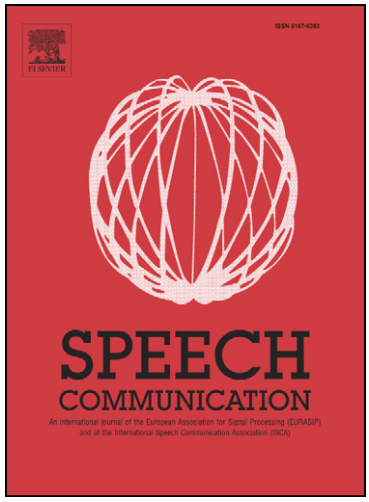
Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports

Hanny den Ouden, Leo Noordman, Jacques Terken



Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports

Hanny den Ouden, Leo Noordman, Jacques Terken

PII: S0167-6393(08)00100-3
DOI: 10.1016/j.specom.2008.06.003
Reference: SPECOM 1737

To appear in: *Speech Communication*

Received Date: 11 January 2008
Revised Date: 27 June 2008
Accepted Date: 27 June 2008

Please cite this article as: den Ouden, H., Noordman, L., Terken, J., Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports, *Speech Communication* (2008), doi: 10.1016/j.specom.2008.06.003

Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports

Hanny den Ouden [a], Leo Noordman [b], Jacques Terken [c]

[a] Hanny.denOuden@let.uu.nl
Utrecht University, Faculty of Humanities, Trans 10, 3512 JK Utrecht, The Netherlands

[b] noordman@uvt.nl
Tilburg University, Faculty of Arts, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

[c] j.m.b.terken@tue.nl
Technische Universiteit Eindhoven, Department Industrial Design, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Corresponding author:
Hanny den Ouden
Utrecht University
Faculty of Humanities
Trans 10
3512 JK Utrecht
The Netherlands
Telephone: 0031-30-2536012
Hanny.denOuden@let.uu.nl

Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports

Abstract

The aim of this research is to study effects of global and local structure of texts and of rhetorical relations between sentences on the prosodic realization of sentences in read aloud text. Twenty texts were analysed using Rhetorical Structure Theory. Based on these analyses, the global structure in terms of hierarchical level, the local structure in terms of the relative importance of text segments and the rhetorical relations between text segments were identified. The texts were read aloud. Pause durations preceding segments, F0-maxima and articulation rates of the segments were measured. It was found that speakers give prosodic indications about hierarchical level by means of variations in pause durations and pitch range: the higher the segments are connected in the text structure, the longer the preceding pauses and the higher the F0-maxima are realized. Also, it was found that speakers articulate important segments more slowly than unimportant segments, and that they read aloud causally related segments with shorter in-between pauses and at faster rate than non-causally related segments. We conclude that variation in pause duration and F0-maximum is a robust means for speakers to express the global structure of texts, although this does not apply to all speakers. Speakers also vary pause duration and articulation rate to indicate importance of sentences and meaning relations between sentences.

Keywords:
discourse, text structure, prosody, global structure, local structure, rhetorical relations

Running title:
Text structure and prosody

1 Introduction

The focus of this paper is on the relation between text structure and prosody. Text structure pertains to the organization of a text. A text is a collection of sentences that cohere: each sentence is related to another sentence or to a group of sentences. Prosody pertains to the suprasegmental aspects of speech, i.e. characteristics beyond the level of the individual speech sounds of vowels and consonants. Prosody is made up of a heterogeneous set of features which includes at least pausing, speaking rate, phrasing, intonation, rhythm, accentuation, and loudness. The study of text prosody focuses on the prosodic characteristics of clauses and sentences in relation to their position and function in text.

Early research in the area of text prosody was concerned with the prosodic realization of sentences at different positions in text. It was shown, for example, that first sentences of paragraphs have longer preceding pauses and higher pitch range than sentences within paragraphs, and that parenthetical and final sentences of paragraphs are articulated with lower pitch range and at faster rate than sentences at other locations in the text (Brubaker, 1972; Lehiste, 1975; Yule, 1980; Thorsen, 1985; Silverman, 1987; Swerts, 1997; Hirschberg & Grosz, 1992; Hirschberg & Nakatani, 1996). Later research also addressed content-related aspects of text. For example, Van Donzel (1999) applied Prince's distinction between new, inferable and evoked information (Prince, 1981), and found a systematic relation with prosody: new information is realized with a more prominent prosody than the other types of information. While in most previous studies paragraph boundaries were determined empirically by a panel of judges, Noordman, Dassen, Terken and Swerts (1999) aimed to arrive at a segmentation of text on a more principled basis. In a pilot study they applied two theories to capture both structure and content of text, namely Story Grammar (Thorndyke, 1977) and Rhetorical Structure Theory (Mann & Thompson, 1988). This study provided preliminary evidence that the durations of pauses preceding sentences and the heights of fundamental frequency peaks in those sentences gradually decrease as hierarchical levels decrease, where high levels are associated with more important information.

The present study elaborates on Noordman et al. (1999) using more texts, using naturally occurring texts in stead of constructed ones, and taking into account more factors. It is concerned with the question of how the structure and content of a text affect prosody when the text is read aloud. Texts are analysed with Rhetorical Structure Theory, which allows to distinguish between various text characteristics: the global structure of text in terms of

hierarchy, the rhetorical relations between the sentences, and the local structure in terms of more and less important information. The prosodic features measured are pause durations between sentences, pitch range and articulation rate of sentences. The aim of this research is to study effects of global and local structure of texts and of rhetorical relations between sentences on the prosodic realization of sentences in read aloud text.

2 Rhetorical Structure Theory

Rhetorical Structure Theory is a descriptive theory of the organization of natural text, characterizing its structure and content primarily in terms of rhetorical relations that hold between parts of the text (Mann & Thompson, 1988; henceforth: RST). Analyzing a text with RST leads to a binary tree-like structure that manifests how the individual units are related to each other, and how each unit contributes to the communicative goals of the text as a whole. RST was originally developed during the late 1980's to explain the textual coherence needed to drive automatic text generation. Later on it has been used successfully for a variety of tasks, ranging from linguistic text analyses to computational applications in language generation and automatic summarization.

Before an RST analysis can start, the text has to be divided into elementary units or segments. The theory specifies a syntactic criterion for that: segments are essentially clauses, where a clause is defined as (a part of) a sentence that contains a finite verb. Clausal subjects and complements, and restrictive relative clauses are considered parts of their host clauses rather than separate segments. For most clauses this segmentation criterion succeeds, but it is not optimal in a clause like segment 7 of the example text (see Table 1) that contains two successive complement clauses, the last one encasing a clause that is considered part of it. By taking the clause as the basic component for segmentation, RST does not distinguish between main clauses, coordinate main clauses and subordinate clauses. After segmentation, the individual segments are grouped into text spans and the rhetorical relations between adjacent text spans are specified. Finally, a hierarchical structure of the text is built.

RST explicitly describes about 25 rhetorical relations, for example, Volitional and Non-volitional Cause, Result, Consequence, Background, Motivation, and Elaboration. The rhetorical relations are defined in terms of conditions on two - occasionally more - segments. One of the segments in such a relation is a nucleus, and the other the satellite. The conditions are defined for the nucleus, the satellite, and the combination of nucleus and satellite, and in

terms of the effect on the reader. The nucleus is the central part of a text span, the most important part of it; the satellite is peripheral in the sense that a text without satellites can still be understood. These basic concepts of RST are illustrated with the help of one of the texts used in this study (see Table 1). The RST analysis of this text is presented in Figure 1.

