Supplementary material for Detecting directional forces in the evolution of grammar: A case study of the English perfect with intransitives across EEBO, COHA, and Google Books

Shimpei Okuda[1], Michio Hosaka[*2], and Kazutoshi Sasahara[†3]

[1]Graduate School of Informatics, Nagoya University, Nagoya, Japan

[2]Department of English Language and Literature, Nihon University, Tokyo, Japan

[3]School of Environment and Society, Tokyo Institute of Technology, Tokyo, Japan

Corresponding authors:

[*] hosaka@chs.nihon-u.ac.jp

[†] sasahara.k.aa@m.titech.ac.jp

1 *be*/*have* and PP forms used for search

Tables 1–3 summarize all the forms for *be*/*have* and for PP, respectively, used in our study.

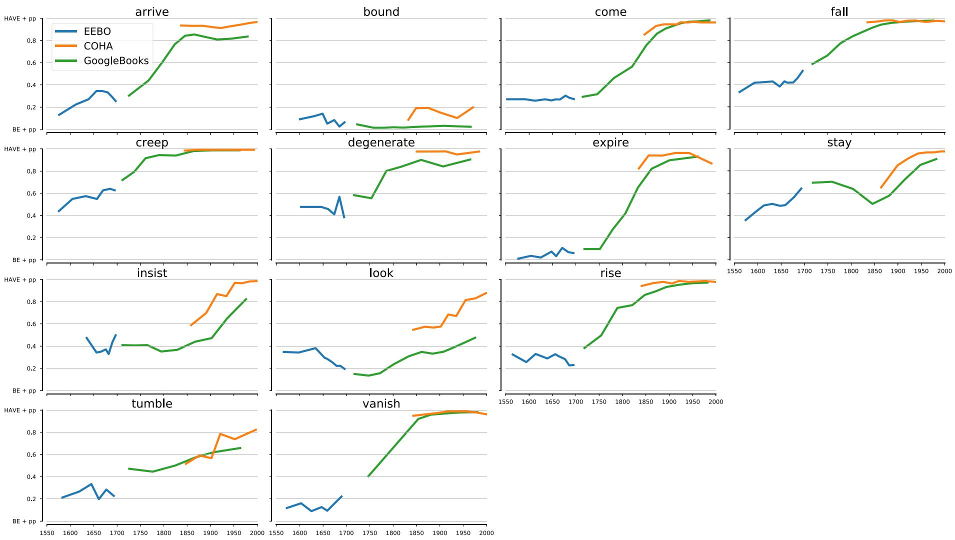

Figure S 1: Frequency changes of *be*/*have*+PP by dataset in Group A

2 Historical frequency changes of *be*/*have*+PP by dataset

Figure 1 shows the historical changes in the frequency of *be*/*have*+PP regarding the 13 intransitive verbs (Group A) selected on the basis of frequency in EEBO, COHA, and Google Books, respectively. Figure 2 shows the same analysis on the six intransitive verbs (Group B) analyzed in prior studies and which fell outside the selection criteria for Group A. These figures indicate that the scaling and binning used here are valid.

Table S 1: *be*/*have* forms used for search

lemma	word	usage
be	be	be, ybe
	am	am, ’m
	are	are, ’re
	is	ys, iss
	was	was, vvas, wes, wass, wast, vvast
	were	were, wer, vvere, weere, vveere, weren, wert, werst, vvert
have	have	have, ’ve, hast, hauest, havest, haue, hav, haf, haif
	has	has, haues, haves, haues, haueth, haveth, havethe, hath, hathe
	had	had, hadde, hadd, haddyst, hade, hadst, haddest, hadest, haddeste, hadste, haddst, hadyste, haved, haued, havd, havid, hauyd, hauid

Table S 2: PP forms used for search (Group A)

lemma	past participle
arrive	**arrived**, arriued, arriv'd, arryued, ariued, arived, arriu'd, aryued, ariu'd
bound	**bound**, bounded, bounde, bownd, boud
come	**come**, came, cum, com, comen, coome, coomen, comest, cumme, comm, commen
creep	**crept**, crepte, creeped
degenerate	**degenerated**
expire	**expired**, expir'd, expyred, expired
fall	**fallen**, falne, faln, fall'n, fell, fal'n, falne, fel, felle
insist	**insisted**
look	**looked**, look'd, lookt, loked, look't, lookte, lookd, lokyd
rise	**risen**, rysen, rose, rised
stay	**staied**
tumble	**tumbled**, tumbl'd
vanish	**vanished**, vanish'd, vanisht, vanysshed, vanish't

*Bold represents the current form.

