



## Using *uh* and *um* in spontaneous speaking

Herbert H. Clark<sup>a,\*</sup>, Jean E. Fox Tree<sup>b</sup>

<sup>a</sup>*Department of Psychology, Building 420, Stanford University, Stanford, CA 94305-2130, USA*

<sup>b</sup>*Department of Psychology, Social Sciences II, Room 277, University of California, Santa Cruz, CA 95064, USA*

Received 20 September 2000; received in revised form 30 August 2001; accepted 27 February 2002

---

### Abstract

The proposal examined here is that speakers use *uh* and *um* to announce that they are initiating what they expect to be a minor (*uh*), or major (*um*), delay in speaking. Speakers can use these announcements in turn to implicate, for example, that they are searching for a word, are deciding what to say next, want to keep the floor, or want to cede the floor. Evidence for the proposal comes from several large corpora of spontaneous speech. The evidence shows that speakers monitor their speech plans for upcoming delays worthy of comment. When they discover such a delay, they formulate where and how to suspend speaking, which item to produce (*uh* or *um*), whether to attach it as a clitic onto the previous word (as in “and-uh”), and whether to prolong it. The argument is that *uh* and *um* are conventional English words, and speakers plan for, formulate, and produce them just as they would any word. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Language production; Disfluencies; Spontaneous speech; Uh, um; Conversation; Dialogue

---

### 1. Introduction

Models of speaking and listening, and of language generation and parsing, are often limited to fluent speech. But in conversation – the prototypical form of language use – fluent speech is rare. Consider the answer by a British academic named Reynard to the question, “And he’s going to go to the top, is he?”:

- (1) Well, Mallet said he felt it would be a good thing if Oscar went.

---

\* Corresponding author.

*E-mail addresses:* herb@psych.stanford.edu (H.H. Clark), foxtree@cats.ucsc.edu (J.E. Fox Tree).

This sentence, with its standard syntax and semantics, could in principle have been generated or parsed within these models. But what Reynard actually produced was this:

- (2) well, . I mean this . uh Mallet said Mallet was uh said something about uh  
you know he felt it would be a good thing if u:h . if Oscar went, (1.2.370)<sup>1</sup>

Reynard took first one direction (“Mallet said something about”) and then another (“he felt it ...”). He replaced phrases (*Mallet said* by *Mallet was*), made clarifications (marked by *I mean* and *you know*), repeated words (*if if*), and added delays (silences and *uh*). Let us call the features present in (2) but not in (1) *performance additions*.

Performance additions such as these have been viewed in three main ways. One view, promoted by Chomsky (1965), is that they are “errors (random or characteristic) in applying [one’s] knowledge of language in actual performance” (p. 3). They therefore lie outside language proper and must be excluded from linguistic theory. Under Chomsky’s influence, performance additions have been excluded from many accounts of speaking and listening as well (e.g. Ferreira, 1993, 2000; Frazier & Clifton, 1996; Kintsch, 1998; Marslen-Wilson & Tyler, 1980, 1981; Mitchell, 1994). A second but related view (e.g. Goldman-Eisler, 1968) is that although performance additions are errors, they are worthy of study for what they reveal about performance.

The third view is that at least some performance additions are genuine parts of language. One example is self-repairs (Levelt, 1983, 1989; Schegloff, Jefferson, & Sacks, 1977). When Reynard says “Mallet said” and then changes his mind, he makes his intentions clear by replacing the entire constituent with *Mallet was*. Even if Reynard’s *said* were classified as an error, his selection of *Mallet was* is not an error, and it is governed by linguistic principles (Levelt, 1983). Likewise, Reynard’s *I mean* and *you know* are conventional English expressions, so they, too, are part of language – even if they aren’t part of (1). In this view, the issue becomes: which performance additions are part of language, and which are not? And for those that *are* part of language, how do speakers formulate and produce them?

In the theory of performance we will work from (Clark, 1996, in press), speakers proceed along two tracks of communication simultaneously. They use signals in the *primary track* to refer to the official business, or topics, of the discourse. They use signals in the *collateral track* to refer to the performance itself – to timing, delays, rephrasings, mistakes, repairs, intentions to speak, and the like. By signal, we mean an action by which one person means something for another in the sense of Grice (1957). In this view, Reynard creates two sets of signals. His primary signals are represented in (1). His collateral signals are represented by many of the performance additions in (2) (e.g. *I mean* and *you know*) plus certain other features of (2). There is already much evidence for such a division of labor (Allwood, Nivre, & Ahlsén, 1990; Clark, 1994b; Clark & Wasow, 1998; Fox Tree, 1995, 1999, 2001; Fox Tree & Clark, 1997; Fox Tree & Schrock, 1999; Levelt, 1983; Smith & Clark, 1993).

<sup>1</sup> We describe the notation conventions later.

Among the commonest performance additions in English are *uh* and *um* (usually spelled *er* and *um* in British English).<sup>2</sup> *Uh* and *um* are characteristically associated with planning problems. But are they collateral signals by which speakers refer to these problems, or are they mere symptoms, or natural signs, of the problems? And if they *are* signals, are they part of language, like *I mean* and *you know*, or not part of language, like sighs and tongue clicks? We will argue that *uh* and *um* are, indeed, English words. By words, we mean linguistic units that have conventional phonological shapes and meanings and are governed by the rules of syntax and prosody. We will also argue that *uh* and *um* must be planned for, formulated, and produced as parts of utterances just as any other word is. Still, these processes are not the same for *uh* and *um* as they are for words in the primary track because *uh* and *um* are used collaterally to refer to performance problems. We begin with three common views of *uh* and *um* and then take up evidence for their status as words and for their role in spontaneous speech.

## 2. Conceptions of *uh* and *um*

*Uh* and *um* have long been called *filled pauses* in contrast to *silent pauses* (see Goldman-Eisler, 1968; Maclay & Osgood, 1959). The unstated assumption is that they are pauses (not words) that are filled with sound (not silence). Yet it has long been recognized that *uh* and *um* are not on a par with silent pauses. In one view, they are *symptoms* of certain problems in speaking. In a second view, they are *non-linguistic signals* for dealing with certain problems in speaking. And in a third view, they are *linguistic signals* – in particular, words of English. If *uh* and *um* are words, as we will argue, it is misleading to call them filled pauses. To be neutral and yet retain a bit of their history, we will call them *fillers*.

### 2.1. Three views of *uh* and *um*

In the *filler-as-symptom* view, *uh* and *um* are the automatic, or involuntary, consequence of one or another process in speaking. One characterization is this: *uh* gives evidence that “at the moment when trouble is detected, the source of the trouble is still actual or quite recent. But otherwise, [*uh*] doesn’t seem to mean anything. It is a symptom.” (Levelt, 1989, p. 484; see also Mahl, 1987; O’Donnell & Todd, 1991). This view has several problems. As we will show, speakers have control over *uh* and *um*, so they are not automatic. Also, when speakers detect trouble in speaking, they often produce items other than *uh* and *um* (Levelt, 1983, 1989). If they do, the appearance of *uh* and *um* must be conditional on other factors, and we would need to know what those factors are. The most intriguing problem is that English has at least two fillers, *uh* and *um*, and so do all other languages we have examined (see later). A priori, *uh* and *um* must have

<sup>2</sup> *Uh* and *um* are pronounced with schwas in both British and North American English. In most British dialects, *gopher* rhymes with *sofa*, so *er* and *um* are both pronounced with schwas as well. *Er* does not rhyme with *cur* or *burr*, as many North American readers of British novels assume. The London–Lund corpus of British English, on which we rely for most of our analyses, transcribes *uh* and *um* with schwas. In the *Oxford English Dictionary* (OED) (2000), a British dictionary, the entry for *uh* says “U.S. = *er*”. *Uh* is also sometimes spelled *ah* in North American English (e.g. Kasl & Mahl, 1965), and *um* is sometimes spelled *erm* in British English (e.g. Watts, 1989). We assume that all of these vowels are dialect variants of schwa.

distinct causes, just as any two options in behavior do, and we must account for the difference.

In the *filler-as-nonlinguistic-signal* view, *uh* and *um* are signals. The oldest and best known proposal is that fillers are used for holding the floor (Maclay & Osgood, 1959, p. 41):

Let us assume that the speaker is motivated to keep control of the conversational “ball” until he has achieved some sense of completion... Therefore, if he pauses long enough to receive the cue of his own silence, he will produce some kind of signal ([m, er], or perhaps a repetition of the immediately preceding unit) which says, in effect, “I’m still in control – don’t interrupt me.”

A related proposal is that fillers are elements “whereby the speaker, momentarily unable or unwilling to produce the required word or phrase, gives audible evidence that he is engaged in speech-productive labor” (Goffman, 1981, p. 293). In both proposals, fillers are signals, though not true words. They are like clearing one’s throat, which might be used to mean “Why don’t you introduce me to your friend?” or “Stay away from that topic of discussion”.

In the *filler-as-word* view, *uh* and *um* are English interjections. This view was originally proposed by James (1972), who placed *uh* alongside *oh*, *well*, *ah*, and *say* as interjections for commenting on a speaker’s on-going performance. She didn’t elaborate on the view, so let us examine what it entails.

## 2.2. Interjections

An interjection is (1) a conventional lexical form (sometimes a phrase) that (2) conventionally constitutes an utterance on its own and (3) doesn’t enter into constructions with other word classes (Wilkins, 1992).<sup>3</sup> Although interjections are sometimes defined as “purely emotive words which have no referential content” (Quirk, Greenbaum, Leech, & Svartvik, 1972, p. 413), they serve many other functions too. They are used not only to express current emotions (*ugh*, *damn*, *hell*, *bravo*, *hooray*), but also to describe current states of knowledge (*huh*, *indeed*, *oh*, *well*), especially surprise (*ah*, *aha*, *boy*, *wow*, *oops*, *gosh*, *hah*), and to request attention (*ahem*, *hey*, *yoo-hoo*) and other actions (*sh*, *whoa*, *shoo*, *enough*). They are used to greet (*hello*, *hi*), bid farewell (*bye*, *so long*, *cheers*), and carry out parts of other routines (*okay*, *thanks*, *bingo*, *checkmate*, *amen*).

### 2.2.1. Meaning

Nouns, verbs, adjectives, and adverbs are ordinarily defined with *paraphrases*. In the *American Heritage Dictionary (AHD)* (*American Heritage Dictionary of the English Language*, 2000), *boy* is defined as “a male child”, *leave* as “to go out or away from”, and *sad* as “affected or characterized by sorrow or unhappiness”. When these words are combined, so are their paraphrases. To say “The sad boy left” is like saying “The male child affected or characterized by sorrow or unhappiness went out”. Interjections, in contrast, are

<sup>3</sup> An interjection is “the most primitive type of sentence” (Curme, 1935) or a “minor sentence ... entering into few or no constructions other than parataxis” (Bloomfield, 1933). See Wilkins (1992).

defined by the *conventional practices* they are used for. In the *AHD*, *well* is defined as “used to express surprise”, *hello* as “an informal expression used to greet another”, and *ah* as “used to express various emotions, such as surprise, delight, pain, satisfaction, or dislike” (our emphases). To say “Hello” is *not* like saying “An informal expression used to greet another”, but like saying “I greet you”, reflecting the conventional practice for *hello*. If *uh* and *um* are interjections, they, too, should be defined by conventional practices.

Most interjections have many uses, making their meanings difficult to pin down. To deal with this problem, we distinguish between *basic meanings* and *implicatures*. A basic meaning of *good-bye*, for example, is “used to express farewell”. Speakers can use *good-bye* to signal other things too, but by implicature. If Ann says “good-bye” to Ben as he walks up to her, she can mean “Go away!”. In Grice’s terminology (see Grice, 1975; Horn, 1984; Levinson, 1983, 2000; Sperber & Wilson, 1986), she is saying farewell and, based on the relevance of that comment in her and Ben’s current common ground, she is implicating that she wants him to go away. “Go away” isn’t a basic meaning of *good-bye*, but an implicature of its use.<sup>4</sup> If *uh* and *um* are interjections, they, too, should have basic meanings and be useful for implicating other things.

### 2.2.2. Timing

When speakers use interjections, they make reference to “one or more of the following basic deictic referencing elements: *I*, *you*, *this*, *that*, *now*, and perhaps *here* and *there*” (Wilkins, 1992). Take *ah* in (3):

- (3) William I’m on the academic council,  
 Sam ah, very nice position (1.2b.1397)

When Sam says “ah”, according to the *ADH* (2000), he “expresses mild surprise”. But he is doing something more. He is asserting, roughly, “I am mildly surprised *now* at the information I have *just now* learned [namely, that you are on the academic council]. Each utterance of *ah* contains indices to the current speaker (I), the current addressees (you), the current moment (now), and other elements in the current common ground. The same holds for other interjections.

Our main interest is in the *temporal index* (Clark, 1999, in press). When Sam produces *ah*, he does it at a particular moment in time. We will denote his index to that moment by  $t(\text{‘ah’})$ . What Sam is asserting is, roughly, “I am mildly surprised at  $t(\text{‘ah’})$  at the information I have just learned”. The temporal index  $t(\text{‘ah’})$  marks the precise moment at which Sam wants to say that he is surprised. If he had delayed *ah* by one second, that would have changed how soon he claimed to have been surprised and therefore, perhaps, what he was

<sup>4</sup> Even many dictionary definitions are best viewed as implicatures. The basic meaning of *hello*, for example, is “used to greet someone”. Via implicatures, it can be “used to welcome into one’s home” or “used to express surprise” (*ADH*, 2000). The basic meaning of *oh* may be “used to propose that its producer has undergone some kind of change in his or her locally current state of knowledge, information, orientation or awareness” (Heritage, 1984, p. 299). Via implicatures, it can be “used to express strong emotion, such as surprise, fear, anger, or pain” or “used to indicate understanding or acknowledgment of a statement” (*ADH*, 2000).

surprised about. By hypothesis, all interjections require *t*(utterance) as part of their meaning. If *uh* and *um* are interjections, they should too.

