INSTITUTE FOR THE
STUDY OF
INTELLECTUAL
BEHAVIOR

An Interactive Approach to Reading and Recognizing Sentences

Michael E. J. Masson and Linda S. Sala

University of Colorado

Program on Cognitive Factors in Human Learning and Memory Report No. 66

Institute for the Study of Intellectual Behavior

University of Colorado

Boulder, Colorado 80309

June, 1977

ABSTRACT

Two experiments are reported which clarify the status of surface and semantic aspects of sentences read in normal versus inverted typography. Experiment 1 provides evidence that Kolers' procedure of having subjects read sentences aloud does not result in very deep levels of processing when normal typography is used. Performing a sentence continuation task significantly increased recognition of originally normal sentences. No such task was required to attain high levels of recognition with originally inverted sentences. Sentence recognition was also strongly affected by repetition of wording and typography, supporting Kolers' procedural interpretation. It is concluded that reading inverted sentences emphasized the interaction between data driven and conceptually driven processes. The semantic nature of conceptually driven processes resulted in deep processing of inverted sentences regardless of task instructions. Experiment 2 replicated the results with the reading aloud task and showed that the second reading of an originally inverted sentence is equally swift when a paraphrased or verbatim test sentence is used. We conceptualize semantic and surface information processing as interacting perspectives of comprehension and memory processes.

Recent investigations of language understanding have spawned a number of models of the reading process (e.g., Cooper, 1972; Gough, 1972; LaBerge & Samuels, 1974; Rumelhart, Note 1; and Smith, 1971). While these models differ in their processing details, they generally share the view that a multilevel memorial representation is formed during the reading of a sentence--one that contains information about the meaning of the sentence, about its surface structure, and sometimes about its perceptual characteristics. However, the consensus among cognitive theorists seems to be that the most durable component of the memorial code is the set of propositions which represents sentence meaning. In most cases, information about surface structure is remembered less well than semantic information (Kintsch, 1974; Kintsch & Bates, 1977; Sachs, 1967).

Contrasting sharply with this line of thinking is the model of reading developed by Kolers (e.g., Kolers, 1975a, b, 1976a, b; Kolers & Ostry, 1974), which holds that the memorial representation of a sentence produced during reading consists of a record of the pattern analyzing operations which were used to encode the sentence. Kolers argued that, instead of viewing reading as a process in which we retain the surface structure of a sentence only long enough to extract meaning from it, we should view it as a process in which the procedures used to encode it are "part and parcel" of the memory representation (Kolers, 1975a). From this viewpoint, what is stored in memory are the procedures applied to the linguistic input at encoding, rather than a propositional representation of sentence meaning. Code durability is therefore contingent on the complexity and elaboration of the encoding operations themselves.

The two experiments described in this paper were motivated by some striking results reported by Kolers (Kolers & Ostry, 1974; Kolers, 1975a, b; Kolers, 1976a, b), who has examined memory for the orientation of the letters, or graphemes, in which sentences are embodied. Typically, Kolers has subjects read aloud two decks of sentences, a <u>read deck</u> and a <u>recognize deck</u>. In the read deck half of the sentences are presented in a transformed typography (e.g., they may be upside down or backwards) and half are presented in normal typography. The recognize deck contains all of the sentences in the read deck plus some new sentences. After reading each sentence in the recognize deck out loud, subjects must classify it either as a new sentence which they have never seen before, as an old sentence in different typography, or as an old sentence in the same typography.

Three findings from such studies are particularly important to current theorizing about memory. First, sentences that were originally read in transformed typography were recognized better than sentences originally read in normal typography. Second, the memorial representations formed during these experiments contain at least semantic and graphemic information for periods over a year (Kolers 1976a, b). Third, with repeated presentation, reading speed shows marked practice effects on the inverted typography.

These results pose several significant questions for conventional theories of memorial representation. If meaning is the most durable component of a multilevel representation, and if other components decay fairly rapidly, how can we explain the retention of graphemic information for periods over a year? Furthermore, if sentences are represented in memory propositionally--in terms of their meaning--how can we account for the fact that sentences originally read in inverted typography are better recognized than those originally read in normal typography? Why, in other words, should memorability be so significantly affected by typography if both normal and inverted sentences are similarly represented? Finally, what part does meaning play in producing practice effects on the inverted typography? Does recognition of meaning facilitate the second reading of a sentence, or is the facilitation due merely to the reinstantiation of pattern analyzing operations?

In explaining his results, Kolers relied heavily on notions about the procedural nature of the memory representation. He reasoned that, because the inverted typography is novel and unfamiliar to the reader, it requires more complex pattern analyses than does normal typography. As a consequence, a more elaborate procedural trace is laid down, making originally inverted sentences more memorable. Moreover, because the nature of the perceptual pattern analyzing operations is determined at least in part by the orientation of the graphemes in which a sentence is presented, information about typography as well as meaning is represented by the stored set of encoding operations. According to Kolers, recognition of a sentence occurs when the set of pattern analyzing operations that mediated its initial encoding are reinstantiated during its second reading. That is, meaning is reinstantiated whenever the appropriate set of encoding operations is reactivated. This suggests that at least some of the procedures that are represented in memory are sentence specific, rather than general and applicable to all similarly inverted sentences. As for reading speed, Kolers (1975a) attempted to show that the second reading of an inverted sentence is faster due to the reinstantiation of pattern analyzing operations rather than to the recognition of sentence meaning. Indeed, less facilitation is found that can be attributed to the semantic component of sentence processing than can be attributed to the processing of the transformed graphemes. The reasons for this may be less straightforward than Kolers has suggested, as will be discussed later.

We would like to propose that an alternative explanation for Kolers' results must be ruled out before it is necessary to discard conventional ideas about the representation of meaning in memory. This alternative explanation arises from a theoretical framework which is based on a fundamental distinction between "conceptually driven" and "data driven" processing (Bobrow & Norman, 1975; Norman, 1976).

We suggest that, in the process of reading, a reader uses the meaning and structure of a sentence or text to help him read. On the basis of his knowledge of the world, of the discourse, and of the language, he generates expectations about the kinds of words and meanings that are likely to occur next in a sentence or paragraph. In other words, as a reader deciphers the written message, information is sent down from memory to lower order processes that are operating on the incoming data, directing and facilitating their execution. This kind of processing has been referred to as conceptually driven. It contrasts with processes that are largely data driven, or perceptual, in nature. Kolers has emphasized graphemic pattern analyzing operations, but we can conceptualize pattern analyzing operations as occurring at a semantic level as well. A parallel can be drawn between what we mean by conceptually driven processes and semantic pattern analyzing operations.

The important feature of the framework presented here is that data driven and conceptually driven processes interact with and require each other (cf. analysis by synthesis, Neisser, 1967). The conceptually driven processes act as a means of anticipating input and the data driven processes can be used to confirm or reject these higher order hypotheses. Processing of a given stimulus event is likely to be data driven until enough information has been activated in memory to provide a basis for the predictive conceptually driven activity (Schank, 1973). Thus, we contend that sentence meaning is explicitly stored in memory, in addition to whatever might be represented in the way of surface structure or perceptual characteristics. Semantic memory is used in a conceptually driven way during reading.

The interaction of conceptually and data driven processes can be best understood in the context of the reading process itself. For someone who is reading aloud, this interaction can be automatic (cf. Shiffrin & Schneider, 1977) and meaning may be processed only to a superficial level. However, inverted typography represent a novel set of stimuli with which the data driven reading processes are not practiced. Because of its unfamiliarity, the operation of data driven processes on the inverted typography is slow and elaborate, and efficient decoding relies heavily on the use of conceptually driven processes. Unless the reader can generate meaning based hypotheses to aid him in analyzing the graphemes, successful reading of the text will be very difficult. Reading of inverted type involves extensive, consciously controlled interactions of conceptually driven and data driven processes.

