DOCUMENT RESUME

ED 233 336 CS 207 456

AUTHOR McCutchen, Deborah; Perfetti, Charles A.
TITLE Coherence and Connectedness in the Development of Discourse Production.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 82
NOTE 44p.
PUB TYPE Information Analyses (070) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Coherence; Computer Assisted Testing; *Connected Discourse; *Developmental Stages; Elementary Secondary Education; *Models; *Reading Comprehension; Reading Research; Reading Skills; Writing Processes; *Writing Research; Writing Skills
IDENTIFIERS *Reader Text Relationship

ABSTRACT

Arguing that discourse coherence can provide a window on the writing processes that produce it, this paper describes developmental differences observed in children's writing and proposes a procedural model to account for those differences. The first section of the paper points out that a developmental model of the writing process should specify the sources of coherence in a written discourse and describe the processes that produce it and how they change. It then suggests four aspects of such a model: knowledge of topic, knowledge of text form constraints, topic coherence, and local connectedness. The second section reviews key issues in the development of discourse production, including local, remote, and unsuccessful connections and then presents research findings on connectedness in children's expository and narrative text production. The third section discusses devices of local connectedness, including references, comparisons, lexical ties, implied event, conjunctions, and complex-syntactic devices. The last section details the development of a computer simulation that was designed to specify the workings of several aspects of the model within the writing process. (FL)

COHERENCE AND CONNECTEDNESS IN THE DEVELOPMENT OF DISCOURSE PRODUCTION

Deborah McCutchen and Charles A. Perfetti

Learning Research and Development Center

University of Pittsburgh

1982

The research reported herein was supported by the Learning Research and Development Center, supported in part as a research and development center by funds from the National Institute of Education (NIE), Department of Education. The opinions expressed do not necessarily reflect the position or policy of NIE and no official endorsement should be inferred.

Coherence and Connectedness in the Development of Discourse Production

Deborah McCutchen

and

Charles A. Perfetti

Generally speaking, text research has been concerned very much with comprehension and very little with writing. However, the study of text production may be able to draw on many of the same text models and analytic procedures that have served comprehension. This is because both production and comprehension deal with how meaning is represented within texts. Both the writer and the reader are involved with the meaning representation, one encoding it into a written discourse and one decoding it. Furthermore, the writer's success is generally evaluated in terms of the reader's comprehension. A well written text i comprehensible; a poorly written text is not. This, of course, is not to say that production is simply the reverse of comprehension and that processes of comprehension can be turned on their head to describe writing. However, we suggest that certain specific text features serve an analysis of writing. In this paper we will provide one example of such an analysis.

We do this first by suggesting four aspects of a model of coherence in writing that we believe are important, especially in the development of writing.

We then discuss several studies of writing development in terms of the model of coherence we propose. Finally, we describe a computer simulation which attempts to specify the workings of several aspects of the model within the writing process.

Aspects of a Model of Discourse Coherence in Writing

Text coherence, from our perspective, is important because it implies the writer's awareness of writing as a communicative act. Coherence reflects the writer's attempt to ensure that the meaning decoded by the reader matches the writer's intended meaning. The text mediates the writer's intention and the reader's comprehension, so a coherent text makes comprehension of the intended message much easier.

A developmental model of the writing process should specify the sources of coherence in a written discourse and describe the processes which produce it and how they change. Although details of such a model are not yet possible, it is possible to suggest components of such a model. In particular, we propose four major sources of discourse coherence and discuss in detail the possible processes we see involved in one.

Topic knowledge. The writer's knowledge of the topic is a clear prerequisite for coherence in the written discourse, and increased topic knowledge has implications for all the other processes we will discuss. There is little question about the difficulty of writing about an unfamiliar topic. Therefore, the importance of topic knowledge is presupposed in our considerations of other components of writing, and knowledge is assumed as a constant.

Text form constraints. Related to topic knowledge is the writer's knowledge of the text form (e.g. expository form or narrative form), and the kind of information it implies. Much text form knowledge undoubtably depends on considerable text processing experience. However, some implicit knowledge of the narrative form may be available to even the most immature writers.

For both topic and text form knowledge, we emphasize not the details of the knowledge structures, but the way they constrain and facilitate what gets written. Topic knowledge is essentially the semantic information available in memory. The activation of this semantic information is part of writing. Text form constraints, we will suggest, modify activation by providing directions for organizing and retrieving semantic knowledge. For example, narrative forms direct the writer's attention to temporal and causal links relating semantic concepts, events, etc. in memory.

Topic coherence. Other aspects of a text may be considered quite apart from the global knowledge available to the writer who produced it. For example, a coherent written discourse is clearly distinguished from a list of unrelated sentences because of its topic coherence, i.e., its underlying semantic integrity. Topic coherence refers to the semantic unity of the information which provides the basis of a well formed discourse. Each concept within a coherent discourse fits within the semantic framework provided by the others so that relations among concepts are clearly specified. Sentences in a coherent discourse are not only relevant to the topic, but are part of the underlying structure of the discourse as well. Thus, topic relevance is necessary but not sufficient for coherence.

Topic coherence is, however, a prerequisite for a well formed discourse. It has, in fact, been described as a defining feature of a discourse (van Dijk, 1977b). For that reason, it is difficult to describe the development of topic coherence separate from more local coherence devices which reflect, not the underlying semantic unity of the discourse as a whole, but the way in which that semantic unity is expressed in the connections among smaller units of the written text, i.e., sentences.

Local connectedness. Sentence-to-sentence connectedness reveals the way underlying semantic relations are realized in the written discourse. A coherent written discourse may be defined as one in which the local connections make the underlying topic coherence as explicit as possible at the surface level. The ideas expressed in one sentence are tied to those expressed in neighboring sentences through explicit and implicit connection devices. Connectedness thus presupposes topic coherence. While there is more to coherence than connectedness, we have more well-defined tools with which to study connectedness within the written discourse itself (e.g., the cohesive ties of Halliday and Hasan, 1976). Because these local connections serve discourse coherence, they are critical components of text production and the primary focus of the present analysis.

Developmental Issues

Connectedness among the sentences of a discourse and the semantic and syntactic properties of the connections themselves are key issues in the development of discourse production. We suggest that one hallmark of the development of writing is the increased use of local sentence connections. It is also likely that the use of particular connective devices shows developmental changes.

In the studies of children's writing that we will describe, we have observed some of these developmental differences in local connectedness. Older children write a higher percentage of locally connected sentences than do younger children. Differences in connectedness are also correlated with differences in topic coherence, that is, with the topic constraints that the writers honor in their written discourse. Again, older children are more sensitive to topic constraints.

These two aspects of coherence, local connectedness and topic coherence, are also influenced by text form constraints. While developmental sequences do not seem to change, children write more coherent discourses within the constraints of a more familiar text form such as narrative, compared with the expository text form.