### 2.3. Primary and collateral signals

Speakers, we assume, refer to the official business, or topics, of the discourse with primary signals, and to the performance itself with collateral signals (Clark, 1996, in press). They use the collateral signals, in effect, to *manage* the on-going performance.

People in discourse recognize the difference between primary and collateral messages, a point made by Goffman (1981) in different terminology. In an analysis of radio talk, he noted that radio announcers are expected “to produce the effect of a spontaneous, fluent flow of words – if not a forceful, pleasing personality – under conditions that lay speakers would be unable to manage” (p. 198). So when they run into problems, as they inevitably do, they often comment on them in *parenthetical asides* that correct, poke fun at, apologize for, or otherwise explain their problem. Consider (4) (p. 290):

- (4) Announcer Seventy-two degrees Celsius. I beg your pardon. Seventeen degrees Celsius. Seventy-two would be a little warm.

The announcer’s job is to report the weather, which leads to his official messages – “Seventy-two degrees Celsius” (in error) and “Seventeen degrees Celsius” (corrected). But to maintain his self-image, he inserts two unofficial messages *within* his official performance – the apology and the joke – a change in stance that both he and his audience recognize. Changes in stance are often marked by intonation or tone of voice. In this light, consider *I mean* in (5):

- (5) Sam is there a doctrine about that, - - I mean a doctrine about u:h – disfavoring American applicants, (2.6.978)

Like the radio announcer, Sam inserts a parenthetical aside (“I mean”) to comment on a problem in his official performance. With it he says that what follows (“a doctrine about disfavoring American applicants”) is what he really wants to say (see Fox Tree & Schrock, in press). We suggest that Sam inserts “u:h” for similar reasons.

The collateral signals that are added to utterances fall into four main categories (Clark, in press):

(a) *Inserts*. Inserts are parenthetical asides placed *between* elements of a primary utterance. These include: editing expressions such as *I mean*, *you know*, *that is*, *no*, and *sorry* (Erman, 1987; Levelt, 1983, 1989); certain discourse markers such as *well*, *now*, *oh*, and *like* (DuBois, 1974; Fox Tree & Schrock, 1999; Schiffrin, 1987; Schourup, 1982; Underhill, 1988); and even laughter, sighs, and tongue clicks.

(b) *Juxtapositions*. These signals are produced by *juxtaposing* one stretch of speech against another. In (2), Reynard juxtaposed “Mallet was” against “Mallet said” as a signal to replace *Mallet said* with *Mallet was*. Replacements are perhaps the commonest form of speech repair (Levelt, 1983; Schegloff et al., 1977). And in (2), Reynard repeated *if*, another common juxtaposition (Clark & Wasow, 1998).

(c) *Modifications*. These signals are produced by *modifying* a syllable, word, or phrase within a primary utterance. They include prolonged syllables and non-reduced vowels, which we take up later (Fox Tree & Clark, 1997; Koopmans-van Beinum & van Donzel, 1996), and try markers (Sacks & Schegloff, 1979).

(d) *Concomitants*. These are collateral signals produced *at the same time as* the speech they comment on but in another form or modality. They include certain head nods, eye gaze, smiles, over-speech laughter, grimaces, iconic gestures, and pointing (Bavelas & Chovil, 2000; Bavelas, Chovil, Lawrie, & Wade, 1992; Goodwin, 1981; Goodwin & Goodwin, 1986).

Most of these signals are self-evident parts of spoken language – conventional words or phrases, and features of prosody. It would be perfectly consistent for *uh* and *um* to be parts of language as well.

Interjections are used mostly as primary signals. In (3), Sam uses *ah* to comment on the topic William has just spoken about. But many interjections can be used as inserts – as collateral signals – such as *I mean* in (5). Although speakers tend to be aware of primary uses of interjections, they tend not to be aware of collateral uses (Watts, 1989). Indeed, it has taken lexicographers years to discover these functions. *You know*, *like*, and *oh* are no less words for that, and the same would hold for *uh* and *um*.

#### 2.4. *Uh* and *um* as collateral interjections

We are now in a position to state the *filler-as-word hypothesis*. It is really a refinement of the James (1972) hypothesis, although it owes much to Allwood et al. (1990), Goffman (1981), and Levelt (1983, 1989). It grew out of evidence (Smith & Clark, 1993), described later, that *uh* and *um* project further delays – *uh* brief ones, and *um* longer ones. The hypothesis, expressed in standard dictionary definitions, is this:<sup>5</sup>

*Filler-as-word hypothesis*. *Uh* and *um* are interjections whose basic meanings are these:

- (a) *Uh*: “Used to announce the initiation, at *t*(‘uh’), of what is expected to be a minor delay in speaking.”
- (b) *Um*: “Used to announce the initiation, at *t*(‘um’), of what is expected to be a major delay in speaking.”

Producing *uh* itself constitutes a brief delay, and *um*, a longer delay (according to evidence described later). If speakers are accurate in their expectations, the delays should often extend beyond *uh* and *um*, and be longer after *um* than after *uh*. *Uh* and *um* can be used for other functions too. The hypothesis is that most other functions are implicatures that follow from the relevance of announcing minor or major expected delays in the current situation.

Another way to signal a delay is to prolong a syllable. Speakers can prolong almost any

<sup>5</sup> Compare the *ADH* (2000), in which *uh* is defined as “Used to express hesitation or uncertainty”, and *um* as “Used to express doubt or uncertainty or to fill a pause when hesitating in speaking”. These definitions are based entirely on written sources – novelists’ and playwrights’ attempts to represent spontaneous dialogue (see the *OED*, 2000). They are *not* based on evidence from spontaneous speech.

syllable beyond its normal, or expected, length, and they often do. Evidence from the Chafe (1980) pear stories, also described later, leads to this hypothesis:

*Prolongation hypothesis.* Speakers prolong a syllable or its parts to signal that they are continuing a delay that is on-going at *t*(syllable).

Speakers often prolong *uh* and *um*, and the London–Lund corpus distinguishes between “u:h” and “uh” and between “u:m” and “um” in which the colons mark a prolongation of one or more segments. By these two hypotheses, the choice of filler and prolongation signal different things. A prolonged *uh*, for example, signals: (1) “I am continuing a delay that is on-going at *t*(‘uh’)”; and (2) “I am initiating, at *t*(‘uh’), what I expect to be a minor delay in speaking”. An alternative proposal is that choice of filler is explained by the prolongation hypothesis: “um” and “u:m” are simply prolonged “uh” and “u:h”.

Speakers plan utterances in three main stages: they conceptualize a message, formulate the appropriate linguistic expressions, and articulate them (Levelt, 1989). If *uh* and *um* are words, speakers must plan these too. They would conceptualize the message “I am now initiating what I expect to be a minor delay”, formulate the word *uh* to express it, and produce “uh”. The formulation process may seem trivial, but *uh* is usually inserted into an on-going utterance, as in “if u:h. if Oscar went,” and that complicates the process.

The rest of the paper divides into two main parts plus a conclusion. The first part takes up evidence for *uh* and *um* as conventional English words: how *uh* and *um* contrast in basic meanings; how they are used to implicate other things; and how they are conventional and under the speaker’s control. The second part takes up evidence about how speakers plan and produce *uh* and *um*: how they monitor for and detect imminent delays; and how they formulate *uh* and *um* as parts of on-going utterances. First, we describe our principal sources of evidence.

### 3. Corpus evidence

The primary evidence for our proposal comes from the London–Lund corpus (hereafter LL corpus). It consists of 170,000 words from 50 face-to-face conversations (numbered S.1.1 through S.3.6) from the Svartvik and Quirk (1980) corpus of English conversations. The conversations were recorded between 1961 and 1976 among British adults, mostly academics, in two- to six-person settings. Although some of the speakers knew they were being recorded, most didn’t, and we excluded those who did. Each example is identified by its line in the corpus; 1.3.334 means conversation 1.3, line 334.

The computerized transcripts of the LL corpus represent words, word fragments, fillers, pauses, tone units, overlapping speech, stress, and prosodic information such as rising, flat, and falling intonation. In the examples we cite, we retain only some of these markings. Ends of tone units are marked with a comma (,) for non-rising intonation and with a question mark (?) for rising intonation. Brief pauses “of one light foot” are marked with periods (.), and unit pauses “of one stress unit” with dashes (-). When we need a measure of pause length, we treat the unit pause as 1 unit long, and the brief pause as 0.5 units long, so “. -” is a 1.5 unit pause, and “- - -” is a 3 unit pause. Overlapping speech is marked with



matched pairs of asterisks (\*). Prolonged syllables are marked with colons (:), as in “u:m”. *Uh* and *um* were sometimes pronounced in brief or normal form, which we will write “uh” and “um”, and other times in prolonged form, which we will write “u:h” and “u:m”. The surreptitiously recorded speakers produced 3904 fillers (“uh” 898, “u:h” 1213, “um” 530, “u:m” 1263).

For auxiliary analyses, we draw on an answering machine corpus (AM corpus), the switchboard corpus (SW corpus), and the Pear stories (Pear corpus). The AM corpus consists of 5000 words in 63 calls to telephone answering machines, section S.9.3 in the full computerized version of the LL corpus. It contains only 319 fillers (“uh” 69, “u:h” 166, “um” 6, “u:m” 78). The SW corpus is a 2.7 million word corpus of telephone conversations (Godfrey, Holliman, & McDaniel, 1992). It marks *uh*, *um*, and sentence boundaries, but not prolongations or pauses; it contains 79,623 fillers (*uh* 67,065 and *um* 12,558). The Pear corpus (Chafe, 1980, Appendix) came from 20 people who each watched a dialogue-free movie about pear pickers and then recounted what happened. The transcripts include “uh”, “um”, “u:h”, “u:m” (spelled *uh*, *um*, *u-h*, and *u-m*) prolonged segments (e.g. a–nd), and pause lengths measured to the nearest 0.05 s. It contains only 282 fillers (“uh” 57, “u:h” 68, “um” 54, and “u:m” 103).

The four corpora differ in the type of speech they represent: face-to-face conversations (LL corpus); telephone conversations (SW corpus); speech to answering machines (AM corpus); and narratives (Pear corpus). They also differ in whether the speakers were British (LL and AM corpora) or North American (SW and Pear corpora). We took the LL corpus as the main source of evidence because it represents face-to-face conversation, includes both pauses and prolongations, and contains enough fillers for reliable analyses.

The LL, AM, and Pear corpora represent the judgments of trained coders who examined and re-examined each audio recording in detail. Note that prolongation, for example, is in principle a continuous variable – *uh* could vary from, say, 50 ms to 5 s. Still, the coders turned it into a binary contrast: a syllable or word was judged in its immediate context to sound either normal, or abnormally long (no matter how long). The same holds for length of pause. Although the Pear coders measured pauses in seconds, the LL and AM coders judged length relative to the speaker’s rate of speech (as “one light foot” or “one stress unit”), reflecting how long the pauses would feel to listeners. So it is ultimately the *perception* of pause length and prolongation that we are studying here. In every case the coders made their judgments far in advance of the hypotheses we consider here.

#### 4. Basic meanings of *uh* and *um*

The core hypothesis is that speakers use *uh* and *um* to signal delays. Although some delays come at grammatical junctures, many come at disruptions within utterances. Schematically, disruptions have three parts (Clark, 1996): (1) a *suspension* of fluent speaking; (2) a *hiatus* in speaking, which may contain nothing, a stretch of silence, a filler, or other collateral actions; and (3) a *resumption* of fluent speaking. If we represent the points of suspension and reduction with left and right curly brackets { and }, then we can represent the disruptions in (2) as follows:

(2') well, { . I mean } this { . uh } Mallet said { } Mallet was { uh } said something about { uh you know } he felt it would be a good thing if { u:h . } if Oscar went,

This notation makes it easy to identify patterns of pauses, delays, and fillers. By pause we mean a *silent* pause, and by delay we mean any combination of pauses and fillers. In (2'), the first *uh* is in a hiatus preceded by a brief pause; “u:h” is in a hiatus followed by a brief pause. Fillers were also sometimes followed or preceded by another filler with or without a pause, as in “um - - uh”. Fillers at grammatical junctures (boundaries of intonation units) can be classified the same way. We tallied these patterns for all fillers in the LL, AM, and Pear corpora. In our statistical tests, we used data only from those speakers who produced fillers in more than one category of filler.

#### 4.1. Delays after *uh* and *um*

By hypothesis, *uh* and *um* are used to announce the start of what are expected to be minor and major delays. Fig. 1 plots the percentages of times in the LL corpus (93 speakers) that *uh* and *um* had delays after them, and Fig. 2 plots the mean lengths of pauses after *uh* and *um*. *Um* was followed by delays far more often than *uh*, 61 to 29% of the time ( $F(1, 342) = 106, P < 0.001$ ). It was also followed by much longer pauses, 0.68 to 0.25 units ( $F(1, 57) = 51.3, P < 0.001$ ).

By the prolongation hypothesis, speakers can also signal delays by prolonging syllables. In one version of the hypothesis, speakers would buy time with the prolongation itself, and the delay would end when the prolongation ended. In another version, the prolongation would signal an additional delay as well. The percentages in Fig. 1 support the second version of the hypothesis, though not by much. There were slightly more delays after prolonged fillers than after normal fillers, 49 to 42% of the time ( $F(1, 342) = 4.92, P < 0.03$ ). There were also slightly longer pauses after prolonged than normal fillers, 0.55 to 0.38 units ( $F(1, 57) = 6.5, P < 0.01$ ).