Table S 3: PP forms used for search (Group B)

lemma	past participle
ascend	**ascended**, assended, ascendyd, ascendid, asscended, assendyd, assendid
become	**become**, becom, becomme, becum, bycome, beecome, bicome, becomne, becoom, becommed, becomed, becomd, beecomed, becomen, becommen
depart	**departed**, departyd, departid, depertid
descend	**descended**, discended, desended, descendyd, dyscended, decende, discendyd, deseended, dyscendyd, dessended, desendid, decendyd
escape	**escaped**, escapt, eskaped, ascaped, escapyd, eschaped, escapd, escapid
go	**gone**, gon, goon, gonne, gooun, govn

*Bold represents the current from.

3 Detection of evolutionary forces using FIT

For reference, we conducted the Frequency Increment Test (FIT) to our dataset. The result is shown in Table 4 — the nominal p-values of FIT results and the p-values adjusted by the Benjamini-Hochberg (BH) correction (called 'q-value') are listed in ascending order. Here, we selected the BH correction for multiple testing, because verbs are competing factors in the evolution of grammar and their frequency changes are unlikely to be independent. In such a case, the Bonferroni correction would be too conservative and thus increase false negatives [1].

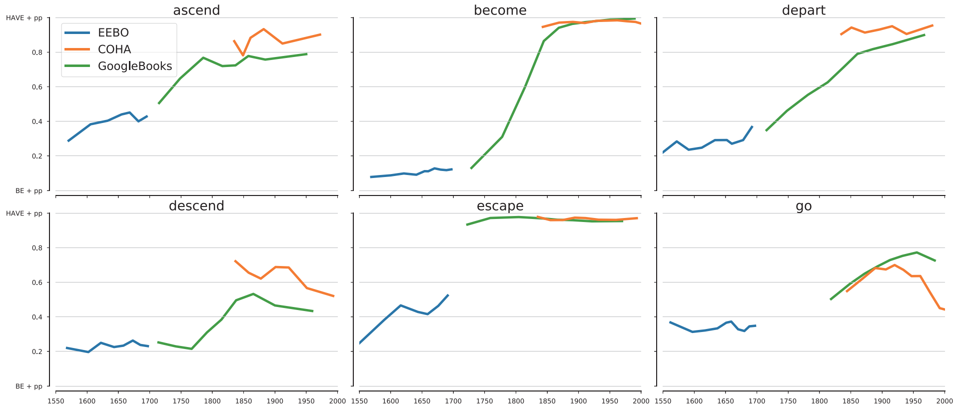

Figure S 2: Frequency changes of *be*/*have*+PP by dataset in Group B

Of the 20 verbs, two (i.e., *creep* and *fall*) rejected the null hypothesis of FIT ($\alpha = 0.05$). In other words, the probability of random drift from *be*+PP to *have*+PP for these two verbs is highly unlikely ($q < 0.05$), which strongly suggests that nonrandom selection was involved.

However, as shown in Table 4, our post-hoc power analysis revealed that the estimated power is below 0.8 for all verbs, indicating that the data size was insufficient for FIT, even though we used the largest historical data sources [2]. Therefore, we might not be able to detect an effect even if it existed.

As mentioned in the body, prior studies have pointed out the problems with FIT. Karjus et al. (2020) reported that the FIT results on the frequency time series can be sensitive to how the corpus is organised into temporal segments (i.e., banning), although the overall observation by Newberry et al. (2017) was successfully replicated. Karsdrop et al. (2020) also demonstrated some problems of the FIT, including the instability on binning and distorted frequency time series, and proposed a deep neural network model-based classification (called the neural TSC) to solve it. Therefore, we used the TSC for our analysis.

