

## Clause order as a discourse marker

Renkema, J.

*Published in:*  
The Berkeley Conference on Dutch Linguistics 1997

*Publication date:*  
2000

[Link to publication](#)

*Citation for published version (APA):*  
Renkema, J. (2000). Clause order as a discourse marker. In T. F. Shannon, & J. P. Snapper (Eds.), *The Berkeley Conference on Dutch Linguistics 1997: The Dutch language at the millennium* (pp. 63-77). Lanham, MD: University Press of America.

### 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.

# Clause order as a discourse marker

Jan Renkema

*Tilburg University*

**1. Introduction.** The central question in discourse studies is: what makes a set of sentences a text, or what makes a discourse a discourse?<sup>1</sup> Therefore the central concept in discourse studies is connectivity, which is usually subdivided into mental coherence and textual cohesion. In my view, coherence and cohesion are in fact two sides of the same discourse coin.

In this paper I will deal with cohesion. This is not to say that this concept is completely clear. I want to make a threefold contribution to cohesion research. First, I want to make the concept more clear. The cohesion between sentences or utterances is often referred to as information flow or topic continuity, but it is by no means clear what these concepts actually describe. See, for example, Van Dijk (1977), Givón (1983), and Ford (1993). Second, in cohesion analysis the current approach is to look back: how is a proposition or a larger piece of discourse linked to the preceding text? See, for example, Jou & Harris (1990) and Tomlin (1985). This is the so-called incremental approach. But I hope to show that it is also necessary to investigate the links between a piece of text and the following text. I will present a plea for backward and forward analysis, for an anaphoric as well as a cataphoric approach. Third, I want to show how after conceptual clarification and a two-sided analysis, hypotheses about discourse phenomena can be tested in corpus-linguistic and experimental research. I will restrict myself to a stylistic phenomenon on the sentence level that is hypothesized as being influenced by cohesion phenomena: the clause order, main > subordinate or subordinate > main. See, for example, Lehmann (1988) and Matthiessen & Thompson (1988). And within this phenomenon I will further restrict myself to the Dutch connector *omdat* 'because'.

I would like to start with an example. Below are two fragments from newspaper articles with two *because*-sentences. In one of them I have changed the order. The examples are originally Dutch. The sentences are translated literally. The question is: which is the version that actually appeared in the newspaper?

---

<sup>1</sup> I would like to thank Hans Hoeken for statistical support and Ineke van Oosten for correcting my English.

- (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 Mynskoff-theater, dat 1600 stoelen telt. *Omdat* het podium daar aanmerkelijk groter is dan de gemiddelde Nederlandse schouwburg, zal de mise-en-scène worden aangepast

'At this moment Cyrano is 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 cannot be reserved more than six months in advance. Van den Ende has expressed interest in the Minskoff theater, which has 1600 seats. *Because* this 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.

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

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

Fragment (1a) is the fragment that actually occurred in the newspaper. And indeed, at least in my view, the cohesion in (1b) seems to be weaker; the connections are rather strange. Fragment (1b) contains a paragraph boundary beginning with *because*. The subclause as it were continues the topic of the preceding paragraph by mentioning Broadway. Thus, a paragraph boundary seems less motivated than in (1a), where the location mentioned in the main clause is the new paragraph topic. It seems that topicality (i.e. being the topic of a passage) has something to do with clause order. In the last sentence of (1b) the cohesion is weaker than in (1a) as a result of the greater distance between Minskoff and its anaphor *this*. It would seem that the length of the link between elements in subsequent sentences influences clause order in some way.

**2. Cohesion, a conceptual clarification.** How could “information flow” be best defined? How can we distinguish between the different kinds of given and new information? How can we decide whether information is foregrounded or backgrounded? In discourse studies these key concepts are too vague; they cannot be used in analyzing discourse or testing hypotheses (see e.g. Renkema 1993). In the diversity of linking phenomena I would propose to distinguish four possibilities. I will refer to them using different symbols as in the following representation. Between the brackets there is a piece of discourse in the context of other discourse.

