Toward an Interactive Model of Reading

David E. Rumelhart, Stanford University*

Reading is the process of understanding written language. It begins with a flutter of patterns on the retina and ends (when successful) with a definite idea about the author's intended message. Thus, reading is at once a "perceptual" and a "cognitive" process. It is a process that bridges and blurs these two traditional distinctions. Moreover, a skilled reader must be able to make use of sensory, syntactic, semantic, and pragmatic information to accomplish his task. These various sources of information appear to interact in many complex ways during the process of reading. A theorist faced with the task of accounting for reading must devise a formalism rich enough to represent all of these different kinds of information and their interactions.

The study of reading was a central concern of early psychologists (see Huey, 1908). Now, after years of dormancy, reading has again become a central concern for many psychologists. It would seem that the advent of the information-processing approach to psychology has given both experimentalists and theorists paradigms within which to study the reading process. The formalisms of information processing, the flowcharts, notions of information flow, and so forth have served as useful vehicles for the development of first approximation models of the reading process. Unfortunately, the most familiar information-processing formalisms apply most naturally to models assuming a series of noninteracting stages of processing or (at best) a set of independent parallel processing units. There are many results in the reading literature that appear to call for highly interactive parallel processing units. It is my suspicion that the serial, noninteracting models have been developed not so much because of an abiding belief that interactions do not take place, but rather because the appropriate formalisms have not been available. It is the purpose of this chapter to adapt a formalism developed in the context of parallel computation to the specification of a model for reading and then show that such a model can account in a convenient way for those aspects of reading that appear puzzling in the context of more linear stage-oriented models. No claim is made about the adequacy of the particular model developed. The primary claim is that this richer formalism will allow for the specification of more detailed models. These will be able to characterize aspects of the reading process that are difficult or impossible to characterize within the more familiar information-processing formulations.

First, I will review two recent models of the reading process. Then, I will discuss some of the empirical evidence that is not conveniently accounted for

This chapter is reprinted from *Theoretical Models and Processes of Reading* (3rd ed., pp. 722–750), edited by H. Singer & R.B. Ruddell, Newark, DE: International Reading Association.

by these models or their natural extensions. Finally, I will develop a reading model that makes use of a formalism allowing highly interactive parallel processing units and then show that this model offers a reasonable account of the Problematic Results section.

Current Models of Reading

Gough's Model

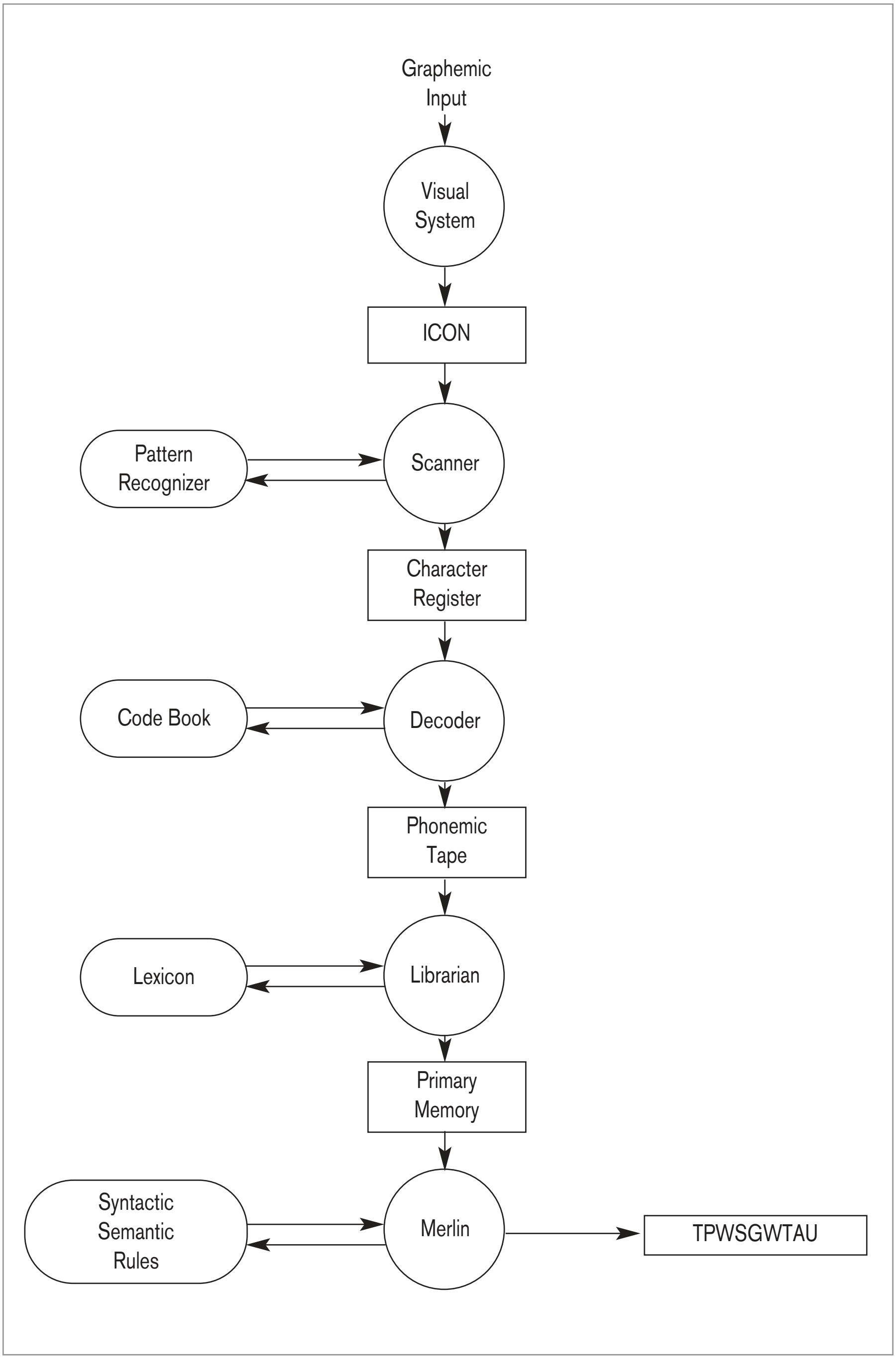
Gough (1972) has proposed a model of reading that is remarkable in the degree to which it attempts to give a complete information-processing account of the reading process. Gough attempts to pin down as completely as possible the events that occur during the first second of reading. A schematic diagram representing the flow of information during the reading process is shown in Figure 1. According to Gough's model, graphemic information enters the visual system and is registered in an *icon*, which holds it briefly while it is scanned and operated on by a *pattern-recognition* device. This device identifies the letters of the input string. These letters are then read into a *character register*, which holds them while a *decoder* (with the aid of a *code book*) converts the character strings into their underlying *phonemic* representation. The phonemic representation of the original character strings serves as input to a *librarian*, which matches up these phonemic strings against the *lexicon* and feeds the resulting lexical entries into *primary memory*. The four or five lexical items held in primary memory at any one time serve as input to a magical system (dubbed *Merlin*), which somehow applies its knowledge of the syntax and semantics to determine the deep structure (or perhaps the meaning?) of the input. This deep structure is then forwarded to its final memory register TPWSGWTAU (the place where sentences go when they are understood). When all inputs of the text have found their final resting place in TPWSGWTAU, the text has been read and the reading is complete.

I do not want to discuss the merits or demerits of Gough's particular model at this point. Instead, I point to the general form of the model. For Gough, reading consists of a sequentially ordered set of transformations. The input signal is first registered in the icon and then transformed from a character-level representation to phonemic representation, then lexical-level representation, and finally to deep structural representation. Thus, the input is sequentially transformed from low-level sensory information into ever higher-level encodings. Note, however, that the information flow is totally "bottom up." That is, the information is initiated with the sensory signal, and no higher level of processing can affect any lower level. The reading process is strict letter-by-letter, word-by-word analysis of the input string. There is no provision for interaction within the system. The processing at any level can directly affect only the immediately higher level.

LaBerge–Samuels Model

In another paper, LaBerge and Samuels (1974) have developed an equally detailed (although somewhat more perceptually oriented) model of the reading process.

Figure 1. Gough's Reading Model

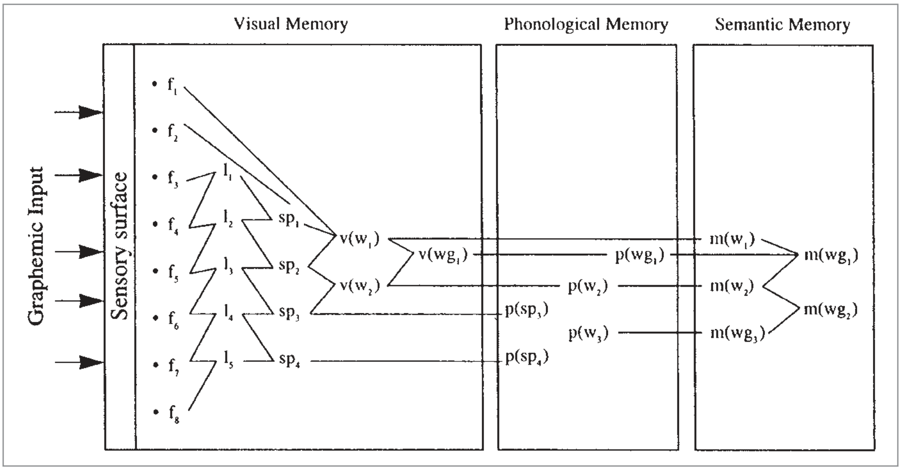


After Gough, 1972.

