

ARTICLE

# Less is more: Diachronic evidence for efficient communication in English noun-to-verb conversion over the past century

Gui Wang  and Bin Shao

School of International Studies, Zhejiang University, China

**Corresponding author:** Bin Shao; Email: [binshao@zju.edu.cn](mailto:binshao@zju.edu.cn)

(Received 16 November 2025 UTC; Revised 10 May 2026 UTC; Accepted 27 June 2026 UTC)

## Abstract

Conversion exemplifies one of the most productive word-formation processes in English. Yet how this phenomenon has evolved over time and what factors drive its development remain underexplored. This study investigates changes in noun-to-verb conversion degree and the semantic, morpho-phonological, stylistic and frequency-based factors shaping this process. Using data from the Corpus of Historical American English, we analyzed 637 noun-to-verb conversion pairs from the 1920s to the 2010s and modeled their trajectories with generalized linear mixed effects models. Results reveal a significant increase in conversion degree over time. This reflects a reconfiguration of how efficient communication is achieved, whereby categorical information is increasingly conveyed by context rather than overt morphological marking. We further found that nouns with higher semantic diversity, lower concreteness, lower frequency and shorter phonological forms display stronger tendencies toward verbal expansion. These lexical properties jointly determine how firmly a noun is anchored at the nominal pole of the noun–verb continuum. Overall, the findings portray the dynamics of conversion as a product of communicative pressures, allowing the English lexicon to achieve greater expressivity with fewer but more flexible forms.

**Keywords:** conversion; denominal verb; efficient communication; language change; word class flexibility; semantic diversity

## 1. Introduction

Conversion is a word-formation process that creates new lexical items by altering word class without overt morphological marking (Plag, 2012). A familiar example is the transformation of the noun *hammer* into the verb *to hammer*, meaning ‘to strike with or as with a hammer’. This process is productive in English (Bauer, 2001; Plag, 1999) and has been part of English grammar for centuries (Pinker, 1994, p. 379). Perhaps no other Germanic language derives as many verbs through conversion as



English does. Recent corpus-based morphological work confirms its central role. Based on 16,624 morphosemantically related noun/verb pairs in WordNet, Mititelu et al. (2023) found that conversion accounts for 46.5% of these pairs and competes with overt suffixes (e.g., *-ize*, *-ify*) across almost all semantic relations.

Despite its central role in English morphology, conversion has predominantly been studied from a synchronic perspective, leaving its diachronic development largely underexplored. In particular, how conversion has evolved over time and what factors drive its development remain open questions. To address this gap, the present study examines diachronic changes in the degree of noun-to-verb conversion in English over the past century and investigates the potential factors shaping this process. Conversion degree is defined here as the proportion of the derived form relative to total lemma use. For example, the word *track* had a conversion degree of only 5% in the 1920s (95% noun, 5% verb) but rose to 35% by the 2010s (65% noun, 35% verb), reflecting a substantial increase in the verbal reuse of the nominal base form. Accordingly, this study is concerned with the dynamics of conversion, aiming to offer insights into the broader mechanisms of language change.

The remainder of this article is structured as follows. Section 2 reviews the theoretical background, covering the nature of conversion as a linguistic process, the debate surrounding its directionality, diachronic approaches and proposed blocking factors. Section 3 describes the methodology, including corpus data, conversion pair selection, variable operationalization and the statistical modeling framework. Section 4 reports the quantitative results for both the diachronic trend in conversion degree and the effects of the predictor variables. Section 5 interprets these findings within the efficient communication framework and cognitive grammar, examining how frequency-based, phonological and semantic properties jointly shape the diachronic change in conversion degree. Section 6 offers concluding remarks.

## 2. State of the art

### 2.1. Conversion as a linguistic process

A longstanding question concerning conversion is what kind of linguistic process it represents. Within a morphological approach, one line of work treats conversion as zero affixation, a regular derivational process involving a phonologically null affix (e.g.,  $N + \emptyset \rightarrow V$ ) (Allen, 1978; Kiparsky, 1982; Marchand, 1969). Others reject this assumption. Plag (2012), for instance, argues on the basis of the ‘overt analogue criterion’ that no overt suffix conveys the same broad range of meanings as conversion, and therefore regards it as a morphological process of derivation without any affix.

Still other researchers emphasize the role of lexical innovation. Clark and Clark (1979) view conversion as a form of pragmatic innovation, in which speakers coin new verbs whose meanings are inferred from contextual cues. Lieber (2001, 2005) develops this idea further, proposing that conversion involves ‘relisting’, whereby an existing lexeme is re-entered in the lexicon under a new grammatical category rather than through affixal derivation. In contrast, syntactic approaches (Arad, 2005; Farrell, 2001; Hale & Keyser, 2002) explain conversion in terms of category underspecification, suggesting that roots are stored without grammatical category and receive part-of-speech labels only in syntactic structures.

While these theoretical frameworks offer different structural accounts of conversion, this study adopts a usage-based perspective. Rather than attempting to resolve questions about its underlying syntactic or morphological architecture, we approach conversion as a measurable distributional property of lexical items. Specifically, we focus on how the degree of word class flexibility, that is, the extent to which established lexical items span multiple grammatical categories in actual language use, evolves over time.

## 2.2. Directionality of conversion

Another major theoretical debate, and one that is directly relevant to the methodological framing of this study, concerns the directionality of conversion: whether the morphological relationship between members of a conversion pair is unidirectional or bidirectional. Standard accounts of derivational morphology assume that one lexeme is structurally and conceptually derived from a single simpler base, with information flowing strictly from base to derivative (Aronoff, 1976; Halle & Marantz, 1993; Plag, 2012). Under this view, conversion pairs have a clearly identifiable base form and a derived form. For example, the noun *hammer* is taken as the base, from which the verb *to hammer* is derived.

Recent paradigmatic approaches, however, have raised fundamental challenges to this unidirectional assumption. These approaches hold that the lexicon constitutes a dense network of multidirectional relationships in which no word holds a privileged status as base (Blevins, 2016; Bonami & Strnadová, 2019). Under this view, the members of a conversion pair stand in a bidirectional, nonhierarchical correspondence, and the identification of one member as the ‘source’ and the other as the ‘derivative’ is theoretically unmotivated. Copot and Bonami (2024) tested these competing views through an acceptability judgment experiment in which participants evaluated pseudowords conditional on a morphologically related form. They found that speakers exploit implicative relationships between words regardless of derivational direction, providing behavioral evidence against the rooted tree view and supporting instead a paradigmatic model of bidirectional lexical correspondence. Cross-linguistic evidence from studies of conversion in French (Tribout, 2010, 2020) further supports the view that directionality is underdetermined or bidirectional.

While we acknowledge these theoretical challenges and the possibility of bidirectionality within the mental lexicon, this study maintains an operational focus on noun-to-verb conversion. Adopting a paradigmatic approach, in which directionality is treated as theoretically unmotivated, would require collapsing noun-to-verb and verb-to-noun conversion into a single undifferentiated phenomenon. This risks obscuring important asymmetries in their historical development. Diachronic evidence indicates that the nominal-to-verbal direction has been the more productive and historically dominant pathway in English (Balteiro, 2007; Cannon, 1985; Marchand, 1969), suggesting that the two directions do not develop symmetrically over time. This historical asymmetry provides independent motivation for investigating noun-to-verb conversion as a directional phenomenon in its own right. Adopting this operational focus allows us to ask a question that a category-neutral framing cannot easily address: what factors determine the degree to which an established nominal form expands its verbal usage over time? By tracking changes in the verbal proportion of lemma use across a century, we can model the gradient, usage-based nature of

word class flexibility while remaining agnostic about the deeper representational question of whether nouns are cognitively prior to verbs in speakers' mental lexicons.

### 2.3. *Diachronic approaches to conversion*

While the aforementioned studies occasionally draw on historical observations to establish directional pathways, theoretical accounts of conversion have traditionally remained synchronic, examining the process as a static grammatical phenomenon. Few studies adopt a diachronic perspective, treating conversion as a dynamic process that evolves alongside broader changes in the lexicon and grammar. Such a perspective is essential, as it allows us to trace how patterns of word class flexibility develop over time and examine how they may reflect broader systemic changes in language. For instance, Biese (1941) surveyed the history of conversions in English from the 13th to 19th centuries, while Balteiro (2007) examined the directionality of conversion through combined synchronic and diachronic evidence. More recently, Neubauer (2024) offered a detailed diachronic study of the English nominal lexicon since 1150, tracking how word-formation processes, including nominalization, have typologically evolved over centuries.

Quantitative corpus-based studies have since begun to track these diachronic shifts on a larger scale. Shao et al. (2024) analyzed 877 noun/verb conversion pairs in English from the 1920s to the 2010s and found that usage across nominal and verbal categories has become increasingly balanced over time. This finding indicates a historical drift toward greater word class flexibility in English. They attribute this trend to the long-term loss of inflectional morphology in English. Their study provides one of the first large-scale quantitative accounts of diachronic change in noun/verb conversion pairs, offering evidence that grammatical category boundaries in English have been gradually loosening.