Development of local connectedness

The studies mentioned above have examined the sentence-to-sentence connections made by second, fourth, sixth, and eight grade children writing expository essays and narratives. Following a procedure used by Hidi (1980), we partially constrained the task by presenting children with initial and final sentences. The sentences, however, contained a blank for the child to choose his or her own topic. The sentences

suggested either an expository text form or a narrative, and the children were instructed to write several sentences in between those two which, when taken all together, would form a "good paper" or a "good story."

To analyze the texts produced by the children, we developed a scoring system which parses a sentence (or the independent clause of a compound sentence containing a coordinate conjunction) into a Given portion and a New portion. The Given portion of a sentence is then examined for the connection(s) from it to the New portion of a prior sentence. That is, we examine what, from the reader's point of view, makes the Given information Given. Interjudge reliability between two judges averaged 94% agreement in the scoring of essays and 88% agreement in the narratives.

Through this analysis, we observed two general types of connections, local and remote, plus a substantial proportion of unsuccessful connections and totally unconnected sentences. We will first describe the types of connections, then the developmental trend in their use.

Local connections. Local connections are connections between adjacent sentences. Frequently sentences were connected both to the immediately preceding sentence and to another prior sentence, usually the topic sentence of the paragraph. These are multiple connections, which we counted as local, since they included connections between adjacent sentences. The following excerpt from a subject sample provides an example of local connections.

There are many things about football that make it fun and exciting. The fun side of football is that you get to score winning touchdowns and be the hero of the game. Another fun thing is playing against people who equal to or better than you to see how good you really are.

	GIVEN	NEW
1.		There are many things about football that make it fun and exciting.
2.	The fun side of football is	that you get to score winning touchdowns and be the hero of the game.
3.	Another fun thing is	playing against people who are equal to or better than you to see how good you really are.

In this example, sentence 1 is the initial sentence we provided, with the topic "football" provided by the writer. The initial sentence is treated as entirely New information, since nothing precedes it.[1] The repetition of "fun" and "football" in sentence 2 establishes a local connection between sentences 1 and 2. Sentence 3 also repeats "fun" and so is connected to sentence 1. However, "another" establishes a connection between the reason to be stated in sentence 3 and one that came before, thus linking sentence 3 to 2 as well. Sentence 3 is an example of a sentence with multiple connections.

Remote connections. If sentence 3 were worded instead "A fun thing about football is . . ." it would be an example of what we call a remote connection. A remote connection refers back to the topic sentence, but not to the immediately preceding sentence. Remote connections are independent comments on the topic sentence, and since one sentence does not build on the preceding one, the order of the sentences is less constrained. In the extreme, the sentences are unordered lists of details. Without the word "another," sentence 3

would not be linked explicitly to sentence 2 and thus would not contain a connection to the preceding sentence. Both sentences 2 and 3 would provide support for the assertion that football is fun, but they would do so independently of one another. The majority of remote connections result from such lists of supporting details which are connected only to the topic and not to one another.

Unsuccessful connections. Unsuccessful connections result when the writer either fails to supply a connection or when the attempted connection fails to honor the Given-New contract (Clark and Haviland, 1977). (Failed attempts are often unclear pronominal reference. The writer can also fail to fulfill the Given-New contract by not providing any New information at all. These kinds of sentences are, in a strict sense, "connected," but they make for tedious, repetitive prose. Since they fail to honor the Given-New contract, however, they are included in this category.) The following subject sample illustrates how the writer can fail to supply connections by not building upon prior information.

	GIVEN	NEW
1.		Ice skating is fun and exciting.
2.	Ice skating is fun	because you don't fall all the time.
3.	You can fall	and break a leg.
4.	You must wear	a helmet.

Sentence 2 contains a local connection back to sentence 1 in the repetition of "Ice skating is fun." Sentences 3 and 4, however, may cause comprehension problems for the reader because their Given information is not connected to New information from prior sentences. In fact, the repetition of "fall" in sentence 3 could be particularly disruptive for comprehension, because it could be read as a

contradiction of the information in sentence 2, that skating is fun because you don't fall. Sentence 2 establishes "not falling" as part of the fun, and sentence 3 discusses disastrous results of falling, without signaling any contrast or change of perspective.[2] This example illustrates how analyses which count simple word repetition can fail to capture local coherence. A simple count of repeated words, or argument overlap (Kintsch & van Dijk, 1978), would overestimate the degree to which coherent texts are produced by young writers.

Connectedness in children's expository writing

We observed in children's writing a developmental trend from unsuccessful connections to remote then local connections among sentences. We will first discuss connectedness within expository texts and then within narratives.

Table 1 displays the percentage of connections falling into our three categories - local, remote, and unsuccessful - for children of various ages in each of the two text forms. As Table 1 shows, sixth and eighth grade children wrote relatively more locally connected sentences in the expository text form, 63% and 64% respectively, than did second and fourth grade children, who wrote 37% and 39% respectively. (A chi square analysis within the expository text indicated significantly different patterns among the four grade levels, $X^2=37.67$, $p<.001$.)

Sixth and eighth grade children also wrote relatively fewer unsuccessful connections, 13% and 18% respectively, than did second and fourth graders. The analysis of unsuccessful connections, however, does not perfectly mirror the trend in local connections. An analysis of variance performed on the unsuccessful connections of the four grade

levels showed a significant decrease in unsuccessful connections between second, 51%, and fourth grade, 38% (Turkey's HSD, p=.05). Since local connections do not show a corresponding increase between these grades, one could hypothesize that an intermediate step in development involves connections of a sort other than local. The intermediate step between unsuccessful and local connections between sentences is reflected in the increase of remote connections between grades two and four.

Table 1

Percentage of Connections in each Category within each Text Form

	EXPOSITORY				NARRATIVE			
GRADE	2	4	6	8	2		6	8
Connection Category								
1. Local	37	39	63	64	49	69	69	81
2. Remote	12	23	24	18	40	25	27	18
3. Unsuccessful	51	38	13	18	11	7	4	2

Global text structure. The trend from unsuccessful to remote to local connections became even more clear in an analysis of the patterns of sentence connections in the discourse as a whole. Sixth and eighth grades essays contained a high proportion of local connections, and when the sentences were parsed and their connections charted, most sixth and eighth grade essays exhibited the zig-zag appearance shown by the subject sample in Figure 1.

