

Coherence structure and lexical cohesion in expository and persuasive texts

Ildikó Berzlánovich, Markus Egg, Gisela Redeker
University of Groningen

1 Introduction

The interplay of coherence and cohesion is an intensely studied, but still not fully understood issue in discourse organization. Both are known to vary with genre (see, e.g., Taboada 2004, Hoey 1991). In expository prose, for instance, the coherence structure is strongly determined by content-oriented relations, while instructive, argumentative, or persuasive texts are structured according to the writer's discursive strategy, involving relations between speech acts and thus what Grosz and Sidner's (1986) call the intentional structure of the discourse. This difference corresponds to the distinction between semantic (or 'ideational') and pragmatic coherence relations (Redeker 1990, 2000; Sanders 1997). Similarly, expository texts have shorter cohesive chains than for instance narratives (Goutsos 1997) and generally can be expected to have more lexical cohesive (thematic) links than other text types.

We take this variation across genres as our starting point for investigating the interaction between coherence and cohesion. In particular, we investigate the hypothesis that lexical cohesion is closely aligned with coherence structure in thematically organized (expository) texts, but less so in texts with a predominantly intentional structure (e.g., persuasive texts). The validity of this hunch has been confirmed w.r.t. local relations in a small pilot study comparing texts from the *Wall Street Journal* (WSJ) corpus (Carlson *et al* 2003) to a sample of fundraising letters (Redeker & Egg 2007). The number of cohesive links between elementary (clause-level) discourse units was greater for units that were directly connected in the discourse structure than for units that had no direct coherence link, and this difference was much larger for the expository (WSJ) texts than for the fundraising letters.

In the present study, we investigate the alignment hypothesis at the *global* levels of discourse structure. We are comparing two well-defined and clearly distinct genres: entries from online encyclopedias on astronomy and fundraising letters. Before presenting our empirical research, we will first discuss our choice of genres and then sketch our view on the role of genre, coherence and cohesion in the organization of discourse.

1.1 Expository and persuasive genres

There is general agreement in the literature that expository texts aiming to enlarge the knowledge of the reader have a strongly informational character. However, it is also a widely observed fact that expository texts are not always easily distinguished from other text types (e.g. Virtanen 1992, Goutsos 1997). Mosenthal (1985) places the different types of expository texts along a continuum, with descriptive texts (e.g. records and reports) at one end, and argumentative texts (e.g. theoreticals) at the other. Similarly, Biber (1989) concludes that there is no single expository text type. He distinguishes three subtypes of exposition: *Scientific exposition* is the most informational, elaborated in reference, technical and abstract in style and content; *learned exposition* is less abstract and less technical; finally, *general narrative exposition* bears both narrative and expository forms and contains informational elaboration. With our choice of encyclopedia entries, we focus on strongly information-oriented learned expository texts which primarily present facts.

Persuasive discourse can be defined as "language that attempts to change or reconfirm the opinions and behaviours of an audience" (Halmari & Virtanen, 2005: 229). However, persuasion is not always explicit throughout a text: many genres that fall under the above definition routinely contain text parts with the characteristics of informative texts (reviews, for instance, usually contain descriptive or narrative passages along with the reviewer's opinions and arguments). Vice versa, there is the recent tendency of the "promotionalization" of informational genres, leading to more extensive genre-mixing or hybridization of genres (Bhatia 2005). We therefore opted for the strongly persuasive genre of fundraising letters: promotional discourse directed at a particular (though not necessarily specified) addressee. The strong persuasive force of fundraising letters (asking for money and aiming to convince the reader to financially support the promoted organization) has been shown in previous studies (e.g., Abelen, Redeker & Thompson 1993, Bhatia 1998, Upton 2002).

1.2 Levels of discourse organization: genre, coherence and cohesion

The concept of *genre* refers to the pragmatic knowledge shared by the members of a discourse community about a more or less conventionalized class of communicative events with common communicative purposes (cf. e.g., Swales 1990). This shared knowledge concerns standard default elements in texts of a particular genre, but also expectations about, e.g., subject matter and stylistic choices. With respect to discourse organization, we focus only on the former, i.e., the implicitly known or explicitly taught *genre-specific schematic structure* containing genre-specific elements like header, lead and body text in newspaper articles (van Dijk 1988).

