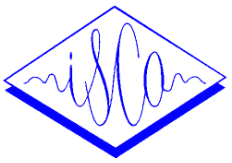

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

A SIMPLE PROCEDURE TO CLARIFY THE RELATION BETWEEN TEXT AND PROSODY

Hanny den Ouden[1,2], Carel van Wijk[1] & Marc Swerts[2,3]

[1] CLS Discourse Studies Group, University of Tilburg
[2] IPO Center for User-System Interaction, Technical University Eindhoven
[3] CNTS Center for Dutch Language and Speech, Antwerp University
J.N.denOuden@kub.nl chvwijk@kub.nl m.g.j.swerts@tue.nl

ABSTRACT

Four texts originally broadcasted on radio, were presented to 52 participants in a paper-and-pencil task: they were instructed to assign a linear structure by placing slashes between segments, and a hierarchical structure by underscoring and crossing out segments. Strength of paragraph boundary had a regular relation with pause duration and F_0 maximum. Hierarchical position tended to be marked in two ways: centrality by pause duration and redundancy by F_0 maximum. Speech rate was independent of text structure.

1 TWO ASPECTS OF TEXT STRUCTURE

Prosodic parameters are dependent on characteristics of the utterances produced by a speaker. Traditionally, most prosodic studies have paid attention to syntactic and informational aspects of individual utterances (see e.g. van Wijk & Kempen, 1984; van Wijk, 1987). But prosodic correlates of the ways utterances combine to form a text, are receiving attention more and more.

So far, this research on textual prosody has mainly concentrated on the *linear* organization of texts. In this type of studies, a text is defined as a concatena-tion of elementary building blocks, the paragraphs, and the focus is on the prosodic marking of their beginning and end. This linear aspect of text structure accounts for the >chunking= function of prosody. But text organization has a *hierarchical* aspect as well. Paragraphs differ in importance: they are not equally essential to the message to be conveyed. And the same goes for the utterances within each paragraph; some of them are crucial, while others could be easily missed. This hierarchical aspect of text structure is expected to have a >levels= effect on prosody.

A lot of evidence has been reported for the chunking function of prosody (see e.g. Sluijter & Terken, 1993; Swerts, 1997). Paragraphs are usually preceded by a relatively long pause and start with a higher pitch range. They also tend to close with a lower pitch range and a higher speaking rate (Hirschberg & Grosz, 1992). For the levels effect, however, evidence is mostly suggestive. It is based, for instance, on having subjects read aloud texts of a specific genre and with carefully designed structures such as narratives based on story grammar (Noordman, Dassen, Swerts & Terken, 1999). The question is whether the same result will be found in situations in which the speaker is free to determine both the genre and the organization of the text.

2 OBJECTIVE OF THIS STUDY

To explore the prosodic correlates of text structure, we should also look at speech obtained in non-laboratory settings. This requires a flexible instrument to analyze texts that differ widely in terms of genre, content and, last but not least, quality. Such an instrument has been developed within the framework of *Rhetorical Structure Theory* (Mann & Thompson, 1988). The application of this technique demands a considerable proficiency and tends to be time-consuming. This has its drawback in terms of reliability and costs (den Ouden, van Wijk, Terken & Noordman, 1999; Sanders & van Wijk, 1996). Therefore we decided to supplement this theoretically based instrument with a practical procedure that would deliver a linear and hierarchical organization of randomly selected texts in a simple way, that is, fast and cheap. In

the following sections we introduce this procedure and test its relevance for prosodic theory on a set of Dutch radio talks.

3 Analytic procedure

The procedure is a simple paper-and-pencil task. A sufficiently detailed result requires several dozens of participants. No specific linguistic expertise is needed. The task can be administered in groups. A one-page text takes about 10 minutes. Each participant gets a paper print of the text and does not hear its spoken version. The text is split into segments (basically clause-type units; for segmentation rules see Sanders & van Wijk, 1996, pp. 126-127). Each segment starts on a new line. Participants are instructed as follows:

- place a slash before the segments that start a new paragraph;
- underscore the segments you consider crucial to the intended message (up to one-third of the total);
- cross out the segments you would delete first in an excerpt (up to one-third of the total).

Thus a participant indicates for each segment whether it starts a paragraph, and whether it is central or redundant. By combining the responses of the participants a segment gets three scores: one for strength of paragraph boundary and two for informational importance: degree of centrality and degree of redundancy, that is, how high respectively how low is the segment located in the hierarchy. [Note that the scores for hierarchical position can not be equated with each other because there is a third way of responding: a segment may be neither underscored nor crossed out.]

4 Material and participants

The task has been tested on four texts: two (descriptive) news reports and two (argumentative) news commentaries. They ranged in length from 25 to 37 segments with a total of 125 segments. The texts were tape recorded from radio broadcastings. The printed versions contained no typographic markers (punctuation, capitals).

A speech processing program was used to determine for each segment the preceding pause (in msec), the F_0 maximum (in Hertz), and the articulation rate (phonemes per second). Since each text had a different speaker (three males, one female), the parameters showed large differences between speakers. This type of variation was removed by transforming the scores of each speaker into standard scores before combining them. In the tables, numbers must be read as follows: a zero-score (0.00) stands for >average=, a negative score for >below average=, and a positive score for >above average=.

A total of 52 advanced university students participated. The scores that each segment received on the basis of their responses, were reduced to four-level classifications ranging from *absent* to *strong* for boundary strength, and from *absent* to *high* for centrality and redundancy. The criterion for these classifications was based on the frequency distributions of the scores. Boundaries were chosen in such a way that each category contained a sufficient number of cases. All segments were classified with respect to the three text features on the basis of the percentage of participants who had scored it for that specific feature, in the following way:

absent:	5 percent or less
low/weak:	6 to 16 percent
moderate:	17 to 50 percent
high/strong:	51 percent or more.

