ISSN: 2088-6799

PROCEEDINGS

International Seminar

LANGUAGE MAINTENANCE AND SHIFT IV

November 18, 2014

Master Program in Linguistics, Diponegoro University
in Collaboration with
Balai Bahasa Provinsi Jawa Tengah

Proceedings
International Seminar
Language Maintenance and Shift IV

November 18, 2014

xviii+488 hlm. 21 x 29,7 cm

ISSN: 2088-6799

Editors:
Zane Goebel
J. Herudjati Purwoko
Suharno
M. Suryadi
Yusuf Al Arief

Master Program in Linguistics, Diponegoro University
in Collaboration with
Balai Bahasa Provinsi Jawa Tengah
Jalan Imam Bardjo, S.H. No.5 Semarang
Telp/Fax +62-24-8448717
Email: seminarlinguistics@gmail.com
Website: www.mli.undip.ac.id/lamas

VISUALISATION OF DIACHRONIC CONSTRUCTIONAL CHANGE USING MOTION CHART

Gede Primahadi Wijaya R.
Universitas Udayana, Denpasar
primahadiwijaya@gmail.com

Abstract

In conjunction with the increasing availability of massive computerised diachronic corpora, advances in computational visualisation apparatus have further equipped diachronic corpus linguists with an array of elegant analytical tools to study and model diachronic constructional changes, which are often times too intricate to be detected through a mere eyeballing on tabular numerical data. Based on the Corpus of Historical American English (COHA), this paper exemplifies the application of one such tool called Motion Chart to visualise recent change in two English Future Constructions, i.e. will+INF and be going to+INF, regarding their infinitival collocates distribution.

***Keywords**: diachronic corpora, COHA, diachronic corpus linguistics, English future constructions, constructional change, collocational change, linguistic motion charts*

Introduction

Diachronic corpora have been widely used to study constructional change in a range of linguistic phenomena (e.g. Hilpert, 2013). While frequency information offered by a diachronic corpus indeed lays a strong foundation for empirically quantitative analysis on how particular linguistic structures change over time, it would become very taxing to grasp such change in an immediate and intuitive way through merely looking at tabular numerical data. The present paper exemplifies the application of an exploratory tool called *motion chart* (Gesmann & de Castillo, 2014; Heylen, Wielfaert, & Speelman, 2013; Hilpert, 2011; Primahadi Wijaya R. & Rajeg, 2014, *inter alia*) that offer a dynamic and condensed visual representation of complex linguistic change into which historical linguists can gain insights in a more effective and holistic fashion.

Motion chart is essentially generated from "a series of diachronically ordered scatterplots" (Hilpert, 2011, p. 435) (see Figure 1 below), whose plotted data points representing the same linguistic phenomenon move around showing how it changes as the temporal reference develops diachronically. As a case study, a recent change in distribution of the infinitival collocates of two English Future Constructions (henceforth EFCx), i.e. *will+INF* and *be going to+INF*, is presented. This study adopts the definition of "constructional change" as proposed by Hilpert (2013, p. 2) in that "Constructional change selectively seizes a conventionalized form-meaning pair of a language, altering it in terms of its form, its function, its frequency, its distribution in the linguistic community, or any combination of these". The form-meaning pair herein refers to the two EFCx whereas the distributional shift of their infinitival collocates may reflect change in the functional or semantic aspect of the two constructions, and, with the help of a motion chart, can be simultaneously indicative of several changes related to the frequency aspects of the constructions (see Hilpert, 2013 for more details).

The diachronic development of the two EFCx have been indeed well-studied (see Hilpert, 2008; Hopper & Traugott, 2003; Mair, 2006 to name but a few). The present paper attempts to contribute to that body of literatures by introducing a graphical tool, through which the various features of constructional change in the EFCx can be explored holistically, and which can thus be extended to further studies on diachronic constructional change. Moreover, this endeavour is framed within broader context of a growing interest in the state-of-the-art visualisation paradigms as well as in the cutting-edge computational tools in corpus linguistic studies (AVML, 2014).

