Causal ambiguity in Natural Language: conceptual representation of 'parce que/because' and 'puisque/since'

Adeline NAZARENKO-PERRIN - IBM Paris Scientific Center
3 et 5 Place Vendome, 75021 Paris Cedex 01, France

Introduction

This research deals with the representation of causal relations found in texts written in natural language, in order for KALIPSOS [1], an NL-understanding and question-answering system, to encode causal information in conceptual graphs so as to handle causal information and reasoning.

Natural languages such as French or English have many ways to express a causal relation. It can be syntactic (*parce que/because*) or lexical (*provoquer/to produce*), explicit or implicit (*Je me suis cassée la jambe et je n'ai pas pu venir/I broke my leg and I couldn't come*), contained in a single sentence or extended over a whole paragraph.

Being particularly interested in explicitly stated causal information, I have studied interpropositional relations introduced by *parce que/because* and *puisque/since* which are the more specifically causal conjunctions.

My approach differs from previous work on causality which was either an attempt to define causal relations in logic ([2] [3] [4]) or the design of AI systems for planning and diagnosis [5]. My research is based on natural language understanding.

If one wants to retrieve and exploit causal information coded in NL-texts, a semantic analyser that builds an adequate representation of causal links is needed. The importance of this point has been underestimated. For KALIPSOS, Sowa's Conceptual Graph theory [6] has been taken as a target representation model; this model can be interpreted in logical terms, thus allowing deduction. Future research will use these interpretation rules and will exploit extra-linguistic knowledge for automated reasoning. This subject is not addressed in this paper which focuses on the semantic analysis problem.

Part I explains that the main difficulty in representing the semantics of *parce que/because* and *puisque/since* is the anchoring of the causal relation. Parts II and III show how to deal with and represent this ambiguity on both syntactic and argumentative levels.

I - Problem

To begin with, it should be noted that this research has been done on the French language. The conjunction *since*, in English, raises other problems in addition to those raised by *puisque* in French, but they are not dealt with here. In this paper, *since* is never used to introduce a temporal relation.

Because of lack of space, I have assumed that the reader is familiar with Sowa's Conceptual Graph model and notations [6].

Traditionally, cause is viewed as a two-argument relation. In the CG-model the relation "X has Y for cause" can be represented as follows (the concepts are written in boxes and the relations in circles):

$$\boxed{X} \longrightarrow \bigcirc\!\!\!\text{CAUSE} \longrightarrow \boxed{Y}$$

When studying such a relation three tasks have to be performed: the first and second arguments (resp. X and Y) have to be identified and the nature of the CAUSE relation has to be determined.

1- Identifying the first argument of *parce que/because* and *puisque/since* relations is trivial because these conjunctions introduce a clearly defined subordinate clause. It would be much more difficult for conjunctions like *en effet/actually*, *donc/thus* or *ainsi/so*, for instance.

2- Determining the nature of the relation is more complex. A cause can be direct or indirect, essential or incidental, deliberate or accidental, several factors might be involved, and a cause may vary according to the point of view. But this question cannot be solved on purely semantic grounds. The conjunctions I have studied, in particular, and natural language, in general, do not make a systematic distinction between these types of causes. I assume that speakers and listeners either make do with a general, basic causal relation or use complementary information about the world. Therefore, in the KALIPSOS project, I have chosen not to raise this question during the text encoding phase but to solve it either during the pragmatic analysis or, if needed, during the question-answering phase (information retrieval process), using extralinguistic knowledge.

3- Identifying the second argument, however, raises difficulties. The question is to know where to anchor the causal relation in the main clause graph because this anchor may be ambiguous for two reasons.

- Syntactic level

  Je veux que mon fils l'épouse parce qu'elle est riche.
  I want my son to marry her because she is rich.

This statement can be interpreted in two different ways depending on the context:

I1 *Je veux que mon fils l'épouse parce qu'elle est riche... ainsi, il ne sera plus à ma charge.*
I want my son to marry her because she is rich... that way I'll no longer have to support him.

