ELSEVIER

Journal of Pragmatics 30 (1998) 485–496

journal of PRAGMATICS

Filled pauses as markers of discourse structure☆

Marc Swerts*

Centrum voor Nederlandse Taal en Spraak, Universitaire Instelling Antwerpen (UIA), Universiteitsplein 1, B-2610 Wilrijk, Belgium and IPO, Center for Research and User-System Interaction, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

Received 2 May 1997; revised version 4 December 1997

Abstract

This study aims to test whether filled pauses (FPs) may highlight discourse structure. This question is tackled from the perspectives of both the speaker and the listener. More specifically, it is first investigated whether FPs are more typical in the vicinity of major discourse boundaries. Secondly, FPs are analyzed acoustically, to check whether those occurring at major discourse boundaries are segmentally and prosodically different from those at shallower breaks. Analyses of twelve spontaneous monologues (Dutch) show that phrases following major discourse boundaries more often contain FPs. Additionally, FPs after stronger breaks tend to occur phrase-initially, whereas the majority of the FPs after weak boundaries are in phrase-internal position. Also, acoustic observations reveal that FPs at major discourse boundaries are both segmentally and prosodically distinct. They also differ with respect to the distribution of neighbouring silent pauses. Finally, a general linear model reveals that discourse structure can to some extent be predicted from characteristics of the FPs.

1. Introduction

One of the typical differences between a well-prepared lecture and daily small-talk is that the latter speaking mode generally contains comparatively more disfluencies. These may to some extent be language-specific such as the hesitation morphemes 'eeto' and 'anoo' in Japanese. Spontaneous speakers of many other languages, like English, regularly produce filled pauses (FPs), though the sounds of these may differ between various dialects or languages (Cruttenden, 1986). This

☆ Marc Swerts is also affiliated with the Fund for Scientific Research – Flanders (FWO) in Belgium. The study reported on in this paper was presented in October 1996 at the International Conference on Spoken Language Processing in Philadelphia (PA). Thanks are due to Liz Shriberg for inspiring ideas, to Robbert-Jan Beun for providing the speech materials, to Anne Wichmann for fruitful discussions, and to René Collier, Hanny den Ouden, and two reviewers of the Journal of Pragmatics for comments on an earlier version of this manuscript.

* Corresponding address: IPO, Center for Research on User System Interaction, Spoken Language Interfaces, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands.

0378-2166/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved

PII S0378-2166(98)00014-9

study will concentrate on FPs in Dutch, where they normally appear as 'uh' [əh] or 'um' [əm].

Since the seminal work by Goldman-Eisler (1968), FPs have received much attention. One reason to study them is that they are believed to shed light on the speech production process. Just as other so-called flaws in spontaneous speech production, like syntactic errors, slips of the tongue and other mispronunciations (Fromkin, 1973, 1980; Levelt and Cutler, 1983; Nooteboom, 1973), FPs are thought to be indicative of the mental processes underlying speech generation. They are viewed as signals that hint e.g. to a speaker's word-searching problems (e.g. Goldman-Eisler, 1968) or to difficulties in conceptualization at major discourse boundaries (Chafe, 1980).

Next to this speaker-oriented approach, there has been some research into how FPs and other disfluencies are processed by a listener. Many people claim that they tend to go unnoticed (e.g. Lickley and Bard, 1996); others maintain that they hinder cognitive processing (e.g. Fox Tree, 1995). Particularly the latter idea could imply that spontaneous speech with all its production errors is a deficient form of language usage. This, however, is in conflict with the finding that spontaneous speech, despite its hesitant nature, is often a better communicative means than fluent, read speech. It has, for instance, been shown that media professionals in interviews tend to speak spontaneously instead of reading their questions in order to secure effective interaction, showing that 'spontaneous' is not necessarily synonymous with 'functionally inadequate' (Kowal et al., 1985).

