, the Brain, and Performance* (New York: Schirmer Books, 1995, 598 pp., $47.00. ISBN 0-02-873320-7)*

Bruno H. Repp

This massive volume is the result of 15 years of theoretical thought, performance experience, and scientific study following the publication of the author's earlier book, *Beyond Orpheus* (Epstein, 1979), where some of the groundwork was laid. Epstein is a conductor and composer as well as professor of music at MIT, and his experience as a musician informs his approach to theoretical issues (and vice versa). As he says in the foreword, he has "long seen performance as the ultimate proving ground of musical verities" (p. xvi). To this fruitful interpenetration of theory and practice Epstein has now added a strong interest in the psychological and neurophysiological underpinnings of music performance, developed during several extended visits to research institutes in Germany. This interest has led him not only to peruse an impressive amount of relevant scientific literature but even to engage in some empirical research of his own. This rare confluence of musicianship, theoretical acumen, and hard science gives the book its unique flavor.

Shaping Time is divided into five parts containing 14 chapters ranging in length from 2 to 165 pages, followed by 83 (!) pages of notes, a bibliography, and an index. The introductory first part, entitled Time, *Motion, and Proportion,* defines some basic concepts and reveals the motivation behind Epstein's investigations. Epstein sees time as "*the* critical element in performance" (p. 3) and believes that shortcomings of performances are often temporal in nature. He says that "judgments about...the way a piece must move...demand extensive experience with the music" (p. 5) and refers to his own performance experience as a source of relevant insights. His theoretical and empirical approach to musical time thus can be seen to arise from a desire not only to communicate his experience to others but also to rationalize and systematize his musical intuitions. There is a certain danger of circularity in this enterprise: If Epstein's theoretical ideas have invaded his intuitions over the years, as they probably have, the latter cannot provide a neutral testing ground for the former.

For Epstein, the central concept in musical time is *motion*, which is the temporal unfolding of a musical structure, accompanied by an "internal sense of motion...a fact widely experienced and confirmed" (p. 487). Epstein discusses quantitative and qualitative (experiential, affective) aspects of musical motion and explains how it is controlled by hierarchic periodicities. He contrasts meter and rhythm, which give structural support, with tempo and its modulations, which pace the motion. Epstein's goal is to understand the *mechanisms* that govern musical motion, in contrast to many earlier authors who have dealt with this concept in a less rigorous or incomplete fashion. In the introductory chapter he provides some glimpses of things to come—a music example, a brief reference to *rubato*—and concludes with a disclaimer that seems appropriate and yet frustrating for the empirically oriented music scientist: "Replication and repeatability are not even desirable, much less applicable. Nothing would bore us faster than a musical system consistently and predictably used in exactly the same way, down to the smallest detail" (p. 17). In other words, it will not be easy to test theories in a rigorous and objective fashion.

Part Two, *Rhythm, Meter, and Motion,* consists of a theoretical chapter that fleshes out some of the concepts mentioned in Part One, followed by a discussion of music examples. The theoretical chapter is entitled *Thoughts for an Ongoing Dialogue*. The partners in this dialogue seem to be mainly other music theorists. To this (psychologist) reader, at least, this part of the book seems relatively conventional and uncontroversial. Epstein distinguishes between chronometric time (meter) and integral or experienced time (rhythm), both of which he sees as parallel, periodic, segmented, hierarchically organized processes whose varying phase relationship creates conflicts in need of resolution, a cyclic process that propels music forward in time. The units of meter are beat, measure, and hypermeasure; those of rhythm are pulse, motive, and phrase. Accent, or structural prominence, is distinguished from stress, or surface prominence. Importantly, motion is described as "the ultimate goal of musical structure, possibly the ultimate goal of music" (p. 26). This is surely one of the topics of the dialogue referred to in the title of the chapter, addressed to those theorists who tend to survey musical structure by eye rather than by ear. Epstein's focus on the temporality of music provides a much-needed counterweight to the abstract analytic discourse that has dominated music theory for decades. By investing musical structures with communicative function through motion, Epstein reinstates performer and listener as essential participants in the musical transaction—one as the controller of motion, the other as its resonator and evaluator.

Epstein provides an apt analogy to structurally guided musical motion (leaving out the performer for the time being) in the form of a roller coaster. The various factors that control its motion—gravity, friction, the slopes of the tracks—need to be in balance, so that the car neither overshoots nor stops short of its final goal. Musical structures need to be similarly balanced in order to result in motion that is goal-directed and terminates smoothly at major structural boundaries. (At the end of the following chapter, Epstein presents an interesting example of a composition—a section of Scriabin's Piano Concerto—in which this balance seems to be absent.) The dualities of beat and pulse, measure and motive, and hypermeasure and phrase are discussed further in considerable detail. Epstein then expands on some broader issues arising from the preceding discussion, including the parallel processing of meter and rhythm, the role of stylistic experience in the perception of meter, and rhythmic ambiguity. At one point he criticizes psychological studies of meter induction because they ignore listeners' pre-experimental experience with musical conventions such as dance rhythms. His point is well taken, but his musical example (2.11, p. 44) rather seems to illustrate that notation (placement of bar lines) and/or the corresponding surface accentuation in performance can determine metrical interpretation, which few would want to dispute.

Epstein further makes the illuminating observation that ambiguity, far from being a problem, makes music interesting by allowing multiple interpretations. He credits Mozart with an especially high "ambiguity quotient" and cites the opening measures of the A-major Sonata, K. 331, as a well-known example. At one point he states that "[a]mbiguity of perception demands decision" (p. 51). This may be true only at the level of conscious analysis, however. In this reviewer's opinion, performers often make decisions for the listener by providing disambiguating surface cues (articulation, accents, agogics), but they can also refrain from resolving ambiguities, if they so choose. Listeners in turn may be unaware of ambiguities unless they are asked to reflect upon what they have heard. In other words, ambiguity resolution may be cognitive rather than perceptual. This chapter concludes with a discussion of phrase prefixes and extensions as devices for anticipating and prolonging the characteristic motion of a phrase.

The following chapter, as already indicated, provides a number of very instructive music examples that illustrate the concepts reviewed earlier. For example, Epstein points out how exaggeration of surface articulations (such as *crescendi* or *sforzandi*) can distort the flow of the music and change its character. By being overemphasized, such "nonstructural" articulations can "begin to feel like structural emphases" (p. 64). Other examples illustrate large-scale harmonic motion in Dvorák's music and compositionally controlled motion in Brahms's scores, among other things.

Part Three, *Tempo,* constitutes the core of the book. Here Epstein presents and justifies his theory of proportional tempo, familiar from his earlier publications and from various historical precedents, which are summarized briefly in the first chapter. The theory claims that most (all?) music—of the Classical and Romantic periods, at least—is structurally designed so that the tempos of successive movements or sections seem most appropriate when they are related by simple integer ratios, such as 1:1, 1:2, 2:3, or 3:4. The simple phase relationships between the periodicities underlying the tempos are said to give coherence to a multi-movement work. The reason why they do so, Epstein says, lies in human neurobiology.

Epstein elaborates on these biological bases in the following chapter. Here he discusses scientific evidence for oscillatory mechanisms in the brain, often drawing on the work of German researchers he has been in contact with. Much interesting literature on biological clocks and time perception is reviewed (often in extensive footnotes), but it all comes down to three crucial claims: (1) Musical behavior is subject to biological constraints, among which periodic oscillatory processes are especially important. (2) Multiple biological oscillators are drawn towards phase synchrony. (3) Phase synchrony is perceived as pleasant, whereas a disturbance of phase synchrony creates tension and displeasure. The first of these claims is hardly controversial if it is interpreted as meaning that humans can only do what they are biologically equipped to do. It is more debatable if it is interpreted as implying that humans will do only what seems easiest or most natural—a minimum effort principle applied to art. It seems to this reviewer that, in the realm of art, much training is devoted to overcoming certain natural proclivities. If phase synchrony were the overriding principle governing musical timing, then the tempos of most performances would be mechanically exact and in proportion. However, musical tempo choices are much more variable, as Epstein is well aware; thus there must be opposing tendencies or considerations that lead performers to deviate from proportionality and phase synchrony. Tension in music is often more pleasant or at least more interesting than resolution, and both composers and performers often delay resolution by prolonging tension. One is led to wonder whether a violation of tempo proportionality may not also create pleasurable tension or desirable contrast between movements. Although such a tension would be without resolution (and this may be the reason why Epstein does not consider it), it presumably would dissipate quickly as the performer or listener adapts to a new tempo.

Actually, it is not certain that deviations from tempo proportionality can generate tension at all. Epstein's theory rests on certain assumptions that can and should be tested experimentally. In order for phase relationship to play a role, two or more oscillators must be active simultaneously. However, it is doubtful whether listeners can (and want to) maintain a regular beat through a final *ritardando* and the pause that typically follows the end of a symphonic movement until the beginning of the next movement, and it is not known how accurate they would be in judging deviations from tempo proportionality under these conditions, especially for ratios other than 1:1. Collier and Collier (1994) found that jazz musicians, whose tempo sensitivity surely is at least equal to that of their classical colleagues, often deviate markedly from tempos intended to be in 1:2 relationship.

Another empirically testable implication of Epstein's theory is that, if an oscillatory process entrained to an ongoing tempo can indeed be maintained through a *ritardando* and a following pause, it should matter when exactly the next movement starts: Performers should want to start in phase with the ongoing oscillatory period, and listeners should prefer such an in-phase start to an out-of-phase start. However, this seems rather implausible: It could hardly make a difference whether the second movement of a symphony starts a fraction of a second earlier or later after a 20-second silence has elapsed. If so, this would imply that phase relationships are really unimportant and that tempo proportionality, if any, is based on *tempo memory*, a concept Epstein mentions only briefly in a later chapter (p. 412). Indeed, Ivry and Hazeltine (1995) found in a recent psychophysical study that interval duration discrimination is not diminished when the comparison interval is presented out of phase with the periodicity defined by a series of standard intervals, which led them to conclude that "timing is interval based rather than beat based" (p. 17). Memory for temporal intervals explains musicians' ability to reproduce the tempo of an earlier performance, and it may just as well operate across breaks between movements (and even within movements) within the same performance. However, unlike the phase relationship of simultaneous oscillators, tempo memory does not offer any strong reason for why simple tempo ratios should be preferred by performers or listeners. Such a preference may be a matter of personal aesthetic choice, and Epstein seems to grant this (p. 155).

