DOCUMENT RESUME

ED 401 720 FL 024 170

AUTHOR Zlatev, Jordan
TITLE A Functionalist-Constructivist Perspective on the Ontogenesis of Grammar: The Acquisition of Two Swedish Particles/Prepositions.
PUB DATE 95
NOTE 16p.; Version of a paper presented at the Groningen Assembly on Language Acquisition (1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; Child Language; Foreign Countries; *Grammar; *Language Acquisition; *Language Patterns; Language Research; Language Usage; Linguistic Theory; Native Speakers; *Prepositions; Stress (Phonology); *Swedish; Uncommonly Taught Languages
IDENTIFIERS *Particles (Grammar)

ABSTRACT

A study explored the functionalist-constructivist approach to acquisition of grammar where word classes emerge as a result of distributional differences related to function. Focus is on acquisition of two Swedish forms, "i" and "pa," which can belong to the categories of either particles or prepositions, in two Swedish children, based on longitudinal corpus data. It was predicted that uses of the two forms with the same basic function (locative or directional) would not be acquired simultaneously as prepositions and particles, and that spatial uses of the forms would be acquired before non-spatial uses. Overall, results failed to confirm these predictions. Data for only one child and one form seemed to confirm both predictions. Both children showed no difficulty in acquiring non-spatial uses of "pa" simultaneously with or even before the spatial ones. Three experiments then tested (1) whether preposition and particle usages could be classified based on simple distributional data only, (2) whether consideration of stress patterns facilitated this analysis, and (3) a hypothesized conceptual structure for mapping acquisition. Results lent some support to the functionalist-constructivist approach to acquisition of word classes. Contains 24 references. (MSE)

A functionalist-constructivist perspective on the ontogenesis of grammar: The acquisition of two Swedish particles/prepositions[*]

Jordan Zlatev
Department of Linguistics
Stockholm University

This paper explores a functionalist-constructivist approach to the ontogenesis of grammar where word-classes emerge as a result of distributional differences relevant to functional goals. The empirical focus is on Swedish particles and prepositions. Two forms, *i* and *på*, which can belong to either category have been studied in the production data of two Swedish children and the conclusions are that their acquisition patterns can be more readily attributed to the structure of the input rather than to a strict "semantics-first" strategy. A number of connectionist simulations were performed with a simplified form of the input to the children. Two experiments which used only formal cues failed to acquire an appropriate category structure. The experiment in which the net performed a form-to-function mapping with the simplified data for one of the children was the most successful one. However, the model failed with the data for the other child and there were obvious difficulties in generalizing to novel sentences. These shortcomings do not appear, though, to be inherent to the approach as such, but rather to the particular simulations.

1. Introduction

Even though "grammar" – and even "language" – has come to mean such different things to linguists of different theoretical persuasions that sometimes it seems as though the line of mutual comprehensibility has been passed, there do remain (a few) phenomena of common interest were contact can still be made. One of these is the acquisition of word-class categories. No matter whether you are a passionate believer in innate Universal Grammar, a dedicated constuctivist, a functionalist or a formalist you will very likely give some significance to good old-fashioned notions such as *noun*, *verb*, *adjective*, *preposition* etc. and entertain it as a project worthy of linguistic inquiry to attempt to provide an account of their ontogenesis and thereby of their nature.

What are the major approaches to word-class acquisition? We may schematically present the contestants in a two-by-two table:

	formalist	**functionalist**
nativist	Pinker	Braine
constructivist	Maratsos & Chalkley	?

In the **nativist-formalist** corner is Pinker's (1984) "semantic-bootstrapping" theory. Labels such as N, V, A and P are innate and triggered by equally innate mappings from the

[*] This is an extended version of the paper with the same title which appears in C. Koster and F. Wijnen (eds.) *Proceedings of the Groningen Assembley on Language Acquisition* (GALA) 1995. I wish to express my gratitude to my colleagues from the *Dynamo* project at the Department of Linguistics in Stockholm – Jan Anward, Gunnar Eriksson and Gunnel Källgren – for their feedback, both positive and negative.

semantic categories of *Name of person or thing*, *Action or change of state*, *Attribute* and *Spatial relation, path or direction*, respectively (cf. ibid: 41). Other, "non-canonical" members such as abstract nouns and temporal prepositions join the respective classes only afterwards, through "structure-sensitive distributional learning".

The **constructivist-formalist** view would be reflected in theories of distributional analysis, where it is the formal markings and co-occurrences of the linguistic items that establish their category membership. This is in the spirit of the "decision procedures" of structuralist linguistics (cf. Harris 1951) and its clearest embodiment in more recent work is perhaps Maratsos and Chalkley (1980). (cf. also Zavrel and Veenstra, this volume).

On the functionalist side, where grammar is not autonomous from meaning (at least for the initial stages of language development) most proposals are "nativist" in the sense that they make "some strong assumptions about the structure of semantic representation at the outset of language development" as expressed by Braine (1990) in his assimilation of Pinker's semantic-bootstrapping theory to Schlesinger's (1982) semantic-assimilation theory. Thus, grammatical categories are a reflection of semantic categories, and the first word-classes of the child are bound to reflect a universal "language of thought".

Finally, there is what according to this typology is to be labeled the **functionalist-constructivist** approach: grammatical categories emerge as a result of the mapping of linguistic utterances to meaningful situations, and the constraints on their formation are functional *as well as* distributional in nature. Furthermore, though assigning a large role to prelinguistic cognition, this approach does not require as much structured pre-linguistic knowledge as its nativist counterpart and thus allows for a formative role of language upon thought *as well as* for being its reflection. Though with some respectable ancestry in the work of Vygotsky (1962) this is not a well-trodden path in Child Language Acquisition. Sinha (1988) has argued for it more generally and some of the work of Bates and MacWhinney (e.g. 1987) may be said to approach it. In Linguistics, Anward (1995) has recently proposed a similar perspective, based on typological data.

Why has this approach been under-represented? One reason is because it makes, what may seem from a narrow-scientific perspective to be, "the weakest claim": linguistic categories are *both* functionalally and distributionally induced, rather than either or; functional factors motivate, but do not determine linguistic structure. The other, related, reason is that such an "epigenetic" story has been generally regarded as vague in its predictions and fuzzy in the details.