5 Results: chunking function

The results in Table 1 show that the strength of a paragraph boundary had a reliable and consistent effect on pause duration and F_0 maximum, and no effect on speech rate. When a segment has been considered to follow a paragraph boundary more often, its preceding pause turned out to be longer and its F_0 maximum higher.

Table 1 Prosodic features of a segment as a function of its paragraph boundary strength (in standard scores)

	absent (n=73)	weak (n=28)	moderate (n=8)	strong (n=12)	Results of ANOVA
pause duration	-0.43	0.35	0.86	1.25	$F_{(3,117)}$ =21.64, p<.001, η5=.36
F_0 maximum	-0.25	0.25	0.61	0.41	$F_{(3,117)}$ =4.11, p<.01, η5=.10
speech rate	-0.01	0.03	-0.09	0.12	F<1

6 RESULTS: LEVELS EFFECT

Before we look at these results, we need to be sure that only hierarchical position can be held responsible for them. An inspection of the classifications shows that this is not the case. There were 20 paragraph boundaries called moderate/strong; 16 of them are followed by a segment high in centrality and absent/low in redundancy. The texts under discussion here tended to start their paragraphs with a segment containing crucial information. This connection between linear and hierarchical position inflates the levels effect. Therefore, segments following a paragraph boundary are kept apart in the tables. They were also excluded from the statistical analyses. The ANOVAs evaluate the levels effect in its pure form.

The results in Table 2 present for degree of centrality a reliable and consistent effect on pause duration, a less regular effect on F_0 maximum, and no effect on speech rate. When a segment has been considered central to the message more often, its preceding pause appeared to be longer and its F_0 maximum tended to become higher.

The results in Table 3 show for degree of redundancy a reliable but irregular effect on F_0 maximum, an unreliable but regular effect on pause duration, and no effect on speech rate. When a segment has been considered redundant more often, its preceding pause was shorter and its F_0 maximum lower.

Table 2 Prosodic features of a segment as a function of its degree of centrality (in standard scores)

	Absent (n=30)	Low (n=26)	Moderate (n=30)	High not following paragraph boundary (n=20)	High following paragraph boundary (n=16)	Results of ANOVA
pause duration	-0.42	-0.35	-0.14	0.27	1.07	$F_{(3,98)}$ =2.89, p<.05, η5=.08
F_0 maximum	-0.45	0.10	-0.10	0.22	0.45	$F_{(3,102)}$ =2.51, p=.06, η5=.07
speech rate	-0.04	0.15	-0.14	0.07	-0.02	F<1

Table 3 Prosodic features of a segment as a function of its degree of redundancy (in standard scores)

	Absent/Low	Absent	Low	Moderate	High	Results of ANOVA
	following paragraph boundary (n=16)	not following paragraph boundary (n=10)	(n=20)	(n=46)	(n=29)	
pause duration	1.02	0.40	-0.21	-0.27	-0.35	$F_{(3,97)}$ =1.97, p=.12, η5=.06
F_0 maximum	0.57	0.57	-0.20	0.05	-0.48	$F_{(3,101)}$ =3.73, p<.025, η5=.10
speech rate	0.13	0.15	-0.13	0.08	-0.14	F<1

7 CONCLUSIONS

Three general remarks can be added to those made in discussing the tables. First, paragraph boundaries are just as hierarchical positions relative features, not absolute ones. Not surprisingly, readers tend to agree strongly on the absence of boundaries and the redundancy of segments. But conversely, they decide far less unanimously on the presence of boundaries and the centrality of segments. Second, the effects of paragraphing and hierarchical position should be carefully kept apart. When it is not separated from the chunking function, the levels effect becomes inflated (as would have been the case here) or suppressed (in texts where paragraphs tend to start with an introductory remark, that is, a >redundant= segment). And third, a simple procedure of the type presented here offers a useful tool for prosodic research and a relevant supplement to the more sophisticated, text-analytic models.

REFERENCES

Hirschberg, J., & Grosz, B. (1992). Intonational features of local and global discourse structure. In *Proceedings of the speech and natural language workshop* (pp. 441-446). Harriman NY: DARPA.

Mann, W., & Thompson, S. (1988). Rhetorical Structure Theory; Toward a functional theory of text organization. *Text, 8,* 243-281.

Noordman, L., Dassen, I., Swerts, M., & Terken, J. (1999). Prosodic markers of text structure. In K. van Hoek, A. Kibrik & L. Noordman (Eds.), *Discourse studies in cognitive linguistics* (pp. 133-148). Amsterdam: Benjamins.

Ouden, J. den, Wijk, C. van, Terken, J., & Noordman, L. (1999). Reliability of discourse structure annotation. *IPO Annual Progress Report*, *33*, 129-138.

Sanders, T., & Wijk, C. van (1996). PISA - A procedure for analyzing the structure of explanatory texts. *Text, 16,* 91-132.

Sluijter, A., & Terken, J. (1993). Beyond sentence prosody: paragraph intonation in Dutch. *Phonetica, 50,* 180-188.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, *101*, 514-521.

Wijk, C. van, & Kempen, G. (1984). From sentence structure to intonation contour; an algorithm for computing pitch contours on the basis of sentence accents and syntactic structure. In B. Müller (Ed.), *Sprachsynthese* (pp. 157-182). Hildesheim: Olms Verlag.

Wijk, C. van (1987). The PSY behind PHI: a psycholinguistic model for performance structures. *Journal of Psycholinguistic Research, 16,* 185-199.