

Word-Monitoring Tasks Interact with Levels of Representation During Speech Comprehension

David J. Townsend,¹ Michael Hoover,² and Thomas G. Bever³

Researchers frequently use data from monitoring tasks to argue that constraints on meaning facilitate lower-level processes. An alternate hypothesis is that the processing level that a monitoring task requires interacts with discourse-level processing. Subjects monitored spoken sentences for a synonym (semantic match), a nonsense word (phonological match), or a rhyme (phonologically and semantically constrained matching). The critical targets appeared at the beginning of the final clause in two-clause sentences that began with if, which signals a semantic analysis at the discourse level, or with though, which maintains a surface representation. Synonym-monitoring times were faster for if than for though, nonsense word-monitoring times were faster for though than for if, and rhyme-monitoring times did not differ for if and though. The results show that conjunctions influence how listeners allocate attention to semantic versus phonological information, implying that listeners form these kinds of information independently.

Phonemes are parts of words, words are parts of phrases, phrases are parts of clauses, and clause meanings are parts of the mental representation of discourse. This hierarchical organization of language suggests that comprehension proceeds serially from phonemes, to words, to phrases, to clauses, to discourses. After all, how can we know what word someone uttered if we have not first identified the phonemes that make up the word? How can we know the meaning of a discourse without knowing the meanings of each of the clauses? Thus, comprehension should involve sequential decisions at several different levels. Yet, we have the intuition of grasping the full meaning of speech immediately and effortlessly. If comprehension requires

This research was supported by BNS-8120463 from the National Science Foundation to Montclair State University, and by a Distinguished Scholar Award from Montclair State University. Requests for reprints should be sent to David J. Townsend.

¹ Department of Psychology, Montclair State University, Upper Montclair, New Jersey 07043.

² Department of Educational and Counselling Psychology, McGill University, 3700 McTavish, Montreal, Quebec H3A 1Y2 Canada.

³ Department of Linguistics, University of Arizona, Tucson, Arizona 85721.

serial computation of the properties at different levels of linguistic structure, how do listeners manage to make decisions at all these levels and arrive at meaning so quickly?

Some simple facts about speech and language reveal the enormity of the dilemma of serial computation. A normal rate of speech is about four words per second (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). If there are three phonemes per word and forty phonemes in English, listeners must decide which one of forty possible phonemes occurred every 83 ms. If listeners know 50,000 words, they must then decide which of these 50,000 words occurred every 250 ms. Once they have decided about a sequence of words, listeners also must assign a phrase structure to the sequence, and then a meaning to the corresponding sentence, all within 2000 ms if there are eight words per sentence. These facts suggest that listeners make at least two dozen decisions every second of connected speech. Clearly, something is wrong with the upward serial model.

In this paper, we contrast two alternatives to serial computation. The *interactionist* hypothesis is that listeners use context to reduce the number and complexity of decisions. The *independence* hypothesis is that listeners compute phonemes, words, and phrases independently and simultaneously.

Interactionism eases the burden of serial computation by proposing that context can facilitate decisions about structure at lower levels. By reducing the number of options that are possible at a lower level, it is easier to make decisions at that level. For example, when listeners encounter a word that context has activated, they can respond more readily to phonological features of the word (Elman & McClelland, 1984; Marslen-Wilson & Welsh, 1978). A similar process may occur at the syntactic level: semantic context may render some hypotheses about syntactic structure implausible, thereby simplifying syntactic decisions (e.g., Crain & Steedman, 1985; McClelland, St. John, & Taraban, 1989; Tyler & Marslen-Wilson, 1977). In both cases, activating information at a higher level facilitates decisions at a lower level by reducing the number of candidates to be considered at the lower level. According to interactionism, information at each level can “cascade” downward to reduce the number of candidates to be considered at all other levels.