Table 1 Example text (translated from of a Dutch news paper article)

segment	
1	The census in China has been extended by five days.
2	Normally it should have ended last Friday.
3	However, millions of people avoided the pollsters
4	or refused to open their doors.
5	This boycott was intended to keep secret illegal children or addresses.
6	During emergency talks last Friday, the government, i.e., the Chinese cabinet, decided to extend the census.
7	An official of the census committee in Beijing said that the committee's employees noticed that it was rather difficult to find active people at home by day or during the evening, but that of course many other people avoided them deliberately.
8	At least eighty million farmers, and that number could even be two hundred million, have squatted in the cities.
9	They had themselves registered at their addresses of origin
10	or they were not counted at all.
11	Although they were officially assured that the census has nothing to do with the police,
12	many people are afraid of reprisals when it is discovered that they do not have residence permits.
13	Also many married people who have more than one child boycotted the census,
14	because they were afraid that the committee of birth control would find this out.
15	The employees of the demographic committee now admit that most people did not keep the one-child policy.
16	One of these days even a family with ten children has been found in the region Shanxi.
17	In the opinion of the authorities, the counting of the homeless is not problematic.
18	It was said that there would not be many homeless people
19	and most of them would have an address in another region.
20	They would be counted at these addresses.
21	However, how this counting should happen is unclear.
22	Other people argue that their privacy is affected.
23	These people were found especially in the random sample of ten percent of the population which had to answer 49 detailed questions.
24	The remaining 90 percent only had to answer nineteen general questions.
25	People who have not yet been counted are encouraged by advertisements to report that to the census committee.
26	The people who have avoided meeting the pollsters thus far will probably not answer this call.
27	The whole affair has not been proved to be very helpful to the accuracy of this fifth census in the 51-year-old history of the People's Republic of China.

Source: *de Volkskrant, 3 November 2000*

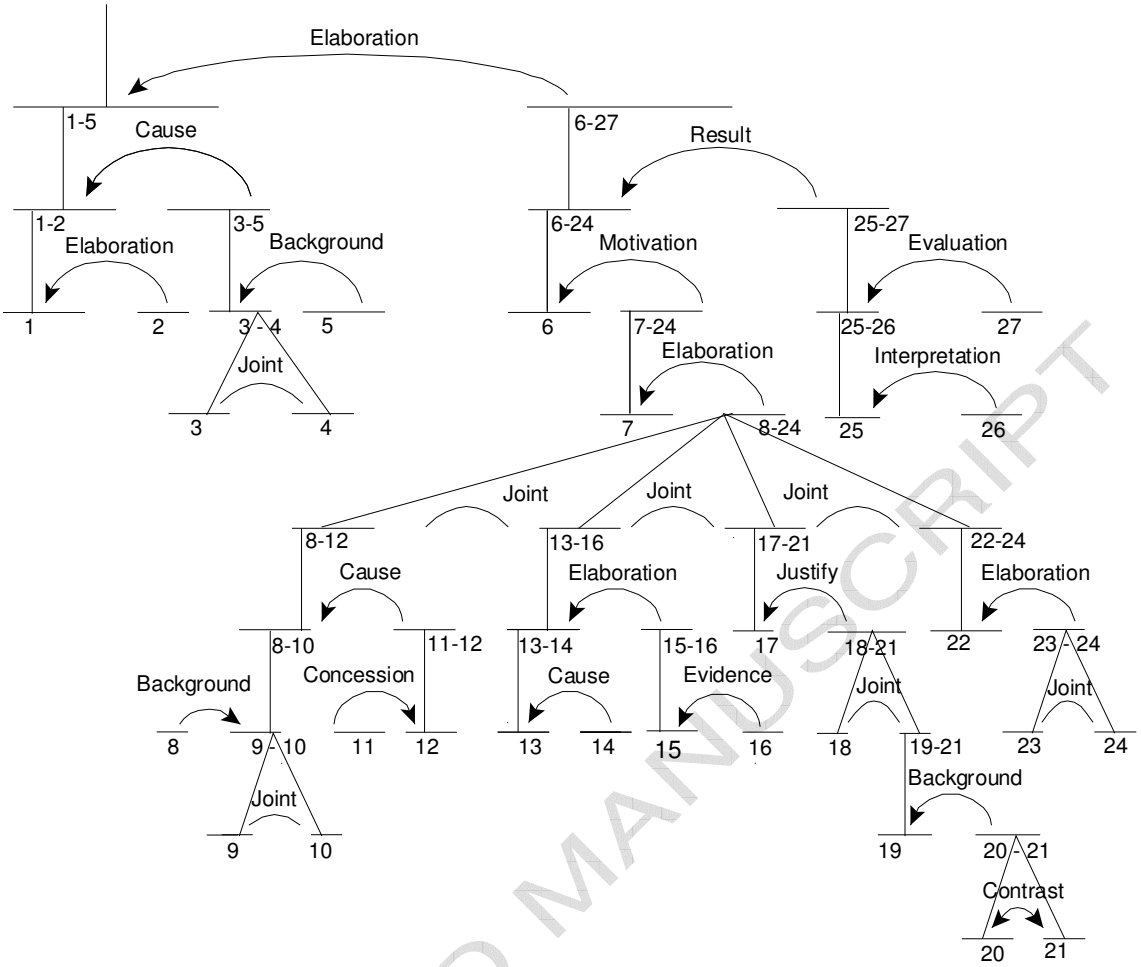


Figure 1 RST analysis of example text

Figure 1 shows the hierarchical organization of the 27 segments of the example text in Table 1. The arrows in the figure connect those parts of the text between which a rhetorical relation holds. A nucleus is represented by a vertical line; a satellite is represented by an outgoing arrow. Some relations like Joint and Contrast, contain two nuclei. The numbers under the horizontal lines indicate the segments that form a text span. The relation between text span 1-5 and text span 6-27 is characterized by Elaboration. This means that the content of segments 6-27 is considered to be an elaboration of the content of segments 1-5, which follows from the definition of the Elaboration relation as described by Mann and Thompson (1988). This relation definition is presented in Table 2.

Table 2. Relation definition Elaboration cited from Mann & Thompson (1988)

Elaboration	
Constraints on N + S combination:	S presents additional detail about the situation or some element of subject matter which is presented in N or inferentially accessible in N in one or more of the ways listed below. In the list, if N presents the first member of any pair, then S includes the second: 1) set: member; 2) abstract: instance; 3) whole: part; 4) process: step; 5) object: attribute; 6) generalization: specific.
The effect:	R recognizes the situation presented in S as providing additional detail for N. R identifies the element of subject matter for which detail is provided.
Locus of effect:	N and S

Note: N = nucleus; S = satellite; R = reader

One level lower in the hierarchy, text span 3-5 is characterized as a Volitional Cause of text span 1-2, based on the definition of the Volitional Cause relation that is presented in Table 3. One level lower, the segments are related to each other by way of an Elaboration: segment 2 elaborates the statement of the nucleus, segment 1. Segment 5 gives background information to segments 3-4, based on the relation definition of Background. In this way, all relations between text spans and between the individual segments on the bottom-level are analyzed on the basis of the relation definitions in Mann and Thompson (1988).