However, "the times they are a'changing", as the poet has said. The reasons for this, I believe, again are two. First there has come up a good deal of *empirical data* which is problematic for the proponents of the more "scientifically pure" approaches. Just to give a few examples: the word-classes of English and the other European languages are not universal; many languages, such as Chinese, do not distinguish formally between verbs and adjectives and a language such as Riau Indonesian may be analysed as having only one (!) lexical category (Gil, 1994). Studies of language development, furthermore, show early sensitivities to the structure of the particular language, rather than anything resembling a "universal stage" (e.g. Bowerman, 1995; in press). This is obviously bad news for nativist theories of both formalist and functionalist varieties. On the other hand, the combinatorial explosion of possible distributional analyses that would result from a straightforward *unbiased* sampling of the distributions of morphemes (with only consequently registering any "semantic entailments") has forced Maratsos (1990) to accept the need for "some mechanism dictating a more limited choice of possibly important encodings of sequences" (ibid: 1375). These developments clearly point towards the need for a synthesis which does not suffer from the respective drawbacks.

And the reason why it may be more feasible to formulate such a synthesis along "functionalist-constructivist" lines now, rather than some ten years ago lies in the current availability of dynamical systems with emergent properties, such as *artificial neural networks*. Both MacWhinney and Bates (though not in a joint publication) have enthusiastically described connectionism along these lines (cf. MacWhinney, 1989; Bates & Elman, 1992). However, the only connectionist simulation that has explicitly addressed the question of word-class acquisition that I am aware of (cf. Elman, 1990) performed an implicit distributional analysis based on simple word-order information only.

In this paper I wish to explore what I here call a "functionalist-constructivist" perspective on the ontogenesis of grammar by focusing on the categories of **verb-particle** and **preposition** in Swedish. The structure of the argument is the following:

– The two classes need to be grammatically distinguished – even where the actual forms are identical. How this is to be accomplished is far from trivial **(section 2)**.

– A study of the development of two Swedish word-forms as prepositions and as verb-particles (cf. Zlatev, 1995) seems to indicate a sensitivity to the actual speech directed to the children (the "input") rather than to universal "bootstrapping" mappings **(section 3)**.

– When a connectionist model is presented with a simplified and standardized version of the parental input to the two children from the study refered to in section 3, it performs less well when only formal cues participate in the analysis, than when there is a "conceptual structure" to guide the learning **(section 4)**.

2. Swedish (verb-)particles and prepositions

Swedish – like most of the other Germanic languages – has two classes of morphemes which are similar in form and function: prepositions and verb-particles. Talmy (1985) calls the latter "satellites" and insists on their categorical separation from the first, despite the fact that this may often be difficult.

> "English ... has come to regularly position satellite and preposition next to each other in the sentence. For some of these juxtapositions, a kind of merged form has developed [e.g. *I drove past him*, note stress on *past*], while for others – especially where two occurrences of the same shape might be expected – one of the two forms has dropped."(ibid: 105)

Swedish can also position particle and preposition next to each other, as in (1) and (2), where it is unproblematic to distinguish the two due to linear order and the fact the particles *in* and *ner* cannot appear as heads in prepositional phrases. Furthermore – and this is their most reliable characteristic in Swedish – verb-particles receive heavy stress (signaled by **bold face** in the examples), prosodically marking them as belonging to the verb-complex, rather than to the prepositional phrase.

1. Pojken gick **in** i rummet.
   boy-DEF went in in room-DEF
   'The boy went into the room.'

2. Pojken ramlade **ner** i vattnet.
boy-DEF fell down in water-DEF
'The boy fell into the water.'

According to these criteria, however, forms such as *i* ('in') and *på* ('on') should be classified either as prepositions, as in (3) and (4) or as verb-particles, as in (5) and (6).

3. Leksakerna ligger i lådan.
toys-DEF lie-PRES in box-DEF
'The toys are in the box.'

4. Boken ligger på bordet.
book-DEF lie-PRES on table-DEF
'The book is on the table.'

5. Spiken sitter **i** (i trädet) /*trädet
nail-DEF sit-PRES in (in tree-DEF) /*tree-DEF
'The nail is (stuck) inside (the tree).'

6. Kaffet är **på** (i köket) /*plattan
coffee-DEF is on (in kitchen-DEF) /*heater-DEF
'The coffee is (turned) on (in the kitchen).'

It is not appropriate to analyse the sentences in (5) and (6) as involving "optional Ground nominals" (cf. Bowerman, 1995) since *i* and *på* as particles cannot be followed by a nominal – while they can be followed by a prepositional phrase. If *i* and *på* were followed by, what I will call, a **landmark nominal** they would be prepositions and the meaning of the utterances would be purely spatial.

In those cases where the particles *can* be followed by a nominal – when they co-occur with transitive and di-transitive verbs – the nominal is usually the direct object which in the cases of spatial descriptions will be the **trajector nominal**.

7. Jag lägger **i** leksakerna (i lådan).
I put-PRES in toys-DEF (in box-DEF)
'I am putting the blocks inside (the box).'

8. Du lägger **på** klossarna (på lastbilen).
You put-PRES on blocks-DEF (on truck-DEF)
'You are putting the blocks on (the truck).'

It is actually possible for the particles *not* to follow the verb and to occupy the position that is typical for the preposition – between the TR.NP and the LM.NP as in (9). But in this case, the requirement for the particle to receive the typical stress pattern is that the LM.NP be a pronoun, while some Swedish speakers accept only the *reflexive* pronoun. But if the particle is "attached" to the verb and, therefore, the indirect object (LM.NP) precedes the direct object (TR.NP) it is possible for it to be a full noun phrase, cf. (10).

9. Han sätter byxorna **på** sig/?mig/?dig/??honom/*Pelle.
he put-PRES pants-DEF on himself/?me/?you/??him/*Pelle
'He is putting the pants on himself/me/you/him/Pelle.'

10. Han sätter **på** sig/mig/dig/honom/Pelle byxorna.
he put-PRES on himself/me/you/him/Pelle pants-DEF
'He is putting the pants on himself/me/you/him/Pelle.'

These observations are far from providing a comprehensive description of the behavior of verb-particles and prepositions in Swedish, the exact relationship between which is still a largely unresolved matter (cf. Wellander 1965). They do, however, illustrate the complex interaction between prosodic, distributional and functional factors involved in the picture.

