ON THE ROLE OF SENTENCE STRESS IN SENTENCE PROCESSING*

ANNE CUTLER AND DONALD J. FOSS
University of Texas

Words bearing high stress appear to be easier to process during sentence comprehension. Since sentence stress typically falls on content words this suggests that comprehension is organized according to a form class bias: process stressed items as content words. The present study measured reaction-time (RT) to word-initial phoneme targets on content and function words in sentence contexts. Half of the words of each type were stressed, half were not. In addition, a variable of "normality" of stress pattern was manipulated. It was found that RTs were shorter for stressed items independent of their syntactic function. No effect for content *v.* function words or normal *v.* non-normal stress pattern was observed. Results were interpreted within the framework of a predictive model utilizing the concept of semantic focus.

When a word receives high stress, it typically has higher pitch, its syllables are lengthened, it is somewhat louder, and, importantly, its vowels occur closer to their citation form (Lehiste, 1970; Tiffany, 1959). Unstressed words, on the other hand, may be contracted in duration, and, in particular, their vowels are likely to be reduced to /ə/. It seems reasonable to assume that words which receive high stress should be easier to comprehend than words which receive lower stress. Indeed, Lieberman (1963) has shown that this is true of isolated words excised from a sentence context. It appears likely that, in the course of sentence comprehension, the percept that arises from a stressed word is "purer" or "clearer" than the percept that arises from an unstressed word. This in turn should render the stressed word easier to process.

Unpublished results from our laboratory tend to support the above assumption. Subjects were asked to listen to lists of unrelated sentences and to comprehend them; in addition, they were asked to press a button whenever they heard a word that began with a specified target phoneme. It has been argued (Foss, 1969, 1970) that time to respond in this phoneme monitoring task is directly related to the difficulty of processing the sentence at the time when the target phoneme occurs. We found that the reaction-time (RT) to the target phoneme was shorter if the word carrying it was a content word (e.g., noun or main verb) than if it was a function word (e.g., preposition or conjunction). Content words typically receive higher stress than do function words; thus the conclusion that high stress words lead to easier processing, while not following directly from these data, does seem plausible. Since high stress words tend to be, as we have pointed out, perceptually purer, the shorter RTs in this phoneme monitoring task may reasonably be ascribed to heightened perceptual clarity.

However, some evidence exists which suggests that this "perceptual clarity" hypothesis may be an over-simplified account. Consider a recent experiment conducted by Shields, McHugh and Martin (1974). They presented sentences which included fragments like those in (1) and (2) to subjects and asked them to perform the phoneme monitoring task (in this case the target phoneme is /b/).

(1) ... the plane to BENkik leaves at noon.

(2) ... the plane to benKIK leaves at noon.

Shields *et al.* found that RT to phoneme targets on nonsense words was shorter when the target-bearing syllable was accented (BENkik) than when it was not (benKIK). In all of the sentences used in their experiment the target phoneme was carried by a nonsense word standing in place of a proper noun.

At first glance it would appear that the above result could be accounted for by a perceptual clarity hypothesis. However, Shields *et al.* included an important control which casts this hypothesis into doubt. They excised the nonsense words from the sentences and presented them to subjects for phoneme monitoring embedded in a string of other nonsense words. The perceptual clarity hypothesis predicts that RT should still be faster when the target is carried by the accented syllable than when it is not. However, no RT difference was found between nonsense words stressed on the first syllable and those stressed on the second syllable. Since the stimuli used in the control task were acoustically identical to those occurring in the experimental sentences, the control argues against the perceptual clarity interpretation of the results in the experimental condition.

In interpreting their results Shields *et al.* made use of a model devised by Martin (1972) which attempts to account for the temporal organization of a sentence. Martin proposed rules for determining a sentence's temporal organization; Shields *et al.* supposed that listeners were, in some sense, employing these rules to predict the location of upcoming accents. Further, they argued that predicting the location of upcoming accents plays an important role in sentence comprehension in that it permits a processing mechanism to direct attention to the "potentially most important elements" of an utterance.[1] Thus, they account for the shorter RT to accented target syllables by appealing to an attention directing mechanism rather than to a perceptual clarity hypothesis. The attention directing mechanism is itself guided by the temporal organization of the sentence.