One kind of evidence that supports interactionism is that subjects can detect a target word faster when there is more semantic or syntactic context. Several studies have found that the time to detect phonological, lexical, and semantic properties decreases as the number of preceding words increases (Marslen-Wilson & Tyler, 1975; Marslen-Wilson, Tyler, & Seidenberg, 1978). For example, to measure the availability of phonological information, Marslen-Wilson *et al.* (1978) used a rhyme-monitoring task in which subjects listened to sentences for a word that rhymed with a cue word. To measure the availability of semantic information, they used a category-monitoring task in

which subjects listened for an instance of a semantic category. Since response times for both rhyme and category monitoring were shorter for targets at the end of a clause than for targets at the beginning, Marslen-Wilson *et al.* (1978) argued that the semantic and syntactic constraints on a word at the end of a clause immediately reduce the number of candidates that listeners need to consider at the phonological and semantic levels.

In contrast to interactionism, the independence hypothesis maintains that listeners simultaneously use level-specific strategies to form representations of phonemes, words, phrases, and meaning, as these become available. Since the functional vocabularies are different at each level, the strategies that apply at various levels are independent. The independence hypothesis implies that processes at different levels do not share limited computational resources, although they may share limited attentional resources.

Analysis of the rhyme-monitoring task can sharpen the distinction between the interactionist and independence hypotheses. The rhyme-monitoring task involves two separate processes that are mediated by the lexical level—one phonological and the other semantic. For example, upon hearing a cue word like DOUBT, subjects can generate a list of candidate words that have the required phonological form: *pout*, *gout*, *trout*, and so on. A semantic and syntactic context like “Though Mary rarely cooks . . .” increases the likelihood that there will be a word that refers to some kind of food, eliminating *pout* and *gout* from the pool of phonologically permissible candidates. Thus, rhyme monitoring is not simply a phonological task; it also can be influenced by syntactic or semantic information from the clause. Consequently, the reduction in rhyme-monitoring performance that Marslen-Wilson *et al.* (1978) observed at the end of a clause may have occurred because semantic and syntactic constraints narrow the pool of possible target words, not because they narrow the pool of possible phonemes.

The interactionist and independence hypotheses differ in their claims about how context affects lower-level processing. According to interactionism, rhyme monitoring informs phonological decisions directly by reducing the number of phonological candidates to consider. The independence hypothesis implies that facilitating the choice of a candidate target word reduces only the number of word candidates that listeners must consider, not the number of phonological candidates. Context may, however, influence how listeners allocate attention toward phonological, lexical, syntactic, or semantic information that may be available at any given time.

Evidence for such shifts of attention between levels comes from studies that have found that increasing the emphasis on higher-level information *decreases* the salience of properties at lower levels. For example, instructing subjects to understand sentences rather than to recite them word for word presumably improves their ability to make judgments about the meaning of

the sentences. However, Green (1975) found that instructions to comprehend *impaired* subjects' ability to detect tones that are superimposed on speech. Such a result suggests that meaning does not facilitate the acoustic task of detecting a tone, but rather, it "competes" for attention with acoustic information (for other examples, see Cairns, Cowart, & Jablon, 1981; Rubenstein, Garfield, & Milliken, 1970; Samuel, 1981). We call these cases in which some factor that makes it easier to respond to higher-level properties (e.g., meaning) also makes it harder to respond to lower-level properties (e.g., nonspeech tones) a *competition* effect. Competition effects present problems for the view that higher-level information reduces the complexity of processing at a lower level. These effects suggest that representations at different levels compete for attentional resources, not computational resources.

In this research we test the interactionist and independence hypotheses by comparing the effects of subordinate conjunctions on various monitoring tasks. A number of studies have found that conjunctions produce a competition effect: conjunctions that increase the salience of meaning also decrease the salience of lower-level properties. For example, Townsend & Bever (1978) found that subjects are faster to judge that the phrase USING THE TELEPHONE is synonymous with the "causal" fragment (1a) than with the "adversative" fragment (1b):

- (1) a. If Pete called up his aunt each . . . USING THE TELEPHONE
 b. Though Pete called up his aunt each . . . USING THE TELEPHONE

Townsend & Bever (1978) found that causal conjunctions have the opposite effect on judgments about lower-level information: whereas subjects indicate more quickly that the word *up* had occurred in sentence fragment (2a) compared to (2b), the location of *up* has little effect when the conjunction *if* is used.