Building on this, Wang and Shao (2025) investigated the mechanisms behind the increased word class flexibility by examining the role of semantic diversity. Using a sample of 607 noun/verb pairs from the 1920s to the 2000s, they found that words with greater overall semantic diversity exhibit higher cross-categorical flexibility. This flexibility is operationalized as the degree of balanced use across noun and verb forms. Crucially, they identified an asymmetry: it is the semantic diversity of the noun senses, rather than the verb senses, that drives this flexibility. They also found that higher permeability in their dataset was driven by an increase in verb usage relative to noun usage. These results provide quantitative, albeit indirect, evidence for a directional preference in noun-to-verb conversion in English.

While Wang and Shao (2025) provide significant insight into the role of semantic diversity in word class flexibility, their approach to directionality was indirect. Like Shao et al. (2024), their initial sample consisted of nondirectional, noun/verb pairs (i.e., words functioning as both nouns and verbs without *a priori* distinguishing the base from the derived form). Their conclusions regarding the directional preference from noun to verb were inferred *a posteriori* from the asymmetric effects of noun versus verb semantic diversity. To the best of our knowledge, a large-scale diachronic study that explicitly selects and tracks a directional sample of established noun-to-verb conversions remains absent from the literature.

Collapsing the two directions yields valuable global trends but risks obscuring potential asymmetries in how each direction develops over time. By focusing on

noun-to-verb conversion, this study moves beyond previous work in two aspects. First, it draws on a directional dataset to directly model the diachronic expansion of verbal usage. Second, while Wang and Shao (2025) focused exclusively on semantic diversity, this study expands the analytical scope by including additional semantic factors (concreteness) alongside morpho-phonological, frequency-based and stylistic predictors, thereby offering a more multidimensional account of the conditions that facilitate or inhibit this diachronic process.

#### 2.4. Blocking factors on conversion

Another underexplored question concerns what factors facilitate or block conversion. This research gap is particularly striking given that conversion is recognized as one of the most productive word-formation mechanisms in English (Bauer, 2001; Plag, 1999) and has served as a major source of lexical innovation since at least the thirteenth century (Gottfurcht, 2008). One reason for this gap may lie in conversion's distinctive properties: its lack of morphological marking (Plag, 2012) and its semantic underspecification (Clark & Clark, 1979; Mititelu et al., 2023). Affixation provides a formal anchor for structural generalizations (e.g., *-ize* selects Latinate bases) and often imposes a narrow semantic profile on its outputs (e.g., agentive *-er* yields persons or instruments). Conversion offers neither: it lacks any overt marker, and converted verbs can express a wide range of semantic relations to their base nouns, including instrument, location, agent and result relations (Mititelu et al., 2023). This combination makes both structural and semantic conditions difficult to identify.

A notable exception is Davies (2004), who conducted a corpus-based investigation using the *Independent/Guardian* newspaper corpus. Davies (2004) extracted all nouns from the 1500 most frequent words in this corpus and used the Oxford English Dictionary (OED) to classify them as having an attested converted verb form or not, then tested a set of proposed blocking factors against these two lists. She proposed 10 potential factors and formulated explicit hypotheses for each, eight of which were operationalized and tested. Specifically, conversion is hypothesized to be inhibited when:

- (1) morphological simplicity holds, meaning a simpler or already existing verb expresses the same meaning; this 'ancestry' relation pre-empts novel conversion and aligns with Gricean economy in avoiding unnecessary duplication in the lexicon (Clark & Clark, 1979; Grice, 1975);
- (2) competing affixation is present, where a derivationally affixed verb (e.g., *hospitalize*) occupies the functional niche; speakers may prefer overt affixes when they convey more specific meanings than the semantically general outcome of conversion (Plag, 1999);
- (3) synonymy applies, such that a well-established near-synonymous verb already conveys the target meaning; 'pre-emption by synonymy' lowers the likelihood that a conversion will be adopted (Clark & Clark, 1979; Plag, 1999);
- (4) homonymy would result, creating potential confusion with an existing word; this is 'pre-emption by homonymy', though its force is context dependent and often weak (Clark & Clark, 1979; Plag, 1999);
- (5) the noun has low concreteness, since abstract nouns have been suggested to resist verbing relative to more concrete ones (Aronoff, 1980);

- (6) domain restriction applies, meaning the noun belongs to semantic fields that tend to be verbed by rival derivations rather than by conversion (e.g., certain proper-name domains; Bladin, 1911);
- (7) word length is greater, longer or structurally more complex bases are dis-preferred on aesthetic and communicative efficiency grounds (Bauer, 2001; Grice, 1975);
- (8) a nominal suffix is present, because overt nominal morphology (such as *-tion*, *-ness*, *-ment*) makes zero derivation to verbs less acceptable (Eschenlohr, 1999; Neef, 2005).

Two additional hypotheses were discussed qualitatively: failure of hypostatization, whereby conversion is inhibited when the noun does not lend itself to an entity construal suitable for predication, and public acceptability, in which lower acceptability norms reduce the likelihood of conversion. Davies (2004) operationalized the eight testable factors through binary coding of their presence or absence for each noun and compared frequencies between converted and nonconverted items. The results showed that morphological simplicity, nominal suffixes and longer word length were the most powerful inhibitors, whereas synonymy and homonymy exerted weaker effects. She concluded that inhibition is gradient rather than categorical.

Davies (2004) made an important contribution to understanding factors that facilitate or inhibit conversion. Several aspects, however, warrant further investigation. First, the study is synchronic and relies on a relatively small sample, which limits the statistical power and precludes examining whether factor effects change over time. Second, it draws on a single-genre corpus (newspaper text), so the extent to which the findings generalize across registers remains uncertain. Research on register variation suggests that more informal contexts (e.g., fiction and conversational media) may foster lexical innovation, whereas more formal prose tends to be conservative (Biber, 1988; Biber & Conrad, 2009; Crystal, 2001). Third, frequency warrants more systematic attention. Usage frequency shapes morphological behavior: high-frequency words tend to resist category change due to entrenchment, whereas low-frequency words are more prone to innovation (Bybee, 2007; Schmid, 2020). Without accounting for such effects, it is difficult to determine whether an observed constraint on conversion reflects a structural limitation or merely differences in how often words occur in discourse.

This study addresses these gaps in three dimensions. First, it adopts a diachronic perspective, examining how the degree of noun-to-verb conversion has evolved over the past century. Focusing on degree rather than the emergence of new conversion items offers two main advantages. Conceptually, it captures conversion as a continuum of word class flexibility rather than a binary outcome of ‘converted’ versus ‘unconverted’, reflecting the gradient nature of grammatical category in actual language. Methodologically, it avoids the data sparsity and definitional ambiguity involved in identifying newly formed conversions, which are rare and difficult to date precisely. Tracking changes in usage proportion for established conversion pairs enables robust statistical inference based on stable lexical items, revealing how the relative strength of nominal and verbal functions shifts across time.

Second, the study complements prior qualitative judgments with quantitative, corpus-based operationalizations of semantic and morphological factors. For example, concreteness is captured using concreteness ratings (Brysbaert et al., 2014) and

semantic diversity is derived from distributional co-occurrence entropy, together reflecting conceptual generality and contextual flexibility in measurable form. Rather than relying on qualitative judgments about ‘domain restriction’, we treat semantic diversity as a continuous proxy for semantic spread. Third, we incorporate frequency and register as two additional predictors, drawing on theories of entrenchment and lexical innovation that posit that high-frequency words resist change while informal contexts promote it.

By analyzing 637 noun-to-verb conversion pairs in a large-scale diachronic corpus, this study (a) quantifies historical change in the degree of conversion in English across the past century and (b) assesses how semantic, morpho-phonological, frequency-based and stylistic factors shape that change. Integrating these dimensions within a unified statistical framework, it offers the first large-scale diachronic account of conversion degree in English and clarifies how multiple factors jointly contribute to the dynamics of conversion, thereby illuminating broader trends in language change.

From a cognitive perspective, noun-to-verb conversion can be understood as a form of word class flexibility governed by general principles of mental representation and language processing. Reinterpreting a noun as a verb requires speakers to activate relational knowledge, infer event structure and integrate context-dependent conceptual mappings (Clark & Clark, 1979). The predictors examined in this study reflect cognitive mechanisms operating at different levels. Word frequency is interpreted as a potential proxy for entrenchment (Bybee, 2007; Schmid, 2020): under the assumption that repeated exposure strengthens mental representations, higher frequency nouns may develop more entrenched nominal representations, making them less susceptible to categorical reassignment. Concreteness and semantic diversity can be understood through the lens of prototype-based construals (Langacker, 1987, 1991): they are theorized to reflect how firmly a noun is anchored at the nominal pole of the noun–verb continuum, with nouns that are more concrete and occur in a more restricted range of contexts being potentially more resistant to verbal reinterpretation. Syllable count is interpreted in terms of processing cost and phonological cuing (Cassidy & Kelly, 1991; Kelly, 1992): shorter words arguably carry weaker phonological commitment to nominal category membership, which may reduce the inferential burden of category reassignment for the listener. Taken together, these cognitive accounts offer a coherent basis for understanding why certain lexical properties determine how readily a noun undergoes verbal reinterpretation.

