JOURNAL OF MEMORY AND LANGUAGE **34**, 709–738 (1995)

The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech

JEAN E. FOX TREE

University of California, Santa Cruz

Speech disfluencies have different effects on comprehension depending on the type and placement of disfluency. Words following false starts (such as *windmill* after *in the* in *the eleventh example is um in the a windmill*) have longer word monitoring latencies than the same tokens with the false starts excised. The decremental effect seems to be limited to false starts that occur in the middle of sentences or after discourse markers. I suggest it is at these points that the repair process is most burdened by the false start. In contrast, words following repetitions (*heart* in *of a of a heart*) do not have longer word monitoring latencies than the same tokens with the repetitions excised. In two experiments, words following spontaneously produced repetitions have faster word monitoring latencies. Two other experiments suggest that this seeming repetition advantage is more likely the result of slowed monitoring after a phonological phrase disruption. Inserting repetitions where they did not occur in a manner that preserved the original phonological phrases resulted in neither an advantage nor a disadvantage or repeating. These studies provide a first glimpse at how speech disfluencies affect understanding, and also provide information about the types of comprehension models that can accommodate the effects of speech disfluencies.

The speech of public speakers, actors, and others who make their livings from spoken words is fluent and clear. Designed to convey information, there are few interruptions, repetitions, or revisions. Of course, these speakers practice many hours to create their error-free utterances. In everyday conversation, people do not carefully plan ahead and practice the articulation and delivery of each utterance. As a result, spontaneous speech is full of stops and starts, repeated words, and restarted ideas. A common assumption is that these hesitations and revisions slow understanding. If fluency is every speaker's goal, then disfluency should be every listener's nightmare. But how do disfluencies really affect comprehension? The experiments I describe here suggest that contrary to expectation, speech disfluencies do not always hinder comprehension. Instead, disfluencies sometimes hinder comprehension but sometimes do not, depending on what kind of disfluencies they are, and on where they fall in the spoken sentence. The effects of disfluencies also have implications for plausible models of speech comprehension.

Speech disfluencies are generally defined as phenomena that interrupt the flow of speech and do not add propositional content to an utterance. This includes long pauses, repeated words or phrases, restarted sentences, and the fillers *uh* and *um*. Disfluencies are common in everyday speech, with estimates ranging from 2 to 26 disfluencies per 100 words (Faure, 1980; French, Carter, & Koenig, 1930, cited in Wingate, 1988; Kasl & Mahl, 1987; Mahl, 1959; Voelker, 1944; Kowal, O'Connell, & Sabin, 1975; Lutz & Mallard, 1986; Clark, Edwards, Liittschwager, & Dorado, 1993). The variance in estimates results largely from whether or not pauses are counted. Because not all pauses are disfluencies (Marek, 1980;

I thank Herbert Clark and Nanda Poulisse for making their tapes and transcripts available for me to use, and the Max Planck Institute for Psycholinguistics for technical and administrative assistance with the Dutch experiments. I also thank Herbert Clark, Willem Levelt, Paul Meijer, Bridget Bly, Elizabeth Wade, Kay Bock, Ellen Bard, Arthur Samuel, David Balota and several anonymous reviewers for helpful feedback. The research was carried out while I was receiving a National Science Foundation Graduate Fellowship, a National Research Service Award, and a Stanford Psychology Department Fellowship.

0749-596X/95 $12.00

Markel, 1990), and it is difficult to tell the difference between disfluent pauses and non-disfluent pauses, I prefer disfluency figures exclusive of pauses, and arrive at an estimate of about 6% non-pause disfluency across these studies.

Despite their frequency, the effect of disfluencies on comprehension has never been directly tested. However, models of language comprehension and indirect studies from which conjectures can be made predict that disfluencies hinder comprehension. This includes all syntax-based models, where interpretations of utterances are extracted from a syntactic base that is built cumulatively on a word-by-word basis following precise rules. In one model, for example, the first word in a sentence is identified and assigned a syntactic category, and then the second word is identified and assigned a category with respect to the first word (Wanner, 1980; Wanner & Maratsos, 1978). In this system, for example, the word after a determiner can be either an adjective or a noun, which in turn can be followed by another limited set of words, including a second adjective, a verb, or a complementizer. This model breaks down when faced with word sequences typical of disfluencies such as *it becomes forms a point,* where a verb follows another verb, or *of a of a heart,* where a preposition follows a determiner.

In another model, phrase structure trees are built according to heuristic rules, such as the rule that new words are attached to the accumulating syntactic tree in a way that minimizes the number of new sentence nodes made (Kimball, 1973; Frazier & Fodor, 1978). So, for example, readers take less time to read *the performer sent the flowers and was very pleased* than to read *the performer sent the flowers was very pleased* because *sent the flowers* has fewer nodes as a verb phrase than as a relative clause (Rayner, Carlson, & Frazier, 1983). The rules are mandatory, to the point of rendering syntactically normal sentences incomprehensible on a first pass reading, as in *the horse raced past the barn fell.* If even syntactically plausible words can slow comprehension, then typical non-filler disfluencies, which are almost always syntactically illegal, should have large negative effects.

One of the most prominent models of spoken language comprehension, that of Marslen-Wilson and Tyler (1980), would also predict that disfluencies hinder comprehension. In this model, the syntax, semantics, and pragmatics of a sentence contribute to the speed at which a word is identified in a stream of speech. For example, *guitar* is identified faster in the normal sentence *the boy held the guitar* than in the semantically implausible *the boy drank the guitar,* the pragmatically implausible *the boy buried the guitar,* or the categorially impossible *the boy slept the guitar* (Marslen-Wilson, Brown, & Tyler, 1988). Disfluencies are often syntactically irregular or semantically nonsensical, so according to this model, they should also slow comprehension. Preliminary support for this disfluency hypothesis comes from evidence that mispronounced words slow people's speed at verbally repeating another person's speech (Marslen-Wilson & Welsh, 1978). In some respects, disfluencies are similar to mispronunciations, and likewise might cause listeners to have difficulty.

In the experiments described here, two types of disfluencies were investigated, false starts and repetitions. False starts occur when speakers start to say something, but then decide to abort their utterances and begin again. For reference purposes, the aborted information will be referred to as the *false start* and the new information replacing it as the *fresh start.* In *for a really champion one you can-it's gonna be twenty cents, you can* is the false start, and the fresh start begins with *it's* (all examples in this paper are from the spontaneous speech of either the Dutch or English corpora that were also used in the experiments). In this example, as in the experimental disfluencies, the fresh start completely replaces the information supplied in the false start. Repetitions occur when speakers repeat words or phrases, such as *of a* in *the shape of a of a heart.* Many disfluencies are a cross between false starts and repetitions, because they both change information and repeat information, as in *you can actually get private insurance-private medical insurance.* In order to isolate the effects of different disfluencies, I limited my investigation to disfluencies that restarted all information, or that repeated words exactly without adding, removing, or replacing information.

If speakers aim for fluent deliveries of speech without backtracking in order to maximize comprehension (Clark & Clark, 1977), then when interruptions and revisions do occur, listeners should experience confusion and slowed comprehension. If comprehension is a process by which listeners build interpretive structures based on the semantics or syntax of incoming words, then false starts should hinder comprehension by forcing listeners to remove words from their structure, like *you can* in the earlier example, and replace them with other words, like *it's gonna be*. Likewise, repetitions should hinder comprehension because they do not fit with the preceding syntactic structure (*of* cannot follow *a* in *of a of a heart*). Changing the interpretive structure to accommodate the disfluency might disrupt speech comprehension. To test whether disfluencies do have this effect, I compared comprehension after a disfluency to comprehension when the disfluency was digitally excised.

As a measurement of comprehension, I used an *identical word monitoring* task (Marslen-Wilson & Tyler, 1980). In identical word monitoring, subjects listen to utterances with a particular target word in mind. When they hear the word they are listening for, they press a response button. The time it takes to press the button reflects their understanding of the utterance up to that point. Marslen-Wilson and Tyler found that subjects' response times get progressively faster with the increasing interpretability of the utterance. Subjects take longest to identify words in incomprehensible sentences. They take less time to identify words in somewhat comprehensible sentences. They are fastest at identifying words in completely comprehensible sentences. The word monitoring task can also reflect more subtle differences in sentence structure, such as whether a noun target is pragmatically or semantically plausible after a verb (Marslen-Wilson, Brown, & Tyler, 1988). In support of its validity as a measurement of comprehension is the fact that word monitoring yields the same pattern of comprehension effects as other measurement techniques, such as learning sentences (Epstein, 1961) or perceiving words in a noisy environment (Martin, 1968a, 1968b), with learning and perceiving enhanced as interpretability increases.

To investigate the effects of false starts and repetitions on comprehension, I compared the time it took people to identify the first viable target words following the disfluencies to the time it took them to identify the same tokens when the disfluencies had been edited out of the speech stream and either substituted with a pause (Experiment 1) or excised entirely (Experiment 2). Only if there is a disfluency effect in both cases can the effect be attributed to the disfluencies and not to the presence of absence of a pause. The design ensures that targets in both conditions are not only matched in form class, number of syllables, frequency, and all other lexical characteristics, but also that they are pronounced in exactly the same way.

The spontaneous aspect of the materials is important. Speech that is read aloud from a script is different from spontaneous speech, and yields different effects. Read speech is more likely to have pauses related to syntax, to be more evenly spoken, and to have more consistent distribution of word accents (Mehta & Cutler, 1988). The time it takes to detect a phoneme in read speech is greatly affected by the length of the word preceding the target phoneme and also by the position of the target in the sentence, but neither of these factors affect recognition in spontaneous speech. Because it is not known how spontaneous disfluencies differ from read or rehearsed disfluencies in either their properties or their effects, only spontaneous speech was used in the experiments described here.

EXPERIMENT 1

Method

Subjects. Thirty people from the Max Planck Institute for Psycholinguistics subject pool of students and community members participated in the experiment. They were each paid fl8,50 for their participation.

Materials. Stimuli were selected from a collection of 40 h of spontaneous Dutch speech gathered in 1984 (Poulisse, 1989). Speakers were high school and college students instructed to describe each of 12 abstract figures. The descriptions were recorded on standard audio cassettes. The experimenter sat in the room with the subjects, acting as a silent listener. The transcriptions note every vocalization, including filled pauses, laughter, and cut-off part-words. Forty-four fresh starts were selected from the corpus, to make 20 critical trials, 20 filler trials, and 4 practice trials. Likewise, 44 utterances containing repetitions were selected. Critical trials were those trials that contained the target words subjects monitored for. Filler trials did not contain the target words. The filler targets were as semantically and syntactically predictable as the critical targets, so subjects could not guess in advance that the target word was not present in the utterance. Target words closely followed the disfluencies in the utterance. In the repetitions, targets were the first word after the repetition in all but one case, which had an intervening *de* (*the*). Transcriptions of critical stimuli are in Appendix A.

Filler trials helped limit the use of strategies. Because targets did not occur in all utterances, subjects could not simply respond at the point where the target could fit into the utterance. They also could not adopt the strategy of responding after hearing a disfluency. Critical trials themselves often contained other spontaneous disfluencies that were not the focus of the experimental investigation. In addition, a phoneme monitoring strategy was of little use, because other words in the utterances began with the same phoneme as the target word. Because disfluencies were anticipated to affect semantic or syntactic integration, it was important that subjects not bypass these processes by only listening for sounds.

Utterances were digitized and then edited using the speech waveform editor of the Speech Laboratory of the Max Planck Institute for Psycholinguistics, replacing the false starts and the second occurrences of words in the repetitions with pauses.

An example of editing in the fresh start utterances follows (stimuli are translated from Dutch):

(1) Unedited: and the next figure, this has- it looks a little like a like a hammer.
(2) Edited: and the next figure, [silence] it looks a little like a like a hammer.

The false start is *this has*, and the fresh start begins with *it looks a little*. The target is the word *looks*. The lead-in to the target, the word preceding the suitable target word, is *it*. The average time between the offset of the last word before the false start and the false start was 833 ms (range of 0 to 2705 ms). The average length of the edited false start was 825 ms (range of 367 to 1988 ms; same as pauses). The average time between the offset of the false start and onset of the target word was 734 ms (range 0 to 2111 ms). The total time from the offset of the false start to the onset of the target includes the lead-in and any pause that may occur between the false start and the fresh start.

