Memory and Levels of Processing in a Psycholinguistic Context

William Marslen-Wilson and Lorraine K. Tyler
Department of Behavioral Sciences, University of Chicago

A levels of processing approach to memory was directly tested in a well-defined processing context by varying the depth to which spoken prose material could be processed. The effects of successively removing the semantic and the syntactic levels of analysis produced decrements in immediate recall that were structured according to the structural segmentation of the material. The results corresponded closely to the joint predictions of a theory of sentence processing integrated with levels of processing memory theory.

The levels of processing approach to memory is, on the face of it, elegantly simple and plausible. As Craik and Lockhart (1972) point out, the perceiver is normally concerned only with extracting the meaning from his sensory input. Equally, it is important for him to continue to remember these meanings. But the preliminary stages of processing necessary to reach the level of comprehension are of little value in themselves, once the input has been processed to this level, and need no longer be retained. The memory system, therefore, is organized primarily to remember meaningful material, so that the nature and the persistence of a memory trace will be a direct function of the depth to which the original stimulus was processed—whether, and to what extent, meaning was extracted from it.

However, for the levels of processing approach to have proper explanatory value, and for it to be testable, it needs to be combined with a theory of perceptual processing, in terms of which the notion of "level of processing" can be independently and unambiguously defined. This added requirement has not been fully taken into account in the experiments so far performed to test the levels of processing hypothesis.

The experiments in question (cf. Bower & Karlin, 1974; Craik, 1973; Gardiner, 1974) have all used an incidental learning paradigm, in which the level of processing is manipulated by varying the orienting task required of the subject. For example, in Bower and Karlin (1974) the subject might be shown a picture of a face and asked to judge whether its gender or its honesty. However, in interpreting the effects of these different orienting tasks on later recognition, it is not clear under what definition of depth the levels of processing hypothesis has been tested. In particular, the level of immediate perceptual processing is not being directly varied because the input to the subjects remains the same in all conditions. Instead, these experiments may be varying the degree of "elaboration coding" of the stimulus (cf. Tulving & Madigan, 1970). But in either case, these experiments do not constitute an adequate test of the levels of processing hypothesis because such a test depends on the precision with which the notion of level of processing has been specified and manipulated.

We propose, instead, to test the levels of processing hypothesis by directly varying the level of immediate perceptual processing, in a processing context where an independently motivated perceptual theory defines a sequence of processing levels ranging from the
preliminary sensory analysis to the extraction of meaning. An appropriate processing context is normal spoken language, in which at least four levels of analysis can be readily distinguished and manipulated and the properties of the memory representation have already been directly linked to the order of processing events at different levels of analysis.

First of all, the use of natural language enables us to control the depth to which the subject can process the material. For example, semantically anomalous material (called syntactic prose) is phonologically, lexically, and syntactically normal but cannot be processed or comprehended semantically because its semantic structure is systematically disrupted. Similarly, by scrambling the word order of normal prose, we can produce material that is syntactically as well as semantically unanalyzable. Differences in the recall of such materials can therefore be directly tied to the available processing levels in the original input, rather than to perceptually ambiguous differences in response task requirements.

Second, the existence of a psycholinguistic model of immediate sentence memory and processing offers a well-structured framework for developing the levels of processing approach. This psycholinguistic model views the memory representation of normal spoken language as changing systematically over time, as a function of the order of speech processing events (Bever, Garrett, & Hurting, 1973; Jarvella, 1971; Sachs, 1967). The listener initially segments the speech into natural linguistic units (sentences and clauses), on the basis of its syntactic structure, and then interprets these units semantically. The completion of this semantic interpretation—presumably at or following the clause boundary—allows the preliminary representation at all linguistic levels to be replaced by just a semantic representation.

