



Cognitive Science 50 (2026) e70202  
© 2026 Cognitive Science Society LLC.  
ISSN: 1551-6709 online  
DOI: 10.1111/cogs.70202

# Efficient Communication in Word Formation: How Syntactic and Lexical Surprisal Jointly Shape English Conversion Over the Past Century

Gui Wang,  Mengyang Yu,  Bin Shao 

*School of International Studies, Zhejiang University*

Received 5 January 2026; received in revised form 11 March 2026; accepted 13 March 2026

---

## Abstract

Conversion degree, defined as the proportion of converted-form use out of total lemma use, has notably increased in English over the past century. For instance, *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). This historical trend presents a communicative paradox: it favors speaker economy (articulatory ease via form reuse) seemingly at the expense of listener economy (increased ambiguity). The present study investigates this paradox through the framework of efficient communication, which predicts that ambiguous forms like conversion should be licensed in predictable contexts. We tested this prediction by analyzing 639 noun-to-verb and 443 verb-to-noun conversion words from the 1920s to the 2010s. Generalized linear mixed models were fitted with surprisal estimates from three architectures: local 5-grams, syntactic POS-grams, and broad-context GPT-2. Results indicate that the historical rise in conversion degree is accompanied by increasingly predictable contexts, suggesting that contextual predictability compensates for the absence of explicit morphological marking. We further demonstrate that syntactic and lexical surprisal exert independent inhibitory effects on conversion degree. Crucially, their positive interaction indicates that the constraining effect of lexical surprisal is stronger when syntactic surprisal is low and attenuated when syntactic surprisal is high, consistent with a resource competition account. These findings portray the evolution of English conversion not as random drift, but as a dynamic optimization reflecting a negotiation between communicative efficiency and cognitive resource limitations, offering a novel information-theoretic perspective on language change.

**Keywords:** Conversion; Efficient communication; Language change; Surprisal; Word formation

---

## 1. Introduction

Conversion is a word-formation process that creates new lexical items by altering word class without overt morphological marking (Plag, 2003). A familiar example is the transformation of the proper noun *Google* into the verb *to google*. The verb has become so entrenched that few speakers pause to consider its nominal origin. Conversion is a productive process in English and has been part of English grammar for centuries (Pinker, 1994: 379). Recent corpus-based morphological work confirms its central role. Based on 16,624 morphosemantically related noun–verb pairs in WordNet, Mititelu, Leseva, and Stoyanova (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. They argue that conversion is not merely frequent but structurally and semantically pervasive.

Recent evidence indicates that conversion has intensified in English over time—not in the sense that more words have acquired conversion status, but in the sense that established conversion words are used more fluidly across lexical categories over time (Shao, Zheng, & De Smet, 2024; Wang & Shao, 2025b). For instance, Shao et al. (2024) investigated 877 noun/verb conversion words in English and observed that their categorical boundaries have become increasingly permeable over the past century. Complementing this, Wang and Shao (2025b) provided a more granular analysis of noun-to-verb conversion by examining 637 items from the 1920s to the 2010s. They identified a significant increase in conversion degree across the century, defined as the proportion of tokens occurring in the converted category out of a word’s total usage across both categorical uses (e.g., for noun-to-verb conversion, the verb token count divided by the combined noun and verb token count). 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 verbal reuse of the nominal base form. Collectively, these findings suggest that English is undergoing a systematic shift toward word reuse, evolving toward a lexicon characterized by greater categorical flexibility.

### 1.1. *The communicative paradox*

The historical rise in conversion degree presents a communicative paradox. From the listener’s perspective, conversion is costly: by reusing the same form across lexical categories, it increases formal ambiguity and places a greater processing burden on the listener, who must infer the intended lexical category from context alone. Under such conditions, one would expect communicative pressure to favor overtly marked alternatives. Yet, the opposite has occurred: conversion degree has not declined but instead increased over the past century.

This paradox is further sharpened by Zipf’s (1949) classic opposition between the speaker’s economy and the auditor’s economy. For speakers, conversion represents the epitome of efficiency: it allows for the recycling of existing forms without the effort of adding affixes (e.g., *to email* vs. *to \*emailize*). This aligns with Zipf’s “Force of Unification,” where speakers prefer a small, versatile vocabulary to minimize memory and articulatory effort. Conversely, for listeners, this strategy introduces ambiguity. According to Zipf’s “Force of

Diversification,” listeners prefer a one-to-one correspondence between form and meaning to minimize decoding effort. An unchecked increase in conversion, using the same form for multiple lexical categories, potentially creates a “noisy” channel, heightening the cognitive burden on the listener to decipher the intended syntactic role. How, then, can English sustain this expansion without compromising communicative success?

