

The temporal structure of spoken language understanding*

WILLIAM MARSLEN-WILSON**

LORRAINE KOMISARJEVSKY TYLER

*Max-Planck Institut für Psycholinguistik,
Nijmegen*

Abstract

The word-by-word time-course of spoken language understanding was investigated in two experiments, focussing simultaneously on word-recognition (local) processes and on structural and interpretative (global) processes. Both experiments used three word-monitoring tasks, which varied the description under which the word-target was monitored for (phonetic, semantic, or both) and three different prose contexts (normal, semantically anomalous, and scrambled), as well as distributing word-targets across nine word-positions in the test-sentences. The presence or absence of a context sentence, varied across the two experiments, allowed an estimate of between-sentence effects on local and global processes. The combined results, presenting a detailed picture of the temporal structuring of these various processes, provided evidence for an on-line interactive language processing theory, in which lexical, structural (syntactic), and interpretative knowledge sources communicate and interact during processing in an optimally efficient and accurate manner.

Spoken language understanding is, above all, an activity that takes place rapidly in time. The listener hears a series of transient acoustic events, that normally will not be repeated, and to which he must assign an immediate interpretation. These interpretative operations are rather poorly understood, and very little is known about the detailed structure of their organization in time.

*The experiments and the original data analyses were carried out while both authors were attached to the Committee on Cognition and Communication at the University of Chicago, and was supported by a grant to WMW from the Spencer Foundation. We thank David McNeill for the use of his laboratory, David Thissen for statistical advice, and Elena Levy for help in running Experiment 2. Additional data analyses and the preparation of the manuscript were carried out after the authors moved to the Max-Planck-Gesellschaft, Projektgruppe für Psycholinguistik, Nijmegen. We thank Tony Ades for helping us think about sentence processing, and the two reviewers, Ann Cutler and David Swinney, for their comments on the manuscript. LKT is also attached to the Werkgroep Taal- en Spraakgedrag, at the Katholieke Universiteit, Nijmegen. A preliminary report of part of Experiment 1 was published in *Nature*, 1975, 257, 784-786.

**Reprint requests should be addressed to: Dr. W. D. Marslen-Wilson, Max-Planck Institut für Psycholinguistik, Berg en Dalseweg 79, Nijmegen, The Netherlands.

This paper, therefore, is intended to provide some basic information about the word-by-word time-course of spoken language processing.

The research will be conducted on the basis of an approach to speech understanding in which the entire range of processing activities are seen as taking place on-line, as the utterance is heard (cf., Marslen-Wilson, 1975; 1976; Note 1; Marslen-Wilson, Tyler, and Seidenberg, 1978). Very generally, it is claimed that the listener tries to fully interpret the input as he hears it, and that the recognition of each word, from the beginning of an utterance, is directly influenced by the contextual environment in which that word is occurring.

These claims about the *on-line interactive* character of speech understanding determine the kinds of experimental questions being asked. Thus the research will have two main foci, examined concurrently in the same experiments. The first focus will be on the global structure of processing — that is, on the overall temporal properties of the flow of communication between different knowledge sources during the perception of an utterance. The second focus will be on the local structure of processing — that is, on the temporally more localized interactions between different knowledge sources during the process of spoken word recognition. These investigations will be carried out in the further context of the general organizational issues that are central to the speech understanding problem.

Determining assumptions in models of language processing

The organization of a language processing model depends on two types of fundamental assumptions — about the distinct types of mental knowledge (or knowledge sources) that are involved in language processing, and about the possibilities for communication and for interaction¹ between these knowledge sources. Our questions here primarily concern this second type of assumption, since it is this that determines the basic structure of the system. We will begin by examining the processing assumptions underlying the type of "serial" metaphor that has dominated psycholinguistic thinking about sentence processing for the past two decades (cf., Carroll and Bever, 1976; Fodor, Bever, and Garrett, 1974; Forster, 1979; Levelt, 1978; Marslen-Wilson, 1976; Tyler, 1980).

¹The term "communication" will be used here to refer to the cases in which the flow of information between processing components is in one direction only, normally from the bottom up. The term "interaction" will be restricted to cases where information is assumed to be able to flow in both directions.

The autonomy assumption

The distinguishing feature of a serial model is its assumption that the flow of information between the different components of a processing system is in one direction only, from the bottom up. A processor at any one "level" has access only to the stored knowledge which it itself contains, and to the output from the processor at an immediately lower level. Thus, for example, a word-recognition component can base its recognition decisions only on the knowledge it contains about the words in the language, and on the input it receives from some lower-level analyser of the acoustic-phonetic input. Given such an input, the word-recognizer will determine which word, or words, in the language are the best match, and then pass this information on to the processor at the next stage of analysis. What it cannot do, in a strictly serial system, is to allow a processor higher in the system to intervene in its internal decision processes – either when ambiguous cases arise, or simply to facilitate the recognition of unambiguous inputs. Each component of the processing system is thus considered to be *autonomous* in its operations (Forster, 1979; Garrett, 1978).

This autonomy assumption means that any apparent effects of higher-level analyses on lower-level decisions cannot be located in the lower-level processor itself. Context-effects in word-recognition, for example, could not be located in the word-recognition component, but at some later stage, at which the appropriate constraints could be used to select among the several words that might be compatible with a noisy sequence of acoustic-phonetic input. Exactly the same logic would apply to the relationships between syntactic and semantic processing components. From the point of view of the temporal organization of processing, this clearly implies that there will be some delay between the point in time at which an input is discovered to be ambiguous at one level of the system and the point at which the resulting ambiguity can be resolved at some higher level.

These processing assumptions have a profound influence on the ways in which a model can implement the further assumptions that are made about the types of mental knowledge (syntactic, semantic, etc.) that need to be distinguished. First of all, they require that each distinct knowledge type be realized as a separate processing component, functioning as a computationally independent processing level within the system. There is no other way, given the autonomy assumptions, that a knowledge type can participate in the analysis of the input. This has normally led to a four-tier processing system, with the four processing components reflecting some version of the linguistic distinctions between the phonological, lexical, syntactic, and semantic aspects of language.

Apart from the equation of knowledge types with processing components, the autonomy assumptions also require the definition of the characteristic processing units in terms of which each component conducts its analyses. For it is the properties of these units that determine the points in the analysis of the input at which one processing component will be able to transmit information to the next, and, equally important, the points at which this further component will itself be able to start analyzing the input it is receiving, and itself generate an output to subsequent levels of the system. Unless the processing units at each level of the system are properly specified, then the precise temporal structure of the flow of information through the system remains ambiguous, and difficult to verify experimentally.

Thus although the autonomy assumptions lead, in principle, to a well-defined theory of the global and local structure of spoken language processing, the effective validity of these claims depends on the further definition of level-by-level processing units. In the case of the lexical aspects of language, and the corresponding word-recognition component, the specification of the processing unit has not seemed to be a problem. The units are assumed to be words, so that the autonomous word-recognition component can only communicate with higher-level processors at those points in the analysis at which it has come to a decision about which word (or words) best match the acoustic-phonetic input it is receiving.

This leads to clear claims about the local structure of processing. These claims can be tested by determining whether context-effects in word-recognition do in fact occur only after the points predicted by the theory. The important question is not *whether* context-effects can be detected,² but rather *when* they occur. If, for example, context-effects are detected before the word-recognition device could have produced an output, then this would be *prima facie* evidence for an interaction of context with processes within the word-recognition device.

The choice of processing components and their associated processing units becomes much more problematical when one comes to consider the global structure of the system. In fact, it is no exaggeration to say that this has been the central theoretical and empirical issue in psycholinguistic research on sentence perception (e.g., Bever, 1970; Carroll, Tanenhaus and Bever, 1978; Fodor *et al.*, 1974; Fodor and Garrett, 1967; Garrett, Bever, and Fodor, 1966; Jarvella, 1971; Marslen-Wilson *et al.*, 1978). This was because modern

²The use of a word's prior and posterior context to resolve problems caused, for example, by a noisy sensory input, has been well established at least since the work of Miller, Heise, and Lichten (1951). We are not concerned here, however, with context effects that may be occurring several words downstream from the original ambiguity in the input.

psycholinguistics was founded on the hypothesis that a transformational generative grammar (Chomsky, 1957; 1965) provided the appropriate conceptual framework for thinking about language processing (cf., Fodor *et al.*, 1974; Miller, 1962; Miller and Chomsky, 1963). The usefulness of this hypothesis depended, historically, on two critical claims; namely, that there was an independent syntactic processor, computing a level of syntactic representation during sentence perception, and that the processing unit for this component was the deep structure sentence (cf., Bever, 1970; Clark, 1974; Fodor *et al.*, 1974; Forster, 1974; Marslen-Wilson, 1976; Tyler, 1980).

These claims, taken together with the autonomy assumption, led to a potentially fully specified processing model, which could make clear predictions about the temporal ordering of sentence processing operations. These predictions were, however, certainly too strong, since they required that semantic interpretation be delayed until the syntactic clause or sentence boundary had been reached, so that any semantically based disambiguating or facilitatory effects on lexical or syntactic processes could not be observed before this point (cf., Marslen-Wilson *et al.*, 1978; Tyler and Marslen-Wilson, 1977). Subsequent serial processing theories (e.g., Carroll and Bever, 1976) have held onto the assumption that there is an independent level of syntactic analysis, but have been much less definite about the processing units involved. Forster (1979), for example, in an otherwise very complete discussion of a serial processing system, states that the products of the autonomous syntactic parser are "relayed piecemeal to the next processor as viable syntactic constituents are identified", but does not identify the notion of a "viable syntactic constituent".

This kind of indeterminacy in the statement of the global structure means that one cannot derive the same type of prediction as was possible for the word-recognition component. Without a clearly specified processing unit, it is difficult to set up a compelling experimental demonstration that semantic analyses are interacting with decisions within the syntactic processing component. One can, however, identify what seems to be the minimal empirical claim that a meaningfully serial model can adopt. This would not be a claim about the presence or absence of interactions, but about the sequential ordering of communication between levels of the system. If a processing system depends on the serial operation of a sequence of autonomous processing components, and if these components include a syntactic processor, then there should be at least some measurable delay in the availability of processing information that would derive from processing components situated later in the sequence than the syntactic processor. If relative delays of this type were to be found, then not only would this be strong evidence in favor of a serial theory, but also the timing of any such delays would help to determine the processing units at each level of analysis.

An alternative processing assumption

The above discussion of the autonomy assumption serves to demonstrate the crucial importance of such assumptions for a processing model, and to provide the context for the rather different processing assumptions made in the on-line interactive approach. This approach does not place any *a priori* limitations on the ways in which different knowledge sources can communicate during language processing. It assumes, instead, a more flexibly structured processing system, such that analyses developed in terms of any one knowledge source can, in principle, be made available to affect the operations of any other knowledge source. We exclude from this the basic acoustic-phonetic analysis of the sensory input. It is implausible, for example, that an acoustic-phonetic analyzer should need to, or would be able to, communicate or interact directly with a semantic analyzer. But the remainder of the processing system, involving the lexical, structural, and interpretative aspects of language understanding, is assumed to allow free communication and interaction during the comprehension process.

The reason for taking this approach is, first, a series of experiments, using the speech shadowing technique, which show that the listener can develop at least a preliminary syntactic and semantic analysis of the speech input word-by-word as he hears it (Marslen-Wilson, 1973; 1975; 1976; Note 1). These findings are inconsistent with the autonomy assumption, which places fixed constraints on the manner in which analyses can develop in time. The data suggest, instead, that the processing system should be organized so as to allow the properties of the input itself to determine how far its analysis can be taken at any particular moment. Given the recognition of a word, the structural and interpretative implications of this word-choice should be able to propagate through the processing system as rapidly and as extensively as possible. This means that an utterance can, in principle, be fully interpreted as it is heard (cf., Marslen-Wilson *et al.*, 1978). Note that this is a claim about the possibilities for communication between knowledge sources, and not, in itself, for two-way interactions between them.

The second type of reason for taking the present approach is, however, the evidence for a genuine interaction between sentential context variables and the process of spoken word recognition. Not only does the listener develop an interpretation of an utterance as he hears it, but also he is apparently able to use this information to influence the recognition of subsequent words in the utterance. The evidence from close shadowers (i.e., subjects repeating back the input at delays of around 250 msec) shows that context can control the listener's lexical interpretation of the input very early on in the word-recognition process – in fact, within 200–250 msec of a word's onset

(Marslen-Wilson, 1975). It is this type of result which leads to the claim that the local structure of processing permits contextual interactions with word-recognition decisions.

These local and global structural claims are very general ones, and clearly need to be much more restrictively stated than the presently available data allow. The purpose of the research here, therefore, is to develop the basic data necessary for a fuller specification of an on-line interactive model. The experiments will, first of all, systematically examine the word-by-word time-course of the development of sentential analyses. Is there an intrinsic temporal ordering in the availability across an utterance of processing information that would derive from different potential knowledge sources? If no ordering is observable, then how early in the analysis of an utterance do the distinguishable types of processing information become available? The second purpose of the experiment is to examine the properties of the postulated interactions at the local structural level. Both these sets of questions will require, at each stage of the argument, a comparison with the claims of the autonomy assumption. To show that knowledge sources can communicate and interact freely means that one has to show that the connections between them are not restricted in the ways that the autonomy assumption would predict.

The next section of the paper lays out the general form of the experiment. We will then go on to discuss in more detail the ways in which the experiment is responsive to the issues we have been raising.

Measuring the local and global structure of processing

Given that the experimental questions crucially involve the precise time-course of speech processing events, this requires us to track these processes in real time. The subject must be able to respond while the operations of interest are being performed, since a delayed measure cannot directly discriminate between operations performed early or late in processing but before the point of testing.

The present experiment, therefore, will use an on-line word-monitoring task, in which the subject listens to a sentence for a target-word specified in advance, and makes a timed response as soon as the target is detected. By placing the target at different serial positions in the test-sentences, and by varying both the structural properties of the sentences and the descriptions under which the targets are defined for the subjects, we can map out the time-course with which different forms of analysis become available to the listener at the local and global levels.

The word-monitoring targets will occur in nine different serial positions in the test sentences, ranging from the second to the tenth word-position. This gives the spread of observations across a sentence that is needed to analyze

the global structure of processing. The further, critical, manipulations involve the prose contexts in which the words are heard, and the types of monitoring tasks the subjects are asked to perform.

Each target-word will occur, between subjects, in three different types of prose context, labelled Normal Prose, Syntactic Prose, and Random Word-Order Prose. These prose contexts systematically vary in the extent to which they allow the development of different forms of sentential analysis during processing. In the Normal Prose condition, the test sentences are syntactically and semantically normal. For example (target-word emphasized):

- (1) The church was broken into last night. Some thieves stole most of the *lead* off the roof.

The use of two-sentence pairs means that the Normal Prose test-sentences also occur in the further context provided by the lead-in sentence (the test-sentences are always the second sentences in the pair).

The Syntactic Prose version of sentence (1) is syntactically still interpretable, but has no coherent semantic interpretation. In addition, the lead-in sentence does not provide an intelligible context for the test-sentence:

- (2) The power was located into great water. No buns puzzle some in the *lead* off the text.

The Random Word-Order version of this is neither syntactically nor semantically interpretable:

- (3) Into was power water the great located. Some the no puzzle buns in *lead* text the off.

Word-monitoring in Normal Prose should be facilitated by the presence of a syntactic and semantic interpretation, and this facilitation should differ across word-positions. Thus it will be possible to measure the parameters of this facilitation by comparing the serial response-curves for the three types of material. The curves for Syntactic and Random Word-Order Prose will diverge from the Normal Prose curves as a function of the on-line contribution of syntactic and semantic information to the Normal Prose monitoring reaction-time – and, therefore, as a function of the global structure of ongoing sentence processing.

The possibility of delineating these global effects depends, however, on the local interactions of the monitoring responses with the overall context manipulations. To ensure that the appropriate interactions occur, and to create the necessary contrasts for the evaluation of questions about the local structure of processing, three different monitoring tasks are used. In each task the subject is asked to listen for a word target, but this target is specified to him in different ways.

First, in the *Identical* monitoring task, the subject is told in advance exactly which word to listen for (in sentences (1)–(3) above, the target would be

specified as *lead*). This means that he is listening for a target that has been both phonetically and semantically defined. In the two additional monitoring tasks – *Rhyme* and *Category* monitoring – these two aspects of the target definition are separated. In *Category* monitoring the subject is told to monitor for a target that is a member of a semantic category given in advance. Thus, in sentences (1)–(3), where the target-word is *lead*, the subject would listen for it under the description *a kind of metal*. In *Rhyme* monitoring he listens for a phonetically defined target; for a word that rhymes with a cue-word given in advance. If the target was *lead*, then the cue-word could be *bread*. A rhyme-monitoring task is used here in preference to a phoneme-monitoring task, since this makes it possible to equate the targets in *Rhyme* and *Category* monitoring. In both cases the subject is listening for a target defined to him in terms of its attributes (phonetic or semantic) as a complete word.

We will assume here that all three tasks involve word-recognition processes, but that the *Rhyme* and *Category* monitoring tasks require a second stage, in which the attributes of the word that has been recognized are matched along the appropriate dimension with the attributes specified for the target. Identical monitoring is assumed not to involve a specific additional attribute-matching stage, since to recognize a word is also, in a sense, to match its attributes against those specified for the target.

This means that the Identical monitoring task should be the most direct reflection of the timing of word-recognition processes, and of the postulated interaction of the different prose contexts with these processes. The function of the other two tasks (*Rhyme* and *Category*) will be to investigate, first, the order in which different types of information about a word are accessed during recognition, and, second, the ways in which the forms of analysis available in the different prose contexts interact with these different aspects of a word's internal representation. The autonomy hypothesis and the on-line interactive approach appear to make quite different predictions here.

In general, however, the purpose of using these three different tasks is to be able to approach the issues we have raised from the perspective of several different experimental probes of on-line analysis processes, so as to develop as richly detailed a picture as possible of these processes. The various questions these tasks allow us to ask are discussed in the next two sections.

Local structural questions and predictions

The timing of word-recognition processes

The basic empirical question about word-recognition in different contexts is the timing of the context-effects that we are predicting. Thus, using the Iden-

tical monitoring task, and measuring latency from the onset of the target-word, the question is precisely when the response can be obtained relative to the temporal duration of the target-word in question. The significance of any contextual facilitation of monitoring responses will depend very much on where in time these responses can be located relative to the word-sized processing units with which we are presumably dealing. The shadowing data (e.g., Marslen-Wilson, 1975) suggest that context-effects should be detectable before all of a word has been heard.

