EVOLUTIONARY FORCES IN THE DEVELOPMENT OF THE ENGLISH PERFECT CONSTRUCTION

MICHIO HOSAKA[*1], SHIMPEI OKUDA[2], and KAZUTOSHI SASAHARA[2, 3]

[*]Corresponding Author: hosaka@chs.nihon-u.ac.jp
[1]Department of English Language and Literature, Nihon University, Tokyo, Japan
[2]Graduate School of Informatics, Nagoya University, Aichi, Japan
[3]JST PRESTO, Japan

1. Introduction

Both directed selection and stochastic drift are the driving forces of biological and cultural evolution, and this is also true for language evolution. The recent argument presented by Newberry et al. (2017) is that drift cannot be rejected and stochasticity has an under-appreciated role in grammatical changes in English, such as the (ir)regularization of the past-tense verb forms.

We focus on the evolution of the English perfect construction (have/be+PP (Past Participle)) and aim to detect signatures of selection and drift working there. Although the have+PP form is used only for the perfect construction in present-day English, earlier English had a variation of have/be+PP, as below.

  a. we <u>ben</u> entred into shippes bond, (Hogg & Denison, 2008)
  b. For ye <u>han</u> entred into myn hous by violence (Hogg & Denison, 2008)

The decline of the BE perfect has been argued for a long time but the exact chronology and nature of the process are still in controversy.

2. Data and Methods

We used three English Corpora—Early English Books Online (EEBO), Corpus of Historical American English (COHA), and Google Books. Here EEBO was used to cover data from the 15th to 17th centuries (755 million words), COHA for data from the 19th to 21st centuries (406 million words), and Google Books for data of 18th century (468 billion words).

From these corpora, we computed the longitudinal frequency changes of have/be+PP forms in 19 target verbs. The target verbs were selected from the verb list of Corpus of Contemporary American English (COCA) (5764 words) if it appears in these corpora more than 200 times and the verbs' transitive usage is rare as the perfect tense and the passive voice are often indistinguishable. Moreover, we tested whether directional evolution occurred in 19 target verbs by using

the Frequency Increment Test (FIT) that detects S-shaped growth. To this end, these corpora need to be comparable in size to properly set the bin size for the FIT test. Thus, we scaled the relative frequencies of the target verbs in Google Books to the frequencies of those in COHA by multiplying constant factors estimated as corresponding frequency ratios for each year as these corpora have year overlaps between 1810 and 2000. We did not scale EEBO data because there is no year overlap with other corpora and the size is approximately comparable to the size of COHA. We applied the FIT test to our corpora as in Newberry et al. (2017).

3. Results and Discussion

Fig. 1 shows that the have+PP form has become dominant over time in most verbs. For example, "become" exhibits a rapid change from 1775 to 1825. The exceptional cases of "go" and "descend" can be explained by the emergence of their adjective usages: "be gone" and "be descended." These results suggest that this auxiliary verb selection is dependent on the nature and grammatical usage of verbs. To examine whether these changes are selection or drift, we applied the FIT test as explained above. As a result, 10/19 verbs passed the FIT test ($\alpha = 0.05$), suggesting that frequency changes from be+PP to have+PP are unlikely due to random drift (i.e., directed evolution) in these verbs. Note that we cannot simply say that the verbs that did not pass the FIT test were subject to random drift; small sample size and binning procedures may affect the result (Karjus et al. 2018), thus we need further research.

We have successfully detected directional forces in the evolution of the English perfect construction. Our finding, however, needs to be further validated by making a complete list of target verbs, testing with various binning methods because several issues have been pointed out (Karjus et al. 2018), and analyzing other available corpora. If these issues are resolved, the evolutionary perspectives given here may shed a new light on grammatical changes in language evolution.

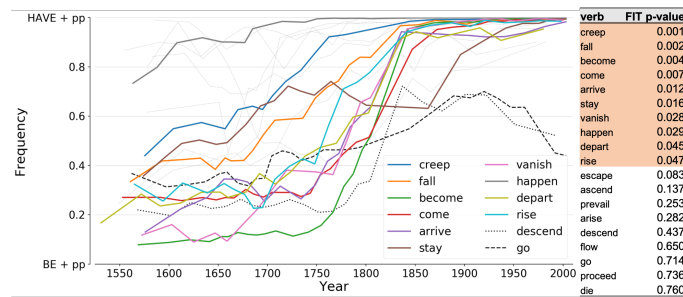| verb	FIT p-value
creep	0.001
fall	0.002
become	0.004
come	0.007
arrive	0.012
stay	0.016
vanish	0.028
happen	0.029
depart	0.045
rise	0.047
escape	0.083
ascend	0.137
prevail	0.253
arise	0.282
descend	0.437
flow	0.650
go	0.714
proceed	0.736
die	0.760

Figure 1. Relative frequencies of (have+PP) / ((have+PP) +(be+PP)) in 19 verbs (left). Result of FIT-test, in which 10 verbs that rejected the null hypothesis ($\alpha = 0.05$) are highlighted (right).

Acknowledgements

References

Hogg, R., & Denison, D. (2008). A history of the English language. Cambridge University Press.

Karjus, A., Blythe, R. A., Kirby, S., & Smith, K. (2018). Challenges in detecting evolutionary forces in language change using diachronic corpora. arXiv preprint arXiv:1811.01275.

Newberry, M. G., Ahern, C. A., Clark, R., & Plotkin, J. B. (2017). Detecting evolutionary forces in language change. Nature, 551(7679), 223.