Table 3. Relation definition Volitional Cause cited from Mann & Thompson (1988)

Volitional Cause	
Constraints on N:	Presents a volitional action or a situation that could have arisen from a volitional action.
Constraints on N + S combination:	S presents a situation that could have caused the agent of the volitional action in N to perform that action; without the presentation of S, R might not regard the action as motivated or know the particular motivation; N is more central to W's purposes in putting forth the N-S combination than is S.
The effect:	R recognizes the situation presented in S as a cause for the volitional action presented in N.
Locus of effect:	N and S

Note: N = nucleus; S = satellite; R = reader; W = writer

In practice, analyzing texts using RST involves top-down and bottom-up analysis at the same time. In general, the analysis starts in a top-down way: the analyst divides the whole text into two large text spans and determines the rhetorical relation between them. These text spans are

in turn decomposed into two smaller text spans and the rhetorical relation between these spans is determined, until finally the level of the individual segments is reached. The bottom-up process works the other way around: the analyst relates two individual segments and assigns a relation definition to the pair of segments, thereby creating a text span; this text span is in turn related to another text span and a relation definition is assigned to it, and so on. Both strategies, top-down and bottom-up, are applied simultaneously until all segments of the text are connected in a tree-like structure. Several studies showed that a text analysis using RST can be assigned with sufficient inter-coder reliability. When people discuss their annotations of RST structures of a particular text afterwards, mostly it leads to consensus about it (Bateman & Rondhuis, 1997), and cases that do not lead to consensus can be explained by ambiguities in the text itself (Den Ouden, Van Wijk, Terken, & Noordman, 1998). Inter-coder reliability between six RST analysts who did not discuss their annotations of hierarchical levels afterwards, ranged from moderate to substantial expressed in terms of kappa (Den Ouden, 2004).

As the discussion of the example text in Table 1 and Figure 1 illustrates, RST-analyses yield the determination of many characteristics of texts. In relation to prosody we are interested in three of them: the global structure of text in terms of hierarchical levels, the rhetorical relations between sentences in terms of relation definitions, and the local structure in terms of nuclei and satellites. How do these characteristics affect the prosody of a text when it is read aloud?

3 Research questions

Concerning the prosodic realization of global text structure, earlier research showed that pause length decreases as transitions in texts are more subtle (for example, Schilperoord, 1996). Accordingly, predictions of the current study with regard to prosody are that pause durations are shorter, pitch range is lower and articulation rate is faster as the hierarchical levels in the global text structure are lower. As we noted earlier with regard to segmentation, RST does not distinguish between simple main clauses and subordinate clauses or coordinate main clauses. However, the syntactic class of the segments has to be taken into account when the relation between text characteristics and prosody is examined, because syntactic status may be correlated with the hierarchical level of segments. For example, main clauses may occur more often at higher levels in the global text structure than subordinate clauses do. In

addition, we may assume that prosody behaves different for various syntactic boundaries (Cooper & Paccia-Cooper, 1980).

Concerning the prosodic realization of content relations, two contrasting pairs of content relations will be examined, namely causal versus non-causal relations and semantic versus pragmatic relations. These distinctions are derived from the taxonomy of text relations by Sanders, Spooren and Noordman (1992), who propose four primitives on the basis of which all text relations can be categorized. One primitive is called Basic Operation, on the basis of which causal relations are distinguished from non-causal relations; another primitive is called Source, on the basis of which semantic and pragmatic relations are distinguished. These primitives will be explained here successively.

With respect to Basic Operation, a non-causal, or additive, relation exists if a conjunction relation can be deduced between two segments, whereas a causal relation exists if a relevant implication relation can be deduced. An example of a non-causal relation between two segments is (1); an example of a causal relation between two segments is (2).

(1) Mary went to the party and her sister stayed at home.
(2) Mary went to the party, because her friend invited her.

In (1) the second segment is an addition to the first segment, whereas in (2) the second segment of the pair caused the first segment. In psycholinguistic research on the way people process text, it has been found that sentences that are causally related have faster reading times than sentences that are non-causally related (Sanders & Noordman, 2000). It is claimed by the authors that non-causal relations are connected less strongly than causal relations. The prediction of the current study on prosody is that speakers also need less time to produce these sentences when they read them aloud. The rationale for the prediction is that speakers - as readers - process causally related sentences faster than non-causally related ones and therefore also produce them at a faster rate, or, alternatively, because faster reading more explicitly expresses the relatedness of the causal relation.

With respect to Source, a semantic relation exists if the coherence between segments is based on the coherence between the events in the world which are described, whereas a pragmatic relation exists if the coherence between segments is based on the illocutionary meaning of one or both of the segments, for example, when a writer or speaker draws a

conclusion. An example of a semantic pair is (3); an example of a pragmatic pair is (4), both cited from Sweetser (1990).

(3) Anna loves Victor because he reminds her of her first love.
(4) Anna loves Victor, because she told me so herself.

In (3) there is a consequence-cause relation of two events in the world, whereas in (4) the second segment can be paraphrased as 'I conclude that she loves him because I know the relevant data'. Sweetser (1990) argues that in a semantically related pair like (3), the consequence in the first segment is presupposed and only the causal relation between both segments is affirmed, whereas in a pragmatic pair like (4) the conclusion in the first segment can not be presupposed. She assumes therefore that pragmatically related segments require comma intonation, i.e., longer pauses, whereas semantic readings do not (Sweetser, 1990: 82). In the present study this claim is investigated empirically. The prediction on prosody is that pause durations between semantically related sentences are shorter than pause durations between pragmatically related sentences.

Concerning the prosodic realization of local text structure, earlier research showed that more important information is realized with a more prominent prosody than other types of information. The predictions of the current study with regard to prosody are that nuclei are prosodically more prominent than satellites: nuclear segments will have longer preceding pauses than satellites, they will have higher pitch range and a slower articulation rate than satellites.

4 Method

4.1 Text material

Twenty newspaper reports were selected from a Dutch national quality newspaper. The mean length of the texts was 28 segments, with a range from 19 to 37. The themes of the reports varied: politics, accidents, crimes, sports, social phenomena. The style of the news reports was informative and non-controversial. Minor syntactic changes were made in some formulations to facilitate the segmentation process. Most changes concerned direct speech that was changed into indirect speech. They were few in number. The texts were divided into

segments (clauses) according to the criteria given by RST. After segmentation, the first author analyzed the twenty texts using RST. The analyses were verified by a second experienced user of RST. In cases of disagreement, consensus was reached after discussion between the first and second analyst. The global structure of the twenty analyses was operationalized in terms of hierarchy; the rhetorical relations between the sentences of the twenty texts were operationalized in terms of causal versus non-causal and semantic versus pragmatic relations; and the local structure of the twenty analyses was operationalized in terms of nuclearity. These operationalisations are explained below. To avoid any confounding of syntactic class of the segments with the text structural characteristic, the syntactic class of the segments is controlled for in all statistical analyses.

4.1.1 Hierarchy

The hierarchical levels of the segments were determined on the basis of the depth of the boundaries between adjacent segments. Each boundary between segments was given a depth score in the following way. First, the RST analyses were aligned in such a way that all individual segments were at the bottom (see Figure 2 for a bottom-up representation of the text analysis in Figure 1). Second, for each boundary the superordinate node connecting the two segments adjacent to the boundary was determined, and the number of subordinate nodes including the connecting node itself was counted; the total number of nodes was considered the score of the boundary. Using this scoring procedure, low boundaries received low scores and high boundaries received high scores. For example, in Figure 2 the boundary between segments 24 and 25 is scored as 8: the superordinate node of segments 24 and 25 dominates seven nodes, five at the left side and two at the right side, and one is added for the connecting node itself. In the same way the boundary between segments 1 and 2 is scored as 1; the boundary between segments 2 and 3 as 4, and so forth.