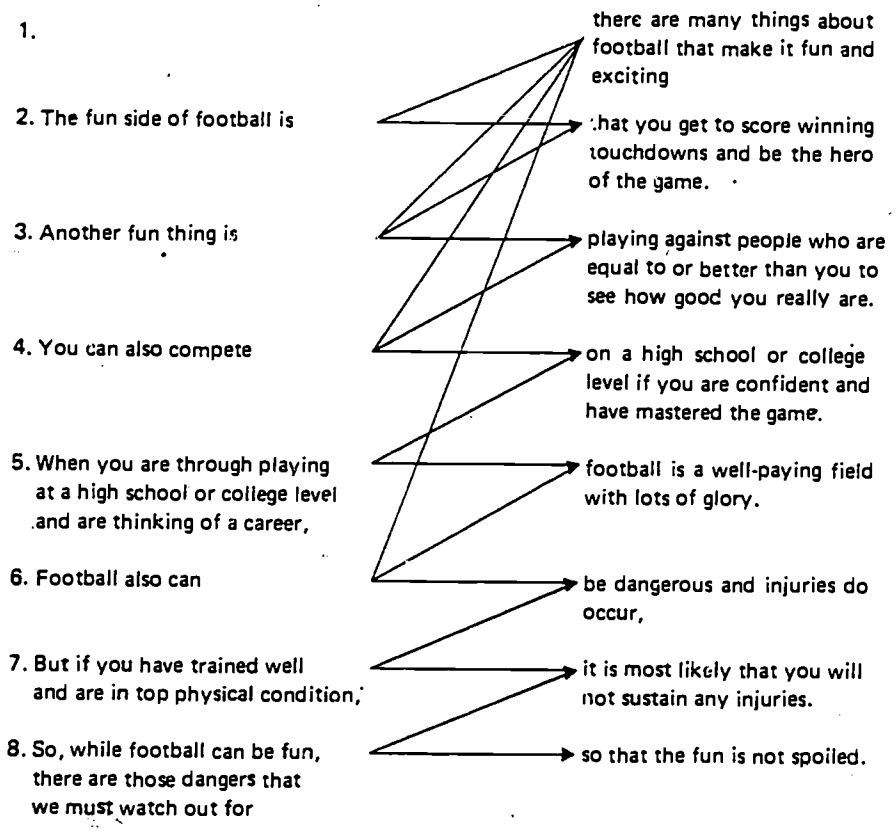

Figure 1. Essay illustrating typical zig-zag structure created by local connections among sentences. Written by an eighth grade subject.

The zig-zag structure of essays such as that in Figure 1 results from the local connection between each sentence and the immediately prior one. Each sentence builds on information from the previous one, and thus the ordered arrangement of sentences is crucial.

Precise sentence order is much less important in essays in which a preponderance of remote connections creates a list-like structure. In these compositions, most frequent in grade four, the topic sentence of a paragraph serves as the sole reference for the Given information in later sentences. As illustrated in the subject sample schematized in Figure 2, later sentences are connected only to the topic, not to each other. Thus the exact order of the supporting sentences is not critical. In Figure 2 sentences 3 and 4, for example, could be interchanged, and since they do not build one upon the other, their meaning would remain intact.

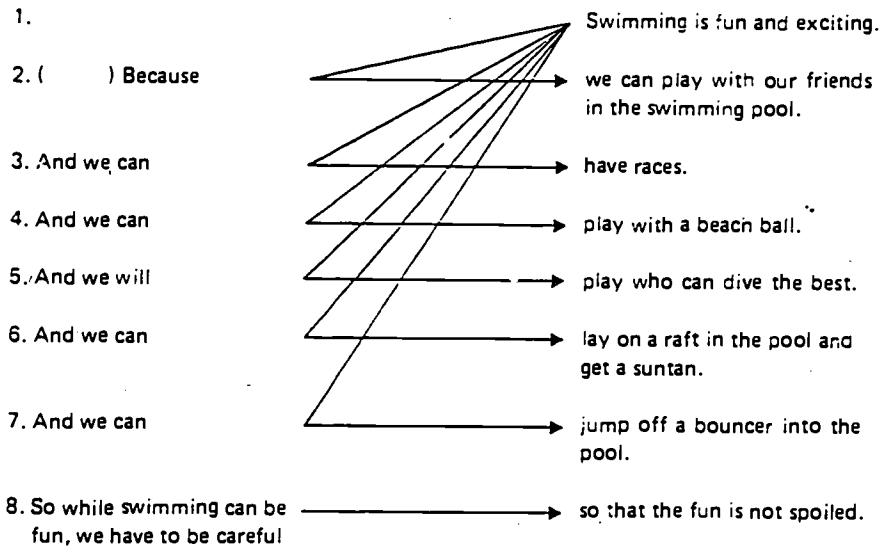

Figure 2. Essay illustrating typical list-like structure created by remote connections to topic sentence. Written by a fourth grade subject.

Not all of the children's essays could be neatly categorized as zig-zag, list-like, or an even mixture of the two. In fact, 43% of second grade essays were too short and contained too many unsuccessful connections for any such pattern to show itself. However, the percentage of essays that could not be classified dropped to 25% or lower in the higher grades. Of those fourth grade essays that could be categorized, the majority, 56%, showed a list-like structure. In contrast, the majority of categorized sixth and eighth grade essays, 67% and 60% respectively, showed a zig-zag structure characteristic of local connections.

The developmental trend in connectedness seems to emerge rather clearly in this analysis of expository discourse. The second grade children in that study failed to make connections in most of their sentences, and fourth grade writers showed an increase in connectedness, but of the remote not local sort. By sixth grade, children were writing a majority of locally connected sentences.

Topic constraints. This developing sensitivity to the demands of local connectedness in discourse seems to be related to the development of the children's sensitivity to topic constraints. As previously mentioned, we provided the children with an initial and final sentence and each child chose a specific topic. An example of one such pair of sentences is presented below:

INITIAL: ( ) is fun and exciting.

FINAL: So while ( ) can be fun, we have to be careful so that the fun is not spoiled.

These sentences imposed three constraints on the children's writing: that they write about a single topic activity, that the activity be fun, and that it have some dangerous aspects to it as well. This third danger constraint is especially apparent if one is sensitive to local connectedness and tries to establish a connection between the final sentence and a previous one. Thus, these sentences determined the topic coherence of possible essays. That is, all three topic constraints had to be met for the final sentence to follow coherently.

Children's ability to honor all three topic constraints reflected their sensitivity to local connectedness. In particular, the danger constraint could be honored by recognizing that the final sentence followed coherently only if the writer made some prior mention of dangers. Thus, the danger constraint was honored by producing local connections at the end of the text.

Second grade children seemed to have only a vague notion of connectedness and the majority of them ignored both the single topic and danger constraints. They wrote lists of various activities they thought were fun. Fourth graders seemed to have a clearer idea of connectedness, but their reliance on remote connections to the initial topic sentence made them focus on only the single topic and fun constraints. Only the sixth and eighth graders were adequately sensitive to local sentence-to-sentence connections to honor the additional danger constraint imposed by the final sentence.

Thus, there seems to be a relation between these two components of discourse coherence: the constraints placed on the underlying topic coherence of a discourse and the local connectedness in the written expression of the topic coherence. If the initial and final sentences

are to be coherent parts of the essay, the topic is determined. It must include both fun and dangerous aspects of the activity. The writer's ability to honor these topic constraints seemed related to the ability to make local connections between sentences.