As we have already noted, it is generally the case that accented items are content words. Content words typically carry high information. Indeed, Berry (1953) found that the amount of stress on a word was proportional to its information content; thus, the higher amounts of stress fell on the words carrying the greatest amount of inform-

---

[1] *While Shields et al. used the neutral phrase "most important elements", they took the elements to be syllables, the basic units in Martin's stress rules. We have changed the elements into words in much of the following. We assume that when a syllable of a word receives primary stress the entire word occupies a privileged (or at least a different) position vis-à-vis the comprehension mechanism as compared to a word that does not receive sentence stress. A complete discussion of the implications of this assumption could lead us off the track, and we will not pursue it here. Cf. Savin and Bever (1970), Foss and Swinney (1973) for discussion of related matters.*

ation in a sentence (generally the nouns and verbs). Therefore we might hypothesize that the *value* of predicting where stress will fall is that it will typically occur on high information (=content) words. Hence, part of the advantage that accrues to content words over function words in the phoneme monitoring task may be due to an attempt ·on the part of the processing mechanism to predict where the content words will fall and to more efficiently process them when they occur in the expected location. For example, procedures of lexical look-up may be biased to search for content words at the place where high stress is expected to occur. Statistically this should be a successful strategy to adopt. Although sentence stress can and does fall on items other than content words, the latter do carry stress in the majority of cases.

The above line of reasoning suggests that accented function words will not reap the same advantage as will accented content words. We have assumed that the processing mechanism is biased to interpret accented words as content words (this is the value of making the prediction about where upcoming accents will fall). This bias may, then, actually lead to a decrease in processing efficiency when the accent falls on a function word (the latter prediction only follows if there is absolutely nothing to the perceptual clarity hypothesis).

We might say that the present line of reasoning constitutes a "form class" hypothesis. The components of this hypothesis are: (1) the sentence processing mechanism makes use of early accents in a sentence to predict later ones; (2) the items that are predicted to be accented are also predicted to be content words; (3) content words in accented location will be processed faster than content words in unaccented location, but (4) no such advantage (and perhaps even a disadvantage) will accrue to accented function words over unaccented function words.

In order to test the "form class" hypothesis we can present sentences of four types to subjects in a phoneme monitoring task. The sentence types are defined by having the target phoneme occur on either a content or a function word and by having the word carrying the target phoneme be stressed or not. The form class hypothesis predicts that RTs will be shorter to stressed content words than to unstressed content words, and that no difference (or perhaps a difference in the opposite direction) should occur between the stressed and unstressed function words. The present study was designed to test the form class hypothesis and thus made use of the above four sentence types.

It may be objected, however, that "normal" intonation places the main accents of the sentence on the content words and that sentences in which function words are accented will sound somehow "unnatural". In fact, while stress typically falls on content words, function words can bear primary stress in quite "normal" sentences: "He looked up the book," versus "He looked it up." For this experiment it was necessary to find sentences which can be spoken in two ways, with and without stress on the target-bearing word, each version sounding "natural" whether the target began a content or a function word. The task proved simple. Each sentence was constructed with a particular stress pattern in mind (half of the sentences accented the word bearing the target, half did not). Another intonation pattern with the opposite accent condition for the target was then constructed for each sentence. For example, the sentence "I'm not sure that Shakespeare's plays are even BY Shakespeare," (in which the target is /b/) was originally

constructed with a contour in which the target word "by" was stressed. The alternative stress pattern, without stress on "by", was: "I'm not sure that Shakespeare's plays are even by SHAKEspeare." If a sentence could not be spoken in a way reversing the accent condition of the targeted word, it was abandoned in favour of another which could.

It is worth mentioning here that this addition to the design opens up the possibility of determining whether the notion of "normal" sentence stress pattern is in fact valid. Although such a notion has intuitive appeal, its usefulness as a concept relevant to actual speech behaviour (as opposed, for example, to its use in citations) has recently been questioned (Schmerling, 1974). The sentences used in this experiment all have a "normal" stress pattern, i.e., the contour with which they were first conceived, as well as another — "non-normal"—contour. Thus, if there is such a thing as "normal" stress pattern it might be expected to exercise an influence in this task. If there is no difference between the "normal" and "non-normal" versions of sentences, then the view that normal stress pattern is a relevant concept for actual speech would not be supported.

METHOD

Materials

Twelve content words (nouns or verbs) and ten function words (prepositions or conjunctions) were chosen as the target-bearing items. The words in these two classes were matched for frequency. The target phoneme was varied across words; the phonemes /b, s, f, k, p, m, n, t/ were used as targets. Two basic or "normal" sentences were constructed for each of the 22 target words. In one, the target received the main stress of the sentence; in the other, the target was unstressed, main stress falling on a nearby word. Then two derived or "non-normal" sentences were constructed by changing the contours of the two basic sentences. The sentence that normally had main stress on the target was, in the derived or "non-normal" version, spoken so that main stress fell on another word. The sentence that normally had main stress on another word was, in the "non-normal" version, spoken so that stress fell on the target.