I2 *Je veux que mon fils l'épouse parce qu'elle est riche et pas parce qu'il l'aime. Les mariages d'amour, je n'y crois pas.*
I want my son to marry her because she is rich and not because he loves her. I don't believe in love matches.

This syntactic ambiguity will be analysed in more detail in the second part of this paper.

- Argumentative level

The statement *Il a de la fièvre/He has a fever* can be considered from two points of view, as a fact or as the speaker's action.

  - Giving the cause of a statement considered as a fact produces an objective explanation as in:
  Il a de la fièvre parce qu'il est malade.
  He has a fever because he is sick.
  where the information 'he is sick' answers the question *Pourquoi a-t-il de la fièvre?/Why does he have a fever?*

  - Giving the cause of a statement considered as a speaker's action produces a justification, implicating the speaker as in:
  Il a de la fièvre puisqu'il est tout rouge/
  He has a fever since he is flushed/
  where the information 'he is flushed' answers the question *Pourquoi dis-tu qu'il a de la fièvre?/Why do you say he has a fever?*

The two interpretations differ with respect to the speaker's argumentation: either he explains a fact or he justifies himself. This argumentative ambiguity will be explained in the third part of this paper where I shall show how it can be represented in conceptual graph structures containing utterance type concepts. Please note that "utterance" is used in the sense of the act of uttering, not that which is uttered.

II - Syntactic ambiguity when anchoring causal relations

(1) *Je veux que mon fils l'épouse parce qu'elle est riche.*

I want my son to marry her because she is rich.

The ambiguity (I1 and I2) corresponds to different syntactic analyses (A1-broad scope and A2-narrow scope):

A1 [*Je veux que mon fils l'épouse*] *parce qu'elle riche.*
[*I want my son to marry her*] *because she is rich.*

A2 *Je veux que* [*mon fils l'épouse parce qu'elle riche*].
I want [*my son to marry her because she is rich*].

Except for the case where the conjunction follows a comma (which rules out the second syntactic analysis), I suggest that the statement is totally ambiguous and that it is impossible to choose between these interpretations on syntactic grounds. The parser must deliver two syntactic trees.

The semantic analyser, however, must try to choose one reading only. There is no guaranteed determining factor but some clues may combine in favour of one or other interpretation. Three of these clues are shown below: coreference of pronouns, temporal correspondence and encyclopedic knowledge.

Coreference of pronouns

(2) *J'ai dit que Pierre partait parce que ça lui chantait.*
I said that Peter was leaving because he felt like it.

(3) *J'ai dit que Pierre partait parce que ça me chantait. I said that Peter was leaving because I felt like it.*

Coreference of underlined pronouns favours the narrow scope interpretation for statement (2) and the broad scope one for statement (3). However, no interpretation is really ruled out.

Although the definition of precise rules seems difficult and still has to be worked out, I think that this coreference is a factor in the cognitive process of natural disambiguation.

Temporal correspondence

It is useful to remember that an effect cannot precede its cause in time and that this temporal information can be computed. It depends on the choice of tenses, on the aspectual indications and on the situation characteristics given in the semantic definitions of the verbs. Several models of temporal representation using conceptual graphs have been designed and implemented [7]. The semantic analyser can use this information to disambiguate a statement.

  Je pense qu'il a mangé parce qu'il y a du fouillis.
  I think he has eaten because there is a mess.

The tenses show that the mess comes after the action of eating. Therefore, the mess cannot be the cause of that action and the narrow scope interpretation is ruled out.

This criterion based on temporal relations is much easier to implement than the first one.

Encyclopedic knowledge

> *J'espère qu'il va faire beau parce que Franck est en vacances.*
> *I hope the weather will be nice because Franck is on vacation.*

Unless we accept a magic causality of meteorological phenomena, everyone knows that clouds do not depend on Franck's activity. This information rules out the narrow scope interpretation for the above statement.

Common sense knowledge may be the most important factor of disambiguation. Sowa's model has tools to encode this kind of knowledge but the relevant information, which is not semantic but pragmatic, must be defined in connection with a specific application.