This may be due in large part, however, to the unnatural aspect of this writing situation--having a predetermined final sentence to write toward. In a sense, the ability of the sixth and eighth graders to honor the danger constraint might also be explained as the ability to plan a composition (i.e., to write with a goal in mind). It has been suggested, in fact, that planning is separate from actual composing only for children of about age 12 and older (Burtis, Bereiter, Scardamalia, and Tetroe, 1981). We will have more to say about this when we discuss possible processes underlying the honoring of constraints of topic and connectedness in discourse.

While these two components of discourse coherence seem related to each other, they are also influenced by the text form constraints. The patterns of connectedness changed somewhat when we observed writers working in the narrative rather than expository text form.

Connectedness in narrative texts

The influence of text forms does not seem to alter the developmental sequence we have just described. Rather, with a familiar text form the sequence appears to be earlier occurring. In the expository texts, remote connections were relatively uncommon until grade four and local connections were not frequent until grade six. By contrast, in the narratives, both connection types were used frequently even by second graders and local connections were prominent by grade

four. Finally, in the narratives eighth graders produced more local connections than they did in the expository texts. These patterns are described in more detail in what follows.

Connectedness. Table 1 displays the percentage of local, remote, and unsuccessful connections written by children of the four ages in the narrative, as well as expository texts.[3] Most interesting is the overall distribution of local and remote connections within the narratives. Second grade writers made almost as many remote connections as local ones, a pattern roughly comparable to that of fourth graders within an expository. The percentage of local connections in the narratives increased between second and fourth, not between fourth and sixth as in the expository texts. Also eighth grade writers showed an increase over the sixth graders in their use of local connections that was not apparent in the expository essays.

The distribution of remote connections within the narratives is the mirror image of that of local connections. The percentage of remote connections decreases between second and fourth grade, and again between sixth and eighth grade (Within the narrative, a chi square analysis indicated significant differences in the distribution of connection categories across grades, $\chi^2=22.57$, $p<.001$.)

Comparing performance in the two text forms, one can see that there are relatively fewer unsuccessful connections overall in the narratives. Unsuccessful connections were infrequent in narratives because any reasonable mention of the main character generally provided a connection, although sometimes a remote one. Since we were not strict in our definition of a story, the story about Bobby could consist of almost any sequence of activities in which he was involved. While some

children wrote better stories than others, most wrote stories of some sort.

Global text structure. The increased connectedness in the children's narratives was also reflected in the global structure of their discourses. Most of the narratives by second grade writers showed either the list-like structure, 38%, or a mixture of list-like and zig-zag, 42%. The zig-zag structure, characterized by local connections, and the mixed structure, characterized by multiple connections, were the dominant structures in narratives from grades four, six, and eight. Thus, the list-like structure, which did not emerge until grade four in the expository texts, was apparent by grade two in the narratives. Similarly, the zig-zag structure emerged in fourth grade narratives, while it had not done so until sixth grade within the expository text form.

Topic constraints. Sensitivity to topic constraints within the narrative also increased with the writer's increased sensitivity to local connectedness. In our study of narrative writing, the initial and final sentences were intended to tap the same knowledge base as did those in the expository study. The sentences were modified, by the introduction of a fictitious character, to suggest to the children a narrative structure. However, like the sentences from the expository study, the final sentence imposed a danger constraint that was not included in the initial sentence:

INITIAL: Bobby always enjoyed ( ) One day...

FINAL: Bobby still thought ( ) was fun, but Bobby knew he would have to be more careful from now on.

Again in the narratives, the children's sensitivity to local connectedness was correlated with their recognition of the topic constraints imposed by the sentences presented them. The narrative text form seemed to provide so much structure that very few children wrote anything other than a narrative about Bobby doing something that was fun. However, recognition of the danger constraint from the final sentence again seemed related to the local connectedness of the discourse.

Second graders wrote slightly fewer than half, 49%, of their sentences with local connections, and 29% of the children in that grade failed to honor the danger constraint. By grade four, children were making local connections in 69% of the sentences in their narratives, and the percentage of children ignoring the danger constraint dropped to 15%. None of the sixth or eighth graders, whose writing was again characterized by local connections, failed to integrate the final sentence into the discourse. (The narrative text form was probably especially helpful for recognition of the complicating danger referred to in the final sentence, since a complication is a major component of a narrative (van Dijk, 1976; van Dijk, 1977a).

Devices of Local Connectedness

While it is interesting to find stability of the developmental sequence within two different text forms, the developmental comparisons suggest something more than the connectedness analysis is needed. Our system of simply tabulating percentages of local, remote, and unsuccessful connections showed fourth and sixth grade performance in the narratives to be very similar. Our holistic judgments of quality, however, were that the sixth grade narratives were better written. We then examined the semantic and syntactic nature of the local connections to see whether the local connections used by the sixth graders differed in kind, if not in number, from those used by the fourth graders.

We evaluated connectedness in a way very similar to that described by Halliday and Hasan (1976). They classify the kinds of cohesive ties occurring between English sentences as those that are semantic--reference and lexical ties--and those that are grammatical (i.e. syntactic)--substitutions, ellipsis, and conjunction. While they focus on the linguistic aspects of the text itself, we focus instead on the development of the writer and how different types of ties suggest different knowledge being used by the writer. We also consider the syntactic complexity of the connection as it is actualized in the written discourse. For these reasons, we tried to keep separate some classifications that we intuitively thought might show developmental differences across the ages we studied. We, therefore, classify connections in groupings slightly different from those of Halliday and Hasan (1976). However, the evaluation of connectedness as such is the same.

Reference. According to our classification scheme, connections of reference are established by a pronoun, or repetition of a word with a demonstrative, which refers to a concept mentioned in a prior sentence. These correspond to Halliday and Hasan's personal and demonstrative reference. The following sentences illustrate both kinds of reference: pronominalization and word repetition.

1. A boy was standing at the cold bus stop.
2. The bitter wind chilled him as he waited patiently.
3. The boy boarded the first bus he could.

In sentence 2, "he" and "him" refer to the boy introduced in sentence 1. The use of the demonstrative "the boy" in sentence 3 also makes clear that this is the same boy mentioned previously.

Comparison. While Halliday and Hasan classify comparison as a third kind of reference, we chose to keep it as a distinct category. We did so because comparison reflects psychological processes that are different from those used in specifying reference. Comparison requires the analysis of the attributes of two or more objects (as in "the most fun of all . . ") or the analysis of their environments in space and time (as in "the next pitch . . ."). We also included "then" in this category, since it too reflects a time sequence.