Figure 2 gives a schematic representation of their model. The basic model consists of three memory systems holding three different representations of the input string. The Visual Memory System holds visually based representations of the features, letters, spelling groups, words, and word clusters. The Phonological Memory System holds phonological representations of spelling groups, words, and word groups. Finally, the Semantic Memory System holds the semantic representation of the words, word groups, and sentences that are read. The reading process begins with the registration of the visual signal on the sensory surface. The information is then analyzed by a set of specialized *feature detectors* that extract information about lines, angles, intersections, and so on, from the physical stimulus. Most of these feature detectors, f_1, feed directly into letter codes, l_1. Thus, the activation of letter codes results naturally from the convergence of a set of feature detectors. These letter codes feed into spelling-pattern codes, sp_1, which in turn feed into visual word codes, $v(w_1)$. Some features (e.g., f_2) map directly into spelling pattern codes and others (f_1) directly into visual word codes. Such features are sensitive to the overall configuration of the words and spelling patterns. There are a number of routes whereby words can be mapped into meanings.

1. Visual word codes can feed directly into word-meaning codes, $m(w_1)$. This route would be necessary for the discrimination of such homophonous word pairs as *pear* and *pair* or *chute* and *shoot*.
2. The visual word codes can pass through a phonological word code, $p(w_1)$, and then into a word-meaning code. This is perhaps the ordinary route of analysis within the LaBerge–Samuels model.
3. The model also allows for word groups, such as *time out*, to be analyzed into visual word-group codes, $v(wg_1)$, from these into phonetic word-group codes, $p(wg_1)$, and finally into group meanings, $m(wg_1)$.

Figure 2. Reading in the LaBerge–Samuels Model



After LaBerge and Samuels, 1974.

4. When a word has not been learned as a visual stimulus, information can be translated directly from visual spelling patterns into phonological spelling patterns, $p(sp_1)$, from these into phonological word codes, and finally into word-meaning codes. In addition, word-meaning codes feed into word-group–meaning codes.

Ultimately, when the entire set of inputs has been presented, a set of word-group meanings will emerge and the reader will be said to have understood the input.[1]

Again, I do not want to discuss the particular merits or demerits of the LaBerge–Samuels model. Rather, I again point out the general form of the model and suggest that it takes that form, at least to some extent, because of the formalisms used to represent the ideas. The LaBerge–Samuels model, like the Gough model, is a strictly bottom-up process.[2] Although there are alternative routes, the basic sequence is from features to letters, to spelling patterns, to visual word representations, to phonological word representations, to word meanings, to word-group meanings—a series of stages, each corresponding to a level of analysis in which no higher level can in any way modify or change the analysis at a lower level. The LaBerge–Samuels model (unlike the Gough model) does allow certain stages to be bypassed. This allows multiple paths of analysis and alleviates some of the empirical problems of the Gough model. Nevertheless, there are a number of results in the literature that are difficult to account for with either model. I turn now to a discussion of a number of these problems.

Problematic Results

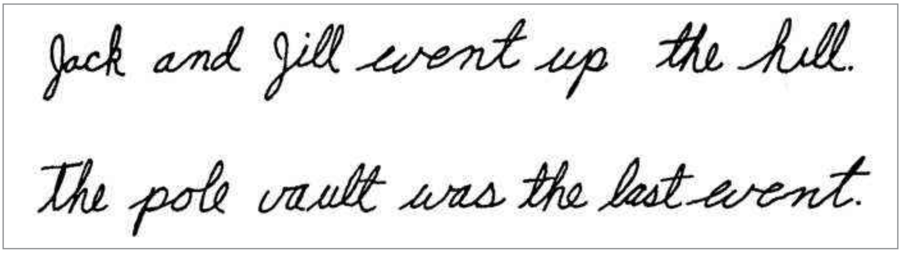
All of the results discussed in this section have one characteristic in common. In each case it appears that the apprehension of information at one level of analysis is partially determined by higher levels of analysis. By and large, such results are very difficult to incorporate in a processing model that assumes that information flows strictly from lower to higher levels. I will begin with a discussion of the effects of orthographic structure on the perception of letters, proceed to a discussion of the effects of syntax on word perception, then to the effects of semantics on word and syntax perception, and finally to the effects of general pragmatic factors on the perception of meanings.[3]

The Perceptions of Letters Often Depend on the Surrounding Letters

The literature on reading abounds with evidence on this point. Perhaps the most difficult of these results for a purely bottom-up model to account for are the well-known context effects illustrated in Figure 3 (after Nash-Weber, 1975). Here we see an ambiguous symbol, , which is interpreted as a *w* in one context and interpreted as an *e* followed by a *v* in another context. It would appear that our interpretation of the sentence has determined our perception of the ambiguous symbol.

The problem with results such as these stems from the fact that we appear to have “word-level” or “phrase-level” perceptions determining our perceptions at

Figure 3. The Dependence of Letter Perception on Context



After Nash-Weber, 1975.

the letter level, a higher-level perception affecting a lower-level one. These results can be accounted for by bottom-up models, but only at some cost. No final decision can be made at the letter level. Either a set of alternative possibilities must be passed on, or the direct feature information must be sent to the higher levels. In either of these cases, the notion that letter perception precedes word perception becomes suspect. Word and letter perception occur simultaneously.

Perhaps the strongest objection to a demonstration such as this one is that it is unusual to find such ambiguous letters and that the norm involves characters that are perfectly discriminable. Although this may be true of printed text, it is not true of handwriting. Characters often can be interpreted only with reference to their context. Yet I would not want to argue that the reading process is essentially different for handwritten than for printed material.

There are many other results that appear to call for this same conclusion. For example, more letters can be apprehended per unit time when a word is presented than when a string of unrelated letters is presented (Huey, 1908/1968). A letter string formed either by deleting a letter of a word or replacing one or two of the letters of the word is often clearly perceived as the original word (Pillsbury, 1897). Even when great care is taken to control for guessing, a letter is more accurately perceived when it is part of a word than when it is among a set of unrelated letters (Reicher, 1969). All of these results appear to argue strongly that letter perceptions are facilitated by being in words. Word-level perceptions affect letter-level perceptions. Here again, the only way that the types of models under consideration can account for these effects is to suppose that partial letter information is somehow preserved and the additional constraints of the word level are brought to bear on the partial letter information.

It is of some interest that these effects can be observed in letter strings that are not words but that are similar to words in important ways. For example, the more the sequential transition probabilities among letters in a string approximate those of English, the more letters can be perceived per unit time (Miller, Bruner, & Postman, 1954). Similarly, even when guessing is controlled (as in the Reicher, 1969, experiment), letters embedded in orthographically regular strings are more accurately perceived than those embedded among orthographically irregular strings (McClelland & Johnston, 1977). Thus, not only is a letter embedded in a word easier to see, but also merely being a part of an orthographically well-formed string aids perception virtually as much. This suggests that orthographic knowledge plays a role nearly as strong as lexical knowledge in the perception of letter strings.

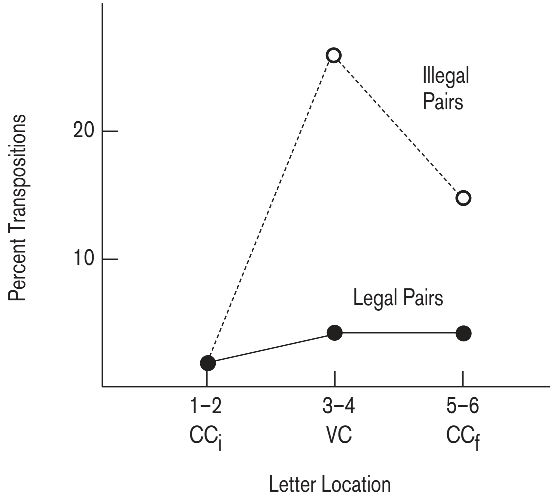
Not only does orthographic structure have a positive effect on the perception of letters embedded in an orthographically regular string, but also our apprehension of orthographically irregular strings often is distorted to allow us to perceive the string as being orthographically regular. This point is nicely illustrated in a recent experiment carried out in our laboratory by Albert Stevens. In this experiment, subjects were presented with letter strings consisting of two consonants (i.e., an initial consonant cluster designated CC_i) followed by two vowels (a vowel cluster, designated VC) followed by two more consonants (a final consonant cluster, CC_f). The initial consonant cluster was constructed from pairs of consonants that can occur at the beginning of English words in only one order (e.g., English words can begin with *pr* but not *rp*). Similarly the vowel clusters used occur as diphthongs in English in one order but not in the other (e.g., *ai* but not *ia*). The final consonant clusters were similarly chosen so that they occur at the end of English words in one order but not the other (e.g., *ck* but not *kc*). Strings were then constructed in which each letter cluster was either in its legal or illegal order. Table 1 illustrates several examples of the various types of letter strings.

Subjects were given tachistoscopic presentations of the various letter strings and asked to name the letters they observed. Of particular interest are the times when they were presented illegal strings but made them legal by transposing the letter pair in their reports. Figure 4 illustrates the comparison of interest. The figure compares the percentage of times an illegally ordered letter cluster is transposed into a legal cluster with the number of times a legal letter cluster is transposed into an illegal one. The results show that although initial consonant clusters are never transposed, illegal vowel clusters are transposed almost 25% of the time as compared to only about 3% transposition for the legal vowel clusters. Similarly, final consonant clusters are transposed almost 14% of the time when they are illegal, but only about 3% of the time when they are legal. These results show clearly the effect that orthographic structure has on our perception of letter strings. The perception of a certain letter in a certain position depends on what we perceive in adjacent

Table 1. Examples of Legal and Illegal Letter Strings

	CC_i			
	Legal VC		**Illegal VC**	
CC_f	Legal	Illegal	Legal	Illegal
Legal	praick	priack	rpaick	rpiack
	stourt	stuort	tsourt	tsuort
Illegal	praikc	priakc	rpaikc	rpiakc
	stoutr	stuotr	tsoutr	tsuotr

Figure 4. Transpositions as a Function of Letter Location



positions as well as on the sensory evidence we have available about that position in the string.

To summarize, then, it appears that no model that supposes that we first perceive the letters in a stimulus and then put them together into higher-order units can be correct. However, models such as the Gough model and the LaBerge–Samuels model can survive such results if they assume that partial information is somehow forwarded to the higher levels of analysis and that the final decision as to which letters were present is delayed until this further processing has been accomplished.