One may wonder, if Swedish linguists have not agreed on the proper characterization of these parts-of-speech, how does the Swedish child manage it? Prosody is most often considered a reliable cue for the particles, but we would also like to know how the child can use this cue as a predictor for the differential grammatical properties of verb-particles and prepositions such as their different "argument-structure". Since, obviously, the partitioning into word-classes is not a goal in itself, but a means to learning the grammar of the language.

3. The INs and ONs of two Swedish children

In an attempt to answer such questions Zlatev (1995) performed an analysis of the longitudinal data of two Swedish children, Markus and Harry, available through the CHILDES database (MacWhinney & Snow 1990). First, I sampled the children's utterances with *i* and *på* from the point of appearance of these forms, until at least 30 utterances (disregarding repetitions) for each form and child were gathered. Thereafter, I classified the forms according to the 6 categories in Table 1: locative, directional, and non-spatial uses of the prepositions and particles, respectively. (Since I had no available information on whether the forms had particle-type stress or not, I made the particle/preposition distinction by (a) structural criteria and (b) by asking adult Swedish speakers to judge whether the utterances would have particle-stress or not.)

Category	Examples
PREP:LOC (preposition, spatial locative)	*bajs i blöjan* (Markus 1;11.12) "doodoo in the diaper"
PREP:DIR (preposition, spatial directional)	*sätta den i en vas* (Markus 2;0.9) "put it in a vase"
PREP:NON (preposition, non-spatial)	*vi titta på bumma* (Harry 2;4.23) "we look at car"
PRT:LOC (particle, spatial locative)	*han måsse ha den på* (Harry 2;8.11) "he must have it on"
PRT:DIR (particle, spatial directional)	*sätta på den* (Markus 1;11.0) "put it on"
PRT:NON (particle, non-spatial)	*hålla i de* (Markus 2;0.16) "hold on to it"

Table 1: Examples of the six categories that the *i* and *på* utterances of Markus and Harry where classified into.

BEST COPY AVAILABLE

Departing, above all, from Pinker's "bootstrapping" model I made the following two predictions.

P1: Uses of *i* and *på* with the same basic function (LOC or DIR) will not be acquired simultaneously as prepositions and particles.

P2: The spatial uses of *i* and *på* will be acquired before the non-spatial ones.

These predictions may be considered to follow, perhaps even more so, from other "semantics-first" approaches such as those which in this paper are labled "nativist-constructivist" (cf. section 1). The reason for this is that while Pinker assumes *instantaneous* word-class formation due to the innate semantic-to-lexical category mappings, Schlesinger's and Braine's theories require a gradual process of assimilation which should result in an observable time-gap.

The results of the study by and large failed to confirm the predictions. The data for only one of the children (Harry), for only one of the forms (*i*), seemed to confirm both predictions. Markus seemed to acquire the forms as particles and as prepositions more or less simultaneously. On the other hand, both children showed no difficulty in acquiring the non-spatial uses of *på* simultaneously or even before the spatial ones. What could this variation depend on? Analysing the speech of the caretakers addressed to the children for the period prior to the emergence of *i* and *på* utterances (and for Markus *during* the period as well – because of the smaller total number of instances) provided one possible influence. As can be seen in Table 2 Markus's family seemed much more fond of using particle-utterances (in bold face) than Harry's. On the other hand, for both children the use of non-spatial *på* (underlined) was proportonally higher than non-spatial *i*.

	PREP: LOC	PREP: DIR	PREP: NON	PRT: LOC	PRT: DIR	PRT: NON	total
Harry i	146 54.07%	43 15.93%	71 26.3%	4 1.48%	4 1.48%	2 0.74%	270
Markus i	70 38.67%	38 20.99%	58 32.04%	**6** **3.31%**	**6** **3.31%**	**3** **1.66%**	181
Harry på	91 33.58%	35 12.87%	<u>115</u> <u>42.44%</u>	9 3.32%	4 1.48%	<u>17</u> <u>6.27%</u>	271
Markus på	50 27.47%	11 6.04%	<u>67</u> <u>36.81%</u>	**16** **8.79%**	**27** **14.84%**	**<u>11</u>** **<u>6.04%</u>**	182

Table 2. Quantitative analysis of the input utterances to Markus and Harry including the forms *i* and *på*.

These results do not, of course, refute the role of semantic cues in the formation of lexical categories. They do, however, like the studies of Bowerman quoted in section 1, indicate an early sensitivity to the patterns of the ambient input. Can this be accounted for from the functionalist-constructivist perspective? Can functional and formal properties collaborate in the induction process and what would the right balance be? Questions such as these call for modeling by dynamic systems which converge to relative equilibria over time – such as the ones described in the next section.

4. The connectionist simulations

The requirements on the type of connectionist model that was to be used for the experiments were that: (a) it should be able to process sequences (b) it should allow the factoring out of constraints corresponding to prosodic, distributional and functional information and (c) it should be simple in the sense of building in as little structure as possible from the start and thus maximizing the role of emergence.

A network model meeting these requirements is the **simple recurrent network** (SRN) model proposed by Elman (1990, 1993) – the same which demonstrated how lexical categories emerging from word-order regularities can be implicitly represented in a neural net (cf. section 1). The basic architecture of an SRN is displayed in Figure 1.

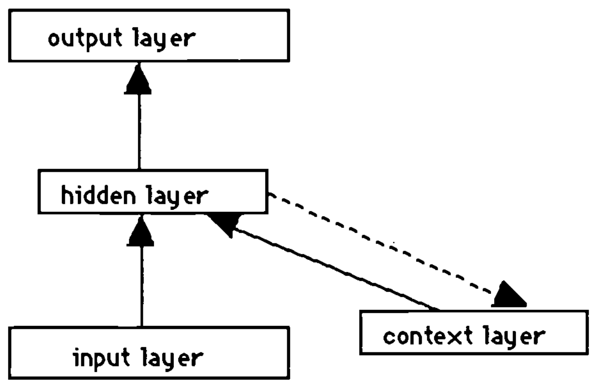

Figure 1. The architecture of a simple recurrent net: the activation pattern during step **n** from the *hidden layer* is copied into the *context layer* and serves as input ("context") during step **n+1**.