Lexical ties. This classification corresponds to Halliday and Hasan's category of the same name. These connections reflect paraphrases and semantic overlap in words from neighboring sentences. Our category also includes some of the ties that Halliday and Hasan classified as substitution. These are specifically nominal and clausal

substitution (substituting "competition" for "playing against people who are equal to or better than you," to borrow an example from an eighth grade writer). Substitutions of this kind also reflect semantic overlap, and so were not sufficiently different, by our view of writer development, to warrant a separate category.

Event-implied. This category includes the ties that Halliday and Hasan would have classified as ellipsis, since the "events" in the preceding discourse imply what has been deleted and may be understood. Implied repetitions of words fall into this category. Also in this category are connections that do not require explicit specification because of the reader's knowledge of real-world events. For instance, when one participant in a dialogue is quoted, the response of the other is expected and thus connected.

These conversational "rejoinders" (as Halliday and Hasan refer to them) are special instances of a broad class of connections that function by fulfilling the reader's expectations at a given point in the discourse. Upon discovering that a character in a story has been hit by a car, for example, the reader expects to then read the character's response or some elaboration of the situation.

This kind of connection does not rely on word or concept repetition. Rather it relies on fulfilling some of the reader's expectations created by the linguistic emphasis of the preceding sentence (Chafe, 1973) or by the nature of the event sequence itself (Clark, 1977; Winograd, 1977). Event-implied connections, then, are based at least as much on real-world knowledge of events as on linguistic knowledge.

Conjunction. While Halliday and Hasan used their conjunction category to specify any sort of grammatical coordination of the meaning relations in adjacent sentences, we restrict our conjunction category to single word conjunctions specifying additive ("also"), adversative ("yet," "but"), and causal ("so") relations. We did so because we hoped to distinguish single-word logical connections from more elaborate syntactic devices.

Because "and" seemed to function often, especially in younger children's writing, as a filler (the written counterpart of the spoken "uhm"), it was not counted as a legitimate conjunction connection for any grade level. This prevented us from counting some appropriate uses by older writers. For consistency's sake, however, "and" as a coordinate conjunction between independent clauses was never classified as a connection.

Complex Syntactic. Because of the likely development of complex syntactic devices, we tried to isolate those instances in which the writer chose special syntactic markings to express connections. With a subordinate clause or a compound predicate, for example, the writer can guide more exactly the reader's understanding of the relations expressed in adjacent sentences. These sorts of connections depend on the writer's linguistic ability and can reveal developing syntactic sophistication. Because of this, we have maintained them as a separate category so as to observe the use of special syntactic connections in writers of various ages.

Table 2 shows how local connections were distributed across our six categories of connections--reference, comparison, lexical ties, conjunction, event-implied, and complex syntactic.

Table 2

Types of Local Connections
Percentage of each Type within each Text Form

	EXPOSITORY				NARRATIVE			
GRADE	2	4	6	8	2	4	6	8
Type of Local Connection								
1. Reference	56	31	17	23	41	22	30	23
2. Comparative	0	18	18	11	9	12	9	10
3. Lexical ties	31	25	30	21	20	18	22	20
4. Conjunction	0	11	9	19	9	24	9	10
5. Event-driven	0	9	7	3	14	17	11	13
6. Complex Syntactic	13	6	19	22	6	7	19	24

Since Table 2 reflects percentages conditional on the writer's use of a local connection, we see differences in the relative frequency of connections of a given type, not overall differences in frequency. Of particular interest is the comparison between grades four and six in the narrative text form. Fourth grade writers were using relatively large percentages of comparisons and conjunctions in their narratives, 12% and 24% respectively. These were generally single-word connections such as "then" (comparison) or "so" and "but" (conjunction). This is not surprising considering the time progression inherent in the narrative form. Sixth graders, however, made less use of single-word conjunction

and comparison connections, 9%, and more use of complex syntactic connections, 19% compared with 7% in the fourth grade narratives. Thus the sixth grade writers were beginning to use subordination to connect ideas across sentences, while the fourth grade writers depended more on single-word connections.

Comparing the relative frequency of these categories of connections across expository and narrative texts, one again sees that text forms influence connectedness. This is especially true in the writing of the younger children.

One may be skeptical of the percentages for the second graders within the expository since they are based on only seven essays.[4] However, local connections in the essays of second graders seemed to come primarily from ties of reference or other semantic overlap in the words in the text. Connections of this sort depend on the reader recognizing the semantic similarities and making the appropriate connections, without explicit direction from the writer. (This is not to say that some lexical ties may not be very sophisticated.) Within their narratives, however, second grade writers were beginning to direct the reader's understanding by including specific conjunctions and comparison connections. There was also an increase in event-implied connections. The facilitating effect of the narrative form is perhaps due to the match it provides with second graders' knowledge of temporal event structure.

In contrast with the second graders, fourth grade writers were using in their essays all of the connections we described. However, they too decreased their use of reference and lexical ties when they worked within the narrative text form. In their narratives, fourth

grade writers began to use more conjunctions to coordinate ideas, as well as the reader's knowledge of the world (in event-implied connections).

In the writing of sixth and especially that of eighth graders, one begins to see less variation in the distribution of connection types from the essays to the narratives. This is not very surprising. By sixth grade, children are reading increasingly more expository texts, and so the expository text form is more familiar to the older children than the younger.

Just as important as their reading experience, however, is their writing experience. By the sixth grade, children seemed to have developed a way to attack a writing problem. They have a repertory of devices to connect their ideas, and they use them in a similar way whether writing an essay or narrative. There remains room for improvement, as indicated by the increased connectedness in the eighth grade narratives. However, these writers showed a constancy in their understanding of the constraints imposed by the two writing tasks and in the way they expressed the underlying topic coherence in their written discourse.

It is difficult, and conceivably unwarranted, to sharply separate this writing skill from other cognitive skills which are developing during this time. Writing skill is certainly related to developing reading and oral skills. In addition to these other language skills, writing must also be influenced by more general cognitive skills. In this analysis of writing we have observed the writer's ability to construct a coherent text, an ability that implies at least some awareness of the text demands on the reader. Taking the reader into

account is a skill not very far removed from perspective-taking, which itself shows striking developmental differences (Piaget, 1926; Flavell, 1977). Bereiter and Scardamalia (1981) point out how critical perspective-taking is in the child's transition from conversation to composition, since the reader, unlike the conversational partner, is unable to ask for immediate clarification. Also, strategies for attacking a writing problem are probably related to other developing metacognitive skills, such as the use of plans (Flavell, 1977) and other strategies (Flavell and Wellman, 1977). In writing, we may be seeing the effects of several developing cognitive skills.

A Computer Model of Development of Writing Procedures

To this point we have been focusing on the written product, with little speculation about the processes that produce it. From these descriptive studies of children's writing we have developed some processing descriptions of writing by children of various ages and how text forms influence these processes.