### 3. Methods

#### 3.1. Corpus data

To investigate diachronic changes in the degree of noun-to-verb conversion in English, we used the Corpus of Historical American English (COHA; Davies, 2012). This corpus was selected for two reasons.

First, COHA provides the temporal depth and scale necessary for robust diachronic analysis. Spanning from the 1820s to the 2010s with over 475 million words, it is the largest structured historical corpus of English. This extensive temporal coverage enables systematic tracking of conversion degree across our target period of the 1920s to the 2010s. As Biber and Gray (2013, p. 108) note, such large-scale

Table 1. Composition of COHA by decade

| Decade | Tokens     |
|--------|------------|
| 1920s  | 23,819,149 |
| 1930s  | 25,710,170 |
| 1940s  | 25,451,525 |
| 1950s  | 26,600,444 |
| 1960s  | 27,023,893 |
| 1970s  | 26,728,218 |
| 1980s  | 27,605,215 |
| 1990s  | 31,034,963 |
| 2000s  | 32,654,928 |
| 2010s  | 32,648,551 |

corpora permit analysis of historical change ‘on a scale previously unachievable with earlier corpus designs’, facilitating nearly continuous monitoring of linguistic developments over time. Crucially, this large historical corpus provides an adequate pool of candidate noun-to-verb conversion pairs for analysis from a diachronic perspective. Table 1 presents the number of tokens across the 10 decades examined in this study.

Second, COHA employs systematic sampling procedures that ensure representativeness across genres. The corpus comprises carefully selected texts from four major categories: newspapers, popular magazines, fiction and nonfiction books, with fiction consistently representing 48–55% of each decade’s content. Davies (2012, p. 123) emphasizes that this balance is maintained across subgenres and domains through systematic sampling (e.g., Library of Congress classification for nonfiction, subgenre categorization for fiction including prose, poetry and drama). This structured approach to genre balance enables confident attribution of observed changes to genuine linguistic development rather than artifacts of changing genre distributions.

3.2. Conversion pair selection

We identified candidate noun-to-verb conversion pairs through a systematic filtering process using the CELEX lexical database (Baayen et al., 1995) and COHA frequency data. Following Kisselew et al. (2016), we first extracted all verbs derived from conversion using the WebCelex interface (<https://webcelex.ivdnt.org/>) with the Perl code: *Class eq ‘V’ && MorphStatus eq ‘Z’*. In this query, *Class = ‘V’* restricts the search to verb lemmas, while *MorphStatus = ‘Z’* specifies zero derivation (i.e., conversion without overt morphological marking). This step yielded 1719 candidate conversion pairs. We then filtered these candidates to retain only those that also contained a noun form using the code: *Class eq ‘N’*, resulting in 1640 lexemes.

At this stage, our list contained words with both verb and noun forms but did not guarantee noun-to-verb derivational direction. For instance, *clean* occurs as both a noun and a verb, but it also has an adjectival form. The presence of an adjective introduces an additional potential derivational pathway (e.g., adjective → verb), making it difficult to determine whether the verb is derived from the noun or from the adjective. To avoid such ambiguity in directionality, we restricted the dataset to lexemes attested in exactly two word classes, namely noun and verb. This filtering step yielded 1361 candidates.

To ensure robust statistical analysis across all factors, we applied additional frequency thresholds. Specifically, we retained only words meeting two criteria simultaneously: (1) the noun form must have a minimum of 50 tokens across all decades (1920s–2010s) and (2) the verb form must have a minimum of 100 total tokens and appear in at least 8 of the 10 decades. These thresholds were chosen to filter out infrequent and potentially idiosyncratic forms while maintaining a sufficiently large sample. A minimum of 50 total tokens ensures the base noun is a well-established lexical item in the corpus. The asymmetry between the noun and verb thresholds is intentional, as the converted form naturally tends to be less frequent than the base form. Applying an equally strict per-decade threshold to the verb would eliminate most legitimate conversion candidates and introduce selection bias toward only the most extreme cases. The verb threshold therefore prioritizes sufficient temporal coverage for reliable diachronic modeling. These criteria yielded 639 valid noun-to-verb conversion pairs. We excluded two additional words (*eddy* and *rein*) due to missing data in the semantic concreteness dataset (Brysbaert et al., 2014), resulting in a final dataset of 637 conversion pairs (see [Appendix 1](#) for the complete list).

### 3.3. Operationalization of variables

To investigate the factors contributing to diachronic changes in the degree of noun-to-verb conversion, we operationalized our dependent variable and predictor variables across four dimensions: semantic, morpho-phonological, frequency-based and stylistic factors. The dependent variable, conversion degree, was operationalized as the proportional usage of the verbal form relative to total lemma use for each conversion pair in each decade, calculated as:  $\text{verb\_freq}/(\text{verb\_freq} + \text{noun\_freq})$ . This ratio produces values bounded between 0 and 1, where values close to 0 indicate low conversion degree with the word predominantly used in its nominal form, and values close to 1 indicate high conversion degree with the word predominantly used in its verbal form. This measure captures the dynamic nature of conversion, allowing us to track how the degree of verbalization evolves over time for individual lexical items.

All predictor variables were calculated based on the noun form of each conversion pair. This focus on the base noun form is theoretically motivated, as we aim to examine how properties of the original noun predict its subsequent tendency to develop verbal usage patterns over time.

As for semantic factors, we examined two dimensions: semantic diversity and concreteness. Semantic diversity was calculated using the PPMI-based method established by Kisselew et al. (2016). We constructed a global context vocabulary of the top 10,000 content words (nouns, verbs, adjectives and adverbs) across all decades to establish a stable semantic coordinate system throughout the century-long timespan. For each noun-to-verb conversion pair in each decade, we collected co-occurrence information within a five-word window, focusing exclusively on contexts where the target word appeared as a noun. Raw co-occurrence counts were transformed using Positive Pointwise Mutual Information (PPMI) to correct for frequency biases, then converted to probability distributions via L1 normalization.

This PPMI-based approach offers several methodological advantages over alternative distributional methods such as word2vec embeddings (Mikolov et al., 2013). PPMI values are directly interpretable as measures of association strength between

words, representing how much more likely two words are to co-occur than would be expected by chance. Additionally, the resulting probability distributions maintain a transparent, probabilistic interpretation where each dimension corresponds to a specific context word, enabling direct examination of which semantic domains contribute to a word's diversity score. Crucially for diachronic analysis, this approach eliminates the need for complex vector alignment procedures across time periods that are required when using neural embedding methods like word2vec, as each word's semantic representation is calculated independently within its temporal context using the same global vocabulary framework. This interpretability is essential for diachronic analysis, as it allows researchers to trace how specific semantic associations evolve over time without introducing artifacts from alignment procedures.

Following Wang and Shao (2025), we quantified semantic diversity using Shannon entropy,<sup>1</sup> which measures the uncertainty or dispersion of a word's co-occurrence distribution across contextual dimensions. This index has been adopted across diverse fields, including biodiversity measurement (Allen et al., 2009; Dickman, 1968) and lexical diversity research (Jarvis, 2013; Shi & Lei, 2022), as it captures both the richness (i.e., number of categories) and the evenness (i.e., relative distribution) of a given system. In this study, higher entropy values indicate that a noun's semantic associations are broadly distributed across many different contextual dimensions, suggesting greater semantic flexibility and polysemy. Conversely, lower entropy values indicate more concentrated, domain-specific usage patterns. To verify that our findings do not depend on this particular diversity metric, we computed two Hill number variants from the same PPMI distributions; the resulting near-perfect rank correlations with Shannon entropy ( $\rho \geq 0.9995$  across all decades) confirm that Shannon entropy is a representative measure of semantic diversity in our data.<sup>2</sup>

To operationalize concreteness, we used concreteness scores from Brysbaert et al. (2014). The dataset provides empirically derived ratings for approximately 40,000 English words, based on crowdsourced judgments from over 4000 participants. These ratings indicate the degree to which word meanings refer to perceptible entities on a 5-point scale (1 = very abstract, 5 = very concrete), with higher scores reflecting greater concreteness. For example, highly abstract words like *value* receive low ratings (1.62), while highly concrete words like *flower* receive high ratings (5.00). This dataset has been adopted across multiple disciplines, including

<sup>1</sup>Shannon entropy is calculated as  $H(X) = -\sum p(x_i) \cdot \log_2 p(x_i)$ , where  $X$  represents the target word's co-occurrence distribution across the corpus,  $x_i$  denotes an individual context word, and  $p(x_i)$  is the probability of the target word co-occurring with context word  $i$ , derived from L1-normalized PPMI values. Higher entropy reflects a more evenly distributed co-occurrence profile across many contextual dimensions; lower entropy reflects concentration within a narrow set of semantic domains.