The primary evidence for this clausal differentiation of the memory representation comes from the experiments of Jarvella (Jarvella, 1970; 1971; Jarvella & Pisoni, 1970). In these experiments, the subjects listened to a continuous prose passage and were asked at unpredictable intervals to recall the immediately preceding section. Jarvella found that the subjects' accuracy of recall was segmented according to the clausal structure of the material. They were only able to repeat with verbatim accuracy the last clause they heard. Memory for earlier clauses was less accurate and faithful to the original only at a semantic level. According to the analysis outlined above, this is because a multilevel representation of the last clause is still available when testing is signaled, whereas processing of the earlier clauses has been completed, leaving only a semantic representation available in memory.

Although not originally stated as such, this psycholinguistic model of sentence memory is effectively a levels of processing model since what is remembered at different delays is determined by the level to which the input is being, or has been processed. The levels of processing approach can therefore be directly tested in this clause-memory context by examining the effects of reducing the depth to which the material can be processed. Although recall of the last clause heard should not be affected by these manipulations, because it is not dependent on a semantic analysis, the preceding clauses should be increasingly poorly recalled, relative to normal prose, as the depth of available processing is reduced. Experiment 1 investigates the effects of removing the semantic level of analysis, and Experiment 2 examines the effects of removing the syntactic level as well.

**Experiment 1**

This experiment tests the levels of processing hypothesis by comparing verbatim recall for normal prose and for prose which cannot be processed semantically (syntactic prose). If, in the clause memory task, recall of the last clause heard is not dependent upon semantic processing of that segment, then performance on syntactic prose should be unimpaired relative to normal prose. However, if recall of earlier clauses is based on their semantic representation, then recall of syntactic prose should be far worse for these clauses than recall of normal prose.
Method

Subjects. The subjects were 10 students at the University of Chicago, who were paid for their services and naive to psycholinguistic experiments. 

Materials. A 1,500-word passage from a paper by Miller (1969) was chosen as the normal prose passage. The material was slightly modified to include 12 test sentences, each consisting of two 8-word clauses, placed at varying intervals through the passage. The syntactic prose passage was directly derived from the normal passage by pseudorandomly replacing all content words by new words of the same form class and word frequency (Kucera & Francis, 1967). In the test sentences, the replacement words matched the originals for syllable length as well. The locations of the test sentences were identical in the two passages. Examples of the two types of prose material follow (test sentence italicized):

Normal: If the total amount of future work to be done were limited, there might be some plausibility to this argument; future work, being limited, should be left for future generations to do. But there is little reason to believe that this is true. There is an enormous amount of work involved in order to achieve full automation of industry; no one who can contribute to it will need to go unemployed for centuries to come.

Syntactic: If the short ground of common men to be done were applied, there might be other felicity to this location; clear state, being interested, should be united for front suggestions to do. But he is good street to plan that they are close. It is an innocent paper of rooms considered in order to mention true solicitor at committee; nothing that may increase at them will open to go reassured for models to last.

Short practice passages, containing two test sentences, were constructed for use before each type of material. A 10-question comprehension test was made up for use after the normal prose passage.

The test and practice materials were recorded by a female reader with a natural intonation pattern at a rate of 100 words per minute. The reader did not know which were the test sentences, so that the only pauses after the test sentences were the natural intersentence breaks. These pauses were long enough to enable the experimenter to stop the recording during testing while the subjects made their responses. The materials were presented to the subjects over headphones as a binaural monophonic signal.

Design and procedure. A modified version of Jarvela's (1971) procedure was followed in this experiment. The subjects were tested individually in an IAC sound-attenuated booth. Each subject heard both types of test material, with the order of presentation balanced across subjects. Each subject heard a total of 12 normal prose test sentences and 12 syntactic prose test sentences. The subjects hearing normal prose first were read the following instructions:

In this experiment, you will hear a passage of English prose read to you. You are to listen to the passage normally, just as if someone was telling you a story. The reading of the passage will be interrupted from time to time by test pauses. During each test pause, your task will be to write down as much as you can remember of the last sentence you heard immediately preceding the interruption. We are interested in learning how much of the immediately preceding sentence you can remember exactly, word-for-word.