As in all artificial neural nets, each layer consists of a certain number of connectionist **units**, which can receive and send activation to other units. In this type of network all units from a certain layer are connected to all units in another layer only in the direction shown by the arrows.

"Training" such a net proceeds as follows: the elements of a given sequence (e.g. the words in a sentence) are presented one-by-one to the **input layer** and a certain pattern (a "teacher signal") is presented at the **output layer** – most often the next element in the sequence (i.e. a "prediction task"). In between is a **hidden layer** which allows for a reorganization of the input and therefore for more complex mappings. The crucial part of an SRN is however the **context layer**, which keeps a copy of the activation pattern in the hidden layer from the *previous* time step. This functions as a gradually receding memory. At the end of every sequence the activation pattern in this layer is "reset", i.e. replaced with a random one, so that the next sequence can begin afresh. The **weights** of the connections between the layers are random when training begins, so the activation pattern that ends up in the output by just passing activation "upward" from the input and context layers will be very different from the teacher signal in the beginning. However, by adjusting the weights with every trial so that the discrepancy is made smaller through an algorithm such as **backpropagation of error** (Rumelhart, Hinton and Williams, 1986), the net gradually converges to a more or less optimal solution.

To apply this type of connectionist model to the questions concerning us in this paper, we need to specify, of course, the nature of the "input" and the "output". In all the experiments the input was a highly simplified encoding of the utterances with *i* and *på* directed to Harry and Markus (cf. Table 2). I will describe this first.

4.1 Encoding the "input"

It would be unrealistic to present as input to the network the available child-directed utterances in "original" form, since that would mean to require from the net to induce all the complexities of grammar from scratch – in effect a *tabula rasa* approach.

On the other hand we do not want to simplify the data so much as to make the task trivial. That is why the following compromise was adopted. Of all the parental utterances with *i* and *på* considered in section 3 only the spatial ones were analysed: 224 for Markus and 326 for Harry. After disregarding differences having to do with non-declarative sentences, topicalization, tense, adverbs and auxiliary verbs it was established that 160 av Markus's input utterances (71%) and 130 of Harry's (40%) could be classified in a number of **types**. Table 3 lists these types with examples and English translations for the input to Markus. The types in Harry's input were similar and the major difference was that there were much fewer instances of the particle utterances (cf. the first six rows in Table 3).

For practical purposes – it is easier to train the net that way – the vocabulary was limited to 32 words, which lead to a further simplification of the input data. All the sentences that were presented to the net could be "generated" by the types and the following "lexicon".

<Agent> -> jag('I'), du('you'), vi('we'), han('he')
<TR.NP> -> jag, du, vi, han, den('it'), mej('me'), dej('you-ACC'), sej('self'), oss('us'), kläder('clothes'), täcke('blanket'), locket('the-lid'), skivan('the-record'), vattnet('the-water')
<LM.rfl> -> mej, dej, oss, sej
<LM.NP> -> bilen('the-car'), tältet('the-tent'), golvet('the-floor'), bänken('the-desk'), vattnet
<static verb> -> sitter('sit'), ligger('lie')
<action verb> -> rullar('roll'), åker('travel'), simmar('swim')
<directional verb> -> går_in('enter'), ramlar_ner(fall-down), häller('pour'), sätter('sit/put'), lägger('lay')

Within these (harsh) limitations I tried to stay as close to the "real" input as possible, so that none of the constructed sentences were semantically anomalous; there was no random generation involved. In this way were formed 109 input sentences for the Markus-simulations and 168 for the Harry-simulations, balanced by type according to the original frequency.

In order not to include any bias concerning the categorization of the word-forms in the input representation, each word-form was encoded with a 32-bit vector, orthogonal to all the others. *i* and *på* had only one vector each, i.e.

i 00000000000000000000000000000010
på 00000000000000000000000000000001

The rationale was that if the net performs an adequate analysis it should nevertheless learn to distinguish between their preposition and particle uses. The question was: what kind of information should it perform the analysis on?

TYPE	Example: "simplified" and original	English translations
<Agent> har <TR.NP> **på** <LM.rfl>	pojken har byxer på sig *ska pojken ha byxerna på sej*	the-boy has pants on self *Will the boy have his pant on?*
<Agent> har **på** <LM.rfl> <TR.NP>	han har på sig byxerna *nu har han på sej byxerna*	he has on self the-pants *Now he has his pants on.*
<Agent> <directional verb> **på** <LM.NP> <TR.NP>	vi sätter på dej blöjan *ska vi sätta på dej blöjan då*	we put on you the-diaper *Are we going to put your diaper on?*
<Agent> <directional verb> **på** <TR.NP>	vi häller på vatten *ska vi hälla på mer vatten här då*	we pour on water *Are we going to pour more water here?*
<TR.NP> <static verb> **i**	borren sitter i *nu sitter borren i där*	the-nail sits in *Now the nail is in.*
<Agent> <directional verb> **i** <TR.NP>	vi lägger i dom *jå nu får vi lägga i dom igen*	we put in them *Well, now we must put them in again.*
<TR.NP> <static verb> på <LM.NP>	hunden sitter på tummen *hunden sitter på Markus tumme*	the dog sits on the-thumb *The dog is sitting on Markus's thumb.*
<TR.NP> <action verb> på <LM.NP>	du bajsar på pottan *sen kan du bajsa på pottan*	you doodoo on the-pot *Later you can sit and doodoo on the pot.*
<Agent> har <TR.NP> på <LM.NP>	du har den på tallriken *kan du ha de på tallriken*	you have it on the-plate *Can you have it on the plate?*
<TR.NP> <static verb> i <LM.NP>	den ligger i lådan *den ska ligga i lådan ja*	it lies in the-box *It should lie in the box, yes.*
<TR.NP> <action verb> i <LM.NP>	vatten rinner i rören å i rören rinner de vatten	water runs in the-pipe *And in the pipe runs water.*
<Agent> har <TR.NP> i <LM.NP>	man har dom i munnen *man ska inte ha dom i munnen*	one has them in the-mouth *One shouldn't have them in the mouth.*
<TR.NP> <directional verb> på <LM.NP>	den ramlar-ner på golvet *ramla- den ner på golvet*	it falls-down on the-floor *Did it fall on the ground?*
<Agent> <directional verb> <TR.NP> på <LM.NP>	vi sätter dej på pappas-axel *kan vi sätta dej på pappas axel*	we put you on dads-shoulder *Can we put you on daddy's shoulder?*
<TR.NP> <directional verb> i <LM.NP>	vi går in i Markus-rum *vi går in i Markus rum*	we go into Markus-room *We go into Markus's room.*
<Agent> <directional verb> <TR.NP> i <LM.NP>	vi lägger nalle i sängen *va ska vi lägga nalle i din säng*	we put teddy in the-bed *Shall we put teddy in your bed?*

Table 3. The types of utterances in the input to Markus, the "simplified" forms that fit them, the original utterances (in *italics*) and English translations of both, the first literal. Note that the "simplified" Swedish sentences are always grammatical.

