The effects of syntactic structure on letter detection in adjacent function words

SETH N. GREENBERG
Union College, Schenectady, New York

ASHER KORIAT
University of Haifa, Haifa, Israel

and

ANNE SHAPIRO
Colgate University, Hamilton, New York

In the present study, we examined letter detection in very frequent function-word sequences. It has been claimed that such sequences are processed in a unitized manner, thus preempting access to their constituent letters. In contrast, we showed that letter detection in the words *for* and *the* (1) was no more difficult when the words appeared in adjacent locations in a sentence (familiar) than when they appeared apart (less familiar sequence) and (2) was contingent upon the words' syntactic roles within the phrase. Thus, letter detection in *for* was easier when the sequence was separated by a clause boundary than when the words were part of the same clause. The advantage derived from clause separation was strongest when a comma divided clauses. These results challenge the unitization account of the "missing-letter" effect in common phrases and support a position where this phenomenon is seen to reflect the extraction of phrase structure during reading.

Very familiar function words in English, such as *the* and *for*, tend to conceal their constituent letters during reading (Corcoran, 1966; Drewnowski & Healy, 1977, 1980; Koriat, Greenberg, & Goldshmid, 1991; Proctor & Healy, 1985; Read, 1983). Thus, when readers are charged with detecting a target letter *t* in continuous text, they miss the letter more often when it appears as part of *the* than when it appears as part of a less frequently occurring word, such as *rather*. Healy (1976) explained this "missing-letter effect" as resulting from the unitization of familiar parts of text. Familiar words or even familiar phrases permit the perceptual system to identify them as whole units before component letters are recognized (see, e.g., Drewnowski & Healy, 1977). The flow of reading then pushes the reader on toward the next unit before the constituents have a chance to be identified (Proctor & Healy, 1985).

One prediction of the unitization position is that factors that destroy the perceptual familiarity of words should enhance letter detection. Thus, misspelling of words reduces or eliminates the missing-letter effect (e.g., Healy, 1980; Healy, Fendrich, & Proctor, 1990), as do unfamiliar graphic presentations of words, such as mixing letter cases within words (Drewnowski & Healy, 1982). In addition, scrambling words in a sentence, so as to disturb familiar word sequences, generally improves letter detection (Drewnowski & Healy, 1980). Healy, Oliver, and McNamara (1987) found that letter detection was also better for high-frequency words when they were presented in isolation than when they appeared in context with one or more other words.

Finally, and of particular interest to the present project, Healy, Conboy, and Drewnowski (1987) studied letter detection in words appearing in very familiar word "frames," such as *for the*. Function-word sequences are among the most frequent sequences in English (see Umeda & Kahn, 1982). When asterisks were inserted in the space between the two function words (e.g., *for*the*), letter detection improved. Meanwhile, inserting asterisks between *the* and a content word that followed it (e.g., *the*boy*) had little effect upon letter detection. They reasoned that if

the units being characterized by the letter detection paradigm are syntactic units, defined in terms of word function, then the interword space after the test word *the* would be more likely to be embedded in the multiple-word unit . . . If the units in question depend more on word familiarity, then the interword space before the test word *the* should be more likely to be embedded (Umeda & Kahn, 1982, p. 285).

Thus, on the basis of their findings, they concluded that it was the perceptual familiarity of the encompassing supraword unit, rather than the more generalized syntactic relationship between words (e.g., the common construction the + noun), that contributes to the missing-letter effect. In sum, these findings demonstrated that letter detection is sensitive to the familiarity of both individual words and the multi-word sequences in which they are embedded. Altogether, these results were consistent with the unitization account of the missing-letter effect, assuming that the perceptual familiarity of an orthographic pattern affects its unitization (see Healy, Conboy, & Drewnowski, 1987).

Recently, however, this account has been challenged by findings suggesting that the missing-letter effect for such words as for, the, and to may in fact be due to their syntactic role in text, rather than to their high familiarity (Greenberg & Koriat, 1991; Koriat & Greenberg, 1991; Koriat et al., 1991). In English, the most frequent words (e.g., the, and) are also function words, which convey syntactic information. Thus, although it is clear that these words are more likely to conceal their constituent letters than do other words, it is not clear that this letter-detection disadvantage derives only from their familiarity. Rather, the linguistic status of these words may also be responsible for the missing-letter effect. Therefore, Koriat and Greenberg conducted a series of studies to disentangle frequency and linguistic function.

In the first of these studies, the investigators took advantage of some of the properties of Hebrew that allow one to separate orthographic frequency from function (Koriat et al., 1991). In Hebrew, function morphemes can appear as single letters prefixed to content words. For example, l (lamed) and m (mem) prefixed to a noun serve as the function morphemes to and from, respectively. Importantly, these function prefix words were no more frequent than their control content words. Nevertheless, they engendered significantly more detection errors. Additional work by Koriat and Greenberg (1991) placed nonwords in function and content word slots in English and Hebrew text. Letter detection was worse when the nonwords appeared in the function slots. Presumably, since nonwords are of zero frequency, letter-detection performance was being affected by the apparent linguistic roles of the embedding strings in the text. Finally, Greenberg and Koriat (1991) showed that letter-detection accuracy in English function words depended on their linguistic role in the sentence. Thus, on produced more omission errors when used in its typical function role than when used in "on switch." In sum, this recent work supported a structural model according to which the missing-letter effect for common function morphemes derives from their role in supporting phrase structure rather than from their perceptual familiarity.