<sup>2</sup>Diversity metrics differ in their sensitivity to rare versus common contextual dimensions (Roswell et al., 2021). Hill numbers provide a unified framework in which different metrics are special cases of a single equation parameterized by an exponent that controls this sensitivity. We computed Hill–Shannon (the exponential of Shannon entropy, weighting rare and common contexts proportionally) and Hill–Simpson (the inverse of the sum of squared probabilities, emphasizing dominant contexts and most robust to sparse distributions). Spearman's rank correlations between Shannon entropy and Hill–Simpson ranged from 0.9995 to 0.9997 across the 10 individual decades ( $n = 637$  words per decade; 6390 word-decade observations in total), indicating that the ordering of words by semantic diversity is essentially invariant across metrics.

cognitive psychology, neuroscience and computational linguistics, to investigate topics ranging from word processing and memory to bilingual cognition and semantic representation, thereby establishing its reliability and validity as a measure of conceptual concreteness (e.g., Pereira et al., 2018; Peters, 2019; Tang et al., 2023; Trott, 2024).

Regarding morpho-phonological factors, nominal suffix presence was determined through systematic manual coding of all 637 conversion pairs by three linguistics graduate students working independently. Each annotator coded the complete list separately as having overt nominal morphology (e.g., *-tion*, *-ment*, *-ness*) or lacking such suffixes. This process achieved 96.9% interrater agreement (20 of 637 cases showed initial disagreement across the three independent annotations), with Fleiss' Kappa = 0.941 and a mean PABAK of 0.958 across the three rater pairs, both indicating almost perfect agreement (Landis & Koch, 1977). Uncertain cases were resolved by consultation with a linguistics professor. Importantly, we relied on synchronic morphological transparency rather than etymological analysis, reasoning that most language users lack detailed etymological knowledge and that contemporary morphological awareness better reflects processing constraints on conversion. Our analysis revealed that 34 of 637 words contained nominal suffixes (5.3%).

Syllable count was automatically calculated using the R package *nsyllable* (Benoit, 2020), which references the Carnegie Mellon University Pronouncing Dictionary and estimates syllables as vowel sequences for unlisted words. We validated this approach by manually checking 50 randomly sampled words against Merriam-Webster pronunciations, achieving 100% accuracy.

For the frequency-based factor, log-normalized frequency was calculated for the noun form of each conversion pair within each decade. Raw frequency counts were divided by the total corpus size per decade, then multiplied by 1,000,000 to obtain frequency per million words (pmw) and subsequently log-transformed to address the characteristic right-skewed distribution of word frequencies (Zipf distribution). This transformation provides a more linear relationship with linguistic processing costs and conversion probability.

Regarding the stylistic factor, informality proportion was operationalized as the ratio of informal register usage: (fiction + magazines + television/movies)/(fiction + magazines + television/movies + newspapers + academic texts). This measure captures the hypothesis that informal contexts promote lexical innovation and may therefore be associated with higher conversion degree. The denominator includes all major genre categories in COHA to ensure proportional representation across registers, while the numerator covers genres associated with informal language use.

All continuous variables were z-score standardized prior to statistical modeling to enable direct comparison of effect sizes across predictors and to facilitate model convergence in the subsequent generalized linear mixed effects model (GLMM) analysis.

### 3.4. Data analysis

To examine the overall diachronic trend in conversion degree, we fitted a baseline GLMM with decade as a fixed effect and conversion pair as a random intercept. To examine the factors contributing to conversion degree, we fitted a comprehensive GLMM that additionally included semantic diversity, concreteness, syllable count,

nominal suffix presence, log-normalized noun frequency and informality proportion as fixed effects.

A GLMM framework was appropriate because conversion degree is a proportion bounded between 0 and 1, and because the same conversion pairs are observed repeatedly across 10 decades. The mixed effects structure accommodates both the proportional nature of the outcome and the within-word dependencies introduced by repeated measurement. We adopted a random-intercept rather than random-slope structure, as the predictors represent intrinsic lexical properties for which item-specific slopes are not theoretically motivated; preliminary models with by-lemma slopes also yielded singular fits without improving model performance. Decade was treated as a categorical factor to allow for nonlinear temporal patterns, with the 1920s as the reference level for the baseline model and the 1930s for the comprehensive model.

To address potential reverse causality in the comprehensive model, we adopted a lag design in which predictors measured in decade  $t - 1$  were used to predict conversion degree in decade  $t$  (i.e., predictors from the 1920s–2000s predicting outcomes in the 1930s–2010s). All continuous predictors were  $z$  standardized prior to model fitting, and multicollinearity was acceptable, with all variance inflation factors below 5 (Shadish et al., 2002).

Both models were initially fitted with a binomial distribution but showed substantial overdispersion. We therefore refitted them with a beta-binomial distribution using the *glmmTMB* package (Brooks et al., 2017) and confirmed model adequacy using simulation based diagnostics from the *DHARMa* package (Hartig et al., 2024). Effect directions and significance patterns were preserved across the binomial and beta-binomial specifications. However, the beta-binomial produced more conservative and better-calibrated standard errors, and is therefore retained as the final specification. Full diagnostic details are reported in [Appendix 2](#).

## 4. Results

### 4.1. Diachronic trend in degree of noun-to-verb conversion

The baseline GLMM revealed a significant and consistent increase in noun-to-verb conversion degree from the 1920s to the 2010s. [Figure 1](#) shows the predicted degree of verbalization for each decade, demonstrating a clear upward trajectory throughout the time series. The predicted conversion degree increased from 11.2% (95% CI: 10.2, 12.3) in the 1920s to 14.4% (95% CI: 13.1, 15.7) in the 2010s, representing a relative increase of approximately 29% over the nine-decade span.

As shown in [Table 2](#), all decade coefficients were positive and significant (all  $p \leq 0.001$ ), providing robust evidence for an upward trend in conversion degree. The coefficients generally increase in magnitude over time, with the 2010s ( $\beta = 0.288$ ) substantially larger than earlier decades. In addition, the model revealed substantial random intercept variance ( $\sigma^2 = 1.780$ ), indicating considerable heterogeneity among conversion pairs in their baseline propensity for verbalization. This lexical-level heterogeneity aligns with the claim that conversion does not form a uniform morphological category (Neef, 2005). In other words, it lacks a consistent set of formal or semantic patterns typical of regular word-formation processes such as affixation or compounding and instead functions as a context-dependent mechanism for lexical innovation.

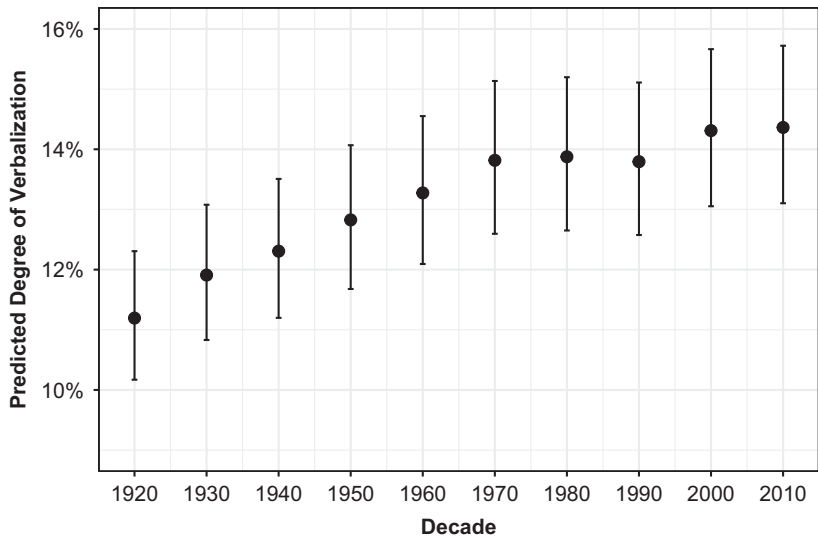


**Figure 1.** Changes in the predicted degree of noun-to-verb conversion from the 1920s to the 2010s. Points represent model-predicted values from the baseline beta-binomial generalized linear mixed effects model and error bars indicate 95% confidence intervals.

**Table 2.** Fixed effects of decade in baseline beta-binomial generalized linear mixed-effects model

| Decade      | $\beta$ | SE    | z      | p      | OR    | 95% CI         |
|-------------|---------|-------|--------|--------|-------|----------------|
| (Intercept) | -2.108  | 0.055 | -38.14 | <0.001 | 0.121 | [0.109, 0.135] |
| 1930s       | 0.071   | 0.022 | 3.27   | 0.001  | 1.073 | [1.029, 1.119] |
| 1940s       | 0.108   | 0.021 | 5.04   | <0.001 | 1.114 | [1.069, 1.161] |
| 1950s       | 0.156   | 0.021 | 7.30   | <0.001 | 1.169 | [1.121, 1.219] |
| 1960s       | 0.196   | 0.021 | 9.20   | <0.001 | 1.217 | [1.167, 1.269] |
| 1970s       | 0.243   | 0.021 | 11.42  | <0.001 | 1.275 | [1.223, 1.330] |
| 1980s       | 0.248   | 0.021 | 11.69  | <0.001 | 1.281 | [1.228, 1.337] |
| 1990s       | 0.241   | 0.021 | 11.39  | <0.001 | 1.272 | [1.219, 1.328] |
| 2000s       | 0.284   | 0.021 | 13.48  | <0.001 | 1.329 | [1.274, 1.386] |
| 2010s       | 0.288   | 0.021 | 13.70  | <0.001 | 1.334 | [1.279, 1.392] |

Note: Reference category is 1920s;  $\beta$  = log-odds coefficient; OR = odds ratio; CI = confidence interval.

