Frequency Accelerates Semantic Change: Evidence from 500 Years of Korean

Cheonkam Jeong[*]
University of California, Irvine
Irvine, CA, United States
cheonkaj@hs.uci.edu

Yeeun Choi
Korea International School
Republic of Korea
lychoi0809@gmail.com

Abstract

The "law of conformity," the finding that frequent words are semantically stable, has been treated as a broad regularity of language change. We show it does not hold for Korean. Using diachronic word embeddings trained on historical corpora spanning 500 years (15th–20th centuries), we find a robust *positive* correlation between frequency and semantic shift: high-frequency Korean words change *more*, not less. The pattern survives six robustness controls and is validated against an English replication. Partial correlation analysis reveals that the role of polysemy in mediating the frequency–change relationship is not fixed but depends on time resolution and corpus homogeneity. We connect the reversal to frequency-driven reductive processes, including grammaticalization, semantic bleaching, and domain shift, that are especially productive in Korean. The frequency–change relationship is not a fixed regularity but varies with language typology and analytical conditions.

1 Introduction

A central question in historical linguistics is what drives words to change meaning over time. Quantitative approaches to this question have accelerated in recent years, enabled by diachronic word embedding methods that measure semantic shift across time periods (Kutuzov et al., 2018; Tahmasebi et al., 2021). Among the most influential findings is the "law of conformity" proposed by Hamilton et al. (2016), which states that more frequent words are more semantically stable: frequency and semantic change are negatively correlated ($\beta_f \in [-1.26, -0.27]$ across languages and corpora, 1800–2000). This has been interpreted as a statistical regularity of language change, alongside a parallel "law of innovation" linking polysemy to accelerated change.

However, this finding rests primarily on evidence from English and a small number of other Indo-European languages observed over roughly two centuries. Whether the law of conformity holds cross-linguistically, particularly for typologically distant languages and over longer time scales, remains an open question. Xu and Kemp (2015) provided early evidence that not all proposed regularities of semantic change generalize uniformly, while Dubossarsky et al. (2017) demonstrated that some observed correlations may reflect artifacts of the embedding process. Even within Hamilton et al. (2016)'s own English data, the relationship between frequency and polysemy is non-trivial: the two are highly correlated, and partial correlation analysis reverses the apparent direction of polysemy's effect, raising the question of whether the internal relationship among frequency, polysemy, and change is itself stable across languages.

Korean provides an ideal test case for these questions. It is agglutinative, head-final, and belongs to a distinct language family, with a well-documented written history spanning over 500 years from the promulgation of the Korean script (*Hunminjeongeum*) in 1446. Crucially, Korean has undergone extensive grammaticalization over this period (Rhee, 2019), a process that preferentially targets high-frequency items (Bybee, 2003) and may interact with the frequency–change relationship in ways not visible in shorter observation windows.

In this paper, we train diachronic Word2Vec models on Korean historical corpora from the 15th to 20th centuries, align them using Orthogonal Procrustes, and measure the correlation between word frequency and semantic shift. We find that the law of conformity does not hold in Korean: high-frequency words undergo significantly *more* semantic change ($r = +0.494$, $\rho = +0.527$). This pattern holds across controls for morphological artifacts, corpus size imbalance, polysemy, light verb

[*]Corresponding author.

Proceedings of the 6th International Conference on Natural Language Processing for the Digital Humanities, pages 13–23

compounds, genre, and period-specific effects, and is further validated by a direct comparison with English using Hamilton et al.'s own pre-trained embeddings (Appendix A). At century resolution, frequency is the independent driver and polysemy plays no role once frequency is controlled, but a controlled comparison across time resolutions and corpus types reveals that this null mediation is partly an artifact of coarse temporal grain and genre heterogeneity. We connect this reversal to Bybee (2003)'s observation that frequency exerts both a conserving force (entrenchment of stored representations) and a reducing force (phonological reduction, semantic bleaching, grammaticalization): in Korean, the reducing force dominates, yielding the positive correlation.

Our contributions are as follows.

1. We provide the first large-scale test of the law of conformity on a non-Indo-European, agglutinative language over a 500-year span.
2. We document a robust positive frequency–change correlation in Korean, reversing the predicted pattern.
3. We show that the role of polysemy in mediating frequency–change depends on time resolution and corpus homogeneity, reconciling apparently contradictory findings across analytical conditions.
4. We identify the dominance of frequency-driven reducing processes (semantic bleaching, grammaticalization, domain shift) over conserving processes as the mechanism underlying the Korean pattern.
5. We provide a direct English–Korean comparison using the same analytical pipeline, confirming that the reversal reflects a genuine cross-linguistic difference.

2 Related Work

2.1 Diachronic Word Embeddings

The computational study of semantic change has grown rapidly since the introduction of neural word embedding methods. Kim et al. (2014) trained neural language models on yearly snapshots of text to track word meaning trajectories, and Kulkarni et al. (2015) proposed statistical tests for detecting significant change using Word2Vec embeddings aligned across time. Building on this line of work, Hamilton et al. (2016) introduced Orthogonal Procrustes alignment (Schönemann, 1966) and proposed two statistical laws: the *law of conformity* (frequent words change less) and the *law of innovation* (polysemous words change more).