A central assumption of the structural model is that the missing-letter effect occurs in the transition from structure to meaning. Function morphemes play a central role in the establishment of syntactic structure, but they recede into the background as the meaning of a phrase or sentence unfolds (see also Aaronson & Ferres, 1983a; Bock, 1990). Presumably, then, the missing-letter effect occurs at a postlexical stage, after the word (or morpheme) has been identified and its linguistic role has been determined (see Koriat et al., 1991).

While the recent work of Koriat and Greenberg (1991) pointed to the linguistic role of individual words (or nonwords) as primary in the missing-letter effect, a systematic comparison between linguistic status and perceptual familiarity has not been made at the level of word frames (i.e., frequently occurring word sequences). Indeed, the observation that letter detection in the was profoundly affected by its orthographic environment testifies to the importance of perceptual familiarity at the supraword level. Of course, the effects of word scrambling on letter detection are also compatible with the structural account, because scrambling presumably impairs phrase structure as well. However, the specific findings reported by Healy, Conboy, and Drewnowski (1987) remain a puzzle. These findings implied, for example, that the perceptual unit involving the includes the preceding space and word. Thus, it was the purpose of the present work to investigate whether letter detection in a highly familiar two-word frame varies with the linguistic structure of the two-word sequence in the sentence. If syntactic structure is critical, as assumed by the structural model, then the particular status of each word within that structure ought to affect letter detection in that word within the frame, and, presumably this should be true even when the embedding word sequence is very frequent.

Specifically, then, the present research focused on the high-frequency frame for the. The first question was whether altering the structural relationship of the function words in this frame would affect letter detection. Thus, within the perceptually familiar frame for the, the words could be part of the same syntactic grouping (e.g., "He was looking for the book in his room . . .") or as part of different phrases (e.g., "The book he was looking for the night of the party . . ."). Because the unitization position holds that perceptual familiarity is the critical determinant of letter detection, manipulation of underlying structural relationships within the sequence ought to have little effect. However, should structural relationships be important, then the positioning of function words within a sentence ought to have implications for letter detection. Kimball (1973) claimed that prepositions are informative about syntactic structure when they appear at the beginning of a phrase. Indeed, Greenberg and Koriat (1991) found that an "elliptical" for appearing at the end of a phrase (e.g., "are you for or against") engenders fewer detection errors than does for beginning a prepositional phrase, presumably because, in the former case, the preposition is less helpful to the on-line extraction of phrase structure. Thus, the structural model makes different predictions regarding letter detection for function words in familiar word sequences. Specifically, in the present research, although for appears immediately before

the in both the aforementioned conditions, its linguistic status differs, because when it is presented at the end of a clause it is less useful than when it initiates a phrase in which both *for* and *the* occur. Thus, according to the structural position, letter detection ought to be better for *for* when *for* and *the* belong to different clauses than when they are part of the same clause. In contrast, because the structural status of *the* is not changed by this manipulation of boundaries, detection of *t* should not differ across these conditions. Indeed, *the* differs from *for* in a variety of ways which we assume might have an impact on their respective structural contributions. Whereas *for* may be followed by as many as seven different lexical-syntactic categories, *the* can be followed by only four. While *for* may be a preposition or conjunction, *the* is always a definite article. Finally, although the former functor may appear in a variety of locations within a syntactic unit, the latter reliably begins a noun phrase. Thus, the present study implicitly took advantage of the malleability of *for* to test the influence of syntactic context on letter detection. The focus of the present research, then, was not potential differences between letter detection in *for* and *the*, but rather the differences engendered by *for* appearing under different structural conditions.

A second question concerned the overall contribution of word frames to the detection of letters in their word constituents. To examine this contribution, Experiment 1 included sentences that contained both *for* and *the* but not in adjacent locations and not in locations where either word was next to another function word. Thus, in the latter condition, *for* and *the* appeared in sequences that were not particularly familiar. A comparison of letter detection in this condition with that obtained in the word-frame condition allows us to examine the contribution of phrase-level unitization. According to the unitization position, letter detection in each of the function words *for* and *the* ought to be even worse when these words appear in *for the*. In contrast, the structural model predicts that letter detection should not be affected by the familiarity of local context.