For statement (1) and similar statements the parser must deliver two different analyses. The semantic analyser tries to remove the ambiguity using a set of combined clues.

III - Argumentative ambiguity when anchoring causal relations

Suggested representation

A statement such as

(4) *Il est malade puisqu'il a de la fièvre.*
He is sick since he has a fever.

cannot be paraphrased by

> *Il est malade parce qu'il a de la fièvre.*
> *He is sick because he has a fever.*

which would contradict our basic knowledge of illness and fever. It would be better to paraphrase (4) with one of the following statements:

> - *Je dis (pense, crois, suppose, etc.) qu'il est malade parce qu'il a de la fièvre.*
> - *I say (think, assume, imagine, etc.) he is sick because he has a fever.* (meaning *He is sick and I say so because he has a fever.*)

These paraphrases bring out the speaker's activity which remains implicit in (4). Although there may be different kinds of activity (epistemic, cognitive, speech, etc.) they shall all be considered here as utterances, since the problem I wish to address is that of finding the proper structure to represent causal relations.

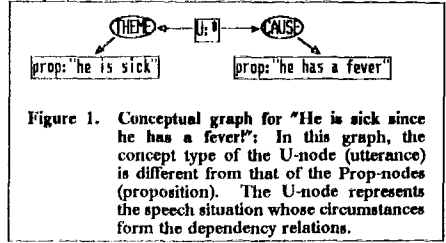The above paraphrases show that statement (4) could be represented as follows:

Figure 1. Conceptual graph for "He is sick since he has a fever": In this graph, the concept type of the U-node (utterance) is different from that of the Prop-nodes (proposition). The U-node represents the speech situation whose circumstances form the dependency relations.

Consequences

Several remarks can be made about the suggested representation.

1- Fever is no longer considered as a cause of illness; it causes the speaker's declaration *Il est malade/He is sick.*

2- In figure 1, proposition 1 is not in the scope of U as it does not belong to its theme. Actually, as O. Ducrot has shown [8], there are two successive utterances in a statement such as (4). This is easily explained if we acknowledge that the information introduced by *puisque/since* is already known. As a presupposed piece of information it cannot be stated; in the representation I have chosen it cannot be the theme of an utterance.

3- The suggested representation shows that it is necessary to introduce a new type of concept (U) in order to represent the speech situation (which is not a proposition) and the related linguistic phenomena. I suggest that this type of concept will enhance the analysis of linguistic phenomena such as modality, temporal relations, reported speech, or any facts dealing with the circumstances surrounding an utterance. For instance, let us consider a woman who says to a child: *Va dans ta chambre!/Go to your room!* This statement and its context could be represented by the graph shown in figure 2.
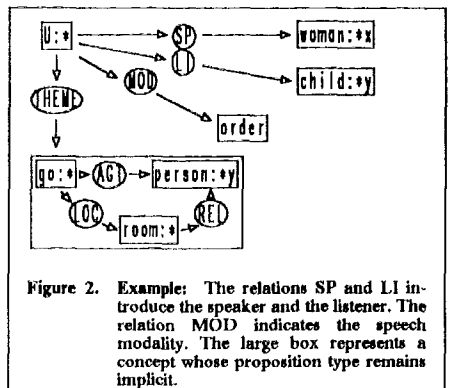

Figure 2. Example: The relations SP and LI introduce the speaker and the listener. The relation MOD indicates the speech modality. The large box represents a concept whose proposition type remains implicit.

I have shown how *puisque/since* statements might be represented. I have also argued that it is necessary to introduce a U-type concept (U-node) into the CG-model so that utterance-related linguistic phenomena can be taken into account. However, in order to represent interpropositional causal links, it must be possible to build the proper representation automatically. Two questions follow: when is it neccessary to have a U-node in the graph of the main clause? Where should a causal relation be anchored in the modified representation?

U-nodes

Any sentence, written, spoken or even reported, is the responsibility of a speaker (at this point, paragraph or text levels are not considered). This leads to a first rule:

Rule 1: A graph representing a sentence must depend on a U-node.