Even so, if Epstein is right about the neurobiological underpinnings and the coherence-lending function of proportional tempo, then most professional musicians should observe the principle, intuitively or deliberately. Therefore, it would be of great interest to examine the tempo choices of famous conductors, chamber ensembles, and soloists, which provide ample and easily obtainable data bearing on the theory of tempo proportionality. In his next chapter, Epstein indeed prepares the reader for such an investigation by discussing methods of empirical tempo measurement. This section is not quite state-of-the-art, as digital waveform editors, now widely available on microcomputers, are mentioned only in passing. Instead, Epstein describes a cumbersome and antiquated method of magnetic tape measurement which he used in his own studies. In order to arrive at a reasonable estimate of the region of uncertainty around observed tempo ratios (or ratios of average beat durations, his preferred measure), Epstein discusses the psychophysical temporal-order threshold (though its relevance is doubtful) and Weber's law. Based on psychophysical findings, he takes the confidence limit to be +/-5%, which seems a reasonable choice. Still, it is important to keep in mind the relatively high probability of finding evidence for tempo proportionality. Epstein permits four simple ratios: 0.5, 0.67, 0.75, and 1.0. Their respective confidence ranges are 0.475–0.525, 0.6333–0.70, 0.7125–0.7875, and 0.95–1.05. The probability that any randomly chosen tempo ratio between 0.5 and 1.0 will support the proportionality theory is (0.025 + 0.0667 + 0.075 + 0.05)/0.5 = 0.43. It also may be noted that Epstein does not deal here with the problem that expressive tempo modulation raises for tempo measurement; he advocates averaging over a number of beats (a procedure for which there is only limited empirical support at present; see Repp, 1994), but in later analyses he often seems satisfied with tempo estimates based on single beat durations.

Primed by Epstein's methodological discussion, this reader was eager to confront the empirical evidence in the following massive chapter of musical examples. He was quite disappointed, therefore, to find that the chapter does not contain any hard data at all. The reason for this is given by Epstein in a footnote, some 30 pages into the chapter:

> It would be methodologically neat...to compile examples of recorded performances and to offer their tempos as proof of the proportional tempo argument advocated here. It would also be unrealistic; for different tempos abound in performances.... To select such an approach would leave us with the fruitless (and unprovable) argument of advocating performer *x* as a "true" advocate of the music, and damning performer *y* as [a] musical infidel.
>
> We have chosen a different approach. Recognizing that most of us probably have a generalized sense of appropriate tempos for this literature, gained in part from our experience of hearing these works, we have designated these tempos as "commonly heard" in examples where such tempos are discrepant from composers' metronome markings. This places the burden of tempo judgment where it properly lies—upon our intuitions, our musical perceptions, our experience with the music (pp. 528–529).

Epstein thus puts the empirical question aside and instead indulges himself in showcasing "our" (i.e., his) scholarship and musicianly insight. This he does brilliantly, and this reviewer, having overcome his initial shock, learned much from reading the chapter. For each of the many works discussed, Epstein provides either the composer's metronome markings or his own estimates of "commonly heard" tempos, or both. (Instead of tempos, he sometimes gives beat durations in milliseconds, but they are derived from the tempos, not actual measurements.) Of course, without empirical data it is impossible to know how accurate the tempo estimates are. For Epstein, tempo proportionality clearly is not just an abstract idea but a recipe for making the "right" tempo choices in order to achieve temporal unification of large works, and for all we know he may have followed this practice for many years. His own preferred tempos surely influence, and perhaps constitute, what he considers to be "commonly heard". Thus it is perhaps not too surprising that example after example yields impressive support for the proportionality theory.

Yet, as Epstein well realizes, great artists often deviate from convention. Famous composers followed individual paths that broke the stylistic rules of their time in one way or another. From that perspective, it seems surprising that all should have followed the principle of (intended) tempo proportionality in all their works. Why did great composers not *deviate* from this tendency in creative ways? Perhaps the answer is that overall coherence was their overriding aesthetic goal. Can performances be perceived as coherent if they violate tempo proportionality? Epstein suggests a negative answer, but this is again an empirical issue, to the extent that coherence can be judged reliably at all. To some extent, the structural coherence of a composition must be independent of the temporal coherence of its performance. Perhaps temporal coherence can be perceived and judged only if one espouses tempo proportionality as an aesthetic goal. In that case, however, the argument would be circular.

Another issue that calls for empirical tests is Epstein's contention that "[i]t is by tempo that the underlying structural shape is heard such that its pacing recalls, indeed identifies, similar elements elsewhere in the work" (p. 172). Indeed, Handel (1993) has found that rhythmical patterns are more difficult to recognize when their tempo is changed. It is unclear, however, whether this is also true for melodic or harmonic patterns, especially when the tempo change is very slight but sufficient to violate proportionality, say from 1:1 to 7:8. Nor is it obvious that a listener must be able to recognize motivic relationships across movements of a symphony in order to appreciate a performance; this may hold only for "musicological listeners" (Cook, 1990) whose aesthetic appreciation rests on analytic insights.

Despite all these reservations, however, it must be said that Epstein's musical examples seem convincing, especially those in which sections having different tempos but containing related motives immediately follow each other. Certainly Epstein makes a strong case for proportional tempos as one possible and even valuable strategy in performing these works. However, it will take a professional music theorist or musician to critically examine the rich and detailed observations in this chapter, which for this reviewer provided mainly an enjoyable educational experience.

The concluding chapter of Part Three finally does contain some empirical data, though not from Western music. Here Epstein summarizes findings (already reported in earlier publications) on tempo changes in the music of non-Western peoples, which he measured from anthropologist's tapes during a research visit to the Max-Planck-Institut in Seewiesen, Germany. The data are reported in meticulous detail, so that readers can follow Epstein's calculations step by step and draw their own conclusions, if they wish. For example, one table and an accompanying graph present beat durations measured at various points during a ritual verbal exchange among Yamomami Indians, lasting some 36 minutes. The durations, apparently of single beats, were "measured by stopwatch, each at a point where a tempo change was detected" (p. 345). From the graph it seems that the tempo accelerated during the initial 18 minutes, though there are many irregularities in the function. From among these irregularities Epstein picked "plateaus" whose beat durations—lo and behold—turned out to be in simple proportional relationships. However, his criterion for what constitutes a plateau seems rather subjective. This reader sees plateaus (if any) in different places. Epstein also sees significance in the finding that an exponential curve fitted to the acceleration portion of the graph passes almost exactly through the chosen plateaus, though this could well have happened by chance. Moreover, the function does not fit the data very accurately; two straight line segments would have done just as well. Thus, while there is not enough space here to discuss every example in detail, it seems that considerable subjectivity was involved in Epstein's analyses of these ethnomusicological data. Even so, tempo proportionality was found in only about 80% of the cases examined. To account for the deviations, Epstein once again refers to the fact that humans are not machines. Returning to Western classical music at the close of the chapter, he points out, however, that "some of us are more gifted than others" (p. 362) in executing precise (proportional) tempo changes. Are these gifted individuals then more machine-like than less gifted ones?

There is indeed a paradox in Epstein's suggestion that the most sophisticated musicians are the ones whose musical behavior is (or should be) governed most strongly by elementary neurobiological principles, whereas the "less gifted" may deviate. Since elementary principles surely must govern elementary behavior, Epstein creates here a "Paradise Lost and Regained" scenario, whereby the experienced musician recovers through insight and conscious effort the innocent perfection he or she has lost along the path of musical training. Why and how that loss has occurred is not clear, however. If biological principles govern music performance, they should govern the activities of all performers, regardless of experience. In fact, the "less gifted" should be more constrained by their innate equipment. By voluntarily submitting to the putative control of biological mechanisms, the "gifted" musician gives up some degrees of artistic freedom. On the other hand, if the neurobiological mechanisms Epstein is envisioning are not innate but are assembled as a function of growing musical experience, then they become merely a scientific (and mechanistic!) metaphor for that experience—the "gift" of neurobiology.

The lost freedom in global tempo choice may be compensated for by exploiting tempo flexibility, which is the topic of Part Four of *Shaping Time*. This is the most empirical part of the book. In an introductory mini-chapter, Epstein announces his strategy: He is going to investigate tempo modulation in selected performances by great artists "that have appeared excellent to one experienced musician's intuitions" (p. 368). This seems a reasonable strategy, assuming that his judgment of excellence was made auditorily, in advance of any measurements. Even so, however, he may have been listening for the very properties that his measurements were expected to confirm, and in that case there is again a certain degree of circularity in the enterprise. A more objective strategy would have been to select a sample of performances at random, measure them, and then have several experienced listeners evaluate the quality of the rubato, to see whether their quality ratings correlate with certain objectively measured properties. But such a larger study was perhaps beyond Epstein's reach at this point, and he appropriately describes his observations as pilot studies. It is interesting to note that his qualms regarding the empirical analysis of artists' tempo choices (see quotation above) did not extend to the analysis of tempo modulation.

The first of the two main chapters in Part Four is on rubato. Epstein begins by distinguishing between classical rubato, where a flexible melodic line weaves around a rigidly timed accompaniment, and romantic rubato, which is totally flexible, yet controlled. This control, Epstein theorizes, derives from the simultaneous operation of two timing mechanisms, a rigid metrical beat and a flexible rhythmic pulse:

> These two time controls, really *systems* of time control, rapidly become dissynchronous and thus in conflict, thereby adding excitement to the performance...A large part of the gratification in good rubato playing lies in the eventual reconciliation of these two systems, their return to phase synchrony....[A]s a general rule this resynchronization of beat and pulse lies within the extreme bounds of the phrase itself. It is at the phrase end (which in its timing is simultaneous with the attack of the next phrase) that the two systems realign. (p. 373)

From this perspective, classical and romantic rubato are similar, except that the explicit ground beat of the former is only implicit in the latter. What Epstein does not say is how such a rigid internal beat in romantic rubato can actually be established and maintained precisely for a number of cycles while a contradictory rhythmic activity is going on. One might suppose that such a ground beat, however it is initiated, would quickly degrade and fade away under these circumstances. Epstein sees no such problem, although he permits some inaccuracy in the system, which he sets arbitrarily at +/-10%. It is important to note that his theory makes no predictions at all about the nature and magnitude of expressive tempo modulations within a phrase, which is what other researchers have been interested in (e.g., Repp, 1992, 1995; Todd, 1985, 1995). The only prediction of his theory is that "the ground beat fits integrally within the phrase" (p. 377), which he proceeds to test in a performance of Chopin's Mazurka in A minor, op. 17, No. 4, by Guiomar Novaes.