(2) ... t a { t a c n } c t ...

t: topical linking

An element can refer to the preceding discourse and can be continued in the following. This element is the best candidate for the topic of the discourse. I will refer to an element with this forward and backward linking as topical (not to be confused with topical in the grammatical sense of fronting an element).

a: anaphoric linking

An element might be only backward or anaphorically linked. Thus, topical and anaphoric elements are both given, but being a disappearing element the anaphoric element has less influence on information flow.

c: cataphoric linking

An element can come up in the discourse and can be referred to in the following discourse. In this case the linking is cataphoric. It can of course become topical afterwards if it is mentioned again.

n: non-bound

An element can come up in discourse without being linked to anything. Such an element is non-bound. A non-bound element is new, just as a cataphoric element is. But the difference is the influence on information flow. A non-bound element is backgrounded in the sense of being less important, like an anaphoric element. In the representation below I have given the relations between these four possibilities and the concepts of fore- and backgrounding and given versus new information.

(3) Cohesion and related concepts

|            | backgrounded | foregrounded | given | new |
|------------|--------------|--------------|-------|-----|
| topical    |              | +            | +     |     |
| anaphoric  | +            |              | +     |     |
| cataphoric |              | +            |       | +   |
| non-bound  | +            |              |       | +   |

**3. Cohesion analysis.** On the basis of the work of, among others, Halliday & Hasan (1976) and Givón (1995), I would propose to distinguish three types of cohesion.

## (4) Three types of cohesion relations:

1. repetition (r)
2. pronominal relations (p)
3. -nymic relations: syno-, anto-, hypo-, mero- (s, a, h, m)

Relations 1 and 2 are more or less grammatical relations. Repetition means that a constituent is repeated literally, e.g. in a story about a girl the constituent *the girl*. In a pronominal relation anaphors are used: in a story about a girl the constituent *she*. The third relation is a semantic one. Four types are distinguished here: synonymy (e.g. *conference* – *meeting*); antonymy (*the old ones* – *the new ones*); hyponymy or hyperonymy (e.g. *tulips* – *flowers*) and meronymy (part – whole relation: *the car was in good condition* – *only the brakes had to be repaired*). With these types of relations each constituent in a discourse can be analysed backward and forward. The analytical framework can be best described by giving an example. Below is an example from the (Dutch) corpus research reported below. It is the beginning of a newspaper article.

## (5) a. sample analysis

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.’

In this stretch of discourse the second sentence is the target sentence. This sentence has the form: *because* (.....)[.....]; the subordinate clause is placed in parentheses, the main clause in square brackets. The clauses are subdivided into constituents. Here follows the target sentence again, now subdivided into constituents.

## (5) b. 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.]

## c. the cohesion structure

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

First the legend for the direction of the cohesion relation. A relation that is both forward and backward (topical) is coded in bold face; an anaphoric relation in lower case and a cataphoric relation in uppercase. The "n" stands for a non-bound relation. The first element in the subordinate clause, *the cabinet*, has a topical meronymic relation with the rest of the discourse. In the preceding text the constituent *vice-premier* occurs and in the following discourse another part of the cabinet: *minister*. The other five constituents have no relation with the rest of the discourse. In the main clause the first constituent has a repetition relation with *buying power* in the preceding sentence. The next constituent *will remain intact* has a synonymy relation with *preserve* in the preceding sentence. The next two constituents have a forward relation; *salaries* to the types of income *minimum wage* and *government benefit*, and *will rise* to *receive more and benefit*. The last constituent *minimally* has no relation at all with the context.<sup>2</sup>

**4. Hypotheses.** The approaches in the literature on information flow and cohesion at sentence level are of two types. In the first type, the issue is the syntactic status of the information: is the information placed in the main clause or in the subordinate clause? Elaborating on concepts derived from more stylistically oriented traditional grammar, it is hypothesized that the main clause tends to contain central or foregrounded information, whereas the subordinate clause usually contains peripheral or backgrounded information. When the order main > subordinate is reversed, the postposed main clause continues to contain the central information, and the preposed subordinate clause serves as an interpretative framework for the main clause. Thus, in both orders the main clause contains the central or more topical information. This approach is called the topicality hypothesis.

1. *The topicality hypothesis:*

The information in the main clause—irrespective of its position in the sentence—is more topical than the information in the adverbial clause.

The concept "topical" 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 of the preceding section—the constituents assigned a boldface symbol or a double symbol (with both lower-case and capital letter). The topicality hypothesis can be schematized as follows.