These findings have been both refined and challenged. Dubossarsky et al. (2017) showed that some observed correlations may be artifacts of the embedding training process, particularly for smaller corpora. More recent studies have employed contextualized embeddings (Hu et al., 2019; Giulianelli et al., 2020) and expanded analysis to languages beyond English, including German, Latin, and Swedish (Schlechtweg et al., 2020). Chen et al. (2023) introduced ChiWUG, the first graph-based evaluation dataset for Chinese semantic change detection, extending computational analysis to a non-Indo-European language. Despite this progress, large-scale tests of Hamilton et al.'s proposed laws on typologically distant languages remain scarce.

2.2 Frequency, Polysemy, and Semantic Change

The relationship between word frequency and meaning change has a long history in linguistics. Bybee (2003) argued that frequency exerts two opposing forces on linguistic items: a *conserving* force, whereby high token frequency entrenches stored representations and resists change, and a *reducing* force, whereby repeated use drives phonological reduction, semantic bleaching, and grammaticalization. Which force dominates may depend on the type of change, the language, and the time scale of observation. The law of conformity captures only the conserving side of this picture; the reducing side predicts the opposite (i.e., high-frequency words should change *more*), precisely the pattern we observe in Korean.

A subtle pattern within Hamilton et al. (2016)'s own analysis is that frequency and polysemy are highly positively correlated, yet after controlling for frequency, polysemy retains a significant positive effect on semantic change. This raises the question of whether the joint structure of frequency, polysemy, and change is itself stable across languages.

2.3 Grammaticalization and Language Change

Grammaticalization, the process by which lexical items develop grammatical functions over time (Kouteva et al., 2019), is one prominent instantiation of Bybee's reducing force: high-frequency items undergo semantic bleaching, category shift,

and phonological reduction as they grammaticalize (Bybee, 2003; Traugott and Dasher, 2001). However, grammaticalization is not the only frequency-driven reductive process; domain shift, meaning extension, and subjectification (Traugott, 1995) also target high-frequency items without necessarily producing grammatical elements.

Korean is known to have undergone extensive grammaticalization over its recorded history (Rhee, 2019): the verb *ha.ta* ('to do') developed from a full lexical verb into a productive light verb, and numerous spatial nouns grammaticalized into postpositions (Lee and Ramsey, 2011). Whether such productive reducing processes can override the conserving force behind the law of conformity has not been tested computationally; our study addresses this question.

2.4 Korean Historical Corpora

Korean has a continuous written tradition dating from the invention of the Hangul script in 1443, but historical texts require extensive orthographic normalization due to archaic jamo and variant spellings (Lee and Ramsey, 2011). To our best knowledge, computational work on Korean diachronic semantics remains limited to qualitative case studies; ours is the first large-scale quantitative analysis using diachronic embeddings.

3 Data

3.1 Corpus

We use the *eon* corpus, a diachronic corpus of Korean texts spanning six centuries (15th–20th), compiled from digitized historical Korean sources originally encoded in UTF-16LE with XML/SGML markup.[1] Texts were grouped by century to form six sub-corpora. Table 1 summarizes the corpus statistics and Table 2 provides the genre breakdown.

The corpus composition shifts from Buddhist and Confucian texts (15–17C) through royal and missionary materials (18–19C) to newspapers and novels (20C); details are in Table 2. The 20th-century sub-corpus is notably dominated by *sinsosel*[2] (60% of texts); we address genre and corpus size imbalances in our robustness analyses.

[1]The corpus was provided by a scholar of classical Korean literature (*hanmunhak*). We gratefully acknowledge their contribution.

[2]*Sinsosel* (신소설, lit. "new novel") is a transitional genre of early modern Korean fiction (ca. 1900–1920) bridging classical and modern Korean prose.

Table 1: Corpus statistics by century (after preprocessing and normalization, as used for model training).

Century	Files	Sentences	Tokens	Vocab
15th	61	29,858	487,967	9,864
16th	61	15,342	245,466	6,536
17th	58	20,384	458,690	12,028
18th	95	22,800	504,160	12,282
19th	65	76,677	1,450,944	31,781
20th	87	157,900	2,322,462	49,885
Total	427	322,961	5,469,689	—

Table 2: Genre distribution by century. Values indicate number of source texts per genre.

Genre	15C	16C	17C	18C	19C	20C
Buddhist	18	8	2	2	–	–
Confucian	3	23	12	15	4	–
Medical	1	7	6	–	2	–
Royal/court	15	2	3	22	6	–
Dict./reference	–	3	4	8	4	16
Translation	–	3	1	4	2	–
Christian	–	–	–	–	8	13
Diary/letter	1	3	9	1	1	1
Literary/poetry	9	–	9	8	–	–
Newspaper	–	–	–	–	–	1
New novel	–	–	–	–	–	52
Other	14	12	12	35	38	4
Total	61	61	58	95	65	87

3.2 Preprocessing

Historical Korean texts require extensive normalization due to orthographic variation across centuries. Our preprocessing pipeline includes Unicode normalization of PUA characters, archaic jamo[3] mapping to modern equivalents, markup removal, whitespace tokenization at the *eojeol*[4] level, and filtering of short tokens (<2 characters) and short lines (<3 tokens).