Whereas it is not too difficult to see how, say, the LaBerge–Samuels model could account for the effects of orthographic structure on letter perception, it is somewhat more difficult to see how the effects of syntax and semantics can be mediated within such a model. I now turn to evidence for syntactic effects in reading.

Our Perception of Words Depends on the Syntactic Environment in Which We Encounter the Words

Perhaps the best evidence for syntactic effects on the level of word perception comes from an analysis of oral reading errors. The most common error in oral reading is the substitution error—when an incorrect word is simply substituted for the correct one. If syntax had no effect on word perception, we would expect that reading errors should be determined by visual similarity and not by part of speech. However, there is a strong tendency for a reading error to be of the same part of speech as the word for which it was substituted. Thus, for example, Kolers (1970) reported that nearly 70% of the substitution errors made by adult readers on geometrically transformed text were of the same part of speech as the correct word. By chance, one would expect only about 18% of the errors should be of the correct part of speech.

In another study, Weber (1970) analyzed reading errors by first graders and found that over 90% of the errors made were grammatically consistent with the sentence to the point of the error. Although it is not clear what percentage to expect under assumptions of random guessing, it is obviously much lower than 90% in most texts. One might argue that these results and those of Kolers occur because words in the same syntactic class are more similar to each other than they are to words outside that class. It is interesting to note in this regard that in the Weber study, the ungrammatical errors were significantly more similar to the correct word than were the grammatical words—at least an indication that this is a syntactic effect and not a visual one.

In another experiment, carried out by Stevens and Rumelhart (1975) with adult readers, an oral reading task showed that about 98% of the substitution errors that were recognizable as words were grammatical. Moreover, nearly 80% of the time the substituted words were of the same syntactic class as the class most frequently predicted at that part in a cloze experiment. Once again, it appears that we have a case of grammatical knowledge helping to determine the word read.

In addition, in an important experiment, Miller and Isard (1963) compared perceptibility of spoken words under conditions in which normal syntactic structure was violated with the case in which syntactic structure was intact. They found that many more words could be reported when the sentences were syntactically normal. Although I do not know of a similar study with written materials, it is doubtless that similar results would occur—another case of a higher level of processing determining the perceptibility of units at a lower level.

It is difficult to see exactly how the models under discussion would deal with results such as these. In the Gough model, syntactic processing occurs only very late in the processing sequence—after information has entered short-term memory. It seems unlikely that he would want to assume that partial information is preserved that far in the process. It is not clear just where syntax should be put in the LaBerge–Samuels model. It is particularly difficult to represent productive syntactic rules of the sort linguists suggest in the LaBerge–Samuels formalism. As I will discuss, it would appear to be essential to be able to represent systems of rules to account for such results.

Our Perception of Words Depends on the Semantic Environment in Which We Encounter the Words

It is even more difficult to incorporate a mechanism for semantic effects on the word-recognition process into a purely bottom-up model than it is to incorporate a mechanism for syntactic effects. There have recently been a number of studies that provide very nice demonstrations of semantic effects on word recognition.

In a series of experiments, Meyer, Schvaneveldt, and Ruddy (Meyer & Schvaneveldt, 1971; Meyer, Schvaneveldt, & Ruddy, 1972, 1974; Ruddy, Meyer, &

Schvaneveldt, 1973; Schvaneveldt & Meyer, 1973) have reported convincing evidence of semantic effects on word recognition. The basic procedure in these experiments involved measuring reaction times to come to a lexical decision about a pair of words. The basic result is that the decision can be made much faster when the pair of words are semantically related (such as *bread–butter* or *doctor–nurse*) than when they are unrelated (such as *bread–doctor* and *nurse–butter*). The most plausible account of these results would seem to be that the process of perceiving the first word somehow allows us to process the second word more quickly just in case it is a semantically related word. Thus, we again have the processing at the semantic level modifying our processing at the word level.

In a series of experiments recently carried out in our laboratory, Graboi (1974) demonstrated this same general effect using quite a different method. In one of his experiments Graboi employed a variation of Neisser's search procedure. First, subjects were trained to search for occurrence of any one of five target words among a list of semantically unrelated nontargets. Half of the subjects searched for any one of the words labeled Experimental Target Set (see Table 2) scattered among lists constructed from the Unassociated Nontargets. The other half of the subjects searched for the words labeled *Control Target Set* against the same background. Notice that neither target set is semantically related to the nontarget background in which it is searched for. After 14 hours of training, the experimental group was searching their lists at a rate of 182.2 milliseconds (msec)

Table 2. Alternative Stimuli in the Target and Nontarget Sets

	A. Associated Nontargets	
Experimental	WORM CHICK NESTS ROBIN CHIRP WINGS FLY	Control
Target Set	EAGLE PARROT SONG BLACK GRAY PURPLE	Target Set
	BROWN GOLD BLUE RED YELLOW GREEN PAINT	
	SAVE SPEND COINS DIME BANK SILVER DOLLAR	
	CASH PENNY PEARL BOOKS SCHOOL READ	
	CLASS WRITE TEACH EXAM NOTES GRADE	
	STUDY ORANGE NUTS GRAPE SWEET PLUM	
BIRD:	APPLE PEACH PEAR FRESH LEMON	:ROCK
COLOR:		:CHAIR
MONEY:	B. Unassociated Nontargets	:HOUSE
LEARN:	HUG PEN SLEEP NIGHT BRIDGE STAPLE LAMP	:SPORT
FRUIT:	RULER LEADER ROAR SUNNY PLACE CORNER	:CLOUD
	ALBUM ABOUT RATE WEEK POINT SWITCH	
	ANKLE TOWN DIAL SPOON TOWEL SHEET STOVE	
	CRUST BRUSH GLASS ROAD WHICH AFTER	
	PASS STORY SIGN CHURCH MURAL PHONE	
	BOOTH CARD STREET MOTOR RADIO KNOB	
	PLUG DRIVE LINE TASK PRINT SHIFT	

per word. The control group scanned at a rate of 180.0 msec per word. At the 15th hour of practice, the background lists were changed. Both groups now searched for their targets against the Associated Nontarget background. Now the experimental group was searching for its targets against a background of nontargets, all semantically associated with the target set. The control group also was switched, but the Associated Nontargets were not semantically related to the control target set. After the change, we found that the control group scanned against the new background at about the same rate as they scanned the old one—179 msec per word scanned. The experimental group, however, scanning through words semantically related to the target set, was slowed to a rate of 197.4 msec per word.

One might suppose that the subjects in the experimental group were just surprised to see related words in the background and a few long pauses accounted for the entire difference. However, on this account one would expect the difference soon to disappear. But this did not happen. Through 5 additional hours of searching (they searched through 2,000 words during a 1-hour session), the difference between the control and experimental subjects remained at about 20 msec per word. It would thus appear that, even when searching for particular words, our expectations are based on meaning as well as visual form.

Using still another experimental procedure, Tulving and Gold (1963) and Tulving, Mandler, and Baumal (1964) both found that the prior presentation of a sentence context lowers the threshold at which a tachistoscopically presented word can be recognized.

Again we have a case of a higher level of processing (meaning) apparently affecting our ability to process at a lower level (the word level). Notice, moreover, that semantic relatedness can either make our processing more efficient (as with the Meyer et al., 1974, and Tulving and Gold, 1963, experiments), or it can interfere with our processing (as with the Graboi, 1974, experiment). It is again difficult to see how a strictly bottom-up, stage-by-stage processing model can account for results such as these.

Our Perception of Syntax Depends on the Semantic Context in Which the String Appears

Although neither Gough nor LaBerge and Samuels have attempted to specify their models much beyond the level of words, a complete model of reading must, of course, account for the way semantics affects our apprehension of the syntax of a sentence we are reading. Experiments at this level are few and far between, but there are numerous examples that seem rather compelling on this general point.

Perhaps the most commonly observed effect of this sort involves the semantic disambiguation of syntactically ambiguous sentences. Consider the following sentences:

(1) a. They are eating apples.
    b. The children are eating apples.
    c. The juicy red ones are eating apples.

At the syntactic level, all three sentences allow for at least two readings:

1. The reading in which the thing referred to by the first noun phrase is performing the act of eating some apples.
2. The reading in which the thing referred to by the first noun phrase is said to be a member of the class of "eating apples."

However, at a semantic level only the first one remains ambiguous—even it would be disambiguated if we had some notion as to the referent of *they*.

Schank (1973) has given a number of similar examples. Consider, for example, the following sentences:

(2) a. I saw the Grand Canyon flying to New York.

b. I saw the Grand Canyon *while I was* flying to New York.

c. I saw the Grand Canyon *which was* flying to New York.

Most readers immediately interpret Sentence (2a) as meaning the same as (2b) rather than Sentence (2c) simply on the grounds that it is semantically anomalous to imagine the Grand Canyon actually flying. On the other hand, Sentence (3a) is ordinarily interpreted to mean the same as Sentence (3c) rather than Sentence (3b):

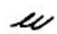
(3) a. I saw the cattle grazing in the field.

b. I saw the cattle *while I was* grazing in the field.

c. I saw the cattle *that were* grazing in the field.

In Examples (1), (2), and (3), semantics play the determining role as to which surface structure we apprehend. Thus, just as orthographic structure affects our ability to perceive letters and syntax, and semantics affects our perception of words, so too does semantics affect our apprehension of syntax.

Our Interpretation of the Meaning of What We Read Depends on the General Context in Which We Encounter the Text

Just as the appropriate interpretation of our ambiguous symbol, , was determined by the sentence in which it was embedded, so too it often happens that the *meaning* of a word is dependent on the words surrounding it. Consider, for example, the following sentences:

(4) a. The statistician could be certain that the difference was significant *since all of the figures on the right-hand side of the table were larger than any of those on the left.*

b. The craftsman was certainly justified in charging more for the carvings on the right *since all of the figures on the right-hand side of the table were larger than any of those on the left.*

Here our interpretation of the second clause is thus quite different depending on the nature of the first clause. In Sentence (4a) for example, the term *figure* is readily interpreted as being a number, the term *table* a place for writing numbers, and the relation *larger* can properly be interpreted to mean >. In Sentence (4b) on the other hand, the term *figure* presumably refers to a small statue, the term *table* refers to a physical object with a flat top used for setting things on, and the relation *larger* clearly means something like *of greater volume*. Here we have a case in which no determination about the meaning of these individual words can be made without consideration of the entire sentence. Thus, no decision can be made about the meaning of a word without consideration of the meaning of the entire sentence in which the word appears.