(6) schematization of the topicality hypothesis

t [t] (...) t  
t (...) [t] t

<sup>2</sup> In this cohesion analysis, relations based on inferred or accessible knowledge (see e.g. Prince 1981) are not dealt with. E.g. the constituent *citizens* in the subordinate clause could be inferred from *everyone* in the preceding sentence. But this relation is more a coherence relation based on reasoning than a cohesion relation based on one of the three types of relations in the discourse itself. This cohesion analysis also does not take into account the distance between linked constituents or the number of linkings between one element and subsequent others.

According to this hypothesis the main clause has more topical information than the subordinate clause.

In the second approach to information flow, the issue is the position of the information: at the beginning or at the end. Continuing the Prague School's theme-rheme approach, discourse can be considered to be a series of sentences in which each subsequent sentence begins with a theme that was introduced at the end of the previous sentence as a 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 order main > subordinate or subordinate > main thus relates to the necessity for backward (anaphoric) or forward (cataphoric) linking. This approach is called the linking hypothesis.

## 2. The linking hypothesis

The position of the clause—irrespective of whether it is a main or a subordinate clause—is related to the information stream, in such a way 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 of the preceding section, these are the elements with the lowercase and capital letters. The linking hypothesis can be schematized as follows:

### (7) scheme of the linking hypothesis

A [A] (B) B and not A (B) [A] B

According to this hypothesis, the first clause has more anaphoric information than the second one, and the second clause has more cataphoric information than the first one.

**5. Corpus research.** From newspapers, occurrences of the relevant constructions with *omdat* 'because' were gathered. Sentences with a postposed subordinate (henceforth *because*-clauses) were much more frequent than sentences with a preposed subordinate (henceforth *Because*-clauses). The ratio *because-Because* was 9:1. Only "clear cases" were selected. The following types were ignored.

- (8) Dit is nodig omdat anders de werkloosheid fors oploopt.  
'This is necessary because otherwise unemployment will increase dramatically.'
- (9) 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.'
- (10) 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 (8) is not interpretable without the preceding discourse due to the anaphor *this*. Sentence (9) contains two subordinate clauses, and sentence (10) begins with a double connective. Furthermore, only those sentences were selected which did not occur at the beginning or the end of a text or paragraph. The corpus contained 65 *because*-clauses and 25 *Because*-clauses.

The sentences were analyzed and encoded within the framework of analysis presented in section 3. Table 1 presents the data testing the topicality hypothesis.

|                                 | topical | non-bound |
|---------------------------------|---------|-----------|
| <i>because</i> -clause (n = 64) |         |           |
| main (first position)           | 1.06    | 2.36      |
| sub (second position)           | 0.73    | 3.14      |
| <i>Because</i> -clause (n = 25) |         |           |
| sub (first position)            | 1.20    | 2.16      |
| main (second position)          | 1.12    | 3.52      |

Table 1. Number of topical and non-bound constituents in relation to type of clause and position.

The topicality hypothesis predicts that the main clause will contain the most topical information, irrespective of its position in the sentence. The first column of table 1 shows that a first-position clause contains more topical information than a second-position one (1.10 vs 0.84;  $z = 1.84$ ,  $p < .05$ ). This means that the main clause scores more highly than the subordinate clause when the main clause is in first position (1.06 vs. 0.73), but not when it is in second position (1.12 vs. 1.20). In other words it is not the syntactic status (main or subordinate) which determines the topicality. So the topicality hypothesis must be rejected.

In table 2 the data are presented for the linking hypothesis.

|                                 | topical | non-bound |
|---------------------------------|---------|-----------|
| <i>because</i> -clause (n = 64) |         |           |
| main (first position)           | 63      | 17        |
| sub (second position)           | 17      | 27        |
| <i>Because</i> -clause (n = 25) |         |           |
| sub (first position)            | 56      | 16        |
| main (second position)          | 52      | 28        |

Table 2. Percentage of clauses with at least one linking constituent in relation to type of clause and position.

The linking hypothesis predicts that the first clause will contain more anaphoric information than the second one, and that the second clause will contain more cataphoric information than the first one. Table 2 shows that the first-position clauses contain anaphoric information more often than the second-position ones ( $z = 4.53$ ,  $p < .001$ ). However, this difference is completely due the sentences in which the main clause is in first position (63% vs. 17%). It is clear that of the four types of clauses, the second-position subordinate clause is the odd one. This



could be explained by assuming that the subordinate clause in second position has a relation only with the preceding main clause.