4 Methods

4.1 Word2Vec Training

We trained separate Skip-gram Word2Vec models (Mikolov et al., 2013) for each century using the Gensim library (Řehůřek and Sojka, 2010), with the following hyperparameters: embedding dimension = 100, context window = 5, minimum word count = 5, training epochs = 10. We used 100 dimensions given the smaller corpus size (cf. 300 in

[3]*Jamo* (자모) are the consonant and vowel letters that compose Hangul syllable blocks. Pre-modern texts contain archaic jamo (e.g., ㆍ *arae-a*) no longer used in modern Korean.

[4]An *eojeol* (어절) is the orthographic word unit delimited by whitespace in Korean; it typically contains a content stem followed by one or more bound morphemes (particles, suffixes, endings).

Hamilton et al. (2016)). All models were trained with a fixed random seed for reproducibility.

4.2 Alignment via Orthogonal Procrustes

We aligned all century models to a reference century (17th, chosen as the chronological midpoint with sufficient data) using Orthogonal Procrustes alignment (Schönemann, 1966; Hamilton et al., 2016), computing the optimal rotation $R^* = \arg\min_R \|X_c R - X_{\text{ref}}\|_F$ s.t. $R^T R = I$ over shared vocabulary words.

4.3 Semantic Shift Measurement

For each word present in at least 4 centuries and with an average frequency ≥ 5, we measured semantic shift as:

$$\text{shift}(w) = 1 - \cos(\mathbf{v}_w^{\text{first}}, \mathbf{v}_w^{\text{last}}) \quad (1)$$

where $\mathbf{v}_w^{\text{first}}$ and $\mathbf{v}_w^{\text{last}}$ are the aligned vectors of word w in its earliest and latest attested century, respectively.[5]

4.4 Frequency and Polysemy Measurement

Word frequency was computed as the average raw token count across all centuries in which the word appeared. Following Hamilton et al. (2016), we used $\log_{10}(\text{frequency})$ in all correlation analyses.

We estimated word polysemy using neighbor *dispersion* (the average pairwise cosine distance among the top 20 nearest neighbors), averaged across all attested centuries. This captures contextual diversity similarly to, though distinct from, Hamilton et al. (2016)'s local clustering coefficient: both measures quantify how diverse a word's usage contexts are, but neighbor dispersion operates directly in embedding space rather than on a co-occurrence network.

5 Results

5.1 Reversal of the Law of Conformity

Figure 1 shows the relationship between log frequency and semantic shift for 2,378 words present in at least 4 centuries. We observe a strong *positive* correlation: Pearson $r = +0.494$ ($p < 10^{-146}$), Spearman $\rho = +0.527$ ($p < 10^{-170}$).

This is the *opposite* direction from the law of conformity, which predicts a negative correlation. Our own English replication confirms the expected negative pattern ($r = -0.263$ for the full 1800–1990 span; Appendix A). In Korean, high-frequency words undergo significantly *more* semantic change over 500 years.

[5]We also computed Hamilton et al.'s displacement rate measure, $\Delta^{(t)}(w) = 1 - \cos(\mathbf{v}_w^{(t)}, \mathbf{v}_w^{(t+1)})$, with pooled regression across all word–period observations; results are reported in Appendix A.

Table 3: Pearson r between log frequency and semantic shift for adjacent century pairs.

Pair	r	p	n
15C–16C	$+0.537$	4.4×10^{-148}	1,974
16C–17C	$+0.228$	3.0×10^{-23}	1,850
17C–18C	$+0.330$	1.5×10^{-52}	2,024
18C–19C	$+0.403$	1.0×10^{-65}	1,658
19C–20C	$+0.445$	9.2×10^{-76}	1,537

5.2 Robustness Check 1: Period-Specific Consistency

To verify that the pattern is not driven by the long time span, we computed the frequency–shift correlation for each pair of adjacent centuries. As shown in Table 3, the correlation is positive and significant for all five pairs, ranging from $r = +0.228$ (16C–17C) to $r = +0.537$ (15C–16C).

5.3 Robustness Check 2: Morphological Control

Korean is agglutinative, so each eojeol may contain a stem plus multiple suffixes. We tested three frequency counting methods to check for tokenization artifacts: original eojeol frequency ($r = +0.494$), approximate stems ($r = +0.441$), and unique stem types ($r = +0.485$). All yield positive correlations ($p < 10^{-100}$).

5.4 Robustness Check 3: Corpus Size Equalization

The 20th-century sub-corpus (2.3M tokens) is nearly 10 times larger than the 16th-century one (245K tokens). We downsampled all centuries to 245,466 tokens, retrained and realigned all models: $r = +0.491$ ($p < 10^{-78}$, $n = 1{,}276$), a change of only -0.003, ruling out the corpus-size artifacts noted by Dubossarsky et al. (2017).

