The processing of underspecified coherence relations

Wilbert Spooren

\(^a\) Discourse Studies Group, Department of Language and Literature, Tilburg University, P.O. Box 90153, Tilburg, NL-5000 LE, Netherlands
E-mail: Published online: 11 Nov 2009.

To cite this article: Wilbert Spooren (1997) The processing of underspecified coherence relations, Discourse Processes, 24:1, 149-168, DOI: 10.1080/01638539709545010

To link to this article: http://dx.doi.org/10.1080/01638539709545010

PLEASE SCROLL DOWN FOR ARTICLE
The Processing of Underspecified Coherence Relations

WILBERT SPOOREN

Discourse Studies Group, Tilburg University

One reason discourse coheres is because coherence relations such as Cause-Consequence can be inferred between the sentences that function as the building blocks of the discourse. This article discusses cases in which the coherence relation remains underspecified: A connective is used that does not "literally" express the intended coherence relation. Underspecification has been dealt with in pragmatics and psycholinguistics. The purpose of this article is to give an analysis of the phenomenon in pragmatic terms, as well as different strategies of how language users deal with these issues, phrased in terms of Horn's (1984) Q- and R-principle. Apart from that, data from the psycholinguistic literature on the interpretation of underspecified relations and data from language acquisition research are presented that suggest that both speakers and hearers tend to be cooperative in using underspecified relations.

There is no one-to-one relation between interclausal (or coherence) relations and connectives. A speaker who is about to encode a message containing a causal link of the type \( \text{CAUSE}(p,q) \) has a number of choices. Languages like Dutch and English contain many causal connectives and, apart from that, the speaker can also opt for non-causal connectives. In the latter case the speaker leaves the coherence relation underspecified. Underspecification is the purpose of this article. Consider, for instance, the example under (1).

(1) D. Sinds 1 juni is Ron Kaal (43) hoofdredacteur van het maandblad O. Kaal werd vorig jaar benaderd door de uitgever, Maurice Keizer, nadat hij in NRC Handelsblad een kritisch stuk had geschreven over het eerste nummer van het blad.
E Since June 1 Ron Kaal (43) has been editor in chief of the monthly O.
Kaal was approached last year by the publisher, Maurice Keizer, after he
had written a critical article on the first issue of the magazine in NRC
Handelsblad.

In example (1) the causal connection between “Kaal was approached by the pub-
lisher” and “he had written a critical piece on the first issue” is signaled by a tem-
poral connective nadat (‘after’). Apparently, interclausal and intersentential
connections can remain underspecified. In the most extreme case, the relation
remains completely implicit, as in (2), but the cases that I would like to discuss
are the ones in which there is a connective, but this connective does not match the
intended relation, as in (1).

(2) We must leave the room. There are too many people inside.

Consequently, by underspecification I mean that the semantics of the connect-
tive that is used to indicate the link does not fully match the semantics of the rela-
tion that is intended by the speaker or writer. In practice this means that
underspecified relations can be recognized by exchanging an underspecified con-
nective by an explicit counterpart.

Underspecification is most often discussed in the domain of causal relations,
but it is surely not restricted to that area. For instance, Meyer (1985) distinguishes
between relations of collection and of contrast. Both can be expressed using the
same connective and.

(3) a This is what happened today. John bought a book, Peter bought a tele-
vision set, and Charlie bought a bike. (collection)
   b John and Charlie both spent their money very differently. John bought a
book and Charlie bought a bike. (contrast)

The replacement criterion indicates that (3)b is a true case of underspecification:
In (3)a and cannot be replaced by but without loosing the collective reading,
whereas in (3)b and can easily be replaced by but.

In the literature, underspecification has been dealt with in pragmatics and in
psycholinguistics. The purpose of this article is to bring these two approaches
together. Different strategies in dealing with underspecified coherence relations
will be discussed. These strategies will be phrased in terms of Gricean implica-
tures. Following Traugott and König (1991), underspecification will be described
as an outcome of the workings of the so-called R-principle (which roughly states
“say no more than necessary”).
One might expect that utterances such as (1), which are very frequent in natural discourse, create interpretive problems for the hearer/reader. After all, how are hearers to know what relation a speaker is hinting at if it is not indicated explicitly in the discourse? The answer must be that the use of underspecified relations is heavily restricted. Some of these restrictions will be discussed. Findings from the psycholinguistic literature on the interpretation of underspecified relations will be reviewed and two studies on the production of underspecified coherence relations by language users varying in language proficiency will be described.

It will be concluded that the language users' use of connectives reflects their conversational cooperativeness: Speakers gradually learn to become more specific in signaling coherence relations, whereas at least for the discourse types and reading tasks discussed, readers and hearers are on the look-out, so to speak, to find underspecified relations.

**THE IMPLICATURE ANALYSIS OF UNDERSPECIFIED RELATIONS**

Underspecified interclausal relations can be analyzed in the framework of pragmatic implicatures. Grice (1981) used examples like (4) to demonstrate the workings of the maxim of manner ("be orderly"): (4) He took off his trousers and went to bed.