It is not altogether clear what predictions the autonomy assumption leads to here. Serial theories of word-recognition, chiefly reflecting the evidence from several studies of lexical ambiguity (e.g., Cairns and Kamerman, 1975; Foss and Jenkins, 1973; Swinney, *in press*),³ claim that the listener depends only on acoustic-phonetic information for his initial identification of a word. Semantic context could only operate later – for example, in helping the system to decide among the two or more readings of the word-choice selected at the first stage. Forster (1976) has made similar proposals, although basing his conclusions primarily on data derived from visual word-recognition studies.⁴

This claim for the autonomy of lexical access should predict that Identical monitoring responses will not be affected by variations in prose context. And if no context effects are obtained, then this would indeed be strong evidence in favor of the autonomy assumption. But if context effects are obtained, there remains at least one loophole for the serial theorist. This is to argue that although lexical access is truly autonomous, the output of this process is not accessible for making a response until it has passed through some subsequent stages that involve contextual variables. A loophole of this type has already been used, in fact, to cope with the demonstrations that monitoring for a phoneme-target involves word-recognition processes, and is not normally executed just on the basis of the acoustic-phonetic analysis of the input.⁵

³This evidence is itself not unambiguous. The studies here using the phoneme-monitoring task have become methodologically somewhat suspect (Cutler and Norris, 1979; Mehler, Segui and Carey, 1978; Newman and Dell, 1978). The lexical decision studies by Swinney (*in press*), while elegantly demonstrating that the contextually inappropriate meaning of a lexically ambiguous word is indeed momentarily activated, do not thereby show that the original identification of the word was itself not affected by context.

⁴We should stress here that our research is concerned, in the first instance, only with spoken language recognition processes. The processing of visible language may involve quite different relationships between stimulus and context variables.

⁵The early work using the phoneme-monitoring response (e.g., Foss, 1969; 1970; Hakes, 1971) clearly was based on the assumption that the output of an acoustic-phonetic analysis (as the earliest stage in the serial sequence of autonomous processing stages) would be available independently of the interpretation of this output elsewhere in the system.

The problem with this type of escape-route is that it seems to render serial theories of word-recognition empirically empty. The theory predicts a certain strict ordering of processing events, but this ordering is claimed not to be actually open to direct confirmation. In the present experiments, the plausibility of this particular escape-route will, again, depend very much on the temporal relationships between monitoring responses and word-durations. In general, an autonomous access process seems to predict that a word cannot be identified until the acoustic-phonetic information corresponding to that word in the sensory input has been analyzed, so that monitoring responses should tend to be rather longer than the durations of the target-words.

Types of context effects

A second question that can be answered here is whether there is any restriction on the type of contextual variable that can be shown to interact with word-recognition decisions. The on-line interactive approach assumes that there need be no such restrictions. It therefore predicts that Identical monitoring responses in Normal Prose will be facilitated relative to Syntactic Prose because of the availability of a semantic interpretation of the Normal Prose test-sentences. Similarly, the syntactic analysis possible in Syntactic Prose should facilitate responses here relative to Random Word-Order Prose.

In addition, the on-line interactive approach also allows for the possibility that the discourse context provided by the lead-in sentence in the Normal Prose conditions could affect word-recognition processes. Cole and Jakimik (1979) have recently reported some evidence to favor this possibility. The second experiment here will in fact directly investigate these between-sentence effects.

Phonological and semantic analyses in word-recognition

A powerful diagnostic for the presence or absence of contextual interactions with word-recognition processes should be the order in which different types of information about a word become activated. On the autonomy hypothesis, the semantic properties of a word are irrelevant to the primary process of word-identification, and should therefore only be accessed relatively late in processing. The on-line interactive approach, however, claims that semantic information about a word does become available early in the identification process, since semantic context is able to interact with this process. The contrast here between Rhyme and Category monitoring allows us to discriminate between these two claims.

We assume that the Category monitoring response requires first identifying a word, and then matching the semantic attributes of this word against the attributes specified for the target. Analogously, Rhyme monitoring involves the matching of a word's phonological attributes against those of the rhyme-target. This, of course, requires the further assumption that Rhyme monitoring will not be based just on a phonetic analysis of the input, available independently of word-recognition processes. Such a possibility seems excluded by the evidence, from recent phoneme-monitoring studies, that phonemically or phonetically defined targets are normally responded to after lexical access, on the basis of the listener's knowledge of the phonological properties of the word containing the target (cf., Cutler and Norris, 1979; Morton and Long, 1976).⁶

Both the on-line interactive approach and a theory incorporating the autonomy assumption could claim that phonological information about a word will be activated early in word-recognition, as part of the process of matching the acoustic-phonetic input against possible word-candidates. But the autonomy assumption requires that this be the only aspect of a word's representation that could be accessed at this point. It is claimed that the listener depends only on acoustic-phonetic variables for his initial identification of the word.

This appears to require that Rhyme monitoring responses will always be faster than Category monitoring responses. The phonological information necessary for attribute-matching in Rhyme monitoring should be available at least as soon as the word has been identified. Whereas semantic attribute-matching, in Category monitoring, depends on a form of analysis of the word which could not even be initiated until after the word had been identified. Thus reaction-times in the two tasks should reflect the sequence in which different aspects of a word become relevant to the operations of successive autonomous processing components.

According to the on-line interactive approach, word-identification involves an interaction between the acoustic-phonetic input and the syntactic and semantic context in which the word is occurring. These early contextual effects could only be possible if syntactic and semantic information was available about the word whose recognition the contextual inputs were helping to facilitate. Accessing information just about the phonological properties of a

⁶Cutler and Norris (1979), as well as Newman and Dell (1978) in fact make a more complex claim: that phoneme-monitoring decisions depend on a race between two processes operating in parallel. One process operates on the phonetic analysis of the string, and the other on the phonological representations of words in the mental lexicon. But this proposal is difficult to evaluate, since it is not clear which process will win under which conditions.

word would not provide a domain within which contextual criteria could apply. Thus, by the time a word had been identified, there should be no asymmetry in the availability of semantic as opposed to phonological information about that word, and, therefore, no necessity that semantic attribute-matching decisions be delayed relative to phonological attribute-matching.

However, this prediction only holds for the Normal Prose conditions. In Syntactic Prose and Random Word Order there is no semantic interpretation of the input which could usefully interact with word-recognition processes. In fact, the semantic attributes of words will be irrelevant to their identification in these contexts. Thus semantic attribute-matching here would depend on a form of analysis of the word-targets that should not begin until after the words had been recognized. But Rhyme monitoring should be relatively much less impaired, since it depends on a form of analysis that will be involved in the processes of word-recognition in the same ways in all three prose contexts.

It is difficult to see how the autonomy assumption could predict this further interaction between prose contexts and the relative differences between Rhyme and Category monitoring. Whatever the processes that make available the syntactic and semantic properties of a word, they should operate prior to the intervention of syntactic and semantic processors. Otherwise sentence processing would not be the strictly bottom-up process that the autonomy assumption requires. Thus the presence or absence of a semantic or syntactic context should not affect the availability of the information needed for semantic attribute-matching.

The properties of sensory/contextual interactions

The issue here is how, and whether, contextual variables affect the listener's utilisation of the sensory (acoustic-phonetic) inputs to the word-recognition process. This question can be examined here because the same test-words will occur, across subjects, in each of the three prose contexts. This means that we can evaluate the ways in which the same acoustic-phonetic inputs, mapping onto the same locations in the mental lexicon, are treated by the recognition system under different conditions of contextual constraint. The central claim of the autonomy hypothesis is that this mapping process is conducted in isolation from contextual variables. If this is so, then the dependency between reaction-time and variations in the sensory input should be unaffected by differences in prose contexts.

The particular dependency we will exploit here is the relationship between reaction-time and the durations of the words being monitored for. The duration of a word measures the amount of time it takes for all of the acoustic-

phonetic information corresponding to a word to be heard. To the extent that all of this information is necessary for word-identification, then response-time should increase with increasing word-length. The autonomy hypothesis requires that, whatever the correlation actually obtained between these two variables, this correlation should be constant across prose-contexts. It also requires that the slope of this relationship between word-length and reaction-time should be constant. This second claim is critical, since, according to the logic of the additive-factors method (e.g., Sternberg, 1969), the assumption of successive autonomous stages in a psychological process is only required if the contributions of these two stages to response-latencies are additive. Additivity, in the present situation, means that the slopes of the regressions of reaction-time on word-length should be parallel across prose-contexts.

The on-line interactive approach claims, in contrast, that sensory and contextual inputs are integrated together at the same processing stage to produce the word-recognition response. This means that the stronger the contextual constraints, then the less the dependence of the word-identification decision upon acoustic-phonetic information. Thus the correlation between word-duration and response-time should differ across prose-contexts, and the slope of this relationship should also change. In Normal Prose, where the contextual constraints are potentially strongest, the dependency between monitoring reaction-time and word-duration will be weakest. But in the other prose contexts the acoustic-phonetic input will be a more important source of information, so that the correlation with word-length will be larger, and the slope of the relationship will be steeper. If these predictions are fulfilled, then the contributions of sensory and contextual variables will not be strictly additive, so that we will be permitted to assume that they interact at the same stage in processing (cf., Meyer, Schvaneveldt and Ruddy, 1975).

Global structural questions and predictions

The issues here are much more straightforward, since we are not trying to test for the presence of an interaction between different types of processing information. The problem here is to determine when different forms of analysis of the input become available to the listener — although, of course, the possibility of tracking these processes does itself depend on the correctness of the on-line interactive claim that global analyses interact with word-recognition decisions.

The three prose contexts used here distinguish, broadly, between two types of sentential analysis — the semantic interpretation of an utterance, and its

syntactic analysis. Given the usual separation of processing functions in a serial processing model, there should be at least some temporal delay in the relative availability of these two types of analysis. In terms of the effects over word-positions in the present experiment, the serial-position curves for Syntactic Prose and Normal Prose should start to separate from the Random Word-Order curves first. At some later point the Normal Prose curves should start to separate from the Syntactic Prose curves, as a semantic analysis starts to become available. Because of the indeterminacy of the syntactic processing unit, recent serial models do not make precise predictions about when these different effects should be detectable; they predict only that there should be some relative delay.

The on-line interactive approach claims that the syntactic and the semantic implications of a word-choice become immediately available to all relevant parts of the processing system, so that there is no reason to predict an intrinsic delay in the availability of one type of sentential analysis rather than another. Note that the use of the three prose contexts does not commit us to the claim that there are separate syntactic and semantic processing components, each constituting a separate computational level of the system. The experiment only requires the assumption that the types of analysis distinguished by the three prose types correspond to distinctions between mental knowledge sources – that is, between the specifically linguistic knowledge involved in the structural analysis of linguistic sequences, and the types of knowledge (not necessarily linguistic in nature) that are involved in the further interpretation of these sequences. We will leave for later discussion the question of how these different knowledge sources might or might not correspond to distinct levels of representation in the language understanding process.

Finally, there is the question of how the three monitoring tasks might differ in their interactions over word-positions with prose context. The Identical and Rhyme monitoring tasks should behave in the same way. The effects of global processes will be detectable in each case *via* their interactions with the word-recognition process. The additional phonological attribute-matching process in Rhyme monitoring can be assumed not to interact with context, and should simply increase monitoring latency by some constant amount.

The Category monitoring task allows for one additional test of the ordering of processing events. The availability of a semantic interpretation in Normal Prose should facilitate the semantic attribute-matching stage in this task, while not affecting attribute-matching in Rhyme monitoring. Thus, to the extent that there is any delay in the availability of a semantic interpretation of an utterance, then this extra facilitation of Category monitoring, relative

to Rhyme monitoring, should only be detectable later in the sentence. This would mean that the word-position effects for Rhyme and Category monitoring should differ early in the test-sentence.

Overview

The proposed experiment combines an investigation of the global temporal ordering of language understanding processes and of the local temporal structure of interactions during word-recognition. We do this by fitting together three prose contexts, which vary the availability of syntactic and semantic processing information over nine word-positions, and three monitoring tasks, which themselves vary in their dependence on different types of analysis of the target-words. The combination of the two sets of questions seems unavoidable, since investigating the one requires a theory of the other, and adequate theories of either are not presently available. One reason for this lack may be precisely because questions about local and global processing issues have not been studied in the context of each other in quite the way we are proposing to do here.

The rest of this paper will be organized as follows. The results and discussion sections following the main experiment will be separated into two parts with the first part dealing only with local structural issues. We need to demonstrate the appropriate effects at this level before being able to interpret the results at the global level. In particular, it will be necessary to show how word-identification processes, as measured here in the word-monitoring tasks, interact with the syntactic and semantic interpretation of an utterance. The second part of the results and discussion will deal with the implications of the word-position effects for the global ordering of language understanding. This will lead to a second experiment, examining the between-sentence effects (due to the lead-in sentence in Normal Prose) that were observed in the main experiment. This will be followed by a general discussion of the implications of the complete global and local structural results for an on-line interactive model of spoken language understanding.

Experiment 1

Method

Subjects

The subjects were 45 students at the University of Chicago, drawn from a volunteer subject pool, and paid for their participation.

Materials and Design

The primary stimulus set consisted of 81 pairs of Normal Prose sentences, with the second sentence of each pair containing a monitoring target-word. The 81 target-words were selected from the Battig and Montague (1969) category norms under the following constraints. Each word was required to be a frequent response to the category label in question, with high frequency defined as being either among the ten most common responses or as being chosen by at least 30% of the respondents. Words from a total of 23 different taxonomic categories were selected. Each target-word was also required to have at least four readily available rhymes. One-syllable words were used, with four exceptions, where two-syllable words were chosen.

The target-words were randomly assigned to one of nine word-positions, such that nine targets appeared in each serial position from the second to the tenth word of the second test sentence in each pair. The Syntactic Prose test-sentences were constructed directly from the Normal Prose sentences by pseudo-randomly replacing all content-words (except the target-words) with new words of the same form-class and word-frequency, such that the resulting sentence had no coherent semantic interpretation. The Random Word-Order sentences were constructed from the Syntactic Prose sentences by scrambling their word-order, while keeping the serial position of the target-words unchanged. The sentences were scrambled in such a way that the same words (though in different orders) preceded the target-word in the Random Word-Order and the Syntactic Prose versions of each sentence. An example of one of the resulting triplets of test-sentences was given earlier, in the Introduction (see p. 8).

Three tapes were then recorded, each containing 81 of the 243 test-sentences. Each tape contained only one occurrence of each target-word, in one of its three prose contexts (Normal Prose, Syntactic Prose, or Random Word-Order). The tapes were further arranged so that each tape contained 27 test-sentences of each prose type, with three targets within each prose type occurring at each of the nine word positions. The order of presentation of the test-sentences (with their respective targets) was the same in all three tapes, with different prose-types and word-positions distributed pseudo-randomly across the entire tape. The materials were recorded by a female reader, at a rate of 160 words per minute. The reader did not know the nature of the experiment, nor which words were targets.

In addition, three sets of instruction booklets were constructed, which informed the subject, before each trial, what kind of prose-type he would be hearing, what the monitoring task would be, and what the cue-word was. There were three monitoring tasks: Identical, Rhyme, and Category monitor-

ing. In Identical monitoring the cue-word was the actual word the subject was to listen for. In the examples given earlier, the Identical cue-word was LEAD. In Rhyme monitoring, where the subject listened for a word that rhymed with the cue-word, the cue-word in the sample case would be BREAD. the cue-words and their target rhymes were, intentionally, not always orthographically identical (e.g., STREET as a cue for WHEAT). In Category monitoring, where the subject listened for a word that was a member of a taxonomic category, the cue-word would specify the category in question. In the present example, the cue-word was METAL.

So that each target-word would be tested, across subjects, under all three monitoring conditions, three sets of instructions for each target were necessary. The combination of tapes and instructions produced nine experimental conditions, so that a block of nine subjects would hear each target-word, in one of nine word-positions, in all nine combinations of prose context by monitoring task. The target-words, therefore, were fully crossed by Prose Context and by Monitoring Task, but were nested with respect to Word-Position. This design fully counter-balances for word and subject effects at the level of Context and Task, but is only partially balanced at the level of Word-Position, since each word-position contains nine words that do not occur in any other position. However, within each word-position, the same words occur in all combinations of Context and Task.

The additional variable of word-length – that is, the temporal duration of the target-words – was not manipulated directly. The large number of different words used assured a reasonable range of variations in duration, and this was assumed to co-vary randomly with the experimental factors of Word-Position, Task and Context.

To allow measurement of monitoring reaction-time, timing-pulses were placed on the second channel of the tape, so as to coincide with the onset of the target-word in the test-sentences. They consisted of a high-amplitude rectangular pulse, at 1300 Hz and of 30 msec duration, and were not heard by the subjects. The location of the timing-pulses was checked on a dual-trace storage oscilloscope, and was accurate to within ± 10 msec. During testing, the pulse started a digital millisecond timer, which was stopped by the subject's detection response (pressing a telegraph key).

A further set of 18 sentences was also constructed, to serve as practice materials for the subjects. These contained examples of all Prose by Task conditions.

Procedure

The subjects were first read a comprehensive set of instructions, describing the experimental situation and the different types of monitoring tasks and

prose contexts they would encounter, illustrated with examples. It was emphasized in the instructions, and during the practice session, that they should not try to guess the category of rhyme words in advance. They were told to respond as quickly and accurately as they could.

Each subject was tested separately in an IAC sound-attenuated booth. They heard the stimulus materials over headphones as a binaural monophonic signal, and responded to the target-words by pressing a telegraph key. There was a two-way voice communication channel open between the subject and the experimenter. To make sure that the subjects had read the instructions for each trial correctly, they were required to read aloud the contents of the relevant page of the instruction booklet before each trial started. For example, the subject would read aloud "Normal Prose; Rhyme; Bread". The subjects were told never to turn the page of the booklet onto the next set of instructions until the test-sentence they were hearing was completed. Thus once the subject was familiar with the situation, the sequence of events would be as follows. At the end of each test-sentence, the subject turned to the next page in the booklet and immediately read aloud the instructions. If he had done so correctly, and had not, for example, mispronounced the cue-word or missed out a page, then the next test-sentence followed at once.

Each experimental session lasted about 35 minutes, with a short break after every 30 trials to change instruction booklets. None of the subjects seemed to have any difficulty in performing the experiment, or in dealing with the variations in monitoring task and prose context.

Part I: The Perception of Words in Sentences

Results

Effects of listening task and prose context

The overall means for the nine Prose Context by Task conditions, summed over word-positions, are given in Table 1. The relationships between these means can be seen more readily in Figure 1.