Others even have presented evidence that FPs themselves have communicative import. Goldman-Eisler already remarked that words following hesitations have a low transitional probability and thus a high information value. Consequently, FPs may be useful for listeners, since they may presignal upcoming important linguistic materials (Fox Tree, 1993; Shriberg and Stolcke, 1996). The idea that FPs have functional validity is even more strongly expressed by investigators who have studied them from a discourse perspective. In a dialogue, they may function pragmatically as indicators of the Feeling-of-Another's-Knowing (Brennan and Williams, 1995), as turn-holders (Stenström, 1994; Maclay and Osgood, 1959) or as a speaker's cue to obtain the gaze of his recipient during the course of a turn at talk (Goodwin, 1981). Other work describing FPs as an interactional phenomenon can be found in Levinson (1983) and Clark (1994).

FPs have also been studied in monologues. Swerts and Geluykens (1994) already noticed that one of their speakers very consistently put 'um' at the onset of a major topical unit. Chafe (1980) claimed that a change of a major idea unit is often accompanied by disfluencies. This is compatible with the finding that unit-initial utterances tend to be more hesitant than other ones (Brubaker, 1972; Swerts and Ostendorf, 1997), which is explained by Brubaker on the basis of a 'reduction-of-uncertainty' hypothesis, i.e., that rate increases and pause decreases as uncertainty about the remaining content of the paragraph decreases.

The present study addresses the question whether FPs may be informative about larger-scale discourse segments, i.e., units that exceed the level of the sentence. The potential relationship of FPs to discourse structure will be tackled here from a double perspective. One question is to what extent structural properties of a given discourse can predict the distribution of (particular types of) FPs. Conversely, one may ask to what extent the occurrence of a (given type of) FP may be a cue to discourse segments. The motivation for such a dual approach is that it should reveal the relevance of FPs for the two main partners in the communication chain, the speaker and the listener. Next to this, the paper not only strives at a quantitative distributional analysis of FPs, but it will also focus on their segmental and prosodic features.

2. The data: Discourse structure

Since FPs have been claimed to be interactional phenomena, it seems logical to investigate them in truly natural conversations. However, in such speech materials, FPs are likely to be the result of many factors (interactive, cognitive, social, ...), which makes these data more difficult to analyze than simple narratives by a single speaker. In the current study, therefore, investigations were based on twelve elicited monologues, i.e., six painting descriptions produced by two female speakers of Dutch (amounting to 46.5 minutes of speech), which contained 310 FPs in total. An earlier analysis of these data (Swerts, 1997) brought to light that the hierarchical level of discourse boundaries is reflected in at least three prosodic features: pause length, amount of pitch reset, and type of boundary tone. In that study, the discourse structure of the monologues was established by instructing nineteen subjects to mark perceived paragraph boundaries in orthographic transcriptions of the spoken texts, while they could hear the speech.[1] Inspired by Rotondo (1984), boundary strength was then defined as the proportion of subjects that marked a specific phrase boundary as a breaking point between two consecutive paragraphs.[2] More details can be found in Swerts (1997).

A typical example of part of a text is given below, followed by a literal translation in English. The digits in round brackets represent the boundary strength estimates larger than 0, computed as the proportion of subjects indicating that there was a break.

(1) het is echt een paard dat [uh] over iets heen springt heel heel snel (0.26) de man die d'r opzit die zit ook helemaal in zo'n gebogen [uh] [uh] ruitershouding met zijn billen omhoog en zijn [uh] hoofd in de manen van het paard (0.95) het paard is wit (0.11) [uh] ruiter is is [uh] rozig rood (0.53)
(it is really a horse that [uh] jumps across something very very fast the man who sits on it he really sits also in such a bent over [uh] [uh] rider's position with his

[1] Actually, for reasons of comparison, Swerts (1997) also had nineteen other subjects perform the same labeling task on a text-alone basis, i.e., subjects could not hear the original speech. The present paper, however, will only focus on the outcome of the text-with-speech labeling.

[2] In this paper, the term 'paragraph' refers to a structural rather than an orthographic unit (Longacre, 1979). In the future, it may be interesting to study FPs in relation to other linguistic macro-units like keyed passages or parenthetical remarks (Goffman, 1981).

bum in the air and his [uh] head stuck in the mane of the horse the horse is white [uh] rider is is [uh] pinkish red)

The current study deals with the boundary strength estimates in the following way. The first part investigates to what extent discourse structure has an influence on the distribution of (particular types of) FPs. In this section, a categorical distinction is made between strong and weak boundaries, the former being defined as those phrase boundaries identified by at least 75% of the labelers as a paragraph transition, whereas the other phrase boundaries are considered to represent weak boundaries. The second part looks at the cue value of given types of FP for discourse segments and treats boundary strength as a continuous feature which may vary between 0 and 1.