The genre-specific schematic structure of fundraising letters has been investigated i.a. by Bhatia (1998) and Upton (2002). In a corpus study on the genre of direct mail letters, Upton (2002) identified seven moves labeled *get attention*, *introduce the cause and/or establish credentials of organization*, *solicit response*, *offer incentive*, *reference insert*, *express gratitude*, *conclude with pleasantries*, all organised around the indispensable central move *solicit response*. Comparable schematic move structures should be identifiable for other genres by systematic corpus analysis.

In contrast to the generalized default structure with genre-specific elements, the *coherence structure* describes an individual text in terms of generally applicable coherence relations. These coherence relations need not be explicitly signalled on the linguistic surface of the discourse, but can follow from world knowledge or pragmatic knowledge (for discussions see, e.g., Redeker 1990, Taboada & Mann 2006).

Lexical cohesive relations, finally, are by definition associated with surface linguistic elements in the discourse. Three subareas of cohesion can be distinguished: relational cohesion (lexical or phrasal elements that signal coherence relations), referential cohesion (anaphoric chains, spatial/temporal chaining), and lexical cohesion arising from structural or collocational linkage between content words. In this paper, we will focus only on the latter.

2 Method

2.1 Corpus

The texts we used for this pilot study have been selected from a larger corpus of Dutch texts from various genres for which we are currently developing an RST-based discourse-structure annotation. The encyclopedia entries are from online encyclopedias on astronomy (<http://www.astro.uva.nl/encyclopedie/>; <http://www.astronomie.nl>); the fundraising letters are a convenience sample from direct-mail campaigns of philanthropic organizations. The four texts in this pilot study (see Table 2) range in length between 217 and 289 words and between 23 and 31 elementary discourse units (EDUs).

Table 1: Texts in this pilot study

Genres	Texts *	Words	EDUs
Encyclopedia entries	EE01: De Zon	289	31
	EE02: Mercurius	288	27
Fundraising letters	FL01: Dierenbescherming	223	24
	FL02: CliniClowns	217	23

* Note: EE = encyclopedia entry, FL = fundraising letter

2.2 Genre-specific global structure

For analyzing the genre-specific global structure of the fundraising letters, we adopted the labels introduced by Upton (2002), which he adapted from Bhatia (1998) (see section 1.2). Upton found that certain moves (i.e., *introduce the cause and/or establish credentials of organization* and *solicit response*) are obligatory (and thus essential for a text to qualify as a direct mail letter), while others are optional. We found the obligatory moves in all the fundraising texts we have looked at so far.

For the texts from the astronomy encyclopedias, no blueprint for the genre-specific move structure was available. We therefore derived labels for the genre-specific moves inductively. There are four standard moves in the encyclopedia entries we have looked at so far: *name the object*, *define the object* (e.g., star or planet), *describe in general* (i.e., describe the stellar object as a whole; e.g., its size, age, or category), and *describe details* (i.e., describe processes and parts of the object; e.g., surface, past/future development, or discovery). The first two of these are obligatory and occur exactly once in each encyclopedia entry, while the two descriptive moves need not both be realized and may each occur in more than one instantiation.

2.3 Coherence analysis

The coherence structure of the texts is described in terms of Rhetorical Structure Theory (RST) (Mann & Thompson 1988; <http://www.sfu.ca/rst>). The smallest units (elementary discourse units or EDUs) in this analysis are clauses, except for intraclausal constituents and restrictive relative clauses, or independent fragments that function as complete utterances. The functional relations between the propositions in a text are defined in terms of semantic constraints on the constituent units and the analyst's plausibility judgements about the writer's purpose in producing those units. The unit that is most central to the writer's purposes is called the *nucleus*; less central supporting or expanding units are called *satellites*. The relations apply recursively to yield a hierarchical structure. RST trees combine subject matter relations (relating states of affairs) and presentational relations (relating illocutions or text parts) in a single representation. This conflation of content structure and intentional structure allows the analyst to choose the contextually most salient relation, that is, the one that maximizes the relevance of a unit to the local or global discourse purpose at hand.