Any variation in the speech situation (change in point of view, in assertiveness, etc.) must also be indicated. The list of variations is open but mainly concerns verbs (*dire/to say, croire/to believe, penser/to think*, etc.) which introduce reported speech or a new point of view.

Rule 2: Verbs that introduce reported speech or a point of view must be defined as an utterance (U-type concept).

Figures 3 and 4 give examples of the application of rules 1 and 2. A further analysis of the utterance activity would refine and extend these rules because modality and other linguistic phenomena may interfere. As we are not concerned here by the distinctions between the different kinds of utterance (epistemic, belief, etc.) and the interpretation rules associated with them, we have left the modality of U-nodes unresolved.
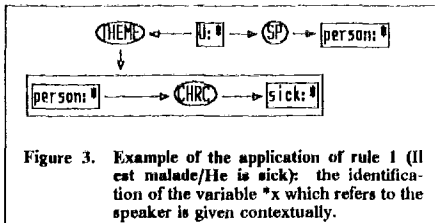
Figure 3. Example of the application of rule 1 (Il est malade/He is sick): the identification of the variable *x which refers to the speaker is given contextually.
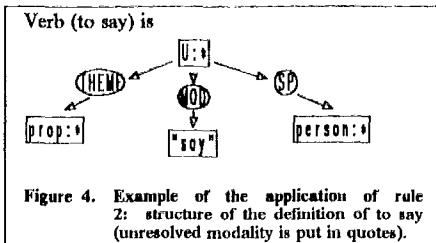
Figure 4. Example of the application of rule 2: structure of the definition of to say (unresolved modality is put in quotes).

Anchoring

As the representation of the main clause becomes more complex the number of anchoring possibilities increases and it is necessary to determine which ones are legitimate.

As already explained above, *parce que/because* and *puisque/since* differ more by the type of their first argument than the nature of the causal relation. *Parce que/because* relations are usually anchored to a proposition whereas *puisque/since* relations are usually anchored to an utterance.

Rule 3: A causal relation expressed by *parce que/because* must be anchored to a propositional node of a conceptual graph.

According to this rule, in the statement

(5) *Je dis qu'il a de la fièvre parce qu'il est malade.*
I say that he has a fever because he is sick.

anchoring can be done in two ways (the statement is syntactically ambiguous) leading to the graphs shown in figures 5 and 6.
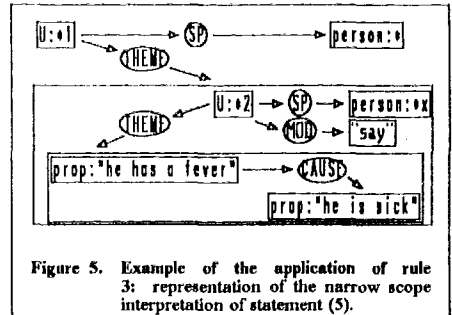
Figure 5. Example of the application of rule 3: representation of the narrow scope interpretation of statement (5).
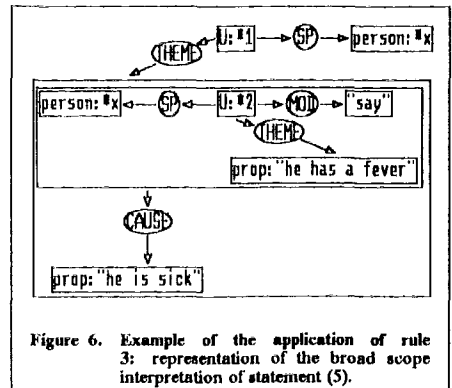
Figure 6. Example of the application of rule 3: representation of the broad scope interpretation of statement (5).

Rule 4: A causal relation expressed by *puisque/since* must be anchored to a U-node of a conceptual graph.

