The Prosodic Realization of Organizational Features of Texts

Hanny den Ouden[1,2] , Leo Noordman[1] & Jacques Terken[2]

[1] Tilburg University, [2]Eindhoven University of Technology
The Netherlands
j.n.d.ouden@tue.nl

Abstract

This study investigates the prosodic realization of organizational features of texts. Twenty read aloud news reports were annotated according to Rhetorical Structure Theory (RST). This theory defines the clustering of elementary units (clauses) into larger segments (hierarchical organization), the relative importance of units (nuclearity) and the rhetorical relations between segments. The prosodic features we considered were pause durations between segments, pitch range and articulation rate of the segments. It was found that pause duration and pitch range reflect the hierarchical organization of a text: the lower a text segment is embedded within the hierarchy of a text, the shorter the pauses and the lower the pitch range. Also, a nuclear segment is read slower than a non-nuclear segment. Finally, rhetorical relations affect pause duration. For example, causal relations are associated with shorter pauses than non-causal relations. We conclude that the organizational features of texts as provided by RST are reflected by prosodic characteristics.

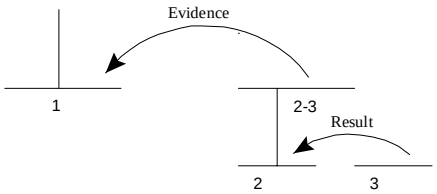
1. Introduction

There is ample evidence that the prosodic features of a spoken text indicate aspects of text organization. Studies such as [1][2][3][4][5][6] have shown that paragraph initial utterances are marked prosodically by long pause durations and high pitch levels, whereas paragraph-final utterances and parenthetical utterances are characterized by relatively high speaking rate and low pitch. These studies imply a model in which the phonological realization component receives information about the location of paragraph boundaries. Theories in the field of discourse linguistics suggest that texts have a more refined organization than simply in terms of the distinction between paragraphs and sentences. For instance, Grosz & Sidner analyze texts in terms of embedded discourse segments, and in fact it has been shown that utterances are associated with different prosodic characteristics depending on their status with respect to Discourse Segment structure, e.g. whether they are Discourse Segment Initial, Discourse Segment Medial or Discourse Segment Final [7][8]. However, the level of embeddedness of discourse segments in the hierarchical representation of the text has not been taken into consideration in these studies. So it remains to be determined whether the prosodic features associated with discourse segments are influenced by the position of the segments in the hierarchical representation of the text.

Several theories of text organization may be used as a starting point for the annotation of text organization, for example Rhetorical Structure Theory [9], Story Grammar [10], Grosz & Sidner [11] and PISA [12], In this study we choose Rhetorical Structure Theory (henceforth RST), for two reasons. First, the reliability of applying RST as a text annotation schema is quite good [13][14]. Second, RST not only accounts for the hierarchical organization of texts, but also for other aspects of textual organization, like the relative importance of segments (nuclearity) and the nature of the rhetorical relations that hold between segments. Our main question is whether hierarchical position of segments, nuclearity and rhetorical relations are reflected in the prosodic features of the text.

1.1. Annotating text organization with Rhetorical Structure Theory

The basic segments that RST works on are syntactic clauses containing a finite verb. Clausal subjects and complements and restrictive relative clauses are considered as parts of their host clauses rather than as separate clauses. The RST analyst determines the rhetorical relations between parts of the text. The process starts with identifying the rhetorical relation that characterizes the text as a whole, i.e. identifying the part of the text that expresses the core and the supporting part, and identifying the relation that holds between the two parts. For each text part this procedure is repeated, until finally all relations between all segments are identified. This process implicates that segments are subdivided into smaller segments and this yields a hierarchical structure of embedded segments. RST defines about 25 rhetorical relations, such as Evidence, Background, Solution, Cause. In each relation at least two segments are involved: two or more nuclei or one nucleus and one satellite. Nuclei are the core of the relations; they are more important for the coherence of the text than satellites. We illustrate the basic concepts of the theory by means of Figure 1.



