

Cohesion analysis and information flow

Renkema, Jan

Published in:
Linguistics in the Netherlands

Publication date:
1996

[Link to publication](#)

Citation for published version (APA):

Renkema, J. (1996). Cohesion analysis and information flow: The case of 'Because' versus 'because'. In C. Cremers, & M. den Dikken (Eds.), *Linguistics in the Netherlands* (pp. 233-244). Amsterdam/Philadelphia: Benjamins.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Cohesion analysis and information flow: the case of 'Because' versus 'because'

Jan Renkema

0. Introduction

In research on textual cohesion, the position of information units is usually related to rather vague contextual concepts relating to information flow, such as: topic continuity, foreground/background information and given/new information. See, for example, Van Dijk (1977), Givón (1983), Grosz and Sidner (1986), and Ford (1993). In this paper an analytical framework for cohesion is presented which makes it possible to do corpus-linguistic research on the basis of operational definitions of anaphoric and cataphoric information linking, topicality and background information. This analytical framework is then used to test hypotheses on the position of the subordinate clause. This paper is restricted to the study of the position of *omdat* (because) clauses, before or after the main clause.¹

1. Because/because-clauses

A because-clause may occur before or after the main clause. In the rest of this paper, these clauses will be designated Because-clauses and because-clauses respectively. In text fragment (1a) below, both variants occur. In (1b) the variants are given in the other position. Which fragment is the one which actually occurred?

- (1) a Cyrano toert op dit moment met groot succes langs de Nederlandse schouwburgen. Komend voorjaar zal in New York met de audities voor Cyrano worden begonnen.

Waar de musical gespeeld wordt is nog niet duidelijk, **omdat** theaters op Broadway pas een half jaar van te voren gereserveerd kunnen worden. Van den Ende heeft zijn oog laten vallen op het Mynskofftheater, dat 1600 stoelen telt. **Omdat** het podium daar aanmerkelijk groter is dan de gemiddelde Nederlandse schouwburg, zal de mise-en-scène worden aangepast.

¹ I would like to thank Brigitte Cleutjens for testing the analysis framework and for the coding. I am also grateful to Leo Noordman, Wilbert Spooren, and Carel van Wijk for their stimulating and critical comments.

'Cyrano is at this moment touring Dutch theaters with great success. Next spring auditions for Cyrano will begin in New York.

The location where the musical will be performed is still unclear, **because** theaters on Broadway may not be reserved more than half a year in advance. Van den Ende has expressed interest in the Mynskoff theater, which has 1600 seats. **Because** that stage is considerably larger than that of the average Dutch theater, the staging will be adapted.'

- b Cyrano toert op dit moment met groot succes langs de Nederlandse schouwburgen. Komend voorjaar zal in New York met de audities voor Cyrano worden begonnen.

Omdat theaters op Broadway pas een half jaar van te voren gereserveerd kunnen worden, is nog niet duidelijk waar de musical gespeeld wordt. Van den Ende heeft zijn oog laten vallen op het Mynskofftheater, dat 1600 stoelen telt. De mise-en-scène zal worden aangepast **omdat** het podium daar aanmerkelijk groter is dan de gemiddelde Nederlandse schouwburg.

'Cyrano is at this moment touring Dutch theaters with great success. Next spring auditions for Cyrano will begin in New York.

Because theaters on Broadway may not be reserved more than half a year in advance, the location where the musical will be performed is still unclear. Van den Ende has expressed interest in the Mynskoff theater, which has 1600 seats. The staging will be adapted, **because** that stage is considerably larger than that of the average Dutch theater.'

Fragment (1a) is the one which actually occurred. The alterations in (1b) do indeed seem to weaken the cohesion of the piece. The new paragraph should start a new topic: the location where the musical will be performed. In (1b), where *Broadway* is placed towards the front of the sentence and the assertion is therefore closely linked to the previous paragraph with *New York*, having a paragraph boundary seems to make less sense. If *Broadway* occurs nearer the end of the sentence as in (1a), the link with *Mynskoff theater* is also closer. In addition, in the last sentence of the fragment the deictic element *daar*, or 'that' in the English, is closer to its antecedent in (1a).