- (2) a. Though Pete called up his aunt each . . . UP
 b. Though Pete called his aunt up each . . . UP

These results demonstrate a competition effect: a conjunction that increases the salience of meaning decreases the salience of the left-to-right order of words. This competition effect suggests that the processes that access meaning and word order *compete* with one another.

Another explanation of competition effects is possible. Since the results in Townsend & Bever (1978) involve an immediate memory task, they may be explained in terms of how listeners organize information in memory. Listeners may organize information in different ways for *if*- versus *though* clauses after they have been understood. Memory studies do not directly

address the issue about the independence of processes at different levels during comprehension.

More recent monitoring studies support an attentional interpretation of competition effects. In Townsend & Bever (1991), subjects listened to discourses for two purposes. Their main task was to make decisions about discourse-level properties, such as the author's personality, but they also had to monitor for a target word uttered by a female rather than the male storyteller. This second task requires accessing information at a more phonological level. The discourses varied the constraints on the sentence that contained the target word. Townsend & Bever (1991) found that stronger constraints on the discourse-level representation of a sentence *increased* the time to detect a change in speaker. Since stronger discourse-level constraints apparently make it easier to decide on a discourse-level meaning, this is another example of a competition effect. Since the effect occurred in an on-line listening task, it suggests that discourse-level and phonological representations are formed independently, but attended with shared resources.

We used three monitoring tasks to determine whether conjunctions produce a competition effect. If the meanings of conjunctions elicit different attentional processes, the on-line processing of initial *if* and *though* clauses will interact with the semantic, syntactic, and phonological properties of monitoring tasks. To test this idea, subjects listened to sentences that began with either *if* or *though*:

- (3) a. If Harry keeps lots of snakes on the farm kids visit every day.
- b. Though Harry keeps lots of snakes on the farm kids visit every day.

The subjects received one of three monitoring tasks. In synonym monitoring, subjects pressed a response key as soon as they heard a word that was similar in meaning to a cue, like YOUNG PEOPLE (e.g., *kids*). In rhyme monitoring, subjects pressed a response key when they heard a word that rhymed with a cue, like BIDS (*kids*). In nonsense word monitoring, subjects pressed a response key on hearing a nonsense word cue like KIG. In the critical cases, the target was the first word or syllable of the second clause.

The interactionist hypothesis maintains that context facilitates word recognition regardless of the monitoring task. Since the contexts are similar across tasks and conjunction, there will be no interaction between conjunction and task.

The independence hypothesis maintains that competition effects occur because representations that have already been formed compete for attention. Thus, the discourse-level demands of conjunctions will interact with monitoring tasks: When the monitoring task emphasizes information at one level, the task will be easier if there is a conjunction that increases the

salience of information at that level. Of the three tasks, monitoring for a synonym most emphasizes the use of semantic and syntactic information from the clause. In this case, subjects may generate and anticipate candidate targets that are appropriate for the semantic and syntactic properties of the emerging sentence. Synonym monitoring will be easier when the demands of discourse-level processing enhance semantic information, that is, when the sentence begins with *if*. Monitoring for a nonsense word emphasizes phonological/lexical information and semantic and syntactic information from the sentence is irrelevant for locating a nonsense word. Thus, nonsense word monitoring will be easier when the demands of discourse-level processing increase the salience of lower-level information, that is, when the sentence contains *though*. If monitoring for a rhyme involves the use of both phonological and semantic/syntactic information, the effects of conjunction on rhyme monitoring will be intermediate between its effects on synonym monitoring and nonsense word monitoring: *if* will enhance the use of semantic/syntactic constraints on the rhyme target, while *though* will enhance the use of phonological constraints on the target.

METHOD

Procedure

Different groups of subjects performed either a synonym-monitoring task, a nonsense word-monitoring task, or a rhyme-monitoring task. Before starting the experiment, the subjects read instructions and examples of the cues and targets in sentences. In the synonym-monitoring task, subjects listened for a word that meant the same as a short cue phrase; in the nonsense word task, they listened for a particular nonsense word; and in the rhyme-monitoring task, they listened for a word that rhymed with a cue word. Before each trial, the subject read the cue for that trial on an index card. The subject then heard a recording of the sentence through headphones in the right ear. In each task, the subject indicated detection of a target by pressing a response key. After hearing the complete sentence, the subject produced a paraphrase of the sentence. Each subject received a short rest every 26 trials.