Separate analyses of variance on subjects and on words were computed, using the untransformed monitoring reaction-times, with missing observations (less than 2%) replaced. Min F' values were derived from these two analyses. All statistics reported are reliable at the 0.05 level unless otherwise noted. The main effect of Prose Context was strongly significant ($\text{Min } F' (2,226) = 72.917$). Targets in Normal Prose, with an overall reaction time of 373 msec,

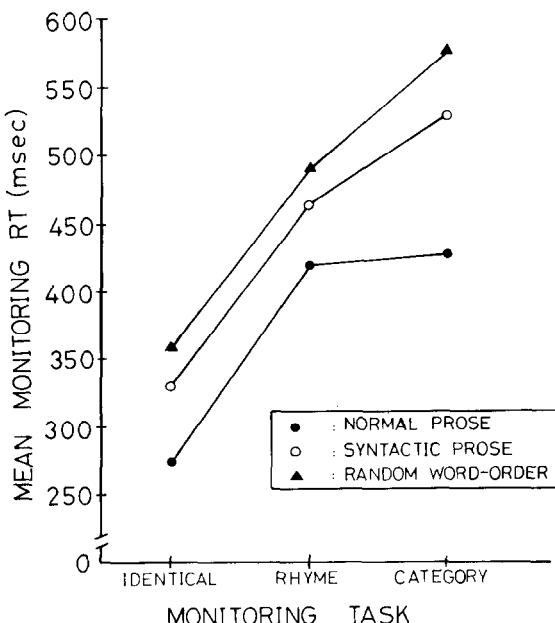
Table 1. *Experiment 1. Mean Monitoring Reaction-times (msec): By Prose Context and Monotoring Task.*

Prose Context	Monitoring Task		
	Identical	Rhyme	Category
Normal Prose	273	419	428
Syntactic Prose	331	463	528
Random Word-Order	358	492	578

Each value is the mean of 405 observations. Overall standard error = 8.75 msec.

were responded to faster than targets in Syntactic Prose (441 msec) or in Random Word-Order (476 msec). A planned contrast showed the difference between Syntactic Prose and Random Word-Order to be significant ($\text{Min } F' (1,226) = 16.262$).

These results confirm, first of all, the basic hypothesis that the monitoring task is sensitive to both syntactic and semantic context variables. Reaction-

Figure 1. *Experiment 1: mean monitoring latencies, by Prose Context and Monitoring task.*

times are fastest when both syntactic and semantic information are present, but are still significantly facilitated even when semantic constraints are absent and only syntactic information is available. The facilitation due to the presence of a semantic level of interpretation (the advantage of Normal over Syntactic Prose) is twice as large (68 versus 35 msec) as the facilitation due to the presence of a syntactic analysis alone (Syntactic Prose versus Random Word-Order).

The second significant main effect, of Monitoring Task ($\text{Min F}' (2,226) = 285.370$), reflects the large overall differences between Identical monitoring (321 msec), Rhyme Monitoring (458 msec), and Category Monitoring (511 msec). There was also a significant interaction between Monitoring Task and Prose Context ($\text{Min F}' (4,462) = 5.998$). This reflects the differing effects of Task across the three Contexts.

The effects of Prose Context are very similar across the Identical and Rhyme Monitoring conditions, with the increase in monitoring reaction-time for the three contexts ranging from 132 to 146 msec. The comparison between Rhyme and Category is more complex. The two tasks do not differ in overall reaction-time under the Normal Prose condition, whereas Category Monitoring is much slower than Rhyme Monitoring in Syntactic Prose and Random Word-Order, with increases of 65 and 86 msec respectively. This pattern is confirmed by pair-wise comparisons between the nine individual means, using the Newman-Keuls statistic, and incorporating an error term derived from the appropriate $\text{Min F}'$ ratio. For all except three cases the means were significantly different from each other at well beyond the 0.01 level. The 27 msec difference between Syntactic Prose and Random Word-Order Identical, and the 29 msec difference between Syntactic Prose and Random Word-Order Rhyme, were both significant at the 0.05 level. The 9 msec difference between Normal Prose Category and Rhyme fell far short of significance. Thus the semantically based decisions required in Category monitoring need take no longer than the phonologically based decisions required in Rhyme monitoring. But this only holds for Normal Prose, where there is a semantic context with which the identification of the target-word can interact.

Effects of word-length

The second aspect of the results that is relevant to the local structure of processing concerns the effects of word-length on monitoring reaction-time. The 81 target-words each appeared in all three prose contexts. The length of the words in milliseconds was measured for each occurrence, using a storage oscilloscope. There were small but consistent differences in mean length be-

tween words said in Normal Prose, which had a mean length of 369 msec, and words in Syntactic Prose (384 msec) and in Random Word-Order (394 msec). An analysis of variance on the word-lengths showed this difference to be significant ($F(2,144) = 5.519$), although none of the overall means differed significantly from each other. There was no effect of Word-Position on word-length ($F < 1$), nor an interaction between Prose and Word-Position ($F < 1$).⁷

The first point about these measurements of word-length is that the mean durations of the words in each prose context are clearly longer than the mean Identical monitoring latencies in each context. Secondly, the relationship between word-length and reaction-time can be shown to change as a function of prose context. The results of a detailed analysis of this relationship are given in Table 2. In this analysis the mean reaction-times for each word were sorted according to the temporal duration of the word. The range of word-lengths covered 300 msec, from 220 to 520 msec (the four two-syllable target-words were not included in this analysis). The mean pooled reaction-times were then computed for all the words falling into each successive 20 msec time-interval (from 220 to 520 msec). A regression analysis was then carried out on the means derived in this way. Table 2 gives the slopes and correlation coefficients for each of the three prose contexts.

There are significant overall effects of word-length in all contexts, and these effects differ across contexts. To the extent that a prose context makes semantic and syntactic information available for interaction with word-recognition decisions, so the dependency between monitoring reaction-time and word-length appears to diminish. Thus the weakest effects are in Normal Prose, where word-length accounts for only 32% of the variance.⁸ The slope, of +0.22 msec, indicates that for every one msec increase in word-length there will be a 0.2 msec increase in reaction-time. Over the 300 msec range of word-durations, this would lead to a total word-length effect of 65 msec. The correlation with word-length is somewhat stronger in Syntactic Prose with 53% of the variance accounted for, and with a slope of +0.27. This would produce an overall effect of about 80 msec. Much the strongest effects are in Random Word-Order, with 87% of the variance accounted for, and a slope of +0.49 — leading to a total effect of 145 msec over the entire range of word-durations. The slopes for all three prose contexts are significantly different from zero, and the Normal Prose ($t(26) = 2.530$) and the Syntactic

⁷ As a further precaution, we also ran an analysis of co-variance, repeating the earlier analyses of reaction-time but with word-length entered as a co-variate. The adjustment for word-length made very little difference to the means, and did not affect the statistical comparisons between them.

⁸ The square of the correlation coefficient gives the percentage of the variance of the raw monitoring latencies that is accounted for by the linear effects of word-length.

Table 2. *Experiment 1. Monitoring Reaction-time and Word-Length: Regressions on Raw and Smoothed Means (by Prose Context)*

	Prose Context		
	Normal Prose	Syntactic Prose	Random Word-Order
Raw Means Analysis			
r =	+0.57	+0.73	+0.93
b =	+0.22	+0.27	+0.49
Smoothed and Raw Means Analysis			
r =	+0.77	+0.89	+0.96
b =	+0.21	+0.28	+0.48

r = correlation coefficient, b = slope in msec (n = 15).

Prose ($t(26) = 2.400$) slopes are both significantly different from the Random Word-Order slope. It is clear that the interactions between word-length effects and prose contexts are not strictly additive.

In addition to the analyses on the raw means, we also took the precaution of computing regressions on a moderately smoothed version of the data, using a smoothing method (running medians followed by running means) recommended by Tukey (1972; Wainer and Thissen, 1975).⁹ The procedure followed was to produce fully smoothed means, and then to combine the raw and the smoothed sets of means. The outcome of the regression analyses based on these combined curves are also given in Table 2. The close similarity between the slopes estimated by the two methods indicates that the estimates are satisfactorily robust.

Individual analyses were also carried out for each of the nine combinations of monitoring task by prose context conditions. The pattern of differences found for the overall means holds for each of the monitoring tasks as well. In each case the slope is shallowest in Normal Prose, about 30% steeper in Syntactic Prose, and twice as steep in Random Word-Order.

Error rate

The error rate in the experiment was very low, with an overall level of 1.3% (counting failures to detect the target and responses to the wrong word). As

⁹The reason for doing this is the notorious sensitivity of linear regression to outlying values (cf., Wainer and Thissen, 1976). The presence of just one or two outliers can have disproportionate effects on the outcome of a regression analysis. Since the questions we are asking here depend on the accuracy with which the slopes of the regression lines are estimated, it seemed important to make sure that these estimates were robust.

far as the small numbers involved allow any firm conclusions, the error-rate increases with mean reaction time. The error-rate was especially low in Identical monitoring (less than 0.5%).

Discussion

The claim we are testing in this part of the experiment is whether or not spoken word recognition involves a direct interaction between the sensory input and contextual constraints. The combination of monitoring tasks and prose contexts allowed a number of tests of this claim, and in each case the results favored the interactive analysis of word-recognition processes. Both syntactic and semantic context facilitate monitoring responses, reaction-times in Identical monitoring are considerably shorter than the durations of the target-words, semantic attribute-matching is not necessarily delayed relative to phonological attribute-matching, and the changes in the effects of word-length across prose-contexts indicates that sensory and contextual inputs interact at the same stage in processing.

These results will now be discussed in more detail, from the point of view of their implications for a more precise statement of an interactive model of spoken word recognition. We will do this in the context of an analysis of the three monitoring tasks. The results will be shown to support our assumption that all three tasks have in common the same initial word-identification stage, followed in Rhyme and Category monitoring by an attribute-matching stage. Identical monitoring appears not to require such a second stage, so that this task will provide the most direct information about the temporal structure of the operations of the first, word-recognition stage.

The Rhyme and Category tasks will provide a somewhat different kind of information about spoken word-recognition. What they reflect are the qualitative properties of the kinds of internal representation that are activated when a word is being recognised. In particular, whether the immediately available lexical representation includes information about the semantic properties of the word.

The relationship between word-monitoring and word-recognition

Identical monitoring

The assumption in using this task was that it would be a relatively pure reflection of word-recognition processes. We need to defend this assumption against two possible objections; first, by showing that the task did indeed involve recognizing the word specified as the target.

Response latencies in Identical monitoring were surprisingly short relative to the durations of the target-words. In Normal Prose, for example, the mean monitoring latency was 273 msec, which is 94 msec shorter than the mean duration of the targets. If, in addition, we estimate the time for response execution to be of the order of 50–75 msec,¹⁰ then in Normal Prose the subject begins to execute the response after only having heard about 200 msec of the input corresponding to the target-word. This means that the subjects were responding after they could only have heard the first two or three phonemes of the word. But even so it is quite clear that they were not treating Identical monitoring as a phoneme-monitoring task.

First, the reaction-times here were much faster than those observed in true phoneme-monitoring tasks. In experiments where phoneme-targets are monitored for in sentence contexts, the latencies fall into the range of 450–500 msec (e.g., McNeill and Lindig, 1973; Morton and Long, 1976). Secondly, Identical monitoring is significantly affected by prose context, with the largest difference being between Normal and Syntactic Prose. It is difficult to see how semantic constraints could affect the detection of a phoneme *per se*.

A second potential worry about the Identical monitoring task is that the listener's advance knowledge of the target-word could in some way enable him to seriously alter or even circumvent normal word-recognition procedures. But this seems excluded on two grounds.

First, there is the general point, drawn from recent phoneme-monitoring studies, that the listener can only have immediate access to the sensory input in the context of its apparently obligatory lexical interpretation (when it can be so interpreted). This means that knowing the target in advance could not enable the listener to match the input against the target in some way which avoided the initial word-recognition process.

The second point is that when word-recognition is studied under conditions where *no* advance information at all is given, then words seem to be recognized just as fast as in Identical monitoring, and equivalent effects of semantic context are obtained. The task in question is speech shadowing, in which some subjects can repeat back normal prose materials accurately and clearly at response delays of 250–275 msec (Marslen-Wilson, 1973; 1975;

¹⁰By "response execution" we mean the component of the reaction-time that involves deciding to make a response and performing the actions required. It is difficult to find a clear estimate of the duration of this component. The conservative value we have chosen here (50–75 msec) is not especially critical, and could be changed in either direction without affecting our arguments. If, for example, response execution requires as much as 125 msec, then this would only strengthen our claims for the earliness of word-recognition. If the estimate is reduced, say to an implausible 25 msec, this still means that words are being recognized well before all of them could have been heard.

Note 1). Again allowing about 50–75 msec for response integration and execution, this level of performance also indicates that words in sentence contexts are being correctly identified within about 200 msec of their onset. Furthermore, when these close shadowers are given passages of Syntactic Prose to shadow, their mean latencies increase by about 57 msec (Marslen-Wilson, Note 1) — a value which is similar to the 58 msec increase found here for Identical monitoring in Syntactic Prose.

This shadowing evidence, and the arguments presented earlier, lead to the conclusion that Identical monitoring not only does reflect word-recognition processes, but does so in a relatively normal manner. We will assume that the same type of word-recognition process underlies performance in the other two monitoring tasks.

Rhyme monitoring

On the analysis assumed here, this task involves two response stages. First, the word needs to be identified, and the speed of this decision is affected by the degree of constraint provided by the prose context. The task then requires an additional attribute-matching stage which adds an average of 137 msec to the response latencies (relative to Identical monitoring). In the matching process, the subject has to obtain access to his internal representation of the word at the level of description specified for the target — in this case, the phonological properties of the terminal segments of the word.

Phonological attribute-matching takes the same amount of time in all prose contexts. The increase from Identical to Rhyme is 146 msec in Normal Prose, 134 msec in Random Word-Order, and 132 msec in Syntactic Prose. Thus whatever advantages accrued to Normal and Syntactic Prose contexts in Identical monitoring are transferred essentially unchanged to the Rhyme monitoring situation. These are the differential advantages that would derive from the presence or absence of syntactic and semantic facilitation of word-identification decisions.

The Rhyme monitoring results are also evidence for a word-identification stage of the same order of rapidity as that postulated for Identical monitoring. In Normal Prose the mean Rhyme monitoring latency is 419 msec. Since the mean word-length in Normal Prose is 369 msec, this means that the subjects are making their responses within 50 msec after the end of the word. Assuming that response execution takes about 50–75 msec, this implies that the subjects were completing the attribute-matching process even before they had heard all of the word.

The most reasonable explanation for this in the present context is that they, in some sense, knew what the word was (and, therefore, how it would

end) relatively early in the word. Thus they could go ahead and begin the matching process independently of when the physical input corresponding to the rhyming syllable was actually heard. The alternative explanation – that the subjects were matching at a sublexical level, segment-by-segment as the word was heard – can be excluded for the same reasons as this kind of explanation was excluded in Identical monitoring. It provides no basis for the differential effects of prose context on performance in the task, nor for why the size of these differences should be the same in both Identical and Rhyme monitoring.

The Rhyme monitoring task, therefore, corroborates the analysis of the time-course of word-identification that was derived from the Identical monitoring results. It also supports the assumption that information about the phonetic properties of spoken words is not directly available for a subject's response independently of the lexical interpretation of the input. Finally, the task provides a base-line for comparison with the Category monitoring task, which depends on access to the semantic rather than the phonological properties of the target.

Category monitoring

Unlike Rhyme monitoring, both response stages in Category monitoring interact with prose context. The word-identification stage is assumed to be similar to that in Rhyme and Identical, and to be facilitated in the same ways by contextual variables. The attribute-matching stage involves finding out whether the semantic properties of the word being identified match the semantic category specified for the target. Unlike phonological attribute-matching, this process is strongly facilitated by semantic context, as is shown by the increase in the difference between Normal Prose and the other two contexts. Comparing Rhyme monitoring with Category monitoring, the advantage of Normal over Syntactic Prose increases by 66 msec, and of Normal Prose over Random Word-Order by 77 msec.

The results also show that semantic information about a word must be available very early in processing, in a normal sentence context. The mean Category monitoring response latency in Normal Prose is 428 msec, which is only 59 msec longer than the mean duration of the word-targets. Thus, just as in Rhyme monitoring, the attribute-matching process must have been started well before all of the word had been heard.

This is exactly what the on-line interactive analysis of word-recognition requires. If semantic context is to affect word-recognition decisions, then it must do so on the basis of the match between the requirements of semantic context and the semantic properties of potential word-candidates. It is this

early accessing of semantic information that makes possible the rapid monitoring decisions obtained here. And the parallel between the Rhyme and Category monitoring results in Normal Prose shows that no special temporal priority should be assigned to the phonological rather than the semantic properties of words being recognized in normal sentence contexts.

Implications for an interactive spoken-word recognition theory

Perhaps the most informative aspect of the results is the evidence (together with the shadowing data) that they provide for the *earliness* of word-recognition in normal sentence contexts. In the Normal Prose conditions the subject is initiating his Identical monitoring response after he has only heard the first 200 msec of the word. With a mean word-length of 369 msec in Normal Prose, the subject is responding well before he has heard the whole word. This finding, taken together with the evidence for contextual interactions with these early recognition decisions, and the failure of additivity for the word-length/prose-context interaction, places strong constraints on the possible functional structure of the spoken-word recognition system.

If facilitatory context-effects are found in a word-recognition task, then this means that, at some point in the lexical access process, the listener's internally represented knowledge of the situation in which the word occurs is able to interact with this process. The central requirement of the autonomy hypothesis is that this interaction takes place after the correct word has been accessed. A word-candidate is selected on the basis of acoustic-phonetic information alone, and only then may be checked against the requirements of context (e.g., Forster, 1976). The evidence for the earliness of word-recognition seems incompatible with this type of model. The word is recognized so early that it is difficult to see how sufficient acoustic-phonetic information could have accumulated to allow a unique word-candidate to have been selected.

The first 200 msec of a word would correspond, in general, to the first two phonemes of the word. It is possible to compute the number of words in the language that are compatible with any given initial two-phoneme sequence. Thus there are, for example, about 35 words in American English that begin with the sequence /ve/ (Kenyon and Knott, 1953). For the 81 words used as targets in the present study, the median number of words possible for each initial sequence was 29. A selection process operating solely on the acoustic-phonetic input would indeed have only a small chance of selecting the single correct word-candidate.

These arguments lead, then, to a recognition system in which the acoustic-phonetic input can be consistent with a range of possible word-candidates at

the moment of recognition, and where, evidently, it is contextual factors that allow a single word-candidate to be selected from among these candidates. The properties of the interaction between word-length effects and prose-contexts further require that contextual and sensory inputs interact continuously at the same stage of processing, and that the interactions are not mediated by a discontinuous, two-stage model.

A discontinuous model could maintain the autonomy of an initial selection process (although giving up the claim that the word was the processing unit) by assuming that a pool of word-candidates is activated at the first stage on the basis of the acoustic-phonetic properties of the input. At the second stage a set of contextual criteria would be applied to select a single choice from among this initial pool. Thus the contributions of sensory and contextual criteria would still be serially ordered, with the contribution of the one terminating before the arrival of the other. This type of model seems to be the last refuge of the autonomy hypothesis.