The conjunctive *and* signals a relation of temporal succession ("and then"). This can be accounted for by the assumption that, without indications to the contrary, events are reported in their "natural order".

In Horn's reformulation of the Gricean maxims (Horn, 1984) the occurrence of underspecified relations falls under the workings of the so-called R-principle, which dictates that a speaker must not say more than necessary. In practice this means that from an utterance ... *p* ..., i.e., an utterance containing *p*, a hearer can deduce that "... more than *p* ...". The R-principle is named after the maxim of Relation ("Be relevant") and generalizes over that maxim and the maxims of Manner ("Be brief and orderly, avoid ambiguity and obscurity") and the second Quantity maxim ("Do not make your contribution more informative than necessary"). It is similar to what Levinson (1983) calls the Principle of Informativeness and generates what Horn has called "lower bounding implicata," in that it sets a lower bound on the amount of information expressed. Examples of the workings of the R-principle are ("+->" is used to indicate (conversational) implicatures):

(5) a John ate the cake +-> "John ate all of the cake"

b I lost a book yesterday +-> "the book was mine"
Do you know the time -> "if you know the time, tell me what it is" (indirect speech act)

The R-principle is complementary to the Q-principle, named after Grice's first Quantity maxim, which states that a speaker say as much as he or she can. This principle generates "upper bounding implicata": From a sequence ... p ... the hearer can deduce "... no more than p ...". The Q-principle works primarily with so-called scalar predicates, i.e., predicates that can be arranged on an implicational scale, such as <all, some>, <hot, warm>, <certain, probable, possible> etc. In general, a linguistic scale <p1, p2, ..., pn> has the property that an utterance ... p1 ... entails the truth of "... pi+1 ..." (the bath is hot entails "the bath is warm"). The Q-principle states that if a speaker utters ... p1, ..., then apparently (s)he was not in a position to commit him- or herself to the stronger statement "... pi+1, ..." (with 0 < i < n), from which one can infer that "not (... pi, ...)"). Examples are:

6. a Some of the boys came -> "not all of the boys came"
   b The bath is warm -> "the bath is not hot"

Horn (1984) notes that the Q-principle can be seen as a principle of hearer economy: The hearer's task is alleviated if all of the information is explicitly present in the discourse (from the hearer's point of view, the ideal case is the one in which a different form is used for every intended meaning). A correlate of this principle of hearer economy in processing terms is found in the work by Britton and Gülğöz (1991), who have shown that readers of an expository text in which text-linking inferences were added performed better on several dependent measures (recall, efficiency of recall, and an inference test).

The R-principle is a principle of speaker economy: The speaker's task is relieved if (s)he can use a small number of forms to express many meanings. (For a more extensive discussion of scalarity and implicatures, see Horn, 1984 and Levinson, 1983).

Traugott and König (1991) argue that the implicature analysis of underspecified relations can account for meaning changes in connectives such as (French) cependant ("yet," originally a phrase expressing co-occurrence) and (Dutch) daientegenwoordig ("consequently," originally expressing a spatial and later a temporal sequence). They describe these meaning changes as a process of conventionalizing of conversational implicatures. These implicatures are generated by the R-principle. For instance, the development of cependant is occasioned by the fact that mere co-occurrence of two situations is rarely highly relevant information, unless the occurrence is contrary to expectation. Gradually, this concessive reading of cependant has been incorporated into the lexicon.
In other connectives, the implicature is still conversational, i.e. it has not been incorporated into the lexical meaning. These are the cases that I would describe as underspecified relations; examples are and expressing a Cause-Consequence relation or a Contrast relation, or and then expressing a Cause-Consequence relation.

Note that the implicature analysis implies that the newer meanings and the underspecified meanings are a specification of the older ones and the explicit ones: The implicature is added, so to speak, to the older or the explicit meaning.

RELATIONSHIPS BETWEEN COHERENCE RELATIONS

What the implicature analysis does not describe is why causal relations are possible interpretations of temporal connectives, and why contrast relations are possible interpretations of list connectives. The relationship between connectives and expressed relations is restricted in an interesting fashion: Whereas the connective and can be used to express contrastive relations (cf. example (3b)), it is difficult to express a concessive relation using and (cf. example (7)).

(7) I'm inclined not to employ John.
He may have all the diplomas but he does not have any experience.
?? He may have all the diplomas and he does not have any experience.

Another example of these restrictions is that it is impossible to use specific connectives like because to express contrastive relations.