A second generation tape was used for all three tasks. For the nonsense word-monitoring task, tape splicing was used to replace the synonym and rhyme target word with a nonsense word.

Materials

There were ten lists of materials. Of the 104 sentences in each list, 96 contained a target. In each list, the critical sentences were drawn from a

pool of 20 sentences that contained an initial subordinate adverbial clause. The target in these sentences was the first word of the main clause and it was always a noun. For each list, a different set of four of the pool of 20 sentences were selected for manipulating the conjunction variable. In each list, two critical sentences contained initial *if* and two contained initial *though*. The target words were the 8th, 9th, or 10th word of the sentence. Conjunctions other than *if* or *though* introduced the remaining 16 sentences in the pool of 20.

The remaining 76 positive trials in each list were used to balance for other properties. The location of the target ranged from the second word of the sentence to the fourteenth word. The grammatical class of the target ranged over nouns, verbs, adjectives, and adverbs. There was a variety of sentence types including single clause, relative clause, complement clause, final adverbial clause, and coordinate clause sentences.

The targets for the nonsense word task were single syllable nonsense words and the targets for both the synonym and rhyme tasks were single-syllable words. The cues for the rhyme task were also single-syllable words, while the synonym cues ranged from 1 to 4 words in length. Each rhyme cue in the critical trials was orthographically similar to the target. To control for differences in number of possible targets for cue words and the strength of the associative relation between cue and target, each cue-target pair appeared with each conjunction across lists.

Subjects

Thirty Columbia College students participated as paid subjects. All subjects were male, right-handed native speakers of English with normal hearing.

RESULTS

Eleven response times (6.1%) were more than two standard deviations beyond a subject's mean response time. To decrease the effects of these outliers, they were replaced with the subject's mean response time plus two standard deviations. There were no misses. Differences in response times were tested with mixed analysis of variance by subjects using conjunction and block as within-subject variables and task as a between-subject variable; differences were also tested with analysis of variance by items using conjunction and task as within-item variables. The mean response times appear in Table I.

The mean response time was 545 ms. Response times were fastest for nonsense word monitoring (367 ms), slowest for synonym monitoring

Table I. Mean Response Times (ms) on Word-Monitoring Tasks

Task	"if"	"though"	Mean
Synonym	624	765	695
Nonsense word	391	342	367
Rhyme	575	569	572

(645 ms), and response times for rhyme monitoring fell between these two extremes (572 ms). The response time difference between the three monitoring tasks was significant, $F_1(2,27) = 39.2, p < .001, F_2(2,38) = 15.2, p < .001$.

The interaction between task and conjunction was significant, $F_1(2,27) = 4.64, p < .05, F_2(2,38) = 3.45, p < .001$. For synonym monitoring, response times were faster for *if* than for *though*, $F_1(1,27) = 39.2, p < .01, F_2(2,38) = 32.0, p < .001$. For nonsense word monitoring, response times were faster for *though* than for *if*, $F_1(1,27) = 6.76, p < .05, F_2(2,38) = 3.86, p < .001$. For rhyme monitoring, conjunction had no effect on response times—both $F_s < 1$.

DISCUSSION

The results confirmed our analysis of the monitoring tasks. In monitoring for a nonsense word, the only information that is relevant is the phonological match between incoming speech and the cue word; we found that subjects used this phonological information more effectively following a *though* clause and less effectively following an *if* clause. In monitoring for a synonym, subjects can use semantic and syntactic information from the clause to predict where the target will occur; we found that subjects use this higher-level information more effectively following an *if* clause and less effectively following a *though* clause. We found no difference between conjunctions in monitoring for a rhyme. The interaction of conjunction with monitoring tasks demonstrates a functional distinction between processing for the sentence-level properties of meaning and syntax and processing for phonological properties.