However, this analysis is rather subjective and ad hoc. The question is whether corpus-linguistic research can be used to demonstrate contextual differences for clauses before and after the main clause with the use of objective criteria. In order to answer this question, a corpus was put together of *omdat*-clauses occurring in newspaper articles. From a corpus of daily newspapers the first hundred passages were taken in which *omdat* 'because' occurred.² This

² Chafe (1984) distinguishes between free clauses and bound clauses. Bound clauses differ from free clauses in that they are not separate intonation units and, in written language, are not separated by a comma from the main clause. In this study of written language this distinction is not held to because the presence or absence of a comma is not a reliable indicator.

corpus contained 88 because-clauses and 12 Because-clauses. This ratio is in keeping with the 90:10 ratio found in *Woordfrequenties* (see Uit den Boogaart 1975) for texts comprising a total of 600,000 words. The selected clauses are independently interpretable, causal sentences within a context with a singular argument and without an extra connective. This means that sentences such as the following are not included in this study:

- (2) Dat is nodig omdat anders de werkloosheid fors oploopt.
'That is necessary because otherwise unemployment will increase dramatically.'
- (3) Jan Engelsman was een groot dichter, omdat hij een groot dichter was.
'Jan Engelsman was a great poet because he was a great poet.'
- (4) Wel doopsgezind, omdat het woord 'gezind' me aanspreekt, het ertoe geneigd zijn.
'Baptist-minded [=Mennonite], however, because I like the word "minded," being inclined towards it.'
- (5) Zonder veel passie dulden ze elkaar omdat ze dat al jaren doen en omdat ze te laf zijn om uit elkaar te gaan.
'Without much passion they tolerate each other because they have been doing it for years and because they are too cowardly to break up.'
- (6) Omdat Ajax dus niet tot het uiterste hoefde te gaan werd het voor de meeste aanwezigen weer zo'n trieste Nederlandse voetbalmiddag.
'So because Ajax [a soccer team] did not need to exert itself, for most of those present it became another one of those dreary Dutch soccer afternoons.'

Sentence (2) is not independently interpretable because the demonstrative pronoun *dat* 'that' is dependent for its meaning on the previous sentence. The previous sentence is therefore necessary in order to analyze the context of *omdat*. In (3) the relationship is not purely causal. In (4) the main clause is not complete; (5) contains a plural argument; and sentence (6) contains an extra connective *dus* 'So'. Sentences at the beginning or end of a text were also left out of consideration.

After this selection process, 64 because-clauses and 11 Because-clauses were left. Subsequently, another 14 Because-clauses were found in other newspaper articles, so that the final corpus consisted of 64 because-clauses and 25 Because-clauses.

2. Cohesion analysis

Inspired by Halliday and Hasan (1976) and Givón (1995), I have developed a system of analysis in which a determination is made for each sentence constituent whether there is a connection with the context (before or after), and if so, what kind of a connection. The context is the 'topic-passage' in which the clause occurs. This may be a complete text (if short), a series of paragraphs on a single topic or a single paragraph.³ Three types of connections can be distinguished:

1. repetition;
2. -nymic relations: *syno-*, *anto-*, *hypo-*, *mero-* (part of a whole);
3. pronominal relations.⁴

The analytical framework can most easily be described using examples from the corpus. Below we discuss first a Because-clause and then a because-clause in context.