An example of editing in the repetition utterances follows:

(3) Unedited: in the shape of a of a heart
(4) Edited: in the shape of a [silence] heart

The second occurrence was excised because the second occurrence of a word in discourse is sometimes a less clear token than the first (Fowler & Housum, 1987). In the data described here, listeners had more difficulty deciphering the second occurrence of words in a repetition when the words were presented in isolation. Twenty people listened to either the first or second occurrence of words in 20 spontaneous repetitions, and to 20 non-repeating words matched in syntactic and semantic form to the repetitions (all items are from Experiment 4 stimuli). The listeners could decipher 52% of the non-repeating words (208/400) and 58% of the first occurring words (116/200). But they deciphered only 38% of the second occurring words (75/200). This difference between the first and second occurrences of words in spontaneous repetitions is significant ($F(1,19) = 6.87$, $p < .02$). If the second occurrence contains less information than the first occurrence, any decrement in comprehension in an edited condition that contained only the second occurrence could be attributed to impoverished input instead of the lack of a repetition. The average length of edited repetition/inserted pause was 416 ms (range of 190 to 697 ms).

Because background noise sounded different for every speaker on every original recording, pauses were selected individually for each stimulus. A pure machine generated silence would have caused a noticeable lack of sound in an edited utterance. Pauses were created by copying long pauses from elsewhere in the speech stream being edited and trimming them to the size of the edited-out speech.

Detectability of editing. When questioned, no subjects from either Experiment 1 or Experiment 2 reported hearing any unnatural-sounding speech. This was not surprising as the machine noise on the recordings was likely to obscure any potentially noticeable junctures. The difficulty of detection of editing has also been noted by another researcher, who found a continuity of perceived intonation when disfluencies were edited out. When *to a gray* was excised from the spontaneous utterance *right of blue to a gray to a pink point*, the sentence sounded as natural as if *to a gray* were never uttered (Levelt, 1984, p. 115).

To test the detectability of editing directly, 12 new subjects listened to the stimuli from Experiment 2 and marked on a transcript where they thought the utterance had been digitally spliced. Six subjects listened to the stimuli from each of the two experimental blocks. The stimuli from Experiment 2 were used because they showed stronger effects in the word monitoring experiments. If editing were causing the effect, then editing should be more easily detected with these items. Subjects could listen to a sentence up to six times. False positives were calculated by multiplying by the number of subjects the total number of positions where a false positive occurred across all subjects. These positions were taken to be the potential false positive points. The proportion of edits correctly detected (53/240) did not differ from the proportion of false positives (385/1776; $t(1,11) = .24$, ns), nor did the proportion of correct detections and false positives vary across disfluencies (false starts: $t(1,11) = 1.10$, ns; repetitions: $t(1,11) = -1.07$, ns).

Design. There were 128 stimuli in all. Eighty stimuli were critical stimuli, 40 were filler stimuli, and 8 were practice stimuli. The critical stimuli consisted of 40 different utterances, 20 fresh starts and 20 repetitions, in two versions, edited and unedited. Each of the filler stimuli came in only one version. Half of the filler trials (10 fresh starts and 10 repetitions) were edited, and the other half were not. The practice stimuli consisted of one example of each of the following eight possible permutations: unedited critical fresh starts, edited critical fresh starts, unedited critical repetitions, edited critical repetitions, unedited filler fresh starts, edited filler fresh starts, unedited filler repetitions, and edited filler repetitions.

Subjects were tested in two groups. The practice stimuli and the filler stimuli in each group were identical. However, the critical stimuli were divided between groups, so that each group heard only one version of an utterance. Each group heard 88 utterances in all, consisting of 8 practice stimuli, 40 critical stimuli, and 40 filler stimuli. The critical stimuli heard by one group (10 unedited fresh starts, 10 edited fresh starts, 10 unedited repetitions and 10 edited repetitions) were heard in opposite versions by the other group.

The experimental order was pseudo-randomized with the constraints that two utterances from the same speaker had at least two utterances from other speakers between them, that no more than four filler trials or critical trials were presented in a row, and that no two descriptions of the same figure were adjacent. The same presentation order was used for both groups.

Procedure. In each experimental trial, subjects heard a warning tone, saw a word on a computer screen, and then heard a recording of a spontaneously produced utterance. They were instructed to press a button immediately upon hearing the target word, with emphasis on both speed and accuracy. The tone lasted for 400 ms, followed by a 600 ms silence. The word then appeared for 500 ms, and 500 ms later the utterance began to play. The utterance played to the finish regardless of whether the subject pressed the response button or not. There was a 1 s silence between trials. The sentences lasted between 3 and 27 s, with an average of 10 s. In the fresh start trials, the target words were heard an average of 6 s into the sentence, after an average of 14 words had been heard. In the repetition trials, the target words were heard an average of 9 s into the sentence, after an average of 19 words had been heard. The experiment lasted about 20 minutes. Here and in all other experiments, reaction time was measured in milliseconds from the onset of the target word.

Results

An experimental error that prematurely timed out responses and the inadvertent use of a neologism as a target word caused the elimination of three items from the false start analysis and seven items from the repetition analysis. Of the remaining responses, 1.7% of the false starts (9/510) and 1.5% of the repetitions (6/390) were null responses, and 6.5% of the false starts (33/510) and 4.9% of the repetitions (19/390) were more than two standard deviations from the mean and treated as outliers. In addition, 1.3% of the filler trials were false positives (16/1200). There was no significant difference across conditions for the error rates.

False starts do appear to hinder comprehension; targets are identified about 22 ms faster in the absence of a false start ($F1(1,29) = 5.25$, $p < .03$; $F2(1,16) = 1.17$, ns; in all analyses, a critical p value of .05 was used to determine significance). In contrast, repetitions appear to speed comprehension; targets are identified about 84 ms faster in the presence of a repetition ($F1(1,29) = 46.6$, $p < .001$; $F2(1,12) = 14.39$, $p < .005$). Table 1 summarizes the results.

Discussion

Results for false starts support the intuitive prediction that disrupting the smooth flow of speech by creating syntactically and semantically anomalous strings slows the comprehension process. When a speaker makes a false start, the target word following the disfluency tends to be harder to recognize. But the results are only significant by subjects. With a new set of items, there may no longer be a false start disadvantage. One possible reason for the insignificant *F2* is that the inserted pauses were also causing a disruption. The disruption would vary depending on the amount of pausing. This variability could reduce the significance of the *F2* analysis.

Results for the repetitions, on the other hand, do not support the prediction. When a speaker repeats, the listener is better able to recognize the word immediately following the repetition, not worse. But there is another explanation for the results. The pauses in the edited versions of the repetitions may have interrupted the flow of the utterance, or added an element of uncertainty, causing a slower detection of targets after pauses, rather than a speeded detection of targets after repetitions.

In Experiment 2, the contribution of pauses was tested by measuring recognition times when pauses were not added to the edited versions of the utterances. If pauses are disruptive, the difference between the unedited and edited false start conditions should increase, and the difference between the unedited and edited repetitions conditions should decrease when the disruptive pauses in the edited versions are removed. For the false starts, comparing monitoring after false starts to monitoring in fluent, pause-less utterances should result in greater condition differences than was found in com-

TABLE 1

EXPERIMENT 1: SUMMARY OF MEAN REACTION TIMES AVERAGED ACROSS SUBJECTS AND ITEMS (IN MILLISECONDS)

	With disfluency	Without disfluency, with pause	Difference
False starts	590	568	22
Repetitions	462	548	−86

paring monitoring after false starts and monitoring after pauses. The benefit of removing the disruptive pauses in the edited condition (*the next figure, it looks*) should result in faster monitoring than in Experiment 1, where pauses were inserted (*the next figure, [pause] it looks*). Monitoring in the unedited version (*the next figure, this has it looks*) should remain the same. For the repetitions, comparing monitoring after a repetition to monitoring in a repetition-free and pause-free utterance should result in smaller condition differences than when comparing monitoring after a repetition to monitoring after a pause. Monitoring might even be faster in the repetition-free and pause-free utterances than in the originally repeating utterances if it is the case that the disruptiveness of pauses is so great it masks a disruptive repetition effect in Experiment 1. Removing pauses might allow this negative repetition effect to surface.

An alternative explanation for the discrepancy between false starts and repetitions is that the false starts considered are longer than the repetitions in ms, and only long disfluencies lead to detectable disadvantages. This explanation is hard to maintain in light of the fact that longer false starts show the same results as shorter false starts (see General Discussion of False Starts). The discrepancy is more likely to result from syntactic or semantic differences between disfluencies rather than from their differences in length.

EXPERIMENT 2

Method

Subjects. Thirty new people from the Max Planck Institute for Psycholinguistics subject pool of students and community members participated in the experiment. They were each paid fl8,50 for their participation.

Materials. Stimuli were constructed from the same utterances used in Experiment 1. As before, each of the critical trials was heard in an edited and unedited version. The unedited version was identical to that of Experiment 1. The edited version consisted of the edited versions of Experiment 1 without the inserted pauses. The targets now fell at the end of the phrase preceding the false start, after the appropriate lead-ins or immediately after the first repetition. The stimulus containing a neologism was replaced.

Design. The design was the same as that of Experiment 1.

Procedure. Experiment 2 differed from Experiment 1 in the timing of each stimulus trial. The warning tone lasted for 500 ms, followed by a silence of 1500 ms, to give subjects enough time to focus their attention on the computer screen. Some subjects in Experiment 1 turned their heads away from the screen while listening to the sentences, and were sometimes not fully prepared for the next trial. Second, the target word appeared on the screen for 715 ms, 215 ms longer than Experiment 1, to make sure all subjects had enough time to read the word. After a 285 ms pause, the sentence started to play. There was a 1 s silence between trials.

Results

Of a total of 600 possible responses, 4% of the false starts (24/600) and 4.5% of the repetitions (27/600) were null responses, and 2.3% of the false starts (14/600) and 1.2% of the repetitions (7/600) were outliers. In addition, 1.4% of the filler trials were false positives (17/1200). There was no significant difference across conditions for the error rates.

The pattern of results is the same as in Experiment 1; word monitoring was slower after

TABLE 2
EXPERIMENT 2: SUMMARY OF MEAN REACTION TIMES AVERAGED ACROSS SUBJECTS AND ITEMS (IN MILLISECONDS)

	With disfluency	Without disfluency, without pause	Difference
False starts	615	585	30
Repetitions	459	510	−51

false starts ($F1(1,29) = 6.24, p < .02$; $F2(1,19) = 4.73, p < .05$) and faster after repetitions ($F1(1,29) = 38.21, p < .001$; $F2(1,19) = 20.89, p < .001$). Table 2 summarizes the results.

Discussion

Removing pauses from the edited condition did not change the results. Word monitoring was still faster without a false start than with a false start, and with a repetition than without a repetition.

The disadvantage of a false start was about the same in the two experiments (22 ms and 30 ms). If pauses were disruptive, then the pauses preceding targets in Experiment 1 should have disturbed listeners, and the difference between conditions in Experiment 2 should have been greater than in Experiment 1. Because the results of Experiment 2, without pauses, were similar to those of Experiment 1, with pauses, it seems that pauses did not create a disadvantage. So at least in the environment of a fresh start, pausing has no effect on word monitoring.

In contrast, pauses may have had an effect in the environment of a repetition. Inspection of means across experiments suggests that pauses might have slowed reaction time in the repetition utterances. The mean reaction time to the target in the unedited condition is almost identical in both experiments. But in the edited condition, the mean reaction time to the target after a pause is 35 ms slower than the mean reaction time without the pause. However, because there is still a strong repetition advantage, the contribution of pausing is at most an additional influence on reaction time and not the source of the effect found in Experiment 1.