Automatic processing is held to be comprised of fewer and qualitatively different sets of operations than unskilled or controlled processing (Kolers, 1975b; Lockhart, Craik, & Jacoby, 1976). Memory is directly affected by the nature of processing: elimination of some operations when a set of procedures becomes automatic may cause problems for retrieval at a later time. For example, originally normal sentences may be remembered less well than originally inverted sentences because the automaticity of processing normal sentences does not produce a memory trace that is extensively elaborated with respect to meaning. Controlled interactions between conceptually and data driven processes lead to semantically elaborated memorial representations.

An alternative explanation for Kolers' results may be derived from the present framework. It rests on three hypotheses. First, we hypothesized that the task of reading aloud which was used by Kolers required relatively automatic processing and did not require deep processing of sentence meaning (cf. Craik & Lockhart, 1972). If this were the case in Kolers' studies, then it is not surprising that sentence meaning had a small effect on task performance. If we were to contrast reading aloud (a highly automatic task) with a task requiring deep processing of meaning, it might be that the difference between memory for inverted and normal typography would be attenuated in a recognition test. In order to deeply process meaning, conceptually driven processing would be required during the reading of both kinds of typography, and, hence, the resulting representations would be elaborate and very durable. Therefore, half of our subjects were asked to perform a sentence continuation task (Bobrow & Bower, 1969) which required them to process meaning deeply. We predicted that the sentence continuation task would reduce differences between the rates of recognition for inverted versus normal sentences.

Our second hypothesis was that the reading of inverted and normal typography differ in the extent to which semantic processing is required during decoding. We postulated that, in reading the inverted typography, a reader would not only <u>process</u> meaning, but would <u>use</u> that meaning to help him decode the transformed symbols. Thus, we expected to find excellent memory for the meaning of originally inverted sentences, irrespective of task demands. In contrast, we postulated that the processing of normal typography under requirements to read aloud is a relatively automatic task, requiring shallow processing of meaning, and producing poor memory for meaning as a consequence. We therefore anticipated that memory for the meaning of normal sentences that were read aloud would be depressed. However, we also expected that, for those normal sentences whose meanings were deeply processed in the sentence continuation task, recognition of meaning would be very high. Thus we predicted that there would be a two way interaction between task demands and the typography in which a sentence originally appeared.

Third, we hypothesized that sentence recognition involves more than simply the reactivation of sentence specific sets of pattern analyzing operations. We reasoned that, if a strong version of Kolers' ideas is correct and only encoding operations (graphemic or semantic) are stored in memory, then subjects should be unable to recognize paraphrased sentences which mean the same as previously read sentences but require different sets of pattern analyzing operations during reading. That is, paraphrased sentences whose wording and syntax differ from originally read sentences should not reinstantiate previously stored sets of pattern analyzing operations and, thus, their meanings should not be recognized. On the other hand, if the products of encoding operations (i.e., meanings) are explicitly stored in memory, subjects should be very good at recognizing paraphrased sentences, even if they are very dissimilar in surface structure from original sentences. Therefore, half of the sentences were tested in paraphrased form so that the pattern analyzing operations used to read them would be as dissimilar as possible from the operations used to read their counterparts in the read deck. We predicted that subjects would be very good at recognizing paraphrases, in spite of the fact that they would not reinstantiate previously stored sets of encoding operations.

Finally, in order to test the recognizability of sentences whose meanings had changed but whose surface structure (and corresponding encoding operations) remained virtually the same, we included a set of verbatim false sentences. These sentences were presented in the read deck but were tested with slightly different versions in the recognize deck. Different versions were created through alteration of the meaning of original sentences by changing one or two words in each sentence. An example of this type of sentence is The students marched angrily on the commons unaware (well aware) that the state troopers awaited them. The test wording of the verbatim false sentences was otherwise identical to the original wording. If reinstantiation of original encoding operations is an important part of recognition, we would expect a high rate of false recognition of this type of sentence.

Experiment 1

Method

Subjects. The subject sample consisted of 30 students at the University of Colorado who participated in partial fulfillment of introductory psychology course requirements.

Design. The experiment followed a 2 X 2 X 2 X 2 mixed design with one between subjects factor, task demand (reading aloud, sentence continuation), and three within subjects factors, typography of first reading (inverted, normal), typography of second reading (inverted, normal), and test sentence wording (verbatim, paraphrase).

Materials. Stimuli consisted of two decks of sentences, a read deck and a recognize deck. All sentences appearing in the read deck were also in the recognize deck. A factorial combination of 3 within subjects factors resulted in 8 conditions to which materials were randomly assigned. They will be designated by a sequence of 3 letters. The first letter will represent the typography in which a sentence was read on its presentation in the read deck (_I_ for inverted, _N_ for normal), the second letter will represent the typography in which a sentence was read in the recognize deck (_I_, _N_), and the third letter will represent the wording of the sentence on its presentation in the recognize deck (_V_ for verbatim, _P_ for paraphrase). Eight sentences were assigned to each of the 8 conditions. If, when we are reporting results, the third letter is omitted, it means that we have collapsed over the levels of the test wording factor.

The read deck also contained a set of verbatim false sentences. These sentences appeared in the same typography in both the read and recognize decks, half in normal and half in inverted typography.

The recognize deck contained a set of new sentences in addition to the sentences that appeared in the read deck. Half of the new sentences were inverted and the other half were typed normally.

Unrelated sentences varying from 13 to 17 words in length were selected from a variety of sources of fiction and nonfiction. From this set of sentences 6 were set aside for use as buffer sentences, 3 at the beginning and 3 at the end of the read deck. Twenty-eight sentences were reserved for presentation as new sentences. For 45 of the remaining sentences paraphrased versions were written so that the wording of two semantically equivalent sentences was as dissimilar as possible (e.g., _When the police arrived they found the thieves had eluded them, leaving the room in shambles_, vs. _The burglars ransacked the room and made their escape before the police arrived at the scene_). Thirty-two students in a cognitive psychology course rated, on a 7-point scale, the degree to which the meanings of these 45 sentences pairs were equivalent. The 32 sentences that received the highest meaning equivalence ratings were chosen for use in the paraphrase conditions. Ratings for the sentences used in the study varied from 5.26 to 3.71, where 6 meant that the sentence meanings were exactly the same, 3 meant that the sentence meanings were moderately similar, and a rating of 0 meant that the meanings were entirely different. From the sentences remaining in the original set, 32 were chosen for use in the verbatim conditions.

Sentences were individually typed on cards. The typographic transformation that we used was the one that Kolers (1968) called *reversed inverted*. This transformation is obtained by rotating each letter 180 degrees around its vertical axis and 180 degrees around its horizontal axis. Each deck was shuffled before presentation so each subject received a different randomization of sentences. Each randomization of the read deck was preceded by the same 3 practice and 3 buffer sentences and followed by 3 more buffer sentences. Therefore the read deck contained 85 sentences (64 originals + 12 verbatim false + 6 buffers + 3 practice). Each randomization of the recognize deck was preceded by 3 practice sentences and thus contained 107 sentences (64 originals + 12 verbatim false + 28 new + 3 practice). In addition, a practice page of text typed in normal form and a practice paragraph typed in inverted form were taken from Miller (1962).

Procedure. Subjects participated individually in the 1.5 hr experimental session. They were asked to read aloud one page of text typed in normal form as practice. Their reading speeds were measured by stopwatch and recorded. This practice text also served as a screening device to ensure that all subjects were skilled readers. None of the practice reading times exceeded 2.5 min. In order to familiarize subjects with the typographic transformation, subjects were then given a paragraph in inverted typography to read aloud. Again, reading times were recorded.