- (7) Vice-premier en minister van financiën Kok zei zaterdag dat behoud van koopkracht voor iedereen volgend jaar 'wenselijk en mogelijk' is. **Omdat** het kabinet de burgers lastenverlichting geeft en de inflatie laag is, blijft de koopkracht intact, terwijl de lonen nauwelijks stijgen. Volgens minister De Vries (sociale zaken en werkgelegenheid) krijgt een modale werknemer er per maand 40 gulden netto bij, iemand met het minimumloon 30 gulden en mensen met een uitkering 20 gulden.
- 'Vice-premier and finance minister Kok said Saturday that it was "desirable and possible" for everyone to preserve their buying power next year. **Because** the cabinet is reducing the tax burden on citizens and inflation is low, buying power will remain intact while salaries will rise only minimally. According to minister De Vries of Social Affairs and Employment, the average wage-earner will receive 40 guilders net more, someone at minimum wage 30 guilders and people with a government benefit 20 guilders.'

³ The analysis framework does not (yet) take the distance between cohesive elements into account. A more detailed analysis would take into consideration not only distance but also the nature of the connection (see also footnote 4).

⁴ In this article the nature of the connections is not under discussion. Research into other types of subordinate clauses will need to be done in order to demonstrate whether or not there is a relationship between the type of connection and the position of the subordinate clauses; for example there might be more pronominal relations in a sentence-initial subordinate clause. For this study, it suffices to know that there is cohesion whenever one of the types of connections given here occurs.

| | | | | | | |
|----------------|--------------|--|--------------|--|-------------------------|--|
| Omdat | (het kabinet | | de burgers | | lastenverlichting | |
| Because | (the cabinet | | the citizens | | reduction of tax burden | |

| | | | | | | | |
|-------|--|-----|--|---------------|--|-------------------------|--|
| geeft | | en | | de inflatie | | laag is,) <u>blijft</u> | |
| gives | | and | | the inflation | | low is,) <u>stays</u> | |

| | | | | | | | |
|------------------|--|----------------|--|---------|--|--------------|--|
| de koopkracht | | <u>intact,</u> | | terwijl | | de lonen | |
| the buying power | | <u>intact,</u> | | while | | the salaries | |

| | | |
|------------|--|-----------|
| nauwelijks | | stijgen.] |
| scarcely | | rise.] |

omdat (m n n n n n) [s r H n S]

m = *vice-premier, minister* (both part of cabinet);

s = *behoud* 'preservation';

r = *koopkracht* 'buying power';

H = *minimumloon* 'minimum wage', *uitkering* 'government benefit' (both types of income);

S = *erbij krijgen* 'get in addition'.

The subordinate clause is between round brackets and the main clause is between square brackets. The subordinate clause consists of six constituents; conjunctions are not included in the count. A constituent that has no contextual connection is coded with an *n* (non-bound constituent).⁵ The other constituents are coded with a letter which indicates the type of connection. The first constituent has a meronymic relation with previously occurring text in *vice-premier* and a similar relation with following text in *minister*. A letter in boldface indicates that the relation is both anaphoric and cataphoric. The other constituents in the subordinate clause bear no relation with the surrounding text. The main clause has five constituents: the underlined elements are seen as a single constituent, and *terwijl* 'while' is not included because it is a conjunction. The first constituent *blijft ... intact* 'remains ... intact' is synonymous with the previously occurring *behoud* 'preservation'. The constituent *koopkracht* 'buying power' is a repetition of a previously occurring constituent. Because in each case these are anaphoric connections, the constituents are coded in lower-case letters. The next constituent, *lonen* 'salaries,' stands in forward, meronymic relation with *minimum-*

⁵ Only if there is a clear lexical connection in the text is a constituent coded otherwise than *n*. This is why, for example, the constituent *burgers* 'citizens' is coded as *n*, even though this element is inferrable from the previously occurring *koopkracht* 'buying power' or *iedereen* 'everybody'. The reason for this purely textual approach is that coding inferrable and situationally evoked elements is rather unreliable.

loon 'minimum wage' and *uitkering* 'government benefit'. A forward (cataphoric) reference is coded with a capital letter. The constituent *nauwelijks* 'scarcely' has no connection with the surrounding text. And the constituent *stijgen* is synonymous with *erbij krijgen* 'get in addition' in the subsequent text; hence the capital letter.