The monologues were fed into a computer with a 16 kHz sampling frequency. For the prosodic analyses discussed below, fundamental frequency (F_0) was determined by the method of subharmonic summation (Hermes, 1986).[3] Duration of the FPs was measured directly in the digitized waveform. Below results for both speakers are pooled; previous analyses (Swerts, 1997) showed that they behave very comparably with respect to prosodic discourse marking.

3. From discourse structure to filled pauses

3.1. Distribution

A first exploration concentrated on how FPs were distributed as a function of different discourse boundaries. Countings of FPs were done by considering all prosodic phrases following either a weak or a strong boundary. A first check was whether or not such a phrase contained one or more FPs. If it did, a distinction was made between phrases with an FP occurring in initial position, and those which did not have an FP at their onset. If an FP occurred at a boundary, it was analyzed as being at the beginning of the next phrase, rather than at the end of the previous phrase. The outcome of these countings is shown in Table 1.

First, the table reveals that FPs are somewhat more typical in the vicinity of stronger boundaries, since the majority of the phrases following a strong boundary have an FP (78%, 50 out of 64), whereas an FP only occurs in about 40% of the phrases following a weak boundary (176 out of 438). This difference in distribution is highly significant ($\chi^2 = 45.809$, df = 2, $p < 0.001$). In accordance with findings by Chafe (1980), it appears that the presence of a boundary makes a hesitation more likely, even though one cannot predict a boundary merely from the presence of a hesitation. Moreover, the majority of the phrases with FPs after a stronger boundary

[3] The fundamental frequency (F_0) is taken as the acoustic correlate of pitch.

Table 1
Distribution of phrases without FP, with FPs only in medial position and with at least one FP in initial position, as a function of weak versus strong discourse boundaries (DB) (further explanations in the text)

	without FP	medial FP	initial FP	Total
weak DB	262	102	74	438
strong DB	14	7	43	64
Total	276	109	117	502

have an FP in initial position. The mirror image is true for the phrases following weaker boundaries, most of which have an FP in non-initial position. In other words, these first countings suggest that discourse structure indeed has an influence on the distribution of FPs.

3.2. Phonetic analysis

From the preceding, it appears that discourse structure to some extent determines the probability that an FP will occur. One observation was that the position of an FP in a phrase is important, with initial FPs being more typical in phrases following stronger discourse breaks. Therefore, a tentative hypothesis would be that phrase-initial FPs reflect earlier/deeper processing (like conceptualization), whereas phrase-internal FPs indicate later processing like lexical search or local syntactic/phonological encoding (Shriberg, 1994). In this view, it is logical to find that initial FPs are more typical at the onset of larger-scale idea units. In order to test the idea that phrase-initial FPs have a special status from a discourse perspective, the next sections explore whether they exhibit specific segmental and prosodic characteristics that set them apart from medial ones.

3.2.1. Segmental features

A first analysis of initial versus medial FPs concentrates on the difference between 'uh' and 'um'. Shriberg (1994) reports that in the three different corpora she investigated, the forms 'uh' and 'um' show systematic differences in sentence position, the latter forms being more typical initially. Shriberg's tentative interpretation is that this distinction reflects different underlying problems in production. "The form 'um' may be used relatively more often during planning of larger units, while 'uh' may be relatively more likely to reflect local lexical-decision making" (1994: 154).

Additional support for this idea comes from Clark (1994; see also Smith and Clark, 1993), who presents evidence that the fillers 'uh' and 'um' differ with respect to their communicative function: the former are often used to signal short interruptions and the latter to signal more serious ones. This result appears from an experiment in which subjects were asked to respond to a set of 40 factual questions. The delay in their answers was longer when their utterances began with 'um' (average: 8.83 seconds), as opposed to responses beginning with 'uh' (average: 2.65 seconds).

On the basis of such previous work, one would therefore hypothesize that initial FPs in the current corpus are more likely to be of type 'um', since they occur relatively more often at major discourse boundaries. The distribution of 'uh' and 'um' as a function of their position in a phrase is given in Table 2. The data in this table basically confirm what one would expect from the literature. The dependency of the form of FP on phrase-position is highly significant ($\chi^2 = 126.967$, df = 1, $p < 0.001$).

