Silence and Discourse Context in Read Speech and Dialogues in Swedish

Sofia Gustafson-Čapková & Beáta Megyesi

Computational Linguistics, Department of Linguistics, Stockholm University, Sweden
sofia@ling.su.se

Centre for Speech Technology, KTH, Stockholm, Sweden
bea@speech.kth.se

Abstract

In this study, we investigate the correlation between silent pauses and discourse boundaries in the notion of theme shift. We examine three speaking styles in Swedish: professional and non-professional reading, and elicited spontaneous dialogues. Considerable attention is given to the syntactic and discourse context in which pauses appear, as well as the characteristics of the discourse structure in terms of pauses.

1. Introduction

During the last decade, researchers have shown an increasing interest in the relationship between prosody and discourse structure. Many researchers have investigated this relationship for different languages in order to detect topic structure.

Swerts and Geleukens [18] show that speakers in monologue use pauses of various length to signal information flow in terms of topic structure.

Hirschberg [8] points out that features used for indicating topic structure in texts include speaking rate, duration of interphrase pause, loudness and pitch. She also reports, that phrases introducing a new topic are characterized by an initial wider pitch range preceded by a longer pause, as well as that they are louder and slower than other phrases.

Shriberg et al. [15] have used a prosodic model for automatic topic segmentation, which performs equally well or better than word-based statistical language models. The authors also report that new topics are realized by some combination of silent pauses, low boundary tones and/or pitch range resets.

The relation between prosody and discourse structure is also investigated by van Donzel [19]. She studied prosodic features of discourse boundaries for Dutch on the basis of clause, sentence and paragraph division, as well as the prosodic features of information structure in the New - Given taxonomy [13]. She reports that discourse boundaries in spontaneous speech are realized by silent pauses and boundary tones similarly to Shriberg et al., but with high boundary tones instead; The stronger the boundary, the more probable the combination of the two cues.

Pauses often indicate prosodic phrase boundaries which highlight the organization of the message [1], [2], [6], [8], [11], [18]. Therefore, we have chosen to study pausing in various speaking styles and relate pausing strategies to discourse structure. More specifically, the aim of our study is to investigate the discourse structure in terms of theme shift and its relation to pausing in three different speaking styles in Swedish: professional news announcement, non-professional reading, and elicited spontaneous dialogues. We analyze the materials from two different perspectives. First, we investigate the discourse position of pauses. Second, we study the discourse context itself and the presence of pauses. The results from the former can be useful to predict discourse boundaries given audio data, and results from the latter might be useful for prediction of silent intervals in text-to-speech systems.

In cases where discourse structure and silent intervals do not coincide, other types of linguistic information, such as part-of-speech and phrasal structure, might help in the prediction of discourse boundaries, and/or in the prediction of pauses in a text-to-speech system.

In the next section, we will give a summary of our data and methodology, as well as a brief overview of the findings reported in our previous studies on the production and perception of pauses ([4], [9]). In section 3, the results on the correlation between discourse structure and pausing are presented. Lastly, in section 4, we conclude the results and suggest directions to future research.

2. Acoustic Pauses and Discourse Contexts

In this study, we use the same speech data for each speaking style as we used in our previously reported studies (see [4] and [9]). The material of read speech consists of recordings of Swedish radio news [14] read by four professional and four non-professional readers. The spontaneous speech data [5] consists of recordings of two Swedish map task dialogues, each with two dialogue participants. The data sets consist of 920 words each.

In order to investigate the duration, frequency, type and position of *acoustic pauses*, the speech data was processed automatically by a pause detector. Silent intervals longer than or equal to 100 ms were defined as acoustic correlate for pausing. Pauses may include natural physical phenomena such as breathing and swallowing. However, particles expressing feedback/back-channelling (e.g. mmm, aaa, aha) in dialogues are not allowed inside pauses. The automatic detection was manually checked in order to obtain consistency.

As mentioned in the first section, various studies have shown that prosody might signal discourse structure in terms of topic structure ([8], [18]). In these studies topic units can be seen as discourse segments. However, the nature of discourse segments is hard to define. In the literature a variety of features is used when describing discourse segment boundaries, such as cue words (e.g. [10]), referring expressions and intonation (e.g. [12]), among others.