Remarkably, prolongation (prolonged vs. normal) and filler choice (*uh* vs. *um*) are *independent* signals of delay in the LL corpus. Consider the percentages of delays in Fig. 1. The difference between prolonged and normal fillers was the same for *uh* (7%) as it was for *um* (7%). That implies that prolongation and filler choice are separate, additive predictors of delay. The point can be made statistically. Of the total variance in percentage of pauses after fillers, 96% can be attributed to the choice of *uh* vs. *um*, 4% to the choice of normal vs. prolonged, and less than 0.01% to a dependence between the two. The first two portions of the variance are significant; the third is not. The same holds for the pauses in Fig. 2. Of the total variance in pause length after fillers, 86% can be attributed to the choice of *uh* vs. *um*, 14% to the choice of normal vs. prolonged, and 0.002% to a dependence between the two. There is nothing in the coding of the LL corpus that requires filler choice and prolongation to be independent, but they are.

*Uh* and *um* differ not only in dialogues (the LL corpus), but also in monologues (the Pear and AM corpora). There were more pauses after *um* than after *uh* both in the AM corpus, 69 to 33% ( $F(1, 54) = 13.5, P < 0.001$ ), and in the Pear corpus, 90 to 76% ( $F(1, 60) = 4.56, P < 0.03$ ). There were also longer pauses on average after *um* than after *uh* in the AM corpus, 0.54 to 0.15 units ( $F(1, 73) = 11.4, P < 0.002$ ), and in the Pear corpus, 0.80 to

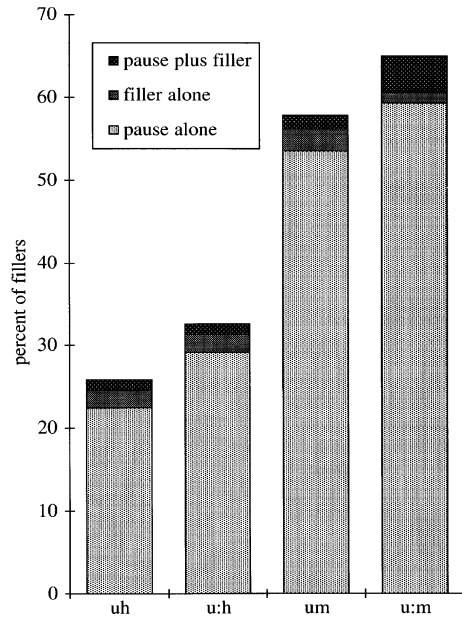


Fig. 1. Percent of fillers followed by delays (LL corpus).

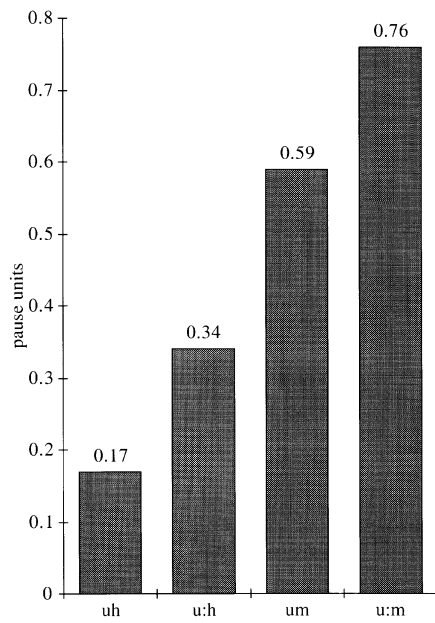


Fig. 2. Mean length of pauses after fillers (LL corpus).

0.52 s ( $F(1, 60) = 2.4$ , n.s.). These data bolster the hypothesis that *uh* and *um* contrast in signaling imminent delays. They also argue against the proposal that *uh* and *um* are simply ways of holding the floor (Maclay & Osgood, 1959). In monologues, there is no issue of holding the floor, yet *uh* and *um* were used just as they were in dialogues.

The filler-as-word hypothesis was originally suggested by a study of answering questions (Smith & Clark, 1993). Twenty-five students were each asked 40 test questions in a one-on-one conversational setting. Example:

- (6) Questioner: In which sport is the Stanley Cup awarded?  
 Student: (1.4 s) *um* (1.0 s) hockey

Although the questions ranged in difficulty, the correct answers were single words with no syntax to plan. When students had difficulty retrieving these answers, they often began their replies with *uh* or *um*, as in (6). There were significantly longer pauses after *um* than after *uh*, 4.12 to 1.00 s (for a similar result, see Barr, 1998). The students' choice of filler appeared to reflect their judgment of how quickly they could retrieve the right answer. The time interval from the end of the question (after *awarded*) to the beginning of the final answer (*hockey*) was 8.83 s for *um*, 2.65 s for *uh*, and 2.23 s for no filler. Students were able to estimate how long it would take them to retrieve the answer even before they had retrieved it.

#### 4.2. Delays before *uh* and *um*

In our proposal, speakers use *uh* and *um* to signal the *initiation* of delays. Still, speakers might delay, anticipate a further delay, and then signal the further delay with *uh* or *um*. If so, there might be delays not only after fillers, but before them. Fig. 3 plots the percentage of times in the LL corpus that *uh* and *um* were preceded by delays, and Fig. 4 plots the average lengths of pauses before *uh* and *um*. There were more delays before *um* than before *uh*, 41 to 34% of the time ( $F(1, 342) = 4.68$ ,  $P < 0.03$ ). There were also slightly longer pauses before *um* than before *uh*, 0.40 to 0.32 units ( $F(1, 57) = 2.7$ , n.s.).

If prolongations signal delays already in progress, there might be more delays, and longer pauses, just before prolonged syllables than before normal syllables. As shown in Fig. 3, there were delays almost twice as often before prolonged *uh* and *um* as before normal *uh* and *um*, 48 to 28% ( $F(1, 342) = 41.2$ ,  $P < 0.001$ ). And as shown in Fig. 4, pauses were twice as long before prolonged *uh* and *um* as before normal *uh* and *um*, 0.48 to 0.24 units ( $F(1, 57) = 22.2$ ,  $P < 0.001$ ). Of the total variance in percentage of delays before fillers (in Fig. 3), 10% can be attributed to *uh* vs. *um*, 90% to normal vs. prolonged, and less than 0.1% to a dependence between filler choice and prolongation. In the total variance in pause length before fillers (Fig. 4), these percentages are 8, 92, and 0.2%. So in the LL corpus, prolongation and filler choice are independent signals of delay *before* the filler as well.

*Uh* and *um* also differ in narratives and on answering machines. There were more pauses before *um* than before *uh* both in the AM corpus, 73 to 54% ( $F(1, 54) = 2.82$ , n.s.), and in the Pear corpus, 62 to 42% ( $F(1, 60) = 4.73$ ,  $P < 0.04$ ). And there were longer pauses on average before *um* than before *uh* in the AM corpus, 0.66 to 0.44 units ( $F(1, 73) = 5.88$ ,  $P < 0.02$ ), and in the Pear corpus, 0.51 to 0.28 s ( $F(1, 60) = 4.51$ ,  $P < 0.04$ ).

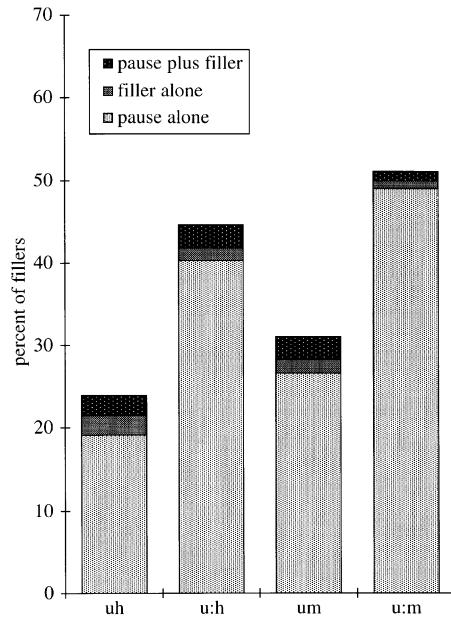


Fig. 3. Percent of fillers preceded by delays (LL corpus).

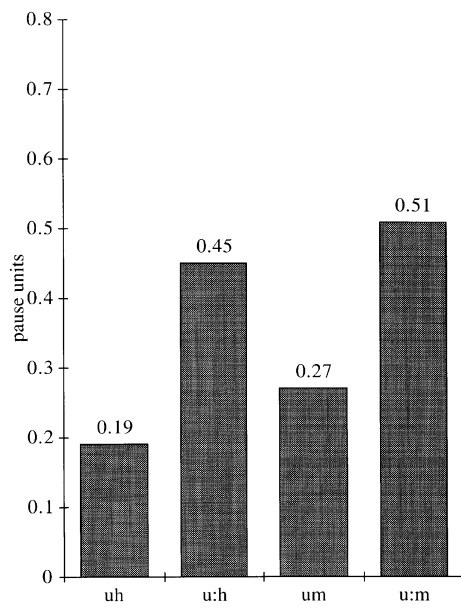


Fig. 4. Mean length of pauses before fillers (LL corpus).

### 4.3. Delays before and after *uh* and *um*

*Um* led *uh* in the number and length of delays associated with it. If *uh* and *um* signal the initiation of a delay, one might expect the lead of *um* over *uh* to be reflected more clearly in *post*-filler delays than in *pre*-filler delays. In fact, *um* led *uh* by 32% in *post*-filler delays compared to 7% in *pre*-filler delays (Figs. 1 and 3). And *um* led *uh* by 0.43 units in *post*-filler pauses compared to 0.08 units in *pre*-filler pauses (Figs. 2 and 4). The lead of *um* over *uh* accounted for 96 and 86% of the variance in delays and pauses *after* fillers, but for only 10 and 8% in delays and pauses *before* fillers. Conclusion: *uh* and *um* contrast mainly in the delays they initiate.

Just the opposite holds for prolonged syllables. Prolonged fillers led normal fillers by 20% in *pre*-filler delays compared to 7% in *post*-filler delays (Figs. 1 and 3). Prolonged fillers led normal fillers by 0.24 units in *pre*-filler pauses compared to 0.17 units in *post*-filler pauses. The lead of prolonged over normal fillers accounted for 90 and 92% of the variance in delays and pauses *before* fillers. It accounted for only 4 and 14% of the variance in delays and pauses *after* fillers. All this is evidence that prolonged *uh* and *um* signal not the initiation of delays, but the continuation of on-going delays.

Words other than *uh* and *um* can also be prolonged, and they show much the same pattern. We examined the pauses around all 360 prolonged words in the Pear corpus (excluding *uh* and *um*) and compared them with a matched set of 360 non-prolonged words ten words downstream from each prolonged word (or upstream if that was impossible). Prolonged words had pauses before them 64% of the time, but they had pauses *after* them only 38% of the time ( $F(1, 19) = 21.3, P < 0.001$ ). For words not prolonged, the numbers were 14 and 11%. These two percentages are significantly smaller than those for prolonged words ( $F(1, 19) = 100.3, P < 0.001$ ). In general, prolonging a syllable signals a delay already in progress.

So far, we have identified two contrasts. The first is between *uh* and *um*. By hypothesis, *uh* signals the initiation of what is expected to be a minor delay, and *um*, what is expected to be a major delay. In our data, *uh* and *um* differed sharply in the delays that followed them, but little in the delays that preceded them. The second contrast is in the prolongation of *uh* and *um*. The hypothesis is that prolongation signals the continuation of an on-going delay. As it happened, prolonged *uh* and *um* contrasted with normal *uh* and *um* mainly in the delays that preceded them, not in the delays that followed. These two sets of signals are independent in the LL corpus. So speakers don't consider *um* to be merely a prolonged *uh*. Selecting between *uh* and *um* is separate from prolonging the item selected.

## 5. Implicatures with *uh* and *um*

If the basic meanings of *uh* and *um* are “used to announce the initiation of what is expected to be a minor, or major, delay in speaking”, other interpretations of *uh* and *um* in context are implicatures. But are *uh* and *um* used to announce just any delay, or are they restricted to certain types? And if the other interpretations are implicatures, where do they come from? We will offer evidence from the LL corpus on both questions.

### 5.1. Delays signaled by *uh* and *um*

Many delays signaled by *uh* and *um* are caused by problems in formulating an utterance. The delays precede the problematic parts, and fillers can be used to signal the delays, as in these examples (see also Goodwin & Goodwin, 1986):

- (7) and there's too much {um you know} the sense of hollowness at the bottom, which {u:m - - - you know}, as though the earth was going to give under your feet, can't find the word (1.8.512)
- (8) unfortunately most of the things there, {.} were {u:m .} I think they were prizes (1.8.566)

In (7) and (8), the speakers are delayed, and know they are delayed, in retrieving the right word (“can’t find the word”) and in formulating the right description (“I think they were prizes”).

Other delays arise as speakers prepare to make repairs, as in these examples (see also Levelt, 1983):

- (9) we're interested in lezh {u:h} religions overseas (1.2.123)
- (10) so I went round to Jackie, {. uh .} to {uh} Pam, at the hospital (2.12.799)
- (11) {u:h} I think he's I think he's got {um -} persuaded Oscar to go to the States (1.2.343)
- (12) but apparently, you {. u:h} they spent {- uh} when it was finished, they went out there to spend a winter there (1.14.744)

In (9) through (12), speakers repair the mispronunciation of *religions*, the name *Jackie*, the verb *got*, and the phrasing of a larger construction. These delays may reflect the time needed to detect the problem, formulate the repair, or both.

Still other delays are caused by problems, at the message level, of what to say next (see Smith & Clark, 1993). These problems are clearest in responses to questions, as here:

- (13) Ann and what sort of rates do you pay for this sort of thing,  
Burton {. u:m -} well, {. u:h - - -} I'm sorry, {.} I ought to know this, I think it's about one fifty an hour (3.2a.240)

Once Ann has asked Burton a question, he is obligated to answer in a timely fashion (Sacks, Schegloff, & Jefferson, 1974). As it happens, he is unable to do that, so he signals delays twice. He is fully aware of the problem (“I ought to know this”) and even apologizes for it (“I’m sorry”). In (14), Geoffrey is still deciding how to answer:

- (14) Thomas well I mustn't go on boringly talking about me, what are you doing  
Geoffrey,  
Geoffrey {- um u:h .} much the same old things, {.} (1.9.960)

And in (15), Connie is still deciding among the alternatives offered, as she makes clear with “oh dear”:

- (15) Barb     what would you like {.} we’ve got {.} vermouth {.} whisky {.} or wine  
           Connie  {uh, .} oh dear, (2.7.30)

These few examples, which are representative of many more, suggest that *uh* and *um* can be used to signal delays of almost any type. It could have been otherwise – as we ourselves had once assumed.