The subjects were then randomly assigned to one of two instructional groups whose task demands were different. Fifteen subjects were instructed to read aloud the sentences in the read deck, and 15 subjects were instructed to perform a sentence continuation task designed to induce processing of sentence meaning (Bobrow & Bower, 1969). Subjects in this group were asked to read each sentence in the read deck silently and then make up a sentence which could logically follow the one they had read. Reading speed was measured and recorded for both groups. For the reading aloud group timing began when they turned over a card exposing a new sentence and ended when subjects finished reading the last word of each sentence. For the sentence continuation group timing began at the same time as for the reading aloud subjects, but ended when the subject began his/her sentence continuation. Reading errors were not recorded but subjects were told that if they made an error they were to reread that word correctly before continuing. If subjects could not decode a word they were asked about troublesome letters. Subjects were informed that some sentences would appear in normal typography and some would appear in inverted typography. However, they were not informed of the recognition test that was to occur.

After processing the sentences in the read deck under the appropriate task demands all subjects read aloud the sentences in the recognize deck and made three judgments on each sentence. Subjects told us whether or not the sentence they had just read generally meant the same as any sentence in the read deck. It was pointed out that some sentences might be worded differently but they would have the same meaning as a sentence that appeared in the read deck. Subjects then gave us a confidence rating from 1 (low) to 7 (high) on the meaning judgment. Finally, if a sentence was judged to mean the same as one in the first deck the subject told us whether it appeared in the recognize deck in the same or different typography as that in which it had appeared in the read deck. After making these judgments the subjects then went on to the next sentence. Subjects were informed of the various kinds of sentences that would appear in the recognize deck and could refer at all times to a card on which the judgments and rating scale were shown. In addition to recording the subjects' judgments, reading speed was measured and recorded for each sentence.

Analysis. In our signal detection analysis of sentence recognition we have categorized hits according to the three sentence variables manipulated in the present experiment. One false alarm rate was used for all hit rates and was based on the rate of false recognition of new sentences regardless of typography. This is contrary to Kolers' (e.g., Kolers & Ostry, 1974) technique of using inverted new sentences in calculating the false alarm rate for old sentences tested in normal form. We believe it best to combine the new sentences to form one false alarm rate because a rate based on more data points can be obtained and it is possible that a new normal or inverted sentence's meaning may be confused with an old sentence's meaning regardless of typography. In our analysis of meaning judgments, both methods of calculating false alarm rates were used and both produced the same pattern of results. We will therefore report in detail only the results obtained using our more general estimate of false alarms. The d' values based on sentence recognition will be labeled d' (sem).

Typography judgments were analyzed using NN sentences, judged to be in the same typography as when first read, as hits, and NI sentences judged to be in the same form were the corresponding false alarms. Memory for inverted typography was calculated in a parallel fashion: II sentences judged as being typed in the same form were hits and NI sentences judged to be in the same form were false alarms. This analysis combines the typography on first reading and on the judgment phase so that only two sentence variables are involved, typography and test wording. The d' values based on recognition of typography will be labeled d' (typ). Kolers and Ostry (1974) used a similar measure of memory for typography except the false alarm rate for inverted typography was based on NI sentences, and for normal typography false alarms were based on IN sentences. Their measure combines the subject's ability to recognize repetition of one form of typography (e.g., II) with his/her ability to recognize changes in the other form of typography (e.g., NI). Our measure is designed to provide a more homogeneous estimate of memory for each form of typography by combining recognition of repetition of one form of typography (e.g., II) with detection of changes in that same form (e.g., IN).

Hit rates of 1.0 and false alarm rates of 0.0 were transformed into more realistic values by use of the following estimate: 1 - (1/2N) for hit rates and 1/2N for false alarms, where N represents the total number of hits or false alarms possible.

Results and Discussion

In order to demonstrate the equivalance of the two groups of subjects, reading aloud and sentence continuation, on the task of reading normal and transformed text, analyses of variance of reading times were calculated for the two practice texts. Neither analysis revealed any difference between the groups. The mean reading times are presented in Table 1. A similar analysis was

----

Insert Table 1 about here

----

carried out on the time taken to read the first set of sentences. There was no difference between the groups. Of course, the normal sentences were read more swiftly than the inverted ones, F (1, 28) = 233.49, $p < .001$.

Recognition of Meaning. The ability of subjects to recognize the meaning of old sentences was reflected in the d' (sem) scores which appear in Table 2. An analysis of variance of these data

----

Insert Table 2 about here

----

revealed no overall difference between the reading aloud and sentence continuation conditions. However, there was a strong interaction between task demands and typography on first reading,

F (1, 28) = 15.93, $p < .001$. As hypothesized, the interaction was the result of similar recognition performance by the two groups on sentences that were originally read in inverted typography (mean d' for reading aloud was 2.67, and for sentence continuation the mean d' was 2.75) and the striking superiority of the sentence continuation subjects in recognizing originally normal sentences (mean d' for reading aloud was 1.86, and mean d' for sentence continuation was 2.49). Subjects who merely read aloud did not process meaning very deeply. Apparently, when a skilled reader reads aloud normally typed sentences, little semantic processing is needed to decode the words. Because the requirements of reading aloud do not require meaning to be deeply processed a semantic representation is less well established than under sentence continuation conditions. Subjects who performed the sentence continuation task were required to elaborately process meaning and, hence, their scores for correctly recognizing the meaning of originally normal sentences were much higher than those of reading aloud subjects. These results were anticipated on the basis of studies on depth of processing done by Craik and Tulving (1975) and Bobrow and Bower (1969). However, the inverted typography was hypothesized to induce more conceptually driven processing than would occur in the reading of normally typed text, leading us to predict the interaction between task demands and original typography. The interaction supports our interpretation of Kolers' findings which suggests that superior memory for sentences originally appearing in inverted typography occurs not only because processing of those sentences is more complex than that of normal sentences, but also because typography is elaborately processed with respect to sentence meaning in the extensive interaction of conceptually and data driven processes. That sentence continuation task demands did not succeed in boosting performance on originally inverted sentences and did succeed in attenuating the differences between inverted and normal typography, suggests that elaborate processing of meaning is what is induced by inverted typography and is the critical factor for sentence memory.

It could be argued that differences in recognition rates between inverted and normal typographies result merely from the fact that different amounts of time are spent processing each of them. Were this the case we would expect the performance of the two groups on the two typographies to parallel each other; the interaction cannot be accounted for by this argument since there were no differences between reading aloud and sentence continuation subjects in initial reading times--hence the superior performance of sentence continuation subjects on originally normal sentences cannot be explained by a processing time argument. Furthermore, Kolers (1973, 1974) has demonstrated that while different typographic transformations took different amounts of time to read, recognition of sentences read in the different typographies did not vary. While the arguments we are making here imply that there should be some relationship between difficulty of reading a transformation and the amount of conceptually driven processing that occurs, it may be that conceptually driven processing is an all-or-none phenomenon in terms of its effects on recognition or that the measures used to assess recognition have been insensitive to subtle differences in the amounts of conceptually driven processing that occur.

The analysis of d' (sem) scores also revealed a three way interaction between task demands, original typography, and test typography, F (1, 28) = 11.12, p < .005. This interaction indicates that the difference between the two groups of subjects in recognizing originally normal sentences was enhanced when the test sentence was inverted (d' difference = .79) compared to when the test sentence was normal (d' difference = .47).

Originally inverted sentences were recognized (M = 2.71) far better than originally normal sentences (M = 2.18), F (1, 28) = 59.59, p < .001. This replicates Kolers findings and provides further substantiation of the fact that processing of normal and inverted typography are qualitatively different. Verbatim sentences were recognized significantly better than paraphrases, F (1, 28) = 20.83, p < .001, although both were recognized at a high rate (mean for V = 2.56, mean for P = 2.33). That paraphrases were recognized at all suggests that what is stored in memory is more than a set of pattern analyzing procedures that are specific to the surface structure of a sentence. But the effect shows that while the semantic component of sentences is important and is indeed represented (allowing recognition of paraphrases to occur), under proper conditions surface structure is also remembered and used in recognition as Graesser and Mandler (1975), Kintsch and Bates (1977), and others have also shown.