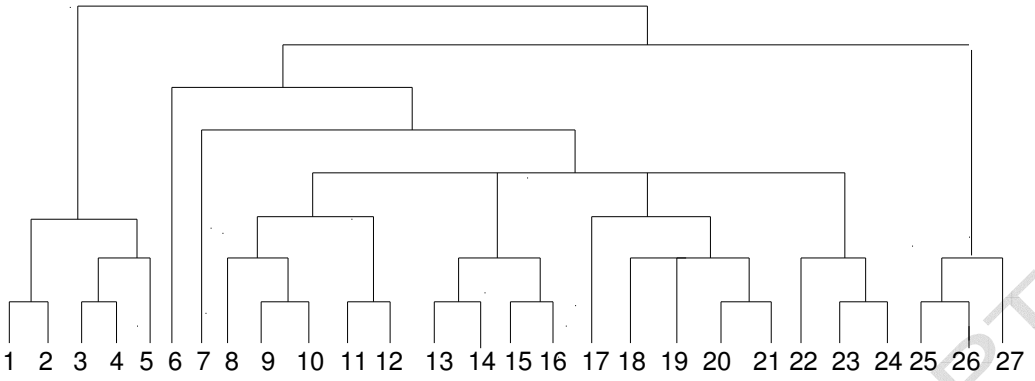


Figure 2 Bottom-up representation of text analysis in Figure 1,
numbers representing the segments as they occur in Table 1

The twenty texts contained 543 boundaries. The level scores ranged from 1 to 10. In the statistical analyses these ten levels were reduced to five, because there were few boundaries with score 4 or 5, and even fewer boundaries with score 6 or higher. Therefore, scores 4 and 5 were put together into one class, and so were scores 6 to 10. This resulted in 210 boundaries for level 1, 133 for level 2, 76 for level 3, 77 for level 4, and 46 for level 5.

4.1.2 Rhetorical relations

The boundaries in the hierarchical structures were classified in accordance with their associated rhetorical relations. For example, the relation between segments 1 and 2 in Figure 1 was classified as an Elaboration; the relation between segments 4 and 5 was classified as Background. Strictly speaking, the Background relation does not exist between segments 4 and 5, but it exists between segments 3 and 4 on the one hand, and segment 5 on the other hand. Nevertheless, the relation name is attributed to the boundary between segments 4 and 5, and segment 5 is considered the second segment of this Background relation.

Twenty-one content relations occurred in the RST analyses: Elaboration (n=119), Joint (n=108), Background (n=48), Cause (n=39), Result (n=36), Concession (n=26), Sequence (n=25), Contrast (n=22), Circumstance (n=21), Interpretation (n=20), Restatement (n=17), Antithesis (n=15), Evaluation (n=14), Justify (n=8), Condition (n=7), Solutionhood (n=6), Purpose (n=6), Motivation (n=2), Enablement (n=2), Evidence (n=1), and Summary (n=1). Only those rhetorical relations that occurred more than ten times were included in the statistical analyses. This was the case for thirteen rhetorical relations.

The rhetorical relations were then classified in terms of Basic Operation, as causal and non-causal relations, using the classification of Mann and Thompson (1988). The causal relations that occurred more than ten times were Cause, Result and Concession (n=101). The non-causal relations that occurred more than ten times were Elaboration, Joint, Background, Sequence, Contrast, Circumstance, Interpretation, Restatement, Antithesis, and Evaluation (n=409).

The rhetorical relations were also classified in terms of Source, as semantic and pragmatic relations based on the classification of Mann and Thompson (1988), who referred to it by 'subject matter' and 'presentational' relations respectively. Semantic or 'subject matter' relations that occurred more than ten times were Elaboration, Circumstance, Cause, Result, Interpretation, Evaluation, Restatement, Sequence, and Contrast (n=313). Pragmatic or 'presentational' relations that occurred more than ten times were Antithesis, Background, and Concession (n=89). The Joint relation is not classified either as semantic or as pragmatic because Mann and Thompson did not classify it as such (1988). Therefore the Joint relation was excluded from the analyses.

4.1.3 Nuclearity

Nuclei and satellites were defined on the basis of the relation definitions. In the RST analysis of the example text (see Figure 1), the nuclei are segments 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, and 18 to 25; and the satellites segments 2, 5, 8, 11, 14, 16, 26, and 27. The twenty RST analyses contained 383 nuclei and 180 satellites. Nuclei outnumbered satellites because many segments were connected by a Joint, Sequence, or Contrast relation. These relations consisted of two (or more) nuclei.

4.2 Procedure

The twenty written news reports were read aloud by twenty native speakers of Standard Dutch, ten males and ten females, most of them advanced students or employees at the former Center for User-System Interaction at Technische Universiteit Eindhoven and the Faculty of Arts at Tilburg University[1]. They were highly educated people with much reading experience. Each speaker read aloud one text. They were instructed to imagine that they would read aloud the text for a blind person as clearly as possible. The speakers prepared the reading aloud carefully. They were encouraged to make notes in the text to improve their reading-aloud. The

1 The speech of the example text in Table 1 is available on http://www.let.uu.nl/~Hanny.denOuden/personal/index.htm

preparation was intended to focus the readers' attention on the content and the structure of the text, and to enable them to read it aloud as much as possible in accordance with their mental representation of the text. The texts presented to the speakers contained capitals and punctuations marks, but no paragraph markers as indentations or blanc lines, to enable them to make their own representation based on the content of the text[2]. The recordings were made in a sound-proof room using a DAT-recorder. The speech was digitized using the speech-processing program Gipos[3]. It contains an algorithm for pitch measurement based on subharmonic summation (Hermes, 1988).

4.3 Speech material

Three prosodic features were predicted to have a systematic relationship with the three text characteristics, namely pause durations between the RST-defined segments, pitch range and articulation rate of segments. They were measured in the following way.

For pause duration, the beginnings and endings of the periods of silence at the boundaries between the segments were determined manually by visual inspection of the waveform. Then, the time in between was determined automatically in milliseconds.

Pitch range was operationalized as the F0-maximum of a segment. First, the pitch contour of each individual segment was inspected for pitch-measurement errors, like voiced-unvoiced errors and incorrect outliers, and for F0-maxima occurring at final rises. The errors were corrected; F0-maxima associated with final rises were removed, because they are not good estimates of the pitch range of the contour. Then the F0-maximum of each segment was measured automatically in hertz (Den Ouden & Terken, 2001).

Articulation rate was defined as the number of phonemes per second. The number of phonemes was computed automatically on the hand-corrected canonical transcriptions of the segments using a program called SampaCount.

Table 4 presents the prosodic characteristics of the twenty texts for both female and male speakers. Pause durations of 0 milliseconds occurred in the speech material, because some sequences of RST-defined segments were read aloud without intervening pauses. Because no

2 In our theory-based approach of text structure, paragraph structure could not be defined independently. Of course, future research may compare the present RST analysis with intuitive assignments of paragraph boundaries by the writers.

3 This programme is not supported any more and cannot be downloaded. It is available on request.

pause preceded the first segment of each text, the number of scores for pause duration is twenty less than that for F0-maximum and articulation rate.

Table 4 Prosodic characteristics of the twenty news reports in relation to gender

			minimum	maximum	mean	standard deviation
pause duration (milliseconds)	female	(n=268)	0	2298	801	374
	male	(n=275)	0	2380	917	426
F0-maximum (hertz)	female	(n=278)	194	364	281	34
	male	(n=285)	99	252	169	29
articulation rate (phonemes/sec)	female	(n=278)	10.5	18.4	14.6	1.50
	male	(n=285)	8.2	21.0	14.6	1.76

Table 5 presents the mean pause duration, F0-maximum and articulation rate of the raw prosodic data for each speaker. Within gender the table is arranged from the shortest to the longest pause duration.