Table S 4: Results of FIT and a post-hoc power analysis for verbs in Groups 1 and 2.

Verb	Group	FIT p-value	BH q-value	Effect size	Power
creep	1	0.003	0.029	0.750	0.660
fall	1	0.003	0.029	0.608	0.624
stay	1	0.013	0.063	0.486	0.457
come	1	0.013	0.063	0.552	0.448
become	2	0.016	0.063	0.548	0.444
arrive	1	0.024	0.080	0.471	0.385
vanish	1	0.029	0.080	0.516	0.372
depart	2	0.032	0.080	0.479	0.356
expire	1	0.046	0.103	0.465	0.313
rise	1	0.073	0.146	0.368	0.256
escape	2	0.091	0.166	0.369	0.232
ascend	2	0.102	0.169	0.386	0.221
tumble	1	0.213	0.328	0.315	0.145
look	1	0.242	0.346	0.231	0.132
insist	1	0.277	0.369	0.233	0.121
bound	1	0.310	0.388	0.228	0.111
degenerate	1	0.391	0.460	0.214	0.093
confer	1	0.437	0.485	0.165	0.085
descend	2	0.612	0.644	0.182	0.063
go	2	0.714	0.714	0.066	0.058

4 Historical frequency changes with a different threshold

In this paper, we set the frequency threshold for verb selection at 200 appearances. Figure 3 is the result when setting the frequency threshold less strictly to 30 appearances. Again, we can confirm that the *have*+PP format became dominant in most verbs.

References

[1] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).

[2] Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Taylor & Francis, 2013).