A third aim was to explore the possible effect of informative punctuation upon letter detection in function words. Specifically, in the sentences where *for* and *the* belonged to different clauses, the frame appeared either with a comma that helped to mark the clause boundaries (*for, the*) or with no comma inserted. The latter comparison checked whether a comma was necessary to prevent processing *for the* as if these words were part of the same clause. As Just and Carpenter (1987) noted, "punctuation is a very strong cue to syntactic analysis, unambiguously indicating sentence boundaries and, with a slightly lower certainty, indicating clause boundaries" (p. 142). Thus, without punctuation readers may miss the clause boundaries and, hence, process the relationship between *for* and *the* incorrectly, at least initially. Perhaps only when a comma is inserted will the terminal position of *for* be determined quickly enough to affect a reduction in *f* detection errors. Of course, inserting a comma may be viewed as "disrupting the frame" (although clearly

it is not the same as inserting a superfluous asterisk; see Healy, Conboy, & Drewnowski, 1987). Under this claim, the unitization position should predict a reduction in detection errors for both *for* and *the*, relative to the condition where they are parts of the same clause. Of course, unitization might suggest that the phrase *for the*, with a comma inserted, still constitutes a unitized perceptual frame since such a sequence is also familiar. Thus, depending upon the significance one assigns to the comma, the unitization position would predict either no relative reduction in detection errors for either word in the syntactically divided frame (i.e., assuming the comma is part of the familiar frame) or a reduction in detection errors for both words (i.e., assuming commas *disrupt* the familiar frame). The structural model, in contrast, predicts the insertion of a comma to reduce errors but only in *for*, not in *the*. The comma should help reveal the linguistic status of *for*, whereas the status of *the* remains the same regardless.

EXPERIMENT 1

Method

Subjects. Sixty Union College students were paid $3 each for participating in this experiment.

Stimulus Materials. Three passages were composed, containing 34, 34, and 32 unrelated sentences (for a total of 100 sentences). Of these, 50 were target sentences and were divided approximately evenly amongst the three passages. The word frame *for the* appeared in 30 of these sentences. In 10 instances, *for* and *the* appeared in the same syntactic phrase (*same*; e.g., "He is going for the gold in a swimming championship" and "Jen worked for the whole hour"). In the other 20 sentences, the words *for the* were embedded in different clauses (*different*; e.g., "Which is she going for the gold plaque?" and "Success is all he worked for the whole decade"). While *for* and *the*, in the *different* sentences belonged to different clauses, they still appeared sequentially in the text. The sentences were a mix of questions and statements to avoid a monotony of structure (about 70% of the *different* sentences and 30% of *same* sentences were questions). In half of the *different* sentences, *for* and *the* were separated by a comma (*different*-comma), and in half, there was no comma (*different*-no comma). Importantly, the 10 sentences in each of the above conditions were matched for the words immediately preceding and following the *for the* sequence. Thus, the entire sequence "going for the gold" (see above) appeared in a sentence in each of the three conditions. A fourth set of target sentences also contained the words *for* and *the*, but, in these sentences, *for* and *the* appeared apart (*apart*)—that is, in different parts of the same sentences. However, each one of these sentences was constructed to match sentences from the other three conditions on the word that preceded *for* and the word that followed *the*. Thus, for example, one *apart* sentence contained the separate sequences "going for" and "the gold." Furthermore, in these sentences, the word following *for* and the word preceding *the* were both always content words (e.g., "going *for home*" and "*wants the gold*"). Lastly, the remaining 10 target sentences contained one three-letter *f*-content word (e.g., *fun*), and one three-letter *t*-content word each (e.g., *toe*). All target sentences displayed the target letters only in the target words assigned to those sentences. In addition to the target sentences, there were filler sentences that contained either one, two, or no *f*s or *t*s. The filler sentences, however, did not contain either critical function word. After having constructed the first set of three passages, a second set was derived where commas were now inserted in the original *different*-no-comma sentences and removed from the original *different*-comma sentences. Thus, what

was a no-comma sentence became a comma sentence, and vice versa. In all other respects, the two sets of passages matched.

Each passage began and ended with a filler sentence. Target sentences were alternated with filler sentences in a passage. Furthermore, the various types of target sentences were evenly distributed within each passage and approximately evenly across the three passages. Passages were printed in paragraph form and used normal presentation (lowercase), so as to give the appearance of a coherent passage. No more than one sentence from each grouping of three that had been matched across the three *for the* sets (*same*, *different*-comma, *different*-no-comma) appeared in a passage. Furthermore, *for*, *the*, *f*-content words, and *t*-content words never appeared at the beginning or end of a sentence or line. Moreover, the lines were constructed so that, in the case of the three sets of sentences containing the *for the* sequence, *for*, *the*, the word immediately preceding *for*, and the one immediately following *the* always appeared together on a line (e.g., "going for the gold").

Procedure. Half of the subjects were randomly assigned to the *t*-target group; the others were assigned to the *f*-target group. Within each group, half received the first set of passages and the others received the second set. In order to familiarize subjects with the task, they were first given a practice passage. Following the reading of the practice passage, the subjects read each one of the three experimental passages. The order of reading the experimental passages was counterbalanced across subjects within the *t*- and *f*-target groups. The subjects were told to read for comprehension but simultaneously circle any target letter spotted during reading. They were told not to go back and circle target letters already missed. To encourage the subjects to read for comprehension, they were informed that they would receive comprehension questions on some of the sentences they read. Nine detailed true-false questions were administered following the reading of the last passage. Analysis of true-false performance indicated that the subjects responded correctly 75% of the time, suggesting that they comprehended much of what they read. To break up the reading session, the subjects were asked to solve three three-digit multiplication problems between passages, resulting in a 2- to 3-min interpassage interval. A session lasted about 20 min.

