TRACING REAL-LIFE AGENTS' INDIVIDUAL PROGRESS IN ONGOING GRAMMATICALIZATION

PETER PETRÉ

FWO / University of Leuven, Blijde Inkomststraat 21, P.O. Box 3308 Leuven, 3000, Belgium

FREEK VAN DE VELDE

FWO / University of Leuven, Blijde Inkomststraat 21, P.O. Box 3308 Leuven, 3000, Belgium

We present a longitudinal corpus-based analysis of 15 authors writing in the 17th and 18th century in an ongoing grammaticalization process (c.q. the auxiliarization of *be going to*). Our aim is to arrive at a fine-grained analysis of the micro-changes involved in this kind of language change. In this way, we hope to help bridge the gap between the agent-based modelling and the more traditional grammaticalization studies.

1. Intro: Breaking down the aggregate view on grammaticalization

Investigating individual differences in language behaviour by looking at non-elicited 'naturalistic' data has recently been made easier by the increasing availability of large-scale corpora, especially for English (Barlow 2013). Recently, some interesting diachronic corpus studies in this field have been carried out (Nevalainen et al. 2011; De Smet, ms.), to arrive at the constraints individual variation is subjected to. These studies, however, do not take a longitudinal perspective, in which individuals are followed through time, to see how they shift their behavior, accommodating to or diverging from particular ongoing changes. The few longitudinal studies that we have (Bergs 2005; Raumolin-Brunberg 2009; Hendriks 2013), are typically small-scale. The present study tries to combine the longitudinal approach with large-scale corpus analysis. We present longitudinal individual data on what is perhaps the most iconic of grammaticalization cases: the rise of *be going to* as a marker for future in English. We make use of the large-scale EEBO corpus, to see how individual

languages users behaved in the seventeenth and eighteenth centuries, a crucial period in the evolution of *going to*.

Breaking down the aggregate view on the grammaticalization of *be going to* into individual users' behavior may help bridge the gap between 'traditional' diachronic linguistics and agent-based modeling. Agent-based models (Steels 2011) are able to show how emergent properties of language structure arise from well-defined individual interactions (Landsbergen et al. 2010; Beuls & Steels 2013), but are sometimes criticized for the allegedly artificial nature of the communicative setting. On the other hand, traditional corpus-based diachronic linguistics often fail to specify the precise conditions of naturalistic settings between real-life agents partaking in ongoing language changes. At present, it is debated whether adults, adolescents or children are the main instigators of language change. Some scholars argue that language change primarily happens over generations (e.g. Lightfoot 1999), while others argue that it takes place during lifetime (Croft 2000; Bergs 2005).

2. Methodology

2.1. *Corpus description and data extraction*

In order to examine if micro-steps in the grammaticalization (or grammatical constructionalization Traugott & Trousdale 2013) of *be going to* occur within real-life agents' lifetimes, we selected 15 prolific authors from EEBOCorp 1.0 (Petré 2013), a half billion+ corpus based on the EEBO-database (eebo.chadwyck.com), containing English books printed between 1473-1700. Selection criteria were: (i) Sufficient material is available for the first and second halves of writer's careers; (ii) Constant register over time; (iii) Writers are from roughly the same social status. Posthumously published works in EEBO not included in EEBOCorp 1.0, and translations done by one of the selected authors were also included. The post-1700 output of Burnet, D'Urfey and Dunton was added from the *Eighteenth Century Collections Online* database (ECCO).

The resulting corpus consists of about 31 million words, with individual author word counts ranging between *ca.* 300,000 and 10,000,000 words. All forms of *going* were extracted from this corpus by means of Perl scripts (n = 5821), taking into account variant forms identified in an exhaustive token list.

Additional scripts and manual analysis was used to filter out a total of 1591 instances of *be going*.

2.2. *Data coding and analysis*

We coded the EEBO datapoints for several formal and semantic features that are commonly associated with the grammaticalization of *be going to*, and can serve as diagnostics to assess the level of grammaticalization reached in a particular individuals, which serves as the dependent variable in our inquiry. Each of these features is analyzed with a level of granularity that allows us to pick up small increments in the level of grammaticalization. In the analysis, we both looked at the behaviour of each feature separately, and at their combined value, by computing a summative measure of grammaticalization. For each of the authors, we divided the collected data in half, to arrive at two categories 'earlier work' and 'later work', in order to check whether differences occurred through the years.