The second-position clause contains cataphoric information more often than the first-position clause, but the result just fails to reach significance ( $z = 1.63$ ,  $p = .052$ ). The trend, however, is the same whether the main clause is first or the subordinate clause is first. So there is some support for the linking hypothesis.

**6. Experimental research.** In the corpus research, hypotheses were tested on the basis of operationalizations of topicality and linking phenomena. One could argue that this approach is not dealing with the reality of the production and the perception of discourse. Is it really the case that writers prefer to continue a main clause or a subordinate clause in postposition? Is it really the case that readers expect that a new sentence will continue the main clause or the last clause of the preceding sentence? Two experiments were conducted to answer these questions, questions on the degree of realism of the findings based on a necessarily small corpus research.

**6.1. An expectation experiment.** In the first experiment subjects were asked about their expectations on the continuation of information after sentences with the structure "main clause > *because*-subclause" or "*Because*-subclause > main clause." The topicality hypothesis predicts that readers expect that the discourse continues the information in the main clause, irrespective of its position. The linking hypothesis predicts that readers expect that the discourse continues on the topic of the last clause (main or subordinate).

**6.1.1. Material.** The material consisted of eight sentences with a main clause and a subordinate clause connected by *omdat* 'because'.<sup>3</sup> The sentences were constructed in such a way that they could occur at the beginning of a text (so without anaphoric links or reference to previously given information). The clauses were as much as possible compatible in length, number, and nature of constituents and information value, in order not to give any bias to subjects in choosing information out of a clause to be continued in the next sentence. The following examples, for instance, were not appropriate as target sentences.

- (11) De werknemers weigerden om hun staking te beëindigen omdat de werkgevers bleven vasthouden aan arbeidsvoorwaarden die veel te streng waren.  
'The employees refused to end their strike because the employers insisted on maintaining terms of employment that were much too stringent.'
- (12) Omdat het regent zal de pont vandaag niet vertrekken.  
'Because it is raining, the ferry will not depart today.'

---

<sup>3</sup> In a pilot experiment discourse fragments were used, ending in the target sentence. But in this approach subjects' expectations were influenced by the preceding discourse

In (11) the subordinate clause gets more attention than the main clause through the words *much too stringent*. In (12) the subordinate clause has less information value than the main clause.

In a pretest the eight target sentences were assessed on naturalness. Out of ten subjects, four persons rated one of the eight sentences as too unnatural. This one was replaced by another. The material was expanded with filler sentences with other connectives. Below is one example of a target sentence in the different conditions:

- (13) a. De gemeenteraad heeft plannen ontwikkeld voor de uitbreiding van de speeltuin omdat ouderverenigingen hadden geklaagd over het gebrek aan recreatieve voorzieningen.  
 'The town council has developed plans for the extension of the playground because parents' associations complained about the lack of recreational amenities.'
- b. Omdat ouderverenigingen hadden geklaagd over het gebrek aan recreatieve voorzieningen, heeft de gemeenteraad plannen ontwikkeld voor de uitbreiding van de speeltuin.  
 'Because parents' associations complained about the lack of recreational amenities, the town council has developed plans for the extension of the playground.'

**6.1.2. Subjects.** There were 96 subjects. They were freshmen college students between 17 and 21 years of age. Of these, 48 subjects were asked to choose either the subject of the main clause or the subject of the subordinate clause as the best candidate for a possible text continuation (subject choice). The other 48 subjects were asked to underline the constituent about which they expected the text to continue.

**6.1.3. Design.** A within-subject design was used. Each subject got four *because*-sentences and four *Because*-sentences (and eight filler sentences). The order of the sentences was varied systematically.

**6.1.4. Procedure.** The experiment was conducted in groups. The subjects were told that the research was about expectations of readers. After succinct instructions on the procedure and a practice sentence, the subjects received a booklet with a sentence and a task on each page. One half of the subjects got a dual-choice task: "in the following sentence I expect information on ..." followed by the subjects of the main and subordinate clauses, of which they were to choose one. The other half of the subjects had a free choice: "Underscore the word or word group about which you expect more information in the following discourse." The experiment took approximately twenty minutes.

