

Same discourse markers, different discourse relations: What do we learn from naïve annotation experiments?

Ivana Didirková, Anne Catherine Simon, George Christodoulides
Centre de recherches Valibel, Université Catholique de Louvain, Belgium

Introduction

We present a preliminary study carried out in the context of a research project investigating the contribution of prosody to the online interpretation of discourse relations. It has been established that a same discourse marker be used to express several discourse relations, leading to different possible interpretations (e.g. Couper-Kuhlen & Kortmann, 2000; Asher & Lascarides, 2003). The aim of this work is to study how naïve subjects identify discourse relations introduced by two different discourse markers (French *alors* and *et*), as compared to experts' annotation. Our hypothesis is that subjects without any experience in annotating discourse relations will tend to associate each DM with its core meaning rather than its alternative meanings. For example, we expect that subjects will tend to identify a connector such as *and* as inducing an addition, rather than as inducing a relation of specification.

Method and participants

In the first experiment, 44 naïve participants were asked to annotate discourse relations in 176 sequences of utterances, using a multiple-choice procedure. Each sequence is constructed according to the following scheme:

$$S1 + DM + S2,$$

where S1 represents the first utterance, the DM is *alors* or *et* (*so, then / and* in French), and S2 is the second utterance. For each DM, four different discourse relations were proposed: two of them were common for both DMs (consequence and specification); in addition, *alors* could also convey topic shift or concession, while *et* could express addition or temporality. Those discourse relations occur in the LOCAS-F corpus, where they have been annotated by experts (see Degand & Simon (2016) for more details).

All participants were university students at the Université catholique de Louvain and had no previous experience in annotating discourse relations. A sheet explaining and illustrating the six discourse relations was presented to subjects prior to the annotation. They were then asked to choose one of the four proposed relations for each sequence. Results were compared to the experts' annotation.

In a second experiment, a different group of participants was asked to judge each sequence for acceptability using a Likert scale.

Results

Annotations were then analyzed for several criteria. When compared to the original expert annotations performed by the authors, the percentage of naïve annotators that chose the same discourse relation exceeds 50% for each of the six discourse relations. However, these scores differ depending on the relation, ranging from 50.2% for the temporality relation expressed by the DM *et* to 75.33% for the consequence relation expressed by *alors*. These two extreme values seem to be linked with the frequency of use of these two connectors, in that *alors* often can

induce a relation of consequence, whereas *et* alone (i.e. without any other temporal cues) is not likely to be perceived as conveying a temporal relation between two segments.

However, the results relating to the expression of addition by *et* were unexpected. Naïve annotators were somewhat reluctant to identify *et* as conveying the relation of addition between S1 and S2, even though it corresponds to the core meaning of this marker. Furthermore, this result also seems to refute the hypothesis that some of the discourse relations would be more or less transparent for non-experts. In cases where *et* was used to express consequence, there was agreement between the expert and naïve annotators in 64% of the cases.

We then analyzed cases of disagreement between naïve and the expert annotators, in order to detect regularities in the identification of relations in these sequences. According to our results, sequences with *alors* annotated as expressing specification and concession by the experts tend to be annotated as expressing consequence by naïve annotators, and topic shift tends to be interpreted as a concession. Interestingly, the acceptability test has shown that the subjects never considered the topic shift sequences as acceptable. We hypothesize that this relation needs a specific prosody, indicating a clear melodic reset. As for the consequence relation, it was mainly interpreted as concession or as specification. In the case of *et*, subjects were inclined to annotate other relations as being either an addition or a consequence.

These results lead to the conclusion that, in some cases, naïve subjects do not base their judgment exclusively on the frequency of use or on the so-called transparency of discourse relations, but on other elements as well (in line with the results of Mak, Tribushinina & Andreiushina (2013)). Some of the additional elements that affect the interpretation of an utterance will be discussed during the presentation. Moreover, inter-group differences observed in annotations underline the importance of using naïve participants to such tasks in order to compare alternative interpretations of linguistic phenomenon (using a methodology similar to the one in Scholman & Demberg (2017), for example).

References

- Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge, UK: Cambridge University Press.
- Couper-Kuhlen, E., & Kortmann, B. (Eds.). (2000). *Cause - Condition - Concession - Contrast : Cognitive and Discourse Perspectives*. Berlin New York: Mouton de Gruyter.
- Degand, L. & A.C. Simon (2016). Variation of Discourse Markers across a multi-genre corpus of spoken French: Annotating function and meaning. Presentation at *Variation et changement pragmatique-discursif 3 (DiPVaC)*, Toronto, 4-6 May 2016.
- Mak, W. M., Tribushinina, E., & Andreiushina, E. (2013). Semantics of Connectives Guides Referential Expectations in Discourse: An Eye-Tracking Study of Dutch and Russian. *Discourse Processes*, 50(8), 557–576.
<https://doi.org/10.1080/0163853X.2013.841075>
- Scholman, M., & Demberg, V. (2017). Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task (pp. 24–33). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-0803>