How is one to account for these restrictions? Sanders, Spooren and Noordman (1992, 1993) have proposed a framework for expressing relationships between coherence relations. The system is discussed in some detail by Sanders in his article "Semantic and Pragmatic Sources of Coherence: On the Categorization of Coherence Relations in Context," this volume. Sanders et al. made use of a restricted set of primitive features of relations, such as Basic Operation (additive versus causal) and Polarity (positive versus negative). In this fashion, it is possible to express proximities between classes of coherence relations: Temporal and causal relations are close because they only differ in the primitive Basic Operation (distinguishing between additive and causal relations). List and Contrast relations are close because they differ only in the primitive Polarity (distinguishing positive from negative relations). On the other hand, List relations and Negative Cause relations (e.g., John is a Republican but he is honest) are distant because they differ both in Basic Operation and Polarity (List: additive, positive; Negative Cause: causal, negative). Most of the relations between coherence relations have a general-specific character: The definitions of the primitives are such that as a rule causal relations are more specific than additive relations. For instance, Cause rela-
tions are more specific than List relations, because a List relation involves mere conjunction of two states of affairs, whereas a Cause relation also involves an implication relation between the two states of affairs. A Contrast relation is more specific than a List relation because a Contrast relation also involves a negation relation between the related states of affairs.

This suggests that the relation between connectives and coherence relations is restricted in such a manner that only connectives expressing relations which are close (in the above mentioned sense) are candidates for expressing a relation. A second restriction is that the connective chosen should at least be as general as the relation intended (a speaker who uses a causal connective to express a temporal relation misleads his audience, to say the least). In practice, this means that Cause relations can be expressed by causal (because), temporal (then) or list (and) connectives, and that Contrast relations can be expressed by contrastive (but) or list (and) connectives, but that List connectives cannot be used to express Negative Cause relations.

INTERPRETING UNDERSPECIFIED RELATIONS

Given that the choice of connectives is restricted along the lines sketched, one can formulate predictions concerning the production and interpretation of underspecified relations on the basis of the implicature analysis presented earlier. There it was stated that the Q-principle is a principle of hearer economy (the hearer’s task is alleviated if all of the information is explicitly present in the discourse), whereas the R-principle is a principle of speaker economy (the speaker’s task is alleviated if (s)he can use a small number of forms to express many meanings).

It seems to follow from such considerations of processing economy that readers/hearers will rely on the strategy that connectives are used explicitly. Such a strategy can be considered as an instantiation of the relatively shallow processing that was reported by Noordman, Vonk and Kempff (1992). Noordman et al. argue that, as a rule, readers of texts distribute their cognitive energy strategically, in that they are reluctant to make inferences and read extra information into a text, unless they are more or less forced to do so because of the experimental task or unless they have a large amount of domain knowledge (see also Simons, 1993).² If readers follow such a strategy, explicit connections will be processed more effectively (faster, deeper) than their underspecified counterpart. This hypothesis seems to be supported by findings like Haberlandt’s (1982), who showed that a sentence was read faster when its connection to the previous sentence was signaled explicitly than when it remained implicit.

However, in reading research there is overwhelming evidence that at least for one type of underspecified relation the prediction should be reversed. Graesser,
Singer and Trabasso (1994) and Graesser and Zwaan (1995) distinguish between several classes of inferences, some of which have been found to be generated online, and some of which are not. Among the online generated inferences are so-called Causal Antecedent inferences, inferences that “explain why an action, event, or state is explicitly mentioned in the text” (Graesser et al., 1994, p. 372). There is an impressive body of evidence that shows that underspecified Causal Antecedent relations are processed fast and retained well. For instance, Myers, Shinjo and Duffy (1987) report an experiment in which the degree of causality between related events was varied. They found that readers recalled events best when they were related by an in-between degree of causality. If the link is either too weak or too strong, it is not stored (in the latter case, the link is too obvious to be stored). Singer (1996) describes experiments in which he used sequences with a causal, a temporal or an adversative link.

(8) a After five minutes, Rich's shower turned cold, so he turned the shower off. (causal)

b Rich showered with cold water for five minutes. He turned the shower up. (temporal)

c After five minutes, Rich's shower turned cold, but he turned the shower up. (adversative)

Subjects read one of these versions and subsequently had to answer a test question Can cold water be uncomfortable? Answers to this question were faster in the causal and in the adversative version than in the temporal version. Singer's explanation is that reading the second clause in the a and c-versions invokes a mediating proposition like "cold water is uncomfortable", that is helpful in answering the question (see also Van den Broek, 1994 for a review of the pertinent literature on causal inferences in narrative texts).

Another finding in the literature is that the degree to which inferences are made is dependent on text type. Whereas the search for causal antecedents is abundant in narrative texts, according to Graesser et al. (1994) readers of expository texts tend to generate fewer inferences. Nevertheless they have been found there, too. Sanders (1992, Chap. 4) manipulated the type of relation of the target sentence with its context. He found that the target sentence was read faster and was yet better retained when it was related to its context in a Problem-Solution structure than when it was related in a List-structure. (Similar results were obtained by Spooren, Mulder & Hoeken, 1996).

