

Scott Grimm, Joan Bresnan (Stanford, California/USA)

Spatiotemporal variation in the dative alternation: a study of four corpora of British and American English

Abstract

Recent work on constructions involving choice between different forms, such as the dative alternation exemplified in (1), has emphasized that multiple variables, such as argument length and pronominality, contribute to the speaker's choice between forms (see Bresnan et al. (2007) *inter alia*).

- (1) a. Who gave that watch to you? (prepositional dative)
b. Who gave you that watch? (double object construction)

This study, employing state-of-the-art tools from NLP parsing technology, tracks shifts in usage patterns in terms of such variables across distinct but comparable corpora culled from different speaker populations and time periods.

Corpora and Methodology

Our study of the dative was based on the "Brown family" of corpora, following Hinrichs and Szmrecsanyi (2007) on the English genitive alternation, which consists of 4 corpora differing in the variety of English spoken and the time period of the sampling: Brown (US, 1960), LOB (British, 1960), Frown (US, 1990), FLOB (British, 1990). Changes in the use of the two constructions were examined against a set of variables: length of arguments (in words), pronominality and thematicity (total occurrences of head noun in entire document).

We provided a further methodological twist by automating the extraction process utilizing NLP parsing tools, namely the Klein-Manning parser (Klein and Manning 2003; de Marneffe et al. 2006). As input to the parser we used versions of the four corpora which had already been tagged for part-of-speech and corrected (developed by the Universities of Lancaster and Freiburg). The parser outputted rich syntactic information, including typed dependencies, which were exploited to extract the prepositional and double object constructions in a far more efficient manner than, say, shallow-parsed corpora permit. These data were manually filtered and automatically labeled for the variables of interest. We further conducted a preliminary error analysis and created a "gold standard" sample to evaluate this method's performance.

Results

We modeled the data with logistic regression techniques and observed a consistent set of changes across the different corpora. Across the US corpora, main effects of thematicity of recipient and length and pronominality of recipient and theme are all harmonically aligned with construction choice, i.e. values indicative of higher prominence aligned with the more prominent syntactic position, and conversely. Additionally, there is significantly more use of the double object construction in the 1990's than in the 1960's, related to greater use of lexical NP recipients and thematic themes in the double object construction. Across the UK corpora, main effects of thematicity, length, and pronominality of recipient and theme also show harmonic alignment with construction choice. There is also a significant change toward greater use of the double object construction, in parallel to the US corpora.

This paper demonstrates how regularities of syntactic change can be observed across a set of large corpora in an efficient manner via exploiting NLP tools, leading to a model of probabilistic changes in linguistic choices across space and historical time.

References

- Joan Bresnan, Anna Cueni, Tatiana Nikitina, Harald Baayen (2007). *Predicting the Dative Alternation*. In: Cognitive Foundations of Interpretation, ed. by G. Boume, I. Kraemer, J. Zwarts. Amsterdam: Royal Netherlands Academy of Science, pp. 69-94.
- Marie-Catherine de Marneffe, Bill MacCartney, Christopher D. Manning (2006). *Generating Typed Dependency Parses from Phrase Structure Parses*. In: LREC 2006.
- Lars Hinrichs, Benedikt Szmeccsanyi (2007). *Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora*. English Language and Linguistics 11(3), pp. 335-378.
- Dan Klein, Christopher D. Manning (2003). *Accurate Unlexicalized Parsing*. ACL 2003, pp. 423-430.