We have modeled these writing processes in a computer simulation based on a version of Anderson's (1976) ACT system, the ACTP production system used by Riley and Greeno (1980). From models of specific subject performance we have constructed a general developmental model with two basic assumptions. First, the model assumes that there is a text-relevant memory network underlying the written discourse and, second, that the information expressed in the discourse corresponds to memory nodes and connections in that network.

The text-relevant memory network is essentially the knowledge upon which the writer bases the discourse. We do not assume that this memory network corresponds to memory of general knowledge, since children often seem to know more that they write in their texts. But by specifying the text-relevant memory network and by holding it constant as we vary the procedures which act on it, we can emphasize the procedural differences we see in the writing behavior of children of various ages. Three different sets of procedures--one modeling the second grade writer, one the fourth, and one the sixth and eighth grade writers--can operate on the same text-relevant memory network, and each will produce a text typical of writers at that grade level. With slight modifications, the system also simulates the influence of text forms.

One may argue that it is the larger knowledge base of the older children, not their procedures, that produces the observed writing differences. We recognize that increased topic knowledge can improve writing. However, we would like to minimize the emphasis on topic knowledge in order to explore the potential of a procedural explanation for the development of local coherence.[5] In the studies described, the effect of knowledge differences was reduced by having the children choose their own topics. (And children did seem to choose activities with which they could be expected to have some experience. There were, for example, no essays on hang gliding or deep sea diving.) Also, the improvement we saw in especially the younger children's narratives over their essays, despite the similarity of topic, suggested that the children had adequate topic knowledge. For example, while younger children failed to mention any dangers in their essays on swimming, children of the same age had no trouble supplying a complication for their narratives: characters hit their heads or scraped their legs or

got splashed. The inclusion of this information in the narratives suggests that the children knew these dangers.

We argue that the younger writers failed to recognize the relevance of that knowledge in their essays because they were not sensitive to the local coherence demands of the final sentence. In the narrative, both the presupposition of dangers in the final sentence and the child's "schema knowledge" of story complications prompted the children to use procedures to access knowledge about dangers. However, a complication is not a necessary part of the schema knowledge of an essay, even if we were to assume (probably incorrectly) that our young writers had such essay schemas. So in the essays, the only clue to the required mention of dangers came in the presupposition of the final sentence, and it could only be recognized if the writer was trying to make the final sentence fit coherently with the text before it. Sensitivity to local coherence had to prompt the appropriate memory scan procedures, without additional direction from a relevant schema.

The importance of procedural differences can be seen in a comparison of the two subject samples schematized in Figures 2 and 3. Notice the similarity in the gist of the details describing the fun of swimming. Differences in the sentence structure, however, reflect differences in the writers awareness of local coherence. These two texts illustrate how similar the knowledge base can be and how different the expression.

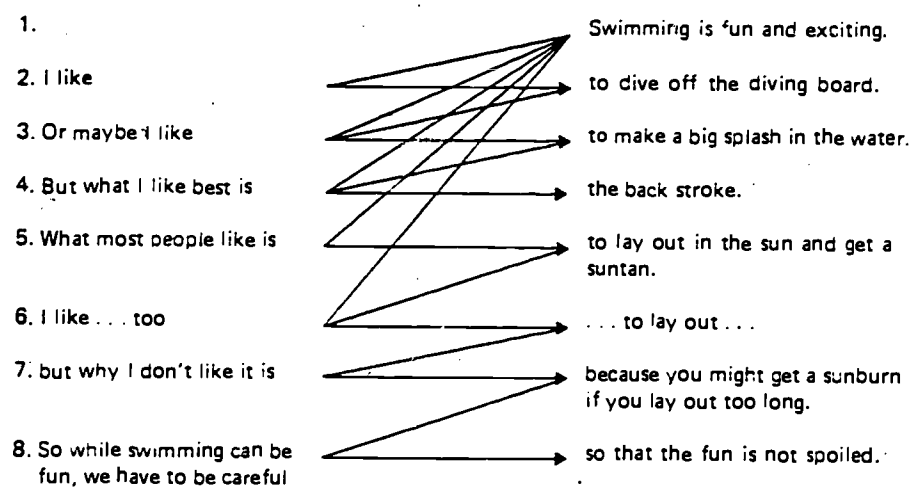

Figure 3. Essentially same information as that represented in Figure 2, but showing a zig-zag structure. Written by a sixth grade subject.

As we saw in our previous analysis of the fourth grade essay in Figure 2, this essay shows a typical list-like structure. Each supporting detail was linked only to the topic sentence. Many of the same supporting details were used by the sixth grader who wrote the essay in Figure 3, but this writer supplied connections between the details with adversative conjunctions ("or" and "but") and clausal subordination. In the following model, we attempt to describe memory access procedures that can account for these difference in local coherence. While a complete model should eventually specify the interaction between topic knowledge and procedures, we will focus this discussion on a procedural description of developmental differences.

Data base

The text-relevant memory network presented in Figure 4 will be used to illustrate how the same knowledge base can produce very different texts when different procedures are used to scan and retrieve information from it. We will first discuss the aspects of the model that apply to the production of expository text and then describe the modifications due to the imposition of the narrative text form.

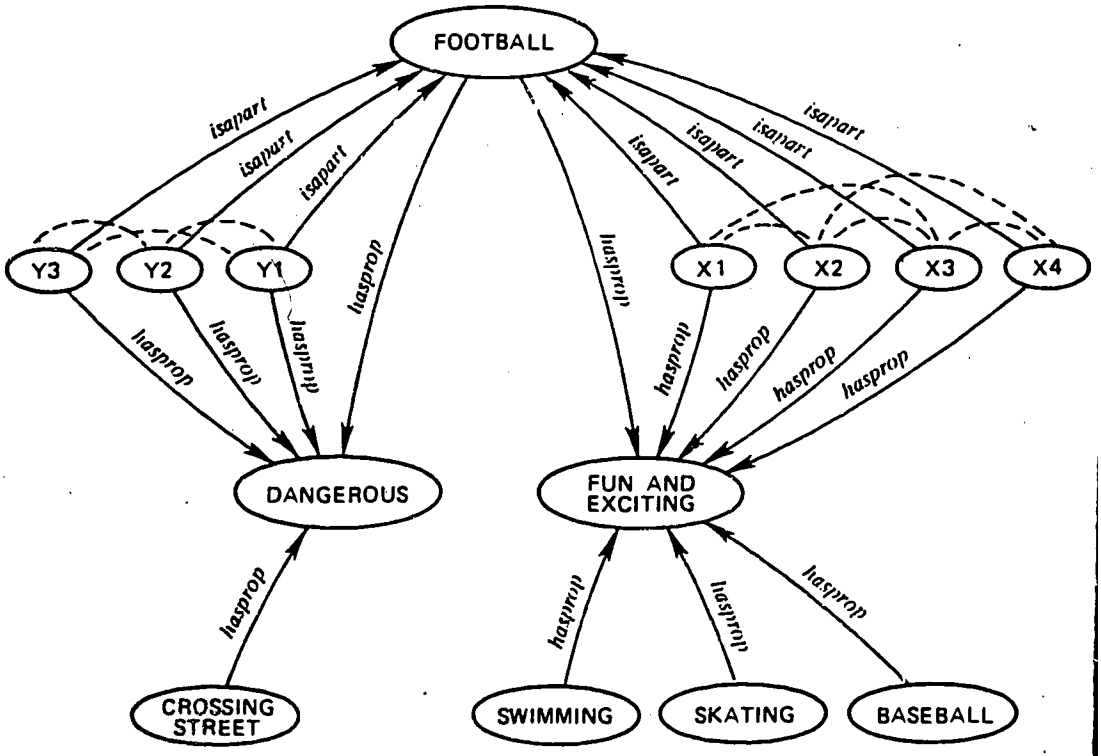

Figure 4. Schematization of text-relevant memory network used in simulation model.