Data and method

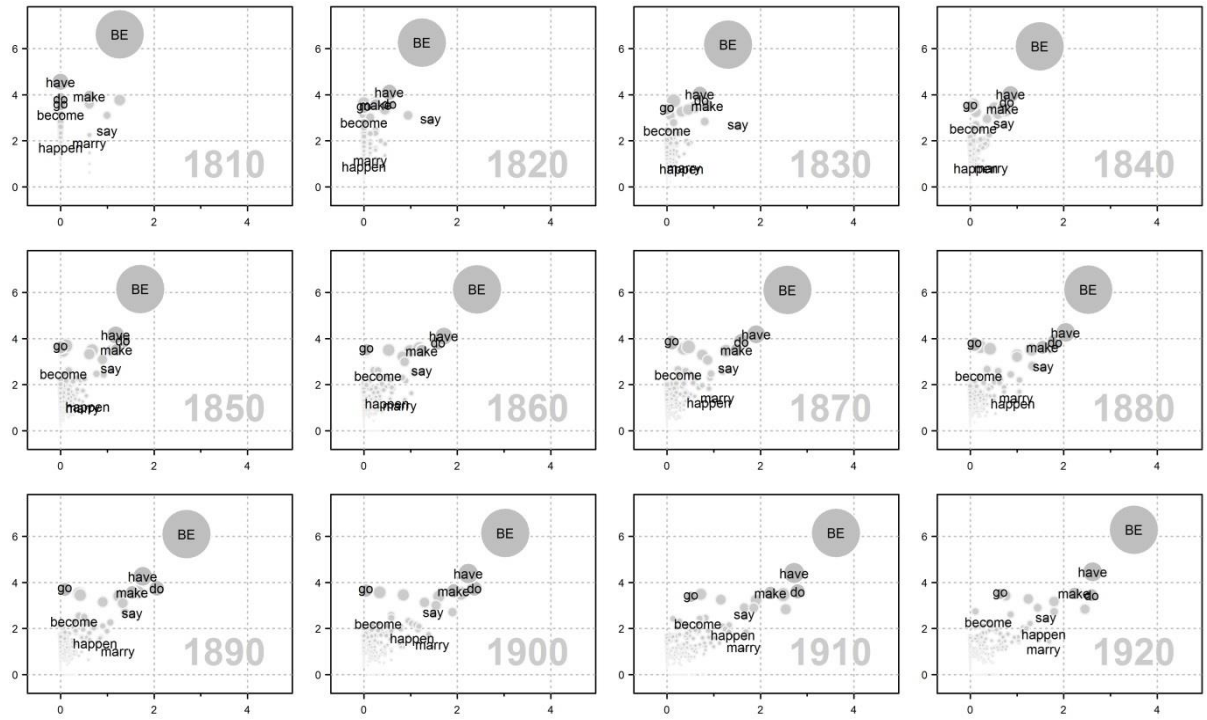
All the data is retrieved from the *Corpus of Historical American English* (COHA) (Davies, 2010), dating from the 1810s up to the 2000s. The co-occurrence frequency lists of the top-200 infinitival complements of the respective EFCx over the 20-decade periods were simultaneously retrieved using the "list" feature in COHA. The two separate frequency lists were then combined using a small R program script (R Core Team, 2014) into a data-frame format as shown in Table 1 required as input for generating the motion chart (Figure 1). Each row of the complete data table contains all dataset of the infinitival collocates for further decades until the 2000s. This study normalised co-occurrence frequency of the infinitives with the constructions into frequency per million words.

Table 1 *Subset of data input for motion chart*

Decade	Collocates	Freq with *BE going to+INF* (*x*-axis values)	Freq with *will+INF* (*y*-axis values)	Joint Freq
1810	accept	0	1.693448556	1.693448556
1810	act	0	3.386897111	3.386897111
1810	add	0	5.080345667	5.080345667
1810	admit	0	11.85413989	11.85413989
…	…	…	…	…

Results and discussion

Figure 1 shows the result of plotting the quantitative information in Table 1 into the so-called *bubble plot*, a special version of a scatterplot[10]. The *x*- and *y*-axes in each graph represent the frequency of the infinitives with *BE going to* and *Will* respectively. Both axes are scaled in logarithmic units for better visibility and understanding of changes occurring. Then, each bubble symbolises all the infinitives and different bubble sizes reflect the joint frequency of the infinitives with the EFCx.