In order to determine whether there is an integral number of ground beats in a phrase, it is necessary to measure the duration of the ground beat independently. This leads Epstein to look for it in the performance, somewhere near the beginning of the phrase, *even though it is supposedly implicit.* To this reader, it is not at all clear why the hypothetical ground beat should be manifest anywhere on the musical surface. Moreover, Epstein admits that "it is not always clear where, or in what unit, the ground beat may be for each phrase" (p. 383), though it should be somewhere at the beginning. From several options (first beat, second beat, first bar, second bar, etc.) he chooses one that happens to fit integrally into the duration of the whole phrase, no matter whether the number of such units in a phrase is 11 or 19 or 23. As a result, he finds different sizes, numbers, and durations of ground beats in different phrases, without being unduly bothered by these inconsistencies. He does not take into account the common phenomenon of phrase-initial lengthening, which makes it unlikely that a ground beat is established in the first downbeat or measure of a phrase. For example, the lengthened initial beat at the beginning of the Mazurka melody (Example 11.1f) is a priori unlikely to be a ground beat; yet Epstein accepts its integral fit into the phrase duration as evidence supporting his theory. Later he does consider the possibility that the ground beat is completely hidden, but his discussion becomes confusing here. For example, he attributes the longer beat durations towards the end of the phrase to "an extended influence of the opening ground beat " (p. 388), despite intervening shorter beats. Yet, such a slowing down is commonly observed in the vicinity of phrase boundaries (see, e.g., Todd, 1985; Repp, 1992), and any resemblance to initial beat durations is likely to be purely coincidental.

Epstein's methods imply extremely high probabilities of finding an integrally fitting ground beat by chance. Take a phrase of duration D and a confidence limit of 0.1(±10%). Then the probability that any randomly chosen ground beat duration d will provide an acceptable integral fit to D is 0.2.[1] Now, if there are m possible candidate units for the ground beat, the probability that at least one will provide an integral fit is $1 - (1 - 2c)^m$. With two candidate units, the probability becomes 0.36, with three 0.49, with six 0.74. In the Chopin Mazurka (Example 11.1d), Epstein finds six different ground beat units in 15 phrases. Although he may not have considered all these units in every phrase, it is not clear what a priori constraints he imposed in each phrase. Thus it is difficult to determine whether his results differ significantly from chance expectations; on the (possibly incorrect) assumption that he considered six possible units in every phrase, they do not.

Several additional performances are analyzed in this chapter, though in less detail. Everywhere Epstein finds evidence for periodicity, though not necessarily exact periodicity. This reader remains unconvinced and frustrated by these analyses, which seem to be based on implausible theoretical assumptions and a disregard of conventional statistical procedures. Yet, the many detailed observations presented in this chapter deserve further scrutiny, and Epstein must be given credit for presenting his data with meticulous care and honesty.

The following chapter deals with *Acceleration and Ritard.* Some of the work presented here was published previously (Feldman, Epstein, and Richards, 1992). Epstein and his colleagues propose that the smooth transition between two different tempi is (or should be) effected via a smooth curve that describes successive beat durations as a function of beat number. The proposed shape of the curve is a cubic spline—that is, the central ∫-shaped (for ritards) or inverted-∫-shaped (for accelerations) portion of a cubic function. The cubic function was apparently chosen because of its mathematical simplicity (p. 422) and "[o]n grounds of neural efficiency, or sheer ease of function" (p. 554), a somewhat dubious rationale. Epstein proceeds to fit this function to timing data from performances of orchestral works by Dvorák and Stravinsky, respectively, chosen because they contain long *accelerandi* or *ritardandi.* His two examples of *ritardandi* are described reasonably well by cubic functions, but it must be noted that neither of the functions is ∫-shaped; rather, they are inverted. This is so because the two tempi in each of these instances are *not* smoothly joined; rather, the *ritardando* progresses to its maximum, whereupon the new tempo commences rather abruptly. In the first example, the new tempo actually represents a return to the original tempo preceding the ritard. Epstein thus uses only the concave half of the inverted ∫-shaped function, and he ignores the convex half which does not fit the data at all. Nevertheless, he points to the symmetry of these functions, which "says something about shaping, finesse, eloquence in performance" (p. 424).

The two examples of *accelerandi* are similarly problematic. The first example (De Falla) exhibits stepwise changes in tempo with *accelerandi* in between, and the piecewise or global fits to cubic curves are not really convincing. The second example (Tchaikovsky) provides a somewhat better fit, to a symmetric cubic spline in this instance, but, in his zeal to capture the data, Epstein extends the curve beyond the flat asymptotes of the inverted ∫-shape, so that the smooth connection with the preceding tempo is lost. The accompanying discussion, which relates performance timing to the kinematics of limb movements, is interesting. Todd's (1992, 1995) work, some (but not all) of it too recent to be taken into account by Epstein, has moved in the same direction but has led to different procedures and conclusions. In a forthcoming paper, Todd (submitted) will discuss Epstein's cubic model in relation to his own theory of linear tempo change.

In another analysis, Epstein examines the final *ritardando* in several performances of Schumann's "Träumerei", unaware (at the time) that Repp (1992) had used the very same music in his detailed studies of expressive timing. Repp found that one portion of this final *ritardando* was generally described well by a quadratic function. Epstein, naturally, prefers to fit cubic functions to his data, but his procedure is questionable: He includes the duration of the final chord as a data point, even though this duration is delimited by key releases, events of uncertain rhythmic significance. He also ignores the motivic structure of the final phrase, which typically results in a segmented *ritardando*, as seen in Repp's (1992, 1995) extensive data. Furthermore, Epstein claims to have found a better curve fit for a professional pianist (Jörg Demus) than for several amateurs, but the differences are small and suggestive at best. This reader's shaken confidence in the data was not restored by a final grand cubic curve fit to the 35-minute *ritardando* observed in the Yamomami data, discussed in an earlier chapter.

In a concluding mini-chapter, Epstein looks back at the premises of his approach to flexible tempo. He says that the nonlinear tempos captured by the cubic functions "embody proportionality, for they set in proportional relationship the tempos that they join" (p. 449). This is a *non sequitur* because tempo proportionality is quite independent of the shape of the tempo transition. Epstein also asks (finally!) whether "gifted performance" is characterized by strict adherence to some "innately determined" model of timing, or whether it is rather the deviation from such a model that marks "giftedness". He suggests that the first statement may apply to *accelerando / ritardando*, the second to rubato. However, this raises additional questions: Do less gifted performers play with less pronounced or less controlled rubato than do gifted performers? (Repp's studies suggest they do not.) Are these deviations from the model not themselves governed by biological constraints on timing? (Todd's recent work suggest they are.) Epstein's conviction that some model is needed to explain performers' exquisite control over tempo and timing is shared by most researchers in the field, but it remains to be seen whether his specific proposals will survive.

The book concludes with Part Five, an epilogue on *Affect and Musical Motion.* To illustrate that "[i]t is motion, with its correlated affect, that makes ultimate sense of the music" (p. 457), Epstein discusses a small number of musical examples, especially the first movement of Mozart's Piano Concerto in D minor, K. 466, a piece of "absolute" music that nevertheless seems to have an affective agenda of struggle and entrapment. In his concluding pages, Epstein stresses the importance of shaping musical motion "in the service of controlled affective statement" (p. 481). Particularly apt is his remark in a footnote that a "neutral" or "literal" performance of a score is itself an interpretation, though one that ignores the affective potential of the music.

A final critical comment is in order regarding the total absence from the book of any attempt to trace the fate of motion and affect in 20th century music and aesthetics. Epstein confines himself, without apology, to the masterworks of the standard repertoire that best illustrate his concerns. However, can one ignore one century of radical change in compositional technique and performance aesthetics? Are Epstein's theories confined to a repertoire that by many is considered part of a museum culture? On the other hand, it is important to ask why the standard repertoire still means so much to contemporary audiences, and Epstein is certainly not alone in restricting his focus to the most beloved music of the past. The fate of musical motion and affect in 20th century music still awaits a detailed scholarly discussion.

Despite its shortcomings as an empirical contribution, *Shaping Time* is required reading for anyone interested in music performance. It is an important milestone in interdisciplinary communication and is likely to stimulate vigorous research and constructive criticism from both psychologists and music theorists. It is richly rewarding as a source of musical insights, which are presented in elegant prose and supported by imaginatively laid out music examples. It is virtually free of technical jargon and readily accessible to readers of various backgrounds. The book is well edited—this reader encountered only a few minor errors along the way—and affordable. Its unusually wide format leaves broad margins that invite the reader's notes and comments. Order your copy today.[2]

REFERENCES

Collier, G. L., & Collier, J. L. (1994). An exploration of the use of tempo in jazz. *Music Perception, 11*, 219–242.

Cook, N. (1990). *Music, imagination, and culture*. Oxford, UK: Clarendon Press.

Epstein, D. (1979) *Beyond Orpheus: Studies in musical structure*. Cambridge, MA: MIT Press.

Feldman, J., Epstein, D., & Richards, W. (1992). Force dynamics of tempo change in music. *Music Perception, 10*, 185–203.

Handel, S. (1993). The effect of tempo and tone duration on rhythm discrimination. *Perception & Psychophysics, 54*, 370–382.

Ivry, R. B., & Hazeltine, R. E. (1995). Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 3–18.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America*, 92, 2546-2568.

Repp, B. H. (1994). On determining the basic tempo of an expressive music performance. *Psychology of Music*, 22, 157–167.

Repp, B. H. (1995). Expressive timing in Schumann's "Träumerei": An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, 98, 2413–2427.

Todd, N. [P. McA.] (1985). A model of expressive timing in tonal music. *Music Perception*, 3, 33-58.

Todd, N. P. McA. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91*, 3540-3550.

Todd, N. P. McA. (1995). The kinematics of musical expression. Journal of the Acoustical Society of America, 97, 1940–1949.

Todd, N. P. McA. (submitted). Time and space in music.

FOOTNOTES

Music Perception, in press.

[1] There must be an integer number n such that $n <= D/d < n + 1$. Within this range of 1, values between n and n + 0.1 and between n + 0.9 and n + 1 provide acceptable fits, whereas other values are unacceptable; the probability of finding an acceptable fit thus is 0.2.

[2] This review was written while the author was supported by NIH grant MH-51230. The author is grateful to David Epstein for several opportunities to discuss his ideas with him, and to Janet Hander-Powers for additional discussion and comments on a draft of the review.

The Dynamics of Expressive Piano Performance: Schumann's "Träumerei" Revisited*

Bruno H. Repp

Ten graduate student pianists were recorded playing Robert Schumann's "Träumerei" three times on a Yamaha Disklavier. Their expressive dynamics were analyzed at the level of hammer (MIDI) velocities. Individual dynamic profiles were similar across repeated performances, more so for the right hand (soprano and alto voices) than for the left hand (tenor and bass voices). As expected, the soprano voice, which usually had the principal melody, was played with greater force than the other voices, which gained prominence only when they carried temporarily important melodic fragments. Independent of this voice differentiation, there was a tendency for velocity to increase with pitch, at least in the soprano and alto voices. While there was an overall tendency for velocities to increase with local tempo, there were salient local departures from this coupling. Individual differences in expressive dynamics were not striking and were only weakly related to individual differences in expressive timing.