**6.1.5. Results.** The linking hypothesis predicts continuation of an element in the last clause. Table 3 presents the results. In the subject choice condition, the percentages were in the expected direction. In the first column, the order main > sub, the subject of the subordinate clause has a higher percentage whereas in second column, the order sub > main, the subject of the main clause has a higher

percentage. But in the free choice condition the picture is reversed; there constituents of the first clause are chosen more frequently. However, in both cases the differences are not significant (subject choice: chi-square (df = 1) = 1.26;  $p > .05$ ; free choice: chi-square (df = 1) = 1.04;  $p > .30$ ).

|             | subject choice |            | free choice |            |
|-------------|----------------|------------|-------------|------------|
|             | main > sub     | sub > main | main > sub  | sub > main |
| main        | 46             | 52         | 53          | 47         |
| subordinate | 54             | 48         | 47          | 53         |

Table 3. Choice of elements from a clause as a function of clause order (in percentages).

**6.2. A rating experiment.** In a second experiment subjects were asked to rate the naturalness of a passage. Subjects received a passage consisting of two sentences—a target sentence and a sentence continuing the main clause or the subordinate clause—and were asked to rate the naturalness of the combination of the sentences. Afterwards they got the four combination possibilities and had to choose the combination which actually occurred in the newspaper. The topicality hypothesis predicts that those passages are rated more natural or realistic in which the second sentence continues the main clause, and the linking hypotheses predicts that the passages in which the second sentence continues the last clause of the preceding sentence will be found more natural or realistic.

**6.2.1. Material.** The material consisted of passages containing two sentences which could have occurred at the beginning of a newspaper article. The first sentence had the form main > *because*-subordinate or *Because*-subordinate > main. The second sentence was a continuation of the main or the subordinate clause. Each passage was constructed in four ways. Below is an example (the arrow means: second sentence continues with ...).

(14)a. main > sub → sub

De Tweede Kamer bepaalde vorig jaar dat tv sponsoring moet worden beperkt, omdat de publieke omroepen hun boekje te buiten gingen in lucratieve contracten met grote ondernemingen. Deze contracten hadden geleid tot discutabele promotieactiviteiten.

'The Dutch House of Commons ruled last year that TV sponsoring is to be restricted, because the public networks were overstepping their bounds in striking deals with big companies. These deals had led to questionable promotional activities.'

b. main > sub → main

De Tweede Kamer bepaalde vorig jaar dat tv sponsoring moet worden beperkt, omdat de publieke omroepen hun boekje te buiten gingen in lucratievecontracten met grote ondernemingen. In een bij grote meerderheid aangenomen wet werd bepaald dat sponsoring alleen is toegestaan als het om sportwedstrijden gaat.

'The Dutch House of Commons ruled last year that TV sponsoring is to be restricted, because the public networks were overstepping their

bounds in striking deals with big companies. In a bill passed by a large majority it was decided that sponsoring would be allowed only during sports broadcasts.'

c. sub > main → sub

Omdat de publieke omroepen hun boekje te buiten gingen in lucratieve contracten met grote ondernemingen bepaalde de Tweede Kamer vorig jaar dat tv sponsoring moet worden beperkt. Deze contracten hadden geleid tot discutabele promotieactiviteiten.

'Because the public networks were overstepping their bounds in striking deals with big companies, the Dutch House of Commons ruled last year that TV sponsoring is to be restricted. These deals had led to questionable promotional activities.'

d. sub > main → main

Omdat de publieke omroepen hun boekje te buiten gingen in lucratieve contracten met grote ondernemingen bepaalde de Tweede Kamer vorig jaar dat tv sponsoring moet worden beperkt. In een bij grote meerderheid aangenomen wet werd bepaald dat sponsoring alleen is toegestaan als het om sportwedstrijden gaat.

'Because the public networks were overstepping their bounds in striking deals with big companies, the Dutch House of Commons ruled last year that TV sponsoring is to be restricted. In a bill passed by a large majority it was decided that sponsoring would be allowed only during sports broadcasts.'

There were 24 passages, and in addition the subjects got 26 filler passages in which the second sentence contained a *Because*- or *because*-clause.

**6.2.2. Subjects.** There were 45 subjects ranging in age between 21 and 25 years. They were students at a School of Journalism.

**6.2.3. Design.** Via the "Latin square design" each subject rated six passages for each of the four combination possibilities and each passage was rated by the same number of subjects. After this task the subjects got the four combinatorial possibilities and had to choose which possibility actually had occurred in the newspaper. The order of the possibilities was varied.