Table 2
Distribution of 'uh' and 'um' as a function of their position in a phrase

	'uh'	'um'	Total
medial position	170	23	193
initial position	29	88	117
Total	199	111	310

3.2.2. Prosodic features

3.2.2.1. Pausal context. Guaïtella (1993) investigated the pausal context of FPs in a corpus of French interviews. She observed that 22.5% of the hesitations were preceded and followed by silence, versus 50% preceded by speech and followed by silence, 5% preceded by silence and followed by speech and 22.5% preceded and followed by speech. In other words, FPs can occur in different contexts, but Guaïtella does not specify which factors determine the distribution. Clark (1994) also reports data on the relationship of FPs with respect to surrounding pauses. Looking at the London-Lund corpus of English conversations, he noticed that the percentages of times that FPs were preceded and followed by perceptible silent intervals are dependent on the form of the FP. In these data, speakers produced 'uh' and 'um' quite often after silent pauses, but they were more likely to use 'um' than 'uh' when they anticipated further pauses.

Given the observation, illustrated in Table 2, that 'um' is much more typical in phrase-initial position, one would expect – given the outcome of Clark's study – that initial FPs are different from medial ones regarding the relative occurrence of surrounding silent pauses. Table 3 presents data on four different classifications, i.e. 'uh' and 'um' (i) without preceding or following pauses, (ii) with a pause only before, or (iii) after the FP, and (iv) with pauses at both sides. The table shows that proportionally more FPs have silent intervals at both sides when they occur in phrase-initial position. Conversely, medial FPs are relatively more often integrated with preceding and following speech. The difference in distribution for initial versus medial FPs is clearly above chance level ($\chi^2 = 137.080$, df = 3, $p < 0.001$).

3.2.2.2. Pitch and duration. Shriberg (1994) presents one of the few studies dealing with the acoustic properties of FPs. First, she found that 'uh' and 'um' are longer than the English indefinite article 'a', which is close to the schwa-like vowel in the FPs. In addition, Shriberg noticed that the F_0 of FPs in clause-internal position can be predicted from the F_0 of the closest preceding peak. Though she did not look at

Table 3
Distribution of filled pauses (FPs) with or without surrounding silent pauses (SP) as a function of their position in a phrase

	without SP	SP after	SP before	SP at both sides	Total
medial position	66	94	15	18	193
initial position	7	11	33	66	117
Total	73	105	48	84	310

tonal and durational features of FPs in onset position, her findings do suggest that the prosody of disfluencies is to some extent under the speaker's control. Therefore, it is useful to explore pitch and duration of FPs from a discourse perspective. These two prosodic parameters have been shown to function as important structuring devices. Therefore, it could be that they also vary on FPs as a function of structurally distinct discourse positions.

Moreover, Guaïtella (1993) reports that vocal hesitations are comparatively longer when they are both preceded and followed by silence. Also, in her data, the type of pitch contour of the FPs appears to be dependent on the pausal context. Given that the analysis of the monologues here reveals that initial and medial FPs differ with respect to surrounding silent pauses, one would expect – on the basis of Guaïtella's findings – that they are also distinct in terms of duration and pitch.

To this end, two types of prosodic analyses were performed: the average F_0 of a filled pause was measured (expressed in Hertz), as well as its duration (in milliseconds). Results are given in Table 4. It shows that initial FPs tend to be higher and longer in duration than non-initial ones; these differences are significant (pitch: $T = 5.01$, df = 308, $p < 0.001$; duration: $T = 11.33$, df = 308, $p < 0.001$). The pitch data are in agreement with the assumption that high F_0 values and rises are associated with the idea of beginning (Vaissière, 1995; Wichmann, 1991; Swerts and Geluykens, 1994).