- (8) Volgens een woordvoerder van de VN-vredesmacht zou het bestand in Sarajevo het resultaat zijn van een politiek akkoord dat in Genève is gesloten. "Alle drie de partijen lieten weten dat zij door hun regering gemachtigd waren om te tekenen," aldus zagsman Frewer.

Bij veel VN-medewerkers bleef de scepsis over het staakt-het-vuren, omdat de Servische delegatie gisterochtend nog weigerde een minder verstrekkend akkoord te tekenen over het openen van corridors over land om Sarajevo te bevoorraden met hulpgoederen. Het staakt-het-vuren werd getekend door kolonel Siber voor het Bosnische leger, generaal Gvero voor de Serviërs en kolonel Blaskic voor de HVO (de Kroatische Defensie Raad).

'According to the spokesperson of the UN Peacekeeping Force, the truce in Sarajevo is the result of a political accord that was reached in Geneva. "All three parties declared that they were authorized by their government to sign," spokesman Frewer asserts.

Many UN workers remained skeptical about the cease-fire, because as recently as yesterday morning the Serb delegation refused to sign a less far-reaching agreement on the opening of land corridors to reach Sarajevo with relief supplies. The cease-fire was signed by Colonel Siber for the Bosnian army, General Gvero for the Serbs and Colonel Blaskic for the HVO (the Croatian Defense Council).'

| | | | | | |
|--------------------------|--|----------|--|----------------|--|
| [Bij veel VN-medewerkers | | bleef | | de scepsis | |
| [Among many UN workers | | remained | | the skepticism | |

| | | | | |
|-----------------------------|--|---------|-------------------------|--|
| over het staakt-het-vuren,] | | omdat | (de Servische delegatie | |
| about the cease-fire,] | | because | (the Serb delegation | |

| | | | | | |
|-------------------|--|-------|--|----------|--|
| gisterochtend | | nog | | weigerde | |
| yesterday morning | | still | | refused | |

| | | | |
|---------------------------------|--|------------|--|
| een minder verstrekkend akkoord | | te tekenen | |
| a less far-reaching agreement | | to sign | |

| | | | | | |
|--------------------------------|--|-----------|--|----------|--|
| over het openen van corridors | | over land | | om | |
| about the opening of corridors | | over land | | in order | |

| | | | | |
|----------|--|----------------|--|---------------------|
| Sarajevo | | te bevoorraden | | met hulpgoederen.) |
| Sarajevo | | to supply | | with aid supplies.) |

[m n n s/R] omdat (m/S n n n n r n n r n n)

m = VN-vredesmacht 'UN Peacekeeping Force';
 s = bestand 'truce';
 (R) = staakt-het-vuren 'cease-fire';
 m = alle drie de partijen 'all three parties';
 (S) = de Serviërs 'the Serbs';
 r = tekenen/getekend 'to sign/signed';
 r = Sarajevo.

The main clause contains four constituents, of which two bear an anaphoric reference; one a meronymic and one a synonymic one. The last constituent bears, in addition, a forward reference in the form of a repetition; hence the coding s/R. The subordinate clause contains eleven constituents; the conjunction *om* 'in order (to)' is not included. Two of these eleven constituents bear a forward as well as a backward reference: the m/S and the boldface r. The constituent *Sarajevo* has an anaphoric connection through repetition.

The coding thus produces configurations such as the following (made-up examples):

- (9) [p s n S a] omdat (r n n H M)
 (10) Omdat (s n) [s r n S n]

The sentences in the corpus were coded by two people independently of each other. The results were in agreement in approximately 95% of the cases. In the other 5% agreement was reached after discussion.