## 5.2. *Implicatures*

When speakers use *uh* and *um*, they often appear to mean something more than “I am announcing the initiation of an expected delay in speaking”. In Grice’s theory, speakers can implicate things by what they say (Grice, 1975; Horn, 1984; Levinson, 1983, 2000; Sperber & Wilson, 1986). Many implicatures are created via what Grice called the *maxim of relevance*. Addressees are intended to see what the speaker is saying as relevant to the current situation. They are to ask themselves, “Why is the speaker saying this now?” and construe the *mutually recognizable* reasons as implicatures (cf. Ducrot, 1980).

The argument here is that, in using *uh* and *um*, speakers often implicate things by appeal to relevance. They presuppose at least three general reasons for using *uh* or *um*:

- A           Speakers have reasons for *wanting*, or for thinking they are *expected*, to be speaking at *t*(filler);
- B           Speakers have reasons for *initiating* a delay in speaking at *t*(filler);
- C           Speakers have reasons for *announcing* that they are initiating a delay in speaking at *t*(filler).

By *presuppose*, we mean that speakers take these propositions to be common ground with the addressees (Stalnaker, 1978). If Reasons A, B, and C are mutually recognizable, addressees should take them as implicatures. If addressees can work out more *specific* reasons behind A, B, and C, they should take those as implicatures too. There is evidence for both predictions.

Consider (16) in which Alan has been asked by Charles in an interview about “recent novels” he has read:

- (16) Alan  I’ve {u:m} recently read {u:m . oh, .} *Lord of the Flies* (3.5a.110)

With his second “u:m”, Alan announces that he is initiating a delay at the end of “I’ve recently read”. By Reason A, he implicates that he would like to continue speaking there, and by Reason B, that he has good reasons for *initiating* a delay. One obvious reason is that he cannot remember a novel, or retrieve the title of a novel, that he has read recently. By Reason C, he implicates that he has reasons for *announcing* the initiation of that delay. An obvious reason is that he wants Charles to know that he is aware of the problem and is in the process of solving it. Indeed, with “oh”, he claims to have just solved it by remembering that he had read *Lord of the Flies*.



One way to highlight the implicatures with *uh* and *um* is to compare announced with unannounced delays. Again consider (16). If Alan had left a naked pause in place of the second “u:m”, that might have suggested that he had become distracted, or was abandoning the answer to start a new direction, or didn’t think Charles deserved an account. By inserting “u:m”, Alan signals that he is working on the continuation and prevents the unwanted inferences.

In the right circumstances, speakers can implicate things by using overlapping speech (see Jefferson, 1973). Consider what Schegloff (1987) has called *recycled turn beginnings*, as here:

- (17) Ken I must allow myself the good time, the first time I do it, it must be terribly  
 \*(1 to 2 sylls) Wednesday at three,\*  
 Jack *\*but I {.* but I thu- { } I\* don’t think I’m going to go on with it,  
 Ken are you doing two or one {. papers this year (1.4.858)

Ken is to ask himself why Jack is deliberately overlapping “but I {. but I thu- I” with the end of his turn, and an obvious reason is that Jack wants the next turn. The implicature is that Jack is requesting the next turn. Speakers can use *uh* or *um* with much the same implicatures, as in this exchange between instructor and student:

- (18) David do you know his remarks on Hamlet,  
 Edward {- - -} yes, I have read them sir,  
 David m, {. what are they, {- -} \*will you give me the gist of his approach,\*  
 Edward *{\*u:h - uh .\** he { } he believes that, {- -} that Shakespeare attempts in Hamlet, something which he {. he didn’t understanding himself – even, (3.5b.972)

David asks Edward to summarize a well known scholar’s remarks on Hamlet with “what are they”. When his request is met with silence (“- -”), he rephrases his request. It is at this point that Edward uses overlapping “u:h - uh .” to implicate that he is preparing to answer.

The specific reasons for stopping, delaying, and announcing the delay change with the situation, so *uh* and *um* can be used with opposite implicatures on different occasions. One example is holding vs. ceding the floor. As we noted earlier, *uh* and *um* have often been described as signaling “I’m still in control – don’t interrupt me” (Maclay & Osgood, 1959; Mahl, 1987; Rochester, 1973; Schegloff, 1982; Siegman, 1979). Although examples (13) through (15) fit this interpretation, others do not. In (19), Helen has just been interviewed by George:

- (19) George and we’ll have to { } we’ll have to let you know, what we, {. decide about your application,  
 Helen thank you very much, and now I have to see the *{u:h .}*  
 George Tutor to Women Students, and the secretary (3.1b.1017)

When Helen takes her turn, she doesn’t know who she is to see next. By using “u:h” to

announce a delay in speaking, she implicates that she wants George to complete her utterance, and he does that (see also Goodwin & Goodwin, 1986). Or consider (20):

- (20) Sam       but {-} the whole object of this, is to talk about, {.} first, naturally the department, {-} but {} but also if anybody wants to raise anything else about the college, {- . u:h} do please do so, {. I mean} it's abs- {} total free for all, {- . u:m - - .} how about things generally, {I mean} have you {uh} let's start with the accommodation, {.} obviously this is a problem {- - u:h - - -}
- William I think it's a problem {.} more and more, as {} as in our new syllabus you see (3.4.21)

Sam inserts *uh* and *um* with long delays several times apparently to invite William to reply, and only after three of these does William reply. Sam uses *uh* and *um* not to hold the floor, but to signal his willingness to give it up. Also, as we noted earlier, speakers use *uh* and *um* in both monologues and dialogues (see also Fox Tree, 1999), and in monologues, there is no issue of holding or ceding the floor. In short, the interpretations of wanting to hold or cede the floor cannot both be basic meanings. Words almost never have two conventional meanings that are antonyms.

It is worth collecting all the interpretations of *uh* and *um* that we have found in the literature. If the typical reason for a delay is that the speaker is *unable* to proceed, speakers should often use *uh* and *um* to implicate "I am unable to proceed". That is reflected in the first five interpretations:

- (a) speakers are currently experiencing a planning problem (Levelt, 1983, 1989);
- (b) speakers are searching memory for a word (Goodwin, 1987; Goodwin & Goodwin, 1986; James, 1972);
- (c) speakers are hesitating about something (*ADH*, 2000; *OED*, 2000);
- (d) speakers are in doubt or uncertain about something (*ADH*, 2000; Brennan & Williams, 1995; *OED*, 2000; Smith & Clark, 1993);
- (e) speakers are still "engaged in speech-productive labor" (Goffman, 1981), such as deciding what to say (Brotherton, 1979; Cook, 1969) or how to say it (Boomer, 1965; Cook, 1971);

If speakers have other reasons for delays, they can implicate them as well, and that accounts for the following 11 interpretations:

- (f) speakers want to keep the floor (Maclay & Osgood, 1959; Mahl, 1987; Rochester, 1973);
- (g) speakers want to cede the floor (examples (30) and (31));
- (h) speakers want the next turn (Beattie, 1983; Sacks cited in Schegloff, 1982, p. 81);
- (i) speakers have completed their turn (Cook & Lalljee, 1972);
- (j) speakers are inviting their addressees to speak (example (31));
- (k) speakers are requesting help in completing the current utterance (example (30); Goodwin & Goodwin, 1986);
- (l) speakers are being polite (Fischer, 1999; Jefferson, 1974);

- (m) speakers are thinking about what was just said (Fischer, 1999);
- (n) speakers are inviting listeners to think about what they are about to say (Jefferson, 1974);
- (o) speakers are marking syntactic or discourse boundaries (Cook, Smith, & Lalljee, 1974; Fischer, 1999; Maclay & Osgood, 1959; Martin, 1967; Swerts, 1998);
- (p) speakers are providing information about their current mental state (Brennan & Williams, 1995; Smith & Clark, 1993).

Speakers can surely use *uh* and *um* with other interpretations as well. The problem of multiple interpretations is solved if all but the basic ones are implicatures.

### 5.3. Listeners' understanding of *uh* and *um*

Listeners should take advantage of what speakers mean by *uh* and *um*, and they do. In one study (Fox Tree, 2001), listeners had to identify words from recordings of speech with spontaneous *uhs* either present or removed. They were faster with the *uhs* present. The *uhs* appeared to heighten attention to the upcoming speech. In another study (Barr, 1998), listeners had to point to abstract shapes on a computer monitor while listening to descriptions of those shapes. Some of the shapes they had seen before, but others they hadn't. When the shapes were new, listeners responded faster and moved their cursors toward them faster when the descriptions were preceded by *um*. In a third study (Christenfeld, 1995), listeners judged speakers with spontaneous *uhs* and *ums* to be more relaxed than the same speakers with their *uhs* and *ums* replaced by pauses.

Listeners also take account of *uh* and *um* in interpreting answers to questions. In one study (Brennan & Williams, 1995; cf. Smith & Clark, 1993), speakers who preceded their answers with *uh* or *um* were judged to be less sure of their answers than speakers who preceded their answers with equivalent pauses. Speakers who preceded "I don't know" with *uh* or *um*, in contrast, were judged to be *more* likely to know the answer. In another study (Fox Tree, in press), listeners were asked to interpret turns that began with a spontaneous *um* or with the *um* removed. When a turn began with *um*, speakers were judged to be having more problems in speech production.

Still, the most direct evidence for interpreting *uh* and *um* is found in addressees' responses, as in (21):

- (21) Roger now what was it, {- - -} sorry, {-} my memory's a little bit { . u:h } hither  
 and yon sometimes, {-} but this is at least what {-} thirty years ago, isn't it,  
 Sam { . } never mind you're doing very well,  
 Roger { um }  
 Sam it'll come  
 Roger { u:m - - - } I think it's a slightly funnier story, (15 s pause) (1.14a.582)

Roger has trouble recalling certain facts and uses *um* in turns 3 and 5 to implicate why he is delaying. Sam explicitly displays his interpretation of Roger's *um* in turn 3 with "it'll come" and waits out Roger's delay in turn 5. The addressees in (18) through (20) are explicit in their understandings of *uh* and *um* as well. Examples like these are common in the LL corpus.

To summarize, if *uh* and *um* have the basic meaning "used to announce the immediate

initiation of what is expected to be a minor, or major, delay in speaking”, speakers can use them to implicate: (1) that they wouldn’t normally expect to delay at this moment; (2) that they anticipated the delay; and (3) that they are aware, at some level, of the reasons for their delay. Depending on the circumstances, they can use *uh* or *um* to implicate further propositions as well, and listeners appear to grasp these meanings and implicatures.

## 6. *Uh* and *um* as conventional

In the filler-as-word hypothesis, *uh* and *um* are conventional English words. If they are, other languages could have evolved other fillers, and *uh* and *um* may even vary across English dialects.

### 6.1. Variation by language

*Uh* and *um* are hardly universal. Table 1 lists fillers from other languages culled from grammars or transcripts of spontaneous speech (spelling from the references). Each of the languages on our list has two or more fillers in contrast. Dutch and German each have two (*uh* vs. *um*; *äh* vs. *ähm*) that are much like English *uh* vs. *um*. Japanese has fillers (*eto*, *ano*, *kono*, and *sono*) that mark distinctions rather different from English (Emmett, 1996, 1998; Hinds, 1975). In several unrelated languages, the fillers are historically derived from demonstratives – e.g. *ano*, *kono*, and *sono* in Japanese and *este* in Spanish (see Wilkins, 1992). Fillers, therefore, may contrast not only in length of delay, as in English, but on other dimensions as well.

The fillers in Table 1 all tend to be brief (one or two syllables) and built around central vowels in the language. The commonest fillers are easy to prolong, as in English “u:h” and “u:m”, Norwegian *e*, *e=*, and *e==*, Swedish *äh* and *ähh*, and Japanese *ano* and *anoo*. And one of the contrasting fillers usually contains a nasal (*m* or *n*), perhaps because these are easy to prolong. Why these properties? Languages need more than one filler if speakers wish to distinguish one type of delay from another. The fillers must be brief if speakers are to plan and produce them quickly midutterance. And they must be prolongable if speakers are to stretch them with little prior planning.

Table 1  
Fillers found in several languages (spelling from sources)

Language	Fillers	References
German	äh, ähm	Fischer (1999)
Dutch	uh, um	Swerts (1998)
Swedish	eh, äh, ääh, m, mm, hmm, ööh, a, öh	Allwood et al. (1990) and Eklund (1999)
Norwegian	e, e=, e==, eh, eh=, m, m=, m==, hm, mm, em=, øhø, aj	Svennevig (1999)
Spanish	eh, em, este, pues	Brody (1987) and van der Vlugt (1987)
French	eu, euh, em, eh, oe, n, hein	Duez (1982, 1991, 1993)
Hebrew	eh, e-h, em, e-m, ah, a-m	Maschler (1997)
Japanese	eeto, etto, ano, anoo, uun, uunto, konoo, sonoo, jaa	Cook (1993), Emmett (1996, 1998) and Hinds (1975)

Table 2  
Percentages of *uh* and *um* in four corpora

Speakers	Discourse type	Corpus	<i>Uh</i>	<i>Um</i>
British academics	Face-to-face	LL	54	46
British academics	Answering machine	AM	76	24
Texans	Telephone calls	SW	84	16
Californians	Narratives	Pear	44	56

## 6.2. Variation by dialect

If *uh* and *um* are conventional, they may also vary from one English dialect to another. From informal observations, their pronunciation seems to differ slightly between British and North American English, and among at least some North American dialects. The division of labor between *uh* and *um* may also vary by dialect. The rates and percentages of *uh* and *um* in our four corpora are summarized in Table 2. The percentage of fillers that were *uh* varies from 44 to 84%, but there is no clear relation to whether the speakers were British or North American, or whether the speech came from dialogues or monologues.