2.4 Cohesion analysis

The analysis of lexical cohesion follows Halliday and Hasan (1976), Halliday and Matthiessen (2004), and Morris and Hirst (1991) (for recent work on genre-specific variation in cohesion and an overview of approaches, see Tanskanen 2006). We distinguish the following categories of lexical cohesive relations:

- Repetition
- Systematic semantic relations: hyponymy (including hyperonymy and co-hyponymy), meronymy (including holonymy and co-meronymy), synonymy, antonymy
- Collocation.

For each content word in a text, we identify its lexical links to preceding lexical items (lemmas), ignoring any links among words inside the same (clause-sized) elementary discourse unit (EDU). This means that if a lexical item is linked to more than one preceding item, all of those relations are registered as cohesive links. Similarly, if a lexical item enters into cohesive relations with more than one item occurring in succeeding EDUs, all those links are counted. As we are interested in the contribution of lexical (i.e., semantic, not referential) cohesive relations, we exclude pronouns and do not follow referential chains (sequences of co-referential items) through the texts. Example (1), translated from Dutch, illustrates (content words are bolded):

- (1) EDU15[At first **sight**, the **surface**_{meronym(MercuryEDU13),coll(surfacetemperatureEDU9)} of **Mercury**_{repetition(MercuryEDU13),meronym(solarsystemEDU12),co-meronym(earthlyEDU11),co-meronym(VenusEDU9),co-meronym(sunEDU7),hyponym(planetEDU4)} **looks** very much like that of the **moon**_{co-meronym(MercuryEDU13),meronym(solarsystemEDU12),co-meronym(earthlyEDU11),co-meronym(VenusEDU9),co-meronym(sunEDU7),hyponym(planetEDU4)}.]

2.5 Centrality

To assess the alignment between coherence structure and lexical cohesion, we use the segments realizing genre-specific moves in the top-level of the RST tree and calculate their centrality in terms of the number of lexical cohesive links to other move segments. We also use the lexical links described in 2.4 to determine the lexical connectedness of each individual EDU. In either case, we count the total number of lexical links (looking backward and forward, but not beyond the nearest occurrence of a related lemma). For the text segments that realize genre-specific moves, the lexical cohesive density is calculated by dividing the number of the external cohesive links by the number of the EDUs for each segment realizing a genre-specific move. For individual EDUs, the more links it has, the more lexically central it is in the text.

3 Results

3.1 Genre and coherence

To identify the genre-specific moves in each text, we scrutinized the top-level segments in the RST trees and found that the mapping (visualized for EE01 and FL01 in Figure 2) was quite straightforward. We did deviate from standard versions of move analysis in that we considered adjacent segments realizing the same move as two instantiations of that move, whereas standard move analysis would identify the combined stretch of text as a single instantiation of that move (cf. the two *getting attention* moves in FL01 in Figure 2 and the multiple *describe* moves in the EE texts, shown in Table 5 below).

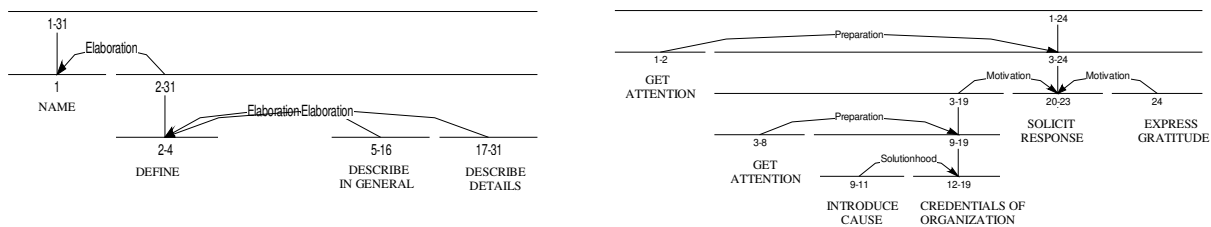


Figure 2: Generic structure and top-level RST structure for EE01 (left) and FL01 (right)