The effects of pausing may be different for the two disfluencies because of the differing environments associated with fresh starts and repetitions. For example, pauses may have differential effects depending on where they fall in a sentence. Eighteen of the 20 repetitions were in the middle of a clause, and 8 of these were in prepositional phrases. In contrast, 8 of the 20 false starts were in the middle of a clause, and 3 of these were in prepositional phrases. Or, pauses may have different effects depending on the form class of the targets. Eighteen of the repetition targets were nouns, compared to 8 of the false start targets.

The lack of a disruptive effect of pauses in the false start utterances rules out the possibility that variable amounts of pause disruption caused the nonsignificant false start items effect in Experiment 1. However, a combination of other factors, such as the balance between beginning and middle false starts (see below) and subjects' poorer performance (perhaps due to the timing of Experiment 1), may be responsible for the null effect. In both Experiments 2 and 3, the false start disadvantage was reliable.

Because the effects of the two disfluencies are different, and, as will be argued, have different etiologies, the remainder of this paper will discuss each disfluency separately. Discussion of Experiments 1 and 2 and follow-up experiments will first be presented for false starts.

False Starts

It is no surprise that false starts slow listeners' ability to monitor for words. After all, they lead listeners to expect one utterance, but then require listeners to abandon the interpretive model they've built and begin again. In a syntactic parsing metaphor, the listener has to restructure part of a syntactic tree built on the basis of the false start to accommodate the revised fresh start. The restructuring takes time, and this is manifested by slower word monitoring. If this explanation is correct, then perhaps different amounts of restructuring will take different amounts of time. So, for example, longer false starts might cause listeners to build more syntactic nodes in the interpretative structure, and removing several nodes might take more time than removing one. Some evidence against this hypothesis is that neither the length of the false start in ms nor in words correlated with the differences in reaction times (Experiment 1: $r = -.15$, ns, ms, $r = -.40$, ns, words; Experiment 2: $r = .12$, ns, ms, $r = -.26$, ns, words; the number of words in the false starts ranged from one to four with a median of two).

An alternative account for the phenomenon is that it is not restructuring that causes slower monitoring, but that it is difficulty making the repair that causes the slower monitoring. When listeners hear a fresh start they have to recall what was said before the false start, identify and abort the false start, and attach the fresh start to the appropriate point in the utterance. The repair process directs attention away from recognition of the upcoming words. In this account, it would not matter how long or complex the false start was, because the difference between conditions does not arise from backtracking in the syntactic tree structure and erasing incorrect nodes, but from the effort of finding where to attach the fresh start.

One way to test the repair hypothesis would be to compare word monitoring when easier or harder repairs need to be made. One way to quantify easier or harder repairs is where the false starts occur in the sentence. False starts occurring at the beginning of sentences might be less disruptive than those occurring in the middle because listeners can abort false starts in their entirety, without searching for the fresh start attachment point, or recalling prior information.

In Experiment 1 there were 11 false starts that began a sentence and 6 that were in the middle of a sentence. In Experiment 2 there were 13 false starts that began a sentence and 7 that were in the middle of a sentence (three items excluded from Experiment 1 were in Experiment 2). These two groups of false starts had markedly different effects on word monitoring. The mean reaction times for targets falling after false starts that occurred at the beginning of a sentence and after false starts that occurred in the middle of a sentence are presented in Table 3.

There was a significant interaction between where in the sentence a false start occurred and whether or not the false start created a word monitoring disadvantage (Experiment 1: $F1(1,29) = 13.17, p = .001$, $F2(1,15) = 6.80$, $p = .02$; Experiment 2: $F1(1,29) = 6.56, p < .02$, $F2(1,18) = 17.92, p < .001$). In the middle of sentences, word monitoring was greatly sped up by the false start's removal. But in the beginning of sentences, monitoring was the same with and without the false start.

The repair hypothesis was tested directly in Experiment 3 by comparing monitoring times after a new set of beginning false starts and middle false starts.

EXPERIMENT 3

Method

Subjects. Thirty men and women from an introductory psychology course at Stanford University participated for course credit.

Materials. Stimuli were selected from a corpus of about 20 h of spontaneous English speech taped in 1989 at Stanford University (Clark, Edwards, Liittschwager, & Dorado, 1990). Speakers were taped on standard audio cassettes using a tape recorder. The speakers spoke in pairs, taking turns retelling stories that they had each heard separately immediately prior to the retelling. The original stories were played to subjects from a prepared audio tape. From the transcript, 44 false starts that began an idea and 44 false starts that occurred within a phrase were selected. Twenty begin-

TABLE 3
EXPERIMENTS 1 AND 2: SUMMARY OF MEAN REACTION TIMES BY PLACEMENT OF FRESH STARTS AVERAGED ACROSS SUBJECTS AND ITEMS (IN MILLISECONDS)

	With false start	Without false start	Difference
Beginning sentence			
Expt 1	614	628	−14
Expt 2	620	625	−5
Within sentence			
Expt 1	567	484	83
Expt 2	618	526	92

ning and 20 middle false starts served as critical trials, and an additional 20 of each served as filler trials. The remaining 8 false starts formed the practice stimuli. The false starts were never the very first words of the stimulus recording. Beginning false starts were usually introduced by the sentence preceding them. This was done to ensure that the effects could not be ascribed to subjects' systematic focus or non-focus of attention at the beginning of a trial. If this precaution were not taken, then some beginning false starts would also begin a trial, but no middle false starts would begin a trial. Each trial always contained a complete thought. Two examples of stimuli follow (the false start is capitalized and the target word is italicized):

(5) Beginning: you could get a little ant swing and an ant ladder and IT COULD EVEN- *there's* a little thing where the ant can climb up a pole and ring a bell

(6) Middle: people have more money there and I'm a business manager and um he says WHY- *but* what about you?

Targets were always the first word after the false start in the middle false start trials. In the beginning false start trials, however, three targets were after one lead-in word due to a shortage of optimum stimuli. As in Experiments 1 and 2, filler trial targets were chosen so that they were as syntactically and semantically predictable as critical trial targets. Transcripts of critical trial stimuli are in Appendix B. To complete the experimental design, 10 false starts preceded by discourse markers and 7 false starts preceded by conjunctions were included. Two of the discourse markers were *I mean* and *you know*. The other eight were words like *and* and *so*, which could serve as either discourse markers or conjunctions. I categorized a word as a discourse marker if it seemed to be unnecessary as a literal connective and if it was uttered in a drawn-out way typical of discourse markers, as if speakers were considering what to say next as opposed to knowing what to say and connecting the ideas with a short, firm *and*. Examples of *and* categorized as a discourse marker and as a conjunction follow (false starts are capitalized, target words are italicized, and the *and*s in questions are in bold-face):

(7) Discourse Marker: This story is about two people, Doris and George, **and** uh THEIR- THE BEGINNING the *first* sentence is that they are married for some amount of years, I forget the exact amount.

(8) Conjunction: You could get a little ant swing and an ant ladder **and** IT COULD EVEN- *there's* a little thing where the ant can climb up a pole and ring a bell.

Although the presence of discourse markers or conjunctions meant these beginning false starts were not truly at the beginning of their idea unit, it did not seem that the definition of beginning false start was stretched too much by the exception. Discourse markers are thought to operate at a meta-discourse level guiding the conversation or the interaction of conversational participants but not affecting the literal level of the information conveyed (Jucker, 1993; Redeker, 1990; Schiffrin, 1987). False starts, in contrast, affect the sentence integration level, at least as far as the experiments described here have indicated. Because discourse markers and false starts were anticipated to affect different levels, I did not expect discourse markers to affect the false start disadvantage. Three of the middle false starts were also preceded by discourse markers.

Utterances were digitized using the MacRecorder audio-digital converter, and edited using the SoundEdit speech editing program. False starts were excised entirely and not replaced with pauses because it was found in Experiments 1 and 2 that the presence or absence of a pause was irrelevant to the false start disadvantage.

The average time between the last word before the false start and the false start was 237 ms (range 0 to 1105 ms). The average length of false start was 711 ms (229 ms to 3 s 272 ms). The average time between offset of the false start and onset of the target word was 204 ms (0 to 573 ms). The average number of words in the false start was 2.35 (1 word to 8 words), and the average number of syllables was 2.68 (1 syllable to 8 syllables). There was a marginal difference between the length of beginning false starts and middle false starts. Beginning false starts were an average of 922 ms long (229 ms to 3 s 272 ms) and middle false starts were an average of 499 ms long (237 ms to 1 s 624 ms; $t(27) = 2.53$, $p =$ ns at the significance level adjusted for the number of tests, .05/7, or .007). Though beginning false starts were only marginally longer than middle false starts, they consisted of significantly more words and syllables (words: $t(27) = 3.64$, $p < .007$; syllables: $t(31) = 3.40$, $p < .007$). Beginning false starts and middle false starts did not differ on any of the other following variables compared: the length of the pauses before the false starts, the length of the pauses after the false starts, the number of words in the utterances up to the false starts, and the number of milliseconds from the onset of the utterance to the false start.

Design. The design of the experiment was the same as that of Experiments 1 and 2. There were 8 practice trials, 40 filler trials, and 40 critical trials. The critical stimuli consisted of 20 false starts that began an idea, and 20 false starts that occurred within a phrase. Each critical trial had two versions, one with the original false start, and one with the false start excised. As previously described, subjects were run in two groups so that each group heard only one version of the critical trials.

Procedure. The procedure was similar to that of Experiments 1 and 2, with some modification in timing. The warning tone lasted for 500 ms, followed by a 1 s silence. The target word then appeared for 1 s, and 500 ms later the utterance began to play. After pressing the response button, the utterance played to the finish, and after a 1 s silence the next trial began. The utterances lasted between 3.5 s and 23.5 s, with an average of about 9.5 s. The target words were heard an average of 6 s into the utterance, after an average of 15 words had been heard. The experiment took about half an hour.

Results

Of a total of 1200 possible responses, 10.8% were null responses (130/1200) and 4.8% were outliers (58/1200). The unusually high number of null responses results from five items in which fewer than a third of the subjects responded. The targets were either hard to hear or hard to monitor for because they lacked syntactic or semantic prominence. One example of an often missed target is *if* in *I wanna if I'm going to. If* did not stick out in the speech stream, and was easily overlooked. Another example is *then* in *and he said that then she asked him.* However, because the pattern of means and F values do not differ when all items are analyzed as opposed to when low-response items are excluded, analyses presented here include all items. In addition, 3.8% of the filler trials were false positives (45/1200). There was no significant difference across conditions for the error rates.

Unexpectedly, word monitoring was slower after both beginning and middle false starts (beginning: $F1(1,29) = 16.15$, $p < .001$, $F2(1,19) = 16.24$, $p < .001$; middle: $F1(1,29) = 17.06$, $p < .001$, $F2(1,19) = 6.25$, $p < .03$). Table 4 summarizes the results.

The presence of discourse markers may have influenced the results. The beginning false

TABLE 4

EXPERIMENT 3: SUMMARY OF MEAN REACTION TIMES AVERAGED ACROSS SUBJECTS AND ITEMS (IN MILLISECONDS)

	With false start	Without false start	Difference
Beginning	637	568	69
Middle	658	597	61

starts preceded by discourse markers tended to show an unexpected false start disadvantage, though the interaction was nonsignificant ($F1(1,29) = 3.57, p < .07$; $F2(1,18) = 3.54, p < .08$). Table 5 summarizes the means of the beginning false starts that were either preceded by a discourse marker or not.

When inspecting the means in Table 5 it is important to remember that the items with a marker and without a marker are not the same, and so means between rows should not be compared. The important variable is the difference between conditions.

At least two important causes of variability among the target words, the form class and the number of syllables, could not account for the differing outcomes of the Dutch and English experiments. Tests of differing effects for items containing targets that were nouns, verbs, or other yielded no differences for either Experiment 2 ($F1(2,58) = .66$, ns; $F2(2,17) = 1.52$), ns) or Experiment 3 ($F1(2,58) = .94$, ns; $F2(2,37) = .88$, ns). Likewise, tests of monosyllabic versus multisyllabic targets yielded no differences for either Experiment 2 ($F1(1,29) = 1.15$, ns; $F2(1,18) = 1.62$, ns) or Experiment 3 ($F1(1,29) = 0$, ns; $F2(1,38) = .25$, ns). Given the inherent variability of spontaneous speech, it is impossible to control all differences in stimuli. But these two analyses of gross differences do bring out the importance of the discourse marker phenomena as opposed to other, random differences among stimuli.