What such results show is that readers are looking for causal connections, so to speak, and thereby follow the R-principle rather than the Q-principle. This strategy reflects the fact that causal relations play a central role in our mental knowledge store: Readers rely heavily on assumptions concerning causality when
processing text. (See Michotte, 1962, for an extensive discussion of the strength of this post hoc ergo propter hoc illusion in the domain of visual perception.) In the analysis of Zwaan, Magliano and Graesser (1995), readers of a text construct a situational model of that text (a representation of the situation depicted by the text). Constructing a situational model is facilitated in case the text conveys a continuous situation, i.e. if the description does not contain breaks on the spatial, temporal and causal dimension. The above-mentioned results indicate that explicit uses of temporal connectives can be considered as breaks on the causal dimension, which impede the processing of the text: If two events are related in a text merely temporally, this will frustrate the information processing. This is reflected in the findings of Bestgen and Vonk (1995), who showed that using an explicit temporal marker is interpreted as a discourse segmentation, that reduces the availability of preceding information.

In sum, the implicit causal reading of temporal connectives is grasped more easily than the explicit temporal reading, because it gives what Van den Broek (1994) calls a connecting inference, “a causal relation between the focal event and information that has remained activated after processing of the prior event” (p. 561). This finding is mostly based on narrative texts, but it is also supported by research in which expository texts were used.

DEVELOPMENTAL ISSUES IN THE USE OF UNDERSPECIFIED RELATIONS

In sharp contrast to the rich literature on the reception of causal underspecifications, the production of underspecified coherence relations has, to my knowledge, as yet not been an issue in psycholinguistic research. A consideration from Horn’s principle of Speaker Economy suggests that speakers and writers will use the R-principle, i.e. they will rely on the willingness of hearers and readers to infer the correct relationship, and, consequently, they will tend to leave coherence relations underspecified. This section and the following sketch an analysis of underspecified uses of connectives in production data. These data stem from two studies on the acquisition of coherence relations. In the first study two groups of primary school children were compared in their use of underspecified coherence relations.

Acquisition data provide information concerning the order in which linguistic structures are mastered and, consequently, concerning the complexity of those constructions. Thus, we can infer processing complexity and processing strategies from developmental patterns.

First-language learners face a double difficulty; They have to master both the concepts and conceptual relations and the linguistic forms that can express this conceptual information. The choice of an adequate linguistic form to express a
specific sort of relation can be seen as a skill that children have to learn. Research reported in such works as Bloom (1991) and Eisenberg (1980) indicates that the capacity to choose a specific connective develops over the years: Semantically more specific connectives such as therefore and although are acquired at a later stage than more general connectives like and and but. Consequently, given a certain relation to be expressed, older children can be expected to use specific forms (following the Q-principle), whereas younger children can be expected to use general forms (following the R-principle). This hypothesis was put to the test in an analysis of utterances of children from different age groups. (This research was carried out in close collaboration with Kiek Tates, cf. Spooren, Tates & Sanders, 1996). In order to investigate the influence of the text type, the context in which the children's utterances were produced was varied.

Method

Children of different ages and with different language-proficiency levels were prompted to produce complex sentences. The setting was more or less controlled, in that the children were given two specific tasks. In one task, they had to describe pictures in which a number of causally related events and situations were taking place (e.g., a street musician playing a violin to an audience crying with emotion). The children were instructed as follows: “In this picture many things are happening. Can you tell me what is happening and how this has come about?”

In the second assignment, the children conversed with the experimenter about controversial topics, such as T.V. programs they were not allowed to see. The children were stimulated to give their opinion and to support it. The researcher's reactions were restricted to back-channel signals.

This design was chosen in order to elicit a sufficient number of different types of coherence relations. The description task was chosen to elicit what are called semantic relations, i.e., relations pertaining to the propositional content of the related clauses. The conversation task was aimed at eliciting so-called pragmatic relations, i.e., relations pertaining to the illocutionary status of the related clauses. (For the terminology see Sanders et al., 1992; Sanders, this volume.) In the experimental session, each task lasted for ten minutes. Two age groups were tested: age 6-7 and age 11-12 (corresponding to the lowest and highest grades in primary school). Each group consisted of six subjects. All of the children were native Dutch speakers. The two subject groups were comparable with respect to reading proficiency.

The children's utterances were recorded on a cassette and subsequently transcribed. Clauses (i.e., main clauses, non-restrictive relative clauses and adverbial
clauses) were taken as the basic unit of analysis. Only clauses which were not (self)interrupted were analyzed.

The resulting utterances were coded for a large number of variables, including the type of usage (implicit, underspecified, explicit). In order to determine whether the use was explicit or not, a list was made of all the connectives used by the children. In addition, a list was made of the relations that these connectives express when used explicitly. Then the coherence relations were coded. (The judges agreed to a high extent concerning the coherence relations in the data. The average agreement percentage was 87.9 %. Agreement varied over the children from 80 % to 96 %). Finally, each coherence relation was coded as implicit, underspecified or explicit.