Results

Means and standard errors for the letter-detection errors are summarized in Table 1. The typical missing-letter effect was replicated: Error rates for the function word conditions pooled together yielded 25.9% and 36.2% errors for *f* and *t* in *for* and *the*, respectively. In contrast, the respective error rates for *f* and *t* for content words were 6.7% and 8.0%. A two-way analysis of variance (ANOVA), Word type (content vs. function) × target letter (*f* vs. *t*) for these data yielded a significant effect for

Table 1
Means and Standard Errors of Percentage of Omission Errors for *for* and *the* Under Different Conditions and for *f* and *t* Content Words

| | Target | | | |
| | *f* | | *t* | |
Condition	M	SE	M	SE
Content	6.7	1.6	8.0	1.9
Function (*for* and *the*)				
Apart	33.3	5.6	38.3	6.6
Same	30.3	6.2	38.3	6.6
Different-Comma	12.3	4.0	33.7	6.1
Different-No Comma	27.6	6.0	34.3	6.7

word type $[F(1,58) = 42.84, p < .0001]$, indicating a function-word disadvantage and no other significant effects.

Of more concern, however, was whether the changing role of *for* within the high-frequency frame would alter letter detection in *for* and *the*. Thus, to evaluate this issue, analyses were carried out only on data pertaining to the effects of structural change. In particular, *f* and *t* detection in *for the* in the *same* sentences was compared with that in the two *different* sentence conditions. It appears from Table 1 that the effect of changing syntactic structure (*same* vs. *different*) depended upon whether a comma appeared to separate phrases. Moreover, the advantage in the *different*-comma sentences was confined to *f* detection. Indeed, the error rate for *for* dropped substantially in the *different*-comma condition (12.3%), as compared with the *same* condition (30.3%), whereas the *different*-no-comma condition yielded only a slight advantage (27.6%). The *t* detection in *the*, meanwhile, was unaffected by structure (33.7% and 34.7% for comma and no-comma sentences, respectively, vs. 38.3% for *same* sentences). An ANOVA comparing detection performance for both targets across the three types of sentences produced an effect of condition (sentence type) $[F(2,116) = 10.75, p < .0001]$ and an interaction of condition with target $[F(2,116) = 5.25, p < .01]$. Furthermore, an analysis comparing only the two types of *different* sentences showed an effect of comma presence (whether or not a comma appeared) $[F(1,58) = 9.38, p < .005]$ and a comma presence × target interaction $[F(1,58) = 7.89, p < .01]$, supporting the contention that the advantage for *different* phrase structures was only for *different*-comma sentences and only for letter *f*. This point is reinforced by two two-way ANOVAs. The first, comparing *different*-comma and *same* sentences, yielded a main effect for condition $[F(1,58) = 16.54, p < .001]$, as well as the expected condition × target interaction $[F(1,58) = 5.72, p < .02]$. Meanwhile, the second ANOVA, comparing the *same* and the *different*-no-comma sentences, indicated no significant effects whatsoever.

Finally, we were interested in the contribution of the frame context to the missing-letter effect. From Table 1, it is apparent that neither *f* detection in *for* nor *t* detection in *the* was affected by the familiarity of local context. A two-way ANOVA that compared the two relevant conditions, *same* and *apart*, yielded no significant effects; all effects resulted in $F < 1$.

Discussion

The results made three important points. First, they confirmed that function words engender more detection errors than do content words, replicating previous findings. Second, and of more interest, they indicated that letter-detection performance in function words is intimately tied to the structural organization of the phrase in which they are embedded: Where *for* appeared at the end of a clause, it presumably contributed little to the extraction of a syntactic framework (*different*-comma) (see Kimball, 1973)

and, consequently, letter-detection performance improved. However, and finally, the results also suggested that it was necessary to include a comma to signal the clause boundary in order to achieve the improvement in *f* detection. Of course, *t* detection in *the* was unaffected by structural manipulation as was anticipated. Implicit in this work was the understanding that while structural changes in a sentence may affect the status of *for*, they were unlikely to affect the status of *the*. Importantly, though, the sequence *for the* did not respond to structural changes as a unit.

These findings are, for the most part, consistent with the structural account of the missing-letter effect, and they raise some problems for the unitization approach. In particular, the pattern of errors in function words was sensitive to the syntactic structure and appeared unaffected by the perceptual familiarity of the embedding phrase. In fact, in the present experiment, sentences were constructed so that the only change made to convert a familiar word sequence to a less familiar one involved replacing the companion function word, *for* or *the*, with a content word. However, detection errors in *for* or *the* were no higher when bonded in a familiar pattern (*same*) than when they appeared apart in less familiar sequences (*apart*). The failure to achieve a difference between these conditions argues that the familiarity of a word sequence may not always be critical to letter detection.