1. The exam was too difficult.
2. Almost everyone made more than ten mistakes.
3. These persons have to take a re-examination.

Figure 1: Example of an RST analysis

The upper level of the hierarchy consists of text part 1-3. The text as a whole is characterized by an Evidence relation between segment 1 and segments 2-3. Segment 1 is the nucleus (represented as a vertical line), because it is the core of the text; text part 2-3 is the satellite. At the second level of the hierarchy the relation between segments 2 and 3 is characterized by Result. Obviously, the hierarchical structures of longer texts have more levels than the hierarchies of short texts.

2. Methodology

2.1. Text material

Twenty news reports were selected from a Dutch national news paper. The reports had a length of about 30 segments; they did not contain direct speech; their contents were objective and non-controversial.

2.2. Scoring of textual parameters

The texts were segmented into basic segments (clauses) according to the criteria given by RST. The twenty texts were analyzed in terms of RST by the first author. In the resulting hierarchical structures the three organizational features to be studied were scored as explained below.

In order to study the effect of hierarchical position, we assigned a number to each boundary between two basic segments, according to the following procedure: For two adjacent basic segments, determine the superordinate node connecting the two segments; then count the number of subordinated nodes dominating the segments, add up all nodes and assign that number to the boundary between the segments. The approach is illustrated by means of Figure 2, which contains a representation that is equivalent to the output of RST, but it has the basic segments all at the bottom level.

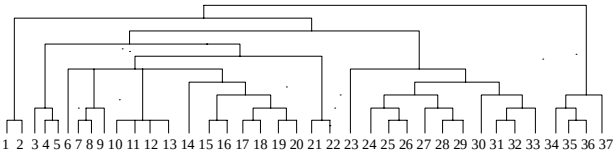


Figure 2: Bottom-up representation of a text's hierarchical organization

Segments 1 and 2 are connected by only one node. Therefore, this boundary is scored as 1 (the lowest score possible for boundaries in the hierarchy). Segments 33 and 34 are connected by nine nodes: six to the left side, two to the right side and one superordinate node. Therefore, this boundary is scored as 9. In the twenty texts there were 543 boundaries. They were scored with a range from 1 to 10. In the statistical analyses these scorings were reduced to a five-level classification, because there were few boundaries scored as 4 or 5, and even less boundaries scored as 6 or higher. Therefore, scores 4 and 5 are taken together and scores 6 to 10 are taken together. This resulted in 210 boundaries scored as 1; 134 boundaries scored as 2; 76 boundaries scored as 3; 77 boundaries scored as 4; and 46 boundaries scored as 5.

Each basic segment was classified as either nucleus or satellite. The distinction between nucleus and satellite is derived directly from the original RST analyses, because each segment is either a nucleus or a satellite, as shown in Figure 1. Segments 1 and 2 are both nuclei (expressed graphically by the vertical lines); segment 3 is a satellite (expressed graphically by the arc). In the twenty texts there were 383 nuclei and 180 satellites.

Discourse relations between segments were derived directly from the RST analyses, as illustrated in Figure 1. The relation between segments 1 and 2 is characterized by Evidence; the relation between segments 2 and 3 is characterized by Result. In our text material 22 relation types were assigned, 13 of which were assigned more than ten times. Only these 13 relations were involved in the statistical analyses. These discourse relations were: Elaboration (n=119), Background (n=48), Circumstance (n=21), Cause (n=39), Result (n=36), Contrast (n=22), Antithesis (n=15), Concession (n=26), Evaluation (n=14), Interpretation (n=20), Restatement (n=17), Joint (n=108), Sequence (n=25). Cause, Result and Concession are causal relations.