In the encyclopedia entries, the title (name of the object) and the definition are the most important units and occur at the highest level of the RST trees. The title is elaborated by the body text, consisting of a nuclear (central) segment presenting the *define* move and at least two Elaboration satellites, one with general information on the object and one with specific information about details. For the fundraising letters, the central nucleus in the RST structure maps onto the move type *solicit response*; the optional *getting attention* is realized by a Preparation satellite, *introduce cause* and the optional *express gratitude* appear as Motivation satellites. When both, *introduce the cause* and *establish credentials of organization*, are realized, they often occur in a Solutionhood relation (as in FL01, see Figure 2). Only one of these moves is needed, however (Upton 2002). Occurring on its own, *introduce cause* relates to the central *solicit response* move as a Motivation satellite and *establish credentials of organization* as a Justify satellite.

To further confirm the informative versus persuasive character of our texts, we examined the complete RST trees and compared the occurrence of subject-matter relations and presentational relations (see 2.3 above). The results are very clear (see Table 2): The encyclopedia texts contain almost exclusively subject-matter relations, reflecting the expository nature of those texts; in the fundraising letters, by contrast, presentational relations account for at least half of all relations, reflecting their strongly persuasive purpose.

Table 2: Coherence relations in the encyclopedia entries (EE) and fundraising letters (FL)

Coherence relations	EE01	EE02	FL01	FL02
Subject-matter (incl. multinuclear) relations	28	26	10	10
Presentational relations	0	1	12	10

The proportion of presentational relations in the fundraising letters is even higher in the top-level structure: In FL01, four of the five relations between the genre-specific moves are presentational relations (see Figure 2); in FL02, this holds for seven of the eight top-level relations. In both texts, the only subject-matter relation between moves is Solutionhood.¹ In the encyclopedia entries, by contrast, none of the top-level relations (of a total of three in EE01 and five in EE02) are presentational. This means that the coherence relations that are closest to the texts' global purpose are exclusively content-related in the encyclopedia texts and predominantly intention-related in the fundraising letters.

Table 3 below shows that the remaining presentational relations in the fundraising letters do not occur with equal frequencies in all the moves. In particular, the move *establish credentials of organization* is realized exclusively with subject-matter relations, while the central move *solicit response* contains only presentational relations in both letters. Still, the presentational relations are sufficiently distributed (within and across the moves) to treat these letters as persuasive throughout.

¹ Note, incidentally, that Abelen *et al* (1993) have argued that the *Solutionhood* relation should (*pace* Mann & Thompson 1988) better be considered a presentational relation in persuasive texts, as its most plausible function there is not simply informative. In fundraising letters, presenting the activities of the philanthropic organization as a solution to a problem evaluates those activities and thus strengthens that move's persuasive force. Following that line of argument, all the top-level relations in the two fundraising letters in this study would be considered presentational.

Table 3: Genre-specific moves and coherence relations in the fundraising letters

Genre-specific moves	EDUs	RST Relations	
		Subject-matter	Presentational
<i>FL01: Dierenbescherming</i>			
Get attention (1)	1-2		1
Get attention (2)	3-8	2	3
Introduce cause	9-11	1	1
Credentials of organization	12-19	6	
Solicit response	20-23		3
Express gratitude	24		
TOTAL		9	8
<i>FL02: CliniClowns</i>			
Offer incentives	1		
Get attention	2-7	3	
Credentials of organization	8-11	3	
Introduce cause	12-16	2	2
Solicit response	17		
Reference insert	18		
Solicit response	19-20		1
Express gratitude	21		
Offer incentives	22-23	1	
TOTAL		9	3