In this study, for a definition of a discourse segment we use the notion of theme. Theme is defined as a chunk with one underlying intention. In other words, a discourse segment is a sequence of utterances aiming to communicate the same intention and thereby the utterances belong to the same theme. *Theme shift* (TS) is the position in the discourse where a new theme is introduced, i.e. it marks a discourse boundary.

As the basis for our investigation of the discourse context regarding theme shift, we asked five subjects to annotate the transliterated text materials for theme shift. The subjects were instructed to annotate the text material in a hierarchical fashion, which means that themes may contain sub-themes. In order to keep discourse and prosody apart, the annotators were not allowed to listen to the audio data. All five subjects except one were naive regarding discourse annotation. Additionally, in order to examine the linguistic characteristics of the text materials, the words in each text material were automatically tagged with their part-of-speech tag including morphological information. Then, each sentence[1] in the texts was automatically parsed on the basis of its phrasal constituent structure. The labels for the constituents include major phrase categories, e.g. adverb (AdvP), adjective (AP), noun (NP) and prepositional (PP) phrase, as well as maximal projections involving several major phrases, e.g. coordinated noun phrases, or a noun phrase with a prepositional phrase attached to it.

Before we go into details on the results concerning the correlation between theme shift and pausing, we will briefly summarize the, for this study, most relevant and important features of pauses in the three speaking styles. These features include the mean duration and frequency of silent intervals longer than or equal to 100 ms. The overview is shown in Table 1 below. For a detailed description see [4] and [9].

Table 1: Features of acoustic pauses

Speaking style	Pause Duration	Word/Pause Ratio
Professional	271 ms	77 (920/12)
Non-professional	561 ms	8.4 (920/110)
Dialogue	538 ms	5.5 (920/167)

Although there are differences in the duration and frequency of pauses between the styles, the total duration is approximately the same for the two reading styles. Hence, the time it takes to pronounce a word in average differs between the speaking styles suggesting greater variation in speech tempo. Our results indicating that pausing patterns vary across speaking styles are consistent with the results reported in [3], [7], [16] and [17]. Next, we will present the results on the correlation between the discourse structure and pausing.

3. Results

To give an overall picture of the correlation between pauses and theme shift in the three speaking styles, recall and precision rates were counted. Recall, given in Definition 1, describes the percentage of pauses that appear in theme shift position according to the majority of the annotators (three or more). Precision, on the other hand, gives the percentage of theme shift that corresponds to pauses in the speaking styles, see Definition 2. 100% recall means that all pauses appear at theme shift, 100% precision means that all theme shifts are realized as pauses, and 100% recall and precision would mean a one-to-one mapping between pauses and theme shift.

$$Recall = \frac{No.\,of\,pauses\,appearing\,at\,theme\,shift}{Total\,no.\,of\,pauses} \quad (1)$$

$$Precision = \frac{No.\,of\,pauses\,appearing\,at\,theme\,shift}{Total\,no.\,of\,theme\,shifts} \quad (2)$$

The correlation between discourse boundaries and pauses for each speaking style is shown in Figure 1. It is clear that theme shift and pauses do not always coincide, and we find clear differences between the speaking styles.

In the reading styles pauses often appear at discourse boundaries; In fact, in professional reading every silent interval is located at theme shift (100% recall). In the dialogues, on the other hand, the majority of pauses can be found in weak or non-existing discourse boundary positions (34% recall). The reader should note that the the amount of pauses are considerably fewer in professional reading than the other speaking styles.

Looking at the discourse boundaries and their acoustic correlate in terms of silent intervals (i.e. the precision rates) we find that non-professional readers and dialogue participants mark the majority of theme shift boundaries with a pause, although the non-professional readers to a greater extent (92% resp. 57% precision). Professional readers on the other hand pause only in few cases at theme shift positions (9%).

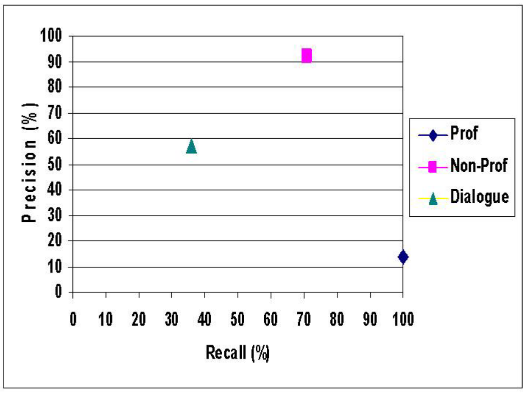

Figure 1: Recall and precision rates for pauses and theme shift in professional and non-professional reading, and in dialogues.