Finally, there was an interaction involving all three sentence variables: original typography, test typography, and test wording, F (1, 28) = 5.83, p < .05. The relative ease of recognizing verbatim copies of old sentences held up across all combinations of original and test typography except for the case in which originally normal sentences were tested in inverted form. The mean d' (sem) for NIV sentences was 2.16 and for NIP sentences was 2.18. When an originally normal sentence was tested in inverted typography information about wording was not used in recognition judgments. The implication here is that processing of normal sentences is so automatic that the resulting memorial representations are poor in quality. Since meaning is not well represented, recognition of sentence meaning as old must rely more on surface features that are represented. When one surface component (typography) is changed there remains only one other surface component (wording) which can be used in matching. It seems that the change from normal to inverted typography precludes the use of surface feature information. Recognition could be aided by repetition of wording if the operations used to encode a normal sentence could be reinstantiated. Since processing is so automatic in this case there is little trace of encoding operations to be reinstantiated. (Note that this explanation implies that the verbatim/paraphrase distinction should be better represented for inverted sentences, and we test this prediction in Experiment 2.) Moreover, the change in typography entails such radically different processing that it may interfere with recognition of the repeated sentence wording. In any case, it is clear that information about surface structure resides in memory as a result of the present experimental manipulations and interacts with semantic components of sentence information in the process of recognition.

Confidence ratings, shown in Table 3, substantiated the major

---

Insert Table 3 about here

---

results of the analysis of meaning judgments. Subjects were more confident in judging the meaning of a sentence as old if it had originally been inverted (mean for I = 6.57, mean for N = 6.12), $\underline{F}$ (1, 28) = 35.24, $\underline{p}$ < .001. As expected, this was especially true for subjects who read aloud (mean for I = 6.57, mean for N = 5.86), since typographical transformation made a bigger difference in the amount of semantic processing activity than it did for sentence continuation subjects (mean for I = 6.57, mean for N = 6.38), as indicated by an interaction between task demands and original typography, $\underline{F}$ (1, 28) = 11.22, $\underline{p}$ < .005. Verbatim copies of sentences in the read deck were also more confidently accepted ($\underline{M}$ = 6.48) at test than paraphrases ($\underline{M}$ = 6.21), F (1, 28) = 19.38, $\underline{p}$ < .001, mirroring the main effect of test wording in the d' (sem) analysis. One unanticipated finding was that sentences tested in normal typography were more confidently accepted ($\underline{M}$ = 6.44) than those tested in inverted form ($\underline{M}$ = 6.25), $\underline{F}$ (1, 28) = 13.03, $\underline{p}$ < .002. It may be that subjects use a higher criterion for acceptance of normally typed sentences such that when they are accepted, they are very sure that the sentences were previously read in one form or another.

Recognition of Typography. Memory for the typography of correctly recognized sentences was estimated by the d' (typ) measure. The mean d' (typ) scores are given in Table 4. An analysis of

---

Insert Table 4 about here

---

variance indicated that subjects were far more accurate at detecting repetitions of or changes in inverted typography (mean d' = 2.41) than in normal typography (mean d' = 1.04), F (1, 28) = 27.42, p < .001. In congruence with the effects of meaning judgments resulting from the nature of the processing induced by inverted typography, this finding suggests that information about the unusual typography is represented in memory in some way. We argue that because the perceptual aspects of an inverted sentence are elaborately processed with respect to the meaning and wording of that sentence, the original typography will be better remembered. The elaborate processing involved the use of conceptually driven processes based on general knowledge of the world and language which interact with data driven processes based on the perceptual aspects of a sentence. Reading a normal sentence requires relatively less elaborate processing of perceptual information since the identification of normally typed words can occur at a reasonably automatic level.

Our claim that the word indentification process in reading involves the use of conceptually driven expectations about meaning and wording would lead to the prediction that memory for typography should be affected by the wording of the test sentence. Repetition of wording, as in verbatim test sentences, should enhance memory for typography scores. This prediction was confirmed by our analysis of d' (typ) scores as testing with verbatim sentences produced greater recognition performance (M = 1.91) than did testing with paraphrases of original sentences (M = 1.55), F (1, 28) = 7.60, p < .02.

The d' (typ) analysis also revealed an interaction between task demands, typography, and test wording, F (1, 28) = 5.60, p < .025. Subsidiary analyses showed that this interaction was a result of the verbatim test sentences allowing higher recognition of typography for both inverted and normal sentences among reading aloud subjects. The effectiveness of repeating sentence wording held up only for originally inverted sentences among sentence continuation subjects. For these subjects memory for the typography of normal sentences was not improved by testing with verbatim sentences. This interaction supports our idea that reading aloud subjects' representations of information about normal sentences is more dependent on surface features than is the case for sentence continuation subjects, and that the two groups of subjects have similar representations of information about inverted sentences.

Reading Time. Sentences were randomly assigned to conditions of typography and wording rather than completely counterbalanced. Therefore we confined our analysis of the sensitive reading time measure to time saved on second reading of sentences. Table 5 presents the mean reading time saved on the second reading of

---

Insert Table 5 about here

---

correctly recognized sentences. These data are based on the difference between time taken on the first and second readings of a sentence which appeared in the same typography on both occasions. Hence, the analysis is restricted to II and NN sentences. Inverted sentences showed a greater savings than did normal sentences (means of 13.33 sec and 0.37 sec, respectively), F (1, 28) = 143.17, p < .001, since savings on normal sentences represent a floor effect. Normal sentences could not show savings due to their reading times being at a minimum on both the original reading and test phases of the experiment. As would be expected by Kolers' notions about pattern analyzing operations, verbatim sentences produced greater savings than did paraphrases, F (1, 28) = 15.32, $p < .001$. Normal sentences could not show this savings effect due to their reading times being minimal. Therefore a typography by test wording interaction obtained, F (1, 28) = 14.05, $p < .001$, in which inverted verbatim sentences showed more savings (M = 15.50 sec) than did inverted paraphrases (M = 11.16 sec).

Although rereading inverted verbatim sentences produced greater savings in reading time, rereading inverted paraphrase test sentences achieved 72% of the savings obtained with verbatim sentences. In both cases the amount of savings for inverted sentences was greater than zero, F (1, 29) = 162.56, $p < .001$, and F (1, 29) = 84.32, $p < .001$, for verbatim and paraphrase test sentences, respectively. Despite the fact that their reading entailed the application of a new set of pattern analyzing operations, considerable amounts of time were saved. This is consistent with the idea that semantic similarity of an original sentence and its paraphrased version was a significant determinant of time taken to read the test sentence. Subjects seem to make effective use of their memory for the meaning of sentences on second reading. Memory for meaning could come into play in helping subjects to streamline their use of conceptually driven processes. If the subject recognizes the meaning of a sentence while reading it, that meaning can be used to suggest specific predictions about what the sentence says and those predictions are likely to be quite accurate. The interaction between conceptually and data driven processes can be refined and made more efficient when memory for meaning is reliable. Consequently, changes in wording do not greatly reduce the ability to reread a recognized sentence.

<u>Verbatim False Sentences</u>. The ability to detect sentences that were nearly verbatim copies of previously read sentences, except for a minor change in wording that altered the meaning, was generally poor. Correct rejection of verbatim false sentences occurred in only 64% of the cases. Normally typed verbatim false sentences were rejected more often (68%) than were inverted verbatim false sentences (60%), $\underline{F}$ (1, 28) = 5.36, $\underline{p} < .05$. Why this should be the case if inverted sentences are better represented in memory is not clear. It may simply be that the changes in meaning were so slight that they were very difficult to detect. Even so, it is necessary to consider what the basis of rejection was if we are to understand the effect. If forgetting of sentences was an important factor we would expect the forgetting rate (and, hence, rejection rate) to be low overall, but to be higher for normally typed sentences (whose representations are poor and more susceptible to forgetting) than for inverted sentences, which is what happened. Another likely possibility is that all of the informational components interact in the judgment of meaning. Since typography was the same on both presentations of verbatim false sentences, and syntax and wording were the same except for one or two words, the match on surface features (which was elaborately represented for inverted sentences) may have induced subjects to accept the sentences as old. The idea here is that the preponderance of

memorial evidence indicated that a verbatim false sentence was an old one. The fact that subjects in the reading aloud condition realized that the typography of inverted verbatim false sentences was repeated (95%) more often than they did for normal verbatim false sentences (5 [illegible]%), $\underline{X}^2$ (1) = 13.47, $\underline{p}$ .001, supports this notion.