Discussion

Results confirm the prediction that false starts in general hinder comprehension, and further show that this phenomenon is not restricted to Dutch. It is not the case that there is a language-specific syntactic problem underlying the false start disadvantage. Some of the Dutch false starts, such as *daar gaat 't* (literally, *there goes it*), had syntactic structures that do not exist in English. But English false starts worked just as well in slowing comprehension. They differed from Dutch false starts, however, in the strength of their effects at different point in the sentence. The Dutch beginning false starts did not hinder comprehension. The English beginning false starts did.

There is one noticeable difference between the beginning false starts in the Dutch and English experiments. In the English experiments, half of the beginning false starts began with discourse markers. None of the Dutch false starts began with discourse markers. There was a tendency for the false start disadvantage to occur in beginning false starts only when those false starts were preceded by discourse markers. The presence of a discourse marker may increase the repair difficulty, in effect changing the false start from a beginning false start to a middle false start. Markers might not be metadiscourse comments bearing no relevance to the literal level of what's spoken. Instead, they may be meaningfully connected to the following information, providing information that listeners must refer back to when making repairs. Conjunctions, in contrast, may be like fulcrums, connecting both preceding and following information equally. False starts after conjunctions are similar to beginning false starts in that they have less effect on word monitoring. When people hear a conjunction, they may conclude the previous idea, but not open a new sentential frame until after the conjunction. Conjunctions may not be the first elements of the next idea. Discourse markers, in contrast, may be.

Suggestive support for this idea is that the word *and*, which made up more than half of both the discourse markers and the conjunc-

TABLE 5

EXPERIMENT 3: SUMMARY OF MEAN REACTION TIMES BY WHETHER OR NOT A DISCOURSE MARKER PRECEDED THE BEGINNING FALSE START AVERAGED ACROSS SUBJECTS AND ITEMS (IN MILLISECONDS)

	With false start	Without false start	Difference
With marker	634	538	96
Without marker	637	599	38

tions preceding false starts, seems to interact with false starts differently depending on the role it plays in the utterance. If *and* is a discourse marker the false start disadvantage seems to be stronger than if *and* is a conjunction. The difference between conditions for *and* as a discourse marker was 114 ms, but only 44 ms for *and* as a conjunction. The interaction was nonsignificant ($F1(1,21) = 1.12$, ns, only 22 subjects had complete cells; $F2(1,11) = 2.92$, ns), but the power of the test was weaker than in the other analyses due to the lower number of items and subjects. A direct test of the interaction between discourse markers and false starts would be a fruitful next step; unfortunately, the corpora used in the experiments described here did not contain enough samples of discourse markers co-occurring with disfluencies, so the clarification of the role of discourse markers must be left to future investigation.

General Discussion of False Starts

In both the Dutch and English experiments, the presence of a false start slowed reaction time to words at the beginning of the fresh start. In the Dutch experiments, the false start disadvantage occurred only when the false start was in the middle of an utterance. I propose that that is where the repair process has the greatest cost to the listener, who must hold part of the utterance in memory, abort the false start, and attach the fresh start. With beginning false starts, listeners can abort the mistaken utterances without holding earlier sentence fragments in memory or executing a costly repair process, so there is no delayed recognition of target words in the fresh starts.

In the English experiment that directly tested the difference between beginning and middle false starts, the beginning false starts unexpectedly caused a false start disadvantage. I suggest that this is because half of the stimuli were not truly beginning false starts. They were false starts near the beginning, but after a discourse marker. The presence of discourse markers may have the same effect on beginning fresh starts as context in the first part of an utterance has on middle fresh starts.

Three alternative accounts for the differing amounts of false start disadvantage are untenable. One is that all processing is harder at the beginning of a sentence, because there is no prior context to speed processing, and so the added difficulty of a false start cannot be detected. But beginning false starts do not take longer to respond to, as might be expected by a beginning-difficulty hypothesis; the means for the beginning and middle false starts are not systematically different (see Tables 3 and 4). Secondly, there is no effect of the speed of reaction time relative to the number of words preceding the target (Experiment 1: $r = .17$, ns; Experiment 2: $r = .13$, ns; Experiment 3: $r = -.12$, ns), which suggests that prior context is not contributing to the reaction times in the first place. This conclusion seems to contradict earlier findings of word position effects in sentences, such as Marslen-Wilson and Tyler's (1980) finding of faster word monitoring the further into the sentence a target word falls. The source of the diverging results is probably the fact that the current studies were done with spontaneous speech. Mehta and Cutler (1988) directly compared word position effects in spontaneous speech and in read speech using a phoneme monitoring task. They found word position effects only with the read speech.

Another explanation for the difference is that beginning false starts are easier to detect on phonological grounds, because their pitch or loudness might more clearly indicate the start of a new idea. If this were true, then fewer beginning false starts should be noticed as edited than middle false starts, because it should be easier to excise information before this proposed beginning-sentence cue than to excise information in the middle of sentences where no phonological information marks the boundary between the false start and the fresh start. But, in fact, just as many beginning false starts were noticed as edited as middle false starts ($t(1,11) = -.74$, ns; see Detectability of Editing under Experiment 1 for more information on how this data was obtained).

A third interpretation is that false starts lead listeners down the wrong path, and it is the amount of confusion that causes the differential disadvantage. Beginning false starts might not show as strong a disadvantage because they might be shorter or structurally less complex, and so not lead listeners as far off. This account further implies that the effects in general do not result from the process of repair, such as finding the fresh start attachment point, but from how much listeners have been led astray.

The first evidence against the confusion hypothesis is that the length of false start does not affect reaction time. This means that short and long false starts, with more or less information to abort, affect reaction time equally. The difference between reaction times in the two conditions does not correlate with the length of the edited-out false start (Experiment 1: $r = -.15$, ns; Experiment 2: $r = .12$, ns; Experiment 3: $r = .22$, ns), with the length of the lead-in phrase (Experiment 1: $r = -.04$, ns; Experiment 2: $r = -.31$, ns; Experiment 3, $r = -.02$, ns), with the sum of the length of the false start and the length of the lead-in (Experiment 1: $r = -.14$, ns; Experiment 2: $r = -.27$, ns; Experiment 3: $r = .21$, ns), or with the number of words in the false start (Experiment 1: $r = -.40$, ns; Experiment 2: $r = -.26$, ns; Experiment 3: $r = .06$, ns). One hypothesis about the correlations presented here is that the variability in the lengths of the false starts biased the results; that is, perhaps the false starts studied actually fell into two groups, short ones that affected lower level lexical processing, and longer ones that affected a higher level sentence integration process. Grouping them as one might mask a real correlation. I investigated this possibility by taking a closer look at the false starts in Experiment 3, which were particularly variable (range of 229 ms to 3 s 272 ms), and contained three false starts that were much longer than the other 17. Eliminating the three false starts from the analyses did not change the results, nor did separate correlations on short and long groups as defined by a median split.

The lack of a correlation between the false start length and the reaction time differences between conditions shows that false starts have a constant effect. Once information needs to be removed from the discourse record, as with a false start, it does not matter how much information is removed, the target's recognition is hindered by the same amount.

The second evidence against the confusion hypothesis is that the severity of the false start does not affect reaction time. False starts that are more different in syntax and semantics from the fresh starts following them may create a greater disturbance than those that are less different, and the amount of discrepancy might vary for beginning and middle false starts. I categorized all the false start and fresh start combinations from Experiment 3 as either (1) similar, where syntactic structure is parallel, (2) moderately similar, where some structure is changed, or (3) not similar where all structure is changed, as in the following examples:

(9) Similar: the lady—the interviewing
(10) Moderately Similar: he asks her—she wants to know
(11) Not Similar: if there—does he have

There is no systematic difference between the reaction times for beginning and middle false

TABLE 6
EXPERIMENT 3: SUMMARY OF MEAN REACTION TIMES BY STRUCTURAL SIMILARITY BETWEEN THE FALSE START AND THE FRESH START AVERAGED ACROSS SUBJECTS AND ITEMS (IN MILLISECONDS)

	Uned	Ed	Difference	N(Items)
Beginning				
Similar	616	558	58	6
In between	604	495	109	5
Not similar	660	624	36	9
Middle				
Similar	607	563	44	9
In between	744	594	150	4
Not similar	647	609	38	7

starts and the structural similarity between the false starts and the fresh starts (beginning: $F1(2,56) = 2.86$, ns, $F2(2,17) = 1.63$, ns; middle: $F1(2,56) = 2.73$, ns, $F2(2,17) = 16.3$, ns; one subject was removed from the $F1$ analysis due to incomplete cells). Table 6 summarizes the results.

There is no support for the hypothesis that the false start disadvantage is caused by false starts leading the listener astray. The source of the disadvantage is more likely to be the repair process. False starts tend to affect comprehension negatively only when they occur in the middle of an utterance or after discourse markers, when listeners have to maintain in memory information preceding the false start. Removing information from the comprehension model does not in itself appear to hinder comprehension, because beginning false starts can be aborted with no cost to the recognition of following words. But the fact that these beginning false starts are rare means that for practical purposes, when listeners hear false starts, comprehension is slowed. The differential results arising from the false starts' locations helps explain why this false start disadvantage occurs.

Repetitions

Despite the fact that repetitions also disrupt the flow of speech and require listeners to make repairs, repetitions appear to aid comprehension. Because the syntax and semantics are repeated and not changed, repairing repetitions may be an altogether different process from repairing false starts. There are at least two ways repetitions can lead to facilitation and not disruption.

One possibility is that having just heard a word facilitates hearing it a second time. Reducing processing on repeated words may facilitate processing of the next word. Another explanation is that repetitions free attention for upcoming information by helping to identify the words in the repetition. When listeners are not sure what a word is, they hold the acoustic information in memory until enough information accumulates for the word to be identified with hindsight (Bard, Shillcock, & Altmann, 1988; Connine, Blasko, & Hall, 1991). Repetitions might confirm that the first tokens of words in the repetition were heard correctly. For example, a listener hearing the sound *uth uh* twice has two chances to interpret the sounds as *of a* and not *over,* allowing memory to be purged earlier than would have been the case with only one token of *uth uh.* With a purged memory, the listener can devote full attention to recognition of the upcoming word, which translates to speeded reaction times after repetitions.

But before accepting a repetition facilitation account, an important alternative explanation for the results needs to be tested: perhaps it is not repeating that speeds word monitoring, but editing that slows it. Previous research has led to conflicting predictions about the effect of editing. Tyler and Warren (1987) found that inserting pauses before target words slowed word monitoring, but Meltzer, Martin, Mills, Imhoff, and Zohar (1976) found that the same manipulation sped phoneme monitoring. In contrast to both, Mens and Povel (1986) found no effect of inserted pauses on phoneme monitoring. They argued further that other researchers' findings of negative effects of temporal displacement may have resulted from local distortions of the speech stream. They considered in particular other experiments by Meltzer et al (1976) and experiments by Buxton (1983). In Experiment 4, the relative contribution of repeating versus editing to the overall effect was measured by comparing word monitoring times for targets following spontaneous repetitions to those following artificially created repetitions, or, *artificial* repetitions.

EXPERIMENT 4

Method

Subjects. Thirty men and women from an introductory psychology course at Stanford University participated for course credit. They were the same subjects who participated in Experiment 3.

Materials. Stimuli were chosen from a corpus of about 20 hours of spontaneous English speech. Details of how the corpus was made can be found in the materials section of Experiment 3. From the transcript, 20 utterances containing spontaneously produced repetitions were selected. Each of these 20 utterances was matched with spontaneously produced utterances of the same syntactic form but without repetitions. In some cases, the target words after originally produced repetitions and their matched non-repeating utterance were identical, as in the following examples:

(12) Originally Repeating: *our own our own medical,* as spoken by Kate to Roland
(13) Originally Non-repeating: *our own medical,* as spoken by Monte to Dinah
(14) Originally Repeating: *when you when you heard,* as spoken by Derek to Oscar
(15) Originally Non-repeating: *when he heard,* as spoken by Marty to Chuck

Targets were always the first word after a repetition. Transcripts of critical trial stimuli are in Appendix C.