3. Hypotheses

The approaches in the literature sources mentioned above — and see also Thompson (1985), Tomlin (1985), Ramsay (1987), Lehmann (1988), Matthiessen and Thompson (1988) and Jou and Harris (1990) — are of two types. In the first type of approach, the issue is the position of the information: at the beginning or at the end. In the second type, the issue is the syntactic status of the information: main or subordinate clause.

a. Position. Continuing the Prague School's theme-rheme approach (in which the theme is normally the first constituent of the sentence and the rheme the rest), one can consider a text to be a series of sentences in which each subsequent

sentence takes as theme what was introduced in the previous sentence as rheme. In other words, the information in each sentence is ordered in such a way that it is placed as closely as possible to related information in previous and subsequent sentences. The ordering variation main/subordinate clause or subordinate/main clause thus relates to the necessity for backward (anaphoric) or forward (cataphoric) linking. This approach is here called the linking hypothesis.

b. Syntactic status. Elaborating on concepts derived from the more stylistically oriented traditional grammars, in which the main clause is seen as the independent and therefore more important entity and the subordinate clause as the dependent and therefore less important one, the relation between the two clauses can be seen as one of nucleus vs. satellite. In this approach, the standard order would be that of main clause followed by subordinate clause. The main clause then contains central information and the subordinate clause, positioned last, contains background information. When the order is reversed, the (postposed) main clause continues to contain the central information, but the (preposed) subordinate clause gets another function, namely that of providing an interpretive framework for the main clause. The preposed subordinate clause then contains less background information.

The approach revolving around the concept 'central information goes in the main clause' is here called the topicality hypothesis. The approach revolving around the concept 'more or less background information goes in the subordinate clause' is here called the information value hypothesis.

Below are given the three hypotheses, made operational in terms of the cohesion analysis.

(i) The linking hypothesis: The position of the clause, irrespective of the question whether it is a main or subordinate clause, is related to the information stream, such that the link between cohesive elements is as short as possible.

The concept of 'information linking' is made operational as anaphoric and cataphoric cohesion. In the coding system these are the elements with the lower-case or capital letters. According to this hypothesis, the first clause has more anaphoric information than the second one, the second clause has more cataphoric information than the first one.

(ii) The topicality hypothesis: The information in the main clause — irrespective of its position in the sentence — is more topical than the information in the subordinate clause.

The concept 'topical information' is made operational as information that occurs not only in the clause in question, but also in clauses before and after, in other words, in the coding system the constituents assigned a boldface symbol or a double symbol (with lower-case and capital letter).

(iii) *The information value hypothesis: When main clause precedes subordinate clause, the subordinate clause contains more background information than the main clause; when subordinate clause precedes main clause, the subordinate clause contains less background information than the main clause.*

The concept 'background information' is made operational as information that occurs only in the relevant clause. In the coding system these are the constituents that are assigned an *n*. This hypothesis would thus predict that in either order, the second clause would contain more *n*-elements than the first clause.

In summary: Hypothesis (i) backgrounds the concept main vs. subordinate clause; all that is important is the position. Hypothesis (ii) backgrounds position and focuses on the main clause. Hypothesis (iii) focuses on the subordinate clause and its position.⁶

In the sentences under study the anaphoric and cataphoric constituents are not frequent. Therefore each clause was considered for the presence or absence of anaphoric and cataphoric constituents. The number of linking constituents was not taken into account. Only the fact of linking counted. Because almost every clause contains a topical or non-bound constituent the testing of the topicality and information value hypotheses was based on the average number of (topical and non-bound) constituents per clause.⁷

4. Results

In Table 1 the data are presented for the anaphoric and cataphoric constituents.

Table 1 Percentage of clauses with at least one linking constituent in relation to type of clause and position

| | anaphoric | cataphoric |
|-------------------------------|-----------|------------|
| <i>because</i> -clause (N=64) | | |
| main (first position) | 63 | 17 |
| subordinate (second position) | 17 | 27 |
| <i>Because</i> -clause (N=25) | | |
| subordinate (first position) | 56 | 16 |
| main (second position) | 52 | 28 |

⁶ Other hypotheses were also tested, such as the hypothesis that a preposed subordinate clause (providing an interpretive framework) is shorter than a postposed one, or that the internal cohesion between clauses is weaker if the main clause is first than if it is second. These hypotheses were not verified in the data under study.