INTRODUCTION

Variation in timing and dynamics are the two primary means a performer has available to make music expressive and communicative. Together they account in large measure for the gestural quality or "motion" in music that engages the listener (Truslit, 1938; Todd, 1992; Repp, 1993b). This is especially true on a keyboard instrument such as the piano, which essentially fixes the pitch, timbre, and amplitude envelope of each note. The artistic controcccccccccccccl of dynamics presents a special challenge for pianists who must deal with several structural layers at once. Yet, expressive dynamics have been given far less attention than expressive timing in scientific studies of piano performance.

The relative intensities with which notes are to be played are indicated only to a very limited extent in the score. Global prescriptions of dynamics such as *forte*, *mezzoforte*, or *piano* are imprecise and relative, comparable in that respect to tempo prescriptions such as *allegro* or *andante*. However, while tempo can be indicated and measured precisely using a metronome, there is no analogous method of calibrating dynamic level in music. (Of course, such a calibration would make little sense, since dynamic level must be flexibly adjusted to the instrument, room size, room acoustics, etc.) At a more local level in a score, one finds hairpins or verbal instructions of *crescendo* and *decrescendo* (or *diminuendo*), which tell the performer to gradually increase or decrease the dynamic levels of successive tones. However, the precise manner in which this is to be done (i.e., the extent and rate of the increase or decrease) is never specified. At the level of the individual note or chord, marks such as *sforzato* or a wedge are occasionally employed to indicate a dynamic accent. Finally, there are the ubiquitous bar lines, which convey the metrical structure. Notes following a bar line usually have a strong metrical accent, and notes occurring in the temporal center of a bar may receive a secondary accent (metrical

This research was supported by NIH grant MH-51230. I am grateful to Charles Nichols for assistance, to Jonathan Berger for permission to use the Yamaha Disklavier at Yale University, to the pianists for their participation, and to Janet Hander-Powers, Nigel Nettheim, and Neil Todd for helpful comments on the manuscript.

subdivision). These are theoretical accents, however, that are not necessarily to be realized by an increase in intensity. If the music has a dance-like or motoric character, regular accents on the downbeat may be appropriate, but in more lyrical and expressive styles this usually leads to undesirable monotony (cf. Thurmond, 1983).

Clearly, there are only very rough guidelines to dynamics in the score, some of which (the barlines) may even be misleading, and it is up to the performer to make the right choices and provide the fine detail. In particular, performers must decide on the basis of what "feels right" to them how strongly to play tones marked as accented, how to shape the dynamic progression of a *crescendo* or *decrescendo*, how to give expressive melodic gestures appropriate dynamic shapes, and how to give repetitive rhythmic figures a characteristic dynamic profile. These are aspects of "horizontal" dynamics, applied to successive notes. In addition, there are aspects of "vertical" dynamics to consider, which apply to simultaneous notes. Particularly important here are the emphasis of the principal melody over less important voices and the proper "voicing" of chords to make them sound rich and balanced. Consequently, the intricate dynamic microstructure of a performance reveals far more about the performer's skill, taste, and grasp of the musical structure than about his or her observance of prescriptions in the score.

Dynamic shaping and differentiation is a difficult and often neglected aspect of the pianist's skills. It involves risks at both extremes of the range (i.e., the possible production of either an ugly or an inaudible tone) and requires adaptation to the instrument and to room acoustics. Most of all, it requires exquisite motor control and a sensitive ear to guide the hands and fingers. For these reasons, even highly skilled pianists' control over dynamics may be less precise than their control over timing (though level of precision is difficult to compare across different dimensions), and masters of dynamic shading are rare and highly esteemed.

Like expressive timing, expressive dynamics can be measured at several different levels: kinematic, acoustic, and perceptual. The present investigation is restricted to the kinematic level—the varying forces of the pianist's finger movements on the keyboard, as reflected in the registered hammer velocities. Such hammer velocity registrations, using photographic means, were obtained already in Carl Seashore's laboratory (Henderson, 1936; Seashore, 1938), though they were not analyzed in great detail and derived in part from music that did not call for much dynamic differentiation. Some of Seashore's basic observations were that (1) similar musical sections tend to show similar dynamic patterns in performance, (2) metrically accented notes are not necessarily played with greater intensity than unaccented notes, and (3) melody notes are played with greater intensity than the accompaniment. (However, this last distinction was confounded with pitch register, the melody being higher than the accompaniment, and to some extent with hand as well, since the melody was usually played by the right hand.)

Similar apparatus to record the piano hammer action during performance was developed later by Henry Shaffer at the University of Exeter, but his subsequent analyses and publications dealt primarily with timing. Shaffer (1981) examined a single performance of a Chopin Etude consisting of a continuous three (right hand melody) against four (left hand accompaniment) rhythm. He noted that the right hand played louder and had a wider dynamic range than the left hand, and that accents were made independently in each hand. Particularly interesting was his finding (implicit in the fact that he fitted straight regression lines to the data) that *crescendi* and *diminuendi* exhibit a linear sequence of intensity values. He calculated these values as the inverses of the upward transit times of the hammers. Hence they were equivalent to hammer velocities in m/s, which more recently have been shown to be linearly related to the peak amplitude of recorded piano tones (Palmer and Brown, 1991).

The nature of dynamic change in piano performance was examined more closely by Todd (1992). Guided by measurements obtained in Shaffer's laboratory as well as by some acoustic data of Gabrielsson (1987), he proposed a model of expression that links dynamics closely to timing variation. Music is often described as consisting of cycles of tension and relaxation, where increasing tension is manifested overtly in both *accelerando* and *crescendo*, whereas increasing relaxation is associated with both *ritardando* and *diminuendo*. Although this coupling is often violated, Todd considers it a default mode that applies whenever there are no contrary instructions in the score. His test cases were two performances of Chopin's Prélude in f-sharp minor, a piece in which the dynamics indeed seemed to follow the timing variations fairly closely at the beat level: The correlation between local tempo and intensity (averaged over all tones within a beat) was about 0.7.[1] Todd also noted the consistency of two different performances by the same pianist, though it was lower for dynamics ($r = 0.85$) than for timing ($r = 0.97$).

On the basis of his admittedly limited data, Todd proposed a model for the covariation between timing and intensity. The model starts with an analysis of the hierarchical grouping structure (Lerdahl and Jackendoff, 1983) of the composition, as did Todd's (1985, 1989) earlier model of expressive timing variation. Each group within this hierarchical structure is assumed to have a *crescendo-decrescendo* shape composed of two linear segments (in terms of the intensity measure used, i.e., inverse hammer velocity), with the temporal location and magnitude of the peak intensity being free parameters. The dynamic shapes of superordinate groups are then linearly superimposed upon those of subordinate groups. Using an analysis-by-synthesis approach, Todd was able to adjust the free parameters in his model until it fit the Chopin Prélude intensity data about as well as the pianist's two performances resembled each other.

Todd's model is an important advance, but it is in need of testing with more extensive data. Although piano performance data are now relatively easy to obtain, thanks to MIDI technology and computer-controlled pianos, little use has been made of these facilities so far for research purposes, particularly with regard to dynamics. Some researchers, however, have obtained amplitude information from acoustic recordings by reading peak amplitudes off visual displays of the waveform envelope (Truslit, 1938; Gabrielsson, 1987), by using a level recorder (Nakamura, 1987), or by computing the root-mean-square (rms) amplitudes of digitized signals (Kendall and Carterette, 1990). These measures naturally include transformations imposed on the radiated sound by soundboard resonances and room acoustics, which introduce considerable variability specific to the instrument and the recording situation (Repp, 1993a). Also, whenever there are several simultaneous tones, their amplitudes are superimposed (but not necessarily additive).

The most interesting of these acoustic studies in the present context is that of Gabrielsson (1987), because it compared five different performances by well-known pianists with regard to both timing and intensity patterns. The music was limited to the first 8 measures of Mozart's Sonata in A major, K.331, with a repeat available for three pianists. The similarity of the amplitude profiles for the repeated passages was striking, although no correlations were reported. There was also considerable similarity of dynamic patterns across artists. Each of the two 4-bar phrases in the music showed a pronounced amplitude peak near the end, followed by a rapid *decrescendo*. This temporal asymmetry, which inspired Todd (1992) to include a "peak shift" option in his model, reflects the motivic and harmonic structure of the composition. The time course of *crescendi* and *decrescendi* does not seem linear in Gabrielsson's data. The correlation between timing and amplitude profiles was not reported, but it is clear that the phrase-final *decrescendo* was coupled with a *ritardando*, whereas earlier in each phrase there was much less correspondence. There appeared to be a lot of fine detail in the amplitude profiles, though it is impossible to tell whether this was intended by the pianist or caused by irregularities in instrument response and sound transmission.

A final phenomenon needs to be mentioned, and that is an increase in amplitude with pitch. Already observed by Truslit (1938) in the expressive playing of scales, it is one of the intuitive rules incorporated in the performance synthesis systems of Sundberg (1988; Friberg, 1991) and Clynes (1987). In Sundberg's system, the increase is about 3 dB per octave, regardless of the instrumental timbre it is applied to. Clynes developed his system using pure tones and only more recently applied it to realistic instrument sounds. According to him, the "pitch crescendo" should be composer-specific, with very little for Beethoven but as much as 6 dB per octave for Schubert. Different values may apply to piano sounds than to pure tones. Surprisingly, there seem to be no data in the literature on the relation between pitch and intensity (or hammer velocity) in actual piano performance, except for recent study by Palmer (in press) who found only a negligible correlation.

From this review of studies of performance dynamics, a few consistent observations emerge:

(1) Repeated performances of the same music generally have highly similar dynamic patterns.

(2) Like timing, dynamic microstructure seems to reflect the hierarchic grouping structure of the music, with *crescendo-decrescendo* patterns within phrases.

(3) There seems to be a coupling of timing and dynamics, which is most evident in phrase-final *ritardandi / decrescendi*.

(4) The change of successive hammer velocities during a *crescendo* or *decrescendo* may be a linear function of metrical time.

(5) Hammer velocity may increase with pitch.

All statements but the first are tentative because of the very limited amount of published data; they may be considered hypotheses in need of further test. The following analyses will address some of them in a more extensive performance data base than has been examined so far. The music chosen, Robert Schumann's "Träumerei", is particularly well suited to an investigation of expressive dynamics because of its polyphonic structure and its slow tempo, which enables pianists to exercise precise control over each individual tone. Repp (1994) analyzed two pianists' performances of "Träumerei" at three different tempos and found expressive dynamics to be quite stable within and across tempos. The focus here is on a different set of performances, played by 10 graduate student pianists and recorded in MIDI format. A detailed comparison of their timing patterns with those of famous pianists' performances has demonstrated that the student pianists have excellent control over expressive timing and are distinguished from the famous artists mainly in terms of their greater conservatism and smaller individual differences (Repp, 1995a). This relative homogeneity—if it applies to dynamics as well—is in fact advantageous for a study in which typical patterns of expression are of primary interest.