### *1.2. Efficient communication and contextual predictability*

The communicative paradox might be reconciled through the framework of efficient communication, which holds that linguistic structures reflect optimization pressures balancing the competing demands of the speaker and the listener (Gibson et al., 2019; Kemp, Xu, & Regier, 2018; Piantadosi, Tily, & Gibson, 2011; Regier, Kemp, & Kay, 2015). These pressures have been formalized in various ways, for instance, as a trade-off between simplicity and informativeness (Kemp et al., 2018; Regier et al., 2015; Y. Xu, Regier, & Malt, 2016), or through the information bottleneck principle, which formalizes the trade-off between compression of meaning representations and communicative accuracy (Zaslavsky, Kemp, Regier, & Tishby, 2018). Despite these distinct formalizations, the approaches converge on the prediction that attested linguistic forms tend toward efficient solutions to the tension between speaker and listener demands.

This framework has successfully explained cross-linguistic variations in word meaning (Kemp et al., 2018; Regier et al., 2015; Zaslavsky et al., 2018), the structures of word forms (Hahn, Degen, & Futrell, 2021; Hahn, Mathew, & Degen, 2021; Mahowald, Dautriche, Gibson, & Piantadosi, 2018), and grammatical form-meaning mappings (Mollica et al., 2021). Most relevantly, A. Xu, Kemp, Frermann, and Xu (2024) examined word reuse and compounding as two common lexicalization strategies and found that historically attested items achieve higher levels of communicative efficiency than hypothetical alternatives. Their findings suggest that word reuse, including conversion, represents an efficient strategy at the lexical level: it minimizes speaker effort by recycling existing forms while maintaining sufficient informativeness for the listener.

While A. Xu et al. (2024) demonstrated that word reuse achieves efficient trade-offs across the lexicon, a complementary question remains at the utterance level: under what contextual conditions is a particular instance of reuse licensed in actual language use? To answer this question, we turn to the uniform information density (UID) hypothesis (Jaeger, 2010; Levy & Jaeger, 2007), which makes predictions about information distribution at the level of individual utterances. The UID hypothesis proposes that speakers structure their utterances to keep information density, quantified as surprisal (i.e., the negative log probability of a linguistic unit given its context), relatively constant across the utterance. Surprisal is directly linked to cognitive effort in comprehension: highly surprising words incur greater processing costs, as evidenced by longer reading times (Hale, 2001; Levy, 2008; Smith & Levy, 2013).

While UID is often operationalized as maximizing the uniformity of information distribution across an entire utterance (e.g., minimizing surprisal variance; Meister et al., 2021), a fundamental corollary of this framework is the pressure to avoid local spikes in information density that might exceed the channel capacity (Levy & Jaeger, 2007). Applying this

framework to conversion yields a specific prediction. As a short, morphologically unmarked form, conversion lacks overt category cues and relies entirely on context for disambiguation. Consequently, in contexts where such cues are weak, the probability mass is distributed across competing category interpretations, thereby increasing surprisal and raising the risk of a local spike in information density. To keep processing demands within acceptable bounds, the surrounding context must, therefore, provide strong “scaffolding” to compensate for the lack of overt marking. Thus, consistent with the pressure to smooth information density, we predict that conversion should be licensed specifically in low-surprisal contexts, that is, contexts with high contextual predictability.

### 1.3. *The present study*

Building on the UID hypothesis, the present study proposes that the historical rise in conversion degree is not random but constrained by contextual predictability. We test this through a hierarchical set of hypotheses.

First, at the global level, we test the universal constraint hypothesis (H1). We posit that a high conversion degree is associated with low contextual surprisal across both conversion directions and across decades. Diachronically, this implies that the observed increase in conversion degree over the past century has been enabled by highly predictive contexts. This constraint should hold across both conversion directions: regardless of whether the shift is noun to verb or verb to noun, the omission of overt marking requires the compensation of strong contextual predictability.