In the subsequent sections, we will describe these results in detail. Above all, we will focus on the linguistic context in which discourse boundaries and pauses appear. The reader should keep in mind that the discourse boundaries can be treated as more or less strong since the annotators do not necessarily agree on the boundary. Therefore, where appropriate, we will refer to the continuum of discourse boundary in the notion of theme shift (TS), and the opposite term theme continuation (TC), as follows:

- No boundary – none of the five subjects labeled a TS, i.e. theme continuation (TC)
- Weak boundary – one or two subjects annotated a TS
- Strong boundary – three or four subjects labeled a TS
- Extra strong boundary – all five subjects agreed on a TS

[1] In the case of dialogues, a sentence is defined as a turn.

3.1. Where can we find pauses?

In Figure 2, the relative frequencies of the position of pauses in thematic context according to the number of discourse annotators are given for each speaking style.

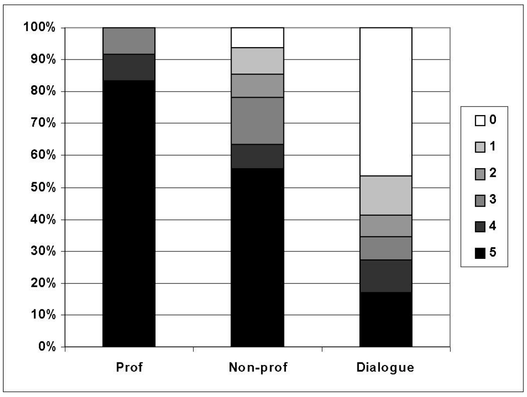

Figure 2: Recall: The distribution of pauses over various types of discourse boundaries according to the number of annotators (0-5) in professional and non-professional reading, and in dialogues.

In *professional reading*, all pauses appear at strong boundaries. In all but one case, the pause can be found at paragraph and sentence boundaries. The exception is one case, where the pause appears between two noun phrases in a list.

In *non-professional reading*, also the majority of pauses can be found at extra strong boundary positions, of which 84% appear at sentence boundaries and 16% at clause boundaries in front of conjunctions. Pauses in strong boundary positions, on the other hand, are located mainly at clause boundaries (41%), but also between noun or prepositional phrases in lists (36%), at sentences boundaries (18%), or before a finite verb when topicalizing PPs (5%). However, the readers pause even in weak boundary positions, mainly between NPs (58%) and clause boundaries (42%). It is also worth noticing that pauses occur even at theme continuation i.e. where no discourse boundary was labeled. These pauses are found between phrases that are sisters in the syntactic tree (such as between NPs, or between AdvPs and PPs), or after topicalizing either NPs or PPs in front of a finite verb.

In *dialogues*, on the other hand, the majority of pauses are found in weak boundary positions, or at theme continuation. The pauses found in extra strong boundary positions are situated between turns (41.4%), clauses (44.8%) of which 70% are located in front of conjunctions, and between phrases (13.8%). Pauses situated at strong boundaries are found between turns (43.3%), clauses (46.7%), and between NPs (10%). As mentioned, pauses are also found in weak boundary positions. Here, we find them between turns (48%), clauses (25.8%), between NPs (6.5%), and between different kinds of nested phrases (19.4%). The pauses appearing at theme continuation are found between turns (13%), clauses (35%), NPs (8%), and in front of PPs (4%). It is also worth noting that pauses at weaker boundaries and theme continuation are to a larger extent found in nested phrases, than are stronger boundaries. Even though the pauses between turns, clauses and phrases are similarly distributed in the different degrees of discourse boundary strength, it should be pointed out, that the clause initial PoS context shows clear differences. Just extra strong boundaries have a preference for conjunctions, while clauses in weaker boundary positions often begins with adjectives, nouns or prepositions. At theme continuation, no tendencies were found.