5.5 Robustness Check 4: Disentangling Frequency and Polysemy

Hamilton et al. (2016) proposed that polysemous words change faster (their law of innovation), and noted that frequency and polysemy are highly correlated in their English data; yet after mutual control, both variables retain independent effects

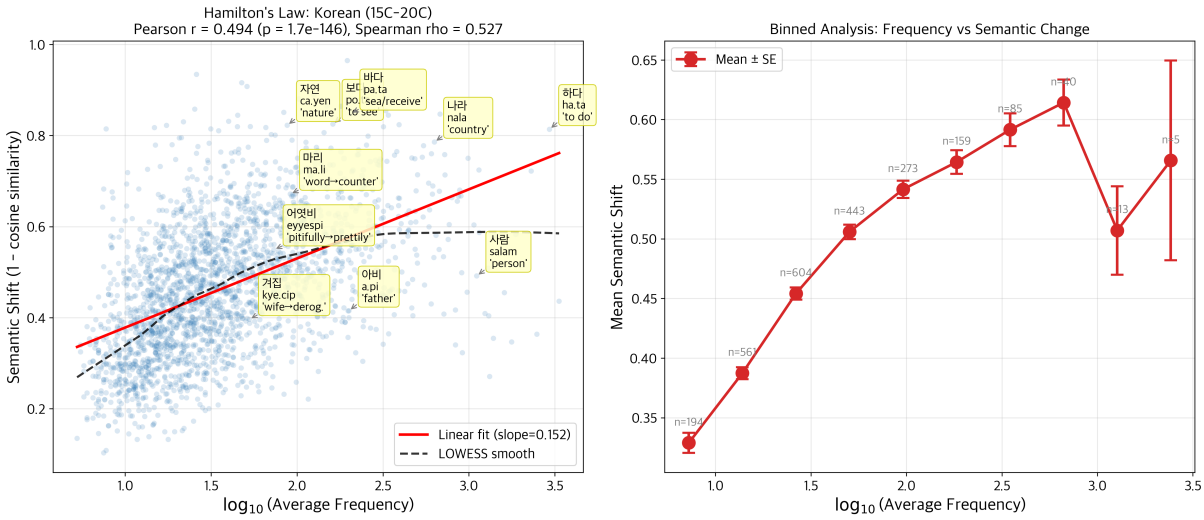

Figure 1: Log frequency vs. semantic shift for 2,378 Korean words (15C–20C). The positive correlation ($r = +0.494$) is the opposite direction from English ($r = -0.263$ in our replication; Appendix A).

Table 4: Pairwise and partial correlations among frequency, shift, and polysemy ($n = 2{,}378$).

Pairwise	r	p
Freq ↔ Shift	$+0.494$	$< 10^{-146}$
Freq ↔ Polysemy	$+0.788$	$< 10^{-300}$
Polysemy ↔ Shift	$+0.394$	$< 10^{-89}$
Partial	r	p
Freq → Shift | Poly	$+0.325$	$< 10^{-59}$
Poly → Shift | Freq	$+0.008$	0.708

(frequency negative, polysemy positive). If high-frequency Korean words are simply more polysemous, the observed pattern could likewise be a polysemy artifact.

We computed pairwise and partial correlations (Table 4). First, controlling for polysemy *reduces* but does not eliminate the frequency–shift correlation (r: $+0.494 \rightarrow +0.325$). Second, controlling for frequency *eliminates* the polysemy–shift correlation entirely ($r = +0.008$, n.s.). This pattern contrasts with Hamilton et al. (2016)'s English data, where both frequency (negative) and polysemy (positive) retain independent effects after mutual control. In Korean, frequency is the only variable that survives mutual control, and its effect is positive.

The pattern persists across polysemy tertiles: low ($r = +0.386$), mid ($r = +0.381$), and high ($r = +0.268$), all significant at $p < 10^{-15}$. Multicollinearity is not a concern (VIF = 2.29), and a permutation test (10,000 iterations) places the partial r 25 SDs above the null ($p < 10^{-4}$).

5.6 Robustness Check 5: Light Verb Exclusion

Korean has a highly productive light verb construction "X-*hata*" (하다 *ha.ta*, 'to do'), which combines with Sino-Korean nouns to form verbs (e.g., 사랑하다 *salang-ha.ta* 'to love', 공부하다 *kongpwu-ha.ta* 'to study'). The number of unique X-*hata* types expanded 22-fold from the 15th to the 20th century.

Excluding all words containing *hata*-related morphology *increases* the correlation slightly ($r = +0.504$, $n = 2{,}295$), confirming that the pattern is not driven by light verb constructions.

5.7 Robustness Check 6: Genre Control

Our corpus spans multiple genres whose distribution shifts across centuries (Table 2).

First, we computed per-genre correlations by restricting the analysis to words attested primarily within each genre. The correlation is positive and significant in all 11 genres, ranging from $r = +0.302$ (Newspaper) to $r = +0.605$ (Confucian) (Table 5).

Second, a linear mixed-effects model with log frequency as a fixed effect and genre as a random intercept confirms that frequency independently predicts semantic shift ($\beta = +0.109$, $z = 16.15$, $p = 1.09 \times 10^{-58}$). The genre random intercept variance is 0.002, indicating that genre contributes negligibly to baseline shift levels.

Third, we trained a separate Word2Vec pipeline exclusively on Confucian texts (15C–18C), which span four centuries with consistent representation.

Table 5: Per-genre frequency–shift correlations. All correlations are positive and significant.