**TABLE 1**
A sample from the corpus (Study 1).

<table>
<thead>
<tr>
<th>I</th>
<th>Jory, informant (boy), age group: 6-7, reading level: high</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task: Picture description</strong></td>
<td></td>
</tr>
<tr>
<td>I: Interviewer</td>
<td>J: Jory, informant (boy), age group: 6-7, reading level: high</td>
</tr>
<tr>
<td><strong>I</strong></td>
<td><strong>J</strong></td>
</tr>
<tr>
<td>Op deze plaat gebeuren een heleboel dingen. Kun jij mij vertellen wat er gebeurt en hoe dat komt?</td>
<td>Hier staat een auto stil</td>
</tr>
<tr>
<td></td>
<td>Here a car is standing still</td>
</tr>
<tr>
<td></td>
<td>En dan botst daar een andere auto tegen aan.</td>
</tr>
<tr>
<td></td>
<td>And then another car bumps into it.</td>
</tr>
<tr>
<td>Ze zijn ramen aan het wassen.</td>
<td>They are cleaning windows</td>
</tr>
<tr>
<td></td>
<td>En dan laat hij zijn emmertje vallen.</td>
</tr>
<tr>
<td></td>
<td>And then he drops his bucket</td>
</tr>
<tr>
<td>En dan valt dat op een meneer zijn hoofd.</td>
<td>And then it falls on a gentleman's head.</td>
</tr>
<tr>
<td></td>
<td>Hier wil een meneer uit zo'n putje klimmen.</td>
</tr>
<tr>
<td></td>
<td>Here a man is trying to climb out of one of those drain holes</td>
</tr>
<tr>
<td></td>
<td>En dan botsen die anderen er allemaal tegen aan.</td>
</tr>
<tr>
<td></td>
<td>And then those others all bump into it</td>
</tr>
<tr>
<td>omdat hij het deksel omhoog houdt.</td>
<td>because he is lifting the cover</td>
</tr>
<tr>
<td></td>
<td>Die mensen huilen allemaal.</td>
</tr>
<tr>
<td></td>
<td>Those people are all crying</td>
</tr>
<tr>
<td>omdat die vioolspeler zo mooi speelt.</td>
<td>because that violin player is playing so beautifully</td>
</tr>
</tbody>
</table>
Table 1 is a sample from a boy from the younger age group. The occurrences of *omdat* ("because") in utterances 8 and 10 were used to signal causal relations; these were therefore coded as explicit. *En dan* ("and then") in 2 and in 5 were used to express *Cause-Consequence* relations (both occurrences can be replaced by an explicit *daarom* ["as a consequence"]) and was therefore in both cases coded as an underspecified usage.

**Results**

The results are summarized in Table 2. In total, 1097 coherence relations were produced by the children (age 6-7: 566, age 11-12: 531). The number of relations produced per child was analyzed in an analysis of variance (ANOVA) with task as a within-subjects factor, and age group as a between subjects factor. There were no main effects of Task ($F(1,10) = 1.54, SE = 54.71, p = .24$) and Age Group ($F(1,10) = 1.18, SE = 43.14, p = .30$), but there was a significant interaction Task x Age Group ($F(1,10) = 7.77, SE = 54.71, p < .025$): Inspection of the means in Table 2 shows that the younger children produced relatively more relations in the description task, whereas the older children produced more relations in the conversation task. Pairwise comparisons using t-tests indicate that the difference between age groups was significant for the description task ($t(10) = 3.01, SE = 3.77, p < .025$), but not for the conversation task ($t(10) = 1.28, SE = 4.29, p = .23$).

Next, the relations which occurred without any connective (the implicit relations) were analyzed. In total, 208 of 1097 relations (18.9%), occurred without any connective. These implicit relations belonged mainly to a restricted number of categories: *Evidence* (37 or 17.8%), *Interpretation* (36 or 17.3%), *Enumeration* (34 or 16.3%), *Concession* (16 or 7.7%), *Elaboration* (14 or 6.7%), and *Circumstance* (12 or 5.8%). The remaining 59 relations belonged to categories occurring less than 10 times (5%).

<table>
<thead>
<tr>
<th><strong>TABLE 2</strong></th>
<th>Mean number of coherence relations and percentages of implicit and underspecified relations as a function of age group (age 6-7, 11-12) and task (description-conversation) (Study 1).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group</strong></td>
<td><strong>Age 6-7</strong></td>
</tr>
<tr>
<td><strong>Task</strong></td>
<td><strong>description</strong></td>
</tr>
<tr>
<td>Number of relations</td>
<td>49.5</td>
</tr>
<tr>
<td>Percentage implicit relations</td>
<td>4.0</td>
</tr>
<tr>
<td>Percentage underspecified relations</td>
<td>82.0</td>
</tr>
</tbody>
</table>
The percentage of implicit relations was calculated per child. The resulting percentages were compared between the two age groups, using the Mann-Whitney U test, with age group as a between-subjects factor. Furthermore, a Wilcoxon Matched Pairs test was carried out to determine the influence of the task. Surprisingly, the younger group left fewer relations implicit than the older age group (age 6-7: 15.2%, age 11-12: 23.2%, z = 2.08, p < .05). There was also a large difference in the degree of implicit relations per tasks: The description task lead to much less implicit relations than the conversation task (description: 5.3%, conversation: 32.3%; z = 3.06, p < .01). The latter difference may account for the difference between age groups. The fact that the older children produced relatively more relations in the conversation task may have inflated their use of implicit relations. Per task, the two groups did not differ in their use of implicit relations (description task: age 6-7: 4.0%, age 11-12: 6.5%; z = 1.04, p = .30; conversation task: age 6-7: 28.3%, age 11-12: 36.2%, z = 1.44, p = 0.15).