The subjects were also warned that they must attend to the meaning of the material because they would be given a comprehension test at the end of the passage. They were then given the normal prose practice passage, and their responses were examined to make sure that they understood the task and that they could identify the preceding sentence. None of the subjects showed any difficulty in doing this.

After they had heard the normal prose test passage and had completed the comprehension test, the subjects were given a modified version of the same instructions for the syntactic prose. Similar instructions, changed as appropriate, were read to the subjects who heard the syntactic prose first. None of the subjects noticed the relationship between the two types of material.

During the experiment, the recording was stopped following each test sentence, and the subjects were allowed 45 seconds to write down their responses. The recording was then restarted without warning. The time allowed for responding was found to be more than adequate by all subjects. Each session lasted about 45 minutes.

Results and Discussion

The subjects' written responses were scored word-by-word according to a single criterion: whether or not they exactly (verbatim) reproduced what they heard. Any divergence, including order errors, was scored as incorrect. The percentages of words correct, by material and clause, are given in Table 1. The mean number of words correct is plotted across word positions in Figure 1.

Because it was not possible to compute F directly, two univariate analyses of variance were performed, using the multivariate program (Finn, 1972), with one variable (mean number of words correct) and four factors (unit, type of material, clause, and word position). The unit factor corresponded to
subjects in the first analysis and to sentences in the second analysis. The results of these analyses were used to derive Min F' values (Clark, 1973). Note that there is no joint error mean square when Min F' values are computed. All significant statistics are reliable beyond the .01 level unless otherwise indicated.

There were strong main effects of type of material, Min F'(1, 31) = 30.626, of clause, Min F'(1, 606) = 59.698, and of word position, Min F'(7, 608) = 7.429, with a significant interaction between type of material and clause, Min F'(1, 608) = 23.417. The overall effects can be seen in Figure 2, which presents the smoothed data derived from this analysis. Only the linear and quadratic components reached significance.

Recall of normal prose was much better than recall of syntactic prose. Clause 2 (the second clause in the test sentence) was recalled better than Clause 1, and there was a significant serial position effect. Both types of material showed a clause effect, but, as the interaction between material and clause indicates, they differed in the size of the effect. In syntactic prose there is a 73% increase in number of words correct from Clause 1 to Clause 2 but a much smaller increase of 9% for normal prose. Using a by-word-position sign test, the clause effect (the increase of Clause 2 over Clause 1) is significant for syntactic prose at the .004 level and for normal prose at the .035 level.

The much greater differentiation across the clause boundary for syntactic as opposed to normal prose is in accord with the predictions of the integrated levels of processing and psycholinguistic model. According to this model, the earlier clause (Clause 1) is normally recalled on the basis of its semantic representation alone. In the clause-memory task, this produces a decrement in verbatim accuracy of recall because a semantic description is indifferent to those syntactic and lexical details which do not change meaning. Nonetheless, since this semantic representation reflects an analysis to the deepest processing levels, an enduring trace remains in memory to enable the subject to recall the earlier material.

### TABLE 1

<table>
<thead>
<tr>
<th>Prose type</th>
<th>Clause 1</th>
<th>Clause 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>79.3</td>
<td>86.4</td>
</tr>
<tr>
<td>Syntactic</td>
<td>43.0</td>
<td>74.6</td>
</tr>
<tr>
<td>Random word order</td>
<td>6.2</td>
<td>67.8</td>
</tr>
</tbody>
</table>
Syntactic prose, however, cannot be comprehended, so that it falls short of the deepest level of processing. Accordingly, the memory representation for syntactic prose fades more rapidly over time, producing poor recall of the earlier material, relative to normal prose.