Both the changing effects of word-length, and more general considerations, argue against this type of solution. The increasing slopes of the regressions on word-length (Table 2) clearly reflect the increasing dependency of the recognition-decision upon acoustic-phonetic information as the available contextual inputs diminish. But on a two-stage model the contribution of acoustic-phonetic information should remain constant. The subsequent context-driven decision process might vary in speed as a function of prose-context, but the basic relationship between acoustic-phonetic variables and the recognition decision should still have the same slope.

This signals a more general problem for any two-stage model with an autonomous first stage: it is in principle not equipped to vary the amount of acoustic-phonetic information that is input to the recognition process. But it is clear that, as contextual constraints vary, so too will the amount of sensory information that is necessary to reach a unique word-choice. This applies not just to the large variations across prose contexts here, but also to the intrinsic variation in the constraints on words at different points in natural utterances. The only way a two-stage model could deal with this would be to allow a return to the first stage whenever the contextual selection process failed to achieve a unique selection from among the first set of word-candidates. But this would violate the autonomy assumption, which was the only motivation for proposing this two-stage model in the first place.

These arguments bring us to the point where we will have to turn to a quite different type of processing model. That is, a distributed processing model in which recognition is mediated by a large array of individual recognition elements, each of which can integrate sensory and contextual information in order to determine whether the word it represents is present in the signal.

This seems the only way to allow contextual factors to interact simultaneously with potentially quite large sets of word-candidates, while at the same time allowing acoustic-phonetic information to continue to be input to the decision process until the point at which a unique word-candidate can be selected.

The best-known model of this type is the logogen theory developed by Morton (1969; 1979; Morton and Long, 1976). There are, however, problems with this model arising from the decision mechanism assigned to the recognition elements (or logogens). In this approach, the recognition-decision within each logogen is based on the concept of a threshold; a word is recognized if and only if the logogen corresponding to it crosses its decision threshold. But, as we have argued in more detail elsewhere (Marslen-Wilson and Welsh, 1978), the notion of an all-or-none threshold leads to incorrect predictions in certain experimental situations.

In the final section of this part of the paper we will describe an alternative recognition model, and show how it accounts for the results reported here. The critical difference with the logogen model is that the recognition elements will be assigned the ability to respond to *mismatches*. Each element is assumed to be able to determine that the word it represents is not present in the signal if either the sensory or the contextual environment fails to match its internal specifications of the properties of the word. This assumption allows us to avoid using decision-thresholds as the basis for the recognition process.

A “cohort”-based interactive recognition theory

The individual recognition elements in this theory are assumed to be directly accessed from the bottom-up during speech recognition. Information about the acoustic-phonetic input is broadcast simultaneously to the entire array of lexical recognition elements. The pattern-matching capacities of these elements allows the relevant elements to become active when they pick up a pattern in the signal which matches their internal specifications of the acoustic-phonetic properties of the words they represent. This means that, early in a word, a group of recognition elements will become active that corresponds in size to the number of words in the language (known, that is, to the listener) that begin with the sound sequence that has been heard up to that point. This preliminary group of word-candidates will be referred to as the *word-initial cohort*.

The subsequent time-course of word-recognition will depend on the ways in which the word-initial cohort can be reduced in size until only a single

word-candidate is left. This can be done even when only acoustic-phonetic information is available (as when a word is heard in isolation). Each recognition element activated early in the word will continue to monitor the incoming acoustic-phonetic signal. As more of the signal is heard, it will tend to diverge from the internal specifications of more and more of the members of the word-initial cohort. These elements will detect the resulting mismatches, and will therefore drop out of consideration as candidates for recognition. This process will continue until only one candidate is left whose properties still match the signal. It is at this point that we assume that the recognition decision can occur, so that the recognition point for any given word can be precisely predicted by examining the sequential structure of the word-initial cohort in which it occurs.¹¹

These recognition elements must also be able to interact with contextual constraints. The semantic and syntactic properties of the members of the word-initial cohort need to be assessed against the requirements of the sentential context in which the word to be recognized is occurring. If this can be achieved, then those recognition elements whose properties do not match the requirements of context can be dropped from the word-initial cohort.

It is not *a priori* clear how these necessary contextual effects should actually be realized in a cohort-based recognition system. One suggestive possibility, which is compatible with the cohort approach so far, is that each recognition element should be thought of as analogous to a procedurally defined "specialist" on the word it represents – in the way that some of the words in Winograd's SHRDLU system have specialists associated with them (Winograd, 1972), and that all of the words in the system proposed by Rieger (1977; 1978) are represented by such specialists. This would mean that each word would have built into its mental representation not simply a listing of its syntactic and semantic properties, but rather sets of procedures for determining which, if any, of the senses of the word were mappable onto the representation of the utterance up to that point.¹²

An arrangement of this type – with the recognition elements being simultaneously sensitive to both contextual and acoustic-phonetic factors – would produce the flexible and continuous interactions between these factors that we have argued for. Each member of the word-initial cohort will continue to accept bottom-up information until the criterial degree of mismatch is

¹¹ Some recent preliminary experiments (Marslen-Wilson, Note 2), testing these types of prediction, have shown a close correlation between recognition-time for a word (heard in isolation) and the point at which that word uniquely diverges from its word-initial cohort.

¹² Note that this type of claim further differentiates the cohort approach from word-detector theories of the logogen type. The extensive knowledge being built into the recognition elements here, would, in the logogen model, be assigned to the "Context System" and not to the logogens themselves.

achieved either with this bottom-up input or with the requirements of the contextual environment. This leads to an automatic optimisation, for any given word in any given context, of the balance between acoustic-phonetic and contextual information sources. Exactly as much acoustic-phonetic information will be analyzed as is necessary for the system to reach a single word-choice given the constraints available from context. And since there is no discontinuity in the contribution of bottom-up information to the decision process, we avoid the complications of trying to restrict the bottom-up contribution to some pre-specified fixed proportion of the relevant signal. The weaker the contextual constraints, the more the system will depend on bottom-up inputs, and this derives perfectly naturally from the interactive distributed processing system we have proposed.

The functional outline of the proposed recognition system is now clear enough, and can account for the major results we have reported. First, it is consistent with the obtained effects of context on word-recognition time. The presence of semantic and syntactic constraints will immediately reduce the size of the initial decision space, so that, on the average, less acoustic-phonetic input will need to be processed before a single word-choice can be selected. The comparisons between the overall means in each prose context show that semantic constraints are almost twice as effective as syntactic constraints alone in speeding response time. This makes sense if we consider that many words are syntactically ambiguous – it is, for example, difficult to find nouns that cannot also function as verbs. This means that word-candidates can less readily be deleted from the initial cohort on the basis of syntactic constraints alone – bearing in mind that in Syntactic Prose the syntactic constraints can do little more than specify the likely form-class of the word to be recognized. Whereas the effectiveness of semantic constraints in reducing the size of the cohort can be potentially much greater.

Secondly, the cohort theory provides a basis for understanding how words can be recognized so early. Our earlier analysis of the target-words used here shows that the median number of word-candidates left in the cohort after two phonemes had been heard should be about 29. If we go one phoneme further, then the median size of the cohort shrinks to about six members. And this is only considering the acoustic-phonetic input to the decision process. This explains how the words heard without context (in Random Word-Order) could be responded to relatively early. The mean Identical monitoring reaction-time in Random Word-Order was 358 msec, which, compared to the mean word length of 394 msec, shows that even without a constraining sentential context a monosyllabic word can often be recognized before all of it has been heard. Set against this result, the earliness of recognition in Normal Prose contexts is no longer very surprising.

Thirdly, the theory allows for the type of continuous interaction between top-down and bottom-up information that is necessary to account for the changes in word-length effects across contexts. But first we have to show where word-length effects come from in a cohort theory, especially in cases where, as in the present experiment, these effects are detectable even when responses are being made before all of the word has been heard. The answer lies in the sources for variations in the duration of monosyllabic words. If we sort the target-words in this experiment according to their relative durations, then we find that the longer words, reasonably enough, are those that begin or end with consonant clusters, and that contain long as opposed to short vowels. Thus "brass" and "floor" are relatively long words, and "bed" and "cat" are relatively short.

Let us further assume, consistent with findings in speech research, that vowels are acoustically-phonetically critical in the interpretation of the input signal — in particular the transitions at the beginnings and ends of vowels into their adjoining consonants. Clearly, this critical information would become available to the listener later in a word when the word began with a consonant cluster. This in turn would lead to a delay, relative to words beginning with a single consonant, in the accumulation of the acoustic-phonetic information necessary for defining the word-initial cohort, and for reducing it in size later in the word. Thus, in so far as variations in duration are due to these effects at the beginnings of words, they will lead to apparent effects of overall duration on monitoring responses that are made before all of the word has been heard.

This would explain how effects of "word-length" are detectable even in the Identical monitoring conditions. Furthermore, the size of such effects will tend to increase when the recognition response requires more acoustic-phonetic information about a word, as in Random Word-Order Prose. For example, if a CVC target-word contains a long vowel, then this will delay, relative to words containing short vowels, the point at which the subject starts to obtain information about the last consonant in the word. This will slow down recognition-time to the extent that the last phoneme needs to be known in order to reduce the cohort to a single candidate. But when contextual inputs to the cohort-reduction processes are stronger, as in Normal and Syntactic Prose, then it becomes more likely that recognition will not require information about this last phoneme. Thus any temporal delay in the arrival of this phoneme will tend to have less effect on response-time.

Finally, the cohort theory has built into it the early accessing of semantic information about a word that is necessary to account for the Category monitoring results in Normal Prose, as well as for the results of earlier experiments (e.g., Marslen-Wilson, 1975). According to the theory, this type of informa-

tion about a word is being activated even before the word has been recognized as part of the process of cohort-reduction. At the point where recognition does occur – no matter how early in the word – semantic information will have played an important role in the recognition decision, and will be immediately available for the making of an attribute-matching decision. But when a word occurs in Syntactic Prose or Random Word-Order contexts, then recognition cannot involve the same dependence on the semantic properties of the word-candidates, since this would lead to the incorrect rejection of candidates that did in fact fit the acoustic-phonetic input. Thus in these contexts there will be an asymmetry in the extent to which semantic as opposed to phonological information about the word can have been actively exploited prior to recognition. This would lead to a relative delay in the availability of the semantic information necessary for semantic attribute-matching. In Syntactic Prose and Random Word-Order the Category monitoring responses are, respectively, 65 and 86 msec slower than the Rhyme monitoring responses. Whereas the two tasks do not differ in the Normal Prose conditions.

In the light of these analyses of the on-line interactive properties of the local structure of processing, we are now in a position to interpret the global structural effects. The next part of the paper will begin by presenting the relevant results – those involving the word-position variable.

Part II: The Global Structure of Sentence Processing

Results

The results reported here derive from the same data as were described in the earlier results section. As before, all statistics are based on the combination of separate analyses on subjects and on words.

Effects of word-position

The third main variable in Experiment One was Word-Position, for which nine word-positions were tested, from the second to the tenth serial position in the second of each pair of test sentences. Nine different target-words occurred in each word-position, and each word was tested, across subjects, in all combinations of Prose Context and Monitoring Task. The effects of Word-Position were investigated using trend analyses to determine the shapes of the word-position curves for the different Prose and Task conditions. Where appropriate, regression analyses of the raw and smoothed reaction-times across word-position were also carried out.

The main effect of Word-Position was not significant ($\text{Min } F' < 1$), nor was the interaction between Word-Position and Monitoring Task ($\text{Min } F' < 1$), or any of the trend components (all trend analyses were carried as far as the quintic component).

The interaction between Word-Position and Prose Context also fell short of full significance ($\text{Min } F' (16,250) = 1.558, p < 0.10$). However the breakdown of this interaction into trend components showed significant linear ($\text{Min } F' (2,250) = 3.026$) and cubic ($\text{Min } F' (2,250) = 4.202$) effects. This was because there were strong linear effects in Normal Prose and Syntactic Prose, but none at all in Random Word-Order (see Table 3). Conversely, there was a strong cubic component in the Random Word-Order curve, but none at all for the other two prose contexts. Random Word-Order also exhibited a fourth-order component ($p < 0.10$). These word-position effects are plotted for each prose context in Figure 2.

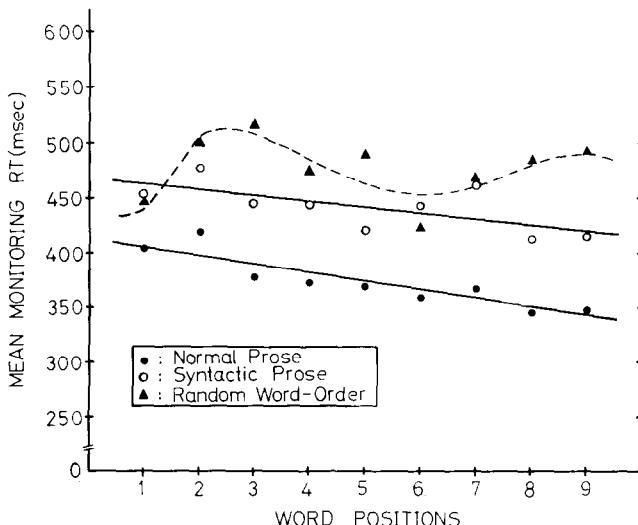
The further discussion here will focus on the relative slopes and intercepts of the linear effects in Normal and Syntactic Prose in each monitoring task, since, with the exception of Syntactic Prose Category, *only* the linear components of the curves in these conditions were significant. We followed here the same precautions as in the analysis of the word-length effects, and computed slope estimates for the combined smoothed and raw word-position curves as well. These are given in Table 4 together with the results of the analyses on the raw means. The combined estimates will be used as the basis for the comparisons between slopes. The three Random Word-Order curves, and the Syntactic Prose Category curve, were treated in the same ways, to allow comparison with the other curves. The individual word-position results for the nine prose by task conditions are plotted in Figure 3. This figure shows

Table 3. *Experiment 1. Monitoring Reaction-Time and Word-Position: Regressions on Raw and Smoothed Means (by Prose Context)*

Prose Context			
	Normal Prose	Syntactic Prose	Random Word-Order
Raw Means Analysis			
r =	-0.90	-0.65	-0.01
b =	-8.05	-5.23	-0.12
Smoothed and Raw Means Analysis			
r =	-0.95	-0.87	-0.22
b =	-7.55	-5.20	-1.42

r = correlation coefficient, b = slope in msec (n = 9).

Figure 2. *Experiment 1: Word-position means for each Prose Context, collapsed across Monitoring Tasks. The linear components of the Normal Prose and Syntactic Prose trends are plotted, and the combined cubic and quartic trends for Random Word-Order. Note: Word-Position 1 on the figure corresponds to the second word in the test-sentence. Thus the zero intercept represents an estimate of monitoring latency at the actual first word of the utterance.*



the raw means together with the linear regression lines derived from the combined raw and smoothed curves. Combined estimates of the overall word-position effect for each prose context (collapsing across monitoring task) are included in Table 3.

The major results can be summarized as follows. In Normal Prose, performance is dominated throughout by linear effects of word-position. In the overall analysis (Table 3), the linear component of the word-position curves accounts for over 80% of the variance in the raw means. This is a surprisingly high value, given the different words at each word-position. The separate analyses for each monitoring task in Normal Prose (Table 4) shows that these linear effects are highly consistent across tasks. In each case word-position accounts for about 61% of the variance in the raw means.

In Syntactic Prose the overall linear effect is somewhat weaker, with only 42% of the raw means variance being accounted for (Table 3). Turning to the individual monitoring tasks, we find strong linear effects in Identical monitoring, weaker but still significant effects in Rhyme monitoring, and no linear effects at all in Category monitoring. The word-position trend for this condition is best described by a fifth-order curve ($p < 0.10$).

Figure 3. Experiment 1: raw Word-Position means for each combination of Prose Contexts and Monitoring Tasks. Best-fitting least-squares regression-lines, based on combined raw and smoothed means (see text) are plotted. Note: Word-Position 1 corresponds to the second word in the test-sentence.

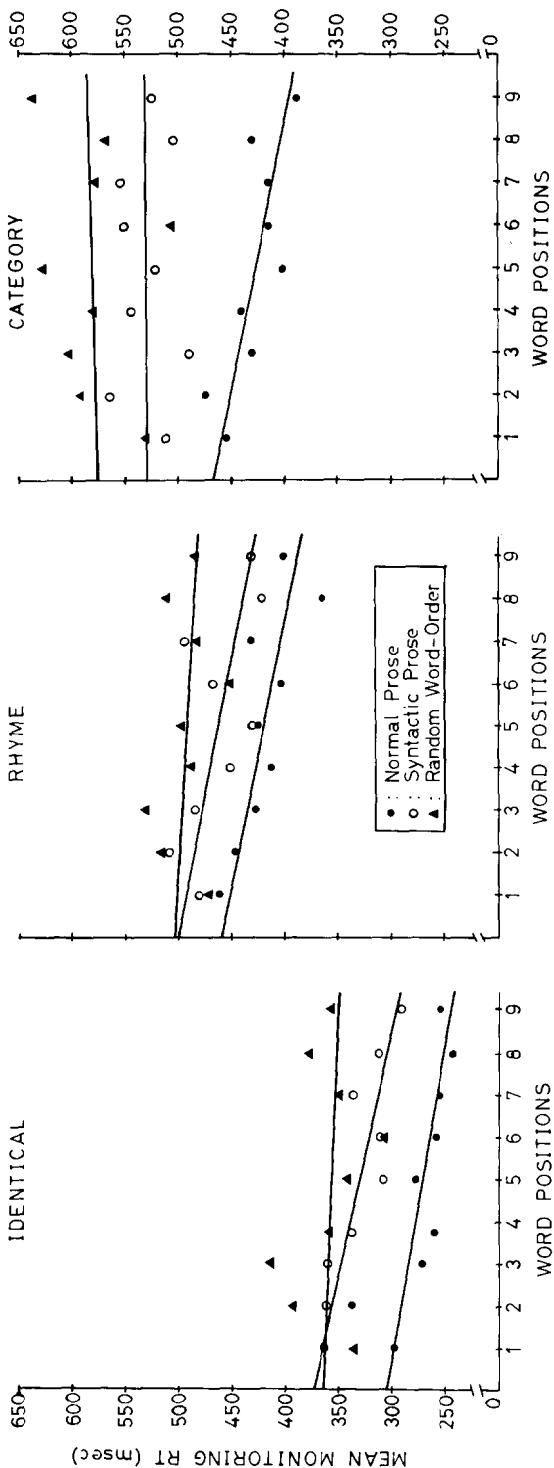


Table 4. *Experiment 1. Monitoring Reaction-time and Word-Position: Regressions on Raw and Smoothed Means (By Prose Context and Monitoring Task)^a*

Prose Context	Monitoring Task		
	Identical	Rhyme	Category
Normal Prose			
r =	-0.78 (-0.88)	-0.79 (-0.89)	-0.78 (-0.91)
b =	-8.20 (-6.88)	-8.23 (-7.92)	-7.67 (-8.00)
Syntactic Prose			
r =	-0.87 (-0.95)	-0.63 (-0.87)	-0.01 (+0.02)
b =	-8.58 (-8.52)	-7.30 (-7.63)	-0.07 (+0.12)
Random Word-Order			
r =	-0.21 (-0.28)	-0.19 (-0.44)	+0.25 (+0.11)
b =	-2.35 (-1.72)	-1.68 (-2.33)	+3.80 (+0.82)

r = correlation coefficient, b = slope in msec (n = 9).

