The need to extend linguistic annotation in a historical corpus: the case of the PPCME2 and contact-induced change

Carola Trips

University of Mannheim
ctrip@mail.uni-mannheim.de

This paper takes research in contact-induced change on the level of argument structure as a textbook example to show that linguistic annotation of historical corpora and preservation of philological information concerning the origin of texts is inevitable to gain valid quantitative results, and concomitantly, crucial insights into processes of copying and language change.

In the framework of the project “Borrowing of Argument Structure in Contact Situations” (BASICS) which investigates to what extent copied Old French verbs (OF; ca. 880-1320) had an impact on the grammatical system of Middle English (ME; ca. 1150-1500), we have enhanced the Penn-Helsinki-Parsed Corpus of Middle English (PPCME2) with (1) lemma information of verbs, (2) information about the origin of the verbs (French or non-French), and (3) information about the origin of the ME texts (French or non-French). Further, we have created the BASICS Toolkit web application where the created resources and associated tools are available (Percillier, 2018). Only the extension of the existing annotation in the ME corpus has allowed us to pursue our main objectives which are to define (1) the spread of argument structures from copied OF verbs to ME verbs; (2) semantic verb classes of ME; (3) the spread of syntactic constructions on the model of OF; (4) language contact through translation (cf. Taylor (2008), Haeberli (2018)). The following two examples serve well to demonstrate not only the advantages but also the necessity to factor in this kind of philological information. Trips and Stein (2018) and Stein et al. (Forthcoming) show that the recipient passive in the history of English starts to occur first with verbs of French origin (for statistical details see Trips and Stein 2018, see also Allen 1995) bringing along a structural property (structural case expressing the RECIPIENT) that was not part of the English system. This seems to be an instance of contact-induced change. We could obtain this result only by systematically adding the information of the origin of the verbs in the corpus. The second case relates to language contact through translation. Trips and Stein (Forthcoming) found a rise in frequency of PPs expressing the RECIPIENT of the verb yeven ‘give’ in ME which shows a peak in the period of M3 (1350-1420) where contact was most intense. Closer inspection of the texts reveals that this peak is due to their origin: ME texts based on an OF original display higher frequencies than those where such a base can be excluded. This effect has also recently been found by Haeberli (2018) investigating object pronoun placement. Without taking this distinction into account, the quantitative results would be skewed.

In the talk I will provide motivation for the steps we took to extend the annotation of the corpus and present the two cases exemplified above in further detail focussing on the necessity of including philological information concerning the origin of verbs and texts.

References


Trips, C. and Stein, A. Forthcoming. “Contact-induced changes in the argument structure of Middle English verbs on the model of Old French”. In Journal of Language Contact. Special Issue on Valency and Transitivity in Contact, E. Grossman, I. Serzant and A. Witzlack-Makarevich (eds), Leiden: Brill.