Speech was digitized and edited using the same system described in the materials section of Experiment 3. There were four conditions in all, presented below (added repetitions are capitalized, and target words are italicized):

(16) Condition 1, Repeating: Where were you when you when you *heard* this news?
(17) Condition 2, Repeating → non-repeating: Where were you when you *heard* this news?
(18) Condition 3, Non-repeating: Where was he when he *heard* the news?
(19) Condition 4, Non-repeating → repeating: Where was he when he WHEN HE *heard* the news?

Condition 1 was edited to create Condition 2 by excising the second occurrence of the repetition. Condition 3 was edited to create Condition 4 by copying the words to be repeated and inserting them into the speech stream with varying lengths of pauses between the two occurrences of the repeated words. As in the previous experiments, pauses were created by copying static from other sections of that speaker's utterance. The standard pause length inserted was 200 ms, which was the average inter-repetition pause length of the spontaneously produced repetitions. However, in many cases, the pause was shortened or lengthened to make the repetition fit with the tempo of the speaker's speech. If the speech sounded very unnatural, it would be likely to disrupt processing no matter what repetition advantage there was.

Although maintaining natural-sounding speech after removing repetitions was easy, creating natural-sounding artificial repetitions was difficult. A native English speaker listened to the recorded utterances and judged if the stimuli sounded natural or not. Though she was aware of the experimental design and purpose, she was not aware of where the stimuli had been edited. All artificial repetitions that were considered unnatural were replaced. A common cause of unnatural repetitions was that the phoneme starting the repetition did not sound natural without the last phoneme of the word preceding it. For example, in the phrase *pay our own,* the *o* of *our own* did not sound natural without the *y* of *pay,* making *pay our own our own* sound unnatural. Repetitions were also easily detectable when the repeated words had marked prosodic forms or pitches that made repetitions sound out of place. A high pitched *if,* for example, would never occur twice in a row in spontaneous speech. First and second occurrences of words in spontaneous repetitions are slightly different, and listeners expect this. Many of the repetitions sounded odd because the sounds that repeated were too similar. Post-experiment measurements of listeners' judgments of editing are discussed under General Discussion of Repetitions.

The average length of repetition was 398 ms (range 171 to 736 ms). There was no difference between the length of the spontaneous repetitions that were excised or the artificial repetitions that were inserted.

Design. The design is the same as that of previous experiments. There were 8 practice trials, 40 filler trials, and 40 critical trials. The critical stimuli consisted of 20 spontaneous repetition utterances and 20 matched utterances without repetitions, each in two versions, edited and unedited. As previously described, subjects were run in two groups so that each group heard only one version of the critical trials.

Procedure. The procedure is described in Experiment 3. The utterances lasted between 4 s and 18 s, with an average of about 10 s. The target words were heard an average of 6 s into the sentence, after an average of 15 words had been heard. The experiment lasted about half an hour.

Results

One subject was replaced because he missed 35% of the critical trials (the average miss rate was 10%). Of a total of 1200 possible responses, 5.2% of repetitions (62/1200) were null responses, and 4.3% of the repetitions (51/1200) were outliers. In addition, 5.2% of the filler trials were false positives (64/1200). There was no significant difference between conditions for the error rates on originally repeating stimuli (Conditions 1 and 2). Originally non-repeating stimuli (Condition 3) did have more errors than the same tokens with artificial repetitions added (Condition 4), suggesting a repetition advantage, but the difference was only significant by items ($F1(1,29) = 3.16$, ns; $F2(1,19) = 4.61$, $p < .05$). Similar items in Experiment 5 show no effect.

Results are different for the spontaneous repetitions and the artificial repetitions. For the spontaneous repetitions, word monitoring was faster after the repetitions ($F1(1,29) = 12.70$, $p = .001$; $F2(1,19) = 16.75, p = .001$). For the artificial repetitions, word monitoring was unaffected by the repetitions $F1(1,29) = 2.58$, ns; $F2(1,19) = .34$, n.s.). Table 7 summarizes the results.

Discussion

With respect to spontaneous repetitions, results replicate effects found in Experiment 2. Targets after spontaneously produced repetitions were recognized faster than the same token when the repetition was edited out. In fact, though the items, subjects, and languages were different, the differences between conditions were nearly identical: 53 ms in the English experiment and 51 ms in the Dutch experiment.

With respect to artificial repetitions, results are unexpected. Targets were recognized just as quickly after artificial repetitions as they were in their original, non-repeating versions. This means that repetition alone is not driving the comprehension advantage observed in the earlier experiments.

But the phenomenon also cannot be ascribed solely to a detrimental effect of editing, because the targets following artificial repetitions were recognized at the same speed as the matched targets not preceded by a manipulation. The artificial repetitions were often noticeably edited and so should have caused a comprehension decrement if editing were driving the effect. This result contrasts with studies mentioned earlier that found disruptive effects of manipulated speech. But these studies were done with read speech, not spontaneous, and so may have been more susceptible to manipulations that changed temporal structure. The possibility remains that there is an editing decrement that is exactly offset by a repetition advantage. However, further data presented under General Discussion of Repetitions suggest that this possibility is unlikely.

One explanation for the different effects of spontaneous and artificial repetitions is that spontaneous repetitions occurred at a point in speech where repeating was useful, whereas artificial repetitions did not. Speakers might repeat when their message is not clear. So repetitions might provide an advantage to listeners only at positions where they naturally occurred. In Experiment 4, artificial repetitions were added where there had been no repeti-

TABLE 7

EXPERIMENT 4: SUMMARY OF MEAN REACTION TIMES ACROSS SUBJECTS AND ITEMS (IN MILLISECONDS)

	With repetition	Without repetition	Difference
Spontaneous repetitions	497	551	−54
Artificial repetitions	493	480	13

tions. In these cases, the extra information of confirming what was heard may have been superfluous.

In Experiment 5, the relative contribution of the location of the repetition to the overall effect was measured by creating artificial repetitions at points where repetitions had occurred spontaneously. These repetitions will be referred to as *reconstructed* repetitions, to contrast them with spontaneous and artificial repetitions. Reaction times to targets after reconstructed repetitions were compared with those after artificial repetitions to see if they showed the same effects. Reconstructed repetitions were made by copying words from the edited versions of the originally repeating utterances in the same way that words in the originally non-repeating utterances were copied to make artificial repetitions. But though the reconstructed repetitions are just as edited as the artificial repetitions, they had the advantage of occurring at the point in speech where the speaker had originally uttered a disfluency. In addition to creating reconstructed repetitions, the artificial repetitions were re-edited to make the editing less noticeable. This was done to test if the repetition effect can be obtained when there is less influence of editing.

Experiment 5

Method

Subjects. Subjects were 30 people from the introductory psychology course at Stanford University who participated for course credit or for $7 pay.

Materials. The stimuli were constructed from the same utterances used in Experiment 3. As before, each of the critical trials appeared in an edited and unedited version. The four conditions are presented below (artificial and reconstructed repetitions are capitalized, and target words are italicized):

(20) Condition 1, Repeating → non-repeating → repeating:
Where were you when you WHEN YOU *heard* this news?

(21) Condition 2, Repeating → non-repeating:
Where were you when you *heard* this news?

(22) Condition 3, Non-repeating:
Where was he when he *heard* the news?

(23) Condition 4, Non-repeating → repeating:
Where was he when he WHEN HE *heard* the news?

Conditions 2 and 3 were identical to those in Experiment 4. Condition 1 repetitions were now reconstructed as opposed to spontaneous. Condition 4 contained re-edited versions of the Condition 4 stimuli from Experiment 4. Condition 1 stimuli were created by editing Condition 2 stimuli in the same way that the Condition 4 stimuli were created by editing Condition 3 stimuli. So, Condition 1 stimuli contained two tokens of the first occurrence of words from the spontaneous repetitions, as the second occurrence had previously been excised. This first occurrence was copied and inserted just before the target word. The originally uttered pauses that occurred between repetitions when the disfluency was first uttered served as the pauses between words in the reconstructed repetitions in most cases. Condition 4 stimuli were re-edited to improve editing quality. Some of the adjustments made were to increase or decrease intervening pause length, to choose more breathy static for the intervening pause, and to add slightly louder static immediately preceding or following an edit juncture, such as after the first *g* in *a big a big mistake.*

Design. The design is described in Experiment 4.

Procedure. The procedure is described in Experiment 4.

Results

One subject was replaced because he misunderstood instructions. Of a total of 1200 possible responses, 4.8% of the repetitions (58/1200) were null responses, and 4.4% of the repetitions (53/1200) were outliers. In addition, 3% of the filler trials were false positives (40/1200). There was no significant difference across conditions for the error rates.

There was no difference between conditions for either the reconstructed repetitions or the ar-

TABLE 8
EXPERIMENT 5: SUMMARY OF MEAN REACTION TIMES ACROSS SUBJECTS AND ITEMS (IN MILLISECONDS)

	With repetition	Without repetition	Difference
Reconstructed repetitions	552	543	9
Artificial repetitions	495	496	−1

tificial repetitions (reconstructed: $F1(1,29) = .68$, ns, $F2(1,19) = .51$, ns; artificial: $F1(1,29) = 0$, ns, $F2(1,19) = .02$, ns). Table 8 summarizes the results.

Discussion

The location of the repetition was not causing the discrepancy between spontaneous repetitions and artificial repetitions in Experiment 4. It does not matter whether the repetitions occurred at natural points of uncertainty or were artificially spliced in where they had never occurred; reaction times to targets following the repetitions were the same in both cases. So it is not the case that the repetition advantage was found only in the originally repeating utterances because those repetitions occurred at particular points in the utterance that were unclear, and so could best benefit from repeating. Likewise, the lack of repetition advantage for artificial repetitions in Experiment 4 was not a result of their being created in an originally repetition-free part of the speech stream. Also, the better editing of the artificial repetitions did not induce a repetition advantage.

The initial hypothesis that repetitions aid comprehension by confirming what was heard and allowing memory to be purged was incorrect.

GENERAL DISCUSSION OF REPETITIONS

Although it is not clear whether or not repetitions aid comprehension, it is clear that repetitions do not hinder comprehension. Though they disrupt fluency and syntactic coherence, they do not slow word monitoring times. In Experiment 1, the targets after spontaneous repetitions were recognized faster than when the second occurrences of words in the repetitions were excised and replaced with pauses. In Experiment 2, the targets were still recognized faster when the repetitions were not replaced by pauses. The results of Experiment 2 ruled out the possibility that disruptive pausing caused the effect in Experiment 1, rather than advantageous repeating. It also ruled out the possibility that repetitions were a hindrance to comprehension. In Experiment 1, the hindrance of repetitions may have been masked by a greater hindrance from pauses, but Experiment 2 showed that this was not the case.

In Experiment 4, there was no repetition advantage when artificial repetitions were inserted into originally smooth stretches of speech. However, the artificial repetitions also did not cause a disadvantage. In Experiment 5, the repetition advantage was found to be a chimera; the advantage cannot be induced by merely repeating words. Even when a repetition was inserted at a point in speech where a repetition had originally occurred, the advantage was gone when editing had taken place.

One explanation for the data is that there is a repetition advantage counteracted by an editing disadvantage. This explanation predicts an effect size that changes based on severity of editing. This prediction was tested by correlating the difference between conditions in Experiment 4 and 5 with the severity of editing. If this explanation is correct, then when the editing is less noticeable, there should be a repetition advantage, but when the editing is more noticeable, there should be an editing disadvantage.

Thirty-two people judged whether or not they thought the repetitions in the critical trial utterances were spliced or otherwise artificially manipulated. Sixteen people listened to the critical trials of Experiment 4, which contained natural and artificial repetitions. Sixteen other people listened to the critical trials of Experiment 5, which contained reconstructed and better-edited artificial repetitions. Each item's mean difference between reaction times was correlated with its severity of editing score, which was the number of people who judged the item as spliced minus the number who judged it as spontaneous. This judgment experiment differs from the judgment experiment reported earlier (see Experiment 1: Detectability of Editing) in that subjects were told in advance to concentrate on repetitions.