The model's second assumption is that relations expressed in the written discourse are part of the text-relevant memory network. As Figure 4 illustrates, the concept node FOOTBALL is connected both to the FUN AND EXCITING node and to the DANGEROUS node by HASPROP links. This represents the knowledge that football is fun but dangerous as well.

The reasons football is considered fun are represented by memory nodes X1, X2, X3, and X4. These are, in some sense, components of football, thus ISAPART links extend from these nodes to the concept node FOOTBALL. The components are themselves fun, and HASPROP links connect them to the FUN AND EXCITING node. These component nodes represent the supporting details that the writer will use to describe the fun of football.

There are also components of football that make it dangerous, as illustrated by the HASPROP links from nodes Y1, Y2, and Y3 to DANGEROUS. These components also have ISAPART links to FOOTBALL. Thus these are the details supporting the assertion that football is dangerous.

Component nodes may also have links among themselves, as represented by the dashed lines. While not explicitly labeled in Figure 4, these links may specify connections such as temporal, causal, or contrastive relations. These dashed lines, then, represent the knowledge that the writer uses to connect one supporting detail to the next within the written discourse.

The writer also knows that other activities can be fun or dangerous, and this knowledge is represented in the network by the other concept nodes with HASPROP links to FUN AND EXCITING and DANGEROUS. These other concept nodes may also have component nodes attached to

them, but these are irrelevant to our discussion and so are not represented in Figure 4.

Developmental differences in expository procedures

As previously described, there were developmental differences in writers' ability to honor topic constraints imposed by our tasks. These differences can be captured in the type of pattern the model's procedural system attempts to match in the memory network and the way the retrieved information is handled.

The second grade writers ignored the single topic and danger constraints and simply wrote lists of activities they thought were fun. The system models their performance by searching memory for activities with HASPROP links to the FUN AND EXCITING node. When this pattern is found, a sentence expressing this relationship is produced and the search begun again. No supporting details are examined for any activity, and the DANGEROUS node is never considered. With the network in Figure 4, this procedural system produces one sentence about football being fun, one about skating, one about swimming, and one about baseball.

Fourth grade writers were able to handle the single-topic and fun constraints in their essays, so the model of their writing performance searches memory for a single activity node that is connected to FUN AND EXCITING by a HASPROP link. The selected activity also must have some predetermined number of component nodes attached to it and to FUN AND EXCITING (to justify its choice as a topic). After a topic activity is chosen, the component nodes are scanned, and as each is found, it is expressed in a sentence, with no examination or mention of any links from it to another component node. Using the network in Figure 4, this

system thus produces a list-like essay about football, with each detail (component nodes X1, X2, X3, and X4) connected only to the topic sentence.

Supporting details are connected, however, in the model of the sixth and eighth grade writers. (These two grade levels performed similarly, so one procedural system models both.) Since these writers were able to handle the danger constraint, as well as fun and single topic, the model of their performance searches memory for an activity node having HASPROP links to both FUN AND EXCITING and to DANGEROUS. Again the activity must have enough component nodes to make it a justifiable topic choice. Links among the component nodes are then examined. The relations expressed in these links are used to order the supporting details in a planning stage and to build connections among the sentences expressing those details in the written discourse.

In this sense, creating local connections is very similar to writing with a goal in mind. The general planning stage just described allows the writer to know what idea is to follow the current one, as well as what precedes it. However, the system might also be built to run sequentially--finding and expressing one detail and only then examining connections from it to another detail. This procedure, with its more narrow planning, may allow the writer to find supporting details overlooked in the initial memory scan. It remains an empirical question whether the more valid description of mature writing is the former or the latter, or a combination of the two.

There is some research, however, supporting these procedural descriptions. Specifically, writing behavior changes between grades four and six. Planning begins to show itself as a stage separate from actual composing only in children from grade six and up. For younger children, there was very little difference between the expression of information when it was first accessed (during an experimenter-directed planning stage) and when it appeared in the final composition (Burtis et al, 1981). In addition, prewriting tasks designed to activate the appropriate semantic network resulted in longer and more elaborated arguments by sixth grade children (Anderson, Bereiter, and Smart, 1980).

Procedural differences in narrative texts

The systems just described must be modified to model performance in the narratives. The narrative text form provides certain information that is incorporated into the pattern matching routines and influences sentence generation.

Since the narrative text form implies a main character (and our initial sentence provided one), even second graders generally kept Bobby as a topic throughout their narratives. Thus the narrative text form required that Bobby be mentioned often and this repetition of Bobby supplied the list-like structure we saw in the second grade narratives. Even many second graders recognized the danger constraint, and, again, we suspect this is due to the complication inherent in the familiar structure of a narrative. The narrative text form builds into the pattern matching routine a search for dangerous aspects of Bobby's activity, since these would be the source of the complication.

The narrative text form provided further cues to the fourth grade writers, enabling them to make local connections in their narratives while they had not done so in their expository texts. Implicit in the structure of a narrative is temporal progression--one event follows another in time. Thus the writer is cued to look for temporal relations among the details of Bobby's activity and, having found them, to express those relations in the text. This accounts for the fourth graders' increased use of "then" and "so" as local connections in their narratives. The pattern matching routine of the fourth grade model explicitly searches for temporal connections among details in the memory network when working within the narrative text form, while it ignores connections in the expository.

The procedural model of the sixth and eighth grade narrative writing remains much the same as that of their expository writing, with the exception of the introduction of the main character Bobby. This model continues to examine all sorts of connections among component nodes—causal and contrastive, as well as temporal. Thus the model of sixth and eighth grade writers has the semantic information to use in "For that reason, ..." and "in spite of ..." subordinate clauses, constructions increasingly used by sixth and eighth grade writers. Thus, the information in the connecting links among supporting details is the semantic substance of the local connections in the written discourse of these writers.