3.2. How frequently does theme shift correlate to pauses?

In the last section, we described the position of pauses with regard to their linguistic context. The main question we address here is whether we can predict the pauses (e.g. in TTS systems) given information on discourse boundaries, and if not, whether we can get help from the PoS and phrasal structure to predict silent intervals.

In Figure 3, the percentage of each type of discourse boundaries having a pause correlate in the audio data is given for the various speaking styles. The reader easily can see from the figure that there are large differences between the speaking styles, and especially between the readings, as was described in Section 3.

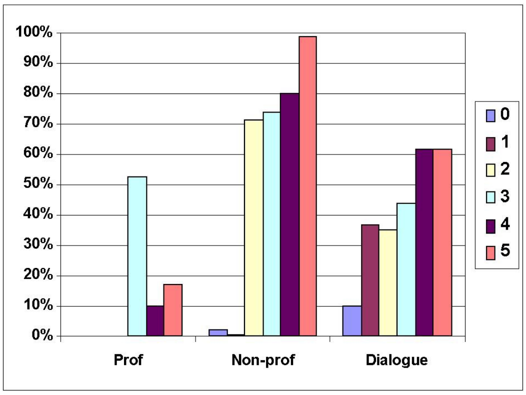

Figure 3: Precision: The percentage of discourse boundaries according to the number of annotators (0-5) that is marked as a pause by the speakers in professional and non-professional reading, and in dialogues.

In *professional reading*, extra strong discourse boundaries alone do not imply pauses since only 17% of the extra strong boundaries have a pause correlate in our data. However, in the case of the location of strong boundaries, the news announcers use pausing as markers for theme shift. In weak or no boundary positions, on the other hand, pauses are not present. We have seen in the previous section that strong boundaries with pause correlates occur at paragraph and sentence endings. To be able to predict pauses on the basis of syntax and discourse boundaries, it is also necessary to examine the syntactic context of strong boundaries which do not have a pause correlate in the audio data. These boundaries are located between phrases or clauses, as well as at sentence boundaries. This means that information about the position of strong discourse boundaries in the text is helpful in the prediction of pauses but it is not enough. We need additional syntactic information to be able to reduce the amount of pauses which would be predicted by the discourse boundaries. Unfortunately, syntactic information only partly helps to limit the amount of pauses by not allowing pauses inside sentences. We also would need to limit pauses between sentence boundaries.

In *non-professional reading*, the stronger the boundary the more certain that it has a pause correlate in the audio data; All extra strong boundaries and a high percentage of strong boundaries are marked as pauses, and only a few weak boundaries have a pause correlate. The examination of those positions in which strong boundaries appear and in which pauses are absent are interesting. These positions can be found between phrases: between clauses (37%), between coordinated NPs and VPs (37%), and in front of finite verbs (25%).

In *dialogues*, the tendency is the same as in non-professional reading; the stronger the discourse boundary, the more probable that it co-occurs with a pause. However, in dialogue it is more frequent with cases where a stronger boundary does not have a correlate in pausing. The positions where an extra strong boundary is not co-occurring with a pause are mainly in front of conjunctions (64%) and by turn boundaries (11%). The positions where a strong boundary does not have a correlate in pausing are in front of conjunctions (28,6%), in front of adverb phrases (28,6%), by turn taking (14,3%) and in front of prepositional phrases (14,3%).

4. Conclusion and Future Research

In this study, we examined the correlation between pauses and discourse boundaries in terms of theme shift in three speaking styles in Swedish: spontaneous elicited dialogues, non-professional news reading, and professional news announcement.

The results, reported in this paper, show clear differences between the speaking styles. The non-professional reading shows the largest overlap between the theme shift markings and pauses, whereas the professional reading shows the same coverage just in one direction, i.e. all pauses occur at theme shifts, but not the other way around. In dialogue, we also noted a correlation between pauses and the discourse structure but to a considerably lesser extent than in non-professional reading.

Our results might be useful in text-to-speech systems for the prediction of pauses on the basis of parsed text, annotated with discourse boundaries, as well as for automatic detection of discourse boundaries given the parsed and transliterated audio data with pauses detected.