The segmentation criterion applied by RST does not take into account the clausal status of segments in terms of main or subordinate clause and position in the sentence. For prosody however, we know that clausal status is relevant. The second clause in the example *'John is hungry. He can't stop eating'* is an independent sentence, while the second clause in *'John is hungry, but he has no time to eat'* is the second member of a pair of two coordinate main clauses, and the second clause of *'John is going to eat early, because he is hungry'* is a subordinate clause. On the basis of the syntactic characteristics, we may expect differences in the prosodic realizations of these three types of clauses. For that reason, we included clausal status as an additional factor in this study. There were 467 main sentences, called 'simple segments', 47 segments which were the second part of a complex sentence consisting of two coordinate main clauses connected by 'but', 'since' or 'and', called 'complex coordinate segments', and 47 segments which were the subordinate clauses of complex sentences consisting of a main clause and a subordinate clause, called 'complex subordinate segments'.

2.3. Speech materials

The twenty written news reports were presented to twenty native speakers of Dutch, ten males and ten females, all highly educated people. Each speaker read one text. The texts were presented without paragraph markers. The speakers were asked to prepare the reading session carefully. They were instructed to imagine that blind people were their listeners and that these people should understand the content of the text fully. The speakers were encouraged to make notes in the text to facilitate the reading aloud.

The recordings were made in a sound-treated room. The speech was digitized with the speech processing program Gipos (http://www.ipo.tue.nl/ipo/gipos).

2.4. Scoring of speech parameters

Prosodic features relevant for the marking of text structural notions are durations of pauses, pitch range and articulation rate. Pause durations of the boundaries between segments in milliseconds were measured by hand. Pitch range, operationalized as the F0 maximum, was measured per segment automatically in Hertz [15]. To avoid errors of the pitch measurements each individual segment was inspected with LPC analysis beforehand. Pitch measurement errors were mostly voiced-unvoiced errors; they were corrected in the speech signal by hand. F0-maxima associated with final rises were removed before applying the automatic procedure. Articulation rate was defined as the number of phonemes per second. The number of phonemes in a segment was calculated automatically (SampaCount) on the hand-corrected canonical transcription.

Since each text was produced by a different speaker, prosodic parameters showed large differences between speakers. This variation was removed by transforming the prosodic measurements into standard scores per speaker, the mean score being zero.

3. Results

3.1. Relation with hierarchy

To study the effect of hierarchical position with five levels on the three prosodic parameters, pause duration, pitch range and articulation rate, a MANOVA was conducted. Also the effect of clausal status was studied. Effects were significant for hierarchical position ($F(12,1587) = 2.88$, $p<.001$, $\eta^2 = .02$) and clausal status ($F(6,1056) = 4.31$, $p<.001$, $\eta^2 = .03$). There was no interaction between the two factors ($F(18,1587) = 1.05$, $p = .40$). Univariate analyses showed that both factors had an effect on pause duration (Hierarchical position: $F(4, 529) = 5.35$, $p<.001$, $\eta^2 = .04$; Clausal status: $F(2, 529) = 10.17$, $p<.001$, $\eta^2 = .04$), and on pitch range (Hierarchical position: $F(4, 529) = 4.54$, $p<.001$, $\eta^2 = .03$; Clausal status: $F(2, 529) = 4.09$, $p<.025$, $\eta^2 = .02$). There was no effect on articulation rate (Hierarchical position: $F<1$; Clausal status: $F<1$). These results show that height in the hierarchies affects the duration of the pauses and the height of the pitch range. This is also the case for the clausal status of the segment. The means of the standard scores of the prosodic parameters in relation to hierarchical position and clausal status are presented in Table 2. Both pause duration and pitch range show a highly consistent pattern: the higher the boundary is embedded within the hierarchy, the longer the pause of that boundary and the higher the pitch range of the segment following that boundary. Also the pause duration is longer and the pitch range is higher in segments that constitute simple sentences than in segments that are part of complex sentences. Whether a segment is the second part of a coordinate sentence or whether it is a subordinate clause, does not make a difference for pause duration and pitch range.