^aResults of regressions on the combined raw and smoothed means are given, in parentheses, following the raw means results.

In none of the Random Word-Order results is more than 7% of the variance in the raw means analyses accounted for by the linear word-position component -- either in the overall analysis (Table 3), or when each monitoring task is considered separately (Table 4). Instead, the curves are best described by an assortment of cubic and quartic components. Given the strong correlation between Random Word-Order responses and word-length, it is likely that these curves primarily reflect variations in the mean durations of the words at different word-positions.

Table 4 and Figure 3 give the slopes estimated for each test condition. The slopes for the three Normal Prose conditions and for Syntactic Prose Identical and Rhyme are very similar, ranging from 6.9 to 8.5 msec decrease in reaction-time per word-position, and in each case significantly different from zero. In Rhyme and Identical monitoring the Normal Prose and Syntactic Prose slopes are not significantly different from each other, while both differ significantly in each context from the Random Word-Order slope. In Category monitoring, however, the Normal Prose slope is significantly different from the Syntactic Prose slope ($t(14) = 3.202$), as well as from the Random Word-Order slope ($t(14) = 2.246$). But the Syntactic Prose and Random Word-Order slopes did not differ in this condition.

The parallel slopes for Normal and Syntactic Prose in Identical and Rhyme monitoring mean that the differences in overall reaction-time between these conditions are not due to divergences in the degree of facilitation across

word-positions. Instead, the differences derive from the intercepts of the word-position effects — that is, the estimated value at word-position one. Table 5 lists the estimated intercepts for the nine prose by task conditions, as well as for the overall prose-context curves. In the Identical and Rhyme monitoring conditions, Normal Prose has a significant advantage ($p < 0.01$) over Syntactic Prose at the beginning of the sentence of about 50 msec. This difference remains fairly constant over the rest of the sentence. Only in Category monitoring is the intercept difference ($p < 0.01$) between Normal and Syntactic Prose accompanied by a difference in slope (see Figure 3 and Table 4).

In the comparison between Normal Prose and Random Word-Order we find significant differences ($p < 0.01$) in both intercept and slope in all monitoring conditions. Reaction-times differ at the beginning of a sentence, and continue to diverge throughout the rest of the sentence (see Figure 3). Comparing Syntactic Prose and Random Word-Order, the intercepts are the same in Identical and Rhyme monitoring, but the scores diverge throughout the rest of the sentence (Figure 3). There is little to be said about Category Monitoring performance in Syntactic Prose and Random Word-Order — reaction-times differ unsystematically by different amounts at various points in the sentence.

Discussion

Before discussing the global structural results, some preliminary points need to be made. First, it should be emphasized that we are concerned here only with the general structure of global processing. The word-position curves in this experiment are the outcome of responses collapsed over 81 different sentences, all varying somewhat in their global structural properties. It is presumably for this reason that only the linear components of the word-position

Table 5. *Experiment 1. Monitoring Reaction-time and Word-Position: Combined Raw and Smoothed Means Estimates of Intercept (msec)*

Prose Context	Monitoring Task			
	Identical	Rhyme	Category	Overall
Normal Prose	305	459	466	410
Syntactic Prose	373	500	529	465
Random Word-Order	364	504	576	488

curves in Normal and Syntactic Prose were significant.¹³ Averaging over many different sentences, the results reflect the overall tendency for structural constraints to be stronger later in a sentence than earlier. The experiment was not intended to capture the detailed properties of global processing for any one sentence, or for any one sentence type. Had it been restricted in this way, then higher-order word-position effects would no doubt have been observed (as for example in the recent work by Harris (Note 3)).

A second preliminary point concerns the possible role of a serial-position response-bias in the word-position curves. In an experiment of this sort, it is at least possible that the subjects' readiness to respond will increase the later in the string that the target occurs. But the complete absence of any linear effect in the Random Word-Order curves suggests that this kind of response bias was not affecting performance in this experiment. This was presumably because of the length, complexity, and variety of the stimuli, as well as the wide distribution of targets over word-positions. Post-test questioning showed that most subjects had not even noticed that targets never occurred in the first sentence.

The third point concerns the interactions of the three monitoring tasks with the word-position effects in the three prose contexts. Identical and Rhyme monitoring behave in almost identical ways (see Figure 3). The slopes of the three prose-context curves in each monitoring task are very similar (Table 4), and so are the relative differences between the intercepts (Table 5). The only difference in the tasks is that the Rhyme monitoring curves are shifted about 140 msec upwards relative to Identical monitoring. This reflects the contribution of the additional attribute-matching stage postulated for the Rhyme monitoring task, and this second stage clearly does not interact with prose context. This was not the case for the semantic attribute-matching stage in Category monitoring, since the Syntactic Prose and Normal Prose curves in this task differ not only in intercept but also in slope. Thus Category monitoring interacts in a different way than the other tasks with the presence or absence of a semantic interpretation, and will, therefore, provide additional information about this aspect of global processing.

The next section of the paper discusses the implications of the word-position effects for the global structure of processing. The questions here center around the earliness with which syntactic and semantic information become

¹³This averaging over many different sentences also accounts for the relatively small size of the mean word-by-word linear facilitation effects. The important point about these word-position effects is not, in any case, their absolute size, but the fact that significant effects are obtained, and that these effects differ across prose-contexts.

available in a sentence, and, in particular, around the possibility of a relative lag in the availability of semantic as opposed to syntactic analyses.

Syntactic and semantic factors in sentence processing

The results show that the semantic dimensions of processing are dominant throughout a sentence, and that they have a significant effect on monitoring responses even at the very beginning of the test-sentence, where the syntactic dimension has only a marginal effect. The basis for this conclusion is, first, the nature and the location of the differences between Normal Prose and Syntactic Prose. In Identical and Rhyme monitoring the only significant difference between the two contexts is in the intercepts of the word-position curves (see Figure 3 and Table 5). At the beginning of the sentence, Normal Prose Rhyme and Identical monitoring responses have an advantage over Syntactic Prose of about 50 msec, which is maintained unchanged over the rest of the sentence. Only Normal Prose can be normally semantically interpreted, and it is presumably this variable which produces the differences in the curves in the two prose contexts.

This means that those semantic aspects of processing that interact with word-recognition decisions are essentially fully developed at the beginning of the sentence, and do not significantly increase over the rest of the sentence in the constraints that they impose on the word-recognition component of Identical and Rhyme monitoring. Semantic constraints operate in word-recognition by enabling word-candidates to be rejected because they map inappropriately onto the meaning representation that is available at that point in the utterance. If such constraints are operating so strongly on the first one or two words of a sentence, then they could hardly be based on a meaning representation derived only from within that sentence. This representation must in large part have been derived from the preceding sentence. Thus the first sentence in each Normal Prose test-pair is playing a crucial role in the immediate interpretation of the second sentence.

The Category monitoring results for Normal and Syntactic Prose reinforce these conclusions. In Normal Prose, semantic information about a word was immediately available for the attribute-matching process, as a consequence of the interaction between the semantic interpretation of the utterance and the process of word-recognition. This meant that semantic attribute-matching in Normal Prose could be carried out as rapidly as phonological attribute-matching. The importance of the close parallels between the slopes and the intercepts in Normal Prose Rhyme and Category is that they show that these semantic interpretation effects, which made it possible for semantic and pho-

nological attribute-matching to be so similar, were operating just as strongly at the beginning of the test-sentence as elsewhere. This inference is clearly also supported by the large intercept difference between Normal Prose and Syntactic Prose Category Monitoring (see Table 5).

The slope differences between Normal and Syntactic Prose in Category monitoring (see Table 4 and Figure 3) serve to underline the importance of the availability of semantic context for performance in this task. Semantic attribute-matching in Syntactic Prose is in fact so severely disrupted that the facilitatory effects of syntactic context on the word-recognition stage of the task are no longer detectable. The separation of the semantic analysis of the word-targets from the process of word-recognition – in contrast to what is assumed to be the case for Normal Prose – evidently produced a cumulative slowing down of the attribute-matching decision across word-positions.¹⁴

In summary, the differences between Normal and Syntactic Prose in all three monitoring tasks show that semantic processing factors are operative immediately the sentence begins, and seem to be as strong there as anywhere else in the sentence. Turning to the syntactic aspects of processing, these appear to develop on a quite different basis.

Syntactic Prose and Random Word-Order show no differences in intercept, and consistently diverge over the rest of the string (in Identical and Rhyme monitoring).¹⁵ To interpret these effects we need to say something about the differences between the two prose contexts. The Random Word-Order materials consist of unordered strings of words, which, because they are simply the words from Syntactic Prose in scrambled order, not only have no syntactic organization, but also cannot be linked up, in an associative manner, to give some overall semantic interpretation (in contrast to the “crash, hospital, ambulance” kind of sequence).

Syntactic Prose materials also have no overall semantic interpretation.¹⁶ It is quite certain that no word-candidate could be rejected in a Syntactic Prose context because it failed to map appropriately onto some “meaning” representation of the sentence at that point. What Syntactic Prose materials

¹⁴This cumulative slowing-down of attribute-matching could be caused by a combination of factors. First, Category monitoring responses in Syntactic Prose average some 150 msec longer than the durations of the words involved. Attribute-matching for one word would still be going on when the next word arrived, so that an increasing backlog could build up over the test-string. Second, the listener has no semantic context available to enable him to focus his attribute-matching efforts selectively; that is, on those word-candidates that are the contextually most likely members of the appropriate category. Thus attribute-matching in Syntactic Prose would not only take longer, but would also have to be applied more thoroughly to more of the words in the test-string.

¹⁵We exclude the Category monitoring task from the discussion of syntactic effects, because of the obscuring effects of the semantic attribute-matching stage of the task.
(continued opposite)

do permit is the assignment of some form of surface syntactic analysis; that is, of the major and minor constituent structure of the string. Syntactic Prose sequences are organized in such a way that lexical items of the appropriate form-class appear in the appropriate slots in the syntactic framework of the string. We assume that it is constraints deriving from this type of structuring that separate Syntactic Prose from Random Word-Order. These constraints develop relatively slowly within the test-strings, and do not show any traces of a carry-over from the lead-in string. The close parallels between the slopes of the Syntactic Prose and the Normal Prose curves (in Rhyme and Identical monitoring) suggest, furthermore, that similar syntactic constraints are accumulating over word-positions in the Normal Prose test-sentences as well.

The word-position effects in this experiment present, then, a clear picture of the availability of different types of analysis during the process of a sentence. This picture is dominated by the processing relationships between sentences in the Normal Prose conditions. Because of the important implications of these between-sentence effects for the basic structure of language processing, we will delay further discussion until these effects have been investigated in more detail. In particular, we need to show directly that the advantage of Normal Prose early in the sentence is indeed due to carry-over effects from the lead-in sentence. This can be done by running the same stimuli on the same tasks, but without the first sentence. If our interpretation of the results has been correct, then removing the lead-in sentence should have the following effects.

First, any change in the pattern of responses should primarily affect Normal Prose. In neither Syntactic Prose nor Random Word-Order could the lead-in material have had any implications for the structural processing of the test-sentence. Second, if the intercept differences between Normal and Syntactic Prose are due to information carried forward from the first sentence, then this difference should disappear when this first sentence is removed. Third, there should be a change in the Normal Prose slope but not in the Syntactic Prose slope, since semantic constraints will be able to develop across the test-sentence in Normal Prose but not in Syntactic Prose. Fourthly, if the carry-over effects in Experiment 1 were primarily semantic effects, then the consequences of removing the lead-in sentence should be most severe for

¹⁶This is quite clear from the example given earlier, and from these additional samples (target-words emphasized):

- (1) It snatched with Mary into it. He's easily your plate if a small *floor* still types to be projected.
- (2) A human light has been stopped out before. Just inside the days they make the *knife* which a fluid had put.
- (3) Janet has been retailing about the levels. The *corn* it posted isn't trying enough pool to kill properly.

Normal Prose Category, because of the greater sensitivity of this task to semantic variables.

Experiment 2

Method

Subjects

36 additional subjects were tested, none of whom had participated in Experiment 1. The subjects were drawn from the same University of Chicago volunteer subject pool, and were paid for their services.

Materials and Design

Direct copies were made of the three stimulus tapes recorded for use in Experiment 1, but omitting the first sentence or word-string in each pair. Thus the test-sentences were acoustically identical to those used in the first experiment. The tapes were combined with sets of instruction booklets to produce exactly the same combinations of task and context conditions as in Experiment 1.

Procedure

The same procedures were followed as in Experiment 1.

Results

Effects of listening task and prose context

The data were analyzed in the same way as before. Separate analyses of variance on subjects and on words, using the untransformed raw data and with missing observations replaced, were combined to give $\text{Min F}'$ ratios. All values are significant at the 0.05 level unless otherwise noted.

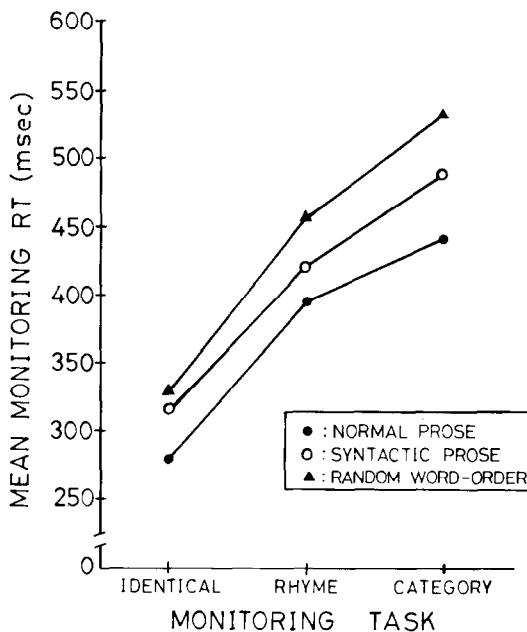
The overall mean monitoring latencies for the nine combinations of Prose Context and Monitoring Task are given in Table 6 (see also Figure 4). As in Experiment 1, there was a strong main effect of Monitoring Task ($\text{Min F}'$ (2,172) = 199.556). This reflects the large overall differences between Identical Monitoring (308 msec), Rhyme Monitoring (423 msec), and Category Monitoring (486 msec). There was also a significant main effect of Prose Context ($\text{Min F}'$ (2,196) = 29.456), which was somewhat smaller than in

Table 6. *Experiment 2. Mean Monitoring reaction-times (msec): By Prose Context and Monitoring Task.*

Prose Context	Monitoring Task		
	Identical	Rhyme	Category
Normal Prose	279	394	442
Syntactic Prose	315	420	487
Random Word-Order	329	456	531

Each value is the mean of 324 observations. Overall standard error = 8.39 msec.

Figure 4. *Experiment 2: mean monitoring latencies, by Prose Context and Monitoring Task.*



Experiment 1. Targets in Normal Prose, with a mean reaction-time of 372 msec, were responded to significantly faster than targets in Syntactic Prose (407 msec) and in Random Word-Order (439 msec). The difference between Syntactic Prose and Random Word-Order is also significant. Note that while the overall Normal Prose latency is unchanged relative to the Experiment 1 mean of 373 msec, the overall Syntactic Prose and Random Word-Order means are significantly faster, by 34 and 38 msec respectively.

Unlike Experiment 1, there was no interaction between Prose Context and Monitoring Task ($\text{Min } F'$ (4,385) = 1.735, $p > 0.10$). As can be seen from Figure 4, the behavior of the three monitoring tasks is now much more similar in each context. The main reason for this divergence from Experiment 1 is the change in Category Monitoring performance. Normal Prose Category is now significantly slower than Normal Prose Rhyme (by 48 msec). Thus the difference between Rhyme and Category in Normal Prose is now of the same order of magnitude as the difference in Syntactic Prose (67 msec) and in Random Word-Order (75 msec).

Effects of word-position

The overall effects of Word-Position are similar to those in Experiment 1. Neither the Word-Position main effect, nor the interactions with Prose or Task reached significance. There were, however, significant linear components to the Word-Position main effect ($\text{Min } F'$ (1,88) = 5.660), and to the Word-Position by Task ($\text{Min } F'$ (2,395) = 4.564) and the Word-Position by Prose ($\text{Min } F'$ (2,265) = 5.076) interactions. The linear effect in the interaction with Task was due to the presence of strong linear effects in Identical and Rhyme monitoring, but none in Category monitoring. There was also a tendency here towards a quadratic component in the interaction, with quadratic effects in Identical and Rhyme monitoring ($p < 0.10$) but none in Category monitoring. In Experiment 1 there had been no trace of a quadratic effect in any of the analyses.

The significant linear component to the Word-Position by Prose interaction was due, as in Experiment 1, to the strong linear effects in Normal and Syntactic Prose but not in Random Word-Order. Linear regression analyses of the overall word-position curves in each prose context showed similar effects to Experiment 1,¹⁷ although with small increases in slope in Syntactic and Normal Prose. The overall curves, do, however, differ in one important way from the earlier results – the intercepts of the three prose context curves are now much closer together, ranging from 427 msec in Normal Prose to 435 msec in Random Word-Order to 447 msec in Syntactic Prose. None of these values differ significantly from each other. Overall, the mean difference in intercept between Normal Prose and the other two contexts is reduced to 14 msec, compared with 67 msec in Experiment 1 (see Table 5).

¹⁷In Normal Prose ($r = -0.88$, $b = -10.82$), 77% of the variance is accounted for by the linear effects of word-position. In Syntactic Prose ($r = -0.80$, $b = -7.68$), 64% of the raw means variance is accounted for. There is no trace, again, of an overall linear effect in Random Word-Order ($r = +0.03$, $b = +0.23$).

The similarities in slope between Experiment 1 and 2 break down when the individual results for each of the nine prose by task conditions are analyzed. Table 7 gives the correlation coefficients and slopes derived from linear regression analyses on both the raw and the combined smoothed and raw word-position means. Figure 5 plots the raw word-position means together with the regression-lines derived from the combined analyses.