The predicted picture is complicated by the superiority of normal over syntactic prose in Clause 2 as well. This difference, although much smaller than in Clause 1, is quite consistent over word positions ($p = .035$, sign test). This suggests that recall of the last clause heard is influenced by the availability of a semantic level of analysis. Such a possibility is perfectly compatible with the levels of processing model but requires some refinement of the psycholinguistic model.

It is not clear from Jarvella’s formulation just when semantic analyses are initiated or terminated. The present results suggest that semantic processing of the last clause is already in progress at the time of testing. But because it has not been completed, lower-level information is still available. This results in superior verbatim recall of the last clause in normal prose because the subject still has available a full spectrum of processing levels. All of these levels, save the deepest, are shared with syntactic prose.

Therefore, to the extent that Clause 2 (as opposed to Clause 1) recall is based on a multilevel representation, syntactic prose recall for Clause 2 differs relatively little from normal prose. That there is a significant difference indicates that, even at the short delay between the end of the last clause and the onset of testing, the lower levels of this multilevel representation have started to fade. Note that even in normal prose, recall of the last clause is not perfect.

This overall interpretation of Experiment 1 can be further tested by presenting the subjects with syntactically unanalyzable material (random word-order prose). If the effects of semantic factors on Clause 1 and 2 recall are levels of processing effects, then removing the syntactic level as well should produce a similar dissociation across the clause boundary. Recall of Clause 2 should be much less impaired, relative to syntactic and normal prose, than recall of Clause 1.

This additional reduction in processing level also tests an alternative explanation of the poor recall of syntactic prose, namely, that syntactic prose is remembered less well because it was originally perceived less accurately. However, if this is the case, then recall of random word-order material should be much worse in Clause 2 than for syntactic prose. Miller and Isard (1963) showed...
that syntactic and semantic disruptions together produced a much stronger effect on intelligibility than did semantic disruptions alone. On the levels of processing account, recall in Clause 2 should only be marginally affected since lexical and phonetic information is still available as a basis for recall.

**Experiment 2**

To maintain continuity with Experiment 1, it is necessary to present random word-order material, which has no syntactic structure, in the same “clause memory” situation. To do this, a clausalike segmentation needs to be imposed on the material, since it has no intrinsic clausal structure. This can be done by reading the material with the appropriate intonation pattern, which signals the beginnings and ends of successive segments. To further encourage the subjects to use the same strategies as in normal and syntactic prose, they are first exposed to the normal prose passage before hearing the random word-order material.

**Method**

**Subjects.** The subjects were 10 students recruited from the same population as in Experiment 1.

**Materials.** The 1,500-word syntactic prose passage in Experiment 1 was transformed into random word-order prose by scrambling the word order of each sentence. For the twelve 16-word test sentences, the word order within each clause was scrambled. Thus recall for the same words was being tested as in the previous experiment. The relative locations of the test sentences were the same as in the syntactic prose. A practice passage containing two test sentences was also constructed. The random word-order materials were read by a female reader at a rate of 160 words per minute, with an approximation to a normal intonation pattern.

**Design and Procedure.** The same procedures were followed as in Experiment 1. All subjects heard the normal prose passage first. This was followed by the random word-order practice and test passages.

**Results and Discussion**

The same scoring procedures were used as in Experiment 1. The percentage of words correctly recalled in each clause is given in Table 1, and the number of words correct over word positions is plotted in Figure 1. Min $F'$ values were computed from two separate multivariate analyses of variance, with the significance level set at .05. There were significant effects of clause, Min $F'(1, 249) = 388.778$, and of word position, Min $F'(7, 225) = 4.190$. The quadratic curves derived from these analyses are plotted in Figure 2.

Recall of random word-order material from Clause 2 did not differ significantly from Clause 2 recall of syntactic prose but was significantly worse than normal prose recall ($p = .035$; by-word-position sign test). Random word-order performance in Clause 1 was extremely poor and much worse than either syntactic or normal prose ($p = .004$; sign test).