3.2 Genre and cohesion

Table 4 shows that the encyclopedia entries are much more densely populated by lexical cohesion relations than the fundraising letters. This difference is too large to be explained by the fact that the fundraising letters are slightly shorter than the encyclopedia entries (cf. the totals rows in Table 5 below, which show overall density measures of 4.6 and 6.4 for the encyclopedia entries, but only 3.3 and 3.6 for the fundraising texts). Note further in Table 4 that the genres differ in the type of lexical linkage. In both encyclopedia texts, about three quarters of the links are based on systematic semantic relations (mainly hyponymies and meronymies), while 29 % the links in FL01 and 45 % in FL02 are based on nonsystematic semantic relations (e.g., *star – light*, *hospital – doctor*), which we are subsuming under the label of collocation.

Table 4: Lexical cohesive links

Type of cohesion	EE01		EE02		FL01		FL02	
Repetition	30	21 %	35	20 %	23	29 %	39	48 %
Systematic lexical links	104	73 %	131	76 %	33	42 %	6	7 %
Collocation	8	6 %	6	4 %	23	29 %	37	45 %
TOTAL	142	100 %	172	100 %	79	100 %	82	100 %

3.3 Alignment of coherence and cohesion

The left-hand columns of Table 5 below show the results of mapping the top-level units in the RST analyses onto the genre-specific moves. In the RST trees, the moves *name* and *define* are the most central parts of the encyclopedia entries, and the text part realizing the *solicit response* move is the most central in the fundraising letters (see 3.1).

In the remainder of Table 5, we use two measures for the lexical centrality of each genre-specific top-level segment: the total number of external lexical links and, to take the varying size of the segments into account, the average number of external links per EDU. Total *external linkage* is highest (up to 62) for the general information segments in the encyclopedia texts (in EE01 even the rather short definition segment has 45 external links). The definition segment scores highest on the *cohesion density (per EDU)* in both EE texts, closely followed by the first *describe in general* move. In the fundraising letters, the two *get attention* segments and the *credentials of the organization* show higher total external linkage (20-31 links in FL01 and 26-27 in FL02) than the discourse-structurally most central segment realizing the *solicit response* moves (11 links in FL01 and 7+3 links in FL02). The correction for segment lengths yields an even more extreme picture of the non-alignment of discourse-centrality and lexical centrality in the fundraising letters: for FL01, the EDUs in the first *get attention* segment average 11.0 links to elements in other segments, while the EDUs

in *solicit response* score very low (2.8 external links per EDU). In FL02 the first *solicit response* move has a moderate level of lexical cohesive density (7 links comparable to e.g., *credentials of organization* with 6.8 or *offer incentives* with 6.5 links); its second instance, however, has an extremely low density score (1.5 external lexical cohesion links per EDU).

Table 5: Genre-specific moves and lexical cohesion

Genre-specific moves	EDUs	Number of EDUs	Cohesion links		Cohesion density (external links per EDU)
			internal	external*	
<i>EE01: De zon</i>					
Name	1	1	0	2	2.0
Define	2-4	3	26	45	15.0
Describe in general (1)	5-8	4	12	49	12.3
Describe in general (2)	9-16	8	3	19	2.4
Describe details (1)	17-22	6	11	29	4.8
Describe details (2)	23-31	9	7	20	2.2
TOTAL		31		142	4.6
<i>EE02: Mercurius</i>					
Name	1	1	0	3	3.0
Define	2-3	2	3	25	12.5
Describe in general (1)	4-6	3	6	35	11.7
Describe in general (2)	7-13	7	34	62	8.9
Describe details (1)	14-21	8	25	35	4.4
Describe details (2)	22-25	4	5	21	5.3
Describe details (3)	26-27	2	0	13	6.5
TOTAL		27		172	6.4
<i>FL01: Dierenbescherming</i>					
Get attention (1)	1-2	2	3	22	11.0
Get attention (2)	3-8	6	11	31	5.2
Introduce cause	9-11	3	1	11	3.7
Credentials of organization	12-19	8	11	20	2.5
Solicit response	20-23	4	3	11	2.8
Express gratitude	24	1	0	5	5.0
TOTAL		24		79	3.3
<i>FL02: CliniClowns</i>					
Offer incentives	1	1	0	14	14.0
Get attention	2-7	6	6	26	4.3
Credentials of organization	8-11	4	7	27	6.8
Introduce cause	12-16	5	3	18	3.6
Solicit response	17	1	0	7	7
Reference insert	18	1	0	12	12.0
Solicit response	19-20	2	0	3	1.5
Express gratitude	21	1	0	12	12
Offer incentives	22-23	2	0	13	6.5
TOTAL		23		82	3.6