#### 4.2. Effects of semantic, morpho-phonological, stylistic and frequency-based factors

The comprehensive beta-binomial GLMM incorporating lagged predictors revealed significant effects for most factors of conversion degree, while also confirming the persistent temporal trend observed in the baseline model. Table 3 presents the complete fixed effects results covering both log-odds coefficients and odds ratios.

Semantic factors demonstrated robust predictive effects on conversion degree. Specifically, semantic diversity exhibited a strong positive effect ( $\beta = 0.294$ , OR = 1.340,  $p < 0.001$ ). This indicates that a one-standard deviation increase in semantic diversity in the previous decade was associated with a 34% increase in the odds of verbal usage. Conversely, concreteness showed a significant negative effect ( $\beta = -0.253$ , OR = 0.777,  $p < 0.001$ ), indicating that more abstract nouns show higher conversion

**Table 3.** Fixed effects of potential factors on predicting noun-to-verb conversion degree

| Variables                          | $\beta$ | SE    | $z$    | $p$    | OR    | 95% CI         |
|------------------------------------|---------|-------|--------|--------|-------|----------------|
| (Intercept)                        | −2.004  | 0.057 | −35.19 | <0.001 | 0.135 | [0.121, 0.151] |
| Semantic diversity ( $t - 1$ )     | 0.294   | 0.046 | 6.46   | <0.001 | 1.340 | [1.230, 1.470] |
| Concreteness                       | −0.253  | 0.048 | −5.21  | <0.001 | 0.777 | [0.706, 0.854] |
| Syllable count                     | −0.146  | 0.053 | −2.73  | 0.006  | 0.864 | [0.778, 0.960] |
| Nominal suffix presence            | 0.016   | 0.124 | 0.13   | 0.896  | 1.020 | [0.798, 1.300] |
| Frequency ( $t - 1$ )              | −0.630  | 0.051 | −12.45 | <0.001 | 0.533 | [0.482, 0.588] |
| Informality proportion ( $t - 1$ ) | −0.013  | 0.010 | −1.29  | 0.196  | 0.987 | [0.969, 1.010] |
| Decade: 1940                       | 0.006   | 0.020 | 0.31   | 0.755  | 1.010 | [0.967, 1.050] |
| Decade: 1950                       | 0.059   | 0.020 | 2.91   | 0.004  | 1.060 | [1.020, 1.100] |
| Decade: 1960                       | 0.079   | 0.021 | 3.75   | <0.001 | 1.080 | [1.040, 1.130] |
| Decade: 1970                       | 0.115   | 0.021 | 5.44   | <0.001 | 1.120 | [1.080, 1.170] |
| Decade: 1980                       | 0.114   | 0.022 | 5.30   | <0.001 | 1.120 | [1.080, 1.170] |
| Decade: 1990                       | 0.107   | 0.022 | 4.85   | <0.001 | 1.110 | [1.070, 1.160] |
| Decade: 2000                       | 0.139   | 0.025 | 5.65   | <0.001 | 1.150 | [1.100, 1.210] |
| Decade: 2010                       | 0.136   | 0.026 | 5.17   | <0.001 | 1.150 | [1.090, 1.210] |

Note: Reference category for decades is 1930s;  $\beta$  = log-odds coefficient; OR = odds ratio; CI = confidence interval; all continuous predictors were z-standardized prior to analysis.

degrees. A one-standard deviation increase in concreteness was associated with a 22% decrease in the odds of verbal usage.

Morpho-phonological factors showed mixed results. Syllable count exhibited a significant negative effect ( $\beta = -0.146$ , OR = 0.864,  $p = 0.006$ ), indicating that phonologically more complex words tend toward lower conversion degrees. A one standard deviation increase in syllable count was associated with a 14% decrease in the odds of verbal usage. However, nominal suffix presence showed no significant effect ( $\beta = 0.016$ , OR = 1.020,  $p = 0.896$ ), suggesting that overt morphological marking does not substantially affect conversion degree patterns in our dataset.

Frequency-based factors revealed the most substantial effect magnitude in the model. Log-normalized noun frequency demonstrated a strong negative relationship with conversion degree ( $\beta = -0.630$ , OR = 0.533,  $p < 0.001$ ), with a one-standard deviation increase in frequency associated with a 47% decrease in the odds of verbal usage. This finding indicates that higher frequency nouns tend to maintain stronger nominal usage patterns.

The informality proportion of noun usage showed no significant effect on subsequent conversion degree ( $\beta = -0.013$ , OR = 0.987,  $p = 0.196$ ). This finding suggests that register variation in nominal usage contexts does not predict future verbalization patterns, contrary to theoretical expectations about informal contexts promoting lexical innovation. The temporal effects remained robust after controlling for all predictor variables, with significant decade coefficients from the 1950s onward (all  $p < 0.01$ ). The magnitude of temporal effects generally increased in later decades, with the 2000s ( $\beta = 0.139$ , OR = 1.150) showing the largest effect and the 2010s ( $\beta = 0.136$ , OR = 1.150) the second largest, relative to the 1930s baseline. This pattern confirms that historical trends toward increased verbalization operate independently of the measured lexical factors.

The comprehensive model also demonstrated improved fit compared to the baseline model (AIC = 57,572.1 vs. 64,190.6), while maintaining acceptable multicollinearity levels with all VIF values below 5. The reduced random intercept variance ( $\sigma^2 = 1.372$  vs. 1.780) indicates that the included predictors account for 22.9% of between-word variation in baseline conversion propensity.

## 5. Discussion

### 5.1. Efficient communication and the rise of conversion degree

This study reveals a significant increase in the degree of noun-to-verb conversion in English from the 1920s to the 2010s. This long-term trend reflects growing word class flexibility, supporting the view that grammatical category boundaries are fuzzy rather than absolute (Rosch, 1973; Taylor, 1995).

One existing account links the rise in word class flexibility to the historical loss of inflectional morphology, or ‘deflexion’, on the grounds that poor inflectional morphology blurs categorical boundaries (Bauer & Valera, 2005; Berg, 2014; Kastovsky, 1985). Shao et al. (2024) similarly attribute the increasing balance between noun and verb use in conversion pairs to the long-term morphological simplification of English. This account, however, cannot directly explain the trend observed in this study. The major phase of deflexion in English occurred during the Middle English period (C. Allen, 2016), and English inflectional morphology has not undergone further meaningful simplification over the past 100 years. Deflexion is therefore better understood as a structural precondition than as a direct cause: the relative poverty of English inflectional morphology removes a formal barrier that, in morphologically richer languages, keeps grammatical categories more clearly delimited.

A more direct explanation can be developed within the efficient communication framework (Gibson et al., 2019; Regier et al., 2016; Xu et al., 2024). This framework holds that language structure reflects a balance between the competing needs of speakers and listeners. As Zipf (1949) observed, speakers favor economy of form by minimizing articulatory and cognitive effort through form reuse, while listeners require sufficient distinctiveness to recover intended meanings accurately. At first glance, the historical rise in conversion degree appears to violate this balance. Reusing the same form across grammatical categories potentially increases ambiguity and shifts more interpretive responsibility onto the listener.

However, recent work suggests that ambiguity can be a functional property of language, provided that context offers sufficient compensatory support (Piantadosi et al., 2012). Xu et al. (2024) extend this principle to form reuse, showing that reuse achieves efficient trade-offs by maintaining sufficient informativeness for listeners while minimizing speaker effort. Conversion can be regarded as a special case of form reuse, one in which the shared form also shifts grammatical category. Therefore, the rise in conversion degree remains compatible with efficient communication: speakers gain articulatory and cognitive economy through form reuse, while listeners, despite the absence of overt morphological marking, can still recover the appropriate category from contextual information.

The increased reliance on contextual information is not specific to conversion but part of a broader historical drift in English. Hawkins (2019) characterizes this drift as a systematic expansion of word-external property assignment, whereby grammatical and lexical properties once realized within individual words have come to depend on neighboring words for their assignment. Hawkins (2019) explicitly identifies noun/verb category ambiguity as one of the central instances of this drift: when a form such as *run* or *play* is no longer disambiguated by inflectional morphology, its category must be recovered from co-occurring articles, infinitival *to* or finite morphology in the surrounding context. The expansion of light verb constructions such as *take a walk* exemplifies the same logic at the level of lexical meaning, where the meaning of the combination is determined more by the nominal complement than by the verb itself.