Genre	r	p	n
Confucian	$+0.605$	2.0×10^{-11}	101
Translation	$+0.541$	7.9×10^{-4}	35
Diary/letter	$+0.495$	1.0×10^{-8}	119
Literary/poetry	$+0.386$	6.9×10^{-6}	128
Other	$+0.383$	1.1×10^{-5}	125
Dict./reference	$+0.377$	6.1×10^{-6}	136
Buddhist	$+0.351$	2.0×10^{-12}	378
New novel	$+0.344$	1.8×10^{-4}	114
Christian	$+0.337$	5.4×10^{-10}	323
Medical	$+0.309$	1.9×10^{-2}	57
Newspaper	$+0.302$	1.3×10^{-4}	155

Table 6: Mediation by time resolution and corpus type. Med. = proportion of total effect mediated by polysemy (Baron–Kenny).

Corpus	Resol.	Span	n	r	Med.
Literary	century	500 yr	2,378	$+0.494$	1.7%
Literary	decade	30 yr	1,448	$+0.323$	19.5%
News	decade	40 yr	40,785	$+0.547$	90.4%

This single-genre analysis yields $r = +0.733$ ($n = 388$), substantially stronger than the full-corpus result over the same time window ($r = +0.495$, $n = 1{,}749$). The mixing of genres in the full corpus thus dilutes rather than inflates the correlation.

5.8 Time Resolution and Mediation

The near-zero polysemy mediation at century resolution (§5.5) could reflect coarse temporal grain rather than a genuine absence of the polysemy pathway. To test this, we conducted a controlled comparison across three analytical configurations: (i) the literary corpus at century resolution (500 years, 15C–20C), (ii) the same literary corpus at decade resolution (30 years, 1880–1910), and (iii) a homogeneous newspaper corpus at decade resolution (40 years, 1920–1960).

Two factors jointly determine mediation strength (Table 6). First, *time resolution*: moving from century to decade resolution on the same literary corpus increases mediation from 1.7% to 19.5%. Second, *corpus homogeneity*: at the same decade resolution, the single-genre newspaper corpus yields 90.4% mediation versus 19.5% for the multi-genre literary corpus (which mixes religious texts, dictionaries, and fiction from 1880–1910). Genre heterogeneity injects noise into the polysemy proxy (neighbor dispersion), attenuating the mediated pathway.

The positive sign is stable across all three configurations ($r = +0.323$ to $+0.547$); what varies is how much of this effect routes through polysemy, a function of analytical resolution rather than the underlying linguistic phenomenon.

5.9 Contextualized Embedding Validation

To test whether the positive correlation is visible to contextualized models, we applied KLUE-BERT (Park et al., 2021) under two configurations. **Frozen BERT**: We extracted contextual embeddings for 1,493 target words (of the original 2,378; the remainder lacked sufficient attestation in BERT-tokenizable contexts) using up to 50 contexts per word per century and measured semantic shift via prototype distance (PRT) and average pairwise distance change (APD). Neither metric shows a significant correlation: PRT $r = -0.010$ ($p = 0.703$), APD $r = +0.032$ ($p = 0.222$). **Continual pretraining**: We further pretrained KLUE-BERT on each century's subcorpus via masked language modeling (3 epochs for 15–18C, 2 epochs for 19–20C). Results remain null: PRT $r = +0.011$ ($p = 0.669$), APD $r = +0.038$ ($p = 0.141$).

The null results are expected: the changes we detect (bleaching, domain shift, light verb development) operate at the *type level*, which static embeddings capture by design but BERT's token-level representations do not. Domain mismatch with pre-modern orthography further limits BERT's sensitivity. Two readings remain possible: (i) the reversal is genuinely type-level—a shift in distributional signature that static embeddings capture but BERT's token-level representations do not—or (ii) the Word2Vec signal partly reflects aligned-embedding artifacts. We favor (i) since corpus-size equalization (§5.4) and Hamilton-style replication (Appendix A) both reproduce the positive sign; the scope of our claim is therefore type-level lexical change.

5.10 Qualitative Case Studies

To illustrate the nature of the detected semantic changes, we examine the diachronic nearest neighbors of selected words (Table 7). The five words below are illustrative rather than exhaustive; the systematic, large-scale evidence for the reversal comes from the 2,378-word correlation analysis (§5.1), the 22-fold expansion of X-*hata* light verb constructions (§5.6), and the per-century shift heatmap (Figure 2).

Table 7: Diachronic nearest neighbors for selected words. Shift = $1 - \cos(\mathbf{v}^{\text{first}}, \mathbf{v}^{\text{last}})$. Yale romanization is provided for all Korean forms.

Word	Gloss	Shift	Span	Earliest neighbors	Latest neighbors
하다 *ha.ta*	'to do'	0.814	15–20C	일 도 *il.to* 'work-also', 그 러 타 *kule.tha* 'be so' (content verb)	것 *kes* 'thing', 엄중하게 *emcwung-ha.key* 'strictly' (light verb)
보다 *po.ta*	'to see'	0.825	15–20C	비취샨 *pichhwi.syan* 'shone upon', 상각하야 *sangkak.ha.ya* 'thinking' (visual)	깔보다 *kkal.po.ta* 'look down on', 살피다 *sal.phi.ta* 'examine' (evaluative)
바로 *pa.lo*	'straight'	0.965	15–20C	씩씩하샤미 *siksik.ha.sya.mi* 'being vigorous,' 너브시고 *nepu.si.ko* 'being broad' (physical)	지 체 *ci.chey* 'delay', 숨 기 지 *swumki.ci* 'not hiding' (abstract)
마리 *ma.li*	'word'	0.671	15–20C	달오다 *tal.wo.ta* 'to say', 듣자오미 *tut.ca.wo.mi* 'hearing' (speech)	송 아 지 *song.a.ci* 'calf', 곰 *kom* 'bear' (animal counter)
비록 *pi.lok*	'although'	0.323	15–20C	이시나 *i.si.na* 'but', 그러나 *ku.le.na* 'however' (concessive)	그러나 *ku.le.na* 'however', 할지언 뎡 *hal.ci.en.tyeng* 'even if' (concessive)