The results have implications for the interpretation of context effects on letter detection as well. They suggest that the improved letter detection that occurs as a result of word scrambling is due to the impairment of syntactic structure rather than to the disruption of familiar word sequences. One finding, however, that appears to favor the perceptual familiarity explanation is that without the comma the linguistic status of *for* had little effect upon letter detection. Perhaps this is because the comma, inserted in the space between the frame constituents, served to disrupt the perceptual familiarity of the frame. If such were the case, though, commensurate improvements should have been observed for both *for* and *the*. In fact, if *the* is processed in terms of a perceptual frame that also includes the preceding interword space (see Healy, Conboy, & Drewnowski, 1987), then inserting a comma in that space ought to improve letter detection in *the*. However, the *different*-comma condition did not lead to a reduction in *t* detection errors, and so the comma evidently did not disrupt the unity of the perceptual frame containing *the*.

Alternatively, *for* followed by a comma (i.e., *for,*) might be regarded as being less familiar than *the* preceded by a comma (i.e., *,the*), and thus the unitization position could claim that the current error pattern, across these conditions, derives from the relative familiarity of these two independent frames (i.e., *for,* vs. *,the*). However, it would then also follow that *the* preceded by no comma would have engendered more errors than would *the* preceded by a comma (*,the*), and, of course, that was not the case here. Thus, it would appear that the structural

explanation is generally more consistent with these findings.

Meanwhile, the failure to obtain an effect without a comma inserted (*different*-no-comma) suggests that when punctuation marks were not available, readers tended to assume that *the* was part of the first clause rather than the following clause. Perhaps, in the present experiment, the use of a comma in some of the *different* sentences induced the subjects to rely upon the comma as evidence that *for* and *the* belonged to different clauses. Therefore, the absence of a comma likely led to a misinterpretation of the *different*-no-comma sentences. While commas are rather routine parts of text, we should not underestimate their importance in normal syntactic analysis (Just & Carpenter, 1987). Regardless, the selective reduction in error rates in *for*, and not *the*, is inconsistent with the notion of perceptually unitized multiple-word frames.

EXPERIMENT 2

The purpose of Experiment 2, then, was to test whether a reduction in detection errors is possible when commas are absent, though presumably desirable. Working on the assumption that subjects are particularly sensitive to the presence of a comma, we experimented with a condition in which commas were never used in the *different* sentences to clarify the correct clause division between our target words. Thus, in addition to dividing subjects into *t* and *f* detectors, we also subdivided them into (1) those who read passages in which the *different* clause condition always included a comma between *for* and *the* and (2) those who read a passage in which a comma was never inserted between *for* and *the*. Possibly, where readers cannot count on commas to help syntactic parsing, they may rely more heavily on other syntactic and semantic cues to determine the correct parsing of the syntactically ambiguous sequence. If that is the case, then in this experiment, readers may show some effect of syntactic parsing even when commas are absent.

Method

Subjects. Forty-eight Union College students were paid $1 for participating in this experiment.

Design and Materials. Four passages of 50 unrelated sentences were constructed. Twenty sentences in a passage contained the target phrase *for the*. Of those sentences, half promoted processing *for the* as part of the same clause (*same*), and the remaining sentences promoted processing the two words as part of different clauses (*different*). As in Experiment 1, sentences from each set were matched for the words that preceded *for* and followed *the*. An additional set of 10 sentences contained two short (3- or 4-letter) target content words, one beginning with *f* and the other beginning with *t*. Target sentences contained no *f* or *t* words other than the target words. The remaining 20 sentences were fillers with one, two, or no *f* or *t* words. There were equal numbers of *f* and *t* words across the filler sentences. Words with an *f* or *t* never began a sentence and never began or ended a line. The various types of sentences were evenly distributed across a passage.

The first two passages were identical except for their ordering of the sentences. In these passages, commas never appeared between *for* and *the*, though they did appear elsewhere in the pas-

sages. The third and fourth passages matched the first and second passages, respectively. However, here commas were always inserted between *for* and *the* in the *different* sentences.

Procedure. The subjects were evenly divided into *f* and *t* target groups. Equal numbers of subjects were assigned to read each one of the four passages, so that half of the subjects within each target group saw passages with commas and half saw passages without commas between *for* and *the* in the *different* condition. The experiment took less than 10 min to run. In all other respects, the procedure was the same as that in Experiment 1.

Results

Mean proportion of detection errors and the associated standard errors for all conditions are presented in Table 2. It is apparent that the function words *for* and *the* engendered more detection errors than did the short content words (30% and 4%, respectively). Summing over the three function conditions, we assessed the effect of word type (content vs. function) and target (*f* vs. *t*). The ANOVA on these data yielded $F(1,44) = 50.80, p < .0001$, for word type, and $F(1,44) = 7.14, p < .05$, for the interaction. Of more interest, though, was the effect of structural context (*same* vs. *different*) on letter detection in *for* and *the*. Although *t* detection in *the* was apparently unaffected by condition (39% and 37%, for *same* and *different*, respectively), *f* detection in *for* was sensitive to the embedding context (28% and 15%, respectively). Overall, there also appeared to be fewer detection errors in the comma group than in the no-comma group, and this was true even for a simple comparison of the *different* sentences for these groups [$F(1,22) = 3.70, p < .08$]. Most importantly, however, both groups showed a modest decline in *f* detection errors in *for* (14% vs. 9%, for the comma and no-comma groups, respectively) when *for* and *the* were parts of *different*, as opposed to *same*, clauses.