4.2 Experiment 1: Word-order

The aim of the first experiment was to see if the preposition and particle uses could be appropriately classified based on simple distributional data only. For this purpose Elman's original set-up was used: a prediction task combined with analysis of the "representations" in the hidden layer. The net had input and output layers of 32 units; the hidden and context layers had 20 units.

The procedure was the following: the net was trained on the Markus-input until convergence (i.e. the error stopped decreasing) for approximately 430 repetitions of the training set (i.e. "epochs"). The weights were frozen and the net was tested on a number of input sentences, while the activation patterns in the hidden layer were saved. Finally Hierarchical Clustering Analysis – a statistical technique for estimating the relative distances between multidimentional vectors – was performed on these patterns. Figure 2 displays the result of the analysis for the words in 7 test sentences as a binary-branching tree in which hidden-layer patterns corresponding to the respective words appear as leaves. The closer they are on a "branch" – the closer they have been categorized by the net.

Looking at Figure 2 we see that indeed there is a reflection of implicit grammatical structure: the 7 subject-NPs are grouped together and separated from the VPs; all the verbs hang on separate branch and at the bottom of the graph we can see a branch for what are – mostly! – landmark nominals (indirect objects and adverbials). As far as *i* and *på* are concerned, the particle uses of *på* appear to be grouped together and separated from the preposition use. For *i*, however, this is not the case. There is also another problem: the word *locket* ('the lid') appeared in the context *vi sätter **på** locket* ('we put the lid on') where it is a direct object and TR-nominal. However it is lumped together with the LM-nominals in the lowest branch, rather than higher up, where the other TR-nominals are. This means, in effect, that the net treats this particular use of *på* as a preposition, rather than as a particle, since it is only prepositions which in a context such as this may take LM-nominal arguments.

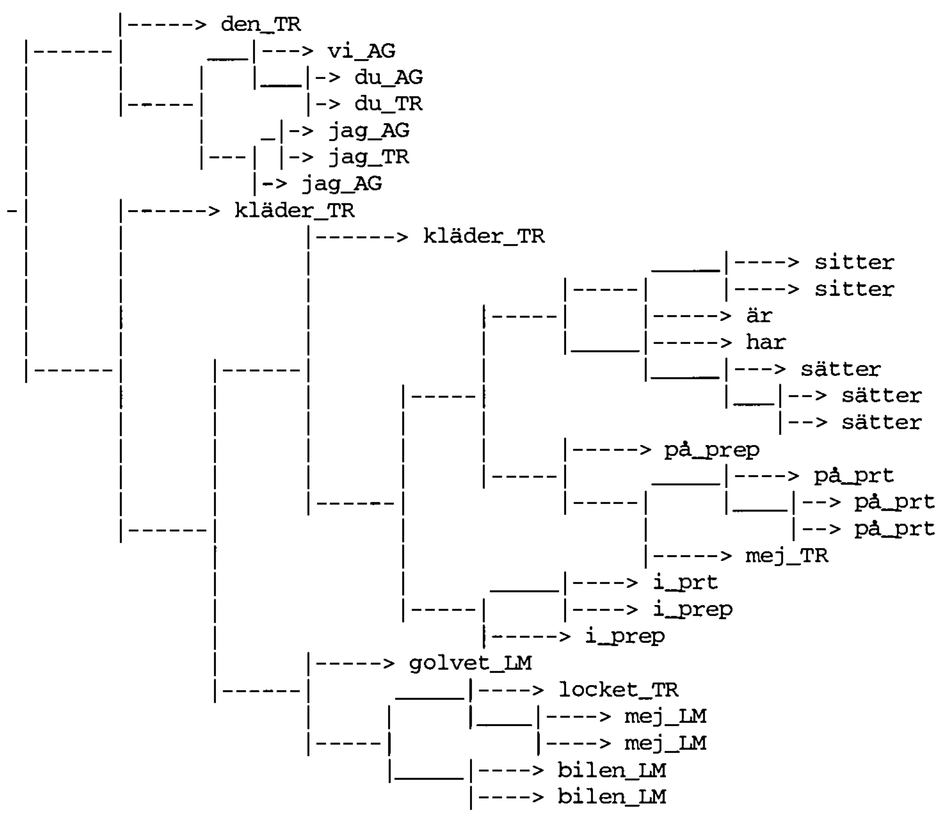

Figure 2. Hierarchical clustering analysis of the activation patterns in the hidden layer of the network which was trained on a prediction task on the Markus-input with word-order information only.

As pointed out in section 2, it is not enough to be able to distinguish the members of word-classes from one another by some superficial characteristic, the classes need to be *structurally* distinct as well, e.g. to participate in the proper classification of the "arguments". And in this respect, the simulation in Experiment 1 obviously failed.

4.3 Experiment 2: Word-order and stress

As described in section 2, the clearest indication for verb-particles in Swedish is their stress pattern. In this experiment the set-up was identical with that in Experiment 1, expect that an extra input node (a "prosody unit") marked stress by being activated for the particle uses, and deactivated for the preposition uses of *i* and *på*. After following the same procedure as in Experiment 1, clustering analysis was performed on the hidden-layer patterns for the same 7 sentences. The graph is displayed in Figure 3.