For the Experiment 4 stimuli, 66% of the artificial repetitions were accurately judged as spliced (210 accurate detection out of 320 possible detections). But, 15% of the spontaneous repetitions were also judged, incorrectly, as spliced (48/320). In addition to the misidentified experimentally relevant repetitions, there were 29 non-experimental repetitions that subjects thought had been spliced. There were 11 repetitions that happened to be in the stimuli that were not a focus of the experimental manipulation, making 176 possible misdetections. This is an error rate of about 16% (29/176). The exact calculation is impossible to determine because some subjects treated non-repeating phrases as repetitions. For example, one subject thought the phrase *ten dol- ten* in *These are uh ten dol- ten for four-fifty* was a spliced repetition, even though *ten dol- ten* is not strictly repeating, and so not a possible response.

For the Experiment 5 stimuli, 53% of the artificial repetitions were correctly judged as spliced (156/320). The drop in percentage reflects the better editing of the stimuli. About the same percentage, 49%, of the reconstructed repetitions were correctly judged as reconstructed (169/320). Like the Experiment 4 stimuli, there were also 27 mis-judged repetitions from among the other repetitions that happened to be in the utterances. This is a rate of 15% (27/176).

In Experiment 4, there was a significant correlation between the severity of editing and the mean difference in reaction times between conditions for the artificial repetitions ($r = -.63, p < .01$). As more people noticed that an artificial repetition had been edited, the repetition advantage disappeared and the editing disadvantage gained prominence. But in Experiment 5, there was no trend whatsoever for either the artificial repetitions or the reconstructed repetitions to show either a repetition advantage or an editing disadvantage ($r = .02$, $p =$ ns, artificial; $r = .37$, $p =$ ns, reconstructed). Of course, the stimuli in Experiment 5 were better edited than those in Experiment 4, so it may be that the stimuli were just too similar in editing quality to show any effect. This explanation entails that the repetition advantage in Experiment 5 was offset in every case by an exactly balancing editing disadvantage. But this is hard to maintain in light of the graded effect in Experiment 4, where the repetition advantage was not offset in every case by an editing disadvantage. This quandary weakens the repetition-advantage/editing-disadvantage explanation.

Another explanation for the data is that all the effects are a result of either a preservation or disruption of the local phonological phrase, or the smallest phrasal unit including the target word. In the following summary of the five conditions tested, the label *x* marks where the offset of a word originally preceded the target, preserving the phonological phrase. The label *y* marks where the offset of a word did not originally precede the target, disrupting the phonological phrase.

(24) Spontaneous Repetition: our own[y] our own[x] medical
(25) Spontaneous Repetition With Repetition Excised: our own[y] medical
(26) Reconstructed Repetition: our own[y] our own[y] medical
(27) No Repetition: our own[x] medical
(28) Artificial Repetition: our own[x] our own [x] medical

The repetition results can be interpreted in the following way. Because the phonological phrase is disrupted in (25), recognizing *medical* takes longer than in (24). But in (25) and (26), the phonological phrase is disrupted in both, so there is no difference between conditions. In (27), the local phonological phrase is intact, so there is no difference between (27) and (28). In this account, the effects of editing severity can be explained as better or worse disruption of the phonological phrase.

There is another way to view what I am calling phonological disruption, and that is that the *y* above marks the location of *editing signals*, or "phonetically recognizable" indicators of self-correction (Hindle, 1983, p. 125). Editing signals before targets might slow responses either by being disruptive or by flagging listeners to listen more carefully for an upcoming error. This approach is an unlikely explanation of the phenomenon, however, because there is no psychological evidence that editing signals exist. Lickley, Shillcock, and Bard (1991) found that listeners did not detect a disfluency at the point of interruption (*y*), but at a point later in the speech stream. In the present research, subjects could not reliably hear where speech had been digitally edited (see Experiment 1: Detectability of Editing).

Arguments against a Prosodic Emphasis Explanation for the Effects

One conjecture is that all the effects described here are caused by differing effects of prosodic emphasis. For example, it could be that when we repeat, we emphasize the repetition. Emphasized words might be understood faster, leaving us more attentional capacity to recognize upcoming information, which translates into faster reaction times after naturally produced repetitions. Neither artificial repetitions nor reconstructed repetitions contain this emphasized repetition, and so the advantage would be lost in these conditions. The prosody argument could likewise be made for fresh starts; perhaps those fresh starts at the beginning of sentences are more prosodically emphasized than those in the middle, and so they are responded to more quickly, counteracting any potential disruptive effect. Middle fresh starts, lacking this advantage, show the natural false start disruption. I have two arguments against this prosodic emphasis explanation.

The first argument is that if such prosodic emphasis were affecting word monitoring, then it should also affect detectability of editing. Prosodically emphasized repetitions and fresh starts should show a marked contrast with immediately preceding material in edited conditions, and consequently editing should be easily detected. Consider the hypothetical emphasis in the fresh start *she asked if they- HOW long they lived. How* should stick out in the edited version, *she asked HOW long they lived,* causing listeners to detect an edit. This was not the case; listeners could not reliably detect where editing had occurred (see Detectability of Editing above). Likewise the hypothetical emphasis in the repetition *in the shape of a OF A HEART* should cause greater discontinuity in the edited version *in the shape of a HEART,* at least for those cases where the emphasis continued to the target words and did not dissipate with the repetition. Listeners did not reliably detect such a discontinuity.

The second argument is that other researchers have found that prosodic emphasis could not be used to reliably identify when a lexical repair was made. Levelt and Cutler (1983) found that 55% of repairs involving erroneous words were not prosodically emphasized. Cutler (1983), in a different corpus, found that 62% were not prosodically emphasized. Repairs involving restructuring the speech but not necessarily making lexical changes, or *appropriateness repairs*, are even less likely to be prosodically emphasized—81% of them were not marked (Levelt & Cutler, 1983). Repairs involving phonetic errors were never prosodically emphasized (Cutler, 1983). Lexical and appropriateness repairs are analogous to the currently defined false starts. Levelt and Cutler did not consider repetition in their analyses, but by analogy to the other results, we could expect anywhere from 55% to 100% of repetitions to not have prosodic emphasis. Because most disfluencies are not prosodically emphasized, prosodic emphasis is unlikely to be the cause of the disfluency effects found here.

CONCLUSIONS

Because disfluencies are generally unwanted in speech and give the impression of adding confusion, intuitive hypotheses lead to the expectation that disfluencies have a negative effect on comprehension. But contrary to expectations, disfluencies are not always detrimental to comprehension. Identifying words in a speech stream does take longer after most false starts, but identifying words is not hindered at all by repetitions. Even in noticeably edited stimuli where repetitions have been inserted into a speech stream, there is no effect on word monitoring.

Studying disfluencies provides information not only about the effect of disfluencies on comprehension, but also about what type of comprehension models are plausible. For example, the current results argue against any process consisting of setting up a syntactic frame and then filling out the empty slots in the frame. In this scenario, listeners might hear the preposition *of* and create the syntactic frame preposition–determiner–(adjective)–noun, and map subsequent information to this frame. But if this were truly happening, then all repetitions and false starts should slow comprehension, because they disrupt syntactic frames.

Instead, the disfluency results suggest that disrupting syntax will only be a disadvantage as far as it is a burden on the repair process. If syntax is altered mid-sentence, the repair process is more costly because the information prior to the false start must be held in memory or recalled, while the false start is identified and aborted, and the fresh start is attached. Altered syntax at the beginning of a sentence does not have the same effect because listeners can abort the false start without simultaneously holding anything in memory. Repetitions may not cause a comprehension disadvantage because they may not engender the same memory-loading repair process as false starts do. Repetitions preserve syntax and semantics, and may be recognized as being the same as something just said, and not requiring any adjustment. The claim of a fundamental difference between repetitions and false starts is supported by informal observation that transcribers miss more repetitions than false starts. The lack of a negative repetition effect also shows that listeners do not automatically enter a repair mode when they hear an incongruity. If the incongruity consists of words identical to something just heard, the repair process is inactive. The process may only begin when the incongruity consists of different words, as in a fresh start.

The repetition results also led to the occlusion that phonological phrases are at least as important in speeding word monitoring times as false starts are in slowing word monitoring times. In all repetition conditions studied, if phonological phrases were disrupted, word monitoring was slowed. More research is needed to determine precisely what the relative contribution of disfluencies, phonological phrases, and syntactic disruptions is to speech comprehension.

The studies described here show that speech disfluencies are not simply stumbling blocks to comprehension. Instead, different disfluencies have different effects. False starts in the middle of a sentence cause more processing trouble than false starts in the beginning, and repetitions cause no trouble at all. Although disfluencies have not been shown convincingly to aid comprehension, they do aid in the process of building a language comprehension model. The effects of disfluencies reflect how language as a whole is processed. For example, if words are processed one by one, incrementally building up syntax, then disfluencies should always distract listeners. Though the answers to how comprehension does operate is still open, disfluencies will probably play a role in constraining possible models.

Appendix A: Dutch Stimuli

Transcript notes for all appendices:

1. Edited-out phrases are capitalized.
2. Target words are underlined.
3. English translations are provided following the original Dutch stimuli.
4. Words in brackets indicate grammatically necessary words omitted from speech.

False Starts

1. 'T 'IS 'N uh ja 't <u>maantje</u> heeft 'n uh 'n soort neus zou je kunnen zeggen, 't lijkt op 'n neus

—IT IS A uh yeah the little moon has a uh a sort of nose you could say, it looks like a nose

2. m de eerste IS 'N die <u>heeft</u> vier punten
—m the first one IS A it <u>has</u> four points
3. en 't volgende figuur DAT HEEFT 't <u>lijkt</u> 'n beetje op 'n uh op 'n hamer
—and the next figure THAT HAS it <u>looks</u> a little like a uh like a hammer
4. nou op die contactlens zijn in 't midden twee uh twee uh boogjes getekend die beginnen in 't midden van de contactlens DIE GAAN DIE IS de <u>voorwerp</u> is symmetrisch
—now on this contact-lens there are two uh two uh bows drawn in the middle which begin in the middle of the contact-lens THEY GO IT IS the <u>shape</u> is symmetrical
5. en de bol van die vaas die lijkt dan 'n beetje op 'n vissekom EN JE MOET uh die <u>handvaten</u> zijn dus geen ringen maar gewoon uh je kan je kan d'r aan vast houden
—the round part of the vase looks a little like a fishbowl AND YOU HAVE TO uh these <u>handlebars</u> aren't rings but ordinary ones uh you can you can hold them
6. um ALS JE DIE je <u>slaat</u> die ster dan dus uh horizontaal doormidden en dan krijg je daar dus 'n plat uh vlak zeg maar
—um IF YOU THAT you <u>turn</u> the star then so uh horizontal through the middle and then you get out of it a uh level plane you could say
7. figuur vier dat is eigenlijk de bovenkant van een beitel 't steeltje dus de steel dus dus niet helemaal te zien maar de de beitel de beitel wel dus de linkse kant daarvan WORDT <u>vormt</u> 'n soort punt
—figure four that is actually the top of a beatle the handle not the handle well well not so easy to see but the the beatle the beatle is so the left side of it BECOMES <u>forms</u> a point
8. de eer- het eerste plaatje is 'n trechter met 'n uh puntje uh 't onderste 't uiteinde IS 'N ja 't <u>loopt</u> 'n beetje uit in 'n in 'n punt
—the fir- the first picture is a funnel with a uh little point uh the bottom-most it goes out IS A yeah it <u>goes</u> in a little out in a in a point
9. twee is AAN DE LINKERKANT VAN dit <u>figuurtje</u> wat op een vlinder lijkt zit aan de linkerkant
—two is ONE THE LEFT SIDE OF this-<u>figure</u> which looks like a butterfly is on the left side
10. 't elfde voorbeeld is erm IN DE 'n <u>windmolentje</u> met drie s- drie schoepen zeg maar
—the eleventh example is um IN THE a <u>windmill</u> with three s- three blades you could say
11. um tien is uh 'n 'n soort van uh ja uh s ALS JE D'R je <u>kunt</u> d'r 'n rondje omheen tekenen en dan haal je d'r zo drie dingen uit
—um ten is uh a a sort of uh yeah uh s IF YOU TAKE IT you <u>can</u> draw a circle out of it and then you get three things out of it
12. uh negen is 'n ja twee driehoekjes op elkaar uh 't eerste driehoekje HEEFT 'N uh de <u>bovenkant</u> is boven of de ja vlakke kant is boven
—uh nine is a yeah two triangles on top of each other uh the first triangle HAS A uh the <u>top side</u> is on top or the yeah the flat side is on top
13. en 'n handvat uh waar normaal 'n spiegel zit daar uh komt 't weer 'n soort kartels DAAR GAAT 'T erm <u>lijkt</u> net of 't van boven s twee oren heeft eigenlijk
—and a handle uh where there's normally a mirror, there uh is again a sort of notches AND IT GOES um <u>looks</u> a little like it has two ears on top actually
14. en 't twaalfde mja DE <u>rechtsonder</u> is rond
—and the twelfth one myeah THE <u>on the right side</u> it's round
15. 't elfde IS 'N uh de <u>bovenkant</u> is rond
—the eleventh IS A uh the <u>top</u> is round
16. als je 'm gewoon hebt IS IT 'N uh ja de <u>onderkant</u> is plat
—if you hold it simply IT IS A uh yeah the <u>bottom side</u> is flat
17. uh op 'n ander figuur zie je als je de lijnen helemaal doortrekt d- k- dan krijg je 'n uh cirkel, maar die cirkel heeft uh drie ui- inhammen en uh 'T ZIJN uh ze <u>eindigen</u> alle drie met spitse uiteinden