Table 5 Mean and standard deviations (within parentheses) for pause duration, F0-maximum and articulation rate for each speaker/text

	gender of speaker	pause duration (in milliseconds)	F0-maxima (in hertz)	articulation rate (in phonemes per second)
Text 6	female	623 (379)	243 (19)	15.8 (1.5)
Text 3	female	712 (274)	258 (17)	15.2 (1.5)
Text 19	female	724 (244)	324 (24)	13.6 (1.0)
Text 5	female	768 (444)	291 (26)	15.8 (1.0)
Text 14	female	795 (393)	261 (21)	13.9 (1.0)
Text 20	female	894 (405)	308 (36)	14.8 (1.3)
Text 7	female	835 (421)	300 (20)	15.6 (1.0)
Text 11	female	852 (465)	291 (23)	14.1 (1.3)
Text 16	female	907 (319)	268 (21)	14.4 (1.4)
Text 9	female	972 (318)	263 (25)	13.2 (1.4)
Text 2	male	587 (325)	146 (15)	13.7 (1.3)
Text 4	male	743 (171)	161 (12)	15.2 (0.8)
Text 12	male	743 (306)	171 (16)	15.0 (1.3)
Text 15	male	833 (267)	201 (22)	14.8 (1.3)
Text 1	male	827 (418)	153 (20)	14.7 (2.5)
Text 10	male	841 (379)	136 (24)	16.4 (1.3)
Text 18	male	867 (274)	154 (18)	13.5 (1.4)
Text 17	male	919 (208)	193 (19)	12.8 (1.3)
Text 8	male	1273 (476)	186 (23)	13.5 (1.1)
Text 13	male	1496 (437)	188 (27)	15.4 (1.5)

As Table 5 shows, there was substantial individual variation in the pause durations, pitch range and articulation rates of the twenty speakers. Therefore, the raw prosodic data of the twenty texts were standardized per speaker/text. All analyses were performed on these standard scores.

5 Results

In this section we examine whether the three text characteristics, i.e. hierarchy, rhetorical relations and nuclearity affect the prosodic features pause duration, pitch range and articulation rate. First, the effect of syntactic class of the segments on the prosodic features is addressed; this is described in section 5.1. In sections 5.2 to 5.4, the investigations of the relations between hierarchy, rhetorical relations and nuclearity on the one hand, and the prosodic characteristics on the other hand, are reported. Syntactic class is included as between-groups factor in these analyses.

5.1 Effect of syntactic class on prosody

Segments of four syntactic classes occurred in the text materials. First, main clauses and main clauses that were the first part of a complex sentence consisting of two coordinate main clauses; these were called 'main segments in initial position'. Second, main clauses that were the second part of a complex sentence consisting of two coordinate main clauses connected by 'but', 'since', or 'and'; these were called 'coordinate main segments in non-initial position'. Third, subordinate clauses preceding main segments; these were called 'subordinate segments in initial position'. Fourth, subordinate clauses following main segments; these were called 'subordinate segments in non-initial position'. For each prosodic parameter a one-way ANOVA was run with Syntactic Class as independent variable with the four levels just mentioned. Table 6 presents the prosodic characteristics of each of these syntactic classes.

Table 6 Pause duration, F0-maximum, and articulation rate in relation with Syntactic Class (in standard scores)

	main segments in initial position (n=469)	coordinate main segments in non-initial position (n=47)	subordinate segments in initial position (n=10)	subordinate segments in non-initial position (n=37)
preceding pause	0.22*	-1.15	0.48	-1.31
F0-maximum	0.17	-0.74	-0.26	-1.09
articulation rate	-0.01	-0.03	-0.24	0.23

* Based on 449 cases since a pause preceding the first segment of each text could not be measured

Syntactic Class affected pause duration ($F(3,539)=71.77$, $p<.001$, $\eta^2=.29$), and F0-maximum ($F(3,559)=33.72$, $p<.001$, $\eta^2=.15$), but not articulation rate ($F<1$). Pairwise comparisons in

post-hoc analyses (Tukey's HSD procedure) showed that pauses preceding main segments in initial position and subordinate segments in initial position were significantly longer than pauses preceding coordinate main segments in non-initial position and subordinate segments in non-initial position. The same pattern was shown for F0-maximum, except that the F0-maximum of subordinate segments in initial position did not differ significantly from that of main segments in non-initial position. In order to simplify the analysis, we decided to divide the segments in two categories: main segments and subordinate segments in initial position on the one hand and coordinate main segments and subordinate segments in non-initial position on the other hand. For reasons of clarity, we changed the name of the factor from Syntactic Class to Position. In the following analyses Position is included as independent factor with two values, 'initial' and 'non-initial'.

5.2 Effect of hierarchy on prosody

In this section the relation between the hierarchical levels of the boundaries between segments and the prosodic characteristics is examined. A one-way analysis of variance was run with Hierarchy (five levels) as independent factor, and, one at a time, the three prosodic parameters as dependent factor. The non-initial segments could not be analyzed, because for the non-initials there was only one observation at level 4, and no observations at levels 3 and 5. Table 7 presents mean standard scores of pause duration, F0-maximum, and articulation rate for each hierarchical level.

Table 7 Prosodic characteristics in relation with Hierarchy (in standard scores); level 1 = lowest level, level 5 = highest level in hierarchical structure

	level 1 (n=138)	level 2 (n=122)	level 3 (n=76)	level 4 (n=76)	level 5 (n=46)
preceding pause	-0.16	0.05	0.50	0.62	0.69
F0-maximum	-0.10	0.01	0.19	0.29	0.36
articulation rate	-0.05	0.11	-0.21	0.11	0.01

For the initial segments, Hierarchy affected pause duration ($F(4,453)=19.85$, $p<.001$, $\eta^2=.15$). Durations of pauses increased as hierarchical levels increased. There was an effect of Hierarchy on F0-maximum ($F(4,453)=4.30$, $p<.005$, $\eta^2=.04$). F0-maximum increased as

hierarchical levels increased. Articulation rate was not affected by Hierarchy ($F(4,453)=1.71$, $p=.15$).

For the initial segments, correlations were computed for the five hierarchical levels and each of the three prosodic parameters. The correlation between Hierarchy and pause duration was .38 ($p<.001$), between Hierarchy and F0-maximum .19 ($p<.001$), between Hierarchy and articulation rate .01 (n.s.). These correlations were based on the mean prosodic realizations of the twenty speakers. Notwithstanding the correlations with pauses and F0-maxima, we will not ignore the fact that the twenty individual speakers differed with regard to the extent to which they realized the hierarchical levels prosodically. Table 8 presents the correlations for each speaker separately. The table is arranged per gender of the speakers in descending order of the correlations between Hierarchy and pause durations.

Table 8 Correlations between Hierarchy and the prosodic characteristics per speaker (n denotes number of initial segments)

			pause duration	F0-maximum	articulation rate
female speakers	speaker 11	(n=24)	.73 **	.31	-.01
	speaker 6	(n=22)	.61 **	.57 **	-.02
	speaker 14	(n=26)	.58 **	.51 **	-.01
	speaker 5	(n=19)	.57 *	.31	.11
	speaker 16	(n=24)	.47 *	-.29	-.06
	speaker 20	(n=21)	.41	-.18	.04
	speaker 7	(n=25)	.12	.01	.41
	speaker 19	(n=27)	.12	.04	-.03
	speaker 9	(n=19)	.06	-.11	.32
	speaker 3	(n=26)	-.03	-.01	.06
male speakers	speaker 1	(n=21)	.61 **	.77 **	.04
	speaker 13	(n=23)	.61 **	.22	-.19
	speaker 12	(n=27)	.56 **	-.05	-.16
	speaker 18	(n=17)	.54 *	.24	-.19
	speaker 17	(n=24)	.45 *	.43 *	.17
	speaker 10	(n=30)	.34	.49 **	.01

speaker 8	(n=25)	.33	.24	-.24
speaker 15	(n=20)	.26	-.12	-.10
speaker 2	(n=21)	.11	-.51 *	.12
speaker 4	(n=22)	.05	.45 *	.12

Note: *: p<.05; **: p<.01

Correlations between Hierarchy and the prosodic characteristics differed among speakers. Ten out of twenty speakers realized longer pauses as the boundaries in the hierarchical structure were higher. Six of the twenty speakers realized higher F0-maxima as boundaries were higher. Four speakers realized both pauses and F0-maxima in the expected way. For articulation rate none of the correlations reached significance.

