This presentation will explore the application of diachronic parsed corpora to questions of information structure and the syntax-pragmatics interface. We present a comparative and diachronic study of the information structure of passive constructions, drawing data from parsed corpora of Icelandic (Wallenberg et al. 2011), English (Kroch et al. 2004; Kroch and Taylor 2000; Taylor et al. 2003) and German (Light 2011). The corpora in question are not annotated for information structural categories, and indeed, recent work suggests that information structural annotation is inconsistent and will require a deeper theoretical understanding of the categories in question (Bech 2013; Cook 2013). Instead, we will highlight several methods for using corpus data for research on information structure without relying on information structural annotation.

In addition to the complete corpora available, we will use data from parsed samples of three parallel New Testament translations: the Tyndale New Testament in the Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME), Luther’s Septembertestament in the Parsed Corpus of ENHG (ENHG), and the Icelandic New Testament translation by Oddur Gottskálksson in the Icelandic Parsed Historical Corpus (IcePaHC). The use of parallel New Testament translations allows us to compare specific verses across texts. Because parallel verses of the New Testament can be assumed to have the same discourse goals, we can consider the different choices made in each language for a single utterance.

Many people share the intuition that the notion “subject” is, at least in part, an information structural one, particularly because subjects frequently appear at the left edge of a clause, which is thought to be associated with given/topical information (cf. Vallduvi 1992). This idea raises interesting questions about the role of the passive construction in the discourse, and how the passive may be used to manipulate information structure by promoting an internal argument to subject.

Seoane (2006) and Los (2009) propose that in English, the passive construction is used more extensively than in the other Germanic languages in order to compensate for the lack of unmarked object topicalization found in languages with verb-seconding (V2). While it appears that the increase in the use of the passive correlates with the loss of V2 word orders in the history of English, quantitative evidence from our corpora undermines proposals like Seoane’s and Los’s. As we will see, V2 languages employ passivization at a lower frequency because they may accomplish the same information structural goals using other means, but object topicalization is not one of those means.

The goal of our presentation will be to apply quantitative data from parsed corpora to the problem sketched above. We will work through the necessary steps, from generating testable predictions through analyzing the results of corpus searches, investigating both comparative and diachronic aspects of use of the passive. In addition to comparing the occurrence of the passive construction across English, Icelandic, and German, we will compare Old, Middle and Modern English. In the end, we will see just how much access to multiple parsed corpora can support the in-depth investigation of a theoretical hypothesis.