## 5.2. Lexical properties shaping conversion degree: anchoring along the noun–verb continuum

Beyond the overall trend, our study identified four lexical properties that significantly predict conversion degree: nouns with higher semantic diversity, lower concreteness, lower frequency and shorter phonological forms display stronger tendencies toward verbal expansion. These four effects can be unified within Langacker's (1987, 1991) cognitive grammar framework. In this framework, nouns and verbs are not discrete categories but prototype-based construals: nouns prototypically designate things, with physical objects as the central case, while verbs prototypically designate processes, defined as sequences of relational states unfolding through conceived time. These two construals represent poles of a continuum, and each significant predictor can be interpreted as an index of how firmly a noun is anchored at the nominal pole. The more weakly a noun is anchored, the more readily it extends into verbal function. The continuum thus provides an interpretive lens for relating the four effects, while the predictor-specific mechanisms that drive each effect are addressed in turn below.

Semantic diversity has a strong positive effect: base nouns that occur across a wider range of contexts are more likely to extend into verbal usage. This positive effect aligns with Wang and Shao (2025), who found that higher noun semantic diversity predicts a more balanced distribution of verbal and nominal usage in noun/verb conversion pairs, with the shift toward balance driven by an increase in verbal usage. Their result supports the interpretation that the semantic diversity of the base noun is a key facilitator of verbal extension.

The mechanism by which semantic diversity facilitates conversion can be understood as follows. Nouns with high semantic diversity might evoke richer frames in the sense of Fillmore (1982): they activate both the entity itself and elements of the broader scene in which it typically participates, such as agents, instruments, locations, manners and outcomes. These frame elements provide candidate metonymic pathways for verbal reinterpretation, given that noun-to-verb conversion is often grounded in metonymic relations between an entity and an action it participates in (Clark & Clark, 1979; Dirven, 1999). For example, *head* participates in multiple scenes: as the leader of a group, as a body part used to strike a ball and as the front-facing part of a moving entity. Each scene licenses a distinct verbal extension, with *head a committee* (LEADER-FOR-LEADING), *head the ball* (BODY-PART-FOR-INSTRUMENT-USE) and *head north* (FRONT-FOR-DIRECTION-OF-MOTION) each drawing on a different element of the noun's associated frame. We suggest, then, that nouns with higher semantic diversity afford richer associated frames, which in turn make a greater number of metonymic routes available in principle, lowering the inferential cost of reinterpreting the noun as a verb.

Concreteness, by contrast, exerts a significant negative effect: more concrete nouns tend to resist further conversion into verbs, whereas abstract nouns are more likely to expand their verbal uses. This finding stands in contrast to Aronoff (1980), who argued that abstract nouns resist conversion. Such an inconsistency may stem from Aronoff's (1980) reliance on highly abstract nouns such as *science* and *justice*, which sit at the extreme end of the abstractness scale and lack the event-relatable content needed for verbal reanalysis.

The negative effect of concreteness may be understood through prototype theory. Verbs prototypically designate relational, temporally evolving processes, while nouns prototypically designate bounded, perceptually grounded things (Langacker, 1987).

Empirically, this categorical difference is reflected in concreteness ratings: verbs are rated as significantly less concrete than nouns (Lievers et al., 2021). Abstract nouns therefore share with verbs a relational, less perceptually grounded semantic profile, which makes them more amenable to verbal reinterpretation than concrete nouns. Villani et al. (2024) further show that abstract concepts tend toward lower specificity and broader semantic inclusiveness, meaning they are not tied to particular perceptual referents and can support a wider range of contextual reanalyses. By contrast, concrete nouns are perceptually specific and categorially rigid, which constrains their extension into verbal use.

Turning to morpho-phonological factors, syllable count shows a significant negative effect, indicating that shorter and phonologically simpler base nouns are more likely to expand their verbal uses. This pattern is consistent with the phonological cuing of grammatical category. Cassidy and Kelly (1991) and Kelly (1992) demonstrate that English verbs contain systematically fewer syllables than nouns, and that both adults and children are sensitive to this distributional asymmetry, using syllable number as a probabilistic cue to grammatical category assignment. Shorter words therefore carry weaker phonological commitment to nominal category membership, leaving their grammatical category more open to contextual reassignment. Longer words, by contrast, are more likely to bear phonological properties associated with nouns, which anchors their categorical interpretation and constrains functional extension. The presence of a nominal suffix, however, shows no significant effect, likely due to the small number of suffixed items in the dataset (about 5%). Nonetheless, the rarity of suffixed bases among conversion pairs indirectly supports the idea that overt nominal morphology inhibits conversion onset (Davies, 2004).

Word frequency shows a significant negative effect on conversion degree, consistent with usage-based theories of entrenchment (Bybee, 2007). Frequent use strengthens linguistic representations, making them highly accessible but less flexible. High-frequency nouns are overwhelmingly experienced in nominal functions and develop stable, automatized representations that resist categorical change (Langacker, 1987). Lower frequency nouns, by contrast, have weaker category associations and are therefore more open to analogical extension and functional shifts. The finding reflects a general trade-off between accessibility and adaptability: frequent words are easier to retrieve (Ellis, 2002; Gahl, 2008) but harder to reanalyze or restructure (Bybee, 2007), whereas less frequent ones remain semantically and syntactically adaptable.

Informality, by contrast, did not show a significant effect. This may be because this study focused on established conversion pairs rather than newly coined ones. Previous studies have shown that informal registers are more conducive to lexical innovation (Biber, 1988; Crystal, 2001). Genre differences may therefore affect the emergence of new conversion pairs but play a limited role in the further development of already conventionalized ones.

Taken together, the significant predictors converge on a unified interpretation: each indexes how firmly a noun is anchored at the nominal pole of the noun–verb continuum, and jointly they predict the degree to which verbal expansion occurs over time. This lexical-level account complements the efficient communication account developed earlier. Specifically, the systematic shift toward word-external property assignment, of which conversion is one instance, creates the conditions under which form reuse becomes communicatively viable, while the lexical properties of individual nouns determine which items participate most readily. More broadly, this study

suggests that English has evolved toward greater word class flexibility, with individual word forms increasingly serving an expanding range of grammatical functions. This long-term trend reflects a structural adaptation in which the semantic, formal and usage-based properties of words jointly facilitate the reuse of existing forms, enabling the language to express more through fewer but more flexible forms.

## 6. Conclusion

This study examined diachronic changes in the degree of noun-to-verb conversion in English, drawing on 637 conversion pairs from the 1920s to the 2010s. The results reveal a significant upward trend in conversion degree, indicating that English has become increasingly flexible in its use of grammatical categories. To our knowledge, this is the first large-scale diachronic study to systematically capture both the temporal trajectory and the underlying constraints of conversion degree.

This long-term trend is best understood as a reconfiguration of how efficient communication is achieved. Speakers rely on form reuse for economy, while listeners recover grammatical category from contextual information. This increased reliance on context is consistent with the broader typological tendency in English toward word-external property assignment. It also aligns with the efficient communication principle that ambiguity is communicatively viable when context provides sufficient compensatory support. Deflexion, while not a direct cause of the observed trend, serves as a structural precondition that removes the formal barrier to word class flexibility.

Beyond the overall trajectory, the results reveal several predictors of conversion degree. Nouns with greater semantic diversity, more abstract meanings, shorter phonological forms and lower frequency are all more likely to expand their verbal uses over time. Together, these factors position individual nouns along the noun–verb continuum: the more weakly a noun is anchored at the nominal pole by its semantic, phonological and frequency profile, the more readily it extends into verbal function.

Several limitations should be acknowledged. This study focuses exclusively on noun-to-verb conversion, leaving open whether similar efficiency-driven patterns occur in other conversion types, such as verb to noun or adjective to verb. Moreover, because the analyses are restricted to English, future cross-linguistic research is needed to test whether the same principles operate in languages with different degrees of morphological richness. Comparative evidence from languages with minimal inflectional morphology, such as Mandarin Chinese, and those with extensive inflectional marking, such as German or Italian, would help to determine whether the mechanisms identified here represent a universal functional tendency or an outcome specific to the structural conditions of English.

**Data availability statement.** The derived data and analysis scripts supporting the findings of this study are available in the Open Science Framework repository at <https://osf.io/vdut8/>. The raw corpus data are from the Corpus of Historical American English and are subject to the access and licensing conditions of the corpus provider.

**Competing interests.** The authors declare none.

**Ethical approval.** This study did not require ethical approval because it involved the computational analysis of publicly available linguistic corpora and did not involve human participants or animal subjects.