I. METHOD

A. The music

A computer-generated score of Schumann's "Träumerei" is shown in Figure 1. The layout on the page illuminates the phrase structure, which is discussed in more detail in Repp (1992). Metrical positions in the music will be referred to by bar number, beat number, and half-beat number; thus "13-3-2" refers to the second eighth note of the third beat in bar 13. An appended "R" refers to the obligatory repeat of bars 1–8.

B. The pianists

Ten pianists participated in the study as paid volunteers. Nine of them were graduate students of piano performance at the Yale School of Music; five were in their first year, one was in her second year, and three were third-year (artist's diploma) students. The tenth pianist was about to graduate from college and had been accepted into the piano graduate program for the coming academic year. The pianists' age range was 21 to 29, and they had started to play the piano between the ages of 4 and 8. Seven were female, three male. They will be identified by numbers prefaced by the letter P (for pianist).

C. Recording procedure

The recording session took place in a fairly quiet room housing an upright Yamaha MX100A Disklavier connected to a Macintosh computer. The music to be played included four pieces, one of which was "Träumerei". The pianist was given the music and asked to rehearse it at the Yamaha for an hour. Subsequently, the four pieces were recorded once, in whichever order the pianist preferred, and then the cycle was repeated twice. If something went seriously wrong in a performance, it was repeated immediately. One pianist, P4, as a result of several retakes and a computer problem, was able to record only two performances of each piece; all others recorded three, as planned. All performances were fluent and expressive. At the end of the session, each pianist filled out a questionnaire and was paid $50.

The questionnaire asked the pianists, among other things, to rate the adequacy of the Yamaha Disklavier in terms of sound quality and responsiveness. From among the five categories provided, nobody chose "excellent"; the choices were "very good" (2), "good" (4), "adequate" (2), and "poor" (2). The questionnaire also asked in some detail how well the pianists knew each of the pieces. Schumann's "Träumerei" had been previously studied by three (P5. P7, P8) and played informally by two; the rest knew it from listening only. The pianists were also asked to indicate how satisfied they were with their performances, choosing from the categories "best effort", "good effort", "average", "below average", and "poor". The distribution of their responses for "Träumerei" was 0/4/5/1/0.

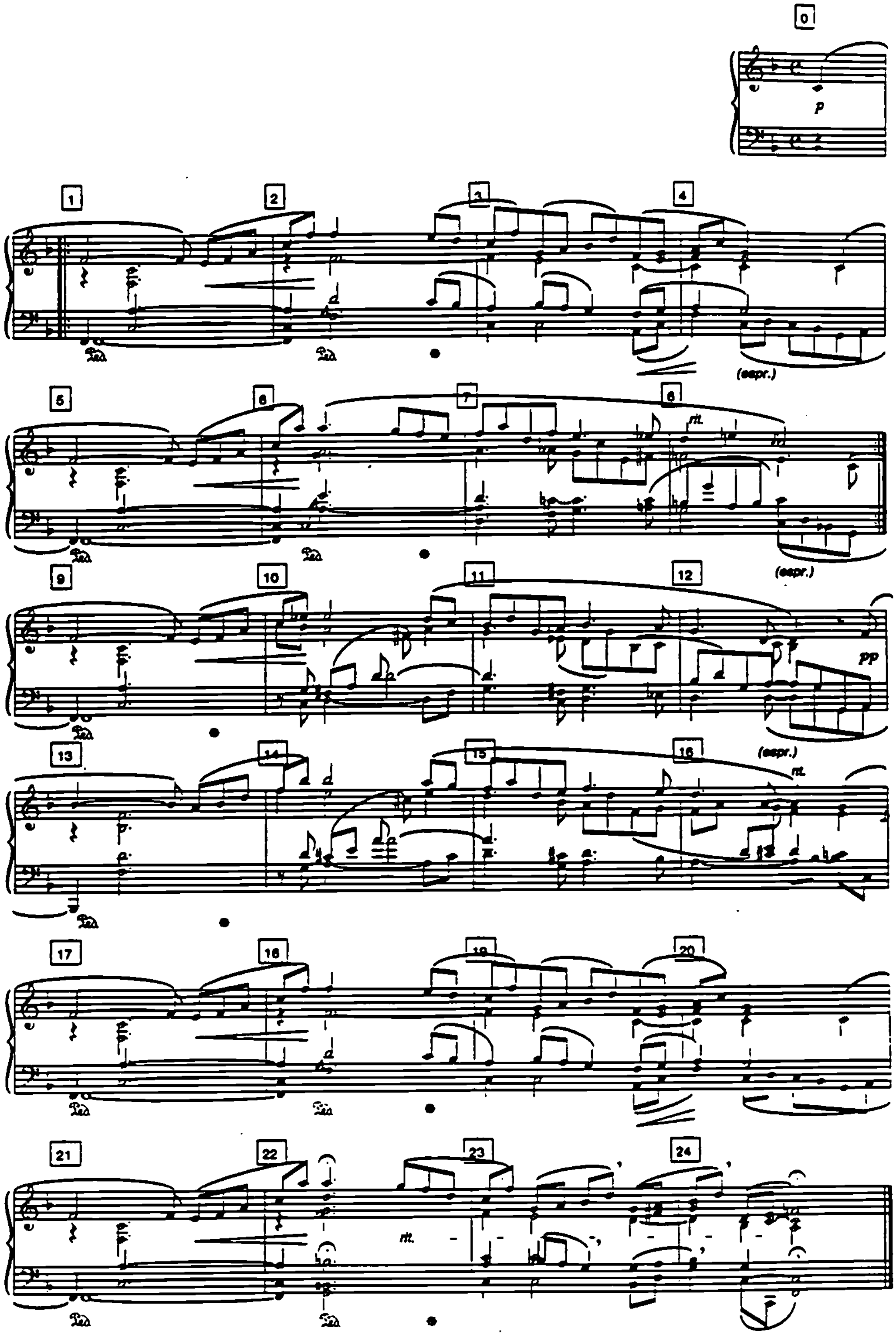


Figure 1. **The score of Robert Schumann's "Träumerei", op. 15, No. 7.**

D. Data analysis

The MIDI data were imported as text files into a Macintosh spreadsheet and graphics program (Deltagraph Professional), where the note onsets were separated from the other events (note offsets and pedal actions) and labeled with reference to a numerical (MIDI pitch) transcription of the score. In that process, wrong pitches (substitutions) were identified and corrected, omitted notes were supplied, and added notes (intrusions) were removed. An analysis of these errors is presented in Repp (submitted). Of the four pieces, "Träumerei" had the smallest number of errors. The 29 performances contained a total of 101 omissions (0.79%), most of which were inner notes of chords.

The parameter of interest in this study was MIDI velocity, or velocity for short, which has a theoretical range from 0 to 127. Its value increases monotonically with hammer velocity, which is picked up by two sensors in the Disklavier, but the precise functional relationship is not known. The relationship between MIDI velocity and peak rms sound level on the Yamaha Disklavier was determined in the course of previous research (Repp, 1993a): For any given pitch, the relationship was nearly linear over the range examined (20–100), though somewhat negatively accelerated towards the higher velocities, with between 3 and 4 velocity units corresponding to 1 dB of sound level for a given pitch.[2] Outside this range nonlinearities do occur. *Fortissimo* notes did not occur in the present performances, but *pianissimo* notes did. Most notes examined, however, had velocities between 20 and 80 and thus were free from any irregularities.

After the initial stages of data analysis, the velocity data for the three performances of each pianist (two for P4) were lined up and averaged to yield an average velocity (or dynamic) profile. From the resulting 10 individual average profiles a grand average profile was then computed, which captures the features shared by most of the performances. This overall profile was then parsed horizontally into "melodic gestures" (Repp, 1992) and vertically into voices. Because of the polyphonic construction of the piece, four voices (soprano, alto, tenor, and bass) can be distinguished throughout, with only a few "secondary" notes not fitting into this scheme. The 457 notes in the piece were assigned to voices as follows: 179 soprano notes (166 primary, 13 secondary), 79 alto notes (67 primary, 12 secondary), 106 tenor notes (89 primary, 17 secondary), and 93 bass notes. A note was considered secondary when it accompanied a seemingly more important (primary) note in the same voice. The principal melody is in the soprano most of the time, but it cascades down through the other voices in bars 7–8, 11–12, and 15–16.

E. Use of the soft pedal

It is common practice among pianists to depress the soft pedal during quiet passages, even though Banowetz (1985) cautions against using the pedal as a substitute for true *piano* playing. Individual differences in the use of the soft pedal were considerable. Five of the pianists (P1, P5, P7, P9, P10) used it almost continuously, with only occasional brief releases that tended to occur in the same places, probably to highlight a chord or brief passage. P6 used the soft pedal in bars 1–8 and from bar 13 on, but not during the repeat of bars 1–8 and during bars 9–12 (except in her first performance). P2 and P3 used the soft pedal only from bar 13 on (where a *pp* appears in the score), and intermittently after bar 18. In the first performance, P3 did not use the soft pedal at all. P4, too, did not make use of the soft pedal, except for a brief episode at the beginning of the first performance. The most unusual case was P8 who used the soft pedal frequently, but only for brief periods, so that it was off for most of the time. This strategy was just the opposite of that of the five pianists who depressed the pedal most of the time. Apparently, P8 used the soft pedal to color specific chords or passages, such as the downbeat and the following chord in bars 1, 5, 9, 13, 17, and 21.

Given these patterns of soft pedal use, the question naturally arose whether the velocity data should be "corrected" to take into account the reduction in sound level caused by the soft pedal. To determine the magnitude of this effect, isolated tones ranging from C2 to C6 in 3-semitone steps, with a fixed arbitrary MIDI velocity of 60, were produced on the Disklavier under MIDI control and were recorded with a microphone, with the soft pedal either raised or depressed. The sounds were digitized, and their peak rms sound levels were measured from their amplitude envelopes. Surprisingly, there was no systematic effect, and individual pairs of tones differed by less than 1 dB. Therefore, no correction was applied for the pianists' use of the soft pedal.[3]

II. RESULTS AND DISCUSSION

A. Reliability

We begin with a consideration of the replicability of dynamic microstructure across repeated performances of the same music. Each performance contained maximally 457 velocity values. (The velocities of omitted notes were left unspecified.) The three performances of each pianist yielded three between-performance correlations whose average was subsequently computed. (For P4 there was only a single correlation.) These average correlations are listed in the first column of Table I. They ranged from 0.732 to 0.897, with a mean of 0.836. Thus they were distinctly lower than the same pianists' between-performance correlations of timing microstructure, whose average was 0.947 (Repp, 1995a).[4] Interestingly, there was a high correlation ($r = 0.826$, $p < .01$) between the two sets of reliabilities: Consistency in timing went hand in hand with consistency in dynamics. The pianist with the highest reliabilities in both domains, P8, happened to be the one who played at the slowest tempo. Overall, however, there was no significant relationship between average tempo and dynamic consistency ($r = -0.153$). Also, there seemed to be no relationship of reliability to familiarity: P5 and P7, who, like P8, had studied "Träumerei" at some time in the past, did not show equally high reliabilities.

