

CRIBLE Ludivine  
PhD Student  
UCL Université catholique de Louvain  
Place Blaise Pascal, 1, box L3.03.33  
1348 Louvain-la-Neuve (Belgium)  
[ludivine.crible@uclouvain.be](mailto:ludivine.crible@uclouvain.be)

December 12th, 2014

## STSM Report

To whom it may concern,

I have just completed a three-week Short Term Scientific Mission at the University of Fribourg (Switzerland) under the supervision of Dr. Sandrine Zufferey. The goal of this research stay was to assess the replicability of a set of guidelines that I have designed for the selection and annotation of DSDs in spoken corpora, and to evaluate their applicability in the written mode. The expected outcome was to identify weaknesses in my protocol and to offer solutions or improvements, as well as to get some experience in the types of issues that we will encounter as members of WG2. At the end of this STSM, all these goals were achieved.

### Work description

During my stay in Fribourg, Dr. Zufferey and I have worked intensively on three specific aspects of discourse annotation: (1) the identification of candidate DSDs; (2) the quantification of inter-coder reliability with a pre-defined tag set of discourse relations; (3) the identifications of factors improving inter-coder agreements in cases where the DRD semantically underspecifies the intended relation. In a first round of annotation, we selected independently all DSDs that we could identify in a sample of four texts combining English and French as well as speech and writing, using the definition and criteria that I provided in my annotation protocol. We then compared our selections and discussed the disagreements to identify weak criteria in the definition. Eventually, we agreed on a common definition that was used for the next two steps of our work.

The second round of annotation focused on identifying the discourse function of 100 DSDs (previously selected by one coder only in order to prevent potential disagreement at this stage) for each sub-corpus (i.e. spoken and written French and English). We used a closed-list of 29 functions previously defined by Crible (in prep.). We briefly discussed the use of each function together. We annotated the four corpora sequentially, each time discussing the disagreements and identifying their causes before annotating the next corpus. The criteria and guidelines were thus progressively improved and disagreements were reduced. This phase revealed the specific difficulty of high-frequency and polysemous DSDs such as *and*, *so*, *but* which are often semantically underspecified and therefore ambiguous.

This finding led us to conduct a third round of annotation that focused on the annotation of about 50 items of *and* / *et* in the same four corpora, using the revised version of the annotation guidelines, which was the output of the previous annotation phase. Our goal was to evaluate whether these improvements in the protocol would help to reduce our disagreements, and whether we could identify regularities (for instance, recurrent paraphrases) that would help us to discriminate between the potential meanings of these DSDs.

A similar, yet smaller-scale task was assigned to an MA research assistant, Christiane Porzig, who annotated the German translations of DSDs found in the English written corpus in its German translation. She also annotated the discourse function of the DSDs that were added in translation, in order to check for possible German-specific functions that were not accounted for in the protocol (results will be analysed soon).

This STSM also represented an opportunity for me to present my scientific project to several research teams : the English Linguistics Department at the University of Fribourg (November 24th) ; prof. Jacques Moeschler at the University of Geneva (December 10th) ; prof. Corinne Rossari and the

researchers in French Linguistics at the University of Neuchâtel (December 11th). These various presentations were designed to introduce myself and my PhD project, as well as to benefit from the expert opinion of researchers in my field of study. All these colleagues are also members of TextLink, and these meetings therefore did not only contribute to my own network, but also encouraged collaborations within the Action.

Finally, I attended two scientific presentations in the field of linguistics at the host institution: Dr. P. Gygax “Why mechanics are always thought of as men?” (December 1st, University of Fribourg), and Prof. P. Athanasopoulos “Representation of motion concepts in multilingual cognition” (December 4th, Institute for Multilingualism, Fribourg).

### **Main results description**

Our annotation experiments helped us confirm that any annotation protocol needs intense training and several rounds of annotation before a reliable agreement can be reached between two coders, since many choices, although theoretically grounded, involve a number of biases in the application of the tag set. Therefore, we cannot stress enough the necessity of documenting all these choices, for transparency and replicability purposes.

More specifically, the first annotation round revealed the respective weight of both functional and syntactic criteria in the selection of DSDs, the latter being the source of most of our disagreements. The question of syntax also raised the question of the identification of the two related segments as a problematic issue. In the perspective of a multilingual project such as TextLink, the use of formal features is potentially problematic, since the same discourse-marking function can be performed by a number of formally diverse devices across languages.

During the second and third annotation rounds, we discovered the impact of modality (speech vs. writing) for one’s annotation biases, namely regarding the distinction between semantic and pragmatic discourse relations. We also found that speech-specific functions were more diverse and ambiguous, due to the unplanned nature of spoken language and the fact that we read spoken discourse in a context devoid of part of its natural cues (prosody, gestures, common ground). These observations resulted in a better operationalization of all functions included in the protocol, namely by providing an unambiguous paraphrase for each possible function that coders were instructed to use during coding, and by providing more details on how to distinguish the meanings of semantically related functions (e.g. concession and contrast). From the perspective of WG2’s upcoming tasks, this confirmed that no protocol can be exhaustive and perfectly replicable, but that there are ways to improve them, and that what is crucial for a project like ours is to provide guidelines on how to use a taxonomy and how to navigate from one protocol to another, by specifying the different choices, where the boundaries are, and when to apply them.

### **Future collaboration**

The goal and results of this STSM will be presented at the first Action Conference (Crible & Zufferey 2015). The revised version of the functional taxonomy will be discussed along other proposals during the WG2 and WG3 meeting in Fribourg next April. Closely related work testing various aspects of this annotation protocol – that could not be tested here due to lack of time and resources - will be presented at the IPrA Conference in Antwerp next July (Crible & Degand 2015), during a panel session co-organised by the host in Fribourg and the home institution (UCL). So far, on the basis of our three-week collaboration, no other output is planned. We are confident that this STSM will be highly relevant to the discussions and deliverables of WG2 by the experience it provides regarding recurrent problems and caveats to look for and partial leads and solutions.

Thank you in advance for your consideration. Best regards,



## **References**

Crible L. (in prep.) « Towards an operational category of discourse markers: A definition and its model ».

Crible L., Degand L. 2015. “Functions and syntax of discourse connectives across languages and genres: towards a multilingual annotation scheme”. Paper presented at the International Pragmatics Association (IprA) conference, July 26-31, Antwerp, Belgium.

Crible L., Zufferey S. 2015. « Assessing the validity of annotation guidelines : Towards multimodal equivalence of Discourse-Structuring Devices ». Paper presented at the Textlink Action Conference, January 26-28, Louvain-la-Neuve, Belgium.