## References

- Allen, B., Kon, M., & Bar-Yam, Y. (2009). A new phylogenetic diversity measure generalizing the Shannon index and its application to Phyllostomid bats. *American Naturalist*, 174(2), 236–243. <https://doi.org/10.1086/600101>
- Allen, C. (2016). Typological change: Investigating loss of inflection in early English. In M. Kytö & P. Pahta (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 444–459). Cambridge University Press. <https://doi.org/10.1017/CBO9781139600231.027>
- Allen, M. (1978). *Morphological investigations* [Doctoral dissertation]. University of Connecticut.
- Arad, M. (2005). *Roots and patterns* (Vol. 63). Springer-Verlag.
- Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.
- Aronoff, M. (1980). Contextuals. *Language*, 56(4), 744–758. <https://doi.org/10.2307/413486>
- Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). CELEX2 (p. 287744 KB) [Dataset]. Linguistic Data Consortium. <https://doi.org/10.35111/GS6S-GM48>
- Balteiro, I. (2007). *The directionality of conversion in English*. Peter Lang AG, Internationaler Verlag der Wissenschaften. <https://www.peterlang.com/document/1102513>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2003). *lme4: Linear Mixed-Effects Models using “Eigen” and S4* (p.1.1-37) [Dataset]. <https://doi.org/10.32614/CRAN.package.lme4>
- Bauer, L. (2001). *Morphological productivity*. Cambridge University Press.
- Bauer, L., & Valera, S. (2005). *Approaches to conversion/zero-derivation*. Waxmann.
- Benoit, K. (2020). *nsyllable: Count syllables in character vectors* (p.1.0.1) [Dataset]. <https://doi.org/10.32614/CRAN.package.nsyllable>
- Berg, T. (2014). Boundary permeability: A parameter for linguistic typology. *Linguistic Typology*, 18(3), 489–531. <https://doi.org/10.1515/lingty-2014-0020>
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Biber, D., & Gray, B. (2013). Being specific about historical change: The influence of sub-register. *Journal of English Linguistics*, 41(2), 104–134. <https://doi.org/10.1177/0075424212472509>
- Biese, Y. M. (1941). *Origin and development of conversions in English*. Suomalainen Tiedekatemia.
- Bladin, V. (1911). *Studies on denominative verbs in English*. Almqvist and Wiksell.
- Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford University Press.
- Bonami, O., & Strnadová, J. (2019). Paradigm structure and predictability in derivational morphology. *Morphology*, 29(2), 167–197. <https://doi.org/10.1007/s11525-018-9322-6>
- Brooks, M. E., Kristensen, K., Benthem, K. J. v., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal*, 9(2), 378–400. <https://doi.org/10.32614/RJ-2017-066>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Bybee, J. (2007). *Frequency of use and the organization of language* (1st ed.). Oxford University Press.
- Cannon, G. (1985). Functional shift in English. *Linguistics*, 277(1), 411–432. <https://doi.org/10.1515/ling.1985.277.1.411>
- Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, 30(3), 348–369. [https://doi.org/10.1016/0749-596X\(91\)90041-H](https://doi.org/10.1016/0749-596X(91)90041-H)
- Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 55(4), 767–811. <https://doi.org/10.2307/412745>
- Copot, M., & Bonami, O. (2024). Baseless derivation: The behavioural reality of derivational paradigms. *Cognitive Linguistics*, 35(2), 221–250. <https://doi.org/10.1515/cog-2023-0018>
- Crystal, D. (2001). *Language and the internet*. Cambridge University Press.
- Davies, C. (2004). *A corpus-based investigation of noun-to-verb conversion in English* [Doctoral dissertation, University of Liverpool]. <https://livrepository.liverpool.ac.uk/3174931/1/416082.pdf>
- Davies, M. (2012). Expanding horizons in historical linguistics with the 400-million word corpus of historical American English. *Corpora*, 7(2), 121–157. <https://doi.org/10.3366/cor.2012.0024>
- Dickman, M. (1968). Some indices of diversity. *Ecology*, 49(6), 1191–1193. <https://doi.org/10.2307/1934512>

- Dirven, R. (1999). Conversion as a conceptual metonymy of event schemata. In K.-U. Panther & G. Radden (Eds.), *Human cognitive processing* (Vol. 4, pp. 275–288). John Benjamins Publishing Company.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Eschenlohr, S. (1999). *Vom Nomen zum Verb: Konversion, Präfigierung und Rückbildung im Deutschen*. Olms.
- Farrell, P. (2001). Functional shift as category underspecification. *English Language and Linguistics*, 5(1), 109–130. <https://doi.org/10.1017/S1360674301000156>
- Fillmore, C. J. (1982). Frame semantics. In Linguistic Society of Korea (Ed.), *Linguistics in the morning calm: Selected papers from SICOL-1981* (pp. 111–137). Hanshin Publishing.
- Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3), 474–496. <https://doi.org/10.1353/lan.0.0035>
- Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5), 389–407. <https://doi.org/10.1016/j.tics.2019.02.003>
- Gottfurcht, C. A. (2008). *Denominal verb formation in English* [Doctoral dissertation], Northwestern University]. <https://www.proquest.com/openview/150d1ed1416a0d53fe194190f2554a2b/1?cbl=18750&pq-origsite=gscholar>
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). BRILL.
- Hale, K., & Keyser, S. J. (2002). *Prolegomenon to a theory of argument structure*. The MIT Press.
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20* (pp. 111–176). The MIT Press.
- Hartig, F., Lohse, L., & Leite, M. de S. (2024). *DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models* (Version 0.4.7) [Computer software]. <https://cran.r-project.org/web/packages/DHARMA/index.html>
- Hawkins, J. A. (2019). Word-external properties in a typology of modern English: A comparison with German. *English Language and Linguistics*, 23(3), 701–727. <https://doi.org/10.1017/S1360674318000060>
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(s1), 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Kastovsky, D. (1985). Deverbal nouns in old and modern English: From stem-formation to word-formation. In J. Fisiak (Ed.), *Historical semantics—Historical word-formation* (pp. 221–262). De Gruyter Mouton. <https://doi.org/10.1515/9783110850178.221>
- Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99(2), 349–364. <https://doi.org/10.1037/0033-295X.99.2.349>
- Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (pp. 131–175). Dordrecht.
- Kisselew, M., Rimell, L., Palmer, A., & Padó, S. (2016). Predicting the direction of derivation in English conversion. In *Proceedings of the 14th SIGMORPHON workshop on computational research in phonetics, phonology, and morphology* (pp. 93–98). Berlin, Germany.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Langacker, R. W. (1987). Nouns and verbs. *Language*, 63(1), 53–94. <https://doi.org/10.2307/415384>
- Langacker, R. W. (1991). *Foundations of cognitive grammar: Volume II: Descriptive application*. Stanford University Press.
- Lieber, R. (2001). The semantics of verb formation. In R. Lieber (Ed.), *Morphology and lexical semantics* (pp. 76–97). Cambridge University Press.
- Lieber, R. (2005). English word-formation processes. In P. Štekauer & R. Lieber (Eds.), *Handbook of word-formation* (pp. 375–427). Springer Netherlands.
- Lievers, F. S., Bolognesi, M., & Winter, B. (2021). The linguistic dimensions of concrete and abstract concepts: Lexical category, morphological structure, countability, and etymology. *Cognitive Linguistics*, 32(4), 641–670. <https://doi.org/10.1515/cog-2021-0007>
- Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach* (2nd Revised & Enlarged ed.). C. H. Beck'sche Verlagsbuchhandlung.

- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 26). Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf)
- Mititelu, V. B., Leseva, S., & Stoyanova, I. (2023). Semantic analysis of verb – Noun zero derivation in Princeton WordNet. *Zeitschrift für Sprachwissenschaft*, 42(1), 181–207. <https://doi.org/10.1515/zfs-2022-2017>
- Neef, M. (2005). On some alleged constraints on conversion. In L. Bauer & S. Valera (Eds.), *Approaches to conversion/zero-derivation* (pp. 103–130). Waxmann.
- Neubauer, M. (2024). *English nouns since 1150: A typological study*. De Gruyter Mouton.
- Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., Botvinick, M., & Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, 9(1), 963. <https://doi.org/10.1038/s41467-018-03068-4>
- Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from Audiovisual input. *TESOL Quarterly*, 53(4), 1008–1032. <https://doi.org/10.1002/tesq.531>
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291. <https://doi.org/10.1016/j.cognition.2011.10.004>
- Pinker, S. (1994). *The language instinct*. Penguin. <https://www.penguin.co.uk/books/16256/the-language-instinct-by-steven-pinker/9780141980775>
- Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. De Gruyter Mouton.
- Plag, I. (2012). *Word-formation in English*. Cambridge University Press.
- R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLoS One*, 11(4), e0151138. <https://doi.org/10.1371/journal.pone.0151138>
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4(3), 328–350. [https://doi.org/10.1016/0010-0285\(73\)90017-0](https://doi.org/10.1016/0010-0285(73)90017-0)
- Roswell, M., Dushoff, J., & Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos*, 130(3), 321–338. <https://doi.org/10.1111/oik.07202>
- Schmid, H.-J. (2020). *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford University Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (p. xxi, 623). Houghton, Mifflin and Company.
- Shao, B., Zheng, J., & De Smet, H. (2024). The blurring of the boundaries: Changes in verb/noun heterosemy in recent English. *Corpus Linguistics and Linguistic Theory*, 20(2), 321–346. <https://doi.org/10.1515/cllt-2022-0053>
- Shi, Y., & Lei, L. (2022). Lexical richness and text length: An entropy-based perspective. *Journal of Quantitative Linguistics*, 29(1), 62–79. <https://doi.org/10.1080/09296174.2020.1766346>
- Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, 26(5), 858–866. <https://doi.org/10.1038/s41593-023-01304-9>
- Taylor, J. R. (1995). *Linguistic categorization*. Oxford University Press.
- Tribout, D. (2010). *Les conversions de nom à verbe et de verbe à nom en français* [Unpublished doctoral dissertation] Université Paris Diderot (Paris 7). <https://shs.hal.science/tel-01577528>
- Tribout, D. (2020). Nominalization, verbalization or both? Insights from the directionality of noun-verb conversion in French. *Zeitschrift Für Wortbildung/Journal of Word Formation*, 4(2), 187–207. <https://doi.org/10.3726/zwjw.2020.02.10>
- Trott, S. (2024). Can large language models help augment English psycholinguistic datasets? *Behavior Research Methods*, 56(6), 6082–6100. <https://doi.org/10.3758/s13428-024-02337-z>
- Villani, C., Loia, A., & Bolognesi, M. M. (2024). The semantic content of concrete, abstract, specific, and generic concepts. *Language and Cognition*, 16(4), 867–894. <https://doi.org/10.1017/langcog.2023.64>
- Wang, G., & Shao, B. (2025). Exploring the effect of semantic diversity on boundary permeability in verb/noun heterosemy using deep contextualized word embedding. *Linguistics Vanguard*. <https://doi.org/10.1515/lingvan-2025-0121>