Table I also shows the reliabilities for the right and left hands separately. The correlations for the right hand (which played the soprano and alto voices) were as high as those for all notes combined, but those for the left hand (which played the tenor and bass voices) were a good deal lower. This could reflect poorer dynamic control in the left than in the right hand, but it could also be due to the lesser importance of the lower voices and/or to a more restricted dynamic range for them (see below). Next, Table 1 lists the reliabilities for the most important voice, the soprano voice, alone. These were somewhat *lower* than those for the whole right hand, which may again be due to a restriction of the dynamic range, as the pianists surely gave special attention to the soprano voice. The dynamic reliabilities for the soprano voice are more directly comparable to the timing reliabilities, which likewise were based only on the highest notes in each chord (Repp, 1995a), but the conclusion remains the same: The dynamic pattern was less reproducible than the timing pattern.

Table 1. *Average correlations among the MIDI velocities in three performances: (a) all notes (n = 457), (b) right-hand notes only (n = 257), (c) left-hand notes only (n = 200), (d) soprano voice only (n = 179), (e) all notes, bars 1–8 only (n = 113). Also, (f) average within-performance correlations for the two renditions of bars 1–8 (n = 112).*

Pianist	(a)	(b)	(c)	(d)	(e)	(f)
P1	0.853	0.837	0.710	0.788	0.877	0.869
P2	0.888	0.903	0.747	0.887	0.882	0.857
P3	0.797	0.792	0.754	0.736	0.873	0.755
P4	0.865	0.888	0.740	0.878	0.872	0.855
P5	0.825	0.847	0.551	0.810	0.824	0.809
P6	0.850	0.843	0.722	0.803	0.869	0.831
P7	0.825	0.793	0.696	0.740	0.870	0.816
P8	0.897	0.911	0.782	0.845	0.874	0.916
P9	0.831	0.813	0.793	0.791	0.850	0.869
P10	0.732	0.702	0.655	0.599	0.835	0.836
Mean	0.836	0.833	0.715	0.788	0.863	0.841

There could be yet another reason for this difference, which is the presence of large *ritardandi* at phrase endings, which inflate the reliabilities for timing. A fair comparison would consider bars 1–8 only, which do not show such extreme timing deviations and therefore have lower timing reliabilities. Therefore, Table I also lists the dynamic reliabilities for all notes in these initial bars (not including their repeat). They are somewhat higher than the dynamic reliabilities for the piece as a whole, but they are still lower (average of 0.863) than the timing reliabilities for bars 1–8 (average of 0.907). Finally, Table 1 shows the average dynamic within-performance reliabilities for bars 1–8. They are somewhat lower than the between-performance reliabilities, suggesting that at least some pianists (most notably P3) intended to play the repeat differently. A similar decrease in reliability was observed for timing (Repp, 1995a), but again the average within-performance timing reliability (0.899) was greater than the average within-performance dynamic reliability (0.841). A similar difference was reported by Palmer (in press) for the repeat in a Mozart Sonata, played by a well-known concert pianist.

B. Dynamic range

We turn now to the dynamic levels and ranges of the performances, both overall and for the individual voices. From this point on, we will no longer consider the three individual performances of each pianist but only their average. The relevant data are shown in Table 2. The first two columns show the mean velocities of all notes and their standard deviations. It is evident that two pianists (P8, P2) played a good deal louder than the others, who played in what seems an appropriate range for the *piano* prescribed in the score. The overall dynamic ranges of the pianists were fairly similar and in the vicinity of 13 dB.[5]

The means and standard deviations for the separate voices are shown in the remaining columns of the table. Not surprisingly, all pianists played the soprano voice more strongly than the other voices; the difference from the alto voice was 14.5 velocity units on the average, or about 3.6 dB. Likewise, all pianists played the alto voice (right hand) more strongly than the tenor voice (left hand), although the average difference was small, only 3.3 velocity units or about 0.8 dB. Finally, all pianists played the bass voice somewhat more strongly than the tenor voice, the average difference being 2.5 velocity units or about 0.6 dB. The alto and bass voices were of similar average intensity. As to dynamic range, the alto voice actually exceeded the soprano voice, which in turn had a wider range than the two lower voices.

C. Dynamic level and pitch

Given the relative prominence of the soprano voice, there was clearly an overall relationship between pitch height and dynamic level. This was confirmed by computing correlations between MIDI pitch and velocity for each individual average performance. These correlations ranged from 0.40 to 0.65, with an average of 0.57. However, this relationship could have been due to the pianists' intention to emphasize the principal melody over the other voices. The relation between pitch and dynamic level is therefore better investigated within each voice.

Table 2. *Mean velocities and standard deviations (in parentheses) for all notes and for each voice separately.*

Pianist	All voices	Soprano	Alto	Tenor	Bass
P1	30.4 (11.5)	39.3 (9.3)	27.4 (11.0)	22.3 (7.8)	25.0 (6.7)
P2	40.5 (14.1)	51.5 (11.7)	34.1 (12.6)	32.4 (10.0)	34.1 (8.8)
P3	33.2 (14.0)	41.8 (13.2)	26.4 (12.3)	24.9 (9.1)	31.9 (12.3)
P4	36.6 (12.8)	45.8 (10.8)	30.8 (11.6)	29.3 (9.6)	32.0 (9.9)
P5	32.6 (11.5)	41.3 (10.0)	30.3 (10.4)	25.5 (7.2)	25.6 (6.7)
P6	34.3 (12.6)	43.8 (9.6)	31.8 (11.4)	26.0 (9.0)	27.6 (9.9)
P7	29.0 (13.0)	40.0 (9.7)	24.2 (8.8)	19.0 (6.8)	23.2 (11.2)
P8	47.0 (14.7)	58.2 (10.0)	41.9 (16.3)	38.5 (12.3)	39.3 (8.8)
P9	33.9 (13.8)	43.2 (11.1)	28.7 (14.1)	26.9 (11.6)	28.5 (10.5)
P10	29.9 (13.1)	39.6 (10.7)	23.9 (10.9)	22.4 (9.1)	24.9 (11.5)
Mean	34.7 (13.1)	44.5 (10.6)	30.0 (11.9)	26.7 (9.3)	29.2 (9.6)

There were indeed positive within-voice correlations between pitch height and velocity, both in the soprano voice (average 0.42, range 0.26 to 0.58) and in the alto voice (average 0.51, range 0.40 to 0.69). In the tenor voice, the correlation was weak (average 0.29, range 0.12 to 0.45), and in the bass it was absent (average -0.04, range -0.10 to 0.31). It may be concluded, therefore, that there is an increase in dynamics with pitch, but the relationship is not very strong. Moreover, it seems to hold only in the higher voices, which play a more important melodic role. The present correlations are substantially higher, however, than the one obtained by Palmer (in press) for the melody voice in a Mozart Sonata, which suggests that style and/or piece-specific structure play a role.

D. Intercorrelations

The reliabilities, which were computed between and within each individual pianist's performances, may now be compared with the intercorrelations among different pianists' average performances. Across all notes (n = 457), these intercorrelations ranged from 0.614 to 0.847, with a mean of 0.748. Thus, all pianists' performances were similar to each other in terms of dynamics, but they were less similar than each pianists' multiple performances were to each other (average correlation of 0.836). This was so despite the fact that random variability was reduced in the average performances. Only one pianist (P10) showed higher correlations with several other pianists' average performances than among his own individual performances. Three other pianists showed a higher correlation than their own reliability with just one other pianist, P1 in each case.

The intercorrelations for the soprano voice only (n = 179) yielded a similar picture. They ranged from 0.452 to 0.814, with an average of 0.644. The average individual soprano voice reliability, by comparison, was 0.788. Three pianists (P3, P7, P10) correlated more highly with other pianists than within themselves. Although the uniqueness of each individual performer's expressive profile had been more striking in the timing domain (Repp, 1995a), there is evidence for individuality of dynamic profiles as well.

Both sets of intercorrelations (overall and soprano only) were subjected to principal components analysis and, as for timing (Repp, 1995a), only a single significant component emerged in each case. This indicates that there was a single underlying dynamic pattern that all performances had in common, and that individual profiles represented variations around this common standard. Therefore, the grand average dynamic profile is representative of the group of pianists as a whole.

E. The grand average dynamic profile

This profile was obtained by averaging the velocities across the individual average performances of the 10 pianists. Furthermore, the velocities for the two renditions of bars 1–8 were averaged. The grand average dynamic profile is shown in Figure 2. The different voices are represented by different symbols. Contiguous eighth notes in the same voice are connected. The close similarity of the patterns in bars 1–4 and 17–20 should be noted; they represent identical passages.

The initial quarter-note upbeat (0-4-1) was played softer than the following downbeat. However, when the quarter-note upbeat recurred, overlapping the momentarily prominent bass voice (4-4-1, 20-4-1), it was played more strongly than the following downbeat. This was also true for the eighth-note upbeat at 8-4-2. At the point of modulation to B-flat major (12-4-2, 13-1-1), where a *pp* is indicated in the score, both upbeat and downbeat were played at a similar, lower intensity, though still more strongly than the bass voice. The grace-note upbeat (16-4-2) was also dynamically close to the following upbeat (17-1-1), which initiates the reprise and was played much more softly than its analogues at 1-1-1 and 9-1-1. The dynamic relationship of phrase-initial upbeat and downbeat thus was sensitive to context.