Despite the variation, British and North American speakers use *uh* and *um* with the same core contrast between major and minor delays. That suggests that *uh* and *um* are conventional for English: they are part of what one learns when one learns English. Speakers of English as a second language often import the fillers from their first language – we have heard examples from native French, Hebrew, Turkish, and Spanish speakers – and that is one reason they continue to be heard as non-native speakers.

## 7. Planning *uh* and *um*

We now turn to planning for and formulating *uh* and *um* in the process of speaking. First, planning. For speakers to select the message “I am now initiating what I expect to be a minor, or major, delay”, they must *monitor* the utterance they are formulating and *detect* an imminent delay. There is much evidence that speakers scan internal representations – a type of inner speech – for material they have formulated but not yet produced (Levelt, 1983, 1989; Wheeldon & Levelt, 1995). But can they monitor for imminent delays, and distinguish major from minor delays?

### 7.1. Locations of *uh* and *um*

One of the most basic units of spontaneous speech is the *intonation unit*, which is a stretch of speech under a single intonation contour.<sup>6</sup> Intonation units range from major constituents, such as sentences and clauses, to single words, such as *uh huh* or *well*. They

<sup>6</sup> Intonation units (e.g. Chafe, 1992) have also been called *tone groups* (Halliday, 1967), *tone units* (Crystal, 1969; Crystal & Davy, 1975; Svartvik & Quirk, 1980) as in the LL corpus, *intonation groups* (Cruttenden, 1986), *information blocks* (Grimes, 1975), *idea units* (Chafe, 1979, 1980), and *lines* (Gee, 1986). In prosodic theory, idealized intonation units are called *intonation phrases* (Nespor & Vogel, 1986; Selkirk, 1996).

may even be fragments of phrases. There is good evidence that intonation units are central to planning. In narratives, they tend to express the basic narrative actions (Chafe, 1979, 1980), and in conversation, they are the units from which turns are built (Clark, 1994a; Ford & Thompson, 1996). According to some analyses, they must be planned as units for prosody to come out right (Grosz & Hirschberg, 1992; Swerts & Geluykens, 1994; cf. Levelt, 1989, p. 400).

Most theories of production predict that the more difficult the planning, the more likely, and the longer, speakers will delay before speaking. Let us define three locations in intonation units: (I) at the boundary; (II) after the first word (ignoring *uh* and *um*); and (III) later. These three locations are marked for the six pauses in examples (22) and (23) (intonation unit boundaries marked with commas):

- (22) and then uh somebody said, . [I] but um - - [II] don't you think there's evidence of this, in the twelfth - [III] and thirteenth centuries? (1.3.334)
- (23) Hamlet um - - - [II] starts, . [I] uh as a noble soul, th- there's no doubt that . [III] that Hamlet has got this nobility of soul (3.5a.315)

Planning should be most difficult in location I, where speakers have to select the message, formulate the syntax for the unit, and create its overall prosody. It should be easier in location II, once they have made many of these selections. It should be easiest in location III, where they have completed even more of the processing. Past research has shown that speakers do indeed pause most (longest and most often) in location I, less in location II, and least in location III (Boomer, 1965; Brotherton, 1979; Chafe, 1980; Deese, 1984; Ford, 1982; Koopmans-van Beinum & van Donzel, 1996; Stenström, 1990).

Speakers should therefore be most likely to signal a delay in location I, less likely in location II, and least likely in location III. We computed the rates of *uh* and *um* in the LL corpus in these three locations. Because there are more opportunities for *uh* and *um* in location III than in I or II, we counted the intonation units of each length in words (excluding *uh* and *um*), calculated the number of opportunities for *uh* and *um* at each location, and computed their rate at each location given the opportunities. The rates are shown in Fig. 5. As expected, the combined rate of *uh* and *um* was highest in location I (43 per 1000 opportunities), next highest in location II (27 per 1000), and lowest in location III (13 per 1000). These differences are highly significant ( $\chi^2(2) = 1268, P < 0.001$ ).

Speakers can signal either a minor or a major delay at each of these locations. They used *um* more often in location I, where delays were presumably longer, than in locations II or III, where delays were shorter. As shown in Fig. 5, *um* comprised 56% of the fillers in location I, but only 39 and 37% of the fillers in locations II and III ( $\chi^2(2) = 126, P < 0.001$ ). Also, prolonged *uh* and *um* occurred more often in location I than in locations II and III, 77 vs. 52 and 54% ( $\chi^2(2) = 224, P < 0.001$ ). So at one extreme, prolonged *um* (“u:m”) comprised almost half (47%) of the fillers in location I, but only 22% of those in locations II and III. At the other extreme, normal *uh* (“uh”) comprised only 13% of all fillers in location I, but 30% of those in locations II and III.

The first intonation unit of a turn might be special because it is the first thing speakers say in taking the floor. But as a location for fillers, it isn't much different from later units.

In the LL corpus it was preceded by slightly more “u:m”s (49 to 45%) and slightly fewer “u:h”s (26 to 32%) than other intonation units.

Fillers from the SW corpus lead to similar conclusions (see also Shriberg, 1994; Swerts, 1998). Although the SW corpus doesn’t mark intonation units, it marks sentences, which represent one or more complete intonation units. By hypothesis, *um* should occur more often at sentence boundaries than within sentences. Indeed, it comprised 32% of all fillers at sentence boundaries, but only 12% of the fillers within sentences ( $\chi^2(1) = 3382$ ,  $P < 0.001$ ).

## 7.2. Predicting delays

There are at least two accounts for the pattern in Fig. 5. One is that speakers monitor for the *location* of an imminent delay. If it is at the boundary of an intonation unit (location I), they prefer *um*. If it is within an intonation unit (location II or III), they prefer *uh*. Let us call this the *boundary account*. The second account is that speakers monitor simply for the *length* of imminent delays. Let us call this the *length account*. Do speakers predict the length of upcoming delays, their locations, or both?

To answer this question, we examined all of the pauses in the LL corpus and counted how often they were preceded by *uh* or *um*. After eliminating pauses that accompanied laughs, sighs, and coughs, we were left with about 19,200 pauses: 12,800 in location I (at unit boundaries), and 6400 in locations II and III (within units). Although these pauses ranged from 0.5 to 3 units in length, there were so few 1.5 and 2.5 unit pauses that we combined them with the 1 and 2 unit pauses, respectively. We then computed the percentage of pauses of each length that were preceded by *uh* or by *um* at unit boundaries (location I) and within units (locations II and III). The percentages are shown in Fig. 6.

Speakers were quite accurate in predicting the length of upcoming pauses. In general, the longer the pause, the more likely they were to precede it with *uh* or *um*. The overall percentages in Fig. 6 increase significantly from short pauses (0.5 and 1 units) to long pauses (2 and 3 units) ( $\chi^2(1) = 76$ ,  $P < 0.001$ ). Also, the longer the pause, the more likely the filler was *um* and not *uh* – although mainly *within* units. At unit boundaries, the ratio of *um* to *uh* increased from 2.6 to 3.3 for short to long pauses ( $\chi^2(1) = 1.16$ , n.s.). But within units, the ratio doubled from 0.9 to 1.9 ( $\chi^2(1) = 11.33$ ,  $P < 0.001$ ). At least within units, speakers are able to predict not only the occurrence of an up-coming pause, but its length.

The percentages in Fig. 6, in fact, reflect a combination of the boundary and length accounts. Speakers were twice as likely, 9.9 to 4.1%, to mark a delay of a particular length *within* a unit as at unit *boundaries* ( $\chi^2(1) = 259$ ,  $P < 0.001$ ). At unit boundaries, the percentages of *uh* and *um* increased only slightly from short to long pauses – from 1.0 to 1.3% for *uh* ( $\chi^2(1) = 1.53$ , n.s.), and from 2.7 to 4.4% for *um* ( $\chi^2(1) = 20$ ,  $P < 0.001$ ). Within units, the increases were much larger. They doubled from 4.3 to 9.6% for *uh* ( $\chi^2(1) = 29$ ,  $P < 0.001$ ), and they increased five-fold from 4.2 to 21.1% for *um* ( $\chi^2(1) = 175$ ,  $P < 0.001$ ). The increases were reliably larger within units than at unit boundaries ( $\chi^2(1) = 426$ ,  $P < 0.001$ ).

At first glance, Fig. 6 seems at odds with Fig. 5. In Fig. 5, speakers seem to use more fillers at unit boundaries than within units, but in Fig. 6, they seem to do just the opposite. The resolution is simple. When there *is* a pause, it is less likely to be signaled if it is at a

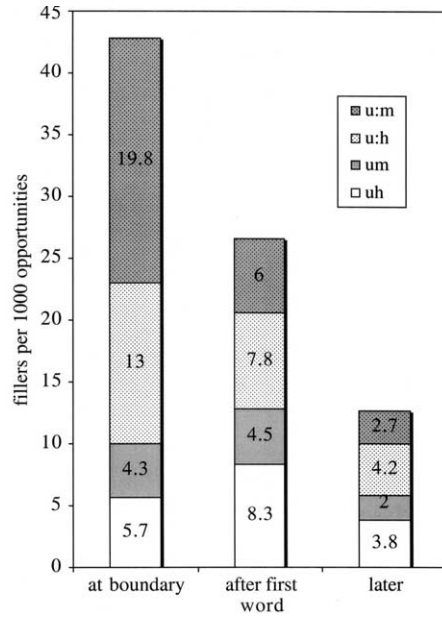


Fig. 5. Rates of *uh* and *um* at three position in tone units (LL corpus).

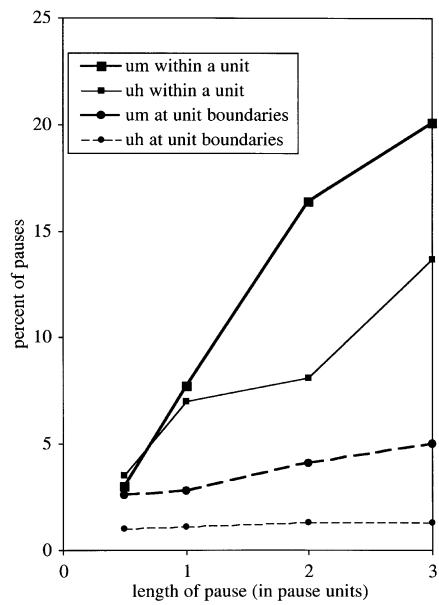


Fig. 6. Percentage of pauses preceded by fillers (LL corpus).



unit boundary (Fig. 6). But there are so many more pauses at unit boundaries that the rate of fillers is, nevertheless, higher at unit boundaries (Fig. 5).

The picture is this. It is acceptable to pause at intonation unit boundaries, so it takes a long pause before speakers feel the need to add *uh* or *um*. It is much less acceptable to pause within intonation units, so it takes a much shorter pause before speakers feel the need to add *uh* or *um*. A one-second delay is major when it comes *within* a unit, but minor, or perhaps not even worth noting, when it comes at a unit boundary. In short, speakers judge imminent pauses not by their absolute length (“brief” vs. “long”), but by their local importance (“minor” vs. “major”) – by how disruptive they would be.

### 7.3. Variation by speaker

Speakers differ enormously in how often they use *uh* and *um*. The 65 speakers in the LL corpus who produced more than 1000 words each ranged from 1.2 to 88.5 fillers per 1000 words (median 17.3). They also varied in which filler they used more often. One used only *uh* (85 instances), and another used only *um* (but just four instances). The median speaker used 52% *uhs* and 48% *ums*. So speakers have characteristic preferences in fillers, just as they do for other words in their vocabulary.

Speakers also differ a great deal in how fast they speak. In the Maclay and Osgood (1959) study, speakers ranged from about 120 to 180 words per minute, and in the study by Goldman-Eisler (1968, p. 25), from 140 to 260 syllables per minute. Some people speak over 50% faster than others. What accounts for the differences? As Goldman-Eisler demonstrated, it isn’t how fast speakers articulate each word, but how much time they leave *between* words. For Goldman-Eisler’s speakers, speech rate correlated  $-0.94$  with the total duration of pauses (including, apparently, “filled pauses”), but only  $-0.17$  with the rate of articulation. Slow speakers are slow because they introduce delays (see also Deese, 1984).

If *uh* and *um* are used to signal delays, slow speakers should find them more useful than fast speakers.<sup>7</sup> To test this prediction, we examined the 65 speakers in the LL corpus with more than 1000 words each, counted their brief pauses, unit pauses, “uh”s, “u:h”s, “um”s, and “u:m”s for each speaker, and calculated the rate of fillers and pauses per 1000 words. As before, we counted brief pauses as 0.5 units and unit pauses as 1.0 units. Speakers ranged from 40 to 190 pause units per 1000 words (median 73). Table 3 shows the correlations among the rates of pausing and the rates of the four fillers.

As expected, the more speakers paused, the more fillers they used. Amount of pausing was correlated with each of the four fillers separately; it had a multiple correlation of 0.50 with them all ( $F(4, 60) = 5.09, P < 0.0013$ ). That is, we can account for 25% ( $= 0.50^2$ ) of the variation of speakers’ pausing by knowing how often they use the four fillers. The same conclusion holds for eight Dutch speakers narrating stories (Koopmans-van Beinum & van Donzel, 1996). The more often these speakers paused before words, the more often they used fillers ( $r = 0.86, F(1, 6) = 17.4, P < 0.006$ ).

<sup>7</sup> This prediction is confirmed in the data of Maclay and Osgood (1959). For their speakers, speech rate was correlated more highly with rate of “filled pauses” ( $-0.71$ ) than with rate of “unfilled pauses” ( $-0.38$ ). But their “filled pauses” also contained unfilled pauses, so their data don’t allow the proper test.