The final analysis concerned the degree of underspecified relations in the case of a connective. Of the 889 relations with connectives, 630 (70.9%) were underspecified. These belonged to a restricted number of categories: Cause (258 or 41.0%), Circumstance (138 or 21.9%), Evidence (53 or 8.4%), Enumeration (34 or 5.4%), and Interpretation (32 or 5.1%). The remaining 115 relations (18.2%) belonged to categories occurring underspecified in less than 3% of the cases.

The type of usage (underspecified, explicit) was compared between the two age groups and per task. The percentage of explicit uses was calculated for each child. These percentages were evaluated using a Mann-Whitney U test, with Age Group as a between-subjects factor. In addition, a Wilcoxon matched pairs test was used with Task as a within-subjects factor.

There was a significant difference between younger and older children in their use of underspecified relations (z = 2.40, p < .05). At age 6-7, 75% of the relations expressed with a connective were underspecified, at age 11-12, 65%. The first thing to note is that the degree of underspecified uses of coherence relations was high. The children frequently used general connectives, relying on the hearer to infer the correct relation. The second thing to note is that the use of the explicit signals increased with age: Children gradually use specific connectives more frequently as they get older. To use maxim terminology: The use of the Q-principle increases with age.

The two tasks differed in the degree of underspecified relations they lead to. In the description task, 81% of the relations was underspecified, in the conversation task, only 57% were underspecified (z = 3.06, p < .01). As in the case of implicit relations, this difference may account for some of the difference between the age groups (underspecification is more frequent in the description task and younger children produce more relations in the description task), but not entirely. This becomes apparent in a comparison of age groups per task: In the description task,
there is no difference between age groups in the degree of underspecified relations (age 6-7: 82%, age 11-12: 79%, z = .16, p = 0.88), but in the conversation task such a difference does exist (age 6-7: 64%, age 11-12: 49%, z = 2.32, p < .05).

Discussion

This study showed that younger children leave relations implicit less frequently than older children and that younger children use underspecified relations more often than older children. The first result can be attributed to the task. Older children produced relatively more relations in the conversation task than in the description task, and implicit relations occurred most often in the conversation task. The fact that the difference between the age groups disappeared in the analysis per task also supports this interpretation.

The second result is central to the present study. It shows that, with age, children become more specific with respect to their relation marking. I have argued that this difference cannot be (entirely) attributed to the task because the two tasks show an interesting asymmetry: The difference between age groups in the use of underspecified relations is only found in the conversation task. This strongly suggests that in the conversational text type children gradually learn to use the Q-strategy.

In a way, these results are complementary to the findings in the literature regarding the reception of underspecified causal relations. In that research narrative and expository texts have been used and these bear the greatest resemblance to the descriptive task used in the present study (see Sanders (1994, this volume) for an analysis of the correlation between text types and relation types). It has repeatedly been found that proficient language users process underspecified causal relations faster than their explicit temporal counterparts, which is in line with an R-strategy. In the production study, it can be seen that, although overall the children are very unspecific in signaling their relations, the Q-strategy is gradually becoming more important, most specifically in the conversational realm.

Together these results suggest that the workings of the Q- and R-strategies apply more generally than just to a specific subset of coherence relations in a specific text type, and that accounts of coherence relations will have to find ways to deal with these various options.

These results confirm the analysis of Bloom and Capatides (1987), who stated that the use of linguistic connectives is a capacity that has to be mastered: A specific use of the connectives occurs only when the child has mastered the coherence relations to such a degree that it is able to fine-tune its language code to the needs of the hearer. If this analysis is correct, then one might expect adult language users to be even more attentive to the conversational needs of the hearer. The question that comes to mind, then, is whether adults might be maximally explicit. There are reasons to doubt such an end line of development, if only for
stylistic reasons: It is generally considered to be a property of high quality language use to vary one's connectives and not to spell out every coherence relation by using very explicit connectives. But adult language learners, i.e., second language learners, for whom stylistic variation is not yet an issue due to insufficient language proficiency, might provide a case in point. Some suggestive data in this respect have emerged from a pilot study of the connective use by an adult Moroccan learning Dutch.

**ADULT LANGUAGE LEARNERS AND THEIR USE OF UNDERSPECIFIED RELATIONS**

Adult second-language learners differ from first-language learners in that they need not learn the concepts and conceptual relations which are to be expressed, but only the linguistic code suited for expressing this conceptual information. Thus it can be stated that it is truly the complexity of the linguistic code that determines the order in which constructions are acquired in adult language acquisition.