The bass note accompanying the downbeat was always considerably softer. The following 4-note chord was even softer. Its constituent tones were not much differentiated, though the highest of them (assigned here to the alto voice) was usually the strongest. The exception was the chord in B-flat major (13-2-1), in which all tones were about equally strong.[6]

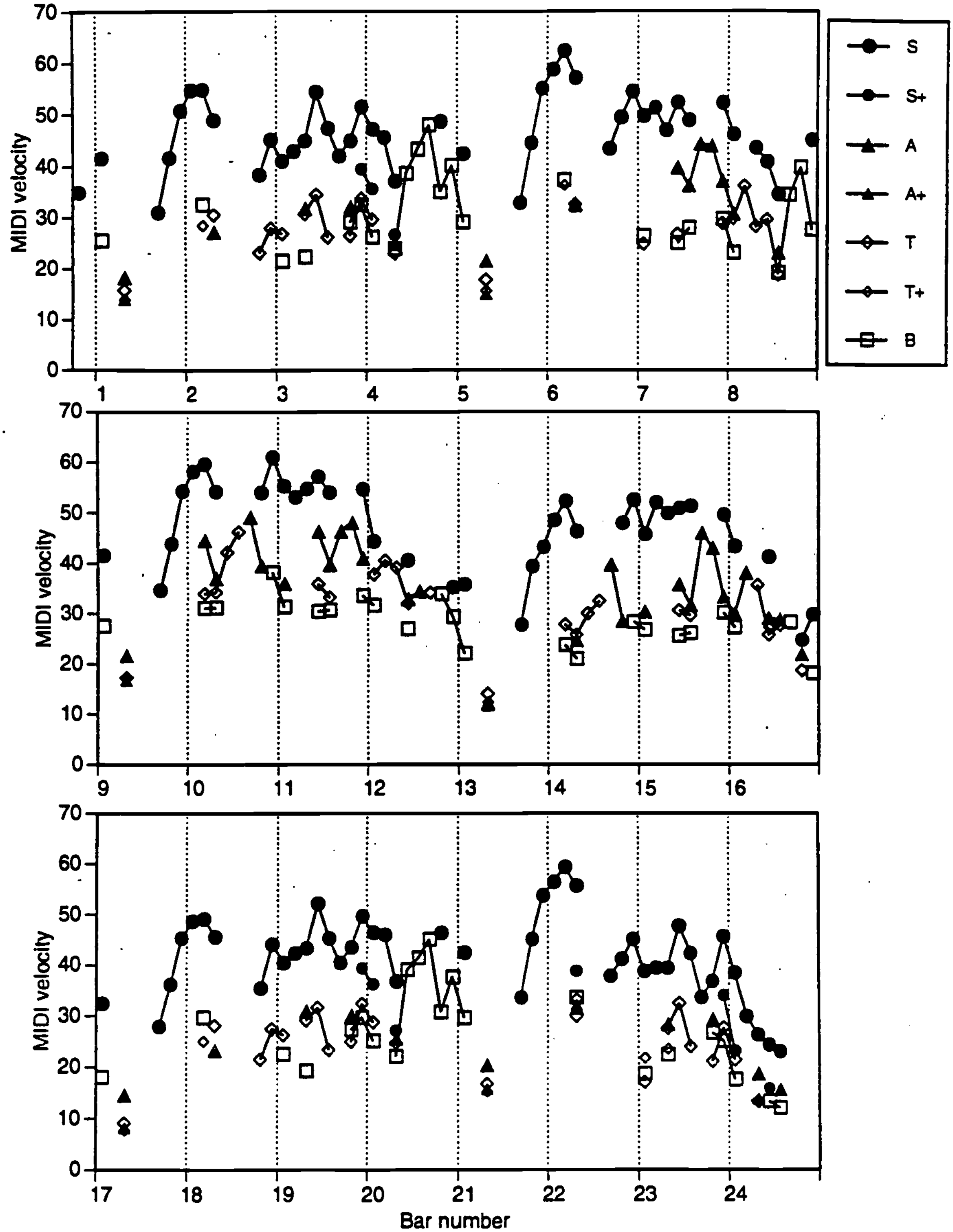


Figure 2. **The grand average dynamic profile for four voices (S = soprano, A = alto, T = tenor, B = bass) and subsidiary notes (S+, A+, T+).**

The following 6-note ascent to the melodic peak in the soprano voice (across bar lines 2, 6, 10, 14, 18, and 22) had a characteristic dynamic shape that was repeated in each of the six phrases. There was a strong *crescendo* over the initial three tones, followed by a much smaller increase to the (unaccented) downbeat and the subsequent pitch peak. The final tone, which repeats and prolongs the highest pitch, was played slightly softer than the preceding tone. The melodies in bars 5–6 and 21–22, which ascend to A_5, were played more loudly and covered a slightly greater dynamic range than those in bars 1–2 and 17–18, which only reach F_5. The one in bars 13–14, which is marked *pp,* was played more softly than that in bars 9–10. It also exhibits a slightly different dynamic shape, probably due to the modulation to B-flat, which necessitated a change in fingering.

Two things are noteworthy about this melodic gesture. First, although intensity increased with pitch, the largest dynamic increase occurred at the beginning, where there was the smallest change in pitch; the largest pitch change (immediately following the downbeat) was accompanied by only a minute dynamic change. This illustrates the loose connection between pitch and dynamics. Second, this phraselet exhibited a pronounced *ritardando,* which culminated at the pitch peak (Repp, 1992, 1995a). The simultaneous *crescendo* is contrary to Todd's (1992) observation of a positive covariation between dynamics and tempo. However, the reduction and slight inversion of the *crescendo* towards the end of the *ritardando* could be due to an underlying trend that counteracts the *crescendo*..

The lower tones accompanying the peak of the phrase were all much softer and not greatly differentiated. In bars 10 and 14, an imitative motive begins in the tenor voice and transfers to the alto. Starting softly, it reached its dynamic peak at the point of transfer (a single note played by the right hand, 10-3-2 and 14-3-2) and then dropped to a lower level for the final tone, which coincides with the resumption of the soprano voice (10-4-1, 14-4-1). The dynamic shape of this imitation motive was quite different from that of its soprano model, especially in that it lacked an initial *crescendo.*

Consider now the soprano voice from 2-4-1 to 4-2-1 and from 18-4-1 to 20-2-1, as well as the nearly identical passage from 22-3-2 to 24-1-2. These are the descents from the melodic peak in the phrases that in previous studies were dubbed Type A (Repp, 1992, 1995a). What is noteworthy here are the dynamic peaks or accents in metrically weak positions immediately *preceding* strong beats (2-4-2, 3-2-2, 3-4-2, and analogous positions elsewhere). The tones in these positions are harmonically and melodically unstable and move strongly towards the following, more stable pitches; they, not the stable and metrically strong tones, were emphasized by the pianists. Parallel patterns at a lower intensity can be seen in the accompanying tenor voice and, in bars 23–24, also in the supplementary soprano voice and in the bass.

At the end of the Type A phrase in bars 4 and 20, the bass voice takes over. Its soft initial tone coincides with the end of the soprano melody, but the following tones were almost at the dynamic level of the soprano. The dynamic peaks again fell on the less stable tones (4-3-2, 4-4-2, 20-3-2, 20-4-2). However, the lower intensity of the bass notes at 4-4-1 and 20-4-1 could also be due to their coincidence with the soprano upbeat to the next phrase, and their low intensity at 5-1-1 and 21-1-1 could be due to their making way for the soprano note on the downbeat. The final soprano tones in bar 24 and their accompaniments trail off towards the final chord, accompanying the extreme *ritardando* at this point.

The second half of Type B phrases is characterized by an overlapping, cascading descent of the melody through the four voices, from soprano to bass. The soprano line (6-3-2 to 7-3-1, 10-4-1 to 11-3-1, 14-4-1 to 15-3-1) was relatively steady but again exhibited a slight accent on the unstable tone preceding the downbeat (6-4-2, 10-4-2, 14-4-2). A similar accent in positions 7-2-2, 11-2-2, and 15-2-2 was merely hinted at, but it emerged more clearly in the alto voice, which takes over at this point. The alto voice then showed a dynamic increase in positions 7-3-2 and 7-4-1 and their later analogues, where it is not accompanied by any other voice. A strong emphasis on the unstable tone at 7-4-2 followed, seen primarily in the soprano voice, which here continues its melodic line while the tenor picks up the cascading motive. The remainder of the phrases in bars 12 and 16 was mainly a *diminuendo* towards the beginning of the next phrase, whereas in bar 8 the bass voice achieved brief prominence (8-3-2, 8-4-1), probably due to the temporary inactivity of the other voices.

F. Dynamics and timing

Todd (1992) postulated a coupling between timing and dynamics, such that the slower the local tempo, the softer the dynamics. This implies a negative covariation between inter-onset interval (IOI) duration and MIDI velocity. (For a graph of the grand average timing profile, see Repp, 1995a, Fig. 2.) Although a local violation of this relationship in the ascent to the melodic peak (bars 1-2, 5-6, 9-10, 13-14, 17-18, and 21-22) has been noted, the overall correlation predicted by Todd nevertheless was confirmed for each individual voice. The coefficients were -0.39 (soprano), -0.52 (alto), -0.49 (tenor). and -0.33 (bass), all significant at $p < .01$. Palmer (in press) recently found a similar relationship for the melody notes of a Mozart Sonata, as played by an excellent pianist.

An additional correlational analysis examined whether individual differences in dynamics might be related to individual differences in timing. The grand average timing profile was subtracted from each pianist's individual average timing profile, and likewise the grand average dynamic profile was subtracted from each pianist's individual average dynamic profile, for the soprano voice only. Correlations were then computed between these residuals. The coefficients were negative for eight of the ten pianists and reached significance in six instances, though they were small. P8 showed the highest correlation (-0.40), P5 the second highest (-0.26). Thus there was a slight tendency for pianists to play relatively soft when they played relatively slow (relative to grand average expressive dynamics and timing).

Todd's model also predicts that velocity during a *crescendo* or *decrescendo* changes as a linear function of metrical distance (see also Shaffer, 1981). Clearly, there are some instance in the present datas where linearity does not hold, particularly in the "ascent to the melodic peak" (bars 1-2, 5-6, 9-10, 13-14, 17-18, and 21-22). The dynamic change may be linear over the first three notes (except in bar 13), but then the velocity increases only by very small amounts (see Fig. 2). A similar observation may be made about the final *decrescendo* in bars 23–24. There are other places, however, where a linear model would seem to capture dynamic changes in the soprano voice quite well, especially across longer IOIs. The dynamic changes in bars 2, 6, 10, 14, 18, and 22 fit a V-shaped pattern that, in bars 10 and 14, includes the tenor and alto voices while the soprano note is sustained. The three-note sequences crossing bar lines 5 and 21, the five-note sequences in bars 12–13 and from bar 7 into bar 8 are other instances where linearity seems to hold. However, a precise evaluation of the model is difficult in the present context because it is not clear whether all velocity minima and maxima should be taken at face value, and whether and how the influence of other factors (pitch, local accents, presence of other voices) should be taken into account.

III. SUMMARY AND CONCLUSIONS

In the Introduction, five hypotheses were deduced from the literature on expressive dynamics. We may summarize now how the present data bear on these hypotheses.