Table 7. *Experiment 2. Monitoring Reaction-time and Word-Position: Regressions on Raw and Smoothed Means (By Prose Context and Monitoring Task)^a*

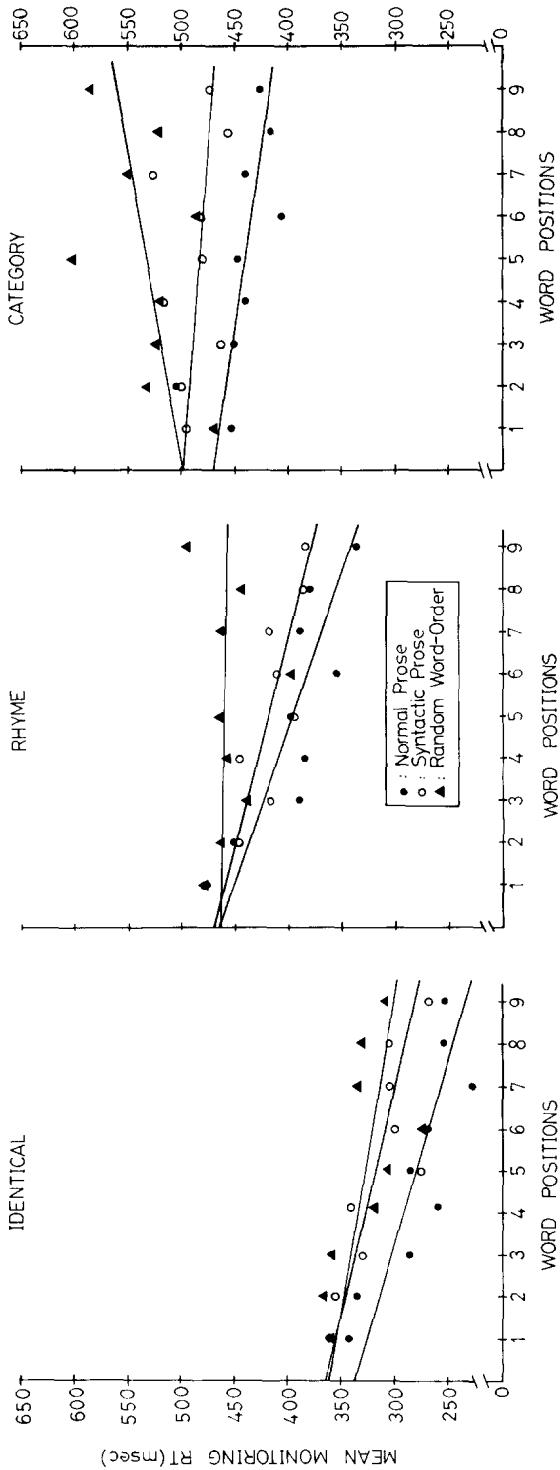
Prose Context	Monitoring Task		
	Identical	Rhyme	Category
Normal Prose			
r =	-0.84 (-0.91)	-0.86 (-0.90)	-0.69 (-0.83)
b =	-11.66 (-11.52)	-13.93 (-13.68)	-7.05 (-5.63)
Syntactic Prose			
r =	-0.84 (-0.92)	-0.86 (-0.95)	-0.27 (-0.60)
b =	-10.08 (-9.63)	-10.10 (-10.12)	-2.43 (-3.20)
Random Word-Order			
r =	-0.61 (-0.76)	-0.00 (-0.08)	+0.47 (+0.67)
b =	-6.73 (-6.75)	-0.03 (-0.55)	+7.39 (+6.62)

r = correlation coefficient, b = slope in msec (n = 9)

^aResults of regressions on the combined raw and smoothed means are given, in parentheses, following the raw means results.

Consider, first, the Normal Prose results. The overall linear effect for this prose context conceals an interaction with task. Relative to Experiment 1 (see Table 4), the Normal Prose slopes in Experiment 2 are considerably steeper in both Identical and Rhyme monitoring (basing these comparisons on the more robust combined estimates of slope). In Identical monitoring, the Normal Prose slope increases significantly from -6.88 msec per word-position to -11.52 msec ($t(14) = 1.889$), and in Rhyme monitoring there is a significant increase from -7.92 msec to -13.68 msec ($t(14) = 2.004$). But in Category monitoring the slope is shallower than in Experiment 1, decreasing from -8.00 msec to -5.63 msec ($t(14) = 1.190$, $p < 0.20$). This Normal Prose Category slope is significantly shallower than the Normal Prose Identical slope ($t(14) = 2.392$) and the Normal Prose Rhyme ($t(14) = 2.825$) slope. In Experiment 1, the Normal Prose slopes for all three conditions had been the same.

Figure 5. Experiment 2: raw Word-Position means for each combination of Prose Contexts and Monitoring Tasks. Best-fitting least-squares of regression-lines, based on combined raw and smoothed means, are plotted. Note: Word-position 1 corresponds to the second word in the test sentence.



There is also one other difference in the Normal Prose curves in the second experiment. This was the appearance, in the trend analyses, of a quadratic component in the Normal Prose Rhyme and Identical curves. This quadratic effect reflects the rather rapid fall off in reaction-time over the early word-positions which can be seen in Figure 5.

In Syntactic Prose there are no significant changes from Experiment 1 to 2. In Identical and Rhyme monitoring there are small, non-significant increases in slope — for the two conditions together, the slope in Experiment 2 is 1.8 msec steeper than in Experiment 1. In Category monitoring there is again no significant linear effect of word-position, and the slope does not differ from zero.

The Random Word-Order conditions show wide variations in the strength and direction of the linear effects. In Identical monitoring, there is a significant negative slope of -6.75, which is almost reliably different from the slope of -1.72 estimated for Experiment 1 ($t(14) = 1.600, p < 0.10$). In Random Word-Order Rhyme there is again no significant linear effect, and the means are best fitted by a quartic curve ($p < 0.10$). In Random Word-Order Category, however, there is a relatively strong positive linear effect, increasing from +0.82 msec in Experiment 1 to +6.62 msec in Experiment 2 ($t(14) = 1.448, p < 0.10$). This curve also contained a significant cubic component ($p < 0.05$).

Table 8. *Experiment 1. Monitoring Reaction-time and Word-Position: Combined Raw and Smoothed Means Estimates of Intercept (msec)*

Prose Context	Monitoring Task			
	Identical	Rhyme	Category	Overall
Normal Prose	339	466	469	427
Syntactic Prose	365	470	500	447
Random Word-Order	362	463	499	435

Finally, as is clear from the comparison between Figures 5 and 3, the estimated intercepts for the regression lines in each task are much closer together than in Experiment 1 (see Table 8). Compared to the first experiment (Table 5), the mean difference in intercept between Normal Prose and the other two prose contexts drops from 64 to 25 msec in Identical monitoring, from 43 to one msec in Rhyme monitoring, and from 87 to 31 msec in Category monitoring. Only in Category monitoring is the Normal Prose intercept still

significantly different from the intercepts for the other two prose contexts ($p < 0.05$).¹⁸

Discussion

Before moving on to the main results of Experiment 2 (the effects on the Normal Prose and Syntactic Prose slopes and intercepts), two other aspects of the results need to be discussed. The first of these is the appearance of sizeable linear word-position effects in Random Word-Order Identical and Category monitoring (see Table 7 and Figure 5). The effect in Category monitoring seems of little importance; the linear fit is rather poor, and the data is better described by a cubic curve. If there is a genuine increase over word-positions here, it presumably reflects cumulative decrements in attribute-matching of the type previously discussed in connection with the Syntactic Prose Category monitoring curves.

The linear decrease over word-positions in Random Word-Order Identical is not only rather stronger (though accompanied by a marginally significant quartic effect), but also raises the possibility that non-linguistic serial-position effects are affecting performance in Experiment 2. In Experiment 1 there had been no evidence for the presence of this type of response-bias. But the test-sequences in Experiment 2 were much shorter, because of the absence of the lead-in sentence, and this may have made it easier for the subjects to build up expectations about where the targets were occurring in the Random Word-Order sequences.

It is unlikely, however, that this kind of non-linguistic predictability effect contributed to performance in the Syntactic Prose and Normal Prose conditions. In Syntactic Prose, in fact, there are no significant increases in the slopes of the word-position effects, comparing Experiments 1 and 2. In Normal Prose, the increases in slope in Experiment 2 are fully accountable for on the basis of the linguistic structure of the test-sentences, and are not distributed in the way that one would expect if a serial-position response-bias was responsible. If the change in slope had been due to the increasing likelihood of the target occurring later in the string, then the facilitatory effects

¹⁸ As a check on the results in Experiment 1, the relationships between word-length and reaction-time in each prose context were also calculated for the Experiment 2 data. The effects were similar to those obtained previously, with the weakest effects in Normal Prose, ($r = +0.40$), stronger effects in Syntactic Prose ($r = +0.74$), and the strongest effects in Random Word-Order ($r = +0.88$). The only difference was an increase in the Syntactic Prose slope ($b = +0.43$), such that it was no longer significantly different from the Random Word-Order slope ($b = +0.49$), while it did differ significantly from the Normal Prose slope ($b = +0.17$).

should be weakest early in the sentence, and become progressively stronger later in the sentence. The observed Normal Prose curves show the opposite pattern. The steepest acceleration is over the first two or three word-positions, and the curve is clearly flattening out later in the sentence (see Figure 5). This is exactly the pattern to be expected if there is a rapid accumulation of semantic and syntactic constraints early in an isolated test-sentence, and provides no evidence that the subjects were incorporating non-linguistic constraints as well.

A second unexpected result involved the expectation that the overall Normal Prose mean should be slower in Experiment 2 than in Experiment 1, and therefore less different from the Syntactic Prose and Random Word-Order means, which should not be affected by the absence of the lead-in sentence. This prediction was confirmed in the sense that there was indeed a smaller overall difference between Normal Prose and the other two contexts. But this was the result not so much of a slowing down of the Normal Prose means as of a speeding up of the Syntactic Prose and Random Word-Order means. The most likely explanation for this is that the group of subjects tested for Experiment 2 were simply faster as a group than the subjects in Experiment 1. If the overall prose context means in Experiment 2 are adjusted by adding to them the difference (24 msec) between the grand means in Experiments 1 and 2, then we obtain the predicted results. That is, the Syntactic Prose and Random Word-Order means are not significantly different in the two experiments, while the Normal Prose mean is significantly slower in Experiment 2.

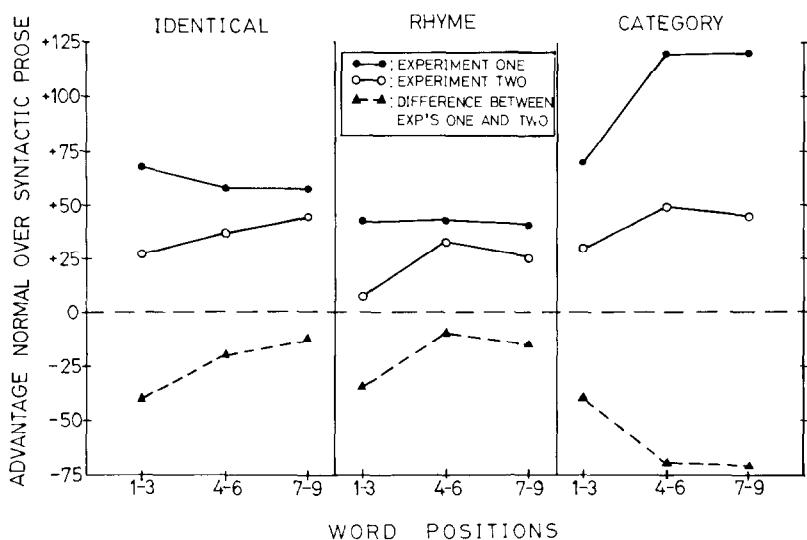
Turning to the main results of Experiment 2, their outcome was entirely consistent with the claim that the advantage of Normal over Syntactic Prose early in the test-sentence in Experiment 1 was due to semantic carry-over effects from the first, lead-in sentence. This claim is supported and enlarged in several ways by the results.

First, there are the changes in intercept and slope in Identical and Rhyme monitoring. In both tasks the Normal Prose and Syntactic Prose intercepts are no longer significantly different. Thus, when the lead-in sentence is removed, responses in Normal and Syntactic Prose start off the sentence with essentially the same degree of facilitation. In addition, there is a significant increase in the steepness of the slopes in Normal Prose but no significant change in Syntactic Prose, in both monitoring tasks. This is consistent with the hypothesis that in Normal Prose, in Experiment 2, both syntactic and semantic constraints are now developing in strength across the sentence, whereas in Syntactic Prose only syntactic constraints can develop, and the rate at which this occurs is unaffected by the presence or absence of a lead-in sentence.

The specifically semantic nature of the effects of removing the lead-in sentence is demonstrated by the second major aspect of the results – the differential effects on Category monitoring performance. The Category monitoring task, as we pointed out earlier, seems especially sensitive to the semantic aspects of processing, and the removal of the interpretative context provided by the lead-in sentence has distinctive effects on performance in this task. The nature of these effects can be seen in Figure 6, which summarizes the differences between Normal and Syntactic Prose over word-positions as a function of monitoring task and of the presence or absence of a lead-in sentence. The upper two curves in each panel show the advantage of Normal Prose over Syntactic Prose at different points in the test-sentences in Experiment 1 and 2. The lowest curve in each panel shows the decrement in the advantage of Normal over Syntactic Prose when the lead-in sentence is removed.

The pattern of the results shown in the lowest curve for each monitoring task is quite different in Category monitoring than in the other two tasks. In

Figure 6. *Experiments 1 and 2: mean advantage (msec) of Normal Prose over Syntactic Prose at early (1-3), middle (4-6), and late (7-9) Word-Positions. The upper panels give the increase for Normal Prose over Syntactic Prose in each experiment separately; the lower panel gives the decrease in the Normal Prose advantage in Experiment 2 as compared with Experiment 1.*



Rhyme and Identical monitoring there is a mean decrement in the facilitation for Normal Prose of about 40 msec early in the test-sentence. This decrement stabilizes later in the sentence at a residual level of about 10–15 msec. But in Category monitoring the opposite occurs. The decrement for Normal Prose increases over word-positions, from about 40 msec early in the sentence to 70–75 msec later in the sentence. This increasing decrement for Normal Prose Category can also be observed in a comparison with Normal Prose Rhyme. Normal Prose Category monitoring is significantly slower than Rhyme monitoring in Experiment 2, but not early in the sentence. The intercepts of the two response curves are identical (see Table 8 and Figure 5), but then diverge over the rest of the sentence by about eight msec per word-position (see Table 7).

These relative decrements in Normal Prose Category monitoring presumably derive from the same source as the increasing decrement over word-positions in Syntactic Prose Category monitoring in Experiment 1 (see footnote¹⁴). That is, the semantic attribute-matching process has become more independent of the word-recognition process and this brings into play a number of factors that cumulatively slow down the attribute analysis process, thereby counteracting the increasing facilitation of responses over word-positions at the word-recognition level. What now has to be determined is how the removal of the lead-in sentence can have these lasting effects on semantic attribute analysis in Normal Prose Category monitoring.

To understand the role of the lead-in sentence in Normal Prose, it is necessary to consider the ways in which the interpretation of an utterance would be changed by locating it in an appropriate discourse context.¹⁹ Take, for example, the test-sentence given earlier:

(4) Some thieves stole most of the lead off the roof.

The possibilities for the on-line interpretation of this sentence are clearly going to be considerably enriched when it is preceded by its lead-in sentence ("The church was broken into last night"). The event described in the test-sentence becomes located in a time and a place, and the processing of each item in the sentence becomes significantly different. The initial noun phrase "Some thieves" would now be processed in an interpretative context which could affirm, early in the recognition process, the semantic appropriateness of the word-candidate "thieves". Later items in the sentence – such as "lead" or "roof" – would also occur in a richer interpretative framework when the

¹⁹The following discussion should not be taken as an attempt at a theory of discourse interpretation. We are simply indicating the general kind of effect that might be interacting with word-recognition processes.

lead-in sentence was present. The listener would, for example, have already encountered a building to which "the roof" could belong.

These effects on the interpretation of a test-sentence would be even stronger in cases where the test-sentence contained an anaphoric pronoun referring back to the previous sentence. Thus the test-sentence "It's only about a *mile* from the university" can never be properly interpreted without knowing that the referent of "It" is "a new apartment" (mentioned in the lead-in sentence). This kind of interpretative framework, simulated here by the use of pairs of sentences, must be a fundamental aspect of the normal analysis of an utterance.

In the case of Category monitoring, the diminished interpretability of the test-sentences in Experiment 2 has clearly interacted with the semantic-attribute matching process. Optimal performance in this task depends on the early semantic analysis of the potential target words, in which the relevant semantic attributes of the words are selected out as part of the process of mapping the attributes of the word being recognized onto the developing interpretative representation of the utterance. When this interpretation is less determinate, as it will be when the lead-in sentence is missing, then this will place a greater load on an additional attribute analysis and matching process, thereby leading to a cumulative decrement in Category monitoring performance.

Taken together, then, the contrasts between Normal and Syntactic Prose in the three monitoring tasks not only confirm the importance of the lead-in sentence in the early processing of the Normal Prose test-sentences, but also bring into sharper focus the nature and the time-course of these carry-over effects. In Rhyme and Identical monitoring, which reflect semantic facilitation of the speed of word-recognition decisions, there is a large decrement in the facilitation of Normal Prose responses over the first two or three words of the sentence (see Figure 6). This decrement reflects the absence of the interpretative framework, provided by the lead-in sentence in Experiment 1, in terms of which the first words in the sentence would naturally be interpreted. Within-sentence constraints, however, build up rapidly in Normal Prose in Experiment 2, so that the word-recognition decrement has been more than halved by the time four or five words have been heard. This rapid build-up in semantic facilitation is reflected in the quadratic component observed in the Normal Prose Rhyme and Identical word-position curves.

Nonetheless, the interpretative framework built up within the sentence in Experiment 2 never quite reaches the level of facilitation observed in Experiment 1. Even over the later word-positions, the facilitation of Normal Prose Rhyme and Identical latencies falls short of the levels achieved when the lead-in sentence was present. And this lasting impairment is clearly seen in the Category monitoring decrement.

In summary, then, Experiment 2 supports and reinforces the conclusions we drew from the pattern of word-position effects in Experiment 1. The discourse context in which a sentence occurs provides an immediate framework in terms of which even the first words of the new sentence are processed and interpreted. We can now return, then, to the implications of the word-position results for the differing claims about the ordering of global processing events.

The temporal ordering of global processes

The question here is whether there is any evidence to support the claims of a serial processing theory for the ordering of communication between global processing components. If no such evidence is found, then we are free to assume that the contributions of different knowledge sources to the analysis of an input are not *a priori* restricted to the kind of serial sequencing required by the autonomy assumption.