Not only is the interpretation of individual words dependent on the sentential context in which they are found, but the meaning of entire sentences is dependent on the general context in which they appear. The following example from Bransford and Johnson (1973) is a case in point:

(5)

Watching a Peace March From the 40th Floor

> The view was breathtaking. From the window one could see the crowd below. Everything looked extremely small from such a distance, but the colorful costumes could still be seen. Everyone seemed to be moving in the same direction in an orderly fashion and there seemed to be little children as well as adults. The landing was gentle, and luckily the atmosphere was such that no special suits had to be worn. At first there was a great deal of activity. Later, when the speeches started, the crowd quieted down. The man with the television camera took many shots of the setting and the crowd. Everyone was very friendly and seemed glad when the music started. (p. 412)

In this passage, the sentence beginning "The landing was gentle..." appears to make no sense. No clear meaning can be assigned to it in this context. As such, when subjects were given the passage and later asked to recall it, very few subjects remembered the anomalous sentence. On the other hand, when the passage was titled "A Space Trip to an Inhabited Planet" the entire passage was given quite a different interpretation. In this case, the anomalous sentence fits into the general interpretation of the paragraph very well. Subjects given the "Space Trip" title recalled the critical sentence three times as often as those given the "Peace March" title. Many other examples could be given. The dependence of meaning on context would appear to be the norm rather than the exception in reading.

To summarize, these results taken together appear to support the view that our apprehension of information at one level of analysis often can depend on our apprehension of information at a higher level. How can this be? Surely we cannot first perceive the meaning of what we read and only later discover what the sentences, words, or letters were that mediated the meaning. To paraphrase a remark attributed to Gough (as cited in Brewer, 1972), it is difficult to "see how the syntax [or semantics, for that matter] can go out and mess around with the print"

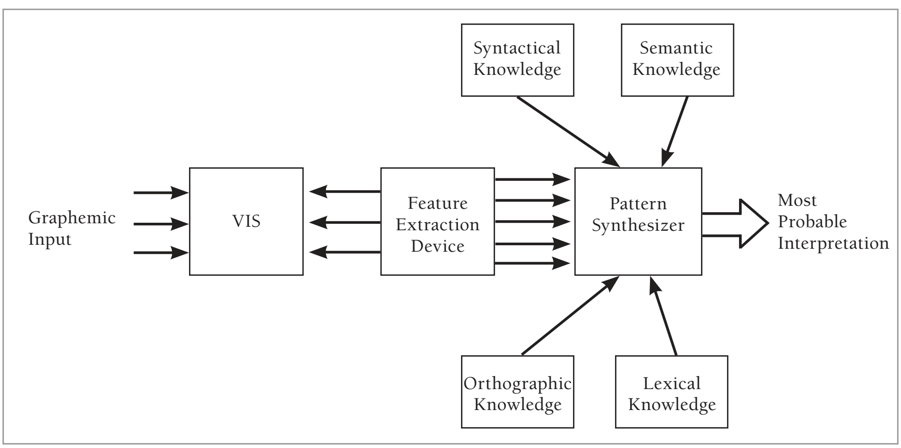
(p. 360). The problem, I believe, arises from the linear stage formalism that has served so well. The answer, I suspect, comes by presuming that all these knowledge sources apply simultaneously and that our perceptions are the product of the simultaneous interactions among all of them.

An Interactive Model

Perhaps the most natural information-processing representation of the theoretical ideas suggested in the previous section is illustrated in Figure 5. The figure illustrates the assumption that graphemic information enters the system and is registered in a visual information store (VIS). A feature extraction device is then assumed to operate on this information, extracting the critical features from the VIS. These features serve as the sensory input to a pattern synthesizer. In addition to this sensory information, the pattern synthesizer has available nonsensory information about the orthographic structure of the language (including information about the probability of various strings of characters), information about lexical items in the language, information about the syntactic possibilities (and probabilities), information about the semantics of the language, and information about the current contextual situation (pragmatic information). The pattern synthesizer, then, uses all of this information to produce a "most probable interpretation" of the graphemic input. Thus, all of the various sources of knowledge, both sensory and nonsensory, come together at one place, and the reading process is the product of the simultaneous joint application of all the knowledge sources.

Although the model previously outlined may, in fact, be an accurate representation of the reading process, it is of very little help as a model of reading. It is one thing to suggest that all of these different information sources interact (as many writers have) but quite another to specify a psychologically plausible

Figure 5. A Stage Representation of an Interactive Model of Reading



hypothesis about how they interact. Thus, it is clear why serious theorists who have attempted to develop detailed models of the reading process (e.g., Gough, 1972; LaBerge & Samuels, 1974) have stayed away from a formulation of the sort illustrated in Figure 5. All that is interesting in the model takes place in the box labeled *Pattern Synthesizer*. The flowchart does little more than list the relevant variables. We need a representation for the operation of the pattern synthesizer itself. To represent that, we must develop a means of representing the operation of a set of parallel interacting processes.

Flowcharts are best suited to represent the simple serial flow of information. They are badly suited for the representation of a set of parallel, highly interactive processes. However, with the advent of the parallel computer (at least as a conceptual device), computer scientists have begun to develop formalisms for the representation of parallel processes. It is interesting that the major problem in each case seems to have been the representation of the lines of communication among the otherwise independent processes.

Of the several different systems of communication that have been proposed, two were developed in the context of language processing by computer and seem to be most promising as a formalism for the development of a reading model. One of these was developed by Kaplan (1973) and is called the General Syntactic Processor (GSP). The second was developed by Reddy and his associates at Carnegie Mellon University (see Lesser, Fennell, Erman, & Reddy, 1974) as an environment for a speech understanding program. This system is called HEARSAY II. These two systems have a good deal in common and solve the communication problem in much the same way—namely, both systems consist of sets of totally independent asynchronous processes that communicate by means of a global, highly structured data storage device. In Kaplan's system the communication center is called a *chart*; in the HEARSAY system it is called a *blackboard*. I use the more neutral term *message center* in my development below. This development is most closely related to the HEARSAY system and could well be considered as an application of the HEARSAY model to reading. However, I also draw from aspects of GSP, and the model as I develop it has the Rumelhart and Siple (1974) model of word recognition as a special case.

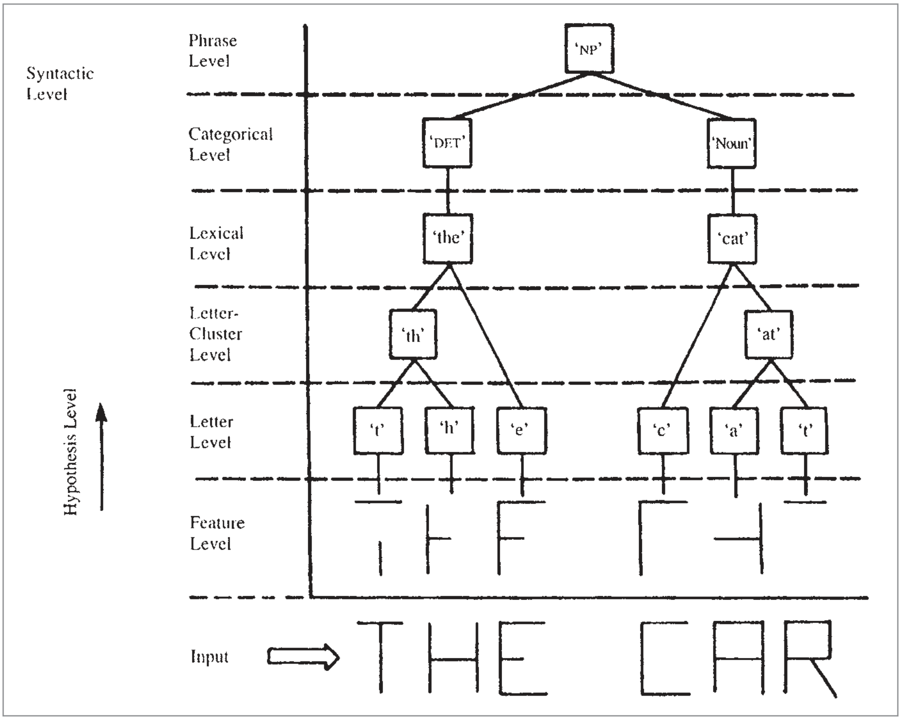
Following HEARSAY, the model can be characterized as consisting of a set of independent *knowledge sources*. (These knowledge sources correspond to the sources of input to the pattern synthesizer in Figure 5.) Each knowledge source contains specialized knowledge about some aspect of the reading process. The message center keeps a running list of hypotheses about the nature of the input string. Each knowledge source constantly scans the message center for the appearance of hypotheses relevant to its own sphere of knowledge. Whenever such a hypothesis enters the message center, the knowledge source in question evaluates the hypothesis in light of its own specialized knowledge. As a result of its analysis, the hypothesis may be confirmed, the hypothesis may be disconfirmed and removed from the message center, or a new hypothesis can be added to the message center. This process continues until some decision can be reached. At that point the most probable hypothesis is determined to be the correct one. To facilitate the process, the message center is highly structured so that the knowledge sources know exactly where to find relevant hypotheses and so that dependencies among hypotheses are easily determined.

The Message Center

The message center can be represented as a three-dimensional space: one dimension representing the position along the line of text, one dimension representing the level of the hypothesis (word level, letter level, phrase level, etc.), and one dimension representing alternative hypotheses at the same level. Associated with each hypothesis is a running estimate of the probability that it is the correct hypothesis. Moreover, hypotheses at each level may have pointers to hypotheses at higher or lower levels on which they are dependent. Thus, for example, the hypothesis that the first word in a string is the word *the* is supported by the hypothesis that the first letter of the string is *t* and supports the hypothesis that the string begins with a noun phrase.