[10] The sequential bubble plots are generated with the open-source statistical programming language R (R Core Team, 2014). The data file and R program script to reproduce the graph in Figure 1 is available from the author upon request. This paper also provides the on-line, animated version of the graph at http://primahadiwijaya.blogspot.com/2014/09/motion-chart-for-futurate-constructions.html

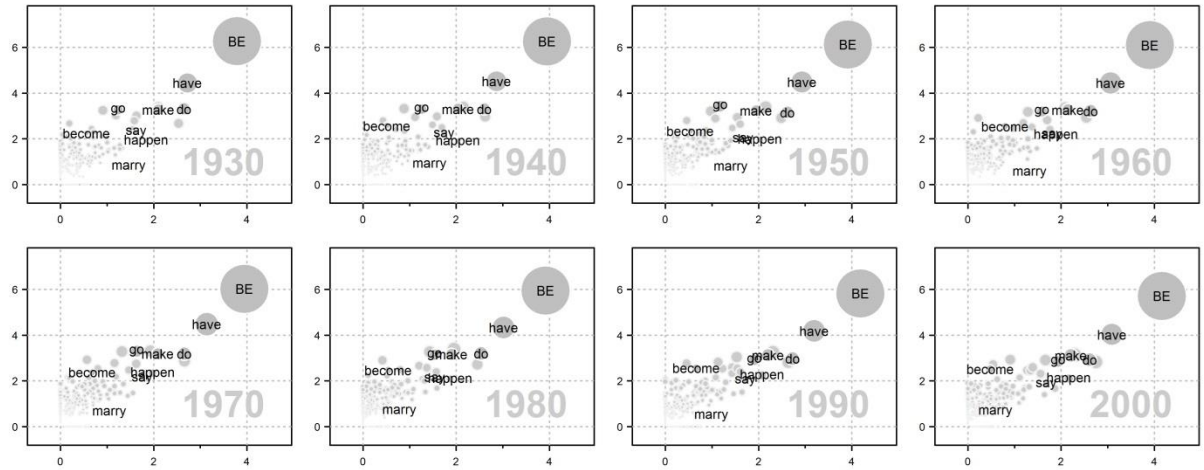


Figure 1 *Changes in the distribution of the infinitival collocates across the two English Future Constructions*

Only nine out of total 263 infinitive types co-occurring with both or one of the EFCx that are labelled in the graph, namely *be*, *become*, *do*, *go*, *happen*, *have*, *make*, *marry*, and *say*.

By looking at Figure 1, it is apparent that stative verb *be* is the most frequent infinitival collocates of the two EFCx across the twenty-decade periods. Moreover, the graph also shows that most infinitives clustered together more closely to the *Will* axis at the beginning of the periods. However, as the time unfolds, a main trend becomes clearly visible, especially with regard to *BE going to* construction. Seen from a bird's-eye perspective, the overall shape of the bubbles has changed quite markedly over time. It is roughly more noticeable after the 1830s, where quite a great deal of infinitive bubbles, including the light verbs, such as *be*, *have*, *do*, as well as telic and dynamic verb, namely *make* (cf. Hilpert, 2008, pp. 119–122), have gently drifted towards the *x*-axis, albeit having their co-occurrence frequency with *Will* remained roughly stable. The other labelled verbs, especially *go* and *happen*, follow this trend only a bit later. It is approximately from the 1850s and the 1910s onwards for *happen* and *go* respectively. By the 2000s, this development results in a diagonal cloud, suggesting that *BE going to* construction gradually expands its usage context, as well as exhibits a progressively rising text frequency, as reported in the previous studies (Hilpert, 2008, 2013, p. 4; Mair, 2006, p. 97 *inter alia*). Another labelled telic verb on the graph, i.e. *say*, exhibits a roughly stabilised trend across the periods, nearly similarly to *become* and *marry*, even though along the way, the latter also show a modest frequency increase with *BE going to* in an unstable fashion.