**6.2.4. Procedure.** The experiment was conducted in groups. The subjects were told that the experiment was about the naturalness of discourse. In the first part of the experiment the subjects received a booklet with a passage with one of the four combinatorial possibilities on each page. For each passage they had to answer, using a seven-point scale, the following question: "How natural does the second sentence sound as a continuation of the first sentence?" The scale was from "very unnatural" to "very natural." In the second part of the experiment the same subjects received the 24 passages with all the combination possibilities. For each passage in four versions they had to answer the following question: "Which of the four versions do you think actually occurred in the newspaper?"

**6.2.5. Results.** Table 4 presents the results for the first part of the experiment, the naturalness ratings. The linking hypothesis predicts that in the order main > sub the version with the continuation on the sub is rated as most natural (and the other way round for the order sub > main).

|                       | order → | main > sub  | sub > main  |
|-----------------------|---------|-------------|-------------|
| continuation ↓        |         |             |             |
| on main clause        |         | 4.32 (0.84) | 4.50 (1.07) |
| on subordinate clause |         | 4.41 (1.07) | 4.19 (1.20) |

Table 4. Rating of naturalness (scores indicate means on a seven point scale; the standard deviation is given in brackets,  $N = 45$ ).

The means are in the expected direction, but the interaction between "order" and "continuation" is not significant ( $p = .12$ ). Analysis of variance proved that the order, main > sub or sub > main, had no effect on the rating of naturalness (4.37 vs. 4.35;  $p = .89$ ). Also the continuation on main or subordinate clause had no effect (4.41 vs 4.30;  $p = .59$ ).

Table 5 presents the results of the second part of the experiment, the comparative rating, i.e. the answer to the question which of the four possibilities actually occurred in the newspaper.

|                       | order → | main > sub  | sub > main  |
|-----------------------|---------|-------------|-------------|
| continuation ↓        |         |             |             |
| on main clause        |         | 7.86 (3.26) | 5.68 (4.00) |
| on subordinate clause |         | 7.84 (2.47) | 2.61 (2.20) |

Table 5. Comparative rating (scores indicate the mean number of times that a version is chosen; the standard deviation is given in brackets;  $N = 44$ ).

The interaction between order and continuation is significant ( $p < .01$ ). But the pattern is not fully in harmony with the prediction. As predicted, after the order sub > main the continuation on the main clause was chosen more often than the continuation on the subordinate clause. But after the order main > sub there was no difference. Analysis of variance proves that the order main > sub is more often considered as the version which actually occurred in the newspaper (7.85 vs. 4.15;  $p < .001$ ). And a continuation on the main clause is more often considered as more natural (6.77 vs. 5.23;  $p < .001$ ).

**7. Concluding remarks.** This corpus-linguistic and experimental investigation was begun to resolve the controversy formulated in the two hypotheses. Formulated in terms of clause order, the question can be asked: is connectivity based on position or syntactic status? In other words, is there any support for the linking hypothesis or the topicality hypothesis?

The results of the corpus research support the linking hypothesis and not the topicality hypothesis. The topicality hypothesis predicts that the main clause—irrespective of its position (in front or at the end)—contains the most topical elements. As we have seen in table 1, no evidence could be found for

this. The first clause contains significantly more topical information. However, this result does not support the linking hypothesis; it only proves that the preferred position for topical information is in the first clause. If we look at the anaphoric and cataphoric information, however, as we can see in table 2, then the first clause—irrespective of its syntactic status (main or subordinate)—is found to have significantly more anaphoric elements and the second significantly more cataphoric elements. So this supports the linking hypothesis. Nevertheless, it should be emphasized that the differences are caused solely by those sentences in the corpus in which the main clause is in first position.

The first experiment, the expectation experiment, does not confirm the findings from corpus-linguistics research. No support was found for the linking hypothesis. In their selection of elements, subjects did not make any distinction between main clause and subordinate clause, irrespective of their position in the sentence. Thus, the expectations about discourse continuation are not influenced by the position of the clause. However, on closer inspection, the results do not refute the linking hypothesis convincingly. There is another explanation for the fact that no differences were found. Below is another example of the stimulus material (literal translation from Dutch):

- (15) a. De werknemers weigerden om hun staking te beëindigen omdat de werkgevers vasthielden aan de arbeidsvoorwaarden.  
 'The employees refused to end their strike because the employers insisted on maintaining the terms of employment.'
- b. Omdat de werkgevers vasthielden aan de arbeidsvoorwaarden weigerden de werknemers om hun staking te beëindigen.  
 'Because the employers insisted on maintaining the terms of employment, the employees refused to end their strike.'