Table 3  
Correlations among rates of “uh”, “u:h”, “um”, “u:m”, and pauses for 65 speakers in the LL corpus

Variable	uh	u:h	um	u:m
uh				
u:h	0.41*** <sup>a</sup>			
um	0.22	-0.10		
u:m	0.26*	0.56***	0.24*	
Pauses	0.25*	0.20	0.37**	0.40***

<sup>a</sup> \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

If *uh* is used for brief delays, and *um* for longer delays, speakers' rates of pausing should be predicted better by their rate of *um* than by their rate of *uh*. In Table 3, amount of pausing is correlated 0.48 with the rate of *um* (“um” and “u:m” together) ( $F(1, 63) = 19.04$ ,  $P < 0.001$ ), but only 0.25 with *uh* (“uh” and “u:h” together) ( $F(1, 63) = 4.39$ ,  $P < 0.04$ ). That is, we can account for 23% ( $= 0.48^2$ ) of people's rates of pausing by knowing how often they use *um*. We gain only 2% by also knowing how often they use *uh*.

#### 7.4. Speakers' control of *uh* and *um*

From the speaker's perspective, using fillers has both its pluses and its minuses. On the plus side, fillers warn addressees about impending delays. When speakers are delayed by *preparedness problems* – finding words, formulating utterances, deciding what to say – they can use fillers to help addressees deal with these problems. And, as just reviewed, there is much evidence that listeners make use of these fillers.

On the minus side, whenever speakers use fillers, they are *announcing* that they are having preparedness problems, something they may not want to admit in public. Speakers on the radio, on television, and in formal speeches are expected to be knowledgeable and competent, so it might undermine their authority to admit to preparedness problems (Goffman, 1981). Courses on public speaking train people to speak without *uh* and *um*, and the best public speakers are successful. In all of the recorded inaugural speeches by US presidents between 1940 and 1996, for example, there is not a single *uh* or *um* (Kowal et al., 1997; see also Duez, 1982, 1991).

If speakers have control of *uh* and *um*, they should use them less often in formal than in informal registers, and there is much evidence that they do. In a study by Schachter, Christenfeld, Ravina, and Bilous (1991), Columbia University lecturers averaged 3.23 fillers per minute in lectures (a formal register), but 5.18 fillers per minute in face-to-face interviews (an informal register). Other lecturers averaged 2.53 fillers per minute in lectures to undergraduates (a formal register), but 3.68 fillers per minute in lectures to graduate students (a less formal register). There are similar findings by Broen and Siegel (1972), Duez (1982), and Kowal et al. (1997).

These findings undermine two alternative accounts for fillers. One is that people *automatically* produce fillers whenever they pause. The Columbia lecturers used fewer fillers

in lectures than in informal interviews even though they produced more pauses in their lectures (they spoke more slowly). The second account is that people automatically produce fillers whenever they run into planning problems. The Columbia lecturers presumably encountered more planning problems in lectures – which is one reason they were slower – yet they used more fillers in informal interviews.

There is other evidence that speakers have control over *uh* and *um*. In a study by Christenfeld (1996), people were asked to speak either freely or in time to a metronome. These people produced over three times as many fillers speaking freely as with the metronome (8.98 vs. 2.06 fillers per minute), even though in the two conditions they spoke at the same rate (i.e. with the same number of pauses) and with the same range of vocabulary. When timing was regulated by the metronome, we suggest, these speakers didn't need to account for their delays, so they used fewer fillers. Also, when people get drunk, they presumably find it harder to monitor for upcoming pauses, or care less about helping their addressees, or both. In a study by Christenfeld and Creager (1996), the drunker people got, the fewer *uhs* and *ums* they used. Fillers are not the automatic outcome of problems or delays.

To summarize, speakers have selective control over *uh* and *um*. They are quite accurate in projecting minor vs. major delays, inserting *uh* before minor ones and *um* before major ones. Speakers who have need of fillers – speakers who pause more – use more fillers, especially *um*. And yet speakers can reduce or eliminate their use of fillers when the circumstances require it.

## 8. Formulating *uh* and *um* within utterances

Formulating and articulating *uh* and *um* would seem simple. Once speakers discover an imminent delay and select the message “I am now initiating what I expect to be a minor delay”, all they need to do is formulate *uh* and produce it. But the process is not that simple. If they are already speaking, they must first suspend speaking. Should they stop mid-word, after the current word, after the current phrase? Should they mark the suspension before producing *uh*, or the resumption following it? If so, how? What intonation should they give *uh*? And even though interjections, by definition, do not combine with other words in *syntactic* constructions, they often combine with other words in *prosodic* units, as in *ah yes*, *oh no*, and *hi there* (Bolinger, 1989). If *uh* and *um* combine with other words, their formulation becomes even more complicated. Simply put, speakers must formulate *uh* and *um* as part of the current utterance or discourse, making whatever accommodations are necessary to do that.

### 8.1. Intonation of *uh* and *um*

As speakers struggle with the competing demands of formulating and articulating utterances, they often create ad hoc intonation units – intonation units that are *not* syntactic constituents. In (23), “Hamlet *um* - - - starts” is an incomplete clause but also an intonation unit, and so are “he believes that” and “what we” in examples (18) and (19).

One type of ad hoc intonation unit consists of *uh* or *um* alone. There were 13 of these in the LL corpus, as in this example:

- (24) Ann but then Tom’s reaction to this is, - um, . ? well they’re only trying to distance themselves, from literature (1.6.250)

The LL corpus marks rising, level, and falling intonation (and their combinations) on the nucleus of each intonation unit – “the peak of greatest prominence”. The nucleus fell on *uh* or *um* only when the intonation unit consisted of *uh* or *um* alone. The *um* in (7) had level intonation; others had rising, level, or falling intonation. So speakers can formulate *uh* and *um* as separate intonation units. They could do that only if they formulated *uh* and *um* as words.

A more common ad hoc intonation unit is one that ends with *uh* or *um*. There were about 300 of these in the LL corpus, as in this example:

- (25) Alice I must admit that um, the book-club, offered to buy us a special pre-Christmas gift (2.12.1)

Alice formulated the intonation unit “I must admit that um” by placing a unit-final contour on *um*. Over half of these examples (166 of them) consisted of a single word plus *uh* or *um*, as here:

- (26) Beth and, . my auntie Flora went . uh . Elsie went home, and u:h, - - - my mother said, - was going on about her to me, over the phone (1.12.1033)

The first word was *and* in 59% of these cases and *but* in 20%; all but two cases had level intonation. Intonation units like Beth’s “and u:h” in (26) are so short that they must each have been formulated as a piece.

Within intonation units, speakers formulate *uh* and *um* with a prosody that makes them distinguishable from the surrounding words, according to a discovery by Shriberg and Lickley (1993). Every clause by a speaker has a melody with a *maximum* pitch, or fundamental frequency  $F_0$ . Likewise, every speaker has a *minimum*  $F_0$ . As Shriberg and Lickley documented, the  $F_0$  of each *uh* and *um* within a clause lies at about 60% of the distance up from the minimum  $F_0$  for that speaker to the maximum  $F_0$  for that clause.<sup>8</sup> The same percentage held for every speaker in their sample, both British and North American.

This prosody helps distinguish *uh* and *um* from the surrounding words, we suggest, by allowing *uh* and *um* to be perceptually segregated from the rest of the clause. Consider (27):

- (27) Reynard this is this is one of the things that - uh one of the many things, - uh in English structure, which is - u:m - - - an item in a closed system (1.1.748)

Each of these clauses has two parts: (a) a melody for the clause proper; and (b) a monotone

<sup>8</sup> Note what a remarkable feat this is at the articulatory level. Speakers are able to alter the pitch of their larynx from the on-going melody of the clause to the monotone pitch of the filler and back again within milliseconds.

accompaniment by *uh* and *um*. Listeners should be able to hear *uh* and *um* separately from the clauses by a perceptual process known as *auditory streaming* (Bregman, 1990): they would hear the primary clause as one auditory stream and *uh* and *um* as another.

## 8.2. *Uh* and *um* as clitics

In many languages, brief unstressed words tend to get cliticized onto adjacent words (see Nespor & Vogel, 1986; Selkirk, 1984). In English, function words of one syllable normally get cliticized onto the following content words, and that often requires a readjustment of syllable boundaries (Selkirk, 1996). The process can be illustrated (following Kenstowicz, 1993) in standard orthography with periods to mark syllable boundaries. It creates *prosodic words* such as these: “to.school” (*to school*) but “to.windia” (*to India*); “the.dog” (*the dog*) but “the.yegg” (*the egg*); and “a.cat” (*a cat*) but “a.negg” (*an egg*). Other function words, such as *not* and *will*, get cliticized in contracted form onto the *prior* word, as in “couldn’t” and “I’ll”. *Uh* and *um* are brief and unstressed, so if they are words, they might also get cliticized onto adjacent words.

We find that *uh* and *um* are often cliticized onto prior words and never onto following words. Here is an utterance recorded from a radio interview (with silences measured to the nearest tenth of a second):

(28) Alan but-uh (0.2) we-um (1.1) uh have-uh (0.1) eight to twelve airplanes  
that-uh enter the airspace right-uh in front of the crowd

In (28), Alan produces five of his six fillers (those marked with hyphens) cliticized onto the prior word. He produces the prosodic words “bu.tuh”, “we.yum”, “ha.vuh”, “tha.tuh”, and “righ.tuh”, each with adjusted syllable boundaries. In the corpora we have listened to, *uh* and *um* are especially common as clitics on introductory conjunctions, as in “an.duh”, “bu.tuh”, “so.wuh”, and “i.fuh”. They are also found on words such as *that* and *right* in (28). In a two and a half minute interview, Alan used 22 of 25 fillers as clitics, including “air.plane.suh” and “re.por.te.duh”. In the corpora we have listened to, many speakers used cliticized *uhs* or *ums* at least some of the time.

Cliticizing *uh* or *um* onto a previous word is strong evidence that *uh* and *um* are used to signal the *initiation* of an expected delay. In (28), when Alan anticipated a delay after *but*, he was able to formulate *but* plus *uh* as the prosodic word “bu.tuh”. On the other hand, he should *not* have been in a position to formulate *uh* plus *we* as the prosodic unit “uh.we”. If he had been in such a position, he could have produced “we” immediately with no need to signal a delay. Put another way, “bu.tuh” was possible because, at that moment of formulation, Alan had already selected both *but* and *uh*. “Uh.we” was not possible because he had not yet selected *we*, or he would have formulated *we* by itself. So if *uh* and *um* are cliticized onto an adjacent word, it should be onto the previous word.

It is possible *in principle* to cliticize *uh* and *um* onto a following word. *Oh*, for one, is regularly cliticized onto following words. *Oh indeed* can be pronounced “o.win.deed”. In the corpora we have listened to, we have never heard *uh* or *um* cliticized onto a following word. It is also possible in principle to cliticize *uh* and *um* onto word fragments. Consider this example:

(29) Reynard no, but fr- uh but from that point of view it would be odd (1.1.384)

Phonologically, Reynard could have combined *uh* with *fr-* to produce “fr-uh”. Yet in the corpora we have listened to, we have never heard *uh* or *um* combined with a word fragment. This finding is consistent with *uh* and *um* viewed as words. It is legitimate prosodically to combine two words, such as Alan’s “bu.tuh”, but not a fragment and word, such as the non-existent “fr-uh” (see Clark & Wasow, 1998).

### 8.3. Strong and weak forms

The words *to*, *the*, and *a* come in two forms – strong and weak – that contrast in use. The strong forms are “tuw”, “thiy”, and “ei”, which have non-reduced vowels and rhyme with *glue*, *see*, and *day*. The weak forms are “tuh”, “thuh”, and “uh”, which have reduced vowels (schwas) and rhyme with the second syllable of *sofa*. In prosodic theory, the strong forms have two morae, and the weak forms have one. According to Selkirk (1996), the weak forms are used whenever the words are cliticized onto the following word, as in “tuh.school”, “thuh.dog”, or “uh.cat”. The strong forms are used only in special contexts – in accented positions (e.g. *He was walking to the house, not from it*), at the ends of phonological phrases (*Where has George gone to?*), and in isolation (*The word you want is ‘to’*). That is, the weak forms are unmarked prosodically, and the strong forms are marked.

Speakers often use the strong, marked forms as collateral signals (Fox Tree & Clark, 1997). Consider this example:

(30) Sam and when you come when you come to look at thiy . thuh literature, - I mean you know the actual statements, (1.2.220.B)

Here Sam produces “thiy”, suspends speaking, and then resumes speaking with “thuh”. Evidence shows that, in cases like (30), speakers use “thiy” to signal that they are suspending their speech at the end of *the*. They often use “ei” and “tuw” analogously.

*Uh* and *um*, too, differ in prosodic weight, and again it is the weak form that is unmarked in use. Normal *uh* is a reduced syllable of one mora, and *um* is a full syllable of two morae. Indeed, *uh* is nearly twice as long as *uh*. In various corpora, *uh* and *um* averaged 327 and 384 ms (Fox Tree, 2001), 344 and 527 ms (Brennan & Williams, 1995), 476 and 615 ms (in Dutch; Fox Tree, unpublished data), and 361 and 586 ms (in Dutch; Swerts, 1998).<sup>9</sup> *Um* also tends to be slightly higher in pitch than *uh* (Swerts, 1998). And like the strong forms of *to*, *the*, and *a*, it is *um* that is more specialized. As we noted earlier, almost all speakers who use only one of the two forms use *uh*. This suggests that speakers are expected to signal delays with *uh* unless they have reason to use *um*.

Perhaps the best evidence that *uh* and *um* are words phonologically comes from the manner in which they cliticize onto *to*, *the*, and *a* (see Clark & Wasow, 1998). Consider these three recorded utterances (Alan is the same speaker as in (28)):

<sup>9</sup> We used times provided by Swerts (pers. commun.) correcting an error in the paper.