Table 4
Mean pitch (in Hz) and mean duration (in millisec.) of filled pauses as a function of their position within phrase

Position	Pitch			Duration		
	$\bar{x}$	*s.d.*	*n*	$\bar{x}$	*s.d.*	*n*
medial	179.15	31.76	193	36.17	14.48	193
initial	196.10	23.37	117	58.67	20.36	117

Because the measurements in Table 4 comprise analyses of both 'uh' and 'um', data on pitch and duration were analysed for these forms separately, too (see Table 5). A General Linear Model Fit reveals that there are significant effects of the form of FPs on both pitch and duration (pitch: $F_{(1,306)} = 9.07$, $p < 0.001$; duration: $F_{(1,306)}$

= 98.80, $p < 0.001$) and of position (pitch: $F_{(1,306)} = 4.15$, $p < 0.05$; duration: $F_{(1,306)} = 15.48$, $p < 0.001$). The interaction between form and position was not significant, neither for pitch nor for duration. This implies that these factors had an independent effect on the prosodic features.

Table 5
Mean pitch (in Hz) and mean duration (in millisec.) of filled pauses as a function of their position within phrase and of their form ('uh' versus 'um')

Position	Form	Pitch			Duration		
		$\bar{x}$	*s.d.*	*n*	$\bar{x}$	*s.d.*	*n*
medial	uh	177.94	31.77	170	34.08	13.44	170
medial	um	188.04	30.92	23	51.65	12.56	23
initial	uh	183.76	30.34	29	38.07	15.39	29
initial	um	200.17	19.09	88	65.45	16.99	88

4. From filled pauses to discourse structure

The previous section has shown that discourse structure has an effect on the relative occurrence of different types of FPs. Phrases following major discourse boundaries contain FPs more often. Additionally, FPs after stronger breaks tend to occur phrase-initially, whereas the majority of the FPs after weak boundaries are in phrase-internal position. Also, acoustic observations reveal that initial FPs are both segmentally and prosodically different from non-initial ones.

The current section presents an approach which is opposite to the one sketched above. The discourse boundary strength is treated here as the 'dependent' variable and it is checked whether it is significantly influenced by the absence versus presence of different types of FPs. For the analyses, a number of simplifications were introduced. First, the prosodic features pitch and duration were not taken into consideration, because there was no predetermined and natural way to categorize these continuous variables into separate classes; moreover, variation in pitch and duration is already partly captured by the distinction between 'uh' and 'um', since these forms differ prosodically. Second, pausal context was reduced to a difference between FPs that were fully separated from preceding and following speech context, and those that were partly (or fully) integrated.

The average scores for discourse boundary strength (DBS), operationalized as proportion of paragraph boundary indications (see above), in different conditions is given in Table 6. It is clear that DBS varies as a function of the FP characteristics in the preceding phrase, with larger scores at the bottom of the table.

The distinction in average DBS between the cases with and without FPs present is significant ($T = 9.0587$, df = 569, $p < 0.001$). For those cases where an FP was present, the following general linear model was fitted to the data:

Table 6
Average discourse boundary strength (DBS) as a function of presence versus absence of filled pauses, type of filled pauses, differing with respect to position in phrase, occurrence of surrounding pauses and form (further explanations in the text)

present	position	separated	form	DBS
no				0.13
yes	medial	no	uh	0.29
yes	medial	no	um	0.29
yes	medial	yes	uh	0.22
yes	medial	yes	um	0.23
yes	initial	no	uh	0.19
yes	initial	no	um	0.41
yes	initial	yes	uh	0.58
yes	initial	yes	um	0.66

(2) DBS = position + separation + form

with the two-way and three-way interactions added. The outcome of applying this model is given in Table 7, which shows a significant influence of position on discourse boundary strength, whereas the effect of separation is only a trend. Form does not have a significant influence. The only significant interaction is that between position and separation: this is probably due to the fact that separation is only relevant for discourse boundary strength, when FPs are in initial position (see Table 6).