Grammaticalization The verb 하다 *ha.ta* ('to do') shows a clear trajectory from content verb to light verb (Rhee, 2019): from 일 *il* ('work') in the 15th century to the nominalizer 것 *kes* ('thing') in the 20th, reflecting its modern role in Sino-Korean constructions (e.g., 공부하다 *kongpwu-ha.ta* 'to study'). Similarly, 바로 *pa.lo* shifts from physical attributes to abstract/discourse functions, a textbook case of semantic bleaching (Bybee, 2003).

Semantic domain shift The noun 마 리 *ma.li* undergoes a complete domain shift: from speech-related neighbors in the 15th century (consistent with its original meaning of 'word/speech') to animal-related neighbors in the 20th, reflecting its modern use as a counting unit for animals.

Stability of function words In contrast, 비록 *pi.lok* ('although') maintains consistent concessive neighbors across centuries (그러나 *ku.le.na* 'however'), yielding a low shift score (0.323).

The qualitative picture matches the quantitative one: high-frequency content words undergo substantial change through diverse reductive processes, while already-grammaticalized function words remain stable. Figure 2 provides a fine-grained view of per-century-pair shift magnitudes.

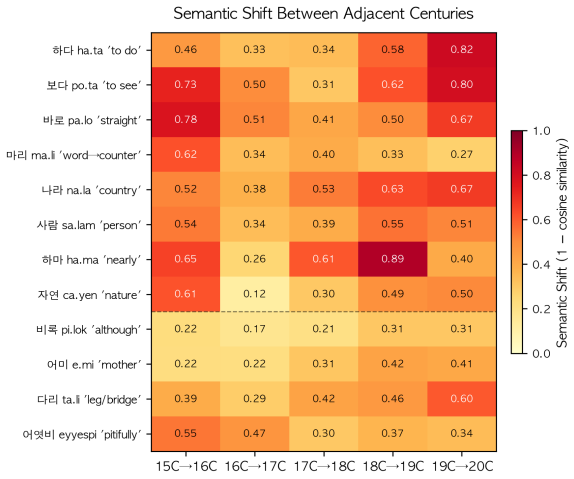

Figure 2: Pairwise semantic shift ($1 - \cos$) between adjacent centuries for selected words. Darker cells indicate greater shift. The dashed line separates high-shift content words (above) from more stable items (below).

6 Discussion

Our central finding is that the law of conformity appears not to hold for Korean: the positive correlation ($r = +0.494$) is robust across all six controls (§5.2–5.7), including corpus size equalization that partly addresses the embedding artifact concerns of Dubossarsky et al. (2017) ($r = +0.491$ vs. $+0.494$). We frame the linguistic interpretations below as compatible with our results rather than as causally established by them.

6.1 The Frequency–Polysemy–Change Triad

Our partial correlation results speak directly to the joint structure of frequency, polysemy, and semantic change identified by Hamilton et al. (2016). In their English data, frequency and polysemy are highly correlated, yet both retain independent effects after mutual control: frequency negative, polysemy positive—the latter opposite to what the positive frequency–polysemy correlation would predict. At century resolution, Korean shows a different structure: frequency retains an independent positive effect ($r = +0.325$ after controlling for polysemy), while polysemy's effect vanishes ($r = +0.008$, n.s.; VIF = 2.29, permutation test 25

SDs above null).

However, the time resolution experiment (§5.8) reveals that this near-zero mediation is partly an artifact of analytical conditions: at decade resolution mediation rises to 19.5%, and in a homogeneous newspaper corpus to 90.4%. Two implications follow. (i) Claims about *whether* polysemy mediates frequency–change are not separable from the conditions under which they are tested; century-level bins and genre mixing both attenuate the polysemy proxy, which may also explain divergent conclusions across English studies. (ii) The *positive* sign of the correlation is stable across all configurations ($r = +0.323$ to $+0.547$); what varies is the causal pathway, not the direction. Consistent with Bybee's dual-force account (§2), the reducing force appears to dominate in Korean.

6.2 Frequency-Driven Reductive Processes

Our case studies (§5.10) instantiate this pattern through grammaticalization, semantic bleaching, and domain shift, while already-grammaticalized items remain stable. These are frequency-driven reductive processes (Bybee, 2003; Rhee, 2019); the 22-fold expansion of X-*hata* (§5.6) provides population-level evidence that this pathway is productive across the corpus, not just in selected examples.

The 19C–20C pair ($r = +0.445$) spans roughly the same time depth as our English replication ($r = -0.668$ for 1850–1990; Appendix A), yet shows the opposite sign—ruling out observation window as the cause and pointing to typological differences (§6.3). The heatmap also shows shifts intensifying in the 18C–19C and 19C–20C transitions, coinciding with Korean modernization and Western-language contact, suggesting interaction with external sociolinguistic pressures.