3. Findings

The scatterplot in the left panel in Fig. 1 brings out the aggregate view on grammaticalization: the score on the Y-axis is a summative measure of how many grammaticalization features a certain datapoint displays. The regression line (lowess) has an s-shaped curve, typical of language change. The rise is significant (Kendall tau = 0.126, $p < 0.0001$ – the relatively weak effect size is not surprising, considering that we only look at a time window of 50 years). The right panel breaks the data down into the two periods for each author. Authors with an increased grammaticalization score in their later work are indicated in red. As can be appreciated, they form the majority of the individuals investigated. Overall, we see an increase in grammaticalization scores through time (lowess regression line). In our paper, we will investigate the differences between the authors in depth.

References

Barlow, M. (2013). Individual differences and usage-based grammar. *International Journal of Corpus Linguistics*, *18*, 443-478.

Bergs, A. (2005). *Social networks and historical sociolinguistics: studies in morphosyntactic variation in the Paston letters (1421-1503)*. Berlin: Mouton de Gruyter.

Beuls K. & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLoS ONE*, *8*, 3, e58960.

Croft, W. (2000). *Explaining language change. An evolutionary approach*. Harlow: Longman.

De Smet, H. (manuscript). How gradual change progresses. The expansion of -ing-clauses with begin through time and across individuals.

Hendriks, J. (2013). Stability of idiolects in unstable times. Life stages and lifespan changes of immigrants in the early modern Dutch urban context. Paper presented at *ICHL*-21, Oslo, 5-9 August 2013.

Landsbergen, F., Lachlan, R. Ten Cate, C. & Verhagen, A. (2010). A cultural evolutionary model of patterns in semantic change. *Linguistics*, *48*, 363-390.

Lightfoot, D. (1999). *The development of Language. Acquisition, change, and evolution*. Oxford: Blackwell.

Nevalainen, T., Raumolin-Brunberg, H. & Mannila, H. (2011). The diffusion of language change in real time: progressive and conservative individuals and the time depth of change. *Language Variation and Change*, *23*, 1-43.

Petré, P. (2013). *EEBOCorp 1.0*.

Raumolin-Brunberg, H. (2009). Lifespan changes in the language of three early modern gentlemen. In: A. Nurmi, M. Nevala & M. Palander-Collin (Eds.), *The Language of daily life in England (1400-1800)* (pp. 165-196). Amsterdam: Benjamins.

Steels, L. (2011). 'Modeling the cultural evolution of language'. *Physics of Life Review*, *8*, 339-356.

Traugott. E. & G. Trousdale (2013). *Constructionalization and constructional changes*. Oxford: Oxford University Press.

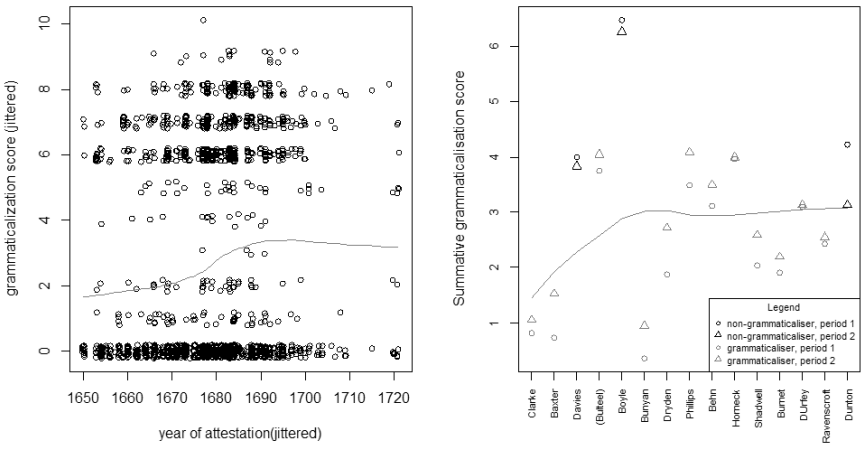

Figure 1: Grammaticalization of *be going to* in EEBO