In view of the pattern found in the previous study, one might expect adult language users to be even more specific in marking coherence relations than children age 11-12. But one can also expect this effect to be mediated by the complexity of the task that second language learners are faced with because they have only a limited repertoire of linguistic means at their disposal.

The material for this analysis comes from the ESF corpus, a corpus that was collected in a longitudinal, cross-linguistic project (the project is described by Klein and Perdue, 1991). The Dutch part of the corpus consists of data from speakers of Moroccan and Turkish origin. The data were collected over a period of approximately three years, starting shortly after the speakers had arrived in Holland. At regular intervals, interviews were held with the informants, who had to perform different tasks, such as conversing, describing a film, role playing, etc.

Input for this pilot study was a subset of the data of one of the Moroccan subjects, who was given the code name Mohammed. This particular subject was chosen because he used a considerable degree of complex sentences in his utterances. The data come from the last year of the data collection because in the first two years, there were relatively few complex utterances. The analyzed samples were the first 2000 lines from sessions 1, 3, 5, 6, 7, and 9 of the last year.

Table 3 gives a sample from Mohammed's utterances. The sample comes from the last stage of the period under investigation (session 7). The topic of conversation is discrimination. A number of Mohammed's utterances contain connectives. For instance, the connective and in 2 is used to signal the Circumstance relation between utterance 1 and 2. And in 3 is used to signal the List relation between 2 and 3, and but in 8 signals the Concession relation between 7 and 8. All of these
I: Interviewer
M: Mohammed, informant (Moroccan)
Period: last
Genre: conversation
Topic: discrimination

M 1 Bijvoorbeeld wij daar
M 2 en zeggen “Stom turk”
M 3 en uh ja “Wat jullie doen hier?”
M 4 “Moet terug naar eigen land” en eh
I Zeggen ze dat tegen jou?
M 5 Ja.
M 6 Hm?
M 7 Niet tegen mij
M 8 maar hun zeggen zo he.
I Tegen jou zeggen ze *Stomme turk*?
M 9 Ja.
M 10 En dachten dat ben ik turk was.
M 11 Dat is di/discriminatie.

For example, we [were] there [in a discotheque]
and [they] say “Stupid Turk”
And uh yeah “What are you doing here?”
“Must go back to [your] own country” and uh
Do they say that to you?
Yes.
Hm?
Not to me
But them say so right
To you they say “Stupid Turk”?
Yes.
And [they] thought that am I was [a] Turk.
That is di/discrimination.

have been coded as explicit relations. The use of and in 10 is different because it is used where the speaker intends an Explanation relation (and can be replaced by because). This was coded as an underspecified use.

These data were coded by a number of raters. Because of the disrupted character of many of the subject’s utterances, only the sequences containing an explicit connective were taken into consideration. The coding procedure consisted of three steps. A list was made of the connectives and the relations which they express explicitly (the list was made on the basis of the definitions of relations in Mann & Thompson, 1988). Then the relation and the connective between two segments were determined. Lastly, it was determined whether the connective used explicitly expressed the coherence relation used.

A possible problem for all corpus analytic research is the coding reliability of the data. In this case, each utterance was coded independently by three raters. In 64.5% of the cases, there was complete agreement among the three raters concerning the “type of use” classification. This relatively low agreement is partly
due to the disrupted character of many of the subject's utterances. In cases of disagreement, the coding was discussed until the raters reached agreement. A total of 554 utterances containing a connective were coded. Of these, 84 (15.2%) were coded as underspecified relations. The type of usage was cross-categorized with part of the year (begin, middle, end) (see Table 4).

The first thing to note is a gradual increase in number of coherence relations over the three periods (begin: 120 relations, middle: 147 relations, end: 287 relations; $\chi^2 = 87.04, df = 2, p < .01$). With time, the number of complex utterances Mohammed uses increases, a finding which obviously reflects Mohammed's improved language proficiency.

There is a strong dependency between part of the year and type of use ($\chi^2 = 7.81, df = 2, p < .05$). Inspection of the table shows that in the beginning of the third year, the percentage of underspecified relations is small (10.0%); in the middle, it increases considerably (21.8%) and in the last part of the year, it decreases again (13.9%). This speaker seems to shift from using the Q-principle to the R-principle and back to the Q-principle.

An explanation for this pattern might run as follows. In the beginning the speaker does not engage in fairly complex utterances (the number of complex utterances is relatively low). At a later stage he is willing to use more complex relations, but he has only few connectives at his disposal and he is forced to be very general, hence the large number of underspecified coherence relations. At the final stage of this period the subject has mastered the language sufficiently to make a real choice between explicitness and underspecification and specific marking of relations prevails again. If this is an adequate picture, it becomes very interesting to look at even later stages of development, to see how this tendency towards a Q-strategy fares. Unfortunately, the ESF corpus does not allow for such extrapolations, in lack of relevant data.