Evidence for the possibility that acceptance of verbatim false sentences was affected by memory for surface structure comes from two other sources. First, in a study of the verbatim/paraphrase factor in normally typed sentences, Toglia (Note 2) found that verbatim false sentences were rejected at a low rate (55%) as in our experiment, while paraphrase false sentences were rejected at a higher rate (88%). This suggests that when wording or syntax is similar on both presentations it sometimes precludes the detection of small changes in meaning, resulting in a low rejection rate. When the surface structure differs on the two presentations, as for the paraphrase false sentences, rejection rates are higher. Second, Thorndyke (1977) also found that subtle meaning changes were rejected at the rate of 55%. Thus, it seems that we again have evidence that meaning and surface features (possibly through the reinstantiation of encoding operations) interact in the process of recognition.

New Sentences. When new sentences were accepted as old during the test phase, subjects provided typography judgments just as they did when they correctly recognized an old sentence. Sentence continuation subjects were more likely (77%) to claim that normally typed new sentences had a different typography than they were to claim that the supposedly original typography had been retained (23%), and for inverted new sentences they claimed that the original typography had been retained more often (76%) than they said it had changed (24%), X^2 (1) = 1[illegible].30, $p < .001$. Subjects in the reading aloud condition distributed their claims about same or different typography equally across normal and inverted new sentences. The tendency for sentence continuation subjects to believe that a new sentence was originally typed in inverted form could reflect the fact that more originally inverted sentences were successfully represented in memory than originally normal sentences. If a new sentence were to be confused with an old one it is more likely that it would be confused with a previously inverted sentence.

Experiment 2

In the first experiment subjects provided us with explicit judgments of meaning and typography, but did not make direct judgments of whether a sentence in the recognize deck was a verbatim copy or paraphrase of a sentence in the read deck. Therefore, Experiment 2 was designed not only as an attempt to replicate the findings of Experiment 1, but as a means of gathering more explicit information about the memorial retention of surface structure and its importance, relative to other sentence components, in the process of sentence recognition. One prediction coming out of Experiment 1 is that the verbatim/paraphrase distinction would be better represented for inverted sentences if meaning and wording are more elaborately processed when inverted typography is read. Experiment 2 allowed us to test this expectation directly.

In addition, Experiment 2 permitted us to look more carefully at the improvement in reading speed on sentences in the recognize deck and at the contributions of typography, wording, and meaning, to this facilitation. Because sentences were randomly assigned to conditions rath r tha counterbalanced in Experiment 1, we were not justified in making direct comparisons of reading time for the sentences in the recognize deck. The sensitivity of reading time to small variations in sentence length and wording suggests the need for counterbalancing.

The methodology of this study was similar to Experiment 1, except that only reading aloud instructions were used, subjects made explicit judgments about the wording of a sentence in the recognize deck (in addition to judgments of meaning and typography), and sentences were counterbalanced rather than randomly assigned to conditions.

Method

Subjects. Participants were 16 adult volunteers from the Boulder community. They were paid at the rate of $2.00 per hour.

Design. The design was a 2 x 2 x 2 within subjects design including the factors of typography of first reading (I, N), typography of second reading (I, N), and test sentence wording (V, P). We omitted the between subjects factor of Experiment 1 (task demand), using reading aloud instructions for all subjects.

Materials. The stimulus materials from Experiment 1 were used for this experiment as well, except that verbatim false sentences were not included.

In addition, sentences were counterbalanced using two 4 x 4 Latin squares to rotate blocks of 8 sentences through conditions.

One Latin square was used for assignment of blocks of 8 sentences to verbatim conditions (IIV, INV, NNV, and NIV), and one square was used for assignment of blocks of 8 sentence pairs (each sentence in the paraphrase condition had two versions), to paraphrase conditions (IIP, INP, NNP, and NIP). Each of the four versions of the read deck were fully specified by a column from the verbatim Latin square and the corresponding column from the paraphrase Latin square. Thus, each verbatim sentence appeared once in each of the 4 verbatim sentence conditions and each paraphrase pair appeared once in each of the 4 paraphrase sentence conditions.

Since there were two semantically equivalent versions of each sentence in the paraphrase conditions (sentence A and sentence B), it was necessary to control the order in which these sentences were read. Hence, there were two versions of each of the 4 arrangements of the decks, an A version, in which sentence A of all paraphrased pairs appeared in the read deck and sentence B appeared in the recognize deck, and a B version, in which sentence B of all paraphrased pairs appeared in the read deck and sentence A appeared in the recognize deck. Recall that for verbatim sentences, an exact replica of the sentence appeared in both decks. Thus, there were a total of 8 arrangements of the decks. Each deck arrangement was presented to 2 subjects, differently randomized for each of them.

Procedure. As in Experiment 1, subjects read, for practice, a page of normally typed text and a paragraph of inverted text. Their reading times were recorded. All subjects were instructed to read the sentences in the read deck aloud. Sentences in the recognize deck were also read aloud and, after reading each one, subjects made the following judgments: whether or not a sentence generally meant the same as one in the first deck, and if so, whether it was a paraphrased or verbatim copy of that sentence, and whether it was in the same or different typography. As in Experiment 1, reading speed was measured and recorded for each sentence in both decks, and subjects' judgments were recorded for each sentence in the recognize deck.

Analysis. In addition to the two signal detection measures used in Experiment 1, we have added a third in order to assess subjects' memory in making verbatim/paraphrase distinctions. This measure will be called d' (VP). A hit occurred when a subject correctly stated that the wording of a recognized old sentence had not changed. A false alarm was made when a paraphrased version of an old sentence was recognized but the subject claimed that the wording had not changed.

As in Experiment 1, whenever hit or false alarm rates of 1.0 or 0.0, respectively, occurred, they were converted to more meaningful estimates: 1 - (1/2N) for hits and 1/2N for false alarms, where N equals the maximum number of hits or false alarms possible.

Results and Discussion

An analysis of the first reading times for all eight sentence types revealed that our counterbalancing procedure successfully equalized the groups of sentences on the reading time measure. As expected, the only variable which proved to have a significant effect on first reading times was original typography, with inverted sentences requiring more reading time (M = 39.3 sec) than normally typed sentences (M = 5.9 sec), F (1, 15) = 139.96, $p < .001$. All other Fs were less than 1.1. Therefore, the effects reported below can be interpreted as resulting from experimental manipulations rather than from sentence specific variables.

Recognition of Meaning. In terms of the recognition of sentence meaning, all major findings of Experiment 1 were replicated in Experiment 2, except one--the three way interaction of original typography, test typography, and test wording. Table 6 presents the mean d' (sem) scores for each sentence type. An analysis of variance performed on these scores revealed that originally

---

Insert Table 6 about here

---

inverted sentences ($M = 2.63$) were better recognized than were originally normal sentences ($M = 1.76$), F (1, 15) = 55.36, $p < 001$, replicating both Kolers' findings and our own findings in Experiment 1. Again, we propose that the superior recognition of originally inverted sentences results from the elaborate, meaning based memorial representations formed during semantic processing of the transformed sentences. Because the inverted typography induces extensive processing of sentence meaning during decoding, its memorial representation is of higher quality when sentences are inverted than when sentences are normally typed and automatically processed. Consequently, the meaning of originally inverted sentences is recognized more often than that of originally normal sentences.