Questions we find important to explore in future concern intonational variation in connection to pausing and discourse structure. Analysis of intonational patterns would shed more light on the importance of the intonational variations and their effect on prosodic phrasing. We also would like to study the hierarchical representation of the discourse and relate it to prosodic phrasing.

5. Acknowledgments

First of all, we would like to thank Petur Helgason for the dialogue corpus, Swedish Radio for the Swedish news recordings, the four non-professional readers, Mattias Heldner for his help with the recordings, and the discourse annotators. Last, but not least, many thanks to Rolf Carlson, Björn Granström, David House, and Sara Rydin for their valuable comments.

6. References

[1] Bruce, G., 1995. Modelling Swedish Intonation for Read and Spontaneous Speech. *Proceedings of International Congress on Phonetic Sciences*. Vol. 2, 28-35.

[2] Deese, J., 1980. Pauses, prosody and the demands of production in language. In *Temporal Variables in Speech, Studies in Honour of Frieda Goldman-Eisler*, Hans and Raupach, Manfred (eds.). Mouton Publishers.

[3] Fant, G.; Kruckenberg, A., 1989. Preliminaries to the Study of Swedish Prose Reading and Reading Style. In *STL-QPSR 2/1989*. Speech Transmission Laboratory (Dept. of Speech, Music and Hearing), KTH, Sweden.

[4] Gustafson-Čapková, S.; Megyesi, B., 2001. A Comparative Study of Pauses in Dialogues and Read Speech. In *Proceedings of Eurospeech 2001*. Vol. 2, 931-935, Aalborg, Denmark.

[5] Helgason, P., forthcoming. *Stockholm Corpus of Spontaneous Speech*. Dept. of Linguistics, Stockholm University.

[6] Hirschberg, J., 1995. Prosodic and other acoustic cues to speaking style in spontaneous and read speech. *Proceedings of International Congress on Phonetic Sciences*. Vol. 2, 36-43.

[7] Hirschberg, J., 1997. Prosodic variation and discourse structure across speaking styles. *Prosody: Theory and Experiment, Studies presented to Gösta Bruce*. Kluwer Academic Publisher.

[8] Hirschberg, J., 2001. Communication and Prosody: Functional Aspects of Prosody. In *Speech Communication: Special Issue on Dialogue and Prosody*, Terken, J.; Swerts, M. (eds.).

[9] Megyesi, B.; Gustafson-Čapková, S., 2001. Pausing in Dialogues and Read Speech: Speaker's Production and Listeners Interpretation. *Proceedings of the Workshop on Prosody in Speech Recognition and Understanding*, 107-113, October 22-24, 2001, New Jersey, USA.

[10] Moore, J. D.; Moser, M., 1995. Investigating Cue Placement and Selection in Tutorial Discourse. *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 130-135.

[11] Ostendorf, M., 1997. Prosodic Boundary Detection. *Prosody: Theory and Experiment, Studies presented to Gösta Bruce*, Kluwer Academic Publisher.

[12] Passonneau, J. R.; Litman, J. D., 1997. Discourse Segmentation by Human and Automated Means. *Computational Linguistics 23:1*, ACL.

[13] Prince, E., 1981. Toward a Taxonomy of Given-New Information. In *Radical Pragmatics*, Cole, P. (ed.). Academic Press, New York, 223-255.

[14] 1999-2000. Recordings of Swedish Radio News, Swedish Radio.

[15] Shriberg, E.; Stolcke, A.; Hakkani-Tür, D.; Tür, G., 2000. Prosody-Based Automatic Segmentation of Speech into Sentences and Topics. *Speech Communication 32*, 127-154.

[16] Strangert, E., 1993. Speaking style and pausing. *PHONUM*. Reports from the Department of Phonetics, University of Umeå.

[17] Strangert, E., 1993. Clause Structure and Prosodic Segmentation. In *FONETIK-93 Papers from the 7th Swedish Phonetics Conference*, John Sören Petterson (ed.). Uppsala, May 12-14, 1993.

[18] Swerts, M.; Geluykens, R., 1994. Prosody as a marker of information flow in spoken discourse. *Language and Speech 37*, 21-45.

[19] van Donzel, M., 1999. *Prosodic Aspects of Information Structure in Discourse*. Ph.D Thesis, Netherlands Graduate School of Linguistics, Holland Academic Graphics.