This very low level of recall in Clause 1 raises the possibility that the subjects were unable to identify the segment in question. Although the strong “clause effect” indicates that they could use intonational cues to identify the mid-sentence break, it is possible that these cues were insufficient to mark off the first clause in memory. Accordingly, in an additional control study, we reexamined performance on random word-order prose, using prompts to indicate which clause the subjects were to recall. However, even when the subjects were required to recall only the seven words following the prompt, and where the prompt was the first word of the first segment, recall of this segment remained at a very low level (percentage words correct $= 5.4\%$).

The results of the control study support the interpretation of Experiment 2 in a level of processing framework. Reducing the available depth of processing from the syntactic to the lexical level produced an asymmetry in recall across the clause boundary. Memory for the last clause in Experiment 2 was unimpaired relative to syntactic prose. It is only at longer delays, marked off by the clause boundary, that the absence of a syntactic level of analysis has a significant effect. The drastic reduction in performance for the first segment demonstrates the importance of the deeper levels of immediate perceptual processing for the formation of an enduring memory trace. The failure— even of prompts to improve performance on
random word-order prose indicates that only the most fragmentary representation of this material is left in memory when testing is signaled. Note that exactly the same words were involved in the random word-order and the syntactic prose conditions.

Experiment 2 also tested a perceptual interpretation of the results, which would predict a larger fall off in Clause 2 for unrelated strings of words than for semantically anomalous but syntactically normal strings. However, no significant difference between random word-order and syntactic prose was obtained in Clause 2. This suggests, again, that the experiments reflect the consequences for memory of available depths of processing rather than just difficulties in immediate perceptual analysis.

**General Discussion**

The principal aim of this research was to test the levels of processing approach to memory by directly varying the depth to which the input could be processed, as opposed to the earlier incidental learning tests of levels of processing, where the input was held constant across processing conditions. We found that memory for spoken prose materials deteriorated as a direct function of the available depth of processing. More significantly, this decrement in performance was clearly structured according to the successive stages of processing specified in a model of sentence perception. As we argued in the introduction, the levels of processing approach can only be adequately tested in the context of an independently motivated processing theory. The pattern of results across the clause boundary, for the different prose types used in this experiment, precisely follow the predictions of a levels of processing memory theory integrated with a psycholinguistic processing model.

Although these experiments were not explicitly intended to test the suitability of other memory theories for a sentence processing context, it is worth considering how a multistore theory would account for our results. Jarvela (1971), in fact, originally stated his model in terms of a dual-storage theory, in which material is shifted clause by clause from a short-term buffer store to a long-term memory store, where it is semantically represented. As stated, such a model provides no basis for any recall of the semantically unanalyzable material in Clause 1, nor does it predict any differentiation in the long-term recall of syntactic as opposed to random word-order prose. A strict two-store model would also predict no differences between prose types in Clause 2 either, because recall of this segment of the material is from a short-term store which can only represent the material at a phonetic or lexical level of analysis, which all three prose types have in common.

However, it is easy enough to emend a strict two-store model to account for the results. One could, for example, allow semantic and syntactic factors to affect recall from the short-term store and postulate degrees of trace duration in long-term memory which would correspond to the differential levels of recall from Clause 1. But that is not the point. As Craik and Lockhart (1972) have lucidly argued, the multistore approach to memory is unsatisfying precisely because it requires these proliferations of additional theoretical constructs, which, apart from anything else, erode the binary distinctions upon which the postulation of separate stores was originally based. We find the levels of processing approach to provide a more fruitful and parsimonious "set of orienting attitudes" (Craik & Lockhart, 1972, p. 681), and this paper is one attempt to exploit its possibilities. And however the results are accounted for, they do graphically illustrate the differential suitability of structured and unstructured material for storage in a natural verbal memory system.

**REFERENCES**


(Received June 16, 1975; revision received August 27, 1975)