—uh on another figure you see if you drag the lines all the way through d- k- then you get a circle, but the circle has uh three ou- inlets and uh THEY ARE uh they end all three in sharp points

18. op 'n ander plaatje zie je 'n op de kop staande driehoek en uh DIE LOOPT uh erm zie je twee kromme lijntjes aan de linker- en de rechterkant zie je naar beneden lopen
—on another plate you see an upside down triangle and uh IT GOES uh um you see two crooked lines on the left and the right-hand side, you see them going downwards
19. 't vierde is 'n rondje MET erm waaruit uh drie andere rondjes uit zign gehaald dus dan krijg je d drie haakjes [zo] ziet het eruit
—the fourth one is a circle WITH um out of which uh three other circle are drawn from so then you get three blades [that's] what it looks like
20. ja nummer een DAT IS erm lijkt op 'n vliegende schotel met aan de linkerkant 'n E d'r aan
—yeah number one THAT IS um looks like a flying saucer with on the left side an E drawn on

Repetitions

1. en daar boven uit uh komen twee uh ja net wormpjes in feite, die dan uh links en rechtsom 'n beetje gebogen naar mekaar toe uh staan in de vorm van 'n VAN 'N hartje
-from the top two little worms are coming out in fact, which are a little bent on the left and right towards each other in the shape of a OF A heart
2. figuur elf dat 's net 'n uh NET 'N broekje
—figure 11 that is just a uh JUST A pair of pants
3. eh en daardoor gaan de uitstekende delen wat lijken op uh op messen aan 'n AAN 'N cirkelzaag of zo
—eh and because of that the protruding parts appear to look uh like knives on a ON A circular saw or something
4. (replaced in Experiment 2) plaatje drie uh ja doet me denken aan 'n uh AAN 'N schertsbaard met met wat piekerige haren
—picture three makes me think of a OF A fake beard with with prickly hair
5. zeven heeft aan de benedenkant 'n hele strakke rechte lijn en ja m 't lijkt op 'n OP 'N pot of zo waar 'n stuk twee stukken uit zijn
—seven has on the bottom a very strong right line and yeah hm it looks like a LIKE A jar or something with a piece two pieces coming out
6. figuurtje zes is rond van boven uh naar beneden toe en daar komen spitse uh ja figuurtjes uit 'n soort 'N SOORT ster maar dan voor de helft
—figure six is round from the top uh to the bottom and there sharp uh yeah little figures come out in a sort of A SORT OF star but only half of it
7. nummer tien lijkt wel op 'n boemerang in de lucht waaraan d'r aan 't rechter RECHTER gedeelte 'n stuk uitgesneden is
—number ten looks probably like a boomerang in the air where on the right RIGHT side a piece is cut out
8. 't heeft wel iets weg van 'n bijl en dat bestaat dan uit uh ja uit drie van die uh VAN DIE deeltjes van 'n bijl
—it has something like an axe and that is composed of uh yeah of three of the OF THE parts of an axe
9. 't onderste gedeelte loopt uit in 'n IN 'N punt of in 'n steeltje
—the bottom-most part goes out in a IN A point or in a prickle
10. en 't binnenstukje dat lijkt op 'n uh op de vorm van 'n VAN 'N schild van 'n wapen
—and the inner part that looks like a uh like the shape of a OF A shield or a weapon
11. 't achtste figuur bestaat uit allemaal kromme lijnen erm die zijn die zo met elkaar verbonden zijn dat er uh stekels onstaan aan de AAN DE bovenkant is 't figuur wat breder meer ja vierkanter
—the eighth figure is completely made out of crooked lines um they are they are so wrapped in each other that uh uh prickles stand up. On the ON THE top the figure is a little broader more yeah square
12. 't zevende voorbeeld bestaat uit uit aan de aan de recheterkant 'n ovaal met 'n hapje

d'r uit en aan de linkerkant 'n stukje van 'n 'n kromme 'N KROMME lijn
—the seventh example is made of on the on the right side an oval with a bite taken out and on the left side a piece of a a crooked A CROOKED line

13. 't tiende voorbeeld is 'n figuurtie in de vorm van 'n oog met aan de onderkant 'n stengeltje met waar aan WAAR AAN de rechterkant twee scherpe puntjes pinnetjes zitten
—the tenth example is a figure in the form of an eye with on the bottom a little stem with where on WHERE ON the right side two sharp points pin-points are
14. um elf uh lijkt 'n beetje op 'n uh OP 'N vossekop
—um eleven uh looks a little like a uh LIKE A fox head.
15. en rechts uh rechtsonder is 'n halve IS 'N HALVE cirkel
—en right uh bottom-right is a half IS A HALF circle
16. uh links in 't midden en rechts komen uitsteeksels naar boven van [wat] is eigenlijk 'n half rondje op 'n plat OP 'N PLAT vlak
—uh on the left side in the middle and on the right projections are coming out from [what] is simply a half circle on a level ON A LEVEL plane
17. uh 't voorwerp in een erm lijkt op 'n erm OP 'N wijnglas waar de onderkant van afgebroken is
—uh the object is a uh looks like a um LIKE A wineglass with the bottom side broken off
18. 't tweede lijkt op 'n uh ja op zo 'n klauwhamer alleen de de voorkant dus waarmee je slaat is 'n beetje 'N BEETJE raar uitgevallen
—the second one looks like a uh yeah like a sort of clutch hammer only the the front part with which you bang is a little A LITTLE strangely shaped
19. uh figuur vijf lijkt [aan] de onderkant op 'n uh erm neus van 'n erm uh NEUS VAN 'N leeuw of van 'n van 'n uh tijger
—uh figure five looks [on] the bottom like a uh um nose of a um uh NOSE OF A lion or of a of a uh tiger
20. 't linker figuurtje is kleiner als je nou 'n hockeystick hebt zeg maar 't onderste gedeelte van die hockeystick waar je mee slaat is 'n krul en 'n stukje steel zou je kunnen zeggen alleen die krul is veel dikker je ziet niet JE ZIET NIET echt de krul d'r in alleen 'n beetje ja 'n beetje alleen veel verdikt
—the left little figure is smaller as if you now had a hockeystick you could say the bottom part of the hockeystick with which you hit is a curl and a piece of stick you could say only the curl is much fatter you don't see YOU DON'T SEE really the curl in it only a little yeah a little only fatter
21. (only in Experiment 2) en aan de rechterkant zie je dan erm meer de platte kant van uh VAN 'n hamer
—and on the right side you see then um more of a flat side of a OF A hammer

Appendix B: English False Starts

False Starts at the Beginning of a Phrase

1. if I buy this will you guarantee that they will actually work? or I mean WILL- is there any chance they'll fail?
2. and THE LADY the interviewing continually puts Kevin on the spot
3. then she says well can you guarantee them? he says no I can't guarantee them. I CA- nobody can do that
4. and then he asked what length of ant she would like and um he recommended that she buy a medium ant and so she thought that a medium ant would be just fine and um then HE BROUGHT- he put out a little ring and he dumped some me- medium sized ants out onto the ring
5. you could get a little ant swing and an ant ladder and IT COULD EVEN- there's a little thing where the ant can climb up a pole and ring a bell
6. and um HE ASKS HER she wants to know what what he has
7. they're discussing her broken leg and she says well I'm busy paying off my leg and she goes he goes what do you mean pay off? AREN'T YOU ON isn't the government paying for yours?
8. he went on to say well what do you get for your taxes? and SINCE- THIS- I believe

the United States has one of the highest tax rates in the world.

9. but wait there aren't any instructions on the box, how am I supposed to use them? and HE SAID WELL YO- YOU TAKE ON- Y- *she* said am I supposed to take them with water?
10. he was playing Pictionary with some friends I guess or some girlfriend and HE SAID THAT- *then* she asked him well di- did it upset you?
11. the female who's name was Doris makes a- just a- just a sexual come-on to George just to try get his attention and just trying to start things so THEY START FI- basically from that they just start talking about their present situations
12. and he says of course I understood I JUST- isn't that an odd way to start a conversation?
13. this story is about two people Doris and George and uh THEIR- THE BEGINNING the first sentence is that they are married for some amount of years I forget the exact amount
14. and then he says my goodness are there any other countries um that also have um such liberties as the United States? and um and then I BEL- U- either it's Alfie or Jane that- that I believe it's Jane that says that Afghanistan and the Congo also have um high taxes
15. and there was a Afghan and it was playing with a flitbat and THE AFGHAN- the man commented that he thought the Afghan had a special appeal for the woman
16. so he tells her that that's a good idea because the five cent ants are a little mangy and um THEY BEGIN TO DISC- she asks him about uh keeping the ant
17. this lady walks in to the gift department of a um uh department store and GOES UP- she wants to buy and ant
18. and she says just not a good judge of character of whether a person will be a good president if he uh does have affairs and uh IN DOING THAT HE WAS RE- O- after that he responded to the question that she had made about whether um Mrs. Hart was really a token
19. he can go in the house and- and look at something and know exactly how to do it uh change it to their tastes and needs um so HE SAID THAT- he asked if he's a better interior decorator and she agreed
20. she says well are these the only ones that you have? I NEED you know I WANNA if I'm going to buy contraceptives I'm gonna buy them you know if they guarantee that they'll work

False Starts in the Middle of a Phrase

1. she questions WHETHER USING- if Gary Hart has used his wife in order to be elected
2. he made some comment about her political involvement I'm not quite sure what it was but she came back with THE- so far he had turned down sex and politics and would he like to try for religion?
3. she asked IF THEY- how long they lived and and he said they didn't
4. and that um that was this- a benefit of OWNING- having ants as a pet because uh you had a great variety of little companions all the time
5. anywhere from five cents ants ten cent fifteen cents and for a really champion one YOU CAN- it's gonna be twenty cents
6. and she says well THAT'S- we have to pay our pay our own.
7. and Mary Lou talks about how SHE WANTS- they've been going out together for two years and it takes at least two years to know whether or not you want to spend the rest of your life with somebody
8. and then she asked him IF THERE- does he have any more expensive ones?
9. and um right now they're they're in a hotel room and she walks in, gives him a peace sign and um walks in in a leather jacket and gives him a peace sign and SAYS- asks him if he wants to hop in the sack
10. people have more money there and I'm a business manager and um he says WHY- but what about you?
11. the man told her that if her ant Afghan

should die prematurely that- just to ring him up and he'd PUT- <u>send</u> a couple in an envelope