As expected, the net could now distinguish better between the preposition and particle uses than in Experiment 1. However, *locket*, though closer to the direct objects was still grouped together with the indirect objects and adverbials. The personal pronoun *mej* having the role of trajector is totally miscategorized. On the whole there is very little improvement, if any, compared to Figure 1.

```
          |-----> den_TR
 |--------|     ____|---> vi_AG
 |        |     |   |___|-> du_AG
 |        |-----|       |-> du_TR
 |              |    _|-> jag_AG
 |              |---| |-> jag_TR
 |                  |-> jag_AG
-|         _______|-------> på_prep
 |        |       |_______|-----> på_prt
 |        |               |_____|-> på_prt
 |        |                     |-> på_prt
 |        |               ______|--> sätter
 |--------|              |      |__|--> sätter
          |       |------|         |--> sätter
          |       |      |       ______|---> sitter
          |       |      |------|      |---> sitter
          |       |             |______|------> är
          |-------|                    |------> mej_TR
                  |       ______|------> i_prt
                  |      |      |______|---> i_prep
                  |      |             |---> i_prep
                  |------|       ______|-----> kläder_TR
                         |      |      |-----> kläder_TR
                         |------|       ______|----> locket_TR
                                |      |      |____|----> mej_LM
                                |------|           |____|----> har
                                       |                |----> mej_LM
                                       |______|----> golvet_LM
                                              |____|--> bilen_LM
                                                   |--> bilen_LM
```

Figure 3. Hierarchical clustering analysis of the activation patterns in the hidden layer of the network trained on the Markus-input with information on word-order and stress on the particles.

4.4 Experiment 3: Mapping form and function

The third experiment was intended as an implementation of the "functionalist-constructivist" approach. For this purpose, we need a representation of a **conceptual structure** which could plausibly – given the facts of human embodiment and pre-linguistic development – be in place by the onset of language acquisition. However, we do not want any simple mapping of the "bootstrapping" or "assimilation" type from this structure to linguistic categories. On the contrary, we would expect a *many-to-many mapping* between form and function allowing for variations between languages.

Since all input sentences expressed spatial meaning the following very simple scheme for capturing their corresponding "conceptual structure" was used. There were 3 argument **roles** with individual units for all entities that could appear in them (and were present in the input): **Agent**, **Trajector**, and **Landmark**. Two similar "roles" indicated the **Manner** of motion or stasis and the basic nature of the **Relation** between the TR and LM – here only Inclusion and Support. Finally there were two units signaling the value of **Motion** and **Directionality** in a binary fashion.

Training the net proceeded as follows: as the words of the input sentence are presented one-by-one to the input layer the corresponding conceptual structure is held as "teacher signal" in the output layer (consisting now of 36 units). With the first word of a new sentence, not only is the context layer reset as before, but the appropriate conceptual structure is placed in teacher position. (Similar experiments with SRNs have been performed e.g. by Stolcke (1990)). This, in effect, implements a very strong constraint: at the onset of learning there are whole sequences and there are whole conceptual structures and there is *nothing to tell the net which word corresponds to which part of the structure*. In other words vocabulary and grammar are learned simultaneously. While almost certainly *too* strong, this constraint assures the initial many-to-many mapping.

The net was trained on the Markus-input for a longer time than in the first experiments and when the error stopped decreasing (after approximately 1500 epochs) it was tested by presenting input sentences and monitoring the activation patterns in the output layer. (No Hierarchical Clustering on the hidden layer was performed this time since this type of serial-to-parallel mapping is known not to result in particularly transparent internal representations). The sole criterion for appropriate categorization was proper assignment of the argument structure.

The performance of the net on the data it was trained on was close to perfect as can be seen in Figure 4. This at first look intimidating diagram should be interpreted in the following way: On the top row are the names of the "roles" and below each is a capital letter standing for each of the possible "fillers" (spelled out in the bottom of the figure). Then each row of numbers shows the activation of the corresponding units in the output layer (with "0" representing lowest and "*" highest value) as the words to the right of each row are presented in the input layer, one by one.

It may be helpful to go through the first sentence. When the first word, *jag*, is presented we see that the A nodes under **Agent** and **Trajector** become activated – the net does not yet know the appropriate assignment. The other activations are at this point rather spurious. Then the second word *har* comes in and the situation changes: apart from the Z unit for **Manner** (HAVE) going up to *, the **Agent** unit A receives top activation, the **Trajector** unit A goes down to zero, while there is a strong expectation that the **Trajector** will be F (CLOTHS) and the **Relation** will be Y (SUPPORT). In the third step, the first of these expectations is met, F gets top activation and the expected landmark is A (SELF). In the forth step *på* comes in and solidifies the value of Y and in the fifth step *mej* finishes the picture.

The situation was not always so harmonious: in the last sentence, for example, the net is almost "certain" that the **Landmark** will be L, (TENT). Instead *bilen* ('car') comes in and the net manages to activate the K unit (CAR) only slightly higher than L.