6.3 Cross-linguistic Implications

The law of conformity was derived from English (1800–2000) and later replicated on Indo-European languages. Our English replication using HistWords (Appendix A; $r = -0.263$ to -0.776) confirms the negative pattern, while the Korean data yield a positive correlation of comparable magnitude. The frequency–change relationship is thus conditioned by both *language typology* and *time scale*.

On Bybee's dual-force account, every language exhibits both a conserving and a reducing force; the observed sign reflects which dominates. Agglutinative morphology may amplify the reducing force in two ways: (i) the productive X-*hata* construction and the grammaticalization of high-frequency verbs and spatial nouns into bound morphemes (§5.6; Rhee, 2019) create recurring constructional slots into which high-frequency items are repeatedly recruited, shifting their distributional neighborhood; (ii) eojeol-level contexts are dominated by bound-morpheme neighbors, so even modest functional re-analysis substantially changes a word's top neighbors. The reversal is thus a prediction of which force dominates given a language's morphosyntactic resources, not a denial of the conserving force.

Together with Xu and Kemp (2015)'s finding that regularities vary across word classes and Chen et al. (2023)'s work on Chinese, our results suggest that cross-linguistic variation in frequency–change dynamics may be more common than previously assumed. Whether other agglutinative languages (e.g., Turkish, Japanese, Finnish) show similar reversals remains an open question.

7 Conclusion

We have presented the first large-scale computational study of the frequency–semantic change relationship in Korean, spanning 500 years of recorded history (15th–20th centuries). Our central finding is that high-frequency Korean words appear to change *more*, not less, a pattern that reverses the law of conformity and is robust across six controls and validated against an English replication. The role of polysemy in mediating the frequency–change relationship is not absent but depends on analytical conditions: time resolution and corpus homogeneity jointly shape how much of the effect routes through polysemy, while the positive sign of the correlation remains stable. We interpret the reversal as consistent with frequency-driven reductive processes, including grammaticalization, semantic bleaching, and domain shift, that are especially productive in agglutinative Korean (Bybee, 2003; Rhee, 2019). The frequency–change relationship thus appears not to be a fixed regularity but to vary with language typology and analytical conditions.[6]

Limitations

Our study has several limitations. First, without a reliable morphological analyzer for historical Ko-

[6]Code and trained models will be made available upon publication.

rean, we tokenize at the eojeol (spacing unit) level, though our lemmatization check suggests this does not drive the main finding. Second, our BERT validation (§5.9) yielded null results, which we attribute to type-level vs. token-level sensitivity and domain mismatch; future work with historically pretrained language models may reveal more nuanced patterns. Embedding quality may also be lower for the smaller sub-corpora (15C: 488K tokens, 16C: 245K tokens), though our corpus-size equalization directly addresses this concern ($r = +0.491$ after downsampling all centuries to 245K tokens); an independent intrinsic evaluation of embedding quality on historical Korean is left to future work. Third, replication with independent historical Korean corpora would strengthen the findings. Fourth, we used average frequency across all attested centuries rather than lagged frequency at timepoint $t-1$; while our adjacent-century pair analysis (Table 3) partially addresses this by showing the positive correlation holds across all five pairs, a fully lagged design would be more appropriate for causal interpretation. Fifth, our neighbor-dispersion approach to polysemy provides only an approximate proxy for true lexicographic polysemy, reported here as a centuries-averaged score, and our time resolution experiment shows that mediation estimates are sensitive to temporal grain and corpus homogeneity. Sixth, the newspaper corpus used in the decade-resolution comparison (1920–1960) is limited in temporal range and may not be representative of all registers of modern Korean.

Acknowledgments

We thank the two anonymous reviewers for their thoughtful and constructive feedback, which substantially improved this paper. We thank the anonymous scholar of classical Korean literature (*hanmunhak*) for generously providing the historical Korean corpus used in this study. We also acknowledge the use of Claude Opus 4.6 (Anthropic) for assistance with manuscript revision and data visualization.

References

Joan Bybee. 2003. Mechanisms of change in grammaticization: The role of frequency. In Brian D. Joseph and Richard D. Janda, editors, *The Handbook of Historical Linguistics*, pages 602–623. Blackwell Publishing, Oxford.

Jing Chen, Emmanuele Chersoni, Dominik Schlechtweg, Jelena Prokic, and Chu-Ren Huang. 2023. ChiWUG: A graph-based evaluation dataset for Chinese lexical semantic change detection. In *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change*, pages 130–138, Singapore. Association for Computational Linguistics.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145, Copenhagen, Denmark. Association for Computational Linguistics.

Mario Giulianelli, Marco Del Tredici, and Raquel Fernández. 2020. Analysing lexical semantic change with contextualised word representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3960–3973, Online. Association for Computational Linguistics.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Renfen Hu, Shen Li, and Shichen Liang. 2019. Diachronic sense modeling with deep contextualized word embeddings: An ecological view. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3899–3908, Florence, Italy. Association for Computational Linguistics.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal analysis of language through neural language models. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 61–65, Baltimore, MD. Association for Computational Linguistics.

Tania Kouteva, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog, and Seongha Rhee. 2019. *World Lexicon of Grammaticalization*, 2nd edition. Cambridge University Press, Cambridge.

Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*, pages 625–635, Florence, Italy. ACM.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: A survey. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, pages 1384–1397, Santa Fe, NM. Association for Computational Linguistics.

Ki-Moon Lee and S. Robert Ramsey. 2011. *A History of the Korean Language*. Cambridge University Press, Cambridge. Originally published in Korean as *Kugosa Kaeseol* (Thaehaksa, 1972; revised editions through 2015).

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26 (NeurIPS 2013)*, pages 3111–3119. Curran Associates, Inc.

Sungjoon Park, Jihyung Moon, Sungdong Kim, Won Ik Cho, Jiyoon Han, Jangwon Park, Chisung Song, Junseong Kim, Yongsook Song, Taehwan Oh, Joohong Lee, Juhyun Oh, Sungwon Lyu, Younghoon Jeong, Inkwon Lee, Sangwoo Seo, Dongjun Lee, Hyunwoo Kim, Myeonghwa Lee, and 18 others. 2021. KLUE: Korean language understanding evaluation. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

Radim Řehůřek and Petr Sojka. 2010. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA.

Seongha Rhee. 2019. Grammaticalization in Korean. In Jaehoon Yeon and Lucien Brown, editors, *The Cambridge Handbook of Korean Linguistics*, pages 507–530. Cambridge University Press.

Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. SemEval-2020 Task 1: Unsupervised lexical semantic change detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation (SemEval-2020)*, pages 1–23, Barcelona, Spain (Online). Association for Computational Linguistics.

Peter H. Schönemann. 1966. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10.

Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2021. Survey of computational approaches to lexical semantic change detection. In Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen, editors, *Computational Approaches to Semantic Change*, pages 1–91. Language Science Press, Berlin.

Elizabeth Closs Traugott. 1995. Subjectification in grammaticalization. *Subjectivity and subjectivisation: Linguistic perspectives*, 1:31–54.

Elizabeth Closs Traugott and Richard B. Dasher. 2001. *Regularity in Semantic Change*. Cambridge University Press, Cambridge.

Yang Xu and Charles Kemp. 2015. A computational evaluation of two laws of semantic change. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, pages 2703–2708.

Table 8: Frequency–shift correlations: Korean vs. English.

Window	**Language**	r	n
Full span	Korean (15C–20C)	$+0.494$	2,378
Full span	English (1800–1990)	-0.263	99,963
$\sim$150 yr	Korean (19C–20C)	$+0.445$	1,537
$\sim$150 yr	English (1850–1990)	-0.668	99,963
$\sim$100 yr	Korean (19C–20C)	$+0.445$	1,537
$\sim$100 yr	English (1900–1990)	-0.776	99,963

A English Replication with HistWords

To enable direct comparison, we replicated the frequency–shift correlation analysis on English using the pre-trained HistWords embeddings released by Hamilton et al. (2016).[7] These are decade-level Skip-gram embeddings (300d, 100K vocabulary per decade) trained on the Google Books English Fiction corpus. We aligned all decades to the 1900s reference via Orthogonal Procrustes, computed semantic shift as $1 - \cos(\mathbf{v}^{\text{first}}, \mathbf{v}^{\text{last}})$, and used vocabulary rank as a frequency proxy (following the Zipfian rank–frequency relationship).

Table 8 summarizes the results alongside the Korean findings from the main paper. The English data consistently show a *negative* correlation across all analysis windows, confirming the law of conformity for English, while the Korean data show a *positive* correlation of comparable magnitude.

The sign reversal is consistent across every adjacent time-period pair: all five Korean century pairs yield positive correlations ($r = +0.228$ to $+0.537$), while all 19 English decade pairs yield negative correlations ($r = -0.322$ to -0.822).

Hamilton-style pooled regression. To ensure that the reversal is not an artifact of our shift measurement (first–last cosine distance), we also replicated Hamilton et al.'s exact procedure: computing displacement rates between adjacent time periods, $\Delta^{(t)}(w) = 1 - \cos(\mathbf{v}_w^{(t)}, \mathbf{v}_w^{(t+1)})$, pooling all word–period observations, and regressing displacement on log frequency. Table 9 reports the results.

Our English replication yields $\beta_f = -0.263$, closely matching Hamilton et al.'s reported $\beta_f = -0.27$ and validating our pipeline. Korean yields $\beta_f = +0.435$ (literary, century resolution) to $+0.491$ (news, decade resolution), confirming that the sign reversal holds under the identical measure-

[7] SGNS embeddings from Google Books Fiction (1800–1990), available at https://nlp.stanford.edu/projects/histwords/.

Table 9: Hamilton-style pooled displacement rate regression ($\Delta^{(t)}(w) \sim \beta_f \cdot \log f(w)$). Our English β_f closely replicates Hamilton et al.'s reported value; Korean β_f is positive across all configurations.

Corpus	β_f	**Pooled** r	n
Hamilton (2016) Eng. SGNS	-0.27	—	—
English Gutenberg (ours)	-0.263	-0.263	633,157
Korean literary (century)	$+0.435$	$+0.435$	34,610
Korean literary (decade)	$+0.426$	$+0.426$	20,119
Korean news (decade)	$+0.491$	$+0.491$	467,370

ment procedure. These results rule out the possibility that the reversal is driven by differences in how semantic shift is operationalized.