12. then he asked HOW- <u>what</u> she should feed the ants and he said you didn't feed them
13. she asked him where he where he came from, if it was in Connecticut and uh he said no he'd moved to Los Angeles cause he couldn't stand uh the snow and PICKING AT THE- <u>scraping</u> the windshield with his credit card
14. and he said that he was upset for about twenty minutes and she said well is that all? and he's like I- <u>that</u> was the beginning of the bad week for me
15. the uh salesman asks asks her WHAT WOULD YOU- <u>can</u> I help you? She says well I'd like to buy an ant
16. he said that um that he really you know BLAME- <u>thinks</u> the press was sortof at fault for uh what happened
17. and she said will when Gary Hart WAS- <u>took</u> the nomination for the president he was standing next to his wife Lee Ha- wife Lee Hart
18. and she realized she needed to go back to school when Harry's um boss WAS HOME- <u>came</u> over for dinner
19. and so a lot of people WERE- <u>objected</u> to the fact that uhm he was using her as a prop you know like he's a family man he's got a wife
20. and he says no I can't guarantee 'em and she said well I ONLY WANT- <u>you</u> seem to have all different kinds of um tablets

Appendix C: English Repetitions

Spontaneously Repeating

1. she goes well what's the most expensive ones you have? and he says well I've got I'VE GOT <u>these</u> right here which are um ten of 'em for eight dollars
2. kay she asked the question that if a IF A <u>president</u> isn't um oh no what what did he think of the incidences surrounding his uh the information that came out about him being with Donna Rice
3. okay there's this this girl was talking to her friend Alfie who lives in WHO LIVES IN <u>England</u>
4. and she says well for instance we have the we have the moon probe and he says what's the moon probe and she says well in nineteen seventy they're going to THEY'RE GOING TO <u>put</u> men on the moon
5. they were just talking about I guess she just bought a house with her boyfriend of something and her and Johnny air- basically during the entire DURING THE ENTIRE <u>thing</u> she'll say something and Johnny Carson will make a crack
6. alright this lady walks into a the gift THE GIFT <u>department</u> of a store and goes straight to the ant counter
7. and so you know he knew HE KNEW <u>about</u> these rumors an- and heard what was going on
8. what other countries are there that are- THAT ARE <u>free</u> like yours and she says well there's Afghanistan
9. guess Jane then mentions something about private insurance and how um private insurance only pays for a small part of the of the medical bills. uh she gave some examples of her OF HER <u>daughter</u> of some of the accidents that she had
10. and he says again how he do how he thinks that it might be too expensive for him to live in a LIVE IN A <u>country</u> that has that has so much so much freedom
11. and she said no this is a free country and we pay for our own OUR OWN <u>medical</u> bills
12. and so her friend says if the if the astronaut breaks his BREAKS HIS <u>leg</u> landing on the moon does he have to pay for it?
13. by asking that question di- it showed she was believing everything that was in the Miami Herald article and that she she believed that this one THIS ONE <u>incidence</u> with Donna Rice meant that he was he was womanizing all over the country
14. and su- so so what are you going to do next and and she says uh well I'm going to uh have a new a A NEW <u>show</u>
15. and he says yes and it has of course some

ant furniture in it. it has an ant AN ANT *ladder* where you can climb where you can climb and ring a bell on the top an ant wheel and an ant ts- ant swing

16. and people like Dwight Eisenhower and Thomas Jefferson if their IF THEIR private lives had had um been explored would not have been able to contribute what they were able to contribute as presidents
17. and uh she says it'll take her like eighteen months to pay for her leg get everything paid off and she's broke she goes but I'm BUT I'M free
18. yes sh- she her initial questions um dealt with the fact well what you know where were you when you WHEN YOU heard this news and how di- how did it strike you?
19. and he said well you're not gonna YOU'RE NOT GONNA find all these answers in a classroom
20. she says that she and her boyfriend just bought a house and he says you bought a house? u- what about WHAT ABOUT getting married?

Non-repeating

1. he said what kind do you want and she said what kind do you have? and um he said well WE HAVE some that are ten dol- te- these are uh ten dol- ten for four fifty
2. and he said well what's you know what would happen IF THAT astronaut broke his leg? and she said oh well the government would pay for it cause he's on government service
3. um this is an interview between Kevin WHO WORKS FOR Gary Hart i- in his presidential campaign and I think Susie is her name
4. this is in nineteen sixty three. she said in nineteen seventy WE'RE SUPPOSED TO land on the moon
5. that was a story of a kind of a kind of a satire on democracy and freedom to a kind of play on words on freedom free to pay bills and free IN A DEMOCRATIC sense type thing
6. a woman walked into a department store and approached THE ANT counter
7. SHE ASKS about Gary Hart's his recent dealings with women and how it affects his campaign
8. okay this story is about Doris and George WHO ARE both married to different people um they start having a fling
9. he says he's sick and tired of living knee deep in snow and having to scrape the ice OFF HIS windshield with a credit card
10. um so the story picks up where Doris WALKS IN THE room wearing a leather jacket
11. and she said oh no the government will pay for that he works for the government only only us taxpayers have to pay for OUR OWN medical bills
12. and he said oh it's just on what if one of the astronauts BREAKS HIS leg while he falls on while he's on the moon
13. he felt that Gary Hart had made A BIG mistake
14. and she said she was going to school and he thought that she was A LITTLE old to be going to school
15. and there were a lot of tricks that the um tricks and toys that the ant could play with like AN ANT wheel
16. um Susan asked Kevin whether he is angered by Hart's behavior and Kevin said that he was for about twenty minutes while he was watching the news WHEN HE first heard of the scandal
17. um she's been taking part in demonstrations against the war um AND HE'S skeptical saying that's not gonna stop the war
18. and then she changed the subject and went and asked uh um how did how did he where was he WHEN HE heard the news about the the Miami um scandal?
19. and she said oh because I just broke my leg and uh I'm paying off the bills for that and IT'S GONNA BE hard for me to buy presents this Christmas
20. so he brings out a ring of three ants um a Doorshire a Cambridge and an Afghan WHO IS eating a little kindof flea or something

REFERENCES

BARD, E. G., SHILLCOCK, R. C., & ALTMANN, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perceptions and Psychophysics*, **44**(5), 395–408.

BUXTON, H. (1983). Temporal predictability in the perception of English speech. In A. Cutler & D. R. Ladd (Eds), *Prosody: Models and measurements*. Berlin: Springer-Verlag.

CLARK, H. H., EDWARDS, F. H., LITTSCHWAGER, J. C., & DORADO, J. S. (1990). *Quotation in retelling stories*. Unpublished Manuscript.

CLARK, H. H., & CLARK, E. V. (1977). *Psychology and language*. San Francisco: Harcourt Brace Jovanovich, Inc.

CONNINE, C. M., BLASKO, D. G., & HALL, M. (1991). Effects of subsequent sentence context on auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, **30**, 234–250.

CUTLER, A. (1983). Speaker's conceptions of the functions of prosody. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements*. Berlin: Springer-Verlag.

EPSTEIN, W. (1961). The influence of syntactical structure on learning. *American Journal of Psychology*, **74**, 80–85.

FAURE, M. (1980). Results of a contrastive study of hesitation phenomena in French and German. In H. W. Dechert, & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 287–290). The Hague: Mouton.

FOWLER, C. A., & HOUSUM, J. (1987). Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, **26**, 489–504.

FRAZIER, L., & FODOR, J. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, **13**, 291–325.

FRENCH, N. R., CARTER, C. W., & KOENIG, W. (1930). The words and sounds of telephone conversations. In *Bell System Technical Journal*.

HINDLE, D. (1983). Deterministic parsing of syntactic nonfluencies. *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, 123–128.

JUCKER, A. H. (1993). The discourse marker *well:* A relevance theoretical account. *Journal of Pragmatics*, **19**, 435–452.

KASL, S. V., & MAHL, G. F. (1987). Speech disturbances and experimentally induced anxiety. In G. F. Mahl (Ed.), *Explorations in nonverbal and vocal behavior*. Hillsdale, NJ: Erlbaum.

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

KOWAL, S., O'CONNELL, D. C., & SABIN, E. J. (1975). Development of temporal patterning and vocal hesitations in spontaneous narratives. *Journal of Psycholinguistic Research*, **4**(3), 195–207.

LEVELT, W. J. M. (1984). Spontaneous self-repairs in speech: Processes and representations. In M. P. R. Van den Broecke, & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dordrecht, Holland: Foris Publications.

LEVELT, W. J. M., & CUTLER, A. (1983). Prosodic marking in speech repair. *Journal of Semantics*, **2**(2), 205–217.

LICKLEY, R. J., SHILLCOCK, R. C., & BARD, E. G. (1991). Understanding disfluent speech: is there an editing signal? *Proceedings of the ICPhS* (Vol. 4, pp. 98–101). Aix-en-Provence, France: International Congress of Phonetic Sciences.

LUTZ, K. C., & MALLARD, A. R. (1986). Disfluencies and rate of speech in young adult nonstutterers. *Journal of Fluency Disorders*, **11**, 307–316.

MACLAY, H., & OSGOOD, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, **15**, 19–44.

MAHL, G. F. (1959). Measuring the patient's anxiety during interviews from "expressive" aspects of his speech. *Transactions of the New York Academy of Sciences*, **21**, 249–257.

MAREK, B. (1980). Phonological status of the pause. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*. The Hague: Mouton.

MARKEL, N. (1990). Speaking style as an expression of solidarity: Words per pause. *Language in Society*, **19**(1), 81–88.

MARSLEN-WILSON, W., BROWN, C. M., & TYLER, L. K. (1988). Lexical representation in spoken language comprehension. *Language and Cognitive Processes*, **3**(1), 1–16.

MARSLEN-WILSON, W., & TYLER, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, **8**, 1–71.

MARSLEN-WILSON, W., & WELSH, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, **10**, 29–63.

MARTIN, J. G. (1968a). A comparison of ordinary, anomalous, and scrambled word strings. *Journal of Verbal Learning and Verbal Behavior*, **7**, 390–395.

MARTIN, J. G. (1968b). Temporal word spacing and the perception of ordinary, anomalous, and scrambled strings. *Journal of Verbal Learning and Verbal Behavior*, **7**, 154–157.

MARTIN, J. G., & STRANGE, W. (1968). The perception of hesitation in spontaneous speech. *Perception and Psychophysics*, **3**(6), 427–438.

MEHTA, G., & CUTLER, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech*, **31**(Part 2), 135–156.

MELTZER, R. H., MARTIN, J. G., MILLS, C. B., IMHOFF, D. L., & ZOHAR, D. (1976). Reaction time to temporally-displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance*, **2**(2), 277–290.

MENS, L. H., & POVEL, D.-J. (1986). Evidence against a predictive role for rhythm in speech perception. *The*

Quarterly Journal of Experimental Psychology, **38A**, 177–192.

POULISSE, W. (1989). *The use of compensatory strategies by Dutch learners of English.* Doctoral Dissertation, de Katholieke Universiteit te Nijmegen, The Netherlands.

RAYNER, K., CARLSON, M., & FRAZIER, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *The Journal of Verbal Learning and Verbal Behavior*, **22**, 358–374.

REDEKER, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, **14**, 367–381.

SCHIFFRIN, D. (1987). *Discourse markers.* New York: Cambridge Univ. Press.

TYLER, L. K., & WARREN, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language*, **26**, 638–657.

VOELKER, C. H. (1944). A preliminary investigation for a normative study of fluency. *American Journal of Orthopsychiatry*, **14**, 285–294.

WANNER, E. (1980). The ATN and the Sausage Machine: Which one is baloney? *Cognition*, **8**, 209–225.

WANNER, E., & MARATSOS, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan & G. A. Miller (Eds.), *Linguistic theory and psychological reality.* Cambridge: The MIT Press.

WINGATE, M. E. (1988). *The structure of stuttering: A psycholinguistic analysis.* New York: Springer-Verlag.

(Received December 30, 1993)
(Revision received March 13, 1995)