Agent				Trajector										Landmark									MtDr		Manner									Rel		
A	B	C	D	A	B	C	D	E	F	G	H	I	J	A	B	C	D	K	L	M	N	J	/	/	Z	P	Q	R	S	T	U	V	W	X	Y	
4	0	0	0	6	0	0	0	1	1	1	0	1	0	1	1	0	0	2	2	1	0	1	5	4	1	6	0	0	1	0	1	1	0	6	4	jag/I
*	0	0	0	0	0	0	0	0	6	2	0	2	0	4	1	0	0	0	3	0	0	0	0	0	*	0	0	0	0	0	0	0	0	1	9	har/have
*	0	0	0	0	0	0	0	0	*	0	0	0	0	7	0	1	0	0	7	0	0	0	0	0	*	0	0	0	0	0	0	0	0	3	7	kläder/cloths
*	0	0	0	0	0	0	0	0	*	0	0	0	0	9	0	0	0	0	1	0	0	0	0	0	*	0	0	0	0	0	0	0	0	0	*	på/on
*	0	0	0	0	0	0	0	0	*	0	0	0	0	*	0	0	0	0	1	0	0	0	0	0	*	0	0	0	0	0	0	0	0	0	*	mej/me
0	0	6	0	1	0	4	0	1	2	1	0	0	0	0	0	1	0	1	3	1	1	1	5	4	2	5	1	0	0	1	1	0	1	6	3	vi/we
0	0	*	0	0	3	0	0	0	1	1	2	2	0	0	1	0	0	1	0	0	0	0	9	9	0	*	0	0	0	0	0	0	0	0	*	sätter/put
0	0	*	0	0	0	0	0	0	1	2	3	3	0	0	2	2	0	0	0	0	1	0	*	*	0	*	0	0	0	0	0	0	0	0	*	på/on
0	0	*	0	0	0	0	0	0	0	0	9	0	0	0	0	1	0	0	0	0	1	0	*	*	0	*	0	0	0	0	0	0	0	0	*	locket/lid
0	0	1	3	0	0	0	0	*	1	1	0	0	0	1	0	0	1	3	2	2	1	1	3	1	3	2	2	0	1	0	1	0	1	7	3	den/it
0	0	0	0	0	0	0	0	*	0	0	0	0	0	0	0	0	0	6	2	1	1	0	0	0	0	8	0	0	0	0	0	0	0	*	0	sitter/sits
0	0	0	0	0	0	0	0	*	0	0	0	0	0	0	0	0	0	4	6	0	0	0	0	0	0	2	0	0	0	1	0	0	0	*	0	i/in
0	4	0	0	1	7	0	0	1	1	1	0	1	0	0	1	0	0	4	1	1	0	1	4	3	1	6	1	0	1	0	0	0	1	7	3	du/you
0	*	0	0	0	3	0	0	5	2	0	1	0	0	1	0	0	0	2	0	0	0	0	*	*	0	*	0	0	0	0	0	0	0	8	2	sätter/put
0	*	0	0	0	0	0	0	0	3	0	3	3	0	1	0	0	0	0	0	0	0	0	*	*	0	*	0	0	0	0	0	0	0	0	*	på/on
0	*	0	0	0	0	0	0	0	6	0	1	2	0	6	0	0	0	0	0	0	0	0	*	*	0	*	0	0	0	0	0	0	0	0	*	mej/me
0	*	0	0	0	0	0	0	0	9	0	1	2	0	6	0	0	0	0	0	1	0	0	*	*	0	*	0	0	0	0	0	0	0	0	*	kläder/cloths
4	0	0	0	6	0	0	0	1	1	1	0	1	0	1	1	0	0	2	2	1	0	1	5	4	1	6	0	0	1	0	1	1	0	6	4	jag/I
0	0	0	0	*	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	1	0	0	0	0	0	*	0	0	0	0	0	0	0	8	2	sitter/sit
0	0	0	0	*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6	5	0	0	0	0	*	0	0	0	0	0	0	0	0	*	på/on
0	0	0	0	*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	8	2	0	0	0	0	*	0	0	0	0	0	0	0	0	*	golvet/floor
0	4	0	0	1	7	0	0	1	1	1	0	1	0	0	1	0	0	4	1	1	0	1	4	3	1	6	1	0	1	0	0	0	1	7	3	du/you
0	0	0	0	1	*	0	1	0	0	0	0	0	0	0	0	0	0	8	6	0	0	1	0	0	0	7	0	0	1	0	0	0	0	*	0	är/are
0	0	0	0	0	*	0	0	0	0	0	0	0	0	0	0	0	0	0	*	0	0	0	0	0	0	9	7	0	0	0	0	0	0	*	0	i/in
0	0	0	0	0	*	0	2	0	0	0	0	0	0	0	0	0	0	7	6	0	0	0	0	0	0	*	0	0	0	0	0	0	0	*	0	bilen/car

Codes:
A=Self, B=You, C=We, D=He, E=It, F=Cloths, G=Blanket, H=Lid, I=Record, J=Water, K=Car, L=tent, M=Floor, N=Desk, Z=Have, P=Sit/Put, Q=Lie/Lay, R=Roll, S=Travel, T=Swim, U=Enter, V=Fall, W=Pour, X=Inclusion, Y=Support, Mt = Motion, Dr = Directional

Figure 4. Six sentences from the Markus-input tested incrementally after the net has converged after approximately 1500 epochs.

The third sentence with particle *i* is interesting, since what looks like an error – the activation of the P unit (SIT) goes down rather than up – is actually appropriate, since *sitta i* is fairly lexicalized and means approximately "is stuck" (cf. example (5), section 2). On the other hand it is fairly reasonable for the net to assume that the implicit **Landmark** would be CAR or TENT. The net also assigned *locket* appropriately to the **Trajector** role, while *golvet* to the **Landmark** role: the first after the particle, the second after the preposition *på*. Therefore, based on the performance on the "training data" alone the experiment was quite successful.

Unfortunately, categorization of "novel sentences" – such that were not included in the training, but are of the same general types – was not as good. For example, from 13 such sentences (6 with particles and 7 with prepositions) there was at least one error in 8 of them. The errors were 10 misses (non-activated units that should have been activated) and 4 overgenerations (erroneously activated units). On the other hand, of these errors neither one involved the **Landmark** role, which is the role for which the structural distinction between preposition and particles is most relevant. In fact, most of the errors were "lexical": not learning *vi* ('we') properly, for example, was the reason for 7 of the 14 errors.

Therefore, as far as the particle/preposition acquisition is concerned, one may again be fairly content with the model.

Somewhat more worrying for the sanity of the net were, however, experiments identical to the one just described above, but with the simplified Harry-input, instead. The hypothesis was that because of the smaller proportion of particle uses, the net would learn the latter much slower, and for a long period fail to differentiate between particles and prepositions.

The results were that the Harry-simulation performed much poorer for *both preposition and particle examples*. In fact the net never converged at all, i.e. performance was indistinguishable from random.

5. Conclusions

The experiments described in section 4 did lend some support to the argument that a functionalist-constructivist approach to the acquisition of word-classes – and grammar in general – capitalizing on *distributional differences relevant to functional goals* is a likely alternative to the approaches emphasizing either only distributional learning or mechanisms such as "semantic bootstrapping". Rhyming better with empirical data, such as that presented in section 3, it appears to be an approach definitely worth pursuing.

However, the connectionist simulation that came closest to this approach displayed a number of shortcomings. First, the net often allowed distributional regularities to override the need to perform a consistent form-to-meaning mapping (e.g. when one test sentence with *på* activated the Inclusion node, since the other words in the sequence seemed to prefer it). This is, of course, one possible reason for overextension, but the net showed this behavior *too often*, and *too permanently* – further training could seldom change a distributionally induced overgeneration.

Second, variation in the input should be expected to result in differences in the acquisition process over time, not in the possibility vs. impossibility of learning. In other words, the model was *not robust*.