5.3 Effects of rhetorical relations on prosody

This section addresses the question whether prosody is affected by rhetorical relations. Firstly, it is examined whether causal and non-causal relations have different prosodic characteristics. Second, it is examined whether semantic and pragmatic relations have different prosodic characteristics. Both analyses concern the position of the second segment of the relation, because the relation was assigned to the boundary preceding the second segment of the related pair.

5.3.1 Causal and non-causal relations

The distributions of causal and non-causal relations are examined for the two positions to find out whether Basic Operation and Position were confounded. These factors were related ($\chi^2(1)=24.37$, $p<.001$): the proportion of initial second segments was higher for non-causal relations than for causal relations (88 versus 68 percent). The distributions of causal and non-causal relations are also examined for the hierarchical levels. Table 10 presents this distribution.

Table 10 Distribution of causal and non-causal relations per hierarchical level, (1=lowest level; 5=highest level in hierarchical structure)

		level 1	level 2	level 3	level 4	level 5
causal relation	(n=101)	46 (45%)	28 (28%)	11 (11%)	13 (13%)	3 (3%)
non-causal relation	(n=408)	150 (37%)	96 (23%)	60 (15%)	58 (15%)	40 (10%)

There was no reliable association between Basic Operation and Hierarchy ($\chi^2(4)=8.09$, p=.09). Causal and non-causal relations were distributed more or less evenly over the levels.

Two-way ANOVAs were run with Basic Operation (two levels: causal, non-causal) and Position (two levels: initial, non-initial) as independent variables, Hierarchy as a covariate, and, one at a time, the three prosodic parameters as dependent variables. Hierarchy was not included as independent factor, because of too-low frequencies in some cells of the matrix. Table 11 presents the prosodic characteristics of causal and non-causal relations in relation with the position of the second segment.

Table 11 Prosodic characteristics in relation with Basic Operation and Position of the second segment of the pair (in standard scores)

		causal	non-causal
initial	preceding pause	0.01	0.29
	F0-maximum	-0.05	0.12
	articulation rate	0.16	-0.03
non-initial	preceding pause	-1.32	-1.15
	F0-maximum	-0.76	-1.00
	articulation rate	0.32	-0.06

Note: number of cases for initial, causal: 69, non-causal: 360; number of cases for non-initial, causal: 32, non-causal: 48

Pauses between causally related segments were shorter than those between non-causally related segments ($F(1,504)=4.59$, $p<.05$, $\eta^2=.01$). Pauses were longer preceding initial segments than preceding non-initial segments ($F(1,504)=100.71$, $p<.001$, $\eta^2=.17$). There was no interaction ($F<1$).

Causality did not affect the F0-maximum ($F<1$). The position of the second segments affected the F0-maxima ($F(1,504)=38.88$, $p<.001$, $\eta^2=.07$). F0-maxima were higher for initial than for non-initial segments. There was no interaction ($F(1,504)=2.64$, p=.11).

Articulation rate was affected by causality ($F(1,504)=5.07$, $p<.05$, $\eta^2=.01$). Causally related segments were read faster than non-causally related segments, i.e. speakers read more phonemes per second. There was no effect of Position and no interaction (both F's<1).

5.3.2 Semantic and pragmatic relations

The distributions of semantic and pragmatic relations are examined over the positions to find out whether Source and Position were confounded. There was a dependence between Source and Position of the second segment of these relations ($\chi^2(1)=5.55$, $p<.025$). The proportion of initial second segments was higher for semantic relations than for pragmatic relations: 88 versus 78 percent. The distributions of semantic and pragmatic relations are also examined over the hierarchical levels. Table 12 presents this distribution.

Table 12 Distribution of semantic and pragmatic relations per hierarchical level (1=lowest level; 5=highest level in hierarchical structure)

		level 1	level 2	level 3	level 4	level 5
semantic relation	(n=305)	117 (37%)	77 (25%)	46 (15%)	44 (14%)	29 (9%)
pragmatic relation	(n=88)	30 (34%)	21 (23%)	13(15%)	14 (16%)	11 (12%)

There was no association between Source and Hierarchy ($\chi^2(4)=1.09$, $p=.90$).

Two-way ANOVAs were run with Source (two levels: semantic, pragmatic) and Position (two levels: initial, non-initial) as independent variables, Hierarchy as a covariate, and, one at a time, the three prosodic parameters as dependent variables. Hierarchy was not included as independent factor, because of too-low frequencies in some cells of the matrix. Table 13 presents the prosodic characteristics of the semantic and pragmatic relations in relation with the position of the second segment.

Table 13 Prosodic characteristics in relation with Source and Position of the second segment of the pair (in standard scores)

		semantic	pragmatic
initial	preceding pause	0.17	0.30
	F0-maximum	-0.01	0.26
	articulation rate	0.05	0.01
non-initial	preceding pause	-1.33	-1.01
	F0-maximum	-0.92	-0.74
	articulation rate	0.23	0.31

Note: number of cases for initial, semantic: 274, pragmatic: 69; number of cases for non-initial, semantic: 39; pragmatic: 20

Pause duration did not differ for semantic and pragmatic relations ($F(1,397)=2.76$, $p=.10$). Position did affect pause duration ($F(1,397)=67.86$, $p<.001$, $\eta^2=.15$). Initial second segments had longer preceding pauses than non-initial second segments. There was no interaction between Source and Position ($F(1,397)=1.78$, $p=.18$).

Semantic relations had no different F0-maxima than pragmatic relations ($F(1,397)=2.24$, $p=.14$). Position did affect F0-maximum ($F(1,397)=26.13$, $p<.001$, $\eta^2=.06$). Initial second segments had higher F0-maxima than non-initial second segments. There was no interaction ($F<1$).

Articulation rate was not affected by Source ($F<1$), nor by Position ($F(1, 397)=2.10$, $p=.15$). There was no interaction ($F<1$).

5.4 Effect of nuclearity on prosody

This section addresses the question whether prosody was affected by the nuclearity of the segments. First, the distributions of nuclei and satellites are examined over the positions to find out whether Nuclearity and Position were confounded. Nuclearity and Position were related to each other ($\chi^2(1)=26.11$, $p<.001$). Nuclei were more often initial segments than satellites were (90 versus 74 percent). Second, the distributions of nuclei and satellites are examined over the hierarchical levels to control for any confounding of the two factors. Table 14 presents this distribution.

Table 14 Distribution of nuclei and satellites per hierarchical level (1=lowest level; 5=highest level in hierarchical structure)

		level 1	level 2	level 3	level 4	level 5
nucleus	(n=362)	86 (23%)	104 (29%)	65 (18%)	69 (19%)	39 (11%)
satellite	(n=170)	124 (69%)	29(16%)	11 (6%)	8 (5%)	7 (4%)

Nuclearity and hierarchical level were related ($\chi^2(4)=108.12$, $p<.001$). Nuclei occurred more frequently at the higher levels than satellites did. At the lowest level, level 1, most satellites were found.

Two-way ANOVAs were run with Nuclearity (two levels: nucleus, satellite) and Position (two levels: initial, non-initial) as independent variables, Hierarchy as covariate, and, one at a time, the three prosodic parameters as dependent variables. Hierarchy was not

included as independent factor, because of too-low frequencies in some cells of the matrix. Table 15 presents the prosodic characteristics of nuclei and satellites in relation with Position.