Second, at a fine-grained level, we distinguish between syntactic surprisal (structure) and lexical surprisal (content). While general measures of contextual predictability capture overall expectations, they conflate structural constraints with semantic probabilities. Roark, Bachrach, Cardenas, and Pallier (2009) demonstrate that the predictive load of a word can be formally decomposed into two distinct components: the probability of the syntactic category appearing at a given position, and the probability of the specific lexical item given that category. van Schijndel and Schuler (2015) also point out the necessity of making a distinction between surprisal types to determine their individual effects. In addition, this distinction is vital for analyzing conversion, which involves a tension between a word’s typical lexical category and its contextually assigned syntactic category. Therefore, we hypothesize (H2a) that both lexical and syntactic surprisal independently constrain conversion. For instance, a converted verb should be less likely to appear in a position with high syntactic surprisal (where a verb is structurally unexpected) or high lexical surprisal (where the specific word meaning is unexpected).

Furthermore, we examine the interplay between these constraints via the modulation hypothesis (H2b). Beyond their independent effects, recent evidence suggests these two types of surprisal also interact. Lexical surprisal has been shown to modulate the time course of syntactic structure building (Slaats, Meyer, & Martin, 2024), and a syntax-lexicon trade-off has been observed in language production (Rezaii, Mahowald, Ryskin, Dickerson, & Gibson, 2022). However, the precise mechanism underlying this interaction remains underexplored, particularly in the context of word formation. We propose two competing possibilities.

The first possibility, which we term the resource competition account, posits that syntactic and lexical processing compete for limited cognitive control resources. As Rezaii et al. (2022: 8) noted, the planning of an utterance might require shared cognitive control resources to be allocated both to lexical selection and to structural selection. This view aligns with Bock's (1982) foundational framework, which explicitly models sentence production within a limited-capacity processing system where syntactic formulation is inherently sensitive to the transient processing demands of lexical retrieval. Under this framework, we hypothesize that when syntactic surprisal is low, abundant cognitive resources allow strict monitoring of lexical selection, resulting in high sensitivity to lexical surprisal. Conversely, when syntactic surprisal is high, structural processing consumes the shared resources, attenuating the effect of lexical surprisal.

Alternatively, under what we term the UID-based compensatory account, the interaction may be driven by a bottleneck mechanism framed through the UID hypothesis. This perspective views the language system as operating within a fixed channel capacity, necessitating a trade-off between structural and lexical information to prevent overload. Under this framework, we hypothesize that when syntactic surprisal is low, the communication channel has ample remaining capacity, and the constraint on lexical selection is relaxed, predicting a flatter slope (lower sensitivity). Conversely, when syntactic surprisal is high, the system must compensate by minimizing lexical surprisal to keep total information density within the bottleneck, predicting a steeper slope (heightened sensitivity).

By testing these hypotheses, this study aims to elucidate how English has increased its lexical flexibility through conversion without sacrificing communicative robustness. In doing so, we provide an information-theoretic perspective on language evolution and explore the interplay between syntactic and lexical constraints in language processing. More broadly, this study extends efficient communication research from static cross-linguistic typology to language change, demonstrating that word formation is systematically shaped by communicative pressures.

## 2. Methods

### 2.1. Data

#### 2.1.1. Corpus

We used the 2021 release of the Corpus of Historical American English (COHA; Davies, 2010) to investigate the effects of contextual predictability on conversion degree over the past century. Comprising over 279 million words from the 1920s to the 2010s, COHA provides the necessary temporal depth and scale for robust diachronic analysis. As Biber and Gray (2013: 108) note, such large-scale corpora permit analysis of historical change “on a scale previously unachievable with earlier corpus designs,” enabling systematic monitoring of language changes over time. In addition, COHA employs a balanced sampling framework across four major genres: fiction, popular magazines, newspapers, and nonfiction books (Davies, 2012). This systematic composition, maintained consistently across decades, ensures that observed shifts in conversion degree reflect genuine linguistic development rather than

Table 1  
Composition of COHA of the past century by decade

| Decade | Tokens     |
|--------|------------|
| 1920s  | 23,819,149 |
| 1930s  | 25,710,170 |
| 1940s  | 25,451,525 |
| 1950s  | 26,600,443 |
| 1960s  | 27,023,893 |
| 1970s  | 26,728,218 |
| 1980s  | 27,605,215 |
| 1990s  | 31,034,963 |
| 2000s  | 32,654,928 |
| 2010s  | 32,648,551 |

artifacts of changing genre distributions. Table 1 presents the token counts for each decade in the present study.