Correlations between hierarchical position and prosodic features were computed on the original 10 levels in the hierarchy. The results are shown in Table 2. For both hierarchical position and clausal status the correlations with pause duration and pitch range were significant. As height of the discourse boundary in the hierarchy increases, pause duration and pitch range increase as well. As the clausal status between segments becomes more complex, pause duration and pitch range decrease.

Table 2. Standard scores of prosodic parameters related to hierarchy and clausal status between segments

		pause	pitch	rate
Hierarchy	1 = lowest boundary	-0.40	-0.39	-0.01
	2	0.04	-0.04	0.05
	3	0.49	0.18	-0.17
	4	0.68	0.30	0.16
	5 = highest boundary	0.65	0.36	0.09
	Correlation	.43**	.28**	.04
Clausal status	simple	0.17	0.16	0.01
	complex: coordinate	-0.93	-0.73	-0.02
	complex: subordinate	-0.76	-0.93	0.00
	Correlation	-.35**	-.37**	-.00

Note. *: $p < .05$, **: $p < .01$

3.2. Relation with nuclearity

Nuclearity is not independent of the clausal status: the more complex a segment, the more likely it is a satellite (the simple segments were satellites in 27% of the cases; complex coordinate segments were satellites in 40% of the cases; complex subordinate were satellites in 79% of the cases). Therefore both nuclearity and clausal status were included in the MANOVA. For clausal status there was an effect ($F(6, 1106) = 18.81$, $p< .001$, $\eta^2 = .09$), but not for nuclearity ($F(3, 552) = 1.29$, $p = .28$). There was no interaction between both factors ($F<1$). Univariate analyses however, showed an effect of nuclearity on articulation rate ($F(1, 554) = 3.87$, $p <.05$, $\eta^2 = .01$). This result shows that whether a segment is a nucleus or a satellite affects the rate of articulation. The means of the standard scores of the prosodic parameters in relation to clausal status and nuclearity are presented in Table 3. Articulation rate shows a highly consistent pattern: nuclei are read aloud more slowly (less phonemes per second) than satellites, regardless of their clausal status.

Table 3. Standard scores of prosodic parameters related to nuclearity for each clausal status

		pause	pitch	rate
Simple	Nuc.(n=344)	0.20	0.21	-0.05
	Sat. (n=124)	0.07	0.02	0.16
Complex: coordinate	Nucl. (n=28)	-0.92	-0.81	-0.22
	Sat. (n= 9)	-0.94	-0.62	0.27
Complex: subordinate	Nucl. n=10)	-0.84	-0.87	-0.17
	Sat. (n=37)	-0.74	-0.94	0.05

3.3. Relation with discourse relations

In our materials there was a systematic relation between particular discourse relations on the one hand and hierarchical position and clausal status on the other hand. Examples of the dependency with hierarchical position are: Background was associated frequently with high scores on hierarchical position, while Restatement was associated only with low scores on hierarchical position. Examples of the dependency with clausal status are: Evaluation and

Interpretation only occurred as simple segments, while Antithesis and Concession frequently occurred as complex coordinate segments. Because of these dependencies a number of cells were empty. Therefore, we conducted a MANOVA with hierarchical position and clausal status as covariates and not as independent factors. Discourse relation then was the independent factor and the three prosodic parameters were the dependent factors. The analysis of variance did not show a significant effect ($F(36, 1479)=1.17$, $p=.23$). This result means that the prosodic realizations of the thirteen discourse relations did not differ. The group of rhetorical relations however could be separated into causal and non-causal relations. A MANOVA with causality as the independent factor and hierarchical position and clausal status as covariates showed a significant effect ($F(3,502)=2.79$, $p<.05$, $\eta^2 = .02$). Univariate analyses showed that causality has an effect on pause duration ($F(1,504)=6.61$, $p<.025$, $\eta^2 =.01$). Means of the standard scores of the prosodic parameters in relation to causality are presented in Table 4. Segments that have a causal relation to the preceding segment are preceded by a shorter pause than segments that have a non-causal relation to the preceding segment.