Table 15 Prosodic characteristics in relation with Nuclearity and Position (in standard scores)

		nucleus	satellite
initial	preceding pause	0.28	0.08
	F0-maximum	0.13	-0.01
	articulation rate	-0.03	0.08
non-initial	preceding pause	-1.17	-1.26
	F0-maximum	-0.83	-0.94
	articulation rate	-0.17	0.29

Note: number of cases for initial, nucleus: 326, satellite: 132; number of cases for non-initial, nucleus: 37, satellite: 47

Nuclearity did not affect pause duration ($F<1$). Position affected pause duration ($F(1,537)=127.48$, $p<.001$, $\eta^2=.19$). Pauses preceding initial segments were longer than pauses preceding non-initial segments. No interaction was observed for pause duration ($F<1$).

F0-maxima of nuclei and satellites did not differ ($F<1$). Position had an effect on F0-maxima ($F(1,537)=53.67$, $p<.001$, $\eta^2=.09$). They were higher for initial segments than for non-initial segments. There was no interaction ($F<1$).

Nuclearity affected articulation rate ($F(1,537)=6.50$, $p<.025$, $\eta^2=.01$). The number of articulated phonemes per second was less for nuclei than for satellites: nuclei were read aloud more slowly than satellites. No effect of Position ($F<1$) was found and no interaction ($F(1,537)=1.85$, $p=.18$).

6 Conclusion and discussion

To summarize, the global structure, rhetorical relations and the local structure of text are realized prosodically in different ways. The global structure of a text is signalled by variation in pause durations and F0-maxima: pauses at higher boundaries are longer than pauses at lower boundaries and F0-maxima of segments following higher boundaries are higher than F0-maxima of segments following lower boundaries. For rhetorical relations, causality is

signalled by variation in pause durations and articulation rates: pauses between causally related segments are shorter than pauses between non-causally related segments and causally related segments are read aloud faster than non-causally related segments. The local structure is signalled by means of articulation rate: nuclei are read at slower rate than satellites. The three text characteristics did not affect prosody to the same extent. The explained variance was higher for global structure than for causal relations and local structure. In addition, we find that the position of the segments in the sentence has a clear effect: segments in initial position are preceded by longer pauses and have a higher F0 maximum than segments in non-initial position, when effects of hierarchy, causality and nuclearity are partialed out. Finally, we find that effects of hierarchy are not systematically produced by all speakers.

In the following paragraph we discuss the text characteristics under consideration one by one by means of some text examples. For hierarchy, the predictions of the current study with regard to prosody were that pause durations, pitch range and articulation rate would correlate positively with the hierarchical levels in the global text structure: as segments are at lower levels in the hierarchical structure they have shorter pauses, lower pitch and faster rate. These hypotheses are confirmed for pause duration and pitch range, not for articulation rate. By means of fragments of the example text the general pattern of gradually decreasing pause durations from higher to lower levels will be illustrated. In segment 7, the writer notes the difficulty of finding people at home. The writer illustrates this issue with the mentioning of four groups of people: farmers (8-12), married people (13-16), homeless people (17-21), and people who argue that their privacy is affected (22-24). The global structure of the text shows that the four text parts are four instances of segment 7, but that they as a whole cohere as one argument. Within the sequence of segments 8-24, the pause durations were longer at the transitions of the four instances, i.e. the ones between 7-8 (0.96), 12-13 (0.96), 16-17 (1.99) and 21-22 (1.06), than at transitions within these text parts. The global structure in Figure 1 also shows that the solution for the problem posed in segment 7 is introduced in segment 25 at a higher level than the preceding text: at the boundary between 24 and 25 the pause duration was also relatively long (1.36). These results regarding pause duration are in accordance with the findings of Schilperoord (1996) who showed that, in dictated speech, pause durations were shorter when transitions in the text structure were subtler. The effect of hierarchy on the F0-maximum in the current study extends Schilperoord's findings.

Individual speakers differed in the extent to which they marked global structure of text with pauses and F0-maxima. One of the possible reasons for the individual variation in

realizing textual characteristics prosodically is that we did not make a distinction between those readers who are 'skilled' at reading aloud and those who are not (Wichmann, 2000: 20). With regard to the goal of this study, i.e. demonstrating the relation between text structure and prosody, it might have been better to select only skilled oral readers since not all people with much reading experience are good oral readers. Close listening to the speech material gave the impression that some speakers have read aloud the succession of sentences in a sort of 'staccato' way, maybe because they tried so hard to read aloud the text as clearly as possible. Another explanation for the individual differences might be that some speakers use other prosodic characteristics to indicate text structure than those measured in this study, for example loudness or vowel lengthening. For instance, Speaker 2, the one realizing a significant opposite pattern for F0-maximum, was a speaker with a very monotonous voice, and therefore he would need to have recourse to other prosodic features to signal communicatively relevant distinctions; he also was the most salient example of the staccato way of reading aloud. Also the contents of the texts themselves were different: a lot of variation within sentences was due to their meaning and meaning relations. Because speakers read one text only, a correlation between text and speaker cannot be excluded in the sense that we do not know whether an individual speaker would show the same prosodic patterns when he or she would have read another text.

For causal relations between sentences the prediction with regard to prosody was that speakers would need less time to read aloud causally related sentences than non-causally related sentences. This hypothesis is confirmed, because pause durations were shorter for causally related segments, and causally related segments were read aloud with faster articulation rate. In the example analysis presented in Figure 1, the relations between segments 2 and 3, segments 10 and 11, segments 11 and 12, and segments 13 and 14 were characterized as causal ones. The pair of segments 13 and 14 is a very clear example of a causal relation: "Also married people who have more than one child boycotted the census, because they were afraid that the committee of birth control would find it out." The pause duration at the transition between the two segments was very short (-1.44) and the second segment was read aloud at relatively fast rate, i.e. many phonemes per second (0.45). These results on the prosodic realization of causality may be interpreted in accordance with the findings of Sanders and Noordman (2000). They showed that causal relations need a shorter processing time than non-causal relations. Our results show that the reduction of time is also

manifest in the production of speech. The shortening of pauses and increase of articulation rate indicate that causally related segments cohere more strongly than non-causally segments.

We should repeat a critical remark here. The findings may suggest that the causal relations in our material existed between two adjacent segments. However, the text structure in Figure 1 shows that the causal relations existed not only between adjacent segments, but also between larger text spans. For example, the text span containing segments 11 and 12 is causally related to the text span containing segments 8 to 10, so that in fact we should predict that both segments 11 and 12 would be read faster. However, due to the way the rhetorical relations were operationalized, the causal relation was associated with the boundary between segments 10 and 11, whereas strictly speaking there was no causal relation between segments 10 and 11 at all. Further experimental research on the prosodic realization of causality should be conducted.

For semantic and pragmatic relations between sentences the prediction with regard to prosody was that speakers would read semantically related segments in a different way than pragmatically related segments, the last ones being read with comma intonation. We did not operationalize comma intonation as such as a dependent variable, for instance as a high rise at the end of the segment preceding the pragmatically related segment, but one aspect of comma intonation is a relatively long pause duration. The predictions based on Sweetser (1990) were not supported by the present study: we did not find longer pauses for pragmatically related segments than for semantically related segments. No prosodic differences at all were found between semantic and pragmatic relations. The same critical remark as the one mentioned for causality is justified here. The semantic and pragmatic relations in the text material did not only exist between adjacent segments, but also between larger text spans. For example, the Motivation relation between segments 6 and 7 was not restricted to these two segments, but concerned segment 6, on the one hand, and the text span consisting of segments 7 to 24, on the other hand.