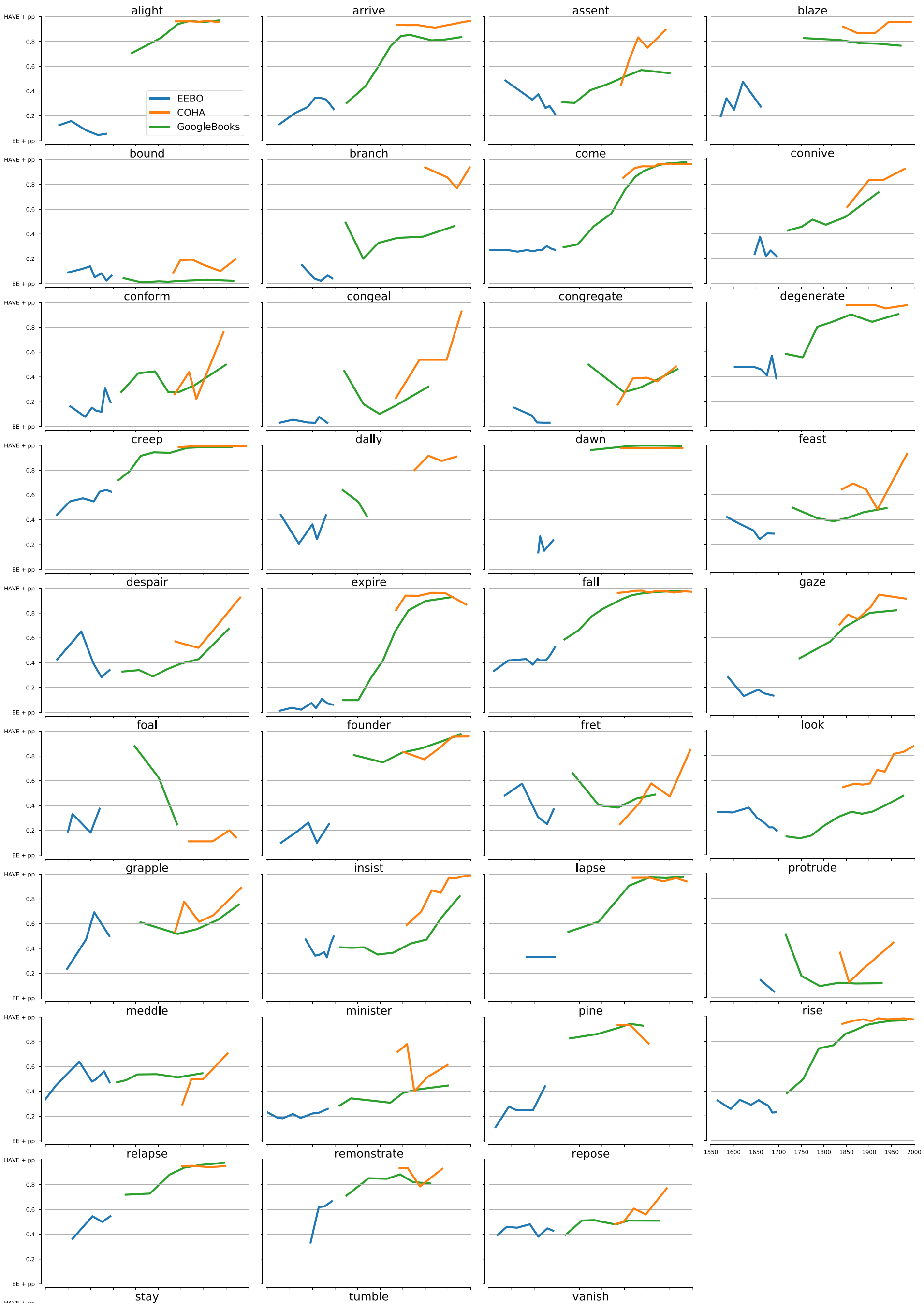

Figure S 3: Frequency changes of *be*/*have*+PP by dataset in Group A

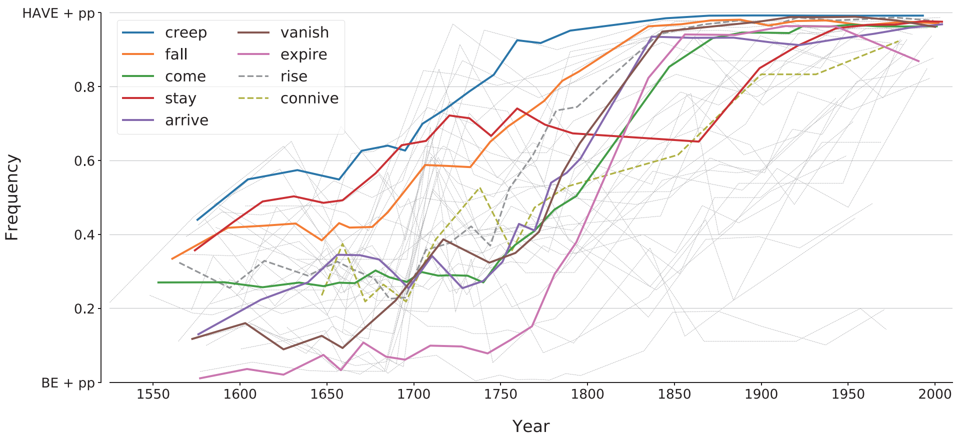

Figure S 4: Frequency changes of *be*/*have*+PP (all aggregated) in Group A (frequency threshold = 30).

Table S 5: Neural TSC result (frequency threshold = 30)

Verb	Probability
alight	1.00
arrive	1.00
blaze	1.00
branch	1.00
come	1.00
congeal	1.00
connive	1.00
creep	1.00
dally	1.00
dawn	1.00
degenerate	1.00
expire	1.00
fall	1.00
foal	1.00
founder	1.00
gaze	1.00
grapple	1.00
insist	1.00
lapse	1.00
pine	1.00
relapse	1.00
remonstrate	1.00
rise	1.00
stay	1.00
vanish	1.00
look	1.00
conform	1.00
tumble	1.00
minister	1.00
congregate	1.00
assent	0.98
despair	0.98
feast	0.94
meddle	0.25
bound	0.07
go	0.01