Two tapes were constructed, each containing all 44 sentences plus ten filler sentences. The latter did not contain the specified target phoneme. Each experimental sentence appeared in its target-stressed version on one tape and in the target-unstressed version on the other. On each tape half of the sentences bore stress on the target item and half did not; likewise, half of the sentences on each tape occurred in their "normal" version and half in their "non-normal" version. (In fact, however, the counterbalancing was inadvertently imperfect. There were equal numbers of "normal" and "non-normal" contours on each tape, equal numbers of stressed and unstressed targets, and equal numbers of content and function word targets, but the normality variable was unbalanced within stress by grammatical function.)

Thus, stress level and grammatical function of the target-bearing item, as well as the

"normal" *v.* "non-normal" comparison, were within-subject variables, with material sets forming the only between-subject variable.

Subjects

Subjects in the initial experiment were 40 undergraduates at the University of Texas who participated as part of a course requirement. The 36 subjects in the replication were obtained from the same source.

Procedure

Subjects were tested in groups of up to four at a time. They were told that they were participating in an experiment on sentence comprehension, and they were instructed to pay careful attention to the content of each sentence since a comprehension test would be given at the conclusion of the experiment. In addition, they were asked to push a button whenever a word in a sentence began with the target phoneme specified for that sentence. The targets varied across sentences, and the particular target was given immediately prior to the presentation of each sentence. The sentences were spoken by a male speaker of standard American. Each sentence was preceded by the word "ready", the target specification, and about two seconds of silence. Three practice items were presented before the experimental set. The sentences were presented binaurally over headphones. A timer was automatically started when the target occurred and was stopped when the subject pressed the button. Timing and data collection were under control of a PDP 8/I computer.

After the experiment the subjects were given a written comprehension test which consisted of 20 sentences, some of which had occurred in the experiment and some of which, while constructed of similar vocabulary to the experimental items, had not occurred. Subjects were required to judge for each sentence whether they had heard it or not.

RESULTS

Performance on the comprehension test maintained a high level (overall mean: 80% correct), indicating that comprehension of the sentences was in general good.

The mean RT for each subject was computed for each of the experimental conditions. (RTs which were longer than 1500 msec. or shorter than 100 msec. were omitted from the calculation of means; it was felt that particularly long RTs might involve a re-processing of the sentences, whereas particularly short ones might have resulted from anticipations or guessing.) The overall means for the comparisons of interest are pre-sented for both the original study and the replication in Table 1.

In performing an analysis of variance on these results, both Subjects and Words within the Content Word-Function Word comparison should be treated as random factors. As Clark (1973) has pointed out, significant effects obtained for such com-

TABLE 1

Mean reaction times (msec.) for the two experiments

Word type	ORIGINAL STUDY (N=40) Target Stress		REPLICATION (N=36) Target Stress	
	Stressed	Unstressed	Stressed	Unstressed
Content	390	432	411	434
Function	394	462	399	483

parisons as the present one between content words and function words may not be generalizable beyond the particular words used in the experiment if the analysis is not carried out in this way. Accordingly, two separate analyses were performed on the results of both the original experiment and the replication. In one analysis, the Subjects variable was treated as a random factor and the means were collapsed across Words within Content and Function; while in the other analysis, Words was treated as a random factor and the means were collapsed across Subjects and Materials Sets. Since the unequal numbers of content and function words meant that this latter analysis had unequal Ns, an unweighted-mean analysis was performed (Winer, 1971). The combined results of these analyses allowed the computation of *min F'*.

The main effect for Stress was significant in both administrations of the experiment: *min F'* (1,32) $=$ 9.15, p <0.01 for the original experiment, *min F'* (1,36) $=$ 6.08, p <0.025 for the replication. The effects for Content v. Function and for Normal v. Non-Normal Stress Pattern, as well as all interactions, were not significant.

In the analysis by Subjects, the additional between-subjects variable of Material Sets was included. This effect was not significant in either administration, but it did enter into two significant interactions which replicated, namely Materials Sets x Normal v. Non-Normal intonation [$F(1,38)$ $=$ 63.88, p <0.001, and $F(1,34)$ $=$ 27.20, p <0.001, in the original experiment and replication, respectively]; and Materials Sets x Content v. Function x Stress level [$F(1,38)$ $=$ 8.58, p <0.01, and $F(1,34)$ $=$ 12.22, p <0.01]. These interactions will be discussed below. One further interaction with Materials Sets appeared in one study but not in the other, and does not seem to be reliable.