Several analyses substantiated these conclusions. For these analyses, we ignored the content-word data. First, a three-way ANOVA, condition (*same* vs. *different*, within subjects) × comma (present vs. absent, between subjects) × target (*f* vs. *t*, within subjects), yielded sig-

Table 2
Means and Standard Errors of Percentages of Omission Errors for *for* and *the* Under Various Contexts and for *f* and *t* Content Words

| | Target | | | |
| | *f* | | *t* | |
Condition	M	SE	M	SE
Comma Present				
Content	4.2	2.6	2.5	1.3
Function				
Same	22.5	7.0	40.8	9.0
Different	8.3	4.6	37.5	8.7
Comma Absent				
Content	5.8	2.6	0.8	0.8
Function				
Same	32.5	8.3	37.5	8.4
Different	23.3	6.3	37.5	8.6

nificant effects for target [$F(1,44) = 5.14, p < .03$], condition [$F(1,44) = 7.54, p < .01$], and the interaction [$F(1,44) = 4.24, p < .05$], supporting the observation that only *f* detection changed with condition. Furthermore, a separate analysis conducted on the *f* detection data in *for* yielded a significant effect of condition [$F(1,22) = 9.00, p < .01$], but unlike in Experiment 1, here there was no hint of an interaction with comma ($F < 1$). Indeed, independent ANOVAs for the comma and no-comma groups, comparing the *same* and *different* sentences, yielded $F(1,11) = 5.42, p < .05$, and $F(1,11) = 3.59, p < .085$, respectively.

Discussion

The critical issue here was to demonstrate that letter detection in *for*, in particular, is altered by structural changes in our sentences. The results of Experiment 2 replicated most of the important findings of Experiment 1. However, here it was also found that *for* appearing in a *different* context may reveal its *f* more readily, though a comma is not present to mark the clause boundary. While there was evidence that subjects do rely on punctuation to mark the boundaries between paragraphs (Koen, Becker, & Young, 1969), the present results suggest that readers can still determine clause boundaries even without informative punctuation, provided they do not expect commas to cue syntactic parsing. Presumably, subjects rely on a variety of cues to determine syntactic parsing, and the present results suggest that the utilization of different cues may be under strategic control. Elsewhere, Aaronson and Ferres (1983a) had indicated that the weighing given to structural attributes of text in reading varied with the memory requirements of a reading task. Under the present, presumably demanding, conditions (i.e., comprehension questions were used), the subjects might have been intent on uncovering all possible available cues to structure.

In our previous work (Koriat & Greenberg, 1991), we discussed several potential sources of information that might be used to extract phrase structure and, hence, contribute to the missing-letter effect. Which of these cues were used by the subjects in the present no-comma group is not clear. However, it is evident from our previous work (Greenberg & Koriat, 1991; Koriat et al., 1991) that significant parsing of context does take place before the occurrence of the missing-letter effect. In particular, Hebrew homographs concealed their initial letters when those letters were functor prefixes but revealed them when they were simply part of the stem (Koriat et al., 1991). Furthermore, Greenberg and Koriat (1991) found that detection of *f* in *for* varied with the context that followed *for* and disambiguated its structural status. The present findings, then, are consistent with the structural account, and conflict with the unitization account, which attributes the effect of context to phrase familiarity. In the no-comma group, in particular, the *for the* frame remained perceptually intact across the *same* and *different* conditions, and yet detection of *f* in *for* appeared to have been modestly

affected by changes in the underlying syntactic structure of the phrase. It may still be true, though, that perceptual frames are unitized and that disruption of those frames can improve letter detection, particularly when "foreign" graphic cues, such as asterisks, are introduced (Healy et al., 1987).

Finally, the results may also be seen as consistent with pause patterns associated with reading. Aaronson and Ferres (1983b) have found readers pausing more at the end of a syntactic unit than at the beginning. Given that many of the sentences in the present *different* condition were questions, this may have accentuated such pausing. Greenberg and Koriat (1991) speculated that structural features may work in tandem with a rhythm in silent reading to produce patterns of letter-detection errors. Such a rhythm may be akin to the prosodic rhythm associated with speech, which is generally characterized by a systematic deemphasis of functors (see Gee & Grosjean, 1983). It is not clear, at present, whether a rhythm during silent reading plays a part in the missing-letter effect, and if indeed it does, whether its contribution to the missing-letter effect is independent of syntactic structure. The sum of the evidence to this point, though, suggests that, at the very least, the role of a rhythm would be intimately tied to the on-line extraction of structure (see Greenberg & Koriat, 1991).[1]