- (31) Alan and from the-uh (0.6) spectator point of view it looks like airplanes going in all directions
- (32) Alan we-um (0.7) have a-uh (0.1) pyro-techniques team that's-uh on the ground or in the water
- (33) Albert Gore I think that the government's role should not be to-uh regulate content, obviously (television, October 2000)

In (31), when Alan inserted *uh* after *the*, he could have produced “*thuh uh*”. Instead, he cliticized *uh* onto *the* and produced “*thi.yuh*”. To cliticize *uh* onto *the*, he had to create a prosodic word with trochaic stress. In a trochee, the first syllable has greater stress than the second, so *the* must be pronounced “*thiy*”, with a non-reduced vowel. Once the syllable boundaries are adjusted, the result is “*thi.yuh*”. Similarly, *a* plus *uh* goes to “*e.yuh*”, as in (32), and *to* plus *uh* goes to “*tu.wuh*”, as in (33). These forms are common in recorded corpora (see also Fox Tree & Clark, 1997; Jefferson, 1974).

*Uh* and *um* aren't always clitics. Some speakers produce “and um” on some occasions and “an.dum” on others. One hypothesis is that they use “and um” to signal a delay after *and*, but “an.dum” to signal both a suspension of speaking (at the end of *and*) plus a delay (Fox Tree & Clark, 1997). The point is that in (31) Alan did not formulate “*thuh*” and then, once he had begun to articulate it, add *uh*. That would have led to “*thuh uh*”. To produce the trochee “*thi.yuh*”, he must have formulated “*thiy*” plus “*uh*” as a piece. In doing so, he signaled simultaneously that he was (1) saying *the*, (2) immediately suspending his speech, and (3) delaying the resumption of speech. Parts (2) and (3) are collateral signals.

To summarize, speakers formulate *uh* and *um* as parts of the current utterance or discourse. Within intonation units, they formulate *uh* and *um* with a pitch, or  $F_0$ , that segregates them from the surrounding words. They may also formulate *uh* or *um* as ad hoc intonation units on their own, or as the final elements of ad hoc units. And speakers regularly cliticize *uh* and *um* onto prior words, forming prosodic words such as “an.dum” and “thi.yuh”. If we assume that speakers must formulate prosodic words as a piece – for example, to get the stress pattern right – they must be formulating *and + um* as a unit as well. To do that, they must be treating *uh* and *um* phonologically as words.

## 9. Conclusions

*Uh* and *um*, we have argued, are not merely “filled pauses”, audible counterparts to silent pauses. They are English words – interjections – with all the properties that this implies. Still, *uh* and *um* are special because they refer to on-going performance. Speakers use *uh* and *um* to announce that they are initiating what they expect to be a minor or major delay before speaking. Producing and interpreting *uh* and *um* are also special because, as interjections, they do not arise as part of the syntax of the sentences uttered. We first review the evidence for *uh* and *um* as words and then suggest a process by which they are produced.

### 9.1. Evidence that *uh* and *um* are words

To be an English word is to conform to the phonology, prosody, syntax, semantics, and pragmatics of English words.

### 9.1.1. Phonology

*Uh* and *um* consist of standard English segments in standard English syllables. As Levelt (1989, p. 483) noted, it would be suspicious if *uh* existed “with only minor phonetic variations, in many if not all languages”, making it “the only universal word”. Indeed, English has at least two contrasting phonological shapes (*uh* and *um*), and so do all of the other languages we have looked at.

### 9.1.2. Prosody

*Uh* and *um* conform to English prosody. When placed within intonation units, they are normally delivered with a parenthetical intonation, a monotone pitch that allows them to be segregated from the melody of the surrounding construction. Some uses of *uh* and *um* finish off fragments of phrases with unit-final intonation; other uses are separated off in their own intonation units. For an item to carry intonation and serve these functions, it cannot be a non-linguistic sound or noise. It must be a word or morpheme.

A crucial feature of *uh* and *um* is that they are often cliticized onto prior words. When cliticized onto *and*, *but*, *the*, *a*, and *that*, for example, they form prosodic words with trochaic stress and adjusted syllable boundaries: “an.duh”, “bu.tum”, “thi.yum”, “ai.yuh”, and “tha.tuh”. *Uh* and *um* could not take part in this process if they were non-linguistic sounds.

### 9.1.3. Syntax

Interjections, by definition, do not take part in syntactic constructions, although they get their meaning in part from their placement within such constructions. When George says “Sunday . the twenty-fifth, - sorry twenty-fourth (211a.173)”, for example, he uses “sorry” to point backward to a recent offense (the error “twenty-fifth”) and to point forward to a fix for the offense (the repair “twenty-fourth”). *Uh* and *um* are no different. Speakers use them to pinpoint the initiation of an expected delay and to point forward to future speech.

### 9.1.4. Semantics

*Uh* and *um* are interjections whose basic meanings are conventional in British and North American English:

(a) *Uh*: “Used to announce the initiation, at  $t$ (‘uh’), of what is expected to be a minor delay in speaking.”

(b) *Um*: “Used to announce the initiation, at  $t$ (‘um’), of what is expected to be a major delay in speaking.”

The meanings of *uh* and *um* contrast in the importance of the delay they initiate. Even if *uh* were the only filler in a person’s lexicon, it would contrast with silence and mean “used to announce the initiation, at  $t$ (‘uh’), of what is expected to be a brief delay in speaking”. In other languages, fillers contrast on other dimensions as well (see, e.g. Emmett, 1998, on Japanese). Speakers can use *uh* and *um* in their basic meanings to implicate a wide range of other things, such as “I want to keep the floor”, “I want to give up the floor”, “I’m uncertain about what I want to say”, or “I’m hunting for the next word”.

### 9.1.5. Use

*Uh* and *um* are signals, not symptoms. One criterion for signals is choice: if one element is selected over another, and that selection contributes to a contrast in what is meant, then



that element is a signal (see Fox Tree & Clark, 1997). Speakers select among *uh*, *um*, and silence (and probably other items too, including gestures), and that selection contributes to what they mean. Indeed, speakers have control over their selection both at the fine-grained level (contrasting minor and major delays) and at coarse-grained levels (contrasting formal vs. informal speaking).<sup>10</sup>

Our findings suggest that *uh* and *um* are controlled in part from top down. When people talk, they make broad, high-level decisions about whether to use formal or informal language, polite or locker-room language, and adult or baby talk, and these guide a range of low-level decisions about words, phonology, and syntax. We assume that people also make broad, high-level decisions about whether to mention or hide preparedness problems, and this guides low-level decisions about *uh* and *um*. At the lower levels, speakers select *the* or *a*, *in* or *on*, and *this* or *a* based on fine-grained representations of what they are trying to say. They are usually unaware of these selections. We assume that the same holds for selecting *uh* or *um* or silence.

*Uh* and *um* are fundamentally different from another signal of delay: syllable prolongation. As we found, speakers can prolong a syllable to signal that they are continuing an on-going delay. Although they select *uh* or *um* to signal a delay that *begins* at the filler, they can prolong the filler to signal a delay already in progress. In the LL corpus, selecting *uh* vs. *um* and prolonging a filler are independent signals. *Um* is not merely a prolonged *uh*, but contrasts with *uh* in its own way. Prolongation also differs from *uh* and *um* in form. Prolongation is not a word or morpheme, but a *process* applied to parts of words. If the same process applies to *uh* and *um*, they, too, must be words.

To sum up, *uh* and *um* satisfy the criteria for being English words. They have conventional forms and meanings, conform to the notion of word syntactically and prosodically, and contrast with another signal of delay, the process of prolongation.

## 9.2. Producing *uh* and *um*

In the model of production we assume, speakers plan two tracks of messages in parallel – a primary and collateral track (Clark, 1996, in press). Take (16), repeated here, in which Alan answers a question about “recent novels” he has read:

(16) Alan I’ve {u:m} recently read {u:m . oh, .} Lord of the Flies (3.5a.110)

Alan answers the question by formulating the clause *I’ve recently read Lord of the Flies*, which represents his primary message. But formulating and producing utterances is an incremental process (Griffin, 2001; Kempen, 1996; Kempen & Hoenkamp, 1987; Levelt, 1989), and Alan produces his clause in three increments (“I’ve” and “recently read” and “Lord of the Flies”), delaying before each increment. He comments on the delays with the asides “u:m”, “u:m”, and “oh”, which represent collateral messages. The issue is how speakers develop messages, formulate expressions, and produce them in the two tracks in parallel.

<sup>10</sup> Experienced users of email and computer chat sessions make extensive use of *uh* and *um*. Their uses are deliberate, have intended interpretations, and otherwise fit the definition of word, not syntom.

The main problem is how to *merge* the two tracks of messages in a single stream of behavior. As already mentioned, speakers solve this problem with four main techniques, often used in combination. With inserts (like “I mean”), they interrupt the primary message or introduce the collateral one. With juxtapositions (like “Max said Max was”), they cut off the primary message and initiate a replacement. With modifications (like prolonging “and”), they combine the two messages in the same expressions. And with concomitants (like certain gestures), they continue speaking while adding the gestures. In this classification, *uh* and *um* are inserts, whereas prolonging *uh* and *um*, and cliticizing them onto previous words, are modifications.

Our proposal is that speakers add these collateral signals by a process that is an elaboration of the Levelt (1983) model for self-repairs. The process has two steps:

*Step 1:* Monitor your speech plans and speech output for problems worthy of comment or repair.

*Step 2:* The moment you discover such a problem:

- (a) formulate a collateral signal for making the comment or repair, and
- (b) select a method for adding the signal to the primary utterance.

Consider “Mallet said Mallet was” in (2). Once Reynard identifies “said” as inappropriate, he formulates a plan for its repair (see Levelt, 1983): he will suspend speaking after “said” (step 2b) and juxtapose a replacement (“Mallet was”) that starts at a previous constituent boundary (step 2a). Steps 2a and 2b depend on each other and may be carried out in either order or simultaneously, depending on conditions.

In this model, producing a collateral signal requires formulating not only the signal itself (part 2a), but also a method of merging it with the primary signal (part 2b). For inserts and juxtapositions, this requires formulating a point of suspending on-going speech. For modifications, it requires specifying which items are to be modified, and for concomitants, over what items the gesture or other signal is to be performed. This model readily applies to repairs of all types (Blackmer & Mitton, 1991; Brédart, 1991; Levelt, 1983), inserts such as *I mean*, repeated words, as in “I {uh} I wouldn’t be surprised at that” (Clark & Wasow, 1998), and non-reduced vowels, as in “thiy { . } the literature” (Fox Tree & Clark, 1997). It also applies to *uh* and *um* and prolongations.

At step 1, speakers monitor for delays that are worthy of comment. According to our data, they look for delays at all major levels of planning – from retrieving a word to deciding what to talk about next. But which delays are “worthy of comment”? Speakers have at least four criteria: (1) delays within clauses are more worthy than delays at clause boundaries; (2) longer delays are more worthy than shorter ones; (3) it is generally preferable to signal delays in spontaneous speech but not in formal speech; and (4) it is desirable to comment on delays that may lead to undesirable inferences.

At step 2a, speakers formulate a signal for commenting on the anticipated delay. If the delay is not yet in play, they can formulate *uh* or *um* as inserts. If the delay is expected to be minor, they will formulate *uh*, and if major, *um*. If the delay is on-going, they can formulate the prolongation of a current word or syllable, including *uh* or *um*.

Speakers also have options at step 2b. Suppose they are preparing to insert *uh* and *um*. One option is *where* to suspend speaking. Speakers could, in principle, interrupt them-

selves the moment they discover the upcoming delay, but they almost always wait for a word boundary, just as they do before all but one type of repair (Levelt, 1983). Another set of options is *how* to suspend speaking. One option is to formulate a simple cut-off and *then* produce *uh* or *um*, as in (21), “you { . u:h } they spent { - uh }”. But when speakers have selected at least one word before the anticipated delay, they can, instead, formulate that word with *uh* or *um* cliticized onto it. In many cases, this requires an adjustment of syllable boundaries, as in “an.duh” and “righ.tuh”. In some cases, it also requires formulating a non-reduced vowel, as in “thi.yum” or “tu.wuh”. Speakers have similar options in prolonging syllables – which syllables to prolong and for how long.

It is the processes at steps 2a and 2b, in short, that lead to the selection of *uh* vs. *um*, cliticized vs. separate, and prolonged vs. normal. Speakers may be able to plan prolongations at a late stage in processing, but they must select *uh* vs. *um*, and cliticized vs. separate, at earlier stages of formulation. The evidence so far leaves much to be specified about when, how, and why speakers make these selections.

In summary, speakers produce *uh* and *um* in much the same way they do other words. They select a message (“I am initiating what I expect to be a minor delay in speaking”), formulate *uh* as an insert that expresses that message, and then produce it. It is just that, unlike most words, they use *uh* and *um* exclusively for comments on the on-going performance.

## Acknowledgements

This research was supported by NSF Grants SBR-9309612 and IRI-9314967, by ATR, by an American Association of University Women Educational Foundation Summer Research Publication Grant, and faculty research funds from the University of California at Santa Cruz, and by Grant N00014-00-1-0660 from the Department of Defense Multi-disciplinary University Research Initiative (MURI) program administered by the Office of Naval Research. We thank Eve V. Clark, Zenzi Griffin, Willem Levelt, Elizabeth E. Shriberg, and Dan Sperber for constructive comments on earlier drafts of the paper, and we are indebted to a host of colleagues for discussion of this research.

## References

- Allwood, J., Nivre, J., & Ahlsén, E. (1990). Speech management – on the non-written life of speech. *Nordic Journal of Linguistics*, *13*, 3–48.
- American heritage dictionary of the English language* (4th ed.). (2000). Boston, MA: Houghton Mifflin.
- Barr, D. J. (1998). Trouble in mind: paralinguistic indices of effort and uncertainty in communication. In S. Santi, I. Guaitella, C. Cave & G. Konopczynski, *Oralité et gestualité, communication multimodal, interaction* (pp. 597–600). Paris: L'Harmattan.
- Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning. An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology*, *19*, 163–194.
- Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*, 469–489.
- Beattie, G. W. (1983). *Talk: an analysis of speech and non-verbal behaviour in conversation* Milton Keynes: Open University Press.

- Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39, 173–194.
- Bloomfield, L. (1933). *Language*. New York: H. Holt.
- Bolinger, D. L. (1989). *Intonation and its uses*. Stanford, CA: Stanford University Press.
- Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, 8, 148–158.
- Brédart, S. (1991). Word interruption in self-repairing. *Journal of Psycholinguistic Research*, 20 (2), 123–138.
- Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: MIT Press.
- Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34 (3), 383–398.
- Brody, J. (1987). Particles borrowed from Spanish as discourse markers in Mayan languages. *Anthropological Linguistics*, 29 (4), 507–521.
- Broen, P. A., & Siegel, G. M. (1972). Variations in normal speech disfluencies. *Language and Speech*, 15 (3), 219–231.
- Brotherton, P. (1979). Speaking and not speaking: processes for translating ideas into speech. In A. W. Siegman & S. Feldstein (Eds.), *Of speech and time* (pp. 179–209). New York: Wiley.
- Chafe, W. (1979). The flow of thought and the flow of language. In T. Givón (Ed.), *Discourse and syntax. Syntax and semantics* (pp. 159–181), 12. New York: Academic Press.
- Chafe, W. (1980). *The pear stories*. Norwood, NJ: Ablex.
- Chafe, W. (1992). *Intonation units and prominences in English natural discourse*. Paper presented at the University of Pennsylvania prosodic workshop, Philadelphia, PA.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Christenfeld, N. (1995). Does it hurt to say um? *Journal of Nonverbal Behavior*, 19 (3), 171–186.
- Christenfeld, N. (1996). Effects of a metronome on the filled pauses of fluent speakers. *Journal of Speech & Hearing Research*, 39 (6), 1232–1238.
- Christenfeld, N., & Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology*, 70 (3), 451–460.
- Clark, H. H. (1994). Discourse in production. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 985–1021). San Diego, CA: Academic Press.
- Clark, H. H. (1994). Managing problems in speaking. *Speech Communication*, 15, 243–250.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H. (1999). *Passing time in conversation*. Paper presented at the workshop on the structure of spoken and written texts, University of Texas, Austin.
- Clark, H. H. (in press). Pragmatics of language performance. In L. R. Horn, & G. Ward (Eds.), *Handbook of pragmatics*. Oxford: Blackwell.
- Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37 (3), 201–242.
- Cook, H. M. (1993). Functions of the filler *ano* in Japanese. In S. Choi (Ed.), *Japanese/Korean linguistics* (pp. 19–38), 3. Stanford, CA: CSLI.
- Cook, M. (1969). Anxiety, speech disturbances and speech rate. *British Journal of Social and Clinical Psychology*, 8, 13–21.
- Cook, M. (1971). The incidence of filled pauses in relation to part of speech. *Language and Speech*, 14 (2), 135–139.
- Cook, M., & Lalljee, M. G. (1972). Verbal substitutes for visual signals in interaction. *Semiotica*, 6, 212–221.
- Cook, M., Smith, J., & Lalljee, M. G. (1974). Filled pauses and syntactic complexity. *Language and Speech*, 17 (1), 11–16.
- Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.
- Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
- Crystal, D., & Davy, D. (1975). *Advanced English conversation*. London: Longman.
- Curme, G. O. (1935). *Parts of speech and accidence*. Boston, MA: D.C. Heath.
- Deese, J. (1984). *Thought into speech: the psychology of a language*. Englewood Cliffs, NJ: Prentice-Hall.
- DuBois, J. (1974). Syntax in mid-sentence. In C. Fillmore, G. Lakoff & R. Lakoff (Eds.), *Berkeley studies in syntax and semantics* (pp. III.1–III.25), I. Berkeley, CA: University of California, Institute of Human Learning and Department of Linguistics.
- Ducrot, O. (1980). *Dire et ne pas dire: principes de sémantique linguistique*. (2nd ed.). Paris: Hermann.
- Duez, D. (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, 25, 11–28.

- Duez, D. (1991). *La pause dans la parole de l'homme politique*, Paris: Editions du Centre National de la Recherche Scientifique.
- Duez, D. (1993). Acoustic correlates of subjective pauses. *Journal of Psycholinguistic Research*, 22 (1), 21–39.
- Eklund, R. (1999). *A comparative study of disfluencies in four Swedish travel dialogue corpora*. Paper presented at the International Conference of Phonetic Sciences satellite meeting on disfluencies, Berkeley, CA.
- Emmett, K. S. (1996). *Contribution of the interactional markers ano(o), eeto, uun(to), and nanka to the development of Japanese conversation*. Unpublished master's thesis, University of Minnesota, Minneapolis.
- Emmett, K. S. (1998). *Ano(o)* is more than “um”: interactional functions of *ano(o)* in Japanese conversation. In M. C. Chalasani, J. A. Grocer & P. C. Haney (Eds.), *Proceedings of the fifth annual symposium about language and society* (pp. 136–148), Vol. 39. Austin, TX: University of Texas, Department of Linguistics.
- Erman, B. (1987). *Pragmatic expressions in English: a study of you know, you see, and I mean in face-to-face conversation*, Stockholm: Almqvist & Wiksell International.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100 (2), 233–253.
- Ferreira, F. (2000). Syntax in language production: an approach using tree-adjointing grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Philadelphia, PA: Psychology Press/Taylor & Francis.
- Fischer, K. (1999). *From cognitive semantics to lexical pragmatics: the functional polysemy of discourse particles*. Unpublished PhD dissertation, Universität Hamburg.
- Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134–184). Cambridge: Cambridge University Press.
- Ford, M. (1982). Sentence planning units: implications for the speaker's representation of meaningful relations underlying sentences. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 798–827). Cambridge, MA: MIT Press.
- Fox Tree, J. E. (1995). Effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34, 709–738.
- Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, 27 (1), 35–53.
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition*, 29 (2), 320–326.
- Fox Tree, J. E. (in press). Interpreting pauses and ums at turn exchanges. *Discourse Processes*.
- Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, 62 (2), 151–167.
- Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: oh what a difference an oh makes. *Journal of Memory and Language*, 40 (2), 280–295.
- Fox Tree, J. E., & Schrock, J. C. (in press). Basic meanings of you know and I mean. *Journal of Pragmatics*.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Gee, J. P. (1986). Units in the production of narrative discourse. *Discourse Processes*, 9, 391–422.
- Godfrey, J. J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: telephone speech corpus for research and development. *Proceedings of the IEEE conference on acoustics, speech, and signal processing* (pp. 517–520). San Francisco, CA: IEEE.
- Goffman, E. (1981). Radio talk. In E. Goffman (Ed.), *Forms of talk* (pp. 197–327). Philadelphia, PA: University of Pennsylvania Press.
- Goldman-Eisler, F. (1968). *Psycholinguistics: experiments in spontaneous speech*. New York: Academic Press.
- Goodwin, C. (1981). *Conversational organization: interaction between speakers and hearers*, New York: Academic Press.
- Goodwin, C. (1987). Forgetfulness as an interactive resource. *Social Psychology Quarterly*, 50 (2), 115–131.
- Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62, 51–75.
- Grice, H. P. (1957). Meaning. *Philosophical Review*, 66, 377–388.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts Syntax and semantics* (pp. 113–128), Vol. 3. New York: Seminar Press.
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82, B1–B14.
- Grimes, J. E. (1975). *The thread of discourse*. The Hague: Mouton.
- Grosz, B., & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. In J. Ohala (Ed.),

- Proceedings of the international conference on spoken language processing* (pp. 429–432), Vol. 1. Banff, Alberta, Canada.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English. Part 2. *Journal of Linguistics*, 3, 199–244.
- Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: studies in conversation analysis* (pp. 299–345). Cambridge: Cambridge University Press.
- Hinds, J. (1975). Interjective demonstratives in Japanese. *Descriptive and Applied Linguistics*, 8, 19–42.
- Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. In D. Schiffrin (Ed.), *Meaning, form and use in context: linguistic application* (pp. 11–42). Washington, DC: Georgetown University Press.
- James, D. (1972). Some aspects of the syntax and semantics of interjections. *Papers from the eighth regional meeting of the Chicago Linguistic Society* (pp. 162–172). Chicago, IL: Chicago Linguistic Institute.
- Jefferson, G. (1973). A case of precision timing in ordinary conversation: overlapped tag-positioned address terms in closing sequences. *Semiotica*, 9, 47–96.
- Jefferson, G. (1974). Error correction as an interactional resource. *Language in Society*, 2, 181–199.
- Kasl, S., & Mahl, G. (1965). The relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, 1, 425–433.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11 (2), 201–258.
- Kempen, G. A. M. (1996). Computational models of syntactic processing in language comprehension. In T. Dijkstra & K. de Smedt (Eds.), *Computational psycholinguistics: AI and connectionist models of human language processing* (pp. 192–220). Philadelphia, PA: Taylor & Francis.
- Kenstowicz, M. (1993). *Phonology in generative grammar*. Oxford: Blackwell.
- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. Cambridge: Cambridge University Press.
- Koopmans-van Beinum, F. J., & van Donzel, M. E. (1996). Discourse structure and its influence on local speech rate. *Proceedings of the Institute of Phonetic Sciences* (pp. 1–11), Vol. 20. Amsterdam: Institute of Phonetic Sciences.
- Kowal, S., O'Connell, D., Forbush, K., Higgins, M., Clarke, L., & D'Anna, K. (1997). Interplay of literacy and orality in inaugural rhetoric. *Journal of Psycholinguistic Research*, 26 (1), 1–31.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, S. C. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.
- Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15, 19–44.
- Mahl, G. F. (1987). *Explorations in nonverbal and vocal behavior*. Hillsdale, NJ: Erlbaum.
- Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8 (1), 1–71.
- Marslen-Wilson, W. D., & Tyler, L. K. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society of London*, 295 (1077, Series B), 317–332.
- Martin, J. G. (1967). Hesitations in the speaker's production and listener's reproduction of utterances. *Journal of Verbal Learning and Verbal Behavior*, 6 (6), 903–909.
- Maschler, Y. (1997). Discourse markers at frame shifts in Israeli Hebrew talk-in-interaction. *Pragmatics*, 7 (2), 183–211.
- Mitchell, D. C. (1994). Sentence parsing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 375–409). San Diego, CA: Academic Press.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
- O'Donnell, W. R., & Todd, L. (1991). *Variety in contemporary English*. (2nd ed.). New York: Harper Collins Academic.
- Oxford English dictionary*. (2000). Oxford: Oxford University Press.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A grammar of contemporary English*. New York: Seminar Press.
- Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, 2 (1), 51–81.
- Sacks, H., & Schegloff, E. (1979). Two preferences in the organization of reference to persons in conversation and

- their interaction. In G. Psathas (Ed.), *Everyday language: studies in ethnomethodology* (pp. 15–21). New York: Irvington.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696–735.
- Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, 60 (3), 362–367.
- Schegloff, E. A. (1982). Discourse as an interactional achievement: some uses of “uh huh” and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: text and talk. Georgetown University Roundtable on Languages and Linguistics 1981* (pp. 71–93). Washington, DC: Georgetown University Press.
- Schegloff, E. A. (1987). Recycled turn beginnings: a precise repair mechanism in conversation’s turn-taking organization. In G. Button & J. R. E. Lee (Eds.), *Talk and social organization* (pp. 70–85). Clevedon: Multilingual Matters.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361–382.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schourup, L. C. (1982). *Common discourse particles in English conversation*. New York: Garland.
- Selkirk, E. O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.
- Selkirk, E. (1996). The prosodic structure of function words. In K. Demuth & J. L. Morgan (Eds.), *Signal to syntax: bootstrapping from speech to grammar in early acquisition* (pp. 187–213). Mahwah, NJ: Lawrence Erlbaum Associates.
- Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Unpublished PhD dissertation, University of California, Berkeley.
- Shriberg, E. E., & Lickley, R. J. (1993). Intonation of clause-internal filled pauses. *Phonetica*, 50, 172–179.
- Siegmán, A. W. (1979). Cognition and hesitation in speech. In A. W. Siegmán & S. Feldstein (Eds.), *Of speech and time* (pp. 151–178). Hillsdale, NJ: Erlbaum.
- Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32, 25–38.
- Sperber, D., & Wilson, D. (1986). *Relevance*. Cambridge, MA: Harvard University Press.
- Stalnaker, R. C. (1978). Assertion. In P. Cole, *Pragmatics Syntax and semantics* (pp. 315–332), 9. New York: Academic Press.
- Stenström, A.-B. (1990). Pauses in monologue and dialogue. In J. Svartvik (Ed.), *The London-Lund corpus of spoken English: description and research* (pp. 211–252). Lund: Lund University Press.
- Svartvik, J. & Quirk, R. (1980). *A corpus of English conversation*, Lund: Gleerup.
- Svennevig, J. (1999). *Getting acquainted in conversation: a study of initial interactions*, Amsterdam: John Benjamins.
- Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30, 485–496.
- Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37 (1), 21–43.
- Underhill, R. (1988). Like is, like, focus. *American Speech*, 63 (3), 234–246.
- Van der Vlugt, M. (1987). *Een inleidende studie van het gebruik en de functie van de stopwoordjes ‘pues’, ‘este’ en ‘eh’ in de gesproken taal van Mexico City*. Unpublished Scriptie, Universiteit Leiden, The Netherlands.
- Watts, R. J. (1989). Taking the pitcher to the ‘well’: native speakers’ perception of their use of discourse markers in conversation. *Journal of Pragmatics*, 13, 203–237.
- Wheeldon, L., & Levelt, W. J. M. (1995). Monitoring the time course of spoken word production. *Journal of Memory and Language*, 34, 311–334.
- Wilkins, D. P. (1992). Interjections as deictics. *Journal of Pragmatics*, 17, 119–158.