⁷ In this study every constituent was considered of equal value. It would, however, be plausible to suppose that the grammatical function of a constituent is relevant to its position in information flow. Normally a subject is more 'topical' than an object. The problem remains how to take possible functional differences into account within a quantitative approach. It does not seem likely, however, that any difference that might be found would materially affect the results presented here, particularly since clause-level results are what is at issue here, and in a clause any and all grammatical functions may occur.

(i) *The linking hypothesis* predicts that the first clause contains more anaphoric information than the second one, and that the second clause contains more cataphoric information than the first one. The hypothesis is confirmed for backward linking: the first-position clauses contain more often anaphoric information than the second-position ones ($z=4.53$, $p<.001$).⁸ This difference is completely due to the sentences in which the main clause is in first position (63% vs. 17%). If the subordinate clause is in first position, the scores are 56% vs. 52%. It is clear that of the four types of clauses, the second-position subordinate clause is the odd one. The other three types of clauses contain an anaphoric link in about 60% of the cases, but only 17% of the second-position subordinate clauses contain an anaphoric link. This means that a because-clause in this position is primarily locally embedded, i.e. primarily has a relation with the (preceding) main clause.

The second-position clause contains more often cataphoric information than the first-position clause, but the result just failed to reach significance ($z=1.63$, $p=.052$). The trend, however, is the same whether the main clause is first (17% vs. 27%) or the subordinate clause is first (16% vs. 28%).

In Table 2 the data are presented for the topicality and information value hypotheses.

Table 2 Number of topical and non-bound constituents in relation to type of clause and position.

| | topical | non-bound |
|-------------------------------|---------|-----------|
| <i>because</i> -clause (N=64) | | |
| main (first position) | 1.06 | 2.36 |
| subordinate (second position) | .73 | 3.14 |
| <i>Because</i> -clause (N=25) | | |
| subordinate (first position) | 1.20 | 2.16 |
| main (second position) | 1.12 | 3.52 |

(ii) *The topicality hypothesis* predicts that the main clause contains the most topical information, irrespective of its position in the sentence. The first column of Table 2 shows that a first-position clause contains more topical information than a second-position one (1.10 vs. .84; $z=1.84$, $p<.05$). This means that the main clause scores more highly than the subordinate clause when the main clause is in first position (1.06 vs. .73), but not when it is in second position (1.12 vs. 1.20). In other words, it is not so much the syntactic status (main clause or subordinate clause) which determines topicality but the position of the clause.

⁸ Given the direction of the hypotheses one-tailed levels of significance are reported in all cases. Differences between percentages are tested as differences between two binominal parameters using the normal probability distribution; differences between numbers of elements are tested with the Wilcoxon rank sum test.

(iii) *The information value hypothesis* predicts a difference in the number of non-bound elements in the preposed and postposed subordinate clause. This hypothesis is confirmed. The number of non-bound constituents in the first-position clause is less than in the second one (2.30 vs. 3.35; $z=2.79$, $p < .05$). This is true whether the order is main-subordinate (2.36 vs 3.14) or subordinate-main (2.16 vs 3.52). In other words, when the order is main-subordinate, the subordinate clause contains more non-bound constituents than the main clause. When the order is subordinate-main, the subordinate clause contains fewer non-bound constituents than the main clause.

5. Discussion

The results of the three hypotheses are remarkably compatible. The study of the linking hypothesis shows that not the syntactic status but the position of the clause is determinative for backwards and forwards linking. The study of topicality confirms this picture: at issue is not the question whether a clause is main or subordinate, but whether it is in initial position. The analysis of information value supports this conclusion: the preposed subordinate clause is the one that is most firmly embedded in the text. The choice between the order 'Main clause *omdat* subordinate clause' and '*Omdat* subordinate clause main clause' is related to a significant extent to the link with the preceding or subsequent text, and not to the syntactic status.