GENERAL DISCUSSION

The present study was motivated by the notion that words may be read in terms of units larger than the single word, provided that the units represent highly familiar frames. The best examples of such frames are function-word sequences such as *for the*, which are among the most familiar two-word sequences in English (Umeda & Kahn, 1982). According to Drewnowski and Healy (1977, 1980), the missing-letter effect can reveal the size of the unit that is used in text processing. Assuming that text processing occurs at the level of familiar multiple-word patterns, then destroying such patterns ought to lead to improved letter detection. In addition, since unitization depends solely upon the familiarity of the sequence, so long as the frame is kept, manipulations of more subtle structural factors within the frame should have little effect upon letter detection. Neither of these predictions was supported by the results of the present study. In contrast, the letter-detection patterns were reflective of the linguistic structure of the sentence in which the function words appeared and did not vary with the locations of *for* and *the*—that is, regardless of whether they appeared together in a familiar sequence or apart. Thus, the present findings reinforce the claims of the structural position regarding the missing-letter effect by extending our earlier work (e.g., Greenberg & Koriat, 1991; Koriat, et al., 1991) to show that structure is crucial even when words appear in highly familiar sequences.

It should also be apparent that all functors are not processed equally. As described earlier, there are a variety of ways in which *for* and *the* differ. At present, we are not entirely clear as to which aspect(s) of these words is most salient to structural extraction. In some recent work, though, Koriat and Greenberg (in press), found that while manipulating a preposition's location in a sequence of several functors in a sentence dramatically affected its letter detection, the same manipulation had no effect upon letter detection in *the*. Thus, we speculate that the definite article is probably accessed earlier than are other functors, because it appears that *the* is more narrowly defined than are other functors. As suggested earlier, *the* has potentially less syntactic roles to play and less opportunity to appear in a position other than one that heads a phrase. Hence, *the* is likely to be a more reliable indicator of syntactic organization.

The present findings also pose problems for yet another accounting of the missing-letter effect—namely, the attentional redundancy approach (see Corcoran, 1966; Schindler, 1978). According to this position, the missing-letter effect for function words occurs because textual redundancy permits readers to skip highly predictable orthographic units. Presumably, then, the attentional redundancy position would argue that *for* in the *different* sentences, where errors were lower, would be less predictable. However, a test of our sentences with the cloze procedure did not support that contention. Fourteen subjects were given a mixture of 25 *same* and 15 *different* sentences sampled from both of the present experiments, along with 15 noncritical sentences. Each critical sentence was truncated after the word preceding *for*; each noncritical sentence was cut after a word that preceded another preposition. (Noncritical sentences were included to encourage the subjects to think of prepositions other than *for*.) The presentation sequence of the various types of sentences was random. The subjects were charged with determining the next word in each of the sentences. For Experiment 1 sentences, the subjects selected *for* for 50.0% of *same* sentences and 48.6% of *different* sentences. Meanwhile, for Experiment 2, *for* was chosen in 25% of *same* sentences and 38% of *different* sentences [$F(1,13) = 5.51, p < .05$]. Clearly, these patterns are at odds with the redundancy explanation. Letter detection was better in *different* sentences, though *for* was certainly no less (and perhaps more) predictable in those sentences.

Finally, the present work points out a potentially important role for punctuation in the missing-letter effect. Koriat and Greenberg (1991) postulated that function words are particularly informative regarding sentential structure (see also, Bock, 1990; Garrett, 1980; Kimball, 1973). Hence, readers automatically monitor text for function words in order to construct frames around them. Others have also stressed the role of punctuation marks in conveying the structure of a sentence (e.g., see Just & Carpenter, 1987). Punctuation may be particularly useful in handling ambiguous or complex structural patterns on-line. A comma or period, according to Just and Carpenter (1987), may indicate to the reader when to "wrap up" syntactic analysis of one clause, and when the word

downstream begins a new syntactic grouping. From the present results, it would appear that readers are indeed sensitive to the utility of certain graphic cues, such as commas. Exactly how and when these pieces of information enter the process and precisely how they interact with other structural cues, such as function words, is not yet clear. Undoubtedly, though, the interaction between different cues in conveying structure warrants further investigation.

The difference between the results of the present experiments discloses another important property regarding the extraction of phrase structure. In Experiment 1, the advantage of *different* sentences occurred only when commas were used to cue phrase boundaries, whereas in Experiment 2, this advantage generalized to a no-comma condition when the presence of a comma in *for the* frames was manipulated between subjects. It would appear that readers are responding to such cues in the text. Hence, while not conclusive, the present work suggests that readers can exercise some control over the cues they select to determine phrase structure.

In sum, the present study draws us deeper into the kinds of complex structural cues that help shape a reader's processing. Clearly, the results send further signals suggesting that how words are incorporated into the syntactic structure is critical to the missing-letter effect and indicate that, in understanding this effect, we are also gaining some insight into how readers process text on-line.

REFERENCES

AARONSON, D., & FERRES, S. (1983a). Lexical categories and reading tasks. *Journal of Experimental Psychology: Human Perception & Performance*, 9, 675-699.