Main clause and subordinate clause are as similar as possible in terms of structure and information load so as to rule out any other factors than the position of the clause in the task of choosing a constituent to continue the discourse. There are no adjectives which can attract extra attention; the subjects of the clauses are real alternatives, etc. The only valid conclusion to be drawn from this experiment is the following: there is no difference for expectations regarding continuation of main clause or subordinate clause when the clauses are equivalent in syntactic structure and information load and when they are of the same length. A comparison with the sentences in the corpus research revealed that, due to the "boreness" of the structure, the experimental sentences might have to be considered as too deviant from real sentences, in spite of the fact that they scored well on naturalness in the pretest.

In the second experiment, the evaluation experiment, in which subjects were asked about the naturalness of a combination with a continuation sentence, the only difference that was found occurred in the design with a forced choice. If the subjects could not compare different versions, neither the position of the clause nor the continuation on main clause or subordinate clause had any effect. The fact that an effect could only be found when versions could be compared means that the effect is only found in an artificial task. This is probably caused by the

fact that in an experiment only the factor under investigation can be manipulated: the position of the clause in combination with a continuation on main clause or subordinate clause. All the accompanying and more natural phenomena like difference in length, marked adjectives, etc. are controlled for. It is likely that with more natural stimulus material, effects will also be found with more natural tasks. But in this case it would be hard to determine which of the phenomena could serve as an explanatory factor. In any case, the evaluation experiment only partially confirms the corpus-linguistic research. Only in the order subordinate > main is a continuation on the main clause rated as more natural. This can be considered as support for both hypotheses. In the order main > subordinate, however, there is no difference in the ratings of the continuations on main and subordinate clause. This can be considered as a lack of support for the topicality hypothesis. So the controversy between position and syntactic status is still unresolved, although corpus-linguistic and experimental research suggest that there are more arguments for the linking hypothesis.

Equally important, if not more important, is the central aim of my contribution. I wanted to show how the concept connectivity could be clarified, and how a two-sided cohesion analysis could be conducted. I have chosen the phenomenon of "clause order" to illustrate how—on the basis of this clarification and cohesion analysis—hypotheses can be tested in which connectivity concepts play a central role. I hope that I have sufficiently underscored the necessity of conducting corpus-linguistic research and analyzing isolated phenomena in experimental settings. In this respect my contribution is a bit like a research program in which other phenomena could be investigated related to the concept of connectivity.

### References

- Chafe, Wallace L. 1984. "How people use adverbial clauses." In Claudia Brugman & Monica Macaulay (eds.), *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, 437–449. Berkeley: Berkeley Linguistics Society.
- Dijk, Teun A. van. 1977. *Text and context. Explorations in the semantics and pragmatics of discourse*. London: Longman.
- Givón, Talmy. 1983. *Topic continuity in discourse*. Amsterdam: Benjamins.
- \_\_\_\_\_. 1995. *Functionalism and grammar*. Amsterdam: Benjamins.
- Haiman, John & Sandra A. Thompson (eds.). 1988. *Clause combining in grammar and discourse*. Amsterdam: Benjamins.
- Halliday, M. A. K. & Ruqaiya Hassan. 1976. *Cohesion in English*. London: Longman.
- Jou, Ierwen & Richard J. Harris. 1990. "Event order versus syntactic structure in recall of adverbial complex sentences." *Journal of psycholinguistic research* 19: 21–42.
- Lehmann, Christian. 1988. "Towards a typology of clause linkage." In Haiman & Thompson, 181–225.

- Matthiessen, Christian & Sandra A. Thompson. 1988. "The structure of discourse and 'subordination'." In Haiman & Thompson, 275–329.
- Pol, Till van der, Leo Noordman & Jan Renkema. 1993. "A functional explanation of variation in Dutch goal-instrument clauses." Unpublished manuscript, Catholic University of Brabant.
- Prince, Ellen F. 1981. "Toward a taxonomy of given-new information." In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
- Renkema, Jan. 1993. *Discourse studies. An introductory textbook*. Amsterdam: Benjamins.
- Thompson, Sandra A. 1985. "Grammar and written discourse: Initial vs. final purpose clauses in English." *Text 5*: 55–84.
- Tomlin, Russell S. 1985. "Foreground-background information and the syntax of subordination." *Text 5*: 85–122.