If the word-position results are to be interpreted within the framework of the claims made by the autonomy assumption, then we will have to do so in terms of three separate higher-level processing components. Apart from a syntactic processor, we will also have to distinguish between a semantic processor and what can be called an "interpretative" processor. The semantic processor is restricted to a linguistic semantic analysis, based on the meanings of the words in the string and on the syntactic relationships between them (e.g., Fodor *et al.*, 1974). The interpretative processor provides a further interpretation of this semantic analysis in terms of the listener's knowledge of the world and of the discourse context in which the utterance is occurring. The availability of all three types of analysis should be delayed in time relative to each other, as well as being delayed relative to the output from the word-recognition component.

We begin with the temporal properties of syntactic processing, since the syntactic processor sits first in the sequence of higher-level processing components. The syntactic contribution over word-positions is estimated from the comparison between Syntactic Prose and Random Word-Order. There are no intercept differences between these two prose conditions in any of the four relevant comparisons (in Rhyme and Identical monitoring in both experiments)²⁰, but there is, overall, a significant divergence of Syntactic Prose from Random Word-Order over word-positions. This is clear evidence for the existence of a non-semantic form of structural analysis, which begins early in

²⁰See footnote 15.

the sentence and builds up over subsequent word-positions. The further results show that whatever role this type of analysis is playing in on-line processing, it is not ordered in the ways dictated by the serial autonomy assumption.

First, there is the finding of a greater intercept difference between Normal and Syntactic Prose in Experiment 1 than in Experiment 2. These intercept advantages in Experiment 1 can only be explained on the basis of an immediate involvement in processing decisions of information carried over from the lead-in sentence in Normal-Prose. This information must be represented at a level of analysis that would correspond to the terminal stage of processing in a serial processing theory. That is, at the point in the analysis of the input at which the output of the semantic processor is integrated with a non-linguistic knowledge base (and with discourse context) by the interpretative processor.

The problem that these results present for the autonomy assumption is the following. If discourse context can interact with the processing of the first words of a new sentence, then this means that these words are already being assessed at a level in the system that is compatible with the level of representation at which the discourse context presumably resides. But within the framework of a serial processing model, the achievement of this interpretative level of analysis of the current input has to be delayed relative to the preliminary processing of the input at strictly linguistic, sentence-internal, levels of structural analysis.

But there is no sign in the data that discourse context is having less effect early in the sentence than elsewhere – which is what a relative delay in reaching the appropriate level of analysis would require. In fact, if we compare the advantage of Normal Prose over Syntactic Prose early in the sentence in Experiment 1 with the early Normal Prose advantage in Experiment 2, then it is clear that discourse context is having its relatively strongest effects at the earliest points at which it is possible to measure processing performance. That is, at the first one or two word-positions, where the developing sentence-internal effects (as observed in the isolated Normal Prose test-sentences in Experiment 2), and, in particular, the developing syntactic effects, have relatively weak consequences for monitoring latencies.²¹

²¹ The present results are not the only evidence for the immediate involvement of discourse context in on-line processing. Recent research by Harris (Note 3), also using the word-monitoring task, shows between-sentence facilitation at the earliest word-positions. His results, in addition, relate the facilitation effects directly to the discourse properties of the text – whether or not the test sentences carried “Given” or “New” information relative to the context-sentence. There are also several experiments by Cole and Jakimik (1979) which show discourse-based effects on word-recognition, using the mispronunciation detection task.

The between-sentence effects in the present experiment demonstrate, then, that the highest level of processing (in a serial sequence of separate processing levels) is in operation at the beginning of a sentence, and certainly as early as the operations of a syntactic processing component. Similar arguments can be made for the immediate engagement of a sentence-internal semantic level of processing. This is the level at which the output of the syntactic processor is used to build a linguistic semantic representation of the word-string. If there is an independent processor at this semantic level, then the Category monitoring results show that this is involved in the analysis of the input as soon as the sentence begins.

The differences between Normal and Syntactic Prose Category monitoring reflect the active construction of a semantic interpretation of the sentence at the point at which the word-monitoring target occurs. We find significant intercept differences between Normal and Syntactic Prose Category monitoring not only in Experiment 1, where discourse context might have been the source of the difference, but also in Experiment 2 (see Table 8), where there is no discourse context. Processing has to start from scratch, as it were, when an isolated test-sentence is heard, and the Normal Prose intercept advantage in Category monitoring shows that this very early processing involves the semantic analysis and assessment of the words being heard. And within the serial framework, as currently defined, this function is assigned to the semantic processor. Note that there is no intercept difference in Experiment 2 between Normal and Syntactic Prose in Rhyme monitoring. This shows that the intercept difference in Category monitoring is specifically due to an interaction with the semantic attribute-analysis of the input.²²

We have, in summary, identified two types of processing information that would derive, in a serial theory, from processing components that would have to be located later in the analysis sequence than a syntactic processor. These are what we have labelled the semantic and interpretative levels of analysis. The results demonstrate that both types of analysis are actively engaged in processing the input from the first word of a sentence onwards, and that there is no sign of any delay relative to the syntactic analysis of the input. Thus there is no evidence that the global structure of sentence processing is ordered in time in the ways which the autonomy assumption requires.

²² An experiment reported elsewhere (Marslen-Wilson *et al.*, 1978) provides evidence for the immediate involvement of a semantic level of processing in a more precisely controlled syntactic context. We contrasted rhyme and category monitoring latencies to word-targets that were located either immediately after or immediately before a mid-sentence clause-boundary. The results confirm that the input is being just as actively semantically interpreted at the first word of a syntactic processing unit (as classically defined) as anywhere else.

General Discussion

The purpose of the research reported here was to determine the validity of certain assumptions about the properties of information flow between different aspects of the language processing system. The results have enabled us to reject the kinds of constraints imposed by the autonomy assumption. In this last part of the paper we will consider what kind of structure we can assign to a more interactively organized processing system. We will be guided in this by the word-recognition theory that the results prompted us to develop, and by the clear evidence for the involvement of discourse variables in the immediate processing of the speech input.

As we noted in the introduction, the description of a language processing system also depends on the distinctions one makes between the types of knowledge that are involved in processing. It is difficult to talk about the structure of a processing system without coming to some decision about what its principal components might be.

The issue of distinguishable knowledge types — and the corollary issue of computationally distinct levels of representation during processing — has to be treated quite differently in an on-line interactive context than in the kind of serial theory that we have rejected. All of these serial theories were clearly rooted in the transformational generative linguistics of the 1960's. Thus the distinctions drawn in the linguistic theory between analytic levels in the formal description were carried over directly into the psychological model — providing the basis not only for the set of processing components selected, but also for the processing units assigned (or not) to these components (cf., Fodor *et al.*, 1974; Levelt, 1974; Marslen-Wilson, 1976; Miller, 1962; Miller and Chomsky, 1963; Tyler, 1980).²³ The earlier formulations of an on-line interactive processing approach — under the name of the interactive parallel theory (Marslen-Wilson, Note 1; 1975; Marslen-Wilson and Tyler, 1975) — were clearly also heavily influenced by linguistic preconceptions. In particular, we took for granted the standard distinctions among types of mentally represented linguistic knowledge, and among a corresponding set of processing components (cf., Tyler, 1980).

It is difficult, however, to maintain this kind of close linkage between the linguistic theory and the process theory in the face of the on-line processing data reported here and elsewhere. Above all, the intelligibility of this linkage depends on the correctness of the autonomy assumption, since it is this that

²³ It should be clear that what we are talking about here is the psycholinguists' interpretation of linguistic theories; we are not claiming that linguists themselves have explicitly made this kind of connection with "performance" questions.

forces, *a priori*, an organization of the processing model that parallels the organization of the linguistic theory. The serial, stage-by-stage, processing system that the autonomy assumption requires can be seen as a sort of process analogy to the hierarchy of analytic levels postulated in early transformational grammars. But this analogy becomes untenable if the processing system, as we have argued here, is organized on an interactive basis.

Our position, then, is that the processing data do not provide any empirical support for the kinds of linkages often proposed between the linguistic theory and a processing model. This linkage can best be seen as a hypothesis about language processing which has turned out to be incorrect. The consequence of this is that we are not justified in deriving any assumptions from a linguistic theory without first examining them in the light of the available process data. Thus in making suggestions about the knowledge types that need to be distinguished in a psychological process model, we will be guided in the first instance only by psychological process data.

This seems, in general, to be the correct strategy to follow in laying out the outline of a process theory. The other disciplines concerned with human language — linguistics, some branches of artificial intelligence — cannot be expected to provide the basic framework for such a theory. These disciplines have neither had as their goal the explanation of psychological processes nor have they been responsive in any meaningful way to process data about language use. Once the framework of a genuinely psychological process theory has been constructed, then — but only then — will one be in position to interpret, from this perspective, the relevant research in other disciplines.

In the light of these remarks, what minimal distinctions between knowledge types do the on-line data require us to make? Note that this question is not the same as the question of computationally distinct levels of representation during processing. The autonomy assumption required that each distinction between mentally represented knowledge types was realized in processing by its own autonomous processing component. But if we allow a more interactive system, then it becomes conceptually and computationally possible to allow knowledge sources to contribute to processing without each source necessarily corresponding to a distinct level of analysis in the processing sequence. We will return later to this question of processing levels.

The available on-line data do not seem to require more than three basic distinctions among knowledge types (leaving aside acoustic-phonetic issues²⁴).

²⁴ Note that the cohort theory, as presented here, makes no specific assumptions about the acoustic-phonetic input to the word-recognition system. The input could be anything from a set of spectral parameters to a string of segmental labels. Thus our use of phonemic descriptions as the basis for determining cohort membership should not be construed as a theoretical claim about the products of acoustic-phonetic analysis.

First, there is clearly a sense in which "words" have a psychological reality, and, in fact, our whole experimental enterprise here has depended on this. We group together here not only words as phonological objects, but also as clusters of syntactic and semantic attributes. It is clear from the kinds of interactions that take place during word-recognition that all these aspects of the internal representation of words must be very closely linked.

Secondly, the present data provide good evidence for the psychological reality in processing of a form of "non-semantic" structural analysis. This conclusion is based, first, on the differences between Random Word-Order and Syntactic Prose. These differences, as we argued earlier, reflect the influence on processing of a form of structural analysis which is not semantic in nature. Second, the similarities between the Normal and Syntactic Prose slopes suggest that whatever is producing the word-position effects in Syntactic Prose is also responsible for the parallel effects in Normal Prose. In Experiment 1, where semantic influences seem to be held constant across the entire Normal Prose test-sentence, the Normal Prose and Syntactic Prose slopes are remarkably similar. The slight divergence in the slopes for the two prose types in Experiment 2 is fully accountable for on the basis of the semantic influences developing within the sentence in Normal Prose.

So far, the distinctions we are making are reasonably uncontroversial, though motivated on different grounds than usual. Contrary to standard assumptions, however, we are not convinced that it is necessary to assume two further functionally distinct knowledge sources. That is, there seems to be nothing in the on-line data to compel a distinction between "semantic" and "interpretative" knowledge sources. The deployment of word-meanings in the interpretation of an utterance may not involve a distinct type of knowledge source – that is, one which is defined within a domain of semantic interpretation based strictly on word-meanings and syntactic relations, and which is separate both from the knowledge source involved in the interpretation of word-strings in terms of a possible world or discourse context, and from the structural syntactic source previously suggested.

In particular, the results here show that words are immediately analyzed with respect to their discourse context. To account for the effects on word-recognition (and on semantic attribute-analysis in Category monitoring), we have to assume that discourse variables are involved in the primary decisions about word-identity. This means that the original selection from among the pool of word-candidates (and, therefore, from among the various senses of individual candidates) is conducted within an interpretative contextual domain. Thus it may be possible to accomplish the semantic analysis of words in utterances through operations of knowledge sources none of which has the properties of the classical semantic interpreter. That is, given the repre-

sentation of word meanings in the lexical knowledge source, the further interaction of these with a structural syntactic source and with an interpretative source appears to be sufficient.

This minimal segmentation of knowledge sources into three types (which we will label *lexical*, *structural*, and *interpretative*) is by no means intended to be the last word on the matter. But it seems to cover the on-line data, and it gives us a basic vocabulary in terms of which we can discuss the processing principles which govern the organization of spoken language understanding.

Processing principles for an on-line interactive theory

The best developed aspect of an on-line interactive treatment of spoken language processing is in the domain of spoken word recognition. This is because the on-line data in this domain bears particularly directly on the internal structure of the processing of a well-defined type of perceptual object. Thus, taking as our model the properties of spoken word-recognition, what do these properties indicate about the general organizational strategies of the processing system?

First, the word-recognition system illustrates two general strategies underlying the organization of language processing: namely, that it allows for interactions between knowledge sources, and that it is designed for optimal efficiency and speed. In the cohort model we can see that it is possible for the knowledge sources involved in processing to cooperate together with respect to a processing target defined within a single knowledge source. From the perspective of the problem of word-recognition, this processing target was the identification of the correct word-candidate. This model for interaction can be extended to cover the case where the ultimate processing target is the interpretation of an utterance.

This possibility for interaction is part of the reason why the system can be optimally efficient in its operations. In the case of spoken word recognition, no more of the sensory input needs to be heard than is necessary to uniquely distinguish the correct word-candidate from among all the other words in the language, beginning with the same sound sequence, that could occur in that particular context. The kind of processing organization postulated in the cohort model allows spoken words to be recognized as soon as they possibly can be — unless one resorts to a potentially errorful guessing strategy. Again, we feel that these claims about the properties of word-recognition can plausibly be extended to cover all aspects of the system's organization.

These two properties, of interactiveness and optimal efficiency, are closely related to the claim we made earlier in the paper that the structural and interpretative implications of a word-choice should be able to propagate through

the processing system as rapidly as possible. If a processing system is to be optimally efficient, then it should allow all sources of information to be mutually available, so that whatever is known at any particular point in processing can be available to resolve indeterminacies in the analysis of the input with respect to any single knowledge source. In the case of spoken word recognition, this means that differences in the structural and interpretative suitability of different word-candidates can be used to optimize the speed with which the correct candidate is selected, and, therefore, the efficiency with which the system is using the available sensory information.

So far, what we have described are evidently desirable properties for a processing system to have. But they are not enough, in themselves, to properly specify the structure of information flow through the system. We can, however, identify two further types of processing principle which are capable of placing strict constraints on the operations of the system. We will refer to these as the principles of "bottom-up priority" and of "obligatory operation".

In the cohort theory, perceptual processing is driven, in the first instance, from the bottom-up. This is not only a matter of greater efficiency, but also of greater security. By defining the set of perceptual alternatives on the basis of bottom-up information alone, a processing system not only restricts the set of alternatives at least as much as top-down pre-selection normally would, but it also ensures (within limits) that the correct word-candidate will be contained within this initial set. Thus the processing system first determines (on a bottom-up basis) what *could* be there, as the proper foundation for then going on to determine what *is* there.

The system allows for top-down effects in the loose sense that contextual information affects the recognition process. But it does not allow for top-down effects in the more precise sense of the term — that is, by allowing contextual factors to pre-select some class of likely words even before any of the relevant sensory input has been received. We do not have any direct evidence for this exclusion of true top-down effects, but the system we have proposed is optimally efficient without needing to invoke such a mechanism. In addition, a top-down pre-selection strategy seems to be inherently both more dangerous than a bottom-up strategy, and in general less effective.

Successive words in an utterance are only rarely fully predictable from their prior context. This means that a top-down strategy would most often be selecting such large classes of possible words as to be of very little use in discriminating between possible candidates. Whereas a bottom-up strategy, as we noted earlier, yields on average no more than about 30 candidates by the time the first two phonemes are known. Of course, when words are more predictable, then top-down selection can be as constraining as bottom-up se-

lection. But here the top-down strategy runs into the danger of failing to include in the pre-selected set the word that actually occurs. Language users are, ultimately, unpredictable, and for any word that a speaker chooses, he could almost always have chosen to use some other word. This is a problem for top-down selection, but clearly not for a bottom-up strategy.²⁵

A further reason for choosing a bottom-up selection process is that it seems more compatible with the second processing principle that we are adopting. This second principle is based on the claim that the operations of the word-recognition system have an "obligatory" or "automatic" property (cf., Forster, 1979; Marslen-Wilson and Welsh, 1978; Shiffrin and Schneider, 1977). If an acoustic-phonetic input can be lexically interpreted, then it apparently must be. Apart from one's own phenomenological experience, the evidence for this comes from several studies which show that even when subjects are asked to focus their attention on the acoustic-phonetic properties of the input, they do not seem to be able to avoid identifying the words involved. This can be seen in the Rhyme monitoring results in the present experiment, in phoneme-monitoring experiments (e.g., Marslen-Wilson, Note 2; Morton and Long, 1976), and in mispronunciation detection tasks (Cole and Jakimik, 1979; Marslen-Wilson and Welsh, 1978). This implies that the kinds of processing operations observable in spoken word recognition are mediated by automatic processes, and are obligatorily applied to any acoustic-phonetic input. We will assume here that the same obligatory processing principle applies to all of the knowledge sources we have identified.

Given these four processing properties – interactiveness, optimal efficiency, bottom-up priority, and obligatoriness – we can now show how these four properties determine the structure of the proposed processing model.

Processing structure for an on-line interactive theory

The two principles that best determine the organization of language processing are those of bottom-up priority and of obligatoriness. In the context of word-recognition, the principle of bottom-up priority describes the reliance of the system on the bottom-up definition of a set of word-candidates. Analogously, the structural and interpretative knowledge sources depend for their operation on a bottom-up input. Structural decisions cannot be made except in the context of the words already recognized and of the set of word-candidates currently being considered. Given this input, a very restricted set of structural assignments will be possible. And what this bottom-up priority

²⁵ These arguments against a top-down selection strategy apply better to "content" words than they do to some "function" words. The latter kinds of words are not only more strongly determined by their structural context (consider, for example, the distribution of words like "the" and "a"), but are also often poorly represented in the speech signal. One can clearly make a stronger case here for the use of top-down constraints (as Ron Cole (personal communication) has pointed out to us).

means here is that the structural knowledge source can only assign an analysis to the string which is compatible with the bottom-up input (the successive words in the utterance). Top-down influences are assumed not to pre-determine possible structural assignments, any more than they are allowed to pre-determine possible word-candidates.

Similarly, the operations of the interpretative source are driven in the first instance from the bottom-up, on the basis of information about the words being recognized and about the structural properties assigned to these words by the structural source. Both sources interact with each other, and with the original selection of the appropriate words (and meanings) from the recurrent bursts of word-candidates at the onset of each word, but they do so strictly under the control of the context defined from the bottom-up by the interpretation of the acoustic-phonetic input.