Figure 6 illustrates a two-dimensional slice of the message center at some point during the reading of the phrase *the car.*

Figure 6. A Two-Dimensional Slice of the Message Center



The figure illustrates hypotheses at five different levels (feature level, letter level, letter-cluster level, lexical level, and syntactic level). The diagram is only a two-dimensional slice inasmuch as no alternative hypotheses are illustrated. In practice, of course, many alternative hypotheses would be considered and evaluated in the course of reading this phrase. It should be pointed out that the tree-like structure should not be taken to mean that the tree was constructed either from a purely bottom-up process (starting with the features, then hypothesizing the letters, then the letter clusters, etc.), nor from a purely top-down analysis (starting with a view that we have a noun phrase and that noun phrases are made up of determiners followed by nouns, etc.). Rather, the hypotheses can be generated at any level. If it is likely that a line begins with a noun phrase, then we postulate a noun phrase and look for evidence. If we see features that suggest a *t* as the first letter, we postulate a *t* in the first position and continue processing. If we later have to reject either or both of these hypotheses, little is lost. The system makes the best guesses and checks out their implications. If these guesses are wrong, it will take a bit longer, but the system will eventually find some hypotheses at some level that it can accept.

An Example. To illustrate the operation of the system, consider the following experimental procedure. A subject is presented with a picture (e.g., Figure 7) and

Figure 7. A Scene

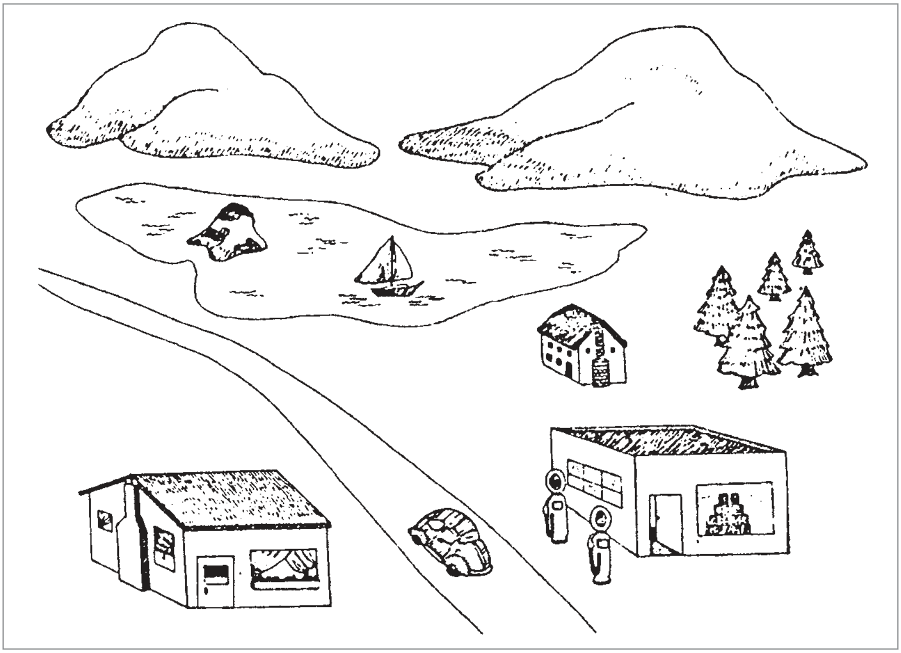
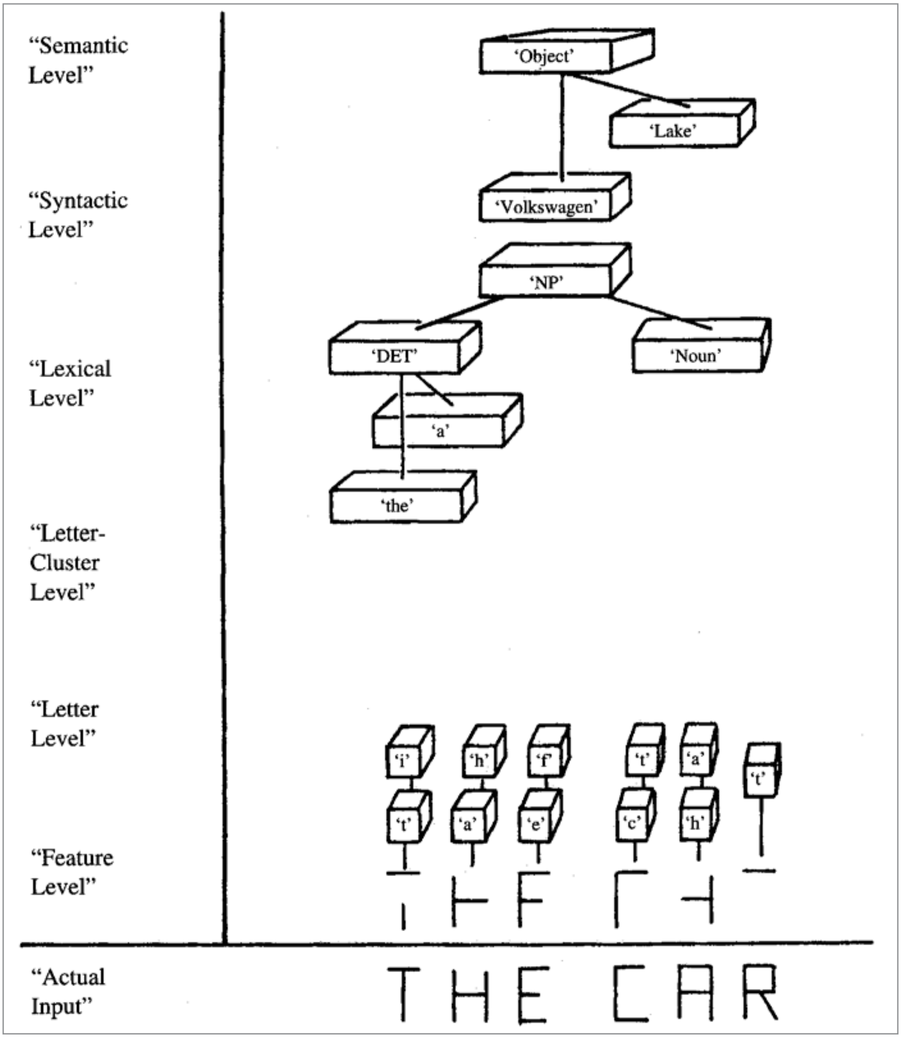


Figure provided by Jean Mandler.

allowed to view it for a few seconds. Then he is given a tachistoscopic presentation of a noun phrase that he knows will refer to one of the objects in the picture. His job is to decide which object was referred to. This experimental procedure is designed to simulate the process of reading a phrase for meaning. (An experimental procedure of this sort is currently under development in our laboratory.) I will illustrate the current model by showing the changes we might expect in the message center as the phrase THE CAR is read after viewing Figure 7.

Figure 8 shows the message center at an early point in the processing of this phrase. The subject knows from the instructions of the experiment that the

Figure 8. The Message Center Shortly After Processing Has Begun on THE CAR



phrase will refer to some object in the picture. Thus, the semantic-level "object" hypothesis can be entered and assigned a high likelihood value from the start. Moreover, through looking at the picture and perceiving certain aspects of it as salient, the subject will develop expectations as to the probable referent of the phrase. In this case, I have assumed that the subject set up special expectations for a phrase referring to *the lake* or to *the Volkswagen*.

Similarly, at the syntactic level, the subject can be quite certain that the input will form a noun phrase. Thus, the hypothesis "NP" is entered into the message center and assigned a high value. Noun phrases have a rather characteristic structure. About 25% of the time they begin with a determiner (DET). Thus, in the example, I have assumed that the hypothesis that the first word was a determiner was entered. Similarly, we can expect the second word of a noun phrase to be a noun about 20% of the time. Thus, I have entered the hypothesis that the second word is a noun. Now, in the case where the first word is a determiner, we could expect it to be the word *the* about 60% of the time and the word *a* about 20%. Thus, I have assumed that these two hypotheses have also been entered.

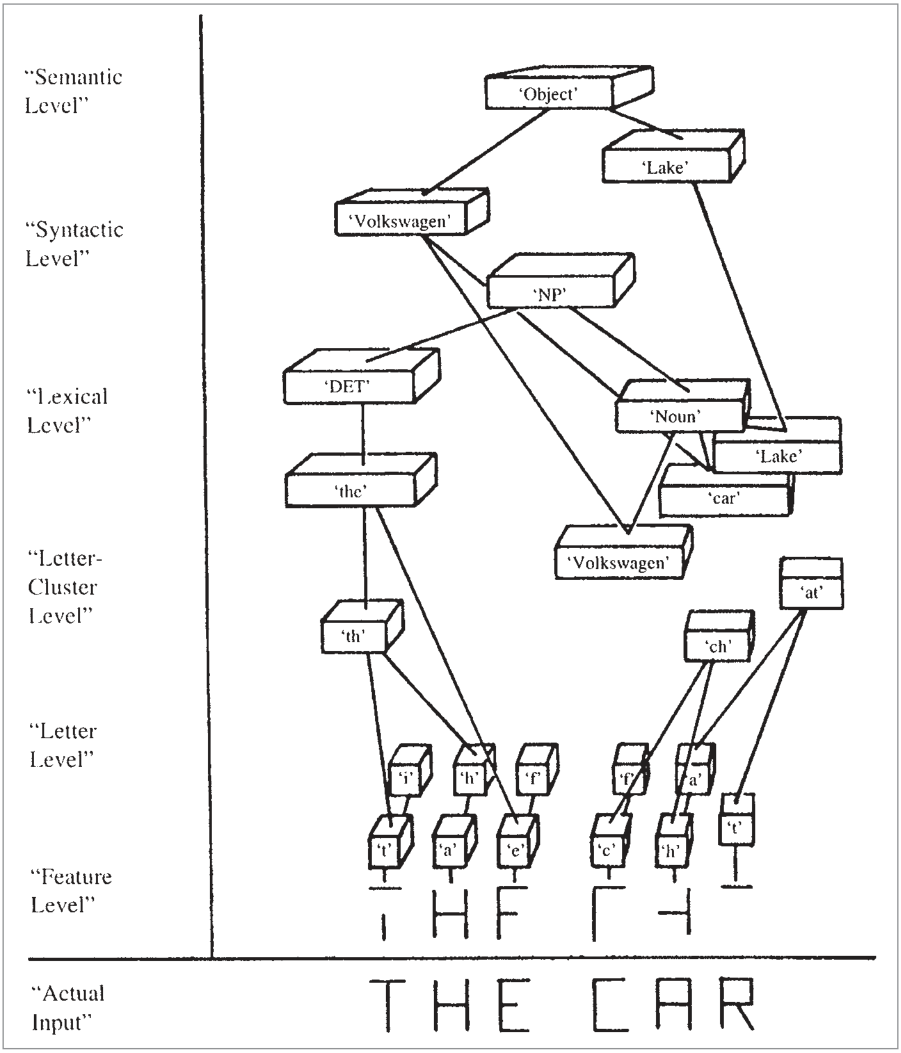
As all these hypotheses are being entered in top-down fashion, hypotheses at the letter level also are being entered bottom-up on the basis of featural information. In the example, I have assumed that for each of the first five letter positions the two most promising letter possibilities were entered as hypotheses. For the sixth letter position, which contains very little featural information, I have assumed that only its most likely letter hypothesis has been entered.