In summary, then, the difference between RTs to targets on content words versus those on function words is not significant when stress is held constant. In contrast, the effect of stress is highly significant. The effect of normality of stress pattern, as it has been defined in this study, is not significant. These variables did not interact with each other.

DISCUSSION

First, these results belie the notion that the role of suprasegmental cues in sentence processing is to indicate the form class of the stressed items. Since no RT difference was found between content and function words when stress was controlled, it seems that it is the stress assigned to most content words, rather than the fact that they are content

words, which has caused the RT difference found in earlier phoneme-monitoring studies.

Second, the failure of the normal—non-normal contrast to reach significance speaks against the viability of the concept "normality" as applied to stress pattern in actual speech situations. The "normal" pattern was defined here as the contour with which the sentence was originally conceived; in each case it was also the contour which the sentence would receive in citation form (the traditional criterion for "normality" of a stress pattern). The "non-normal" pattern was defined as the contour which was constructed as a variation on the original pattern. Since no difference was found between the two conditions, we assume that the sentence processing mechanism does not expect a particular syntactic form to be associated with the particular stress contour of the incoming sentence.

Before turning to further discussion of the role of stress in the processing of a sentence, a brief explanation is in order for the fact that two unpredicted interactions showed up in both the original experiment and the replication. These were, as stated above, the interaction of Material Sets with Word Class (content, function) and Stress level, and the interaction of Materials with the Normal—Non-normal contrast. Both owe their existence to an effect which typically occurs in the phoneme-monitoring task, namely that RT is shorter to targets occurring later in the sentence than to earlier targets (Foss, 1969). This effect was claimed by Shields *et al.* (who manipulated it in their experiment and found it to be highly significant) to be due to the fact that later targets allow the subjects to "lock in" to the rhythmic pattern more securely and therefore to be more certain of upcoming accent locations. It is not clear that this explanation is sufficient to account entirely for the phenomenon, nor did Shields *et al.* report whether the effect was absent from their nonsense word control sequences, which would be predicted by their explanation; however, the effect is quite reliable. In this experiment, while target position was loosely controlled in that approximately half of the sentences in each category had early occurring and half late-occurring targets, it was not an experimental variable and was not counterbalanced across materials sets. This allowed an unfortunate imbalance to occur between the two sets of materials on two of the experimental variables, namely Stress level and Normality. Specifically, it was found on later inspection that Tape I had more late- than early-occurring targets in the "non-normal" cell, while on Tape II this situation prevailed in the "normal" cell. Again, Tape I had more late- than early-occurring targets in the category unstressed content words, whereas this situation obtained for stressed content words on Tape II. (All other cells on each tape were balanced with respect to position of the target.) It could be predicted from the former situation that "normal" intonation would produce faster RT than "non-normal" on Tape II, whereas the reverse would be the case on Tape I. This was, in fact, exactly the nature of the Materials x Normality interaction (see Table 2).

From the second imbalance it could be predicted that unstressed content words might be faster than stressed on Tape I, while the reverse would be the case on Tape II; function words, which were balanced for early and late targets, would show no such imbalance but should show the stress effect in all cells. Again, this was exactly the

TABLE 2

Mean reaction times (msec.) for Material Sets
and Normal—Non-normal Intonation

	ORIGINAL STUDY Materials		REPLICATION Materials	
Stress	*Tape I*	*Tape II*	*Tape I*	*Tape II*
Normal	469	381	426	436
Non-normal	399	429	386	480

nature of the interaction of materials with word class and stress. The magnitude of the target position effect argues strongly for rigorous control of this factor in future phoneme monitoring experiments.

The results of the present experiment support the conclusion that sentence stress plays an important role in sentence comprehension. Further, we can conclude that the role of suprasegmental cues is not merely to indicate the syntactic function of words in a sentence; the "form class" hypothesis is not sufficient to explain the advantage for stressed words in phoneme monitoring RT. Also, it seems unlikely that the difference in RT between stressed and unstressed words can be explained merely in terms of differences in perceptual clarity, since Shields *et al.* found that the RT difference disappeared when the target items were presented without sentence context.

Of course, Shields *et al.*'s control sequences were strings of nonsense words, and it may well be that processing such a string is a task so foreign to subjects that the intelligibility advantage of the stressed syllables is simply masked by an overall rise in the level of difficulty. Thus, their control does not provide strong evidence on which to base the conclusion that the RT advantage of stressed words during sentence processing reflects the prediction of upcoming accent locations.