As in Experiment 1, we also found that old sentences tested in verbatim form (mean d' (sem) = 2.41) were recognized more often than sentences tested in paraphrased form (mean d' (sem) = 1.98), F (1, 15) = 30.66, $p < .001$. Despite the fact that reading paraphrased test sentences does not involve repetition of pattern analyzing operations, paraphrases were recognized at a fairly high rate. Therefore, recognition cannot be completely dependent on specific pattern analyzing operations but is also affected by memory for meaning. A significant interaction between typography on reading and test phases, $F\ (1, 15) = 5.15$, $p < .05$, indicated furthermore that recognition of meaning was improved when a sentence was read and tested in the same typography. The mean d' (sem) for II sentences was 2.69 compared to 2.56 for IN sentences. Similarly, this mean for NN sentences was 1.87 and for NI sentences the mean was 1.65. These effects demonstrate the interactive nature of the processing that occurs both during reading, when sentences are originally encoded, and during recognition, when test sentences are matched to memorial representations. Recognition of sentence meaning is influenced by whether or not surface structures of the original and test sentences match, and by whether or not the typographies of the original and test sentences match. Thus, the matching process that occurs during recognition is not one in which sentence characteristics (meaning, wording, typography) are considered independently, but one in which all three components interact and influence decisions about any particular component, such as sentence meaning.

This concept of interactive processing is important in understanding why the three way interaction of original typography, test typography, and test wording was not significant as it was in Experiment 1. Recall that, in Experiment 1, this interaction resulted from the lack of a difference in recognition between

NIV and NIP sentences (recognition of verbatim sentences was superior to recognition of paraphrased sentences for all other combinations of typography). We argued that NIV and NIP sentences were recognized at approximately the same rate because, when originally normal sentences were tested in inverted type, subjects did not use their memory for the wording of these sentences in making judgments about meaning, perhaps because of the difficulty created by this change in typography. It may be that this interaction disappeared in Experiment 2 because the additional task demand of making explicit judgments about sentence wording induced subjects to use memorial information about surface structure even for NIV and NIP sentences in making judgments about meaning.

If making verbatim/paraphrase judgments accentuated the importance of sentence wording during the recognition phase, it might be expected that this would be reflected in recognition of sentence types other than NI sentences. The expectation is confirmed when the difference in recognizability of NNV and NNP sentences is considered. In Experiment 1, the mean d' (sem) scores for reading aloud subjects on NNV sentences were 12% greater than on NNP sentences. This difference grew to 43% in Experiment 2, supporting our hypothesis that making wording judgments caused subjects to be more aware of and to make more use of their memory for sentence wording. This interpretation points up the interplay between memory for meaning, wording, and typography during recognition.

Another possible explanation for our failure to obtain the three way interaction in Experiment 2 is that the effect may have been an artifact of failing to counterbalance in Experiment 1. Besides being theoretically uninteresting, this hypothesis would lead one to expect that other results in Experiment 1 were artifactual and should therefore not have replicated in the second experiment. However, the extensive similarities between the results of the two experiments suggest that failing to counterbalance in Experiment 1 was inconsequential, at least for the judgments that subjects made about sentences. As might be expected, counterbalancing proved to be more important for reading times, as will later be discussed.

Recognition of Typography. Memory for the typography of recognized sentences was estimated with a d' (typ) measure different from Kolers', as described previously. The mean d' (typ) scores for each sentence type are given in Table 7. With this measure, the

---

Insert Table 7 about here

---

main effect of typography was highly significant, F (1, 15) = 41.81, $p < .001$ indicating that the typography of originally inverted sentences (M = 2.57) was recognized significantly more often than the typography of originally normal sentences (M = 0.41). It should be pointed out that when d' (typ) scores are computed using Kolers' method, this main effect is concealed, though all other effects relating to typography remain significant. We feel that our measure is more representative of the effects that typography has during reading and recognition, insofar as information about the typography of originally inverted sentences is more elaborately encoded than is the typography of originally normal sentences.

The analysis of variance of d' (typ) scores also showed that the typography of an old sentence was better recognized if the sentence was tested with a verbatim (M = 1.74) copy as opposed to a paraphrase (M = 1.24), F (1, 15) = 15.26, p < .005. This **result** is a clear indication of the interdependence of the memorial representations of different sentence components.

Recognition of Wording. Mean d'(VP) scores are presented in Table 8 and represent the subjects' ability to determine whether

---

Insert Table 8 about here

---

a correctly recognized test sentence is a verbatim copy or a paraphrase of an old sentence. An analysis of variance yielded one significant effect. Subjects' memory for sentence wording was more accurate for originally inverted sentences (M = 2.26) than for originally normal ones (M = 0.81), F (1, 15) = 43.60, p < .001. Thus, the judgments which subjects made concerning sentence wording provide further evidence for the interaction between memory for wording and typography. The fact that d'(VP) scores were higher for sentences which were originally read in inverted typography than those in normal type leads to the conclusion that, when an inverted sentence is read, subjects process and encode sentence wording and syntax more extensively than they do when reading normal sentences.

Reading Time. The influence of semantics in the reading process was forcefully demonstrated in the reading time results of Experiment 2. Reading time savings based on the difference between first and second readings of a sentence were calculated for those sentences which were read in the same typography on both occasions. The mean savings for correctly recognized II and NN sentences are presented in Table 9.

Insert Table 9 about here

Inverted sentences ($\underline{M}$ = 15.68 sec) showed greater savings than normal sentences ($\underline{M}$ = 0.08 sec), $\underline{F}$ (1, 15) = 68.68, $\underline{p}$ < .001, since the first reading of a normal sentence cannot usually be improved upon. There was no difference in the amount of savings for verbatim and paraphrase test sentences, $\underline{F}$ < 1. Ninety percent of the amount of savings obtained with verbatim sentences was found with paraphrases. Both amounts of savings for II sentences were significantly more than zero, $\underline{F}$ (1, 15) = 54.80, $\underline{p}$ < .001, and $\underline{F}$ (1, 15) = 42.00, $\underline{p}$ < .001, for verbatim and paraphrase sentences respectively.

Thus, the savings attained by subjects was statistically the same for sentences tested with verbatim versus paraphrase versions. Despite the fact that there is no repetition of surface structure (wording, syntax, or pattern analyzing operations), when an IIP test sentence is read, it is read with the same speed as an IIV sentence which does repeat surface structure. The overriding factor must therefore be the general semantic representation underlying the two sentence types. It is clear that this representation greatly influences the reading process.

The results of Experiment 1 indicated that paraphrase test sentences produced significantly less savings than verbatim sentences. This effect was likely due to the fact that in Experiment 1 we did not counterbalance paraphrase sentence versions across encoding and test phases. If the versions used on the test phase inherently required longer reading times, then savings would be artificially reduced. In Experiment 2, we counterbalanced sentence versions and eliminated the verbatim/paraphrase effect found in the first experiment. It is obvious that reading time is physically constrained by number and length of words. Savings scores are affected by this fact and, consequently, the counterbalancing procedure used in Experiment 2 provides a set of results which is free of this problem.

The analysis of second reading times also supported our notions about the importance of the use of semantic knowledge in reading. Since our analysis of first reading times of sentences indicated comparability of the sentence groupings on that measure, we analyzed the raw reading times of correctly recognized sentences on the test phase. The mean reading time for each sentence type is shown in Table 9. In general, originally inverted sentences ($\underline{M}$ = 14.63 sec) were reread faster than originally normal ones ($\underline{M}$ = 17.26 sec), $\underline{F}$ (1, 15) = 52.42, $\underline{p}$ < .001. Again, the verbatim/paraphrase variable was not significant, showing that, among correctly recognized sentences, testing with verbatim versus paraphrase versions did not affect reading time.