* Note: The sum of the cohesion links exceeds the total because each external link (between two genre-specific moves) occurs twice in the table

Consider, finally, Figure 4 below, which shows the number of the lexical cohesive links in which each EDU is involved. For the encyclopedia entries, the profiles show sharper and higher peaks than for the fundraising letters. EE01 also shows a very clear dominance of the initial moves (*define* and *describe in general*), i.e. very strong alignment of lexical centrality with centrality in the coherence structure. The secondary peaks in the EDUs 10, 17, and 22 are due to a combination of local lexical links with long distance meronymic and hyponymic links to lexical items in the EDUs 3 to 6. The picture is a bit less clear for EE02, where the ‘define’ and ‘describe in general’ moves contain EDUs with few lexical links and the first of the three ‘describe details’ moves (EDUs 14-21) is heavily related to lexical material in the definition and in the general description.

FL01 and FL02 also show a tendency for EDUs early in the text to have more lexical links than later ones, but the peaks are much lower. Significantly, the all-important moves *solicit response* (EDUs 20-23 in

FL01 and EDUs 17 and 19-20 in FL02), appear as rather minor peaks in the centrality profiles, with a height of 12 in FL01 and 7 and 3 respectively for the two instances in FL02, compared to a maximum height of 22 (for EDU 5 in FL01 and also for EDU 8 in FL02). The alignment between centrality in the discourse structure and lexical centrality is thus much less pronounced here than in the encyclopedia texts.

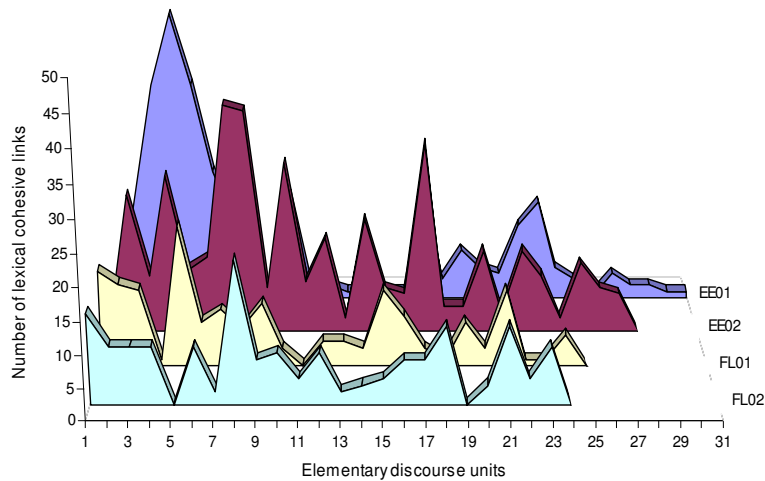


Figure 4: Lexical centrality of the elementary discourse units in the encyclopedia entries (EE) and fundraising letters (FL)

4 Conclusions

We take these results as encouraging evidence for our hypothesis that lexical cohesion is more closely aligned with coherence structure in information-oriented (thematically structured) texts than in reader-oriented (more intentionally structured) texts. We will add more texts to this comparison, both prototypical genres (e.g., advertisements) and so-called mixed genres built up of more than one text type (e.g., reviews). A more fine-grained statistical analysis will take into account the number of content words per EDU.