However, there is other evidence which indicates that such prediction does take place. Consider a phoneme-monitoring experiment in which RT is compared to targets on the same (acoustically identical) word in the same sentence spoken with two differing suprasegmental contours, one which predicted that the target-bearing item should bear high stress, the other which predicted it should bear lower stress. The perceptual clarity explanation would predict no difference between RT to the two conditions in this experiment, whereas the "prediction" hypothesis would lead us to expect that the predicted-high stress targets would elicit faster RTs than the predicted-low stress targets. In fact, Cutler (1975) has demonstrated that the latter result holds, thus providing strong support for the notion that the prediction of upcoming accents is a part of sentence processing.

We agree, then, with Shields *et al.* that the role of suprasegmental cues is to direct attention to the main sentence accents, and we accept the premises that these accents fall on the "potentially most important elements". Both Shields *et al.* and the Cutler study just mentioned used only content words as target-bearing items; however, the present study has shown that importance is not determined by the form class of the items. We will now offer some speculations concerning what the determiners are.

It is not a very big leap to suggest that the functional importance of an item is more concerned with the semantics and pragmatics of the sentence than it is with its syntax. Variations in the suprasegmental contour of an utterance commonly indicate variations in the semantics underlying the utterance. By stressing different words in the same sentence it is possible to convey subtly different messages. For example, sentence (3) is not equivalent to sentence (4).

(3)   JOHN eats caviare for breakfast.
(4)   John eats CAVIARE for breakfast.

In (3) "John" is said to be the focus of the sentence, while in (4) the focus is "caviare." Such differences in sentence focus are assumed to be accompanied by differences in the presuppositions carried by the sentences. For example, in (3) the speaker is presupposing that someone eats caviare for breakfast (he is assuming, that is, that his audience is aware that someone eats caviare for breakfast) and he is asserting that that someone is John. If the speaker places primary stress on "caviare" as in (4), then he is presupposing that John eats something for breakfast and asserting that what John eats is caviare.

The intimate relation between placement of stress within a sentence and the semantic interpretation of the sentence has been discussed extensively by linguists (e.g., Halliday, 1967; Jackendoff, 1972). We wish to advance the claim that prediction of sentence stress locations, an integral component of sentence understanding, allows the sentence processing mechanism to anticipate and direct particular attention to the elements in the sentence which will be focused. It is worthy of note that recent work by Allen and O'Shaughnessy (forthcoming) has shown that reliable correlates of semantic focus assignment are to be found in the fundamental frequency contour. Thus the acoustic prerequisites of this view would appear to be satisfied.

The present data do not, of course, provide direct evidence for this claim. However, they are certainly compatible with it; and by ruling out the "form class" interpretation of the stress effect we have made the semantic or focus interpretation more viable.

REFERENCES

ALLEN, J. and O'SHAUGHNESSY, D. (forthcoming). The effect of focus-shifting transformations on fundamental frequency contours.

BERRY, J. (1953). Some statistical aspects of conversational speech. In W. Jackson (Ed.) *Communication Theory* (New York), 392.

CLARK, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *J. verb. Learn. verb. Behav.* 12, 335.

CUTLER, A. (1975). *Sentence Stress and Sentence Comprehension,* Unpublished Ph.D. dissertation, University of Texas.

FOSS, D. J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty upon decision times. *J. verb. Learn. verb. Behav.,* 8, 457.

FOSS, D. J. (1970). Some effects of ambiguity upon sentence comprehension. *J. verb. Learn. verb. Behav.* 9, 699.

FOSS, D. J. and SWINNEY, D. A. (1973). On the psychological reality of the phoneme: Perception, identification and consciousness. *J. verb. Learn. verb. Behav.,* 12, 246

HALLIDAY, M. A. K. (1967). Notes on transitivity and theme in English. Part II. *J. Ling* 3, 199.

JACKENDOFF, R. A. (1972). *Semantic Interpretation in Generative Grammar* (Cambridge, Mass.).

LEHISTE, I. (1970). *Suprasegmentals* (Cambridge, Mass.).

LIEBERMAN, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Lang. Sp.*, 6, 172.

MARTIN, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychol. Rev.*, 79, 487.

SAVIN, H. and BEVER, T. G. (1970). The nonperceptual reality of the phoneme. *J. verb. Learn. verb. Behav.* 9, 295.

SCHMERLING, S. F. (1974). A re-examination of "Normal stress." *Language*, 50, 66.

SHIELDS, J. L., McHUGH, A., and MARTIN, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *J. Psychol.*, 102, 250.

TIFFANY, W. R. (1959). Non-random sources of variation in vowel quality. *J. Sp. Hrg. Res.*, 2, 305.

WINER, B. J. (1971). *Statistical Principles in Experimental Design* (New York).