Puisque/since has additional syntactic properties that limit the ways the causal relations can be anchored. The following unacceptable (*) statements

- * *Est-ce puisqu'il est malade qu'il a de la fièvre?*
* *Is it since he is sick that he has a fever?*
- *Est-ce parce qu'il est malade qu'il a de la fièvre?*
Is it because he is sick that he has fever?
- * *Il a froid non puisqu'il est malade mais puisqu'il a sommeil.*
* *He is cold not since he is sick but since he feels sleepy.*
- *Il a froid non parce qu'il est malade mais parce qu'il a sommeil.*
He is cold not because he is sick but because he feels sleepy.

show that *puisque/since* relations cannot be questioned, denied or, more generally, inserted in another proposition. As noted above these relations cannot be stated, they cannot combine to form a proposition.

Rule 5: A causal relation referred to by *puisque/since* cannot appear in a propositional node of a conceptual graph.

(6) *Je dois lui dire que Jean est là puisque c'est la vérité!*
I must tell him John is here since it is the truth!

The representation of statement (6) shown in figure 7 is an example of the application of rules 4 and 5.
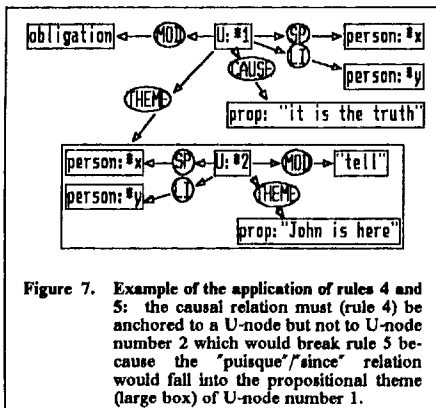

Figure 7. Example of the application of rules 4 and 5: the causal relation must (rule 4) be anchored to a U-node but not to U-node number 2 which would break rule 5 because the "puisque"/"since" relation would fall into the propositional theme (large box) of U-node number 1.

Conclusion

This study of *parce que/because* and *puisque/since* has shown that the question "Cause of what?" is even more important than the question "What kind of cause?". The main task has been to uncover and represent the ambiguity (both syntactic and argumentative) of these conjunctions.

Although linguistic phenomena (coreference of pronouns and temporal correspondence) and encyclopedic knowledge may help to reduce the syntactic ambiguity, there is no guaranteed determining factor.

The importance of argumentative characteristics had to be taken into account; consequently I have introduced a new type of concept to represent the utterance situation. This concept type allows the *parce que/because* and *puisque/since* relations to be properly represented in the form of conceptual graph structures. In addition, I have defined rules to allow the KALIPSOS semantic analyser to build such structures automatically.

Further work has to be done. The example of the definition of *dire/to say* can be extended to encode other verbs (*croire/to believe, penser/to think, supposer/to suppose, nier/to deny*, etc.). It will be necessary to distinguish between different types of utterance nodes in conceptual graphs.

Since the conceptual structures that have been built must enable causal questions to be answered, it is also necessary to define rules so that the question/answering system can handle the utterance type concept nodes.

References

[1] A. Berard-Dugourd, J. Fargues, M.-C. Landau. "Natural Language Analysis Using Conceptual Graphs". *Proc. of the International Computer Science Conf.'88*. Hong-Kong, Dec. 1988.

[2] J. L. Mackey. *The Cement of the Universe: a Study of Causation*. Oxford University Press, 1974.

[3] D. Lewis. "Causation". *Journal of Philosophy*, 70, 1973.

[4] Y. Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*. MIT Press, Cambridge, MA, 1988.

[5] A. C. Graesser, L. F. Clark. *Structures and procedures of implicit knowledge*. Roy O.Freedle (Ed.), Vol. XVII in the series Advances in Discourse Processes, Ablex. Norwood, New Jersey, 1985.

[6] J. F. Sowa. *Conceptual Structures. Information Processing in Mind and Machine*. Addison Wesley Publishing Company. Reading, MA, 1984.

[7] P. Zablit. "Conceptual Graph Representation for Time Reference Interpretation: a Focus on Tense and Aspect". *Proc. of the Fifth Annual Workshop on Conceptual Structures*. Stockholm, 1990.

[8] O. Ducrot. *Dire et ne pas dire: Principes de sémantique linguistique*. Hermann, Paris, 1972.