Figure 9 illustrates the state of the message center at a later point in the processing. In the meantime, the lexical hypothesis "a" has led to a letter hypothesis that was then tested against the featural information and rejected. The hypothesization of an initial "t" has led to the hypothesization of an initial "th" at the letter cluster level—a hypothesis that is given added validity by the possible "h" in the second position. The lexical-level hypothesis of the word "the" also has led to the hypothesization of the letter cluster "th" followed by the letter "e." The prior existence of these hypotheses generated from the bottom up has led to a mutual strengthening of all of the hypotheses in question and a resultant weakening of the alternative letter hypotheses at the first three letter positions.

While this processing was taking place, lexical hypotheses were generated from the semantic level as possible nouns. In this instance I have assumed that the semantic hypothesis "lake" has led to the lexical hypothesis that the word *lake* was in the string and that the semantic hypothesis "Volkswagen" has led to the lexical hypothesis "Volkswagen" and to the lexical hypothesis "car." Meanwhile, the letter hypotheses have led to alternative letter-cluster hypotheses "ch" and "at."

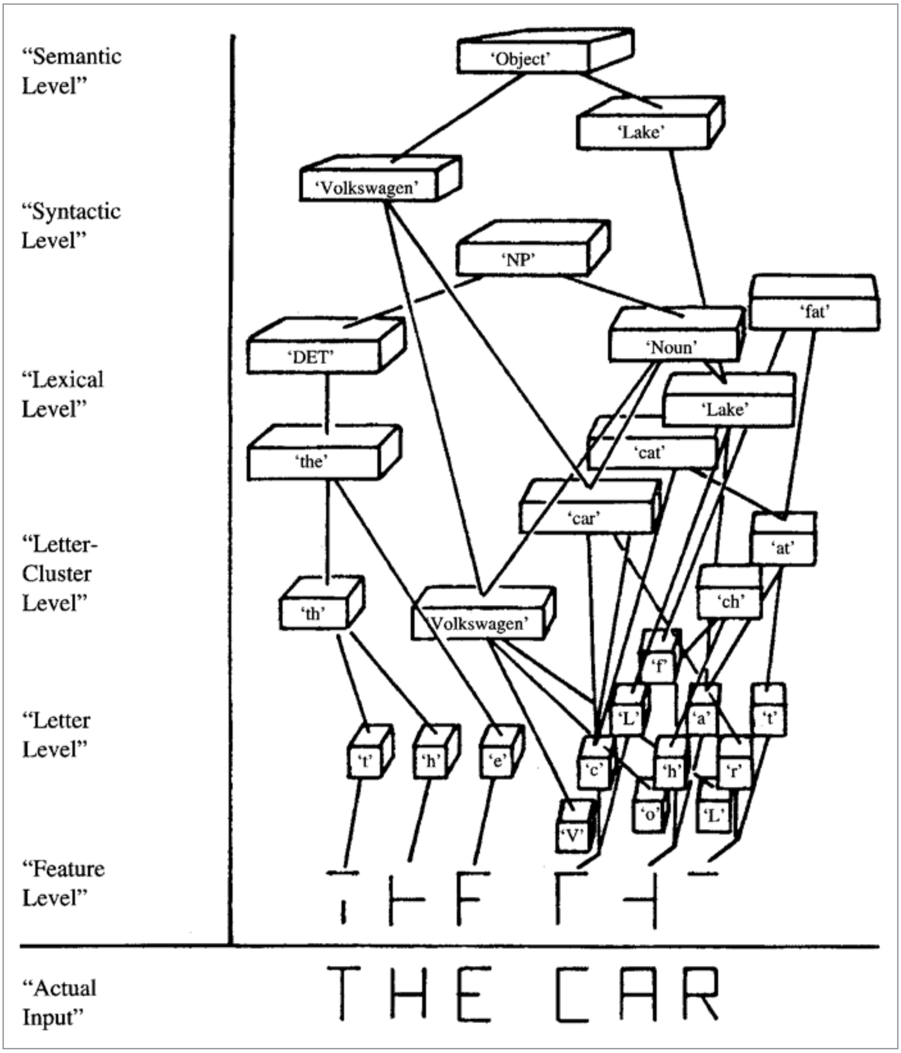
Figure 10 illustrates the state of the message center at a still later point in the processing of the input. By this point, the hypothesis that the first word is *the* has reached a sufficient value that further processing has ceased. No new hypotheses have been generated about the first word. On the other hand, lexical hypotheses on the second word have proliferated. The existence of the letter hypothesis "c" followed by the letter-cluster hypothesis "at" has led to a hypothesization of the

Figure 9. The Message Center Slightly Later in the Processing Sequence



lexical item "cat." Similarly, the letter hypothesis "f" followed by "at" has led to hypothesizing the lexical item "fat." The lexical hypothesis "cat" is consistent with the "noun" hypothesis, thus strengthening the view that the second word is a noun. At the same time, the lexical hypotheses "lake," "Volkswagen," and "car" either have strengthened existing letter hypotheses or have caused new ones to be generated. Notice, in particular, that the prior existence of the letter hypotheses "c" and "a" strengthened the semantically derived lexical hypothesis "car," which in turn strengthened the letter hypothesis "r"—even though the letter hypothesis

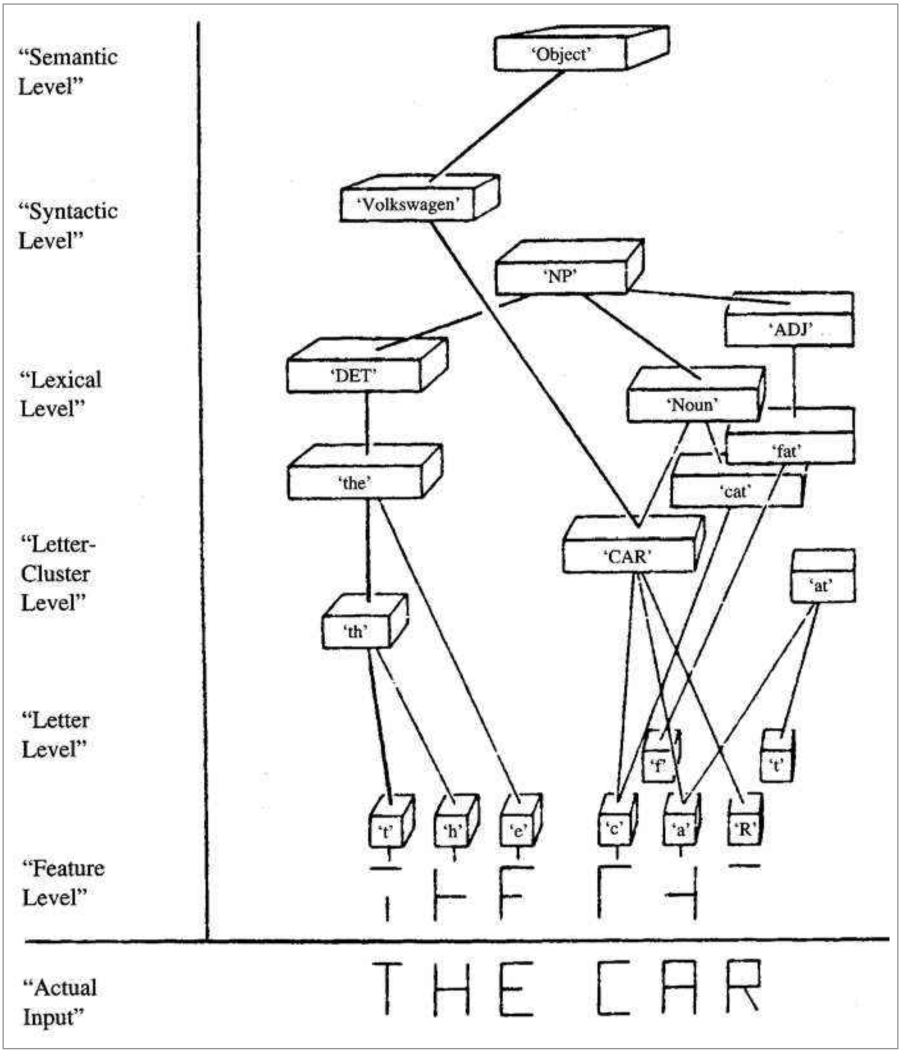
Figure 10. The Message Center Well Into the Processing Sequence



"r" has not yet been evaluated in light of the featural information in the final position.