We believe that originally inverted sentences were reread more swiftly than normal sentences for two reasons. First, reading an inverted sentence would have resulted in a more reliable and elaborate encoding of the meaning of the sentence as evidenced by our d' (sem) results. Second, reading inverted sentences involved the formation and use of a more detailed representation of surface structure. Evidence for this second notion can be found in our results involving memory for typography and wording. The d'(VP) scores, for example showed that wording of inverted sentences is better recognized than that of normal sentences. Furthermore, the analysis of second reading times revealed an interaction between original and test typography, F (1, 15) = 7.30, $p < .02$. This interaction indicated that II sentences (M = 23.35 sec) were reread more swiftly than NI sentences (M = 28.59 sec), with sentences tested in normal type showing a floor effect. This supports our ideas about the more elaborate representation of inverted sentences as compared to normal sentences.

New Sentences. Subjects made typography and sentence wording judgments on distractor sentences which they falsely recognized. Analyses of their judgments about falsely recognized new sentences provide some insight into how false alarms come about under these experimental conditions. Analysis of typography judgments on falsely recognized new sentences showed that subjects indicated that normal new sentences had a different typography more often (74%) than the same typography (26%), while for inverted new sentences subjects were more likely to claim that the typography was the same (71%) than that it had changed (29%), X^2 (1) = 11.67, $p < .001$. As in Experiment 1, subjects thought that a falsely recognized new sentence meant the same as an originally inverted sentence. This finding is consistent with the idea that inverted sentences are more likely to be remembered and represented in memory than normal sentences and, consequently, will more often cause incorrect meaning matches. That is, when a subject attempts to match the meaning of a new sentence with a representation of an old sentence, he will be making more comparisons involving inverted old sentences simply because there are more of them in memory. Second, a similar analysis of sentence wording judgments revealed that subjects were far more likely to claim that a falsely recognized new sentence was a paraphrased version (84%) of an old sentence than they were to claim that it was a verbatim copy (16%), $\underline{X}^2$ (1) = 27.59, $\underline{p} < .001$. Since mismatches during recognition are probably semantically based, and since there is little similarity between the wording of old and new sentences (which in fact have different meanings), subjects concluded that the wording of falsely recognized new sentences was not the same. Therefore, false alarms were almost always judged to be paraphrases of old sentences. The typography of the new sentences did not affect these judgments of wording.

General Discussion

Our results point out the importance of semantic information for sentence comprehension and recognition. In Experiment 1 we demonstrated the necessity of encoding sentences to a deep semantic level (Craik & Lockart, 1972; Craik & Tulving, 1975) in order to achieve a durable representation. Kolers (1975a) attributed his failure to find large improvements in reading inverted sentences which had first been read in normal typography to the relatively small role played by semantics in the reading of inverted sentences. Our results attest to the fact that Kolers' procedure left subjects with an impoverished semantic representation of normal sentences. The reading time results of Experiment 2 are a particularly striking demonstration of the powerful effects of semantics on the reading process. The extensive processing of inverted sentences allowed large improvements on second reading even when wording and syntax (hence, pattern analyzing operations) were quite different. In view of these findings, Kolers' arguments (e.g., Kolers, 1976b) against emphasis on linguistic representations seem to lose much of their force.

Contrary to Kolers' arguments that improvement in reading time results from the reapplication of the procedures that mediated its encoding, it appears that while it may be the reapplication of procedures that is responsible for the improvement, the critical procedures are <u>semantic</u> in nature. It may be that a record of pattern analyzing operations is indeed retained but that the interactive analyses of the conceptually driven and data driven processes form the "core" or most important part of this representation. It is also possible that knowledge of sentence meaning is represented in abstract propositional form (Kintsch, 1974) and is thus relatively independent of sentence specific surface structures. Our results suggest that if meaning is propositionally stored it is not strictly independent of surface structure--as we have seen, the informational components interact with each other in the process of recognition. One or the other semantic explanation must be invoked to explain how paraphrases can be read with essentially the same amount of savings shown on verbatim test sentences. The perceptual aspects of processing appear to be decidedly less important to processing and representation than the linguistic and semantic aspects.

We are not, however, advocating a purely semantic approach to the analysis of comprehension and recognition. Our results also point to the existence and utilization of memorial representations of various aspects of surface structure. Recognition of semantic content was significantly influenced by memory of wording and typography. We found evidence for improved recognition when wording and pattern analyzing operations were repeated, indicating that these surface aspects are reliably encoded as Kolers has argued. Yet recognition was affected more by the quality of the original encoding than by repetition of wording or typography. That is, the strongest effect was the main effect of original typography: thorough processing of inverted sentences resulted in recognition performance superior to that of superficially processed normal sentences. Repetition of wording was not as powerful a factor, and repetition of typography (pattern analyzing operations) was least effective. These findings, along with our results on memory for surface structure, are consistent with our ideas about the interactive nature of semantic, syntactic, and visual processing and the overriding importance of semantics.

We believe that Kolers has been quite correct in pointing out that the processes involved in reading normal and inverted sentences differ markedly. However, our data are not consistent with his view that graphemic pattern analyzing operations alone underlie the processing differences. We offer a different explanation which is consistent both with the data that Kolers has obtained and with the data presented here. Reading sentences in normal typography does not involve extensive visual processing or perceptual analysis. Identification of normally typed words is a somewhat automatic process (cf. Shiffrin & Schneider, 1977) which does not require the careful visual analysis devoted to the reading of inverted sentences. This distinction is consistent with Kolers' (1975b) description of the effects of extended practice with inverted typography: the identification process became automated and recognition performance declined. Lockhart et al. (1976) also view the process of automatization as a reduction in the number of operations required to satisfy some performance criterion, in this case word recognition. Shiffrin and Schneider's (1977) recent finding that automatically processed stimuli are poorly retained parallel our results concerning poor recognition of normal sentences which are simply read aloud.

The controlled processing (cf. Shiffrin & Schneider, 1977) which occurs during the reading of inverted sentences seems to involve a number of different but interrelated processes. In performing this task, the subject must carry out detailed perceptual analyses of the sentence at a word by word or even letter by letter level. These seem to be the pattern analyzing operations emphasized by Kolers. Our view of the procedure is that the results of this type of analysis provide perceptual data which can be used to verify or reject certain aspects of conceptually driven processing. Conceptually driven processes are extremely important in our framework. We believe that a purely perceptual analysis (isolated from semantic or linguistic analysis) of the type thus far proposed by Kolers is not reasonable nor consistent with our data.

A more realistic explanation would include a heterarchical interaction between conceptually and data driven processes (cf. Schank, 1973) in which perceptual information is used almost exclusively until sufficient contact has been established to initiate predictions of a syntactic and semantic nature. Once the context has begun to form, the subject is able to predict with some accuracy the grammatical category to which the next word in the sentence should belong. The subject also has some idea about the identity of certain words based on semantic constraints imposed by the content and general world knowledge. The source of these semantic and syntactic predictions is the subject's general world knowledge or semantic memory and knowledge of grammar and syntax. The visual pattern analysis disconfirms or verifies the predictions and in this way the conceptually and data driven processes interact. The quality of the sentence's context will dictate the extent to which conceptually driven processing will be able to provide accurate and specific predictions. There should be a direct relationship between quality of the context, efficiency of conceptually driven processing, and, hence, speed of reading inverted sentences.

We argue that in our experiments the representation of semantic and surface aspects of a sentence come about at least partly as a consequence of memory for the interaction between conceptually and data driven processes that occur during reading. Surface features of a sentence are elaborated with respect to the semantic content of a sentence, as evidenced by the effects of memory for wording and typography on meaning judgments. Furthermore, judgments about wording and typography were affected in ways consistent with our ideas about how normal and inverted sentences are processed and about the nature of the resulting representations.