The results confirm our hunch that rhyme monitoring is not a simple phonological task. We found clear and opposite effects of conjunction on performing a task that is clearly semantic (synonym recognition) versus a task that is clearly phonological (nonsense word recognition); these conjunction differences suggest that *if* and *though* focus attention on one kind of information at the expense of the other. Using the same materials, we found that conjunctions have no effect on performing the rhyme-monitoring

task. The failure to find a conjunction effect for rhyme monitoring suggests that subjects can use the semantic information from the sentence to eliminate candidate words that are generated from the phonological information that is provided by the cue word. However, since *if* and *though* focus attention on different levels, rhyme monitoring does not make use of all information that is available. For *if*, rhyme monitoring emphasizes the integration of semantic/syntactic information from the clause with potential target candidates, but for *though*, it emphasizes phonological matching of words in the clause with potential target candidates.

The results have implications for the nature of competing levels, the categorical nature of attention, and the role of a sentence-level representation. We discuss these implications in turn.

First, the results indicate that the level of representation that an adverbial clause elicits is truly phonological, not just “surface” or “lexical.” That is, the conjunction effect in monitoring for nonsense words shows that it does not depend on having accessed an entry in the mental lexicon.

Second, these results show that levels of representation mediate shifts in attention during the on-line processing of sentences. According to interactionism, it should be possible to attend to a representation of a clause at one level while processing another clause at a different level. Indeed, this is required if a semantic representation of context is going to inform on-going syntactic, lexical, or phonological processing: the perceptual system must simultaneously attend to the semantic representation of the context and the current phonological representation. However, the results of this study confirm previous findings that, under certain conditions, the shift of attention between the semantic and the phonological levels is categorical (Townsend & Bever, 1978, 1991).

Third, the existence of the conjunction effect for targets in the final main clause suggests that sentences may mediate attentional shifts. It would be possible, and perhaps even functional, if listeners attended to the semantic level of a main clause that follows an initial *though*-clause, which might best facilitate the interpretation of the *though*-clause. Our results for nonsense-word monitoring, however, indicate that listeners attend to information in the main clause that follows an initial *though* clause primarily at the phonological level. This suggests a process in which, after hearing an initial *though* clause, listeners represent both the *though* clause and the following main clause at a phonological level. On the other hand, our synonym monitoring results indicate that listeners attend to main clause information following an *if* clause primarily at the semantic/syntactic level. Thus, whether the initial clause of a sentence directs attention to the syntactic/semantic level or to the phonological level, attention remains directed toward that level during the final clause. This result explains another probe result: for tests of final main

clauses, Townsend & Bever (1978) found that meaning is relatively more accessible when the initial clause is introduced by *if*, while the effects of word position on word probe response times are greater when it is introduced by *though*. The fact that a similar reversal occurs for synonym versus nonsense word monitoring suggests that Townsend & Bever's (1978) results for final main clauses are due to shifts of attention between the semantic level and more superficial levels.

REFERENCES

- Cairns, H. S., Cowart, W., & Jablon, A. D. (1981). Effects of prior context upon the integration of lexical information during sentence processing. *Journal of Verbal Learning and Verbal Behavior*, *20*, 445–453.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*. Cambridge: Cambridge University Press.
- Elman, J. L., & McClelland, J. L. (1984). Speech perception as a cognitive process: The interactive activation model. In N. Lass (Ed.), *Speech and language, Vol. 10*. New York: Academic Press.
- Green, D. W. (1975). The effects of tasks on the representation of sentences. *Journal of Verbal Learning and Verbal Behavior*, *14*, 275–283.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.
- Marslen-Wilson, W. D., & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature*, *257*, 784–786.
- Marslen-Wilson, W., Tyler, L., & Seidenberg, M. (1978). Sentence processing and the clause boundary. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 219–246). New York: Wiley.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.
- McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, *4*, S1287–336.
- Rubenstein, H., Garfield, L., & Milliken, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*, 487–492.
- Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474–494.
- Townsend, D. J., & Bever, T. G. (1978). Inter-clause relations and clausal processing. *Journal of Verbal Learning and Verbal Behavior*, *17*, 509–521.
- Townsend, D. J., & Bever, T. G. (1991). The use of higher-level constraints in monitoring for a change in speaker demonstrates functionally distinct levels of representation during discourse comprehension. *Language and Cognitive Processes*, *6*, 49–77.
- Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 683–692.