(1) Repeated performances of the same music generally have highly similar dynamic patterns. The present results confirm that the dynamic profile is a stable and replicable characteristic of expressive performance, though its reliability is not quite as high as that of the expressive timing pattern. While timing profiles seem to be unique to individual performers (Repp, 1992, 1995a), there were a few instances here in which a pianist's dynamic profile resembled that of another pianist more than it resembled his or her own profile across repeated performances.

An interesting new finding was that the reliabilities of timing and dynamics are correlated. This may indicate a coupling between these two parameters, or it may be due to individual differences in consistency that are reflected in both primary dimensions of expressive microstructure. Also, the dynamics of the left hand were less reliable than those of the right hand. In part this may have been due to the somewhat narrower dynamic range of the left-hand part, though the difference in range was rather small. A genuine difference in dynamic control between the hands would not be surprising in view of the fact that, in most piano music, the right hand is assigned musically more interesting and technically more challenging material than the left hand. Greater attention to the right hand may enhance the difference. Handedness probably did not play a role here: Two pianists (P5 and P7) were left-handed but showed some of the lowest left-hand reliabilities.

(2) Like timing, dynamic microstructure seems to reflect the hierarchic grouping structure of the music, with crescendo-decrescendo *patterns within phrases.* This observation is also supported by the present findings. Within each of the 4-bar phrases, the intensity of the melody rose steeply during the first half (the antecedent part) and fell gradually during the second half (the consequent part). However, this asymmetry parallels the pitch contour of the principal melody and thus may have been due in part to a covariation of dynamics with pitch (see below).

(3) There seems to be a coupling of timing and dynamics, which is most evident in phrase-final ritardandi/decrescendi. This hypothesis was supported in an overall analysis. However, although phrase-final *ritardandi* were indeed accompanied by *decrescendi* (bars 12, 16, and 24), the phrase-medial *ritardandi* were coupled with *crescendi.* This departure from the default pattern may have been due to the steeply rising pitch, whose effect on dynamics overrode that of the coupling to timing. During the second half of each phrase, there was a tendency to stress metrically weak but harmonically dissonant upbeats more than the following downbeats, which parallels a tendency to lengthen these notes (Repp, 1995a).

(4) The change of successive hammer velocities during a crescendo *or* decrescendo *may be a linear function of metrical time.* The present analysis differs considerably from that of Todd (1992) who computed average intensities of all notes in a beat and modelled them by the superposition of several underlying linear functions, according to the hierarchical phrase structure of the music. Here, expressive dynamics were examined at a finer level of detail, and therefore the results may not bear directly on Todd's model. The hypothesis considered here was that successive individual notes exhibit linear increases and decreases in MIDI velocity, because of an aesthetic preference for this manner of change. The evidence regarding this hypothesis is mixed, probably due to a multiplicity of factors that govern expressive dynamics.

(5) Hammer velocity may increase with pitch. Among the four voices, the soprano voice was clearly the most prominent. This was true even in passages where other voices had the principal melody (bars 8, 12, and 16). The lower voices were not much differentiated; only when one of them assumed melodic significance did it rise above the others. The prominence of the soprano voice was probably due to deliberate emphasis, not just to its high pitch: The other voices also differed in pitch register, but barely in average dynamic level. More unambiguous evidence for a relationship between intensity and pitch was obtained within each of the two right-hand voices, suggesting that higher notes indeed tended to be played louder. The correlation was not very strong, however, suggesting that pitch is by no means the only influence on dynamics. The pitch-dynamics relationship may reflect an expressive convention (Friberg, 1991), but in part it may also be a compensation for a decrease in the perceived loudness of piano tones with increases in pitch.

Individual differences in expressive dynamics were not considered here in great detail. According to principal components analyses, there was only a single underlying pattern, so that the individual dynamic profiles can be regarded as variations around a common standard. A similar finding was obtained in an earlier analysis of these pianists' timing profiles (Repp, 1995a), which contrasted with the diverse timing patterns observed in a group of famous concert artists (Repp, 1992). Unfortunately, it is impossible to recover hammer velocities from acoustic recordings, so a future comparison of student and expert dynamics will have to be conducted on the basis of acoustic energy measures or will have to await the availability of a sufficient number of MIDI recordings by experts. It is not known at present whether famous artists also show greater diversity than students in dynamic patterns. Certainly, they may be expected to show more finely differentiated patterns and more precise control over dynamic contours and balances than young pianists, but these differences may be more quantitative than qualitative in nature. In other words, it seems unlikely that individual dynamic patterns will differ as radically as do some individual timing profiles.

Finally, it should be emphasized once more that the present analysis concerned only the level of MIDI velocities, which are a reflection of the forces applied by pianists' fingers to the keys.

While pianists' intentions may be formulated at the level of action, they are also informed by auditory feedback and thus take into account acoustic and perceptual factors. The relationship between the hammer velocities of individual notes and the resulting sound structure is not simple, as the latter includes effects of instrument characteristics, sound transmission, and the interaction of simultaneous tones. Perception, in addition, introduces phenomena such as masking, fusion, and stream segregation. A multilevel study of expressive dynamics—including kinematics, acoustics, and perception—remains a project for the future.

REFERENCES

Banowetz, J. (1985). *The pianist's guide to pedaling*. Bloomington, IN: Indiana University Press.

Clynes, M. (1987). *Computerized system for imparting an expressive microstructure to succession of notes in a musical score* (U. S. Patent No. 4,704,682).

Friberg, A. (1991). Generative rules for music performance: A formal description of a rule system. *Computer Music Journal, 15*, 56–71.

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K.331). In A. Gabrielsson (Ed.), *Action and Perception in Rhythm and Music* (pp. 81-103). Stockholm: Royal Swedish Academy of Music.

Henderson, M. T. (1936). Rhythmic organization in artistic piano performance. In C. E. Seashore (Ed.), *Objective analysis of musical performance* (pp. 281–305). Iowa City: IA: The University Press. (University of Iowa Studies in the Psychology of Music, Vol. IV).

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*, 129–164.

Lerdahl, F., and Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics, 41*, 525–533.

Palmer, C. (in press). Anatomy of a performance: Sources of musical expression. *Music Perception*.

Palmer, C., & Brown, J. C. (1991). Investigations in the amplitude of sounded piano tones. *Journal of the Acoustical Society of America, 90*, 60–66.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America, 92*, 2546–2568.

Repp, B. H. (1993a). Some empirical observations on sound level properties of recorded piano tones. *Journal of the Acoustical Society of America, 93*, 1136–1144.

Repp, B. H. (1993b). Music as motion: A synopsis of Alexander Truslit's Gestaltung und Bewegung in der Musik. *Psychology of Music, 21*, 48–72.

Repp, B. H. (1994). Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research, 56*, 269–284.

Repp, B. H. (1995a). Expressive timing in Schumann's "Träumerei": An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America, 98*, 2413–2427.

Repp, B. H. (1995b). Detectability of duration and intensity increments in melody tones: A partial connection between music perception and performance. *Perception & Psychophysics, 57*, 1217–1232.

Repp, B. H. (submitted). The art of inaccuracy: Why pianists' pitch errors are difficult to hear.

Seashore, C. E. (1938/1967). *Psychology of music*. New York: McGraw-Hill/Dover.

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartók: Studies in motor programming. *Cognitive Psychology, 13*, 326–376.

Sundberg, J. (1988). Computer synthesis of music performance. In J. A. Sloboda (Ed.), *Generative processes in music* (pp. 52–69). Oxford, UK: Clarendon Press.

Thurmond, J. M. (1983). *Note grouping: A method for achieving expression and style in musical performance*. Camp Hill, PA: JMT Publications.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33–58.

Todd, N. (1989). A computational model of tempo. *Contemporary Music Review, 4*, 405–416.

Todd, N. P. McA. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91*, 3540–3550.

Truslit, A. (1938). *Gestaltung und Bewegung in der Musik*. Berlin-Lichterfelde: Chr. Friedr. Vieweg.

FOOTNOTES

* *Journal of the Acoustical Society of America*, in press.

[1] The tempo measure was obtained by taking the inverse beat duration. The intensity measure was inverse hammer flight time, as in Shaffer (1981). The correlation was computed over 128 beats, though data points were spaced two beats apart; the reason for this is not clear.

[2]Across different pitches, the relationship between velocity and sound level is greatly perturbed by factors such as soundboard resonance and room acoustics (see Repp, 1993a).

[3]This was perhaps to be expected, given that the soft pedal of an upright piano only moves the hammers closer to the strings. At the same time, the acoustic analysis confirmed the presence of large differences in peak rms sound level (here up to 13 dB) between tones of different pitch, as observed previously by Repp (1993a) on the same instrument as well as on a well-maintained Bösendorfer Imperial concert grand. This alarmingly large variation was not "corrected" for because it may depend on microphone position and because it is unlikely that the pianists adjusted their playing in response to it, given their limited experience with the instrument. The MIDI velocities are almost certainly a better measure of the pianists' intentions than is the radiated sound.

[4]This may in part be due to the slow rate of note onsets in the piece; as the event rate increases, timing reliability may decrease more rapidly than dynamic reliability. However, this remains to be investigated.

[5]The dynamic range may be taken to be 4 times the standard deviation. Since about 4 velocity units correspond to 1 dB, the standard deviations can thus be interpreted roughly as ranges in dB.

[6]The terms "note" and "tone" are used interchangeably here, following common usage. Strictly speaking, however, notes are graphic symbols that do not have intensities, and the intensities of the tones are inferred from the measured velocities.

Appendix

SR #	Report Date	NTIS #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 07[illegible] 484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 19[illegible]-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-907
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329	**
SR-97/98	January-June 1989	PB 90-121161	ED 32-1317
SR-99/100	July-December 1989	PB 90-226143	ED 32-1318
SR-101/102	January-June 1990	PB 91-138479	ED 325-897
SR-103/104	July-December 1990	PB 91-172924	ED 331-100
SR-105/106	January-June 1991	PB 92-105204	ED 340-053
SR-107/108	July-December 1991	PB 92-160522	ED 344-259
SR-109/110	January-June 1992	PB 93-142099	ED 352-594
SR-111/112	July-December 1992	PB 93-216018	ED 359-575
SR-113	January-March 1993	PB 94-147220	ED 366-020
SR-114	April-June 1993	PB 94-196136	ED 371-423
SR-115/116	July-December 1993	PB 95-154936	ED 378-624
SR-117/118	January-June 1994	PB 96-122403	ED 390-082
SR-119/120	July 1994-December 1995		

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned

Contents—Continued

Haskins Laboratories Status Report on
Speech Research

SR-119/120
JULY 1994-DECEMBER 1995

Contents

—Continued Inside—

NOTICE

REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☑ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").