Third, the "conceptual structure"[1] available from the onset of learning was kept simple in the spirit of constructivism, but perhaps it is *too* simple. For instance, as it stands, the model can not distinguish between a reflexive and non-reflexive 3p pronoun, e.g. *han sätter sej* ('he sits himself') vs. *han sätter honom* ('he sits him'). This means that either more structure should be "build in" from the start, e.g. a "binding mechanism", or that such more complex structure should emerge epigentically. However, it is not clear how this could be achieved, given the chosen representation.

These are serious problems, but it does not seem that they are inherent to the functionalist-constructivist perspective. Rather they are directly connected with the fact that the simulation described in this paper is an example of, so called, "toy models". But even as such it serves a purpose in indicating ways to go on. I will end this paper, by just sketching two such pointers:

A more fine-grained representation of **Relation** would allow the model to capture the non-complete functional equivalence of prepositions and the particles. This, on its part,

[1] It would be mistaken to refer to it as "semantic representation" since – unless one adopts some extreme version of nativism such as Fodor's (1975) – there can not be any linguistic meaning prior to language acquisition.

would give rise to more functional/semantic constraints on the emergent grammatical structure.

It is indeed unrealistic not to presuppose any lexical knowledge prior to grammar. A scheme in which the complexity of the input is increased gradually is very likely to improve the model's performance considerably – there is a point to "starting small" (cf. Elman (1993).

References

Anward, J. (1995) Toward a Typology of Part-of-Speech Systems. Paper presented at the Conference on Functional Approaches to Grammar, July 1995, Albuquerque.

Bates, E. and J. Elman (1992). Connectionism and the study of change. In M. Johnson (Ed.), *Brain Development and Cognition: A Reader*. Oxford: Blackwell Publishers.

Bates, E. and B. MacWhinney (1987) Language Universals, Individual Variation, and the Competition Model. In: B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*, Hillsdale, NJ: Erlbaum.

Bowerman, M. (1995). Leaving things out _ and putting them in _: Optional Ground nominals in the acquisition of English. Paper presented at the Workshop on First Language Acquisition: Functional Constraints on Ellipsis, July 1995, Albuquerque.

Bowerman, M. (in press). Learning how to structure space for language–A crosslinguistic perspective. In: P. Bloom, M. Peterson, L. Nadel & M Garret (Eds.) *Language and space*. Cambridge, MA: MIT Press.

Braine, M. D.S. (1990) What sort of innate structure is needed to "bootstrap" into syntax. *Cognition*, 45, 77-100.

Elman, J. (1990). Finding structure in time. *Journal of Cognitive Science* 14, 179-211.

Elman, J. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71-99.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Gil, D. (1994). The structure of Riau Indonesian. In: *Nordic Journal of Linguistics*, Vol 17, 2, 179-200.

Harris, Z. (1951). *Methods in Structural Linguistics*. Chicago: Chicago University Press.

MacWhinney, B. (1989). Competition and lexical categorization. In: R. Corrigan, F. Eckman, & M. Nooman (Eds.), *Linguistic categorization*. New York: Benjamins.

MacWhinney, B. & C. Snow (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, 17, 457-472.

Maratsos, M, (1990). Are actions to verbs as objects to nouns? On the differential semantics bases of form, class, category. In: *Linguistics* 28, 1351-1379.

Maratsos, M. & M. A. Chalkley (1980). The internal language of children's syntax: the ontogenesis and representation of syntactic categories. In: K.E. Nelson (Ed.) *Children's Language, Vol 2*. New York: Gardner Press.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Rumelhart, D., G. Hinton & R. Williams (1986). Learning internal representations by error propagation". In: J. McClelland & D. Rumelhart (Eds), *Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Volume II*: 318-362. Cambridge: MIT Press.

Schlesinger, I. M. (1982). *Steps Toward Language: Toward a Theory of Native Language Acquisition*. Hillsdale, NJ: Erlbaum.

Sinha, C. (1988). *Language and Representation, A socio-naturalistic approach to human development*. NY: Harvester.

Stolcke, A. (1990). Learning Feature-based Semantics with Simple Recurrent Networks, TR-90-015, International Computer Science Institute. Berkeley, California.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In: T. Shopen (Ed.) *Language Typology and Syntactic Description, Vol III: Grammatical categories and the lexicon*, Cambridge, UK: Cambridge University Press.

Vygotsky, L.S. (1986). *Thought and Language*. Cambridge, MA: MIT Press.

Wellander, E. (1965). Aktionsart och aspekt. In *Nysvenska studier* 44, 1964. Uppsala.

Zlatev, J. (1995). Distributional analysis or 'semantic bootstrapping' in the emergence of grammar?, In: *15th Scandinavian Conference of Linguistics*, January 1995, Oslo.

ERIC REPRODUCTION RELEASE FL024170

I. Document Identification:

Title: A functionalist-constructivist perspective on the ontogenesis of grammar: The acquisition of two Swedish particles/prepositions

Author: Jordan Zlatev

Corporate Source: Proceedings of the Groningen Assembly on Language Acquisition

Publication Date: 1996

II. Reproduction Release: (check one)

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in Resources in Education (RIE) are usually made available to users in microfiche, reproduced in paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. If permission is granted to reproduce the identified document, please check one of the following options and sign the release form.

✓ Level 1 - Permitting microfiche, paper copy, electronic, and optical media reproduction.

Level 2 - Permitting reproduction in other than paper copy.

Sign Here: "I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: J Zlatev

Position: Ph.D-ship

Printed Name: Jordan Zlatev

Organization: Department of Linguistics, Stockholm University

Address: S-106 91 Stockholm, Sweden

Telephone: 46-8-16 23 42

Date: October 1st 1996

III. Document Availability Information (from Non-ERIC Source):

Complete if permission to reproduce is not granted to ERIC, or if you want ERIC to cite availability of this document from another source.

Publisher/Distributor:

Address:

Price per copy: Quantity price:

IV. Referral of ERIC to Copyright/Reproduction Rights Holder:

If the right to grant reproduction release is held by someone other than the addressee, please complete the following:

Name:

Address:

V. Attach this form to the document being submitted and send or fax to:

Acquisitions Coordinator
ERIC/CLL
1118 22nd Street, NW
Washington, DC 20037
FAX: 202-659-5641
TEL: 202-429-9292