Further research must show whether this picture is also applicable to other types of adverbial clauses. The system presented here for the analysis of cohesion has in any case demonstrated its usefulness for the operationalizing and testing of concepts relating to information flow that up till now have remained quite vague. On the basis of these results, it now becomes relatively easy to conduct corpus-linguistic research on the distribution of information over clauses in other complex sentences and text sequences.

In conclusion I would like to propose some directions for future research. The cohesion analysis system presented here can of course be used to study other hypotheses. Two examples follow. First, in the literature (see for example Matthiessen and Thompson 1988) a preposed subordinate clause is seen as providing an interpretive framework for the following main clause and/or text. This intuition leads to a hypothesis such as the following: a preposed subordinate clause contains at least one cataphoric element. When the relation between main and subordinate clauses is also included in the cohesion analysis, it is directly possible to determine how many preposed subordinate clauses fulfill this condition. The issue would then not be a difference in the amount of cataphoric information between the preposed and postposed subordinate clauses, but the question whether one or more cataphoric elements is present, and what their scope is (only the main clause or also the subsequent text).

Second, in the literature on given-new information, no distinction is made between a new element which continues to play a role in the text ('globally new') or one that is dropped again immediately ('locally new'). This distinction is made in the system of cohesion analysis proposed here. If, for example, one wishes to test whether subordinate clauses contain more locally new information than main clauses, one can calculate the number of *n*-elements per clause based on the cohesion analysis, and then determine which clause contains the highest percentage of locally new information.

In summary, the importance of this research is its contribution to: (a) the concretizing of relatively vague notions in information flow theory; (b) a reliable and easily implementable system for the collection of data for the analysis of cohesion phenomena; (c) the development and testing of hypotheses for corpus-linguistic research.

References

- Boogaart, Uit den, P. C., ed. (1975) *Woordfrequenties in Geschreven en Gesproken Nederlands*, Oosthoek, Scheltema & Holkema, Utrecht.
- Chafe, W. L. (1984) 'How People Use Adverbial Clauses', in C. Brugman and M. Macaulay, eds., *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, 437-449.
- Dijk, T. A. van (1977) *Text and Context. Explorations in the Semantics and Pragmatics of Discourse*, Longman, London.
- Ford, C. E. (1993) *Grammar in Interaction. Adverbial Clauses in American English Conversation*, Cambridge University Press, Cambridge.
- Lehman, C. (1988) 'Towards a Typology of Clause Linkage', in J. Haiman and S. A. Thompson, eds., *Clause Combining in Grammar and Discourse*, John Benjamins, Amsterdam/Philadelphia.
- Matthiessen, C. and S. A. Thompson (1988) 'The Structure of Discourse and "Subordination"', in J. Haiman and S. A. Thompson, eds., *Clause Combining in Grammar and Discourse*, John Benjamins, Amsterdam.
- Givón, T. (1983) *Topic Continuity in Discourse*, John Benjamins, Amsterdam/Philadelphia.
- Givón, T. (1995) *Functionalism and Grammar*, John Benjamins, Amsterdam/Philadelphia.
- Grosz, B. J. and C. L. Sidner (1986) 'Attention, Intentions, and the Structure of Discourse', *Computational Linguistics*, 12, 175-204.
- Halliday, M. A. K. and R. Hasan (1976) *Cohesion in English*, Longman, London.
- Jou, J. and R. J. Harris (1990) 'Event Order versus Syntactic Structure in Recall of Adverbial Complex Sentences', *Journal of psycholinguistic research*, 19, 21-42.
- Ramsay, V. (1987) 'The Functional Distribution of Preposed "If" and "When" Clauses in Written Narrative', in R. S. Tomlin, ed., *Coherence and Grounding in Discourse*, John Benjamins, Amsterdam/Philadelphia.
- Thompson, S. A. (1985) 'Grammar and Written Discourse: Initial vs. Final Purpose Clauses in English', *Text*, 5, 55-84.
- Tomlin, R. S. (1985) 'Foreground-Background Information and the Syntax of Subordination', *Text*, 5, 85-122.