Table 4. Standard scores of prosodic parameters related to causality

	pause	Pitch	rate
Causal	-0.31	-0.28	0.14
Non-causal	0.17	-0.01	0.01

4. Conclusion

The aspects of text organization as captured by Rhetorical Structure Theory, i.e. hierarchical position, nuclearity and rhetorical relations, have an effect on prosodic parameters. The duration of a pause preceding a segment and the pitch range of that segment increase as the position of that segment in the hierarchy increases. Besides that the clausal status of segments play a role, i.e. a simple main sentence has a longer preceding pause and a higher pitch range than a non-initial clause in a complex sentence. The more refined distinction between coordinate main clause and subordinate clause does not affect prosody. Nuclearity affects articulation rate, i.e. nuclear segments are read at a slower rate than non-nuclear segments. Pauses between segments which are related in a causal way are shorter than pauses between segments which are related in a non-causal way. The results of the present study provide evidence that the relation between prosody and text organization is more fine-grained than has been demonstrated by earlier research, in which the organization of a text is considered merely as a succesion of sentences and paragraphs.

5. Acknowledgements

The research reported in this paper was financially supported by the Cooperation Unit of Brabant Universities (SOBU). Special thanks go to Leo Vogten and Jan Roelof de Pijper for their assistance with Gipos and SampaCount; to Wilbert Spooren for his critical comments; to Carel van Wijk for his help with the statistical analyses; and to the audience of the Discourse op Dinsdag Meeting at Utrecht University, November 13, 2001 for their constructive criticism.

6. References

[1] Brubaker, R., 1972. Rate and pause characteristics of oral reading. *Journal of psychological research*, 1, 141-147.

[2] Lehiste, I., 1975. The phonetic structure of paragraphs. In *Structure and process in speech perception*, Cohen, A.; Nooteboom, S. (eds.). Berlin: Springer, 195-203.

[3] Silverman, K., 1987. *The structure and processing of fundamental frequency contours*. PhD Thesis, Cambridge University, Cambridge UK.

[4] Sluijter, A.; Terken, J., 1993. *Beyond sentence prosody: Paragraph intonation in Dutch*. Phonetica, 50, 180-188.

[5] Swerts, M., 1995. *Prosodic features of discourse units*. PhD Thesis. University of Technology Eindhoven.

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

[7] Hirschberg, J.; Grosz, B., 1992. Intonational features of local and global discourse structure. *Proceedings of the Speech and Natural Language Workshop*, New York, Harriman, NY, DARPA, Morgan Kaufmann Publishers, 441-446.

[8] Hirschberg, J.; Nakatani, C., 1996. A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the 34th Annual Meeting of the Association of Computational Linguistics*, Santa Cruz, CA, USA, 286-293.

[9] Mann, W.; Thompson, S., 1988. *Rhetorical Structure Theory: Toward a functional theory of text organization*. Text, 8, 243-281.

[10] Thorndyke, P., 1977. *Cognitive structures in comprehension and memory of narrative discourse*. Cognitive psychology, 9, 77-110.

[11] Grosz, B.; Sidner, C., 1986. *Attentions, intentions, and the structure of discourse*. Computational Linguistics, 12, 175-204.

[12] Sanders, T.; Wijk, C. van, 1996. *PISA, a procedure for analyzing the structure of explanatory texts*. Text, 16, 91-132.

[13] Ouden, J. den; Wijk, C. van; Terken, J.; Noordman, L., 1998. *Reliability of discourse structure annotation*. IPO Annual Progress Report 33, 129-138.

[14] Ouden, H. den; Terken, J.; Noordman, L., in preparation. Inter coder reliability of hierarchical organization of texts.

[15] Ouden, H. den,; Terken, J., 2001. Measuring pitch range. *Proceedings of the 7th Conference on Speech Communication and Technology*, Aalborg, Denmark, *vol. 1*, 91-94.