Summary

We have briefly sketched some important aspects of a developmental model of coherence in writing and have argued here that discourse coherence can provide a window on the writing processes which produce it. We have described developmental differences we observed in children's writing, and we proposed a procedural model to account for those differences.

Our model of coherence emphasized four aspects important for the development of writing skill, and there may be others. Those emphasized here were knowledge of topic, knowledge of text form constraints, topic coherence, and local connectedness.

Our descriptive studies revealed that children show a sensitivity to the demand for local connectedness within their written discourse and that this sensitivity increases from the second through the eighth grade. Sensitivity to local connectedness was also related to sensitivity to the topic constraints imposed by the tasks. Children who did not make local connections among their sentences also often failed to recognize the additional constraint in the final sentence we provided. Although the developmental sequence did not change, the text form within which the children wrote, expository or narrative, also influenced their writing. Younger children were able to write more coherently within the narrative form, while older children wrote similarly whether in the narrative or expository text form.

We modeled these developmental differences in a computer simulation. The model emphasized differences in the procedures used to scan and retrieve information from memory. Modifications of the

scanning procedures simulated developmental changes as well as the influence of text forms on the writing process.

Thus the computer simulation emphasized two of the aspects important for discourse coherence--local connectedness and text form constraints--although actual sentence generation was unspecified. A third aspect of coherence, topic knowledge, was explicitly held constant. The fourth, topic coherence, may be related to sensitivity to topic constraints, but our task did not permit strong statements about that relationship. The procedural aspects of the computer model were intended to simulate possible strategies used by the writer to retrieve and organize information from memory, in addition to specifying which part of the text-relevant memory network was the writer's focus.

How closely the simulated strategies correspond to those actually used by writers remains to be seen. In the studies described here, we used the written product to infer the processes which produced it. The accuracy of our inferences may be tested using other, more process-oriented methodologies. Protocol analysis, as used by Hayes and Flower (1980), has yielded useful information about general subprocesses of writing. It may also prove useful in determining what topic constraints are recognized by children in a given task (and when). Similarly, we may also see planning differences during production of remote and local connections. Analyses of process, in addition to product, are necessary for more complete understanding. However, just as text analysis has improved our understanding of comprehension, it can provide useful insights into writing.

Notes

1. This is primarily a simplifying assumption for purposes of our analysis. One could imagine instead a broader analysis which assumes Given-New relative to writer and reader (through the writer's use of deictic terms and such). Thus, individual phrases could have Given and New elements, and a sentence could contain many such Given-New patterns. However, to make our scoring scheme more manageable, we chose to treat sentences as our units and to focus on the ideas the writer included in the text. While we do consider how the writer guides the reader through those ideas, we do not treat the writer-reader relation as a basis of Given and New.

2. One can imagine a coherent reading of sentence 3, with contrastive stress placed on "can." While a speaker can supply contrastive stress and thus be assured of the listener's interpretation, the writer cannot control interpretation without an explicit contrastive marker such as "but" or "however." This is one of the interesting differences between writing and speaking.

3. The percentages of local connections are somewhat inflated in this table. This is especially true in the very short texts typical of the younger writers. The first sentence the child writes, if it is connected at all, cannot contain a remote connection because it is only the second sentence of the text. We did not exclude these necessarily local connections from this analysis because they were informative in a later analysis of the nature of the connections.

4. We collected 28 second grade expository samples, but only seven resembled essays. Most second grade children wrote lists of activities they enjoyed, and such lists of loosely connected activities could not be expected to show the same coherence as a discussion of one activity. Thus, only the second grade samples that honored the single topic constraint (only those that were in fact essays) were used in the analysis of connectedness.

5. While it seems intuitive that increased knowledge improves writing, Scardamalia, Bereiter, and Woodruff (1980) found no quality differences in children's compositions when they were writing about a familiar topic compared with an unfamiliar one. These findings further suggest that factors other than knowledge can play an important role in writing.

References

Anderson, J. R. (1976). Language, Memory and Thought. Hillsdale, N. J.: Erlbaum.

Anderson, V., Bereiter, C., and Smart, D. Activation of semantic networks in writing: Teaching students how to do it themselves. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Bereiter, C., and Scardamalia, M. (1981). From conversation to composition: The role of instruction in a developmental process. In Advances in Instructional Psychology (Vol. 2), R. Glaser (Ed.), 1-64. Hillsdale, N. J.: Erlbaum.

Burtis, P. J., Bereiter, C., Scardamalia, M., and Tetroe, J. The development of planning in writing. In C. G. Wells & B. Kroll (Eds.), Exploration of children's development in writing. Chicester, England: Wiley, in press.

Chafe, W. (1973). Language and memory. Language, 49, 261-281.

Clark, H. H. (1977). Inferences in comprehension. In Basic Processing in Reading: Perception and Comprehension, D. LaBerge and S. J. Samuels (Eds.), 243-263. Hillsdale, N. J.: Erlbaum.

Clark, H. H., and Haviland, S. E. (1977). Comprehension and the given-new contract. In Discourse Production and Comprehension, R. C. Freedle (Ed.), 1-40. Norwood, N.J.: Ablex.

Flavell, J. H. (1977). Cognitive Development. Englewood Cliffs, N. J.: Prentice Hall.

Flavell, J. H., and Wellman, H. M. (1977). Metamemory. In Perspectives on the Development of Memory and Cognition, R. V. Kail and J. W. Hagen (Eds.), 1-33. New York: Erlbaum.

Halliday, M. A. K., and Hasan, R. (1976). Cohesion in English. London: Longman.

Hayes, J. R., and Flower, L. S. (1980). Identifying the organization of writing processes. In Cognitive Processes in Writing, L. W. Gregg and E. R. Steinberg (Eds.), 3-30. Hillsdale, N. J.: Erlbaum.

Hidi, S. How children cope with constraints in writing coherent sentence sequences. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Piaget, J. H. (1926). The Language and Thought of the Child. New York: Harcourt, Brace.

Scardamalia, M., Bereiter, C., & Woodruff, E. The effects of content knowledge on writing. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Riley, M. S., and Greeno, J. G. Details of programming a model of children's counting in ACTP. (Tech. Rep. 1980/6). Pittsburgh, Pa.: University of Pittsburgh, Learning Research and Development Center, 1980.

van Dijk, T. A. (1976). Philosophy of action and theory of narrative. Poetics, 5, 287-338.

van Dijk, T. A. (1977a). Semantic macro-structure and knowledge frames in discourse. In Cognitive Processes in Comprehension, M. A. Just and P. A. Carpenter (Eds.), 3-32. New York: Wiley.

van Dijk, T. A. (1977b). Text and Context. Explorations in the Semantics and Pragmatics of Discourse. London: Longman.

Winograd, T. (1977). A framework for understanding discourse. In Cognitive Processes in Comprehension, M. A. Just and P. A. Carpenter (Eds.), 63-88. New York: Wiley.