Rather than attempting to isolate the construct of pattern analyzing operations from ideas of semantic and linguistic representation, as Kolers has done, we believe that it is more productive (theoretically and practically) to interpret perceptual and semantic analyses as interactive processes. Kolers has succeeded in demonstrating the importance of considering the nature of pattern analyzing operations and their effects on memory. We have attempted to demonstrate how pattern analyzing operations can be conceptualized as part of a general framework for comprehension and memory in which semantic information processing is emphasized.

Reference Notes

1. Rumelhart, D. E. _Toward an interactive model of reading_. (Tech. Rep. CHIP 56) La Jolla, California: University of California, San Diego, March, 1976.

2. Toglia, M.P. _Verbatim and paraphrase memory for sentences_. Unpublished doctoral dissertation, Department of Psychology, University of Colorado, 1976.

References

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding*. New York: Academic Press, 1975.

Bobrow, S. A., & Bower, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology*, 1969, *80*, 55-61.

Cooper, F. S. How is language conveyed by speech? In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, Mass.: M. I. T. Press, 1972.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 671-684.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, *104*, 268-294.

Gough, P. B. One second of reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye.* Cambridge, Mass.: M. I. T. Press, 1972.

Graesser, A., & Mandler, G. Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *1*, 238-248.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W., & Bates, E. Recognition memory for statements from a classroom lecture. _Journal of Experimental Psychology: Human Learning and Memory_, 1977, 3, 150-159.

Kolers, P. A. The recognition of geometrically transformed text. _Perception and Psychophysics_, 1968, _3_, 57-64.

Kolers, P. A. Remembering operations. _Memory & Cognition_, 1973, _1_, 347-355.

Kolers, P. A. Two kinds of recognition. _Canadian Journal of Psychology_, 1974, _28_, 51-61.

Kolers, P. A. Specificity of operations in sentence recognition. _Cognitive Psychology_, 1975, _7_, 289-306. (a)

Kolers, P. A. Memorial consequences of automatized encoding. _Journal of Experimental Psychology: Human Learning and Memory_, 1975, _1_, 689-701. (b)

Kolers, P. A. Reading a year later. _Journal of Experimental Psychology: Human Learning and Memory_, 1976, _2_, 554-565. (a)

Kolers, P. A. Pattern analyzing memory. _Science_, 1976, _191_, 1280-1281. (b)

Kolers, P. A., & Ostry, D. J. Time course of loss of information regarding pattern analyzing operations. _Journal of Verbal Learning and Verbal Behavior_, 1974, _13_, 599-612.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. _Cognitive Psychology_, 1974, _6_, 293-323.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. Depth of processing, recognition and recall. In J. Brown (Ed.) _Recognition and recall_. London: Wiley, 1976.

Miller, G. A. Psychology: The science of mental life. New York: Harper & Row, 1962.

Neisser, U. Cognitive psychology. New York: Meredith, 1967.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

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

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

Smith, F. Understanding reading. New York: Holt, Rinehart, & Winston, 1971.

Thorndyke, P.W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Footnote

Order of authorship is alphabetical to reflect the equal contributions of both authors.

This research was supported in part by National Science Foundation grant BNS 72-02084 to the Institute for the Study of Intellectual Behavior and this is publication No. of the Institute. The research reported here was carried out while M. Masson held a National Research Council of Canada postgraduate scholarship. Preliminary reports of this research were presented at the meetings of the Rocky Mountain Psychological Association, May, 1977, and at the meetings of the Canadian Psychological Association, June, 1977.

The authors are indebted to Walter Kintsch for his advice and encouragement during our work on this project. We are also grateful to Lyle E. Bourne, Jr., Walter Kintsch, and Peter G. Polson for their critical comments on earlier versions of this paper, and to many other colleagues for stimulating and enlightening discussions.

Requests for reprints should be sent to Linda S. Sala, Department of Psychology, University of Colorado, Boulder, Colorado 80309.

Table 1

Mean Reading Times (sec) for Practice Paragraphs and Sentences Used in Read Deck as a Function of Task Instructions

		Task Instructions	
Material	Typography	Reading Aloud	Sentence Continuation
Practice	Inverted	340.4	348.4
Paragraphs	Normal	110.6	103.9
Read	Inverted	32.1	28.9
Sentences	Normal	5.9	6.9

Table 2

Mean Recognition Performance (d')[a] for Original Sentences as a Function of Task Instructions and Sentence Characteristics

	Original Typography							
	Inverted				Normal			
	Test Typography							
	Inverted		Normal		Inverted		Normal	
	Test Wording[b]							
Task Instructions	V	P	V	P	V	P	V	P
Reading Aloud	2.83	2.56	2.85	2.43	1.72	1.83	2.06	1.85
Sentence Continuation	2.85	2.48	2.90	2.76	2.60	2.52	2.68	2.17

[a]See text for an explanation of the derivation of the d' measure.

[b]V = Verbatim; P = Paraphrase.

Table 3

Mean Confidence Ratings for Original Sentences as a Function of Task Instructions and Sentence Characteristics

	Original Typography							
	Inverted				Normal			
	Test Typography							
	Inverted		Normal		Inverted		Normal	
	Test Wording[a]							
Task Instructions	V	P	V	P	V	P	V	P
Reading Aloud	6.78	6.21	6.74	6.54	5.84	5.56	6.13	5.90
Sentence Continuation	6.70	6.37	6.74	6.48	6.36	6.17	6.54	6.43

[a]V = Verbatim; P = Paraphrase.

Table 4

Mean Performance (d')[a] on Recognition of Typography of Correctly Recognized Sentences

	Typography			
	Inverted		Normal	
	Test Wording[b]			
Task Instructions	V	P	V	P
Reading Aloud	2.65	2.24	1.24	0.63
Sentence Continuation	2.74	2.02	0.99	1.31

[a]See text for an explanation of the derivation of the d' measure.

[b]V = Verbatim; P = Paraphrase.

Table 5

Mean Reading Time (sec) Saved on Second Reading of Correctly Recognized Sentences[a]

	Typography			
	Inverted		Normal	
	Test Wording[b]			
Task Instructions	V	P	V	P
Reading Aloud	16.7	11.6	0.0	-0.3
Sentence Continuation	14.2	10.8	1.0	0.8

[a]This analysis is restricted to sentences which appeared in the same typography (II and NN) at encoding and test phases.

[b]V = Verbatim; P = Paraphrase.

Table 6

Mean Recognition Performance (d')[a]

for Original Sentences

	Original Typography			
	Inverted		Normal	
	Test Typography			
Test Wording	Inverted	Normal	Inverted	Normal
Verbatim	2.86	2.78	1.80	2.20
Paraphrase	2.52	2.35	1.51	1.54

[a]See text for an explanation of the derivation of the d' measure.

Table 7

Mean Performance (d')[a] on Recognition of Typography of Correctly Recognized Sentences

	Typography	
Test Wording	Inverted	Normal
Verbatim	2.82	0.66
Paraphrase	2.32	0.15

[a]See text for an explanation of the derivation of the d' measure.

Table 8

Mean Recognition Performance (d')[a] on Recognition of Wording of Correctly Recognized Sentences

	Original Typography	
Test Typography	Inverted	Normal
Inverted	2.36	0.61
Normal	2.16	1.00

[a] See text for an explanation of the derivation of the d' measure.

Table 9

Mean Total Reading Time (sec) and Reading Time (sec) Saved[a]

on Second Reading of Correctly Recognized Sentences.

	Original Typography							
	Inverted				Normal			
	Test Typography							
	Inverted		Normal		Inverted		Normal	
	Test Wording[b]							
	V	P	V	P	V	P	V	P
Reading Time	21.4	25.3	5.8	6.0	27.4	29.7	5.8	6.1
Savings	16.5	14.9	---	---	----	----	0.4	-0.3

[a]Savings were calculated only for those sentences appearing in the same typography (II and NN) at encoding and test phases.

[b]V = Verbatim; P = Paraphrase.