Needless to say, such an explanation is very speculative and these data need to be treated with much care because of the relatively low interrater agreement and the small size of the data set. Nevertheless, in combination with the results of Study 1, they are suggestive.
There is a remarkable difference between the two developmental studies, namely, the relative degree of underspecified uses of connectives (high in Study 1, low in the present pilot study). This suggests that it is indeed the case that adult language learners are, more than children, aware of the needs of their conversational partners. Again, caution is needed here, because other factors covary, such as the fact that many of the samples from the ESF corpus involved free conversation, and none of the tasks of the ESF informants were aimed at the production of complex utterances (as opposed to the children’s tasks in Study 1). Furthermore, the language proficiency of the informants in the ESF corpus is very low, even in the third year, compared to the children investigated in Study 1, which resulted in relatively few complex utterances.

More important than these relative differences between the two studies are the similarities. In both, a comparison was made across time, and in both, a tendency was found toward reliance on the Q-principle.

CONCLUSION

In this article I have discussed a pragmatic analysis of underspecified coherence relations. I have also discussed a number of studies concerning the use of underspecified coherence relations. The studies show a remarkable pattern. In the psycholinguistic literature on the reception of discourse, a predominance of the R-principle is found: Language users process underspecified coherence relations faster than explicit coherence relations. This finding is mainly based on causal coherence relations and on the processing of narrative texts, but it can also be observed in expository texts. This was taken to indicate that readers expect, so to speak, to find causal links between the sentences of a text. One might be inclined to interpret this tendency of readers as typical for temporal connectives that sustain a causal interpretation. To my knowledge no data are available on the reception of other types of underspecification (such as additive connectives like and for contrastive or causal relations), and this is clearly an area of future research.

The production study showed a gradual predominance of the Q-principle: An increase of language proficiency is accompanied by a tendency to become more specific (a finding which was corroborated by some suggestive data from an adult language learner). And this tendency to become more specific was most obvious in a conversation task.

Note that these results are contrary to what one might expect in terms of hearer and speaker economy. An economical speaker would be well advised to use only non-specific connectives, whereas hearer/reader economy dictates not to “read more” into the text than explicitly stated. It is as if acquiring proficiency in language implies, among other things, that the language user takes into consideration the needs of the conversational partner.
Obviously, the generalizability of the results obtained in these studies is a point of further research. And the use of underspecified coherence relations does not depend solely on application of the Q- and R-strategy. In fact, it depends on a large number of factors such as the shared background knowledge of the conversational partners and the exact linguistic form of the related clauses. With respect to the former, if the conversational partners are both aware that Tom and Dick cannot stand each other, a sequence such as *At 10 o'clock Tom came in. Ten minutes later Dick left* is likely to have a causal implication. Such an implication is absent if the partners are trying to make a list of all the guests that were present at a party. And with respect to linguistic information, according to Traugott and König (1991), sequences of states are more likely to give rise to causal interpretations than sequences of events (p. 197). Such interrelations show that the processing of underspecified coherence relations is a very complex phenomenon, that is subject to many restrictions that as yet are not very deeply understood. Yet, however tentative, the data presented in this article suggest the cooperativeness of language users in establishing coherence.

Acknowledgement: This work has benefitted much from comments by Leo Noordman, Ted Sanders and three anonymous reviewers.

NOTES

1. Oversteegen (1993) analyzes temporal and non-temporal uses of connectives from a synchronic point of view and argues that the Traugott and König analysis cannot be the entire story. She cites a number of cases in which non-temporal uses have different, rather than more specific, meanings. A case in point is (i):

(i) D Terwijl Anne jou gisteren de keuken heeft helpen opruimen, weiger je haar nu te helpen.
E While Anne helped you clean the kitchen yesterday, you refuse to help her now.

Part of the meaning of the temporal *terwijl* is that there is a temporal overlap between the related clauses, whereas (i) shows that in the case of non-temporal *terwijl* temporal overlap can be explicitly violated. Consequently the non-temporal meaning of *terwijl* is not more specific than the temporal meaning of *terwijl*: It has a different meaning.

It might be argued, however, that at the level of argumentation, simultaneity does hold in (i), as can be seen from the acceptability of examples such as *Whereas Anne helped you clean the kitchen yesterday, at the same time, you refuse to help her now*, in which at the same time means something like "at the same time it is also true that". This was pointed out to me by Birgit Bekker (personal communication).
2. It is not self-evident that the findings of Noordman et al. (1992) can be applied to the processing of underspecified coherence relations. In all of the experimental texts that Noordman et al. used for instance, an explicitly causal connective was used, and what Noordman et al. showed was that readers, as a rule, do not fully derive "the" causal interpretation of that link. The processing of underspecified relations deals with the derivation of, for instance, a causal relation when a temporal relation is signalled. Yet it seems that in both cases issues of depth of processing are involved.

REFERENCES