These developments point to a number of features of constructional change that can be visually discerned, particularly in the *BE going to* construction. First, over the last 200 years, *BE going to* has accommodated a greater array of co-occurring infinitive sets, hence indicative of analogical extension of the construction (Hilpert, 2013, pp. 9–10). From the semantic point of view, it has undergone semantic change into a more schematised construction, i.e. "in a word, more grammaticalized" (Hilpert, 2008, p. 119, see also 2013; Mair, 2006, p. 96). Two other frequency-related phenomena of constructional change could further be simultaneously explored from the plot. The first one is the increasing type frequency of *BE going to* due to its wide-ranging applicability to different infinitive types, which is then reflective of the second type of change, i.e. its increased productivity (cf. Hilpert, 2013, p. 5; Mair, 2006, p. 97).

Concluding remark

This brief paper has demonstrated how the many faces of constructional change phenomena in a particular form-meaning pair can be explored more effectively with the help of a visual representation exemplified herein, i.e. motion chart. This kind of representation can aid diachronic (corpus) linguists to deal with theoretical questions; such as, when does a particular construction change? How does it change (i.e. whether it changes abruptly or more gradually; whether it has become more or less frequent over time)? When and how are particular elements integrated with the use of a particular construction? It is thus hoped that this method can find further application to explore constructional change based on historical language data other than English.

References

AVML. (2014). *Advances in Visual Methods for Linguistics*. Book of Abstracts, University of Tübingen. Retrieved from http://avml2012.files.wordpress.com/2014/09/avml_2014_bookofabstracts4.pdf

Davies, M. (2010). *The Corpus of Historical American English (COHA): 400+ million words, 1810-2009*. Retrieved from http://corpus.byu.edu/coha/

Gesmann, M., & de Castillo, D. (2014). Using the Google Chart Tools with R: googleVis-0.5.2 Package Vignette (Version 0.5.2). Retrieved from http://cran.r-project.org/web/packages/googleVis/vignettes/googleVis.pdf

Heylen, K., Wielfaert, T., & Speelman, D. (2013). *Tracking change in word meaning. A dynamic visualization of diachronic distributional semantic models*. Presented at the DGfS 2013 - Workshop on the Visualization of Linguistic Patterns, University of Konstanz, Germany. Retrieved from http://ling.uni-konstanz.de/pages/home/hautli/LINGVIS/dgfs13_heylenetal_abstract.pdf

Hilpert, M. (2008). *Germanic Future Constructions: A Usage-based Approach to Language Change*. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Hilpert, M. (2011). Dynamic visualizations of language change. Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics*, *16*(4), 435–461.

Hilpert, M. (2013). Corpus-based approaches to constructional change. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 1–13). Oxford: Oxford University Press. Retrieved from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195396683.001.0001/oxfordhb-9780195396683-e-25

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd edition.). Cambridge, UK; New York, NY: Cambridge University Press.

Mair, C. (2006). *Twentieth-Century English: History, Variation, and Standardization*. Cambridge, UK ; New York: Cambridge University Press.

Primahadi Wijaya R., G., & Rajeg, I. M. (2014). Visualising diachronic change in the collocational profiles of lexical near-synonyms. In I. N. Sudipa & G. Primahadi Wijaya R. (Eds.), *Cahaya Bahasa* (pp. 247–258). Denpasar: Swasta Nulus.

R Core Team. (2014). R: A language and environment for statistical computing (Version 3.1.0). R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Master Program in Linguistics, Diponegoro University
in Collaboration with
Balai Bahasa Provinsi Jawa Tengah

Jalan Imam Bardjo, S.H. No.5 Semarang
Telp/Fax +62-24-8448717
Email: seminarlinguistics@gmail.com
Website: www.mli.undip.ac.id/lamas