AARONSON, D., & FERRES, S. (1983b). A structure and meaning based classification of lexical categories. In S. White, B. Kachunk, & F. Podwall (Eds.), *Reading and reading disability: Scanning the 80's from left to right*. New York: New York Academy of Sciences.

BOCK, K. (1990). Structure in language: Creating form in talk. *American Psychologist*, 45, 1221-1236.

CORCORAN, D. W. J. (1966). An acoustic factor in letter cancellation. *Nature*, 210, 658.

DREWNOWSKI, A., & HEALY, A. F. (1977). Detection errors on *the* and *and*: Evidence for reading units larger than the word. *Memory & Cognition*, 5, 636-647.

DREWNOWSKI, A., & HEALY, A. F. (1980). Missing *-ing* in reading: Letter detection errors on word endings. *Journal of Verbal Learning & Verbal Behavior*, 19, 247-262.

DREWNOWSKI, A., & HEALY, A. F. (1982). Phonetic factors in letter detection: A reevaluation. *Memory & Cognition*, 10, 145-154.

GARRETT, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177-220). San Diego, CA: Academic Press.

GEE, J. P., & GROSJEAN, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.

GREENBERG, S. N., & KORIAT, A. (1991). The missing-letter effect for common function words depends on their linguistic function in the phrase. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 17, 1049-1059.

HEALY, A. F. (1976). Detection errors on the word THE: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception & Performance*, 2, 235-242.

HEALY, A. F. (1980). Proofreading errors on the word THE: New evi-

dence on reading units. *Journal of Experimental Psychology: Human Perception & Performance*, 6, 45-57.

HEALY, A. F., CONBOY, G. L., & DREWNOWSKI, A. (1987). Characterizing the processing units of reading: Effects of intra- and interword spaces in a letter detection task. In B. K. Britton & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 279-296). Hillsdale, NJ: Erlbaum.

HEALY, A. F., FENDRICH, D. W., & PROCTOR, J. D. (1990). Acquisition and retention of a letter-detection skill. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 270-281.

HEALY, A. F., OLIVER, W. L., & McNAMARA, T. P. (1987). Detecting letters in continuous text: Effects of display size. *Journal of Experimental Psychology: Human Perception & Performance*, 13, 279-290.

JUST, M. A., & CARPENTER, P. A. (1987). *The psychology of reading and language comprehension*. Newton: Allyn and Bacon.

KIMBALL, J. P. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15-47.

KOEN, F., BECKER, A. L., & YOUNG, R. E. (1969). The psychological reality of the paragraph. *Journal of Verbal Learning & Verbal Behavior*, 8, 49-53.

KORIAT, A., & GREENBERG, S. N. (1991). Syntactic control of letter detection: Evidence from English and Hebrew nonwords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 17, 1033-1048.

KORIAT, A., & GREENBERG, S. N. (in press). Prominence of leading functors in function morpheme sequences as evidenced by letter detection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

KORIAT, A., GREENBERG, S. N., & GOLDSHMID, Y. (1991). The missing-letter effect in Hebrew: Word frequency or word function? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 17, 66-80.

PROCTOR, H. D., & HEALY, A. F. (1985). A secondary task analysis of a word familiarity effect. *Journal of Experimental Psychology: Human Perception & Performance*, 3, 286-303.

READ, J. D. (1983). Detection of Fs in a single statement: The role of phonetic recoding. *Memory & Cognition*, 11, 390-399.

SCHINDLER, R. M. (1978). The effect of prose context on visual search for letters. *Memory & Cognition*, 6, 124-130.

UMEDA, N., & KAHN, D. (1982). Frequency of occurrence of two- and three-word sequences in English. *Journal of the Acoustical Society of America*, 72, 2031-2033.

NOTE

1. In addition to a functor's location within a phrase, the awkwardness of the structure might also affect rhythm and, perhaps, letter detection. Indeed, *different* sentences were judged as more awkward than *same* sentences. Twelve subjects were asked to rate sentences in both experiments for awkwardness on a 5-point scale. For Experiment 1, *same* sentences were rated as less awkward than *different*-comma sentences, which in turn were rated as less awkward than *different*-no-comma sentences [$F(1,11) = 41.25$, and $F(1,11) = 26.13$, respectively, $p < .001$, in each case]. In Experiment 2, *same* and *different*-comma sentences were rated as not different ($F < 1$), but both *same* and *different*-comma sentences were rated as less awkward than *different*-no-comma sentences [$F(1,12) = 9.73, p < .01$, and $F(1,12) = 7.26, p < .02$, respectively]. While these findings showed that *different* sentences, particularly without commas, were more difficult than were *same* sentences, the pattern of ratings was not reflective of the overall pattern of letter detection. In Experiment 1—where the *different*-no-comma sentences were deemed more awkward than *same* sentences and less awkward than *different*-no-comma sentences—they engendered the least number of *f* detection errors. In Experiment 2, *different*-comma sentences yielded less errors than did *same* sentences, though the two types of sentences did not differ in rated awkwardness. Thus, it would appear that differences in judged awkwardness cannot account for the observed letter-detection patterns.