Table 7
Outcome of a general linear model fit which relates discourse boundary strength to a number of FP characteristics

Factor	*F* value	*p* value
position	$F_{(1,302)} = 7.16$	0.0079
separation	$F_{(1,302)} = 2.97$	0.0858
form	$F_{(1,302)} = 1.13$	0.2882
position × form	$F_{(1,302)} = 0.87$	0.3513
position × separation	$F_{(1,302)} = 6.75$	0.0098
separation × form	$F_{(1,302)} = 0.19$	0.6649
position × separation × form	$F_{(1,302)} = 0.26$	0.6082

5. Discussion and conclusion

As was noted in the introduction, speakers often produce 'errorful' utterances when talking spontaneously, due to e.g. limitations in memory and attention, or to physiological constraints. As a consequence, the resulting speech distinguishes itself considerably from what an 'ideal speaker' in an 'ideal communicative setting' would produce: it typically exhibits flaws in syntactic constructions, articulation errors, hesitations, self-repairs, etc. Nevertheless, numerous studies have indicated that spontaneous speech is very regular, even to the extent that one may distinguish consistencies in the production faults. This is for instance apparent from the fact that the tongue slips into patterns (Nooteboom, 1973), and that mispronunciations are repaired in a rule-governed way (Levelt and Cutler, 1983).

The research reported on in this paper shows that there is also a system in the way speakers produce FPs. From the perspective of discourse structure, it appears that FPs may carry information about larger-scale topical units. Stronger breaks in the discourse are more likely to cooccur with FPs than do weaker ones. The FPs at stronger breaks also tend to be segmentally and prosodically different from the other ones; they more often have silent pauses preceding and following them. This type of FP behaviour is probably not task- or domain-specific, i.e., typical for the sort of descriptive language usage investigated here, since it was partly inspired by observations on other corpora (Chafe, 1980; Swerts and Geluykens, 1994; Swerts and Ostendorf, 1997). It remains to be seen whether there are any cross-linguistic similarities in FP structure.

The results have, first of all, linguistic import. Regarding the speaker, one could interpret the differences in FP characteristics in terms of distinct processing (Shriberg, 1994). It cannot be excluded, however, that the production of different types of FP is to some extent under the speaker's control. Irrespective of the fact whether FPs are undeliberate hesitations or explicitly controlled rethoric devices, it appears that they have cue potential for listeners: if the structure in FP characteristics is picked up by listeners, FPs have a symbolic function in discourse structure. This would mean that FPs are linguistic elements, instead of nonlinguistic events such as grunts, laughter, and coughing. In Dutch, they could be treated as real discourse markers (Schiffrin, 1984), because they function in similar ways as do expressions such as *nou* ('now, well') or *effe kijken* ('let me see, let's see').

The results are also relevant from a technical point of view. The potential use of FPs in automatic speech synthesis is questionable, but one could imagine – at the risk of getting a ridiculous effect – that an adequate modeling of FPs may serve as a cue that the computer is looking for information. For a speech recognition module, on the other hand, FPs are definitely potentially interesting, since they appear to be easily detectable on the basis of only a few prosodic features (duration, pause, F_0) (Shriberg et al., 1997) and since they are very unique speech sounds that are spectrally stable during a relatively long period of time. It has been shown that automatic speech recognizers can exploit FPs to detect self-repairs (Nakatani and Hirschberg, 1994) or to improve statistical language modeling (Shriberg and Stolcke, 1996). From the study presented in this paper, it appears that the occurrence and prosodic structure of FPs, in combination with other prosodic and lexical information, may be useful to automatically delimit coherent passages.

References

Brennan, Susan E. and Maurice Williams, 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: 383–398.

Brubaker, Robert S., 1972. Rate and pause characteristics of oral reading. Journal of Psycholinguistic Research 1: 141–147.

Chafe, Wallace L., 1980. The Pear Stories: Cogniti‹e, cultural and linguistic aspects of narrative production. Norwood, NJ: Ablex.

Clark, Herbert H., 1994. Managing problems in speaking. Speech Communication 15: 243–250.

Cruttenden, Alan, 1986. Intonation. Cambridge: Cambridge University Press.

Fox Tree, Jean E., 1993. Comprehension after speech disfluencies. Ph.D.-dissertation, Stanford University, Stanford, CA.

Fox Tree, Jean E., 1995. The effect of false starts and repetitions on the processing of subsequent words in spontaneous speech. Journal of Memory and Language 34: 709–738.

Fromkin, Victoria A., ed., 1973. Speech errors as linguistic evidence. The Hague: Mouton.

Fromkin, Victoria A., ed., 1980. Errors in linguistic performance. Slips of the tongue, ear, pen and hand. New York: Academic Press.

Goffman, Erving, 1981. Forms of talk. Philadelphia, PA: University of Pennsylvania Press.

