Sara Budts - UAntwerp

Mind Bending Grammars innovatively combines language change research with research on the individual mind. In the past linguistics has often treated change as happening to an abstract object 'language'. But it is the minds of actual people that change language. Mind-Bending Grammars contributes to our understanding of the adaptive powers of adult cognition. The project specifically aims at making a breakthrough in two key issues in linguistic theory by tracing with the utmost detail step-by-step changes in grammatical constructions:

1. How much innovation is possible in the adult cognitive system? Where does change halt only to be continued by a next generation?
2. Are changes in adult cognition one-off events or are they connected? How strong do connections need to be before structures start coevolving as a group?

Studies on language change in living adults so far have been small-scale. Data for a single informant rarely go beyond the size of one short novel, which would only produce a handful of examples of relevant grammatical structures. To solve this, the novel data the project exploits come from the minds of dead people. Specifically, Mind-Bending Grammars compiled the 90 million word corpus EMMA (Early Modern Multiloquent Authors). EMMA is a sample of 50 of the most productive writers who were born in the 17th century, mostly taken from the London-based elite. The corpus draws on massive digitalization projects such as EEBO and 18thConnect, which cover all available English printed matter between 1473-1800. The sheer size of EMMA makes it possible to trace grammatical developments in detail across single lifetimes as well as five different generations. Because of the importance that is attached to individual change, it is vital that the sociohistorical contextualisation of our metadata is as accurate as possible (e.g. authorship check, date of writing rather than date of publication,... ). In the first part of my talk, I will address how we dealt with the trade-off between this need for sociohistorical accuracy and the practical benefits of (semi)automatic processing.

The second part of my talk will be concerned with my own research project. As a spin-off of Mind Bending Grammars, my project investigates the rise of 'do' as an auxiliary in 17th century English (1).

(1) Do you like ice cream? - I do not like it.
(2) Will the bus arrive on time? - I would not bet on it.

Usage-based strands of linguistics have tentatively attributed the rise of auxiliary 'do' to analogical influence of the modal auxiliaries (e.g. 2), which emerged as a category about 50 years prior to the spread of 'do' and typically occurred in the same syntactic contexts. I look into the potential influence of the modal auxiliaries on auxiliary 'do' by charting their paradigmatic relations throughout the Early Modern period in a semi-automated way. In the first stage of the analysis, I have generated word embeddings to find out if the distributional profiles of 'do' and the modal auxiliaries converge throughout the 17th century. During the second stage of the analysis, I intend to look into the reason for this convergence by generating vector representations of tokens-in-context. While the first stage of the analysis has been carried out, the second stage has proven to be especially challenging from a methodological perspective. In the second part of my talk, I will address these challenges as well as the benefits and drawbacks of a couple of strategies to overcome them. These strategies include the generation of word-level LSTM representations, character-level LSTM representations, and convolutional neural networks.