On the use of passives across Germanic

Caitlin Light (University of Pennsylvania) and Joel Wallenberg (University of Iceland)

Introduction There has long been an intuition that the subject position is deeply relevant to information structural concerns cross-linguistically, particularly because they frequently appear at the left edge of a clause, which is thought to be associated with given/topical information (cf. Vallduvi 1992). This raises interesting questions about the role of the passive construction in the discourse, and how it may be used to manipulate the information structure by promoting an internal argument to subject. In this paper, we approach this question from a quantitative crosslinguistic perspective. Using three parsed corpora, we provide evidence that Early Modern English uses passivization at a significantly higher frequency than two closely related languages with more word order flexibility, Icelandic and Early New High German. We will argue that verb-seconding (V2) languages employ passivization at a lower frequency because they may accomplish the same information structural goals using other methods. Finally, we will show evidence from Old and Middle English that further supports this claim. **Passives in the New Testament** This paper uses 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). We were able to compare the frequencies of passivization across all three texts for the Gospel of John. The rate of passivization in the PPCEME is significantly higher than the frequencies in either the ENHG or IcePaHC translations of John, while the difference between the ENHG and IcePaHC is not statistically significant (Table 1). Furthermore, we compared the PPCEME and IcePaHC translations of Acts, which shows the same effect (Table 2). We note that there is a significant difference between the frequency of passives in John and Acts for both texts,[1] which seems to suggest that some stylistic difference exists between the two text samples; but nonetheless, the crosslinguistic pattern is clear. Finally, we compared the rates of passivization in the full PPCEME and IcePaHC corpora, thus confirming that the generalization holds across texts (Figure 3). These results demonstrate conclusively that Early Modern English passivizes at a different rate than both Icelandic and Early New High German. **A closer look at the data** 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 essentially identical discourse goals, we can thus consider the different choices made in each language for a single utterance. We found that in the PPCEME and ENHG translations of John, there were 54 tokens that were translated as a passive in the PPCEME, and as a non-passive in the ENHG corpus. It is known that the Tyndale bible was influenced by the Luther bible, which suggests that the author frequently chose passives to translate non-passives from the German text. In contrast, there were only 16 ENHG passives that corresponded to a non-passive in the PPCEME. Interestingly, the subject position seems to have been highly relevant in this process: of the 54 examples described above, 41 (77%) are constructed such that the tokens share the same structural subject, despite the passive/active distinction. Thus, there appears to be a force of 'subject preservation' between texts which overrides the choice of larger structure. If we look more closely at the examples that demonstrate congruent subjects, we find several patterns emerging: in seven of the 41 cases (17%), a reflexive is translated as a passive in the English. In an additional 11 cases (27%), the ENHG has an instance of the verb *werden* ('become'). We will argue that German is distinguished from English because it has more options in the active voice, and we will consider how this fact relates to the V2 nature of the language. We will extend this argument to Icelandic as well. **Passives in Old and Middle English** Further evidence for our claim can be found by comparing parallel translations of a second text, the Rule of St. Benedict. The translations available are in the York-Toronto-Helsinki Parsed Corpus of Old English (YCOE) and the Penn Parsed Corpus of Middle English (PPCME2). The Middle English translation of this text is well-known for displaying Icelandic-like V2 patterns particular to Northern Middle English texts (Kroch et al. 2000), while we expect the Old English translation to demonstrate the V2 orders known to exist in Old English (Pintzuk 1991). And in fact, we find that these two texts passivize at near-identical frequencies (Table 4).[2] In the final version of this paper, we will compare these results to a translation of the same text from a Southern Middle English dialect, to complete the pattern. **Conclusion** In this paper,

[1] For PPCEME, Chi square = 24.376, $p \approx 0$; for IcePAHC, Chi square = 11.7438, $p = 0.0006$.

[2] Note that this frequency of passivization is higher than those seen in the Icelandic and German texts; we suspect that this is a stylistic effect of the text being considered, as was seen to exist in Acts.

we use a simple hypothesis to quantitatively explore the information-structural properties of three related languages. Our results demonstrate the autonomy of syntax from information structure: different syntactic operations can be used with the same information structural goals in mind. We support this with data of a new and crucial kind: general information structural patterns cannot be explored in such depth without extensive quantitative evidence. This is made possible by parsed corpora like those used in this study.

	Passive	Active	Freq. Passive
PPCEME	146	914	0.138
ENHG	108	1071	0.092
IcePaHC	86	1100	0.073
EME vs. ENHG: Chi-square = 11.3571, $p = 0.0008$			
EME vs. Icelandic: Chi-square = 25.008, $p \approx 0$			
ENHG vs. Icelandic: Chi-square = 2.6136, $p = 0.1059$			

Table 1. The frequency of passives in three translations of John.

	Passive	Active	Freq. Passive
PPCEME	186	636	0.226
IcePaHC	96	714	0.119
Chi-square = 32.394, $p \approx 0$			

Table 2. The frequency of passives in two translations of Acts.

	Passive	Active	Freq. Passive
PPCEME	17264	40108	0.301
IcePaHC	1437	8661	0.142
Chi-square = 1077.378, $p \approx 0$			

Table 3. The frequency of passives in the full EME and Icelandic corpora.

	Passive	Active	Freq. Passive
OE	348	871	0.285
Northern ME	218	514	0.298
Chi-square = 0.2806, $p = 0.5963$			

Table 4. Passives in OE and Northern ME translations of the Rule of St. Benedict.

References

A. Kroch, A. Taylor, and D. Ringe. The middle english verb-second constraint. *Textual parameters in older languages*, page 353, 2000.

Anthony Kroch and Ann Taylor. Penn-Helsinki Parsed Corpus of Middle English, 2000.

Anthony Kroch, Beatrice Santorini, and Ariel Diertani. Penn-Helsinki Parsed Corpus of Early Modern English, 2004.

Caitlin Light. Parsed Corpus of Early New High German, 2011.

Susan Pintzuk. *Phrase structures in competition: Variation and change in Old English word order*. PhD thesis, University of Pennsylvania, 1991.

Ann Taylor, Anthony Warner, Susan Pintzuk, and Frank Beths. The York-Toronto-Helsinki Parsed Corpus of Old English Prose. University of York, 2003.

Enric Vallduvi. *The Informational Component*. PhD thesis, University of Pennsylvania, 1992.

J. C. Wallenberg, A. K. Ingason, E. F. Sigurðsson, and E. Rögnvaldsson. Icelandic Parsed Historical Corpus (IcePaHC). Version 0.3. Size: 260 thousand words., 2011. URL `http://www.linguist.is/icelandic_treebank`.