- Xu, A., Kemp, C., Frermann, L., & Xu, Y. (2024). Word reuse and combination support efficient communication of emerging concepts. *Proceedings of the National Academy of Sciences*, 121(46), e2406971121. <https://doi.org/10.1073/pnas.2406971121>
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.

## Appendix 1. Noun-to-verb conversion pairs examined in this study

Accent, age, air, alarm, anchor, angle, ape, apprentice, arc, auction, awe, axe, bag, bail, bait, balloon, band, bandage, bank, barrel, barricade, bat, bath, beach, bead, beam, bed, beggar, belt, benefit, bias, bicycle, blanket, block, blockade, bloom, blossom, board, boat, bomb, bond, book, border, bottle, bowl, brace, bracket, brain, brake, branch, breakfast, breast, breeze, bridge, brim, brush, bubble, bud, budget, bug, bulk, bunch, bundle, burden, bus, butcher, butter, button, cable, cage, cake, campaign, cap, captain, capture, card, carpet, cart, case, cash, catalogue, caution, cement, chain, chair, chalk, channel, charm, chart, charter, chronicle, circle, clam, class, claw, climax, cloak, clock, cloud, clown, club, cluster, coach, coast, coat, cock, code, coin, comb, commission, corral, counsel, course, court, cradle, craft, crane, crank, credit, crest, crop, crowd, crown, crusade, cup, curb, curtain, cushion, cycle, dam, date, dawn, deck, decree, desert, detail, devil, dial, dice, diet, disguise, distance, ditch, document, dog, doom, dope, dose, dot, draft, dress, drug, drum, duel, dust, dwarf, ease, echo, eclipse, edge, effect, elbow, end, engineer, evidence, experiment, eye, face, factor, fan, farm, fashion, father, fault, feast, feather, fence, ferry, feud, fiddle, field, figure, file, film, finger, fire, fish, flag, flake, flame, flank, flock, flood, floor, flower, foam, focus, fog, foot, force, ford, fork, form, fracture, fragment, freight, fringe, frost, fuel, function, fund, fur, fuss, gag, gas, gauge, gear, gleam, gorge, gossip, grace, grade, grease, grimace, groom, groove, ground, group, guarantee, guard, halt, hammer, hand, handcuff, handicap, harm, harness, harvest, hazard, head, headline, heap, heat, hedge, heel, herald, herd, hinge, hint, hole, holiday, hook, Hoover, hose, host, hound, house, hunger, husband, ice, image, inch, inconvenience, index, influence, ink, interest, interview, iron, jail, jet, jockey, journey, junk, knife, knight, knot, label, lace, land, landscape, lapse, latch, layer, leaf, lecture, letter, lever, levy, limit, line, link, litter, load, loan, lobby, lock, lodge, loft, log, loop, loot, lunch, lure, machine, mail, mandate, manure, map, mark, market, marshal, mask, mass, massacre, match, mate, matter, merchandise, merit, mesh, mess, milk, mill, mind, mirror, mist, moan, mob, model, mop, mortar, mortgage, mother, motion, mouth, mug, mushroom, muzzle, nail, neck, needle, nerve, nest, nickname, nose, note, nurse, occasion, oil, orbit, ornament, outfit, outlaw, outline, pack, package, pad, paddle, page, pain, pair, palm, pan, panic, paper, parade, park, partition, partner, paste, patch, patrol, pattern, pause, paw, pawn, peg, pen, pencil, people, pepper, perfume, pet, petition, phase, phone, photograph, phrase, picket, pickle, picnic, picture, piece, pile, pillow, pin, pioneer, pipe, pit, pity, place, plague, plan, plant, plaster, plate, plot, plug, pocket, point, poison, pole, pool, position, post, posture, pot, powder, power, pressure, prey, price, print, profile, progress, promise, prop, pump, purse, quarrel, quarter, query, question, race, radio, rag, rage, raid, railroad, rat, rate, ration, rattle, rear, recess, referee, regiment, register, remedy, rendezvous, rhyme, riddle, rocket, roof, room, root, rope, rust, rut, sack, sacrifice, saddle, sanction, sand, sap, scalp, scar, scent, school, scout, scrap, screen, screw, sculpture, seal, seam, season, seat, seed, sense, sentence, service, shade, shame, share, shell, shelter, shepherd, shield, ship, shoe, shop, shore, shoulder, shovel, shower, side, sight, sign, silence, silhouette, silver, size, skate, sketch, skin, skirt, slate, slave, slope, slot, smoke, snake, soap, soil, soldier, sort, space, span, spark, spear, spice, spike, spirit, splinter, sponge, sponsor, spoon, sport, spur, stage, stake, stall, star, station, steam, steel, stem, stick, stone, storm, strap, stream, stress, string, stub, stump, style, suit, summer, summons, sun, supplement, surf, surge, swamp, swarm, sweat, switch, table, tackle, tag, tail, tailor, tape, tar, tax, telegraph, telephone, telescope, term, test, thread, throttle, thumb, thunder, tile, time, tone, tool, torment, torpedo, torture, tour, towel, tower, track, trade, trap, treasure, trigger, triumph, troop, trumpet, trust, tug, tune, tunnel, tutor, vacation, vacuum, value, veil, vest, view, voice, vote, vow, voyage, wad, wage, wall, waltz, water, wax, weather, wedge, weed, weight, wheel, whip, whistle, wing, winter, wire, witness, wolf, wonder, word, worm, worship, wound, wreck, wrench, wrinkle, yell.

## Appendix 2. Model diagnostics and robustness checks

### Overdispersion testing

Both initial binomial GLMMs were tested for overdispersion using the Pearson's statistic, computed as the sum of squared conditional Pearson's residuals divided by the residual degrees of freedom. Conditional

residuals were extracted after random intercepts were accounted for, so the test reflects excess variance beyond what the random-effect structure already captures. The dispersion ratio was 13.54 for the baseline model and 11.16 for the comprehensive model (both  $p < .001$ ), confirming overdispersion in both cases.

### Beta-binomial specification

We refitted both models using a beta-binomial distribution via *glmmTMB* (Brooks et al., 2017), retaining identical model formulae and random-effect structures. The beta-binomial directly models extra-binomial variance through an explicit dispersion parameter ( $\phi = 66.6$  for the baseline model;  $\phi = 77.4$  for the comprehensive model). This was preferred over observation level random effects, which failed to converge in preliminary testing.

### Model adequacy

Simulation-based diagnostics from *DHARMa* (Hartig et al., 2024) were used to confirm fit. Bootstrap-based outlier tests, which *DHARMa* recommends as more reliable than standard dispersion tests for integer-valued grouped binomial data, returned nonsignificant results for both beta-binomial models (baseline:  $p = .14$ ; comprehensive:  $p = .36$ ).

### Robustness across distributions

A comparison of the original binomial and beta-binomial models confirmed that effect directions and significance patterns were preserved across both specifications, with the beta-binomial producing appropriately more conservative standard errors. All results reported in Section 4 are based on the beta-binomial specifications.

### Software versions

All models were fitted in *R* version 4.5.1 (R Core Team, 2025) using *lme4* version 1.1–37 (Bates et al., 2003) for binomial GLMMs and *glmmTMB* (Brooks et al., 2017) for beta-binomial GLMMs.

---

**Cite this article:** Wang, G., & Shao, B. (2026). Less is more: Diachronic evidence for efficient communication in English noun-to-verb conversion over the past century, *Language and Cognition*, 18, e55, 1–23. <https://doi.org/10.1017/langcog.2026.10102>