Another issue with regard to the distinction between semantic and pragmatic relations is that the number of pragmatic relations in our text material was about one quarter of the total number of relations. This number seems to be relatively high in newspaper reports that are intended to describe events in an objective way. Although we used Mann and Thompson's list of 'subject matter' and 'presentational' relations as a criterion for classifying the relations as semantic and pragmatic ones, we wonder whether this list is satisfactory. Some authors, like Potter (2007), argue that Background and Elaboration are similar relations, the principal

distinction being that Background precedes the nucleus and Elaboration follows it, and that they both are 'subject matter' relations. If we had classified Background as a 'subject matter' relation in stead of a 'presentational' relation, the result with respect to the semantic-pragmatic distinction might have been different. To demonstrate the pure effect of semantic and pragmatic relations on prosody, follow-up research has to be conducted with a more systematic manipulation of this rhetorical relation.

For nuclearity, the predictions with regard to prosody were that nuclei would be read aloud with more prominent prosody than satellites: nuclear segments would have longer preceding pauses than satellites, they would have higher pitch range and a slower articulation rate than satellites. These hypotheses are confirmed for articulation rate, not for pause duration and not for pitch range. Nuclei were read at a slower rate than satellites. Segment 25 is a clear example of a nucleus that is read aloud at slower rate (-0.83) than many other segments, and that is quite understandable since it is one of the core sentences of the text. In general, nuclei contain more important information for understanding the text as a whole than satellites do. The fact that they are read aloud slower is an indication for that importance.

One of the intriguing questions with regard to the relation between text structure and prosody is what reasons speakers have to vary their prosody in order to signal text characteristics. We think that there are two possible explanations that do not exclude each other. One is concerned with conceptualization, the other with communication.

Firstly, the systematic variation of various prosodic features can be explained in terms of the way a text is conceptualized by a speaker. The speakers in this study were readers in the first place. The reading aloud task forced them to read the text in advance very carefully for themselves. They were encouraged to pay attention to the content and structure of the text. From the perspective of conceptualization, the prosodic marking is the reflection of the speakers' mental representation of the text. A reader's text structure becomes apparent in its prosodic marking. The prosodic realization of text characteristics like global and local structure, and rhetorical relations, gives support to the psychological reality and relevance of them.

A second explanation for the systematic variation of the prosody in relation to text characteristics, is that it is communicatively relevant. According to this view, speakers 'know' that listeners' understanding of the text is facilitated by varying pause durations and F0-maxima. In the same way speakers 'know' that fast readings of the satellites do not prevent

the listener from a clear understanding of the whole text and that slow reading of the nucleus is important for communicating the content of the text. Because the speech material was only acoustically analysed and not perceptually, it remains a question whether or not listeners perceive the prosodic variation indicating text structure.

Wennerstrom (2001) incorporates the conceptualization and communication perspective by saying that 'prosody adds an important element of cohesion to a text to help listeners derive a coherent interpretation' (p. 79). She gives a number of interesting examples showing that prosodic features reveal speaker's assumptions about what information is accessible in the listener's mental representation of the text and how utterances are to be integrated within that representation.

Concerning practical application of the current findings, it remains to be investigated whether text-to-speech systems can benefit from the results. We do not pretend to know the exact prosodic parameters for implementation in text-to-speech systems, but the results may have identified some characteristics of text that could be relevant for the naturalness of these systems. In further research, text-to-speech systems with and without text prosodic parameters should be compared, to answer the question whether or not the implementation of text characteristics is perceived as an improvement. Values for the prosodic realization of text characteristics as emerging from the current study are included in the Appendix.

The questions raised in this study were about how global and local structure, and the rhetorical relations between sentences are reflected in prosody by human speakers. The results show that the prosodic realization of text characteristics has not only to do with the structural position of sentences in texts, but also with the content and the content relations between sentences and text spans. Although this does not apply to all speakers, variation in pause duration and F0-maximum is a robust means for speakers to express these text characteristics.

Acknowledgements

The research reported in this paper was funded by SOBU (Cooperation of Brabant Universities). Our thanks go to Leo Vogten and Jan Roelof de Pijper for making available their programming software.

References

Bateman, J. A., & Rondhuis, K. J. (1997). Coherence relations: Towards a general specification. *Discourse Processes, 24*, 3-49.

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

Cooper, W., & Paccia-Cooper, J. (1980). Syntax and Speech. Harvard University Press.

Donzel, M. van (1999). *Prosodic aspects of information structure in discourse*. Dissertation. University of Amsterdam.

Hermes, D.J. (1988). Measurement of pitch by subharmonic summation, *Journal of the Acoustical Society of America*, *83*, 257-264.

Hirschberg, J., & Grosz, B. (1992). Intonational features of local and global discourse structure*, Proceedings of the Speech and Natural Language Workshop* (pp.441-446). New York: Harriman.

Hirschberg, J., & Nakatani, C. (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the 34th annual meeting Association for Computational Linguistics*, Santa Cruz, 286-293.

Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 195-203). Berlin: Springer.

Mann, B., & 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. 133-145). Amsterdam: Benjamins.

Ouden, H. den (2004). *Prosodic realizations of text structure*. Dissertation. University of Tilburg.

Ouden, H. den, & Terken, J. (2001). Measuring pitch range. *Proceedings of the 7th European Conference on speech communication and technology*, Aalborg, Danmark, 91-94.

Ouden, H. den, Wijk, C. van, Terken, J., & Noordman, L. (1998). Reliability of text structure annotation. *IPO Annual Progress Report, 33*, 129-138.

Potter, A. (2007). *An investigation of interactional coherence in asynchronous learning environments.* Dissertation. Nova Southeastern University.

Prince, E. (1981). Toward a taxonomy of Given-New Information. In: Cole, P. (Ed.), *Radical Pragmatics*, Academis Press, New York, 223-255.

Sanders, T., & Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *29*, 37-60.

Sanders, T., Spooren, W., & Noordman, L. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*, 1-35.

Schilperoord, J., 1996. *It's about time*. Dissertation. Utrecht University.

Silverman, (1987). *The structure and processing of fundamental frequency contours*. Dissertation. Cambridge, UK: Cambridge University.

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

Sweetser, E. (1990). *From etymology to pragmatics*. Cambridge: Cambridge University Press.

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

Thorndyke, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*, pp. 77-110.

Thorsen, N. Gronnum (1985). Intonation and text in standard Danish. *Journal of the Acoustical Society of America*, *80*, 1205-1216.

Wennerstrom, A. (2001). *The music of everyday speech*. Oxford: Oxford University Press.

Wichmann, A. (2000). *Intonation in text and discourse: beginnings, middles and ends*. London: Longman.

Yule, G. (1980). Speakers' topics and major paratones. *Lingua*, *52*, 33-47.

Appendix

Estimates of raw scores of male and female voices for adjusting pause duration, F0-maximum and articulation rate to global structure, rhetorical relations and local structure in generated speech, for segments in initial and non-initial position (only estimates for significant results are presented)

			preceding pause (in milliseconds)		F0-maximum (in hertz)		articulation rate (in phonemes/sec)	
			male	female	male	female	male	female
Levels in global structure								
initial	5	highest	1211	1059	179	293		
	4		1181	1033	177	291		
	3		1130	988	175	287		
	2		938	820	169	281		
	1	lowest	910	795	166	278		
non-initial	5	highest	-	-	-			
	4		525	457	202	320		
	3		-	-	-	-		
	2		406	352	148	257		
	1	lowest	393	341	141	249		
Rhetorical relations								
Basic operation								
initial	causal		921	805			14.8	14.9
	non-causal		1040	909			14.6	14.7
non-initial	causal		355	307			15.8	15.2
	non-causal		427	371			14.5	14.5
Source								
initial	semantic							
non-initial	pragmatic							
Local structure								
initial	nucleus						14.6	14.7
	satellite						14.7	14.7
non-initial	nucleus						14.3	14.3
	satellite						15.0	15.1