Closely related to this bottom-up priority is the obligatory processing property of all three knowledge sources. If the appropriate bottom-up inputs are presented to these sources, then they must run through their characteristic operations on these inputs. It is important here to distinguish between obligatory, bottom-up priority processing, and the type of autonomous processing that we rejected earlier. The obligatory and bottom-up aspects of structural processes, for example, do not mean that they cannot be influenced by interactions with interpretative processes — any more than the obligatory and bottom-up aspects of word-recognition processes mean that they are not fundamentally interactive in nature. In the case of structural analysis, the facilitatory or disambiguating effects of interpretative interactions are usually less noticeable, because of the restricted set of structural alternatives that sequences of words usually permit. But when structural ambiguities can not be directly resolved by subsequent bottom-up information, then it is possible to detect immediate interpretative interactions with structural assignment processes.

In a recent study addressed to this issue, we showed that structurally ambiguous sentence fragments, such as “folding chairs” and “landing planes”, are immediately disambiguated when an interpretative context is present that biases towards one reading rather than another (Tyler and Marslen-Wilson, 1977). Given the design of this experiment, the fragments could not have been disambiguated without some rather sophisticated on-line processing that involved putting together the meanings of the two words in the fragment, their structural relationships, and the interpretative context in which they occurred.

The assumption of bottom-up obligatory processing enables us to answer a number of questions about language understanding. First, returning to the word-monitoring results, these principles explain why the Syntactic Prose

slope was so similar to the Normal Prose slope. In each case the structural knowledge source was operating to provide similar structural assignments, which provided increasingly stronger constraints over word-positions. This overall effect is the same because the gross structure of the materials was identical in both prose contexts, so that the structural knowledge source, with its obligatory and bottom-up priority properties, would reach the same answers in either case. It would, therefore, place equivalent restrictions on word-selection processes across word-positions in both contexts. In general, the resolution of structural analyses may sometimes be slower in Syntactic Prose than in Normal Prose, but the answers will not differ — unless the material is unresolvably ambiguous without an appropriate interpretative context. This is closely analogous to the situation in word-recognition. A word in a Syntactic Prose or Random Word-Order context will eventually be correctly recognized, if the bottom-up input is adequate, but it will take longer for the point of recognition to be reached.

Secondly, it is the obligatory bottom-up properties of the processes within each knowledge source that enable the lexical and structural sources to function when a complete analysis is not possible. That is, the structural knowledge source will still function when there is no interpretative framework available, and the lexical source will still function when neither structural nor interpretative analyses are available. It is essential that they be able to do this. First, so that they can deal with prose materials like Random Word-Order or Syntactic Prose. Second, and more fundamentally, so that they can deal with the natural variation in the word-by-word resolvability of normal utterances. There will inevitably be moments during processing where the interpretative or the structural analysis of the input is delayed; either because of temporarily unresolved ambiguity, or simply because too little information is available to resolve an indeterminacy in the interpretation of the input with respect to one or another knowledge source.²⁶

Thirdly, and related to the previous point, the obligatory bottom-up principle enables us to define the role of "seriality" in an interactive processing system. What the present model has in common with an autonomous processing system is the priority it assigns to bottom-up inputs, and the consequent dependencies between knowledge sources. Thus, for example, the interpretative source normally requires its input to be analyzed in terms of both

²⁶ As these remarks imply, we do not claim that the input is *always* fully interpreted word-by-word as it is heard; only that it in principle *can* be. Whether a particular utterance will be completely interpreted as it is heard depends on the ordering of processing information across the utterance, and on the dependencies between different constituents (see Harris, Note 3; Marslen-Wilson *et al.*, 1978). But there is nothing intrinsic to the structure of the processing system itself that would prevent immediate on-line interpretation.

the lexical and the structural knowledge sources. However, given that the structural and interpretative sources can apparently have access to even the earliest stages of word-recognition processes, this means that the dependencies between knowledge sources need not result in any intrinsic delay in the timing with which their characteristic influences start to become detectable – except, presumably, for the short period at the beginning of a word during which the word-initial cohort is being defined. However, as the second point implies, the present system will appear to behave like a serial system in cases where a complete analysis is not possible. For example, in word-recognition in Random Word-Order contexts the selection of a unique word-candidate does depend solely on bottom-up information, and the availability of semantic analyses of the word do seem to be delayed relative to phonological analyses.

We can, in summary, treat “serial” phenomena in language processing as a special case of a more general on-line interactive system. This more general system is designed to function in the primary domain for spoken language understanding – that is, utterances heard in discourse contexts. The interactive properties of this system enable it to take full advantage of whatever constraints these contexts provide, and the seriality of its operation under certain unusual circumstances simply reflects the robustness of its underlying processing principles.

Fourthly, the obligatory bottom-up processing principles enable us to say something about the question of processing levels; that is, whether we should assume, as in the serial theories and in the interactive parallel theory, that each knowledge source is represented during processing as a computationally distinct level of analysis of the input. In the present context, this question can be seen to be artifactual, generated by the autonomy assumption and by linguistic preconceptions.

The major consequence of the obligatory bottom-up principles for the structure of processing is that they entail that the input will always be represented at the maximal level of representation to which its analysis can be taken. The principles require that the analysis of the input must propagate as far as its properties permit. The extent to which a given input can propagate will determine the “level of representation” at which it becomes perceptually available. This is the only sense in which the notion of “level of representation” can be interpreted in the present system.

If the words being spoken cannot be identified, then the input is perceptually represented as a string of more or less accurately segmented groups of speech sounds – the listener hears nonsense words. If the word can be identified, then it is obligatorily perceived as a word. To the extent that sequences of words can be assigned a unique structural description, then this will

become part of their perceptual representation. To the extent that an interpretation can be achieved as well, then they will be perceived in these terms. In fact, it may be more accurate to say that at this maximally interpreted level of perceptual representation of the input, the structural and lexical aspects of the percept are (at least momentarily) absorbed into this interpretative representation. We can, therefore, understand the implications of the constant effect of interpretative context from the first to the last word-position in Normal Prose in Experiment 1. Utterances in contexts are normally interpretable as they are heard, and it is this that provides the familiar illusion of a seamless communicative process.

The concept of a fixed layering of "processing levels" seems, then, to be inapplicable to the interactive processing of a normal utterance. We should emphasize, also, that the forms of perceptual representation that one obtains for materials that cannot be fully analyzed are not to be confounded with the intermediate levels of analysis proposed in serial theories. Syntactic Prose sequences, for example, are perceived as sequences of words, with meanings, that stand in some specifiable structural relationship to each other, but which cannot be intelligibly represented in terms of the listener's interpretative frameworks. It is clear as one listens to Syntactic Prose that one cannot avoid constantly trying to force an interpretative analysis of it. What gives the material its curiously quasi-poetic quality is that one often succeeds in almost finding a coherent, though bizarre, interpretation of it, only to have this interpretation collapse as more of the sequence is heard (see, for example, sentence (2) in footnote¹⁶). This form of perceptual representation should evidently not be confused with the string of syntactic labels that would make up the representation of a sentence within an autonomous syntactic processor.

If we examine the notion of fixed processing levels a bit further, then we can see that it is also inconsistent with the strategies of interactiveness and of efficiency. Consider, for example, the construction of a strictly lexical level of representation during processing — consisting, presumably, of an unstructured and uninterpreted string of words. Since the evidence shows that the recognition of each word is conducted at least partly in terms of its structural and interpretative suitability, then the subsequent construction of a strictly lexical level of representation would require that each word be then somehow stripped of the context in terms of which it was originally recognized. This seems both unlikely and highly inefficient. In fact, the process of word-recognition should be seen in a different perspective. What is important about the contextual intervention in word-recognition is not just that it facilitates word-identification, but also that it enables the structurally and interpretatively relevant aspects of the word to be directly incorporated into the on-

line interpretation of the utterance. This is, after all, what the system recognizes words for; not as ends in themselves.

Similar arguments apply to the possibility of an independent level of representation in terms of the structural knowledge source. The word-recognition data show that structural influences affect primary word-identification decisions, so that word-candidates are being analyzed in terms of their structural as well as their interpretative suitability. There is no reason to suppose that these two types of contextual influence are then separated in the later processing of the input, only to be brought back into contact again at the final stage of processing. The insertion of an intervening strictly structural level of processing not only serves no purpose in an interactive system, but would also slow down and complicate the analysis process.

A final and important proviso concerns the word "on-line", as we have used it throughout this paper. Everything we have said here about processing structure applies only to what we will call *normal first-pass processing*; that is, to the sequence of obligatory operations that the listener runs through in interpreting a normal utterance as he hears it in a natural context. But once these obligatory routines have been performed, then the listener can have access to the analysis of the input in many different ways for many different purposes. He can, for example, disentangle the acoustic-phonetic properties of the input from the other aspects of its analysis, he can focus on the structural relations between words, and so on. Our claims here have nothing to do with these secondary and, we assume, "off-line" processes.

The important point here is that the term "on-line" should be restricted to the obligatory, first-pass processes, as reflected in true on-line tasks. This is the reason for our methodological emphasis on fast reaction-time tasks, in which the response can be closely related in time to a specific point in the (auditory) stimulus materials. If experimental tasks are used where a close temporal relationship between the sensory input and the response cannot be specified in this way — for example, by using post-sentence measures — then off-line processes may well be mainly responsible for the subject's response. It is only to the extent that a task is, directly or indirectly, primarily determined by the properties of on-line processes that one can be confident that it is tapping the basic properties of language processing operations.

References

- Battig, W.F., and Montague, W.E. (1969) Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms. *J. exper. Psychol. Mono.*, 80, (3, pt. 2).
- Bever, T.G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York, Wiley.

- Cairns, H.S., and Kamerman, J. (1975) Lexical information processing during sentence comprehension. *J. verb. Learn. verb. Behav.*, 14, 170-179.
- Carroll, J. M., and Bever, T. G. (1976) Sentence comprehension: A study in the relation of knowledge to perception. In E. C. Carterette and M. P. Friedman (Eds.), *The handbook of perception. Vol. 5. Language and Speech*. New York, Academic Press.
- Carroll, J. M., Tanenhaus, M. K., and Bever, T. G. (1978) The perception of relations: The interaction of structural, functional, and contextual factors in the segmentation of sentences. In W. J. M. Levelt and G. B. Flores D'Arcais (Eds.), *Studies in the perception of language*. New York, Wiley.
- Chomsky, N. (1957) *Syntactic structures*. The Hague, Mouton.
- Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge, Mass., MIT Press.
- Clark, H. H. (1974) Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics, Volume 12: Linguistics and adjacent arts and sciences*. The Hague, Mouton.
- Cole, R. A., and Jakimik, J. (1979) A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, New Jersey, LEA.
- Cutler, A., and Norris, D. (1979) Monitoring sentence comprehension. In W. E. Cooper and E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, New Jersey, LEA.
- Fodor, J. A., Bever, T. G., and Garrett, M. F. (1974) *The psychology of language*. New York, McGraw-Hill.
- Fodor, J. A., and Garrett, M. F. (1967) Some syntactic determinants of sentential complexity. *Percept. Psychophys.*, 2, 289-296.
- Forster, K. (1974) The role of semantic hypotheses in sentence processing. In Colloques Internationaux du CNRS, No. 206, *Problèmes actuel en psycholinguistique*, Paris.
- Forster, K. (1976) Accessing the mental lexicon. In R. J. Wales and E. C. T. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam, North-Holland.
- Forster, K. (1979) Levels of processing and the structure of the language processor. In W. E. Cooper and E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, New Jersey, LEA.
- Foss, D. J. (1969) Decision processes during sentence comprehension: effects of lexical item difficulty and position upon decision times. *J. verb. Learn. verb. Behav.*, 8, 457-462.
- Foss, D. J. (1970) Some effects of ambiguity upon sentence comprehension. *J. verb. Learn. verb. Behav.*, 9, 699-706.
- Foss, D. J., and Jenkins, C. M. (1973) Some effects of context on the comprehension of ambiguous sentences. *J. verb. Learn. verb. Behav.*, 12, 577-589.
- Garrett, M. F. (1978) Word and sentence perception. In R. Held, H. W. Leibowitz, and H.-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII, Perception*. Berlin, Springer Verlag.
- Garrett, M. F., Bever, T. G., and Fodor, J. A. (1966) The active use of grammar in speech perception. *Percept. Psychophys.*, 1, 30-32.
- Hakes, D. T. (1971) Does verb structure affect sentence comprehension? *Percept. Psychophys.*, 10, 229-232.
- Jarvella, R. J. (1971) Syntactic processing of connected speech. *J. verb. Learn. verb. Behav.*, 10, 409-416.
- Kenyon, J. S., and Knott, T. A. (1953) *A pronouncing dictionary of American English*. Springfield, Mass., G. and C. Merriam.
- Levelt, W. J. M. (1974) *Formal grammars in linguistics and psycholinguistics, Vol. 3: Psycholinguistic applications*. The Hague, Mouton.
- Levelt, W. J. M. (1978) A survey of studies in sentence perception: 1970-1976. In W. J. M. Levelt and G. B. Flores D'Arcais (Eds.), *Studies in the perception of language*. New York, Wiley.
- Marslen-Wilson, W. D. (1973) Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522-523.
- Marslen-Wilson, W. D. (1975) Sentence perception as an interactive parallel process. *Science*, 189, 226-228.
- Marslen-Wilson, W. D. (1976) Linguistic descriptions and psychological assumptions in the study of sentence perception. In R. J. Wales and E. C. T. Walker (Eds.), *New approaches to the study of language*. Amsterdam, North-Holland.

- Marslen-Wilson, W. D., Tyler, L. K., and Seidenberg, M. (1978) Sentence processing and the clause-boundary. In W. J. M. Levelt and G. B. Flores D'Arcais (Eds.), *Studies in sentence perception*. New York, Wiley.
- Marslen-Wilson, W. D., and Welsh, A. (1978) Processing interactions and lexical access during word-recognition in continuous speech. *Cog. Psychol.*, 10, 29–63.
- McNeill, D., and Lindig, K. (1973) The perceptual reality of phonemes, syllables, words, and sentences. *J. verb. Learn. verb. Behav.*, 12, 419–430.
- Mehler, J., Segui, J., and Carey, P. (1978) Tails of words: monitoring ambiguity. *J. verb. Learn. verb. Behav.*, 17, 29–37.
- Meyer, D. E., Schvaneveldt, R. W., and Ruddy, M. G. (1975) Loci of contextual effects on visual word recognition. In P. Rabbit and S. Dornic (Eds.), *Attention and Performance*, V. New York, Academic Press.
- Miller, G. A. (1962) Some psychological studies of grammar. *Amer. Psychol.*, 17, 748–762.
- Miller, G. A., and Chomsky, N. (1963) Finitary models of language users. In R. D. Luce, R. R. Bush and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol II. New York, Wiley.
- Miller, G. A., Heise, G., and Lichten, W. (1951) The intelligibility of speech as a function of the context of the test materials. *J. exper. Psychol.*, 41, 329–335.
- Morton, J. (1969) The interaction of information in word-recognition. *Psychol. Rev.*, 76, 165–178.
- Morton, J. (1979) Word recognition. In J. Morton and J. C. Marshall (Eds.), *Psycholinguistics series 2: Structures and processes*. London, Paul Elek.
- Morton, J., and Long, J. (1976) Effect of word transitional probability on phoneme identification. *J. verb. Learn. verb. Behav.*, 15, 43–51.
- Newman, J., and Dell, G. (1978) The phonological nature of phoneme monitoring: a critique of some ambiguity studies. *J. verb. Learn. verb. Behav.*, 17, 359–374.
- Reiger, C. (1977) Viewing parsing as word-sense discrimination. In W. Dingwall (Ed.), *A survey of linguistic science*. Stamford, Conn: Greylock.
- Reiger, C. (1978) GRIND-1: First report on the Magic Grinder Story Comprehension project. *Discourse Processes*, 1, 211–231.
- Shiffrin, R. M., and Schneider, W. (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.*, 84, 127–190.
- Sternberg, S. (1969) The discovery of processing stages: Extensions of Donder's method. In W. G. Koster (Ed.), *Attention and performance II*. Amsterdam, North-Holland.
- Swinney, D. A. Lexical access during sentence comprehension: (Re)Consideration of context effects. *J. verb. Learn. verb. Behav.*, in press.
- Tukey, J. W. (1972) *Exploratory data analysis*. 2nd preliminary edition. Reading, Mass., Addison-Wesley.
- Tyler, L. K. (1980) Serial and interactive theories of sentence processing. to appear in J. Morton and J. C. Marshall (Eds.), *Psycholinguistics series 3*. London, Paul Elek.
- Tyler, L. K., and Marslen-Wilson, W. D. (1977) The on-line effects of semantic context on syntactic processing. *J. verb. Learn. verb. Behav.*, 16, 683–692.
- Wainer, H., and Thissen, D. (1975) Multivariate semi-metric smoothing in multiple prediction. *J. Amer. Statist. Assoc.*, 70, 568–573.
- Wainer, H., and Thissen, D. (1976) Three steps towards robust regression. *Psychometr.*, 41, 9–34.
- Winograd, T. (1972) *Understanding natural language*. New York: Academic Press.

Reference Notes

- 1 Marslen-Wilson, W. D. Speech shadowing and speech perception. Unpublished Ph. D. thesis, Department of Psychology, MIT, 1973.
- 2 Marslen-Wilson, W. D. Sequential decision processes during spoken word recognition. Paper presented at the Psychonomic Society meetings, San Antonio, Texas, 1978.
- 3 Harris, J. Levels of speech processing and order of information. Unpublished Ph. D. thesis, Committee on Cognition and Communication, Department of Behavioral Sciences, The University of Chicago, 1978.

Résumé

On présente deux expériences faites pour étudier la compréhension du langage parlé en fonction du parcours temporel dans la chaîne. La recherche a porté à la fois sur les processus (locaux) de reconnaissance des mots et sur les processus structuraux et interprétatifs (globaux).

Dans chacune des expériences on a utilisé trois tâches de détection de mots. Ces tâches se distinguaient par la qualification sous laquelle le mot à détecter était décrit (phonétique, sémantique où les deux) et par trois contextes différents (normal, sémantiquement anormal où brouillé).

D'autre part les mots cibles se distribuaient selon neuf positions dans les phrase-tests. La présence ou l'absence d'un contexte phrasique permettait une estimation des effets inter-phrases sur les processus locaux et globaux. L'ensemble des résultats donnent une image détaillée de la structuration temporelle des différents processus. L'hypothèse d'un calcul du langage interactif en temps réel est favorisée. Selon cette hypothèse, durant le traitement, les diverses sources de connaissance, lexicales, structurales (syntaxiques) et interprétatives communiquent et interagissent d'une façon maximalement précise et efficace.