A major methodological challenge is the refinement of the centrality measures. For the centrality in the coherence structure, we have used the (loosely defined) top-level of the RST tree (comprising the upper two to four levels, depending on the mapping onto the genre-specific moves). But each level of the RST tree typically contains nuclei and satellites, with the former being more central than the latter. When looking at the hierarchical coherence structure of the whole text, the most central segment is the nucleus at the highest level in the hierarchy and the most central EDU is found by descending along the nuclei inside that segment. That procedure might allow the assignment of a centrality score to each EDU in the text. For centrality w.r.t. lexical cohesion, type and strength of the relations might be taken into account. The scope of the cohesion relations could be extended beyond the (one) preceding or succeeding related lemma to yield lexical chains, which have been used as measures of centrality by Hoey (1991) and Tanskanen (2006).

A still more challenging extension will be the attempt to directly compare complete RST trees to tree structures based on lexical cohesion density (for a review of algorithms for comparing labeled trees, see Bille 2005). Our earlier pilot study with local relations (Redeker & Egg 2007) has revealed some of the complications such an attempt will have to deal with.

References

- Abelen, E., Redeker, G. & Thompson, S.A. (1993). The rhetorical structure of US-American and Dutch fund-raising letters. *Text* 13 (3): 323-350.
- Bhatia, V.K. (1998). Generic patterns in fundraising discourse. *New directions for philanthropic fundraising* 22: 95-110.
- Bhatia, V.K. (2005). Generic patterns in promotional discourse. In: H. Halmari & T. Virtanen (Eds.), 213-225
- Biber, D. (1989). A typology of English texts. *Linguistics* 27: 3-43.
- Bille, P. (2005). A survey on tree edit distance and related problems. *Theoretical Computer Science* 337: 217-239.

- Carlson, L., Marcu, D. & Okurowski, M.E. (2003). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In J. van Kuppevelt & R. Smith (Eds.), *Current directions in discourse and dialogue*. Dordrecht: Kluwer, 85-112.
- Goutsos, D. (1997). *Modeling discourse topic: Sequential relations and strategies in expository text*. Norwood, NJ: Ablex.
- Grosz, B.J. & Sidner, C.L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3): 175-204.
- Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M.A.K. & Matthiessen, Ch.M.I.M. (2004). *An introduction to functional grammar* (3rd ed.). London: Hodder Arnold.
- Halmari, H. & Virtanen, T. (Eds.) (2005). *Persuasion across genres*. Amsterdam: Benjamins.
- Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.
- Mann, W.C. & Thompson, S.A. (1988). Rhetorical structure theory: toward a functional theory of text organization. *Text* 8(3): 243-281.
- Morris, J. & Hirst, G. (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics* 17: 21-48.
- Mosenthal, P.B. (1985). Defining the expository discourse continuum. Towards a taxonomy of expository text types. *Poetics* 14: 387-414.
- Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14: 367-381.
- Redeker, G. (2000). Coherence and structure in text and discourse. In W. Black & H. Bunt (eds.), *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics*. Amsterdam: Benjamins, 233-263.
- Redeker, G. & Egg, M. (2007). On the interaction of relational coherence and lexical cohesion in expository and persuasive text genres. 10th International Pragmatics Conference, Göteborg, 8-13 July 2007 (abstract and slides available at <http://www.let.rug.nl/~redeker/talks.html>).
- Sanders, T. (1997). Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes* 24: 119-147.
- Swales, J. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.
- Taboada, M.T. (2004). *Building coherence and cohesion*. Amsterdam: Benjamins.
- Taboada, M. & Mann, W.C. (2006). Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies* 8: 423-459.
- Tanskanen, S.-K. (2006). *Collaborating towards discourse: lexical cohesion in English discourse*. Amsterdam: Benjamins.
- Upton, T.A. (2002). Understanding direct mail letters as a genre. *International Journal of Corpus Linguistics* 7(1): 65-85.
- van Dijk, T.A. (1988). *News as discourse*. Hillsdale, NJ: Erlbaum.
- Virtanen, T. (1992). Issues of text typology: Narrative – a ‘basic’ type of text? *Text* 12(2): 293-310.