

Mikko Haule

University of Helsinki (Finland)

November 7, 2016

STSM report

I have completed a Short Term Scientific Mission (2016/10/03 – 2016/10/15) at the Université catholique de Louvain under the supervision of Prof. Liesbeth Degand. The main purpose of this STSM was to discuss the principles of classification of the occurrences of causal connectives and to compare different schemas used to annotate causal relations.

One part of the STSM concentrated on the classification of causal relations and connectives that I have used in my PhD project in which I study the use of three French causal connectives (*parce que*, *car*, *puisque*) and two Finnish connectives (*koska*, *sillä*) in French and Finnish corpora.

On the one hand, I formulated and presented in a schema the criteria that I have used to determine the different categories of both causal relations and other factors that might have an effect on the choice of a specific connective. We discussed the appropriateness of different categorizations and Prof. Degand offered me some valuable comments and advice in this issue.

On the other hand, we discussed some ambiguous cases which are difficult to classify in precisely one category, especially when a fine-grained categorization is in use. This discussion with Prof. Degand helped me to find a solution to these cases and to define and improve the classification.

I also received some valuable advice from Prof. Degand concerning the use of methods and tools in the data analysis. I could use them during my stay.

Another purpose of the STSM was to compare the classification of causal relations that I use with the annotation schemes of CCR and PDTB and to evaluate the advantages and disadvantages of less/more fine-grained classifications.

CCR distinguishes between *objective* and *subjective* relations according to the source of coherence (Scholman, Evers-Vermeul & Sanders 2016). In objective relations, both segments describe situations in the real world (example 1) whereas in subjective relations, speaker/author is actively engaged in the construction of these relations (example 2).

- (1) [The plaintiff received his car.]S1 because [the advertisement was formulated ambiguously.]S2
- (2) [Drugs destroy people's lives.]S1 so [drugs have to be battled judicially.]S2

The categorization of causal relations in PDTB makes a distinction between 1) *cause*, 2) *cause_belief* and 3) *cause_speech-act*. It differs from CCR in that it makes a distinction between *cause_belief* and *cause_speech-act* which both would fall in the category of subjective relations in CCR. *Cause* (in PDTB) corresponds with objective relations (in CCR).

The classification (MH) I use (based e.g. on Simon & Degand 2007, Stukker, Sanders & Verhagen 2009) makes a distinction between 1) non-volitional, 2a) volitional objective, 2b) volitional subjective¹ 3) mental, 4) epistemic and 5) speech-act relations.

The following table shows these three classifications and their approximate relations to each other. Non-volitional and volitional relations correspond to “cause” in PDTB and “objective relations” in CCR whereas mental and epistemic relations fall in the “cause_belief” category in PDTB and in the category of “subjective relations” in CCR.

CCR	PDTB	MH
objective	cause	non-volitional
		volitional objective
		volitional subjective
subjective	cause_belief	mental
		epistemic
	cause_speech-act	speech-act

I analysed a part of my French and Finnish corpus data using these three schemas. On the basis of the analysis, the following conclusion on the advantages and disadvantages of these categorizations could be drawn.

The advantage of the most fine-grained classification is that it enables a more deep analysis on those causal relations that favour certain connectives. The difference between some French and Finnish connectives in terms of frequency in different causal relations indicates that all distinctions that this fine-grained classification includes are justified. If a less fine-grained classification is used, these differences between causal relations/connectives are not observable.

One of the disadvantages of the fine-grained classification is that in some cases the limit between some relations, e.g. mental and epistemic, is not very clear. This is a problem as such and could be even more problematic when several researchers analyse an occurrence of causal relation which has features that might indicate its belonging to two different categories and they could therefore interpret the occurrence differently. A less detailed classification will not face so many difficulties in the interpretation of borderline cases and is therefore more suitable for analysis/annotation in cases where a less detailed analysis is sufficient.

Although in the fine-grained classification some of the occurrences are unclear due to the borderline cases, many of these unclear cases could be categorised according to the less fine-grained classifications. Therefore, it would be justified to primarily use the fine-grained classification and in borderline cases a less fine-grained classification (for example, mental and epistemic relations could

¹ In my terminology, *volitional objective* refers to relations where the cause segment is clearly an observable state of affairs in the real world like in the sentence *He went home because it was ten o'clock*. In a *volitional subjective* relation the cause segment includes some kind of evaluation like in the sentence *He went home because he had to do his homework*.

be put in the same category that would correspond to “cause_belief” relation in PDTB and “subjective relation” in CCR).

References

Stukker, Ninke & Sanders, Ted & Verhagen, Arie (2009) “Categories of subjectivity in Dutch causal connectives: a usage-based analysis.” Sanders Ted and Sweetser Eve (eds) *Causal Categories in Discourse and Cognition*. Mouton de Gruyter, Berlin.

Scholman, Merel C.J. & Evers-Vermeul Jacqueline & Sanders Ted J.M. (2016) “A step-wise approach to discourse annotation: Towards a reliable categorization of coherence relations.” *Dialogue & Discourse* 7(2). 1–28.

Simon, Anne Catherine & Degand, Liesbeth (2007) “Connecteurs de causalité, implication du locuteur et profils prosodiques. Le cas de car et de parce que.” *Journal of French Language Studies* 17. 323–341.