Finally, Figure 11 illustrates a state of processing after the letter hypotheses have been tested against the featural information. At this point only three lexical hypotheses for the second word remain—"fat," "cat," and "car." The lexical hypothesis "fat" has led to the syntactic hypothesis that the second word is an adjective (ADJ), and the lexical hypothesis "cat" has led to the semantic-level hypothesis that there should be a cat in the picture. Meanwhile, the semantic

Figure 11. The Message Center Near the End of Processing: At This Point the Semantics of the Input Have Been Pretty Well Determined



hypothesis "Volkswagen" has been strengthened by the finding that the final featural information is consistent with the hypothesis that the last letter is an *r*. At this point the semantic hypothesis "Volkswagen" is probably high enough to lead to a response. If not, a test of the semantic hypothesis "cat" will lead to the rejection of that hypothesis and the consequent strengthening of the "Volkswagen" hypothesis and thus the lexical hypothesis "car" and the letter-level hypotheses "c," "a," and "r."

It should be clear from this example how, in principle at least, one could build a model of reading that actually would employ constraints from all levels concurrently in the process of constructing an interpretation of an input string. Of course, this example is a long way from the specification of such a model. All I have illustrated here is the nature of the message center and how it is structured to facilitate communication among processes acting at various levels. Before a concrete model of reading can be specified, the nature of the various knowledge sources must be specified as well. I now turn to a brief discussion of the separation of the various knowledge sources.

The Knowledge Sources

I do not yet have a detailed model of the operation of all the knowledge sources. However, I do have ideas about a number of them and will now discuss them.

1. *Featural knowledge.* At this level, I am assuming that features are extracted according to the assumptions of the Rumelhart and Siple (1974) model. Moreover, I am assuming that these critical features are the basic level of processing. In a tachistoscopic experiment, all decisions must be made with respect to the set of features extracted during and shortly after the exposure. In freereading situations, the reader can go back and get more featural information if no hypothesis gets a sufficiently high rating or if some hypothesis does get a high rating at one point and is later rejected. Such occasions probably account for regressions in eye movements.

2. *Letter-level knowledge.* This knowledge source scans the feature inputs, and whenever it finds a close match to a known letter, it posits a letter hypothesis. In addition, whenever a letter hypothesis appears from a higher level, this knowledge source evaluates that hypothesis against the feature information. In addition to information about letters in various fonts, the letter-level knowledge source presumably takes into account the probabilities of letters in the language. Thus, relatively more featural evidence would be necessary to postulate a "z" or "q" than an "e" or a "t."

3. *Letter-cluster knowledge.* This knowledge source scans the incoming letter-level hypotheses, looking for letter sequences that are likely and form units in the language or for single letters that are frequently followed or preceded by another letter (e.g., as *q* is frequently followed by *u*). In either case a letter cluster is postulated. In the latter case a letter-level hypothesis is also introduced. (That is, if a *q* is found, a *qu* is postulated at the letter-cluster level and a *u* is postulated at the letter level.) The value associated with any of these hypotheses depends on the values of the letter-level information on which it is based and on the frequency of such clusters in the language. In addition, the letter-cluster knowledge source looks for the introduction of letter-cluster hypotheses from the lexical level. Whenever it finds these it evaluates them by proposing the appropriate letter-level hypotheses. For this knowledge source, as with all others, the most probable hypotheses that are unsupported from the following or that support no higher-level hypotheses are evaluated first.

4. *Lexical-level knowledge.* The lexical-level knowledge source operates in exactly the same way as the other knowledge sources. It scans the letter-cluster and letter hypotheses for letter sequences that form lexical items or that are close to lexical items. When it finds such information, it posits the appropriate lexical-level hypotheses and any additional letter-cluster or letter-level hypotheses. When evaluating the goodness of any hypothesis, it takes into account the goodness of the evidence on which it is based and the a priori frequency of that item in the language. In addition, whenever a lexical item is postulated from either the semantic or syntactic levels, this knowledge source evaluates that hypothesis by postulating those letter-cluster and letter hypotheses that are not yet present. Those letter and letter-cluster hypotheses that are present are strengthened due to the convergence of lines of evidence. Other alternatives without such convergent information are relatively weakened.

5. *Syntactic knowledge.* Like all the other knowledge sources, this knowledge source is designed to operate in both a bottom-up and top-down mode. Thus, whenever a lexical hypothesis is suggested, one or more syntactic category hypotheses are entered into the message center. In general, not all syntactic category hypotheses consistent with the lexical form would be expected. Instead, those categories that are most probable, given that lexical item, would be entered first. Similarly, sequences of lexical category hypotheses would be scanned, looking for phrase possibilities, and so on. At the same time, the syntactic knowledge source would have the capacity to operate in a top-down fashion. Thus, for example, whenever a noun-phrase hypothesis were entered, the syntactic knowledge source would establish, say, a determiner, syntactic category hypothesis that in turn might initiate lexical-level hypotheses of determiner words such as *a* and *the*. Following Kaplan's (1973) GSP, I assume that this top-down portion of the syntactic knowledge source would be well represented by an Augmented Transition Network (ATN) parser. (See Stevens and Rumelhart, 1975, for an application of an ATN to reading data.) Like all other levels, the syntactic hypotheses are given values dependent on the goodness of the evidence (or prior probabilities) of the hypotheses on which they are based. Moreover, a convergence of top-down and bottom-up hypotheses strengthens both.

6. *Semantic-level knowledge.* This is perhaps the most difficult level to characterize. Nevertheless, I assume that its operation is essentially the same as the others. Whenever strong lexical hypotheses occur, this knowledge source must have the ability to look for semantic-level correlates to evaluate the plausibility of the hypothesis (at both the lexical and syntactic levels). Moreover, it must be able to develop hypotheses about the content of the input and generate lexical-level hypotheses as possible representations of this. The experimental procedure discussed in the previous section was designed as an attempt to reduce the complexity of the semantic component by supposing a relatively simple referential semantics.

Still, of course, after having outlined the functional characteristics of the various knowledge sources, I am still far from the quantitative model I have in mind.

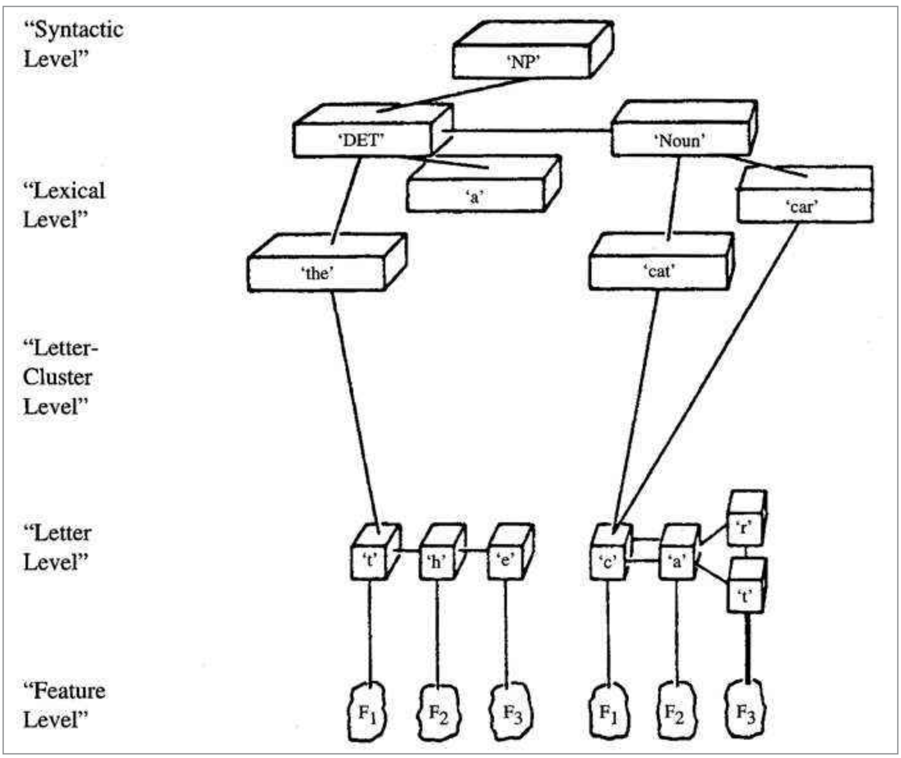
However, it would appear that a HEARSAY-type model such as this offers promise as a framework for the development of serious models of reading that nevertheless assumes a highly interactive parallel processing system.

A Mathematical Model of Hypothesis Evaluation

In this section I will specify in somewhat more detail the nature of the hypothesis-evaluation process I envision. Figure 12 illustrates a simplified version of the message center from the primary example. This figure differs somewhat in format from the previous figures of this type in order to make clearer the sequential dependencies among hypotheses at the same level. Thus, the fact that an "NP" consists of a "DET" and "NOUN" and that the word *cat* consists of *c* + *a* + *t* is illustrated by the arrows connecting those constituents at the same level. Moreover, the dependency arrows have been drawn to only the left-hand member of such hypothesis sequences. In a sense, as we shall see below, the left-hand member is representing the entire sequence of hypotheses.

There are four different types of dependency relationships among hypotheses in this model. These types are illustrated in the figure. First, a hypothesis may

Figure 12. An Illustration of the Relations Among the Hypotheses in the Message Center



have one or more *daughter* hypotheses. A daughter hypothesis is one at a lower level that is connected directly to a higher-level hypothesis. In the figure, the hypothesis "DET" has two daughter hypotheses: "the" and "a." The hypothesis "the" has a single daughter, "t." A hypothesis may have any number of daughters. Each daughter is an alternative way in which the higher hypothesis can be realized. Thus, "the" and "a" are alternative ways in which "DET" can be realized. For any hypothesis, h_i, I shall use the symbol D_i to designate its set of daughters.