Goldman-Eisler, Frieda, 1968. Psycholinguistics: Experiments in spontaneous speech. New York: Academic Press.

Goodwin, Charles, 1981. Conversational organization: Interaction between speakers and hearers. New York: Academic Press.

Guaïtella, Isabelle, 1993. Functional, acoustical and perceptual analysis of vocal hesitations in spontaneous speech. Proceedings ESCA Workshop on Prosody, Lund, September 1993, 128–131.

Hermes, Dik J., 1986. Measurement of pitch by subharmonic summation. Journal of the Acoustical Society of America 83: 257–264.

Kowal, Sabine, Mary R. Bassett and Daniel C. O'Connell, 1985. The spontaneity of media interviews. Journal of Psycholinguistic Research 14: 1–18.

Levelt, Willem J.M. and Anne Cutler, 1983. Prosodic marking in speech repair. Journal of Semantics 2(2): 205–217.

Levinson, Stephen C., 1983. Pragmatics. Cambridge: Cambridge University Press.

Lickley, Robin J. and Ellen G. Bard, 1996. On not recognizing disfluencies in dialogue. Proceedings International Conference on Spoken Language Processing, Philadelphia, PA, October 1996, vol. 3, 1876–1879.

Longacre, Robert E., 1979. The paragraph as a grammatical unit. In: Talmy Givón, ed., Syntax and semantics, vol. 12: Discourse and syntax, 115–134. New York: Academic Press.

Maclay, Howard and Charles E. Osgood, 1959. Hesitation phenomena in spontaneous English speech. Word 15: 19–44.

Nakatani, Christine H. and Julia Hirschberg, 1994. A corpus-based study of repair cues in spontaneous speech. Journal of the Acoustical Society of America 95(3): 1603–1616.

Nooteboom, Sibout G., 1973. The tongue slips into patterns. In: Victoria A. Fromkin (ed.), Speech errors as linguistic evidence, 145–156. New York: Academic Press.

Rotondo, John A., 1984. Clustering analysis of subjective partitions of text. Discourse processes 7: 69–88.

Schiffrin, Deborah, 1987. Discourse markers. Cambridge: Cambridge University Press.

Shriberg, Elizabeth E., 1994. Preliminaries to a theory of speech disfluencies. Ph.D.-dissertation, University of California at Berkeley.

Shriberg, Elizabeth E. and Andreas Stolcke, 1996. Word predicatibility after hesitations: A corpus-based study. Proceedings International Conference on Spoken Language Processing, Philadelphia, PA, October 1996, vol. 3, 1868–1871.

Shriberg, Elizabeth E., Rebecca Bates and Andreas Stolcke, 1997. A prosody-only decision-tree model for disfluency detection. Proceedings Eurospeech '97, Rhodes, September 1997, vol. 5, 2383–2386.

Smith, Vicki L. and Herbert H. Clark, 1993. On the course of answering questions. Journal of Memory and Language 32: 25–38.

Stenström, Anna-Brita, 1994. An introduction to spoken language interaction. London: Longman.

Swerts, Marc, 1997. Prosodic features at discourse boundaries of different strength. Journal of the Acoustical Society of America 101(1): 514–521.

Swerts, Marc and Ronald Geluykens, 1994. Prosody as a marker of information flow in spontaneous speech. Language and Speech 37: 21–43.

Swerts, Marc and Mari Ostendorf, 1997. Prosodic and lexical indications of discourse structure in human-machine interactions. Speech Communication 22(1): 25–41.

Vaissière, Jacqueline, 1995. Phonetic explanations for cross-linguistic prosodic similarities. Phonetica 52: 123–130.

Wichmann, Anne, 1991. Beginnings, middles and ends: Intonation in text and discourse. Ph.D.-dissertation, Lancaster University.

Marc Swerts finished a Ph.D.-thesis in 1994 which dealt with a phonetic analysis of prosodic correlates of discourse structure in spontaneous speech. He has since then worked and published on related research, both fundamental and application-oriented, on the relevance of prosody for automatic speech recognition and for speech synthesis, and on cross-linguistic comparisons of different intonation systems. With Julia Hirschberg, he is currently editing a special issue of Language and Speech on 'Prosody and Conversation'.