

# Production and perception of underspecified connectives: the effect of genre

Ludivine Crible<sup>1</sup>, Vera Demberg<sup>2</sup>

<sup>1</sup>University of Edinburgh, <sup>2</sup>Saarland University

[Ludivine.Crible@ed.ac.uk](mailto:Ludivine.Crible@ed.ac.uk)

Keywords: discourse connectives; corpus pragmatics; experimental; acceptability; genre

It is generally assumed that coherence relies on building discourse relations across utterances. When producing discourse relations, speakers and writers have a choice between implicit and explicit marking (e.g. Taboada 2009), the latter being usually carried out by “connectives”. This paper focuses on a third, intermediate level between implicitness and explicitness, viz. *underspecified* connectives (Spooren 1997). They typically include *and*, *but* or *so*, which can be used to express many different meanings besides their encoded one (addition, contrast and consequence, respectively). Our goal is to determine the contextual and linguistic conditions that favor the production of underspecified connectives. Specifically, we want to test whether underspecification is more frequent in informal and unplanned genres, and whether it equally applies to various discourse relations.

We report on the findings of a corpus study and two experiments. First, the distribution of *and*, *but* and *so* and their different meanings is computed across different genres of spoken English from the *DisFrEn* corpus (Crible 2017), where they were manually annotated. The corpus findings suggest the following hypotheses for the experiments: underspecified connectives are more acceptable in informal genres; contrastive uses of *and* (e.g. *marijuana is glorified and alcohol is condemned*) are unacceptable in formal genres; consequential uses of *and* are acceptable (e.g. *he left and the party was ruined*).

We then discuss the results of two offline perception experiments. The first is a forced choice task targeting the acceptability of utterances containing *and* in two computer-mediated genres (chats and comments to online press articles). Four batches of 28 items (each carried out by 30 remunerated participants) oppose the original connective with a stronger or weaker alternative (e.g. original *so* opposed to the weaker *and* in the same pair of utterances expressing the relation of consequence). These stimuli, extracted from a corpus of chat conversations (Loyola CMC corpus, Goldstein-Steward et al. 2008), were used in their original version and in a modified, more formal version. The interface, prepared on LingoTurk (Pusse et al. 2016) for Prolific Academic, visually represented a “chat” display and a “comment” one, in order to clarify the difference in formality between the versions. Results show that the original connective is mostly preferred, especially when confronted to a weaker alternative. There is a slight but significant preference for stronger connectives (e.g. *however*) in the formal genre, especially for the relation of contrast.

In the second experiment, the participants are presented with the same stimuli, either in their “chat” or “comment” version, and they have to drag-and-drop their preferred connective from a list of options. Regression models show that utterances originally connected by *so* tend to be assigned the weaker *and* by participants, whereas this connective is rarely used in contrastive relations. Genre is not a significant predictor, except in relations originally marked by *but*. Overall, both experiments show little effect of genre except for an incompatibility between *and* and contrastive relations in formal genres. In this respect, perception matches production as measured by the corpus study. However, results from the two experimental studies do not fully map, thus showing that perception and elicitation are not necessarily aligned. The combination of corpus-based and experimental methods has highlighted the role of frequency, cognition and genre expectations on connective variation, while further qualitative analyses should uncover the role of linguistic cues in the context of the connectives.

## References

- Crible, L. 2017. Discourse markers and (dis)fluency in English and French. Variation and combination in the *DisFrEn* corpus. *International Journal of Corpus Linguistics* 22(2): 242-269.
- Goldstein-Stewart J., Goodwin K., Sabin R. & Winder R. 2008. Creating and using a correlated corpora to glean communicative commonalities. In *Proceedings of LREC 2008, Morocco*.
- Pusse, F., Asad, S. & Demberg, V. 2016. LingoTurk: managing crowdsourced tasks for psycholinguistics. In *Proceedings of the 2016 Conference of the NAACL: Demonstrations*.
- Spooren, W. 1997. The processing of underspecified coherence relations. *Discourse Processes* 24(1): 149-168.
- Taboada, M. 2009. Implicit and explicit coherence relations. In J. Renkema (Ed.), *Discourse, of Course*, Amsterdam, John Benjamins: 127-140.