The reciprocal relationship to daughter is *parent*. Any hypothesis may have one or more parent hypotheses. A parent hypothesis is one to which a hypothesis can lend direct support. Thus, in the figure, "NOUN" is a parent of both "car" and "cat." Similarly, the letter hypothesis "c" has two different parents, "car" and "cat." Only hypotheses that are at the left-most position of a sequence of hypotheses may have parents. Thus, the hypothesis "NOUN" has no parent. For each hypothesis, h_i, I shall designate the set of parents P_i.

In addition to parents and daughters, hypotheses may have *sisters*. Sisters are hypotheses in a sequence that either follow or precede a particular hypothesis at the same level. Sisters are not alternatives but are consistent possibilities of the same level. There are two sorts of sister hypotheses: *right sisters* and *left sisters*. Right sisters are hypotheses that follow a given hypothesis in a sequence of hypotheses. Thus, "NOUN" would be a right sister of "DET," and "r" and "t" are right sisters of the letter-level hypothesis "a." I designate the set of right sisters of h_i as R_i. Left sisters are those hypotheses that precede a given hypothesis in a string of hypotheses. Although it is possible for a hypothesis to have more than one left sister, no cases of this are illustrated in the figure. I designate the set of left sisters of hypothesis h_i as L_i.

We are now in a position to develop a measure for evaluating hypotheses. The measure that I will propose is essentially the Baysian probability that the hypothesis is true given the evidence at hand. The evidence favoring a particular hypothesis can be broken down into two parts: contextual evidence, dependent only on sister and parent hypotheses; and direct evidence, dependent solely on the evidence derived from daughter hypotheses and, ultimately, featural evidence. Equation (1) illustrates the assumed multiplicative relationship between these two kinds of evidence:

(1) $$s_i = v_i \cdot \beta_i^4$$

where s_i is the overall strength of the hypothesis h_i, v_i is a measure of the direct evidence for h_i, and β_i is a measure of the contextual evidence for h_i. Now we can define the values of v_i and β_i in terms of the parents and sisters of h_i. Equation (2) gives the value of the contextual strength of h_i:

(2) $$\beta_i = \begin{cases} \Pr(h_i) & P_i = L_i = \varphi \\ \Sigma_{s_k} \cdot \dfrac{\Pr(h_i|h_k)}{v_i} & \text{otherwise,} \end{cases}$$

where the sum is over all $h_k \in P_i$ or L_i. Thus, when h_i has no parents or left sisters, its contextual strength is given by its a priori probability. Otherwise, its contextual strength is given by the sum, over all of its left sisters and parents of the strength of the left sister or parent, h_k, times the conditional probability of the hypothesis given h_k. The sum is then divided by its own direct strength so that its direct strength will not contribute to its contextual strength (because as we shall see, its own direct strength contributes to the strength of its parents and left sisters and is represented multiplicatively in s_k).

Direct evidence for a hypothesis comes only from its daughters. Equation (3) gives the direct evidence for a hypothesis as a function of a value associated with its daughters:

$$(3) \quad v_i = \begin{cases} \Sigma C_{ik} \cdot \Pr(h_k | h_i) & D_i \neq \varphi \\ 1 & \text{otherwise,} \end{cases}$$

where the sum is over all $h_k \in D_i$, and where C_{ik} is the *cumulative evidence* for hypothesis h_i associated with the sequence of hypotheses whose left-most member is the daughter h_k. Thus, in the diagram, the direct evidence for "car" is determined jointly by the direct evidence for "c," for "a," and for "r." The value of C_{ik} is given by the following equation:

$$(4) \quad C_{i,k} = \begin{cases} v_k & R_k = \varphi \\ \sum_j v_k \cdot C_{i,j} \cdot \Pr(h_j | h_i, h_k) & \text{otherwise,} \end{cases}$$

where the sum is over all $h_j \in R_k$. Thus, the cumulative evidence for hypothesis h_i associated with hypothesis h_k is determined by the product of the direct evidence for h_k and the cumulative evidence for its right sister. If its probable right sisters are very strong, then the cumulative evidence is very strong and thus offers good support to its parent. Otherwise, it offers support against its parent.

Finally, we must give special attention to the first-level hypotheses associated with featural-level inputs. For any letter hypothesis h_i, featural-level inputs have cumulative values of C_{iF} given by:

$$(5) \quad C_{i,F} = [\Pr(F)]^{-1}$$

where F is the set of features observed in that location. This, in effect, is a normalizer designed to keep the strengths in the 0 to 1 range.

The equations (1) to (5) define a system of evaluation that makes near optimal use of the information available at any given point in time. Whenever a new hypothesis is postulated and a new connection is drawn, new values must be computed for the entire set of hypotheses. Resources can be allotted to the knowledge sources based upon their momentary evaluations. Effort can be focused on generating hypotheses from the top down whenever we have hypotheses with strong contextual strengths and few daughter hypotheses. Effort can be focused on the generation of hypotheses from the bottom up whenever there is strong direct evidence and few parents. Moreover, the strength values can be signals to stop processing and accept a hypothesis. When some criterion strength value is obtained, a hypothesis can be accepted and no further processing need be required. Then resources can be siphoned to other more critical areas.

Of course, specifying equations such as these does not fully specify our model. We must specify all of the knowledge sources and how they postulate hypotheses. They do, I feel, illustrate that the model under consideration can be quantified and can generate specific predictions—in spite of the enormous complexity of a highly interactive system.

QUESTIONS FOR REFLECTION

1. Why does Rumelhart disagree with linear models of reading processing that suggest letter perception precedes word or syntactic recognition?
2. How does Rumelhart use the concept of hypothesis testing to explain his theory of message center processing?
3. Rumelhart proposed that bottom-up models do not fully describe the influence of higher level processing on lower level perceptions. What is his evidence for this proposition?
4. Why did Rumelhart consider his model interactive rather than bottom-up or top-down?

ACKNOWLEDGMENT

Research support was provided by grant NS 07454 from the National Institute of Health.

NOTES

*When this chapter was written, Rumelhart was at the University of California, San Diego.

[1] LaBerge and Samuels were particularly interested in the role of attention and the notion of automaticity in reading. I also have omitted discussion of episodic memory because neither one of these aspects of their model is relevant to my point here.

[2] Actually, the aforementioned attention mechanism of the LaBerge–Samuels model offers *some* top-down capacity. However, within their model it is limited and serves to speed up certain weak bottom-up paths.

[3] I use the term *perception* rather freely here. In general, it is my opinion that the distinction between the perceptual and conceptual aspects of reading is not that useful. As I will suggest later, there appears to be a continuity between what has been called perception and what has been called comprehension. My use of the term *perception* in the present context is simply the use of the one term to cover the entire process.

[4] This is the same relationship between these two sorts of evidence assumed by Luce (1959) and which is incorporated into the Rumelhart and Siple (1974) model for word recognition.

REFERENCES

Bransford, J.D., & Johnson, M.K. (1973). Considerations of some problems of comprehension. In W.G. Chase (Ed.), *Visual information processing*. New York: Academic.

Brewer, W.F. (1972). Is reading a letter-by-letter process? In J.F. Kavanagh & I.G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Gough, P.B. (1972). One second of reading. In J.F. Kavanagh & I.G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Graboi, D. (1974). *Physical shape, practice and meaning in visual search*. Unpublished doctoral dissertation, University of California, San Diego.

Huey, E.B. (1968). *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press. (Original work published 1908)

Kaplan, R.M. (1973). A general syntactic processor. In R. Rustin (Ed.), *Natural language processing*. New York: Algorithmics.

Kolers, P.A. (1970). Three stages in reading. In H. Levin & J.T. Williams (Eds.), *Basic studies in reading*. New York: Basic.

LaBerge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.

Lesser, V.R., Fennell, R.D., Erman, L.D., & Reddy, D.R. (1974). *Organization of the HEARSAY II speech understanding system* (Working Papers in Speech Recognition III). Pittsburgh, PA: Carnegie Mellon University.

Luce, R.D. (1959). *Individual choice behavior*. New York: Wiley.

McClelland, J.L., & Johnston, J.C. (1977). The role of familiar units in perception of words and nonwords. *Perception and Psychophysics*, *22*, 249–261.

Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

Meyer, D.E., Schvaneveldt, R.W., & Ruddy, M.G. (1972, November). *Activation of lexical memory*. Paper presented at the meeting of the Psychonomic Society, St. Louis, MO.

Meyer, D.E., Schvaneveldt, R.W., & Ruddy, M.G. (1974). Functions of phonemic and graphemic codes in visual word recognition. *Memory and Cognition*, *2*, 309–321.

Miller, G.A., Bruner, J.S., & Postman, L. (1954). Familiarity of letter sequences and tachistoscopic identification. *Journal of Genetic Psychology*, *50*, 129–139.

Miller, G.A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, *2*, 217–228.

Nash-Weber, B. (1975). The role of semantics in automatic speech understanding. In D.B. Bobrow & A. Collins (Eds.), *Representation and understanding*. New York: Academic.

Pillsbury, W.B. (1897). A study in apperception. *American Journal of Psychology*, *8*, 315–393.

Reicher, G.M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274–280.

Ruddy, M.G., Meyer, D.E., & Schvaneveldt, R.W. (1973, May). *Context effects on phonemic encoding in visual word recognition*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL.

Rumelhart, D.E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review*, *81*, 99–118.

Schank, R.C. (1973). Identification of conceptualizations underlying natural language. In R.C. Schank & K.M. Colby (Eds.), *Computer models of thought and language*. San Francisco: Freeman.

Schvaneveldt, R.W., & Meyer, D.E. (1973). Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic.

Stevens, A.L., & Rumelhart, D.E. (1975). Errors in reading: Analysis using an augmented network model of grammar. In D.A. Norman, D.E. Rumelhart, & the LNR Research Group, *Explorations in cognition*. San Francisco: Freeman.

Tulving, E., & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, *66*, 319–327.

Tulving, E., Mandler, G., & Baumal, R. (1964). Interaction of two sources of information in tachistoscopic word recognition. *Canadian Journal of Psychology*, *18*, 62–71.

Weber, R.M. (1970). First graders' use of grammatical context in reading. In H. Levin & J.T. Williams (Eds.), *Basic studies in reading*. New York: Basic.