2.1.2. Conversion word selection

We identified candidate conversion words through a systematic filtering process combining morphological annotations from the CELEX lexical database (Baayen et al., 1995) with frequency data from COHA. Following Kisselew, Rimell, Palmer, and Padó (2016), we extracted candidates via the WebCelex interface (<https://webcelex.ivdnt.org/>) based on zero-derivation morphology. For noun-to-verb conversion, we queried for lemma entries classified as verbs (*Class eq 'V'*) with zero-derivation morphological status (*MorphStatus eq 'Z'*). This initial step yielded 1719 candidate verbs derived from conversion. We then filtered this list to retain only lexemes that also possessed a corresponding noun form (*Class eq 'N'*), resulting in 1640 words. For verb-to-noun conversion, we applied a symmetrical procedure: querying for noun entries (*Class eq 'N'*) with zero-derivation status (*MorphStatus eq 'Z'*) and verifying the existence of a corresponding verb form (*Class eq 'V'*). To control for directionality and exclude ambiguous cases (e.g., *clean*, where the verb derives from an adjective), we restricted the dataset to lexemes associated exclusively with the noun and verb word classes in CELEX.

To ensure sufficient data density for statistical modeling, we applied frequency thresholds based on the usage patterns in COHA. We retained items meeting two criteria: (1) the base form (noun for noun-to-verb; verb for verb-to-noun) appeared at least 50 times per decade (at least 500 tokens in total), ensuring consistent and stable use across the entire century; and (2) the converted form (verb for noun-to-verb; noun for verb-to-noun) appeared at least 100 times in total and occurred in at least 8 of the 10 decades. The asymmetry between these thresholds is intentional, as the converted form tends to be less frequent than the base form. Applying an equally strict per-decade threshold would, therefore, eliminate most legitimate conversion candidates and introduce selection bias toward only the most extreme cases. The threshold for the converted form instead prioritizes sufficient temporal coverage for reliable diachronic modeling. These procedures yielded a final dataset of 639 noun-to-verb and 443 verb-to-noun conversion words (see Supplementary Appendix A1 for the complete list).

### 2.1.3. Preprocessing

To ensure data quality and valid estimates of contextual predictability, we performed two preprocessing steps. First, we removed all sentences containing redaction characters (i.e., “@”). In the offline version of COHA, this symbol is inserted every 200 tokens to replace 10 consecutive words for copyright protection (Davies, 2012). Because statistical language models rely on continuous context to estimate next-word probability, target words appearing in sequences with redacted context would yield distorted surprisal estimates. Consequently, we retained only complete, unredacted sentences for analysis.

Second, to facilitate the calculation of syntactic surprisal, we mapped the original part-of-speech annotation (i.e., CLAWS7 tags; Leech & Smith, 2000) to the Universal Dependencies tagset. Fine-grained CLAWS7 distinctions can introduce noise when estimating the probability of core grammatical categories. For example, treating singular and plural nouns as distinct labels reduces the reliability of probability estimates for the general noun category. To ensure that our metric captures the constraints of the syntactic slot rather than morphological variation, we implemented a mapping scheme that abstracts away from such distinctions while preserving fundamental category boundaries:

- Verbs versus Auxiliaries: Tags starting with VV were mapped to VERB (lexical verbs), while tags indicating *be*, *do*, *have*, and modals (VB\*, VD\*, VH\*, VM\*) were mapped to AUX.
- Nouns versus Proper Nouns: Tags starting with NP were mapped to PROPN; other noun tags (NN\*) were mapped to NOUN.
- Determiners: Articles and determiners (AT\*, D\*) were mapped to DET.
- Adjectives and Adverbs: Mapped to ADJ and ADV, respectively.

This mapping ensures that our syntactic surprisal estimates reflect the predictability of the structural slot itself (e.g., the likelihood of any noun appearing after a determiner) rather than being influenced by morphological variants.

## 2.2. Variable operationalization

### 2.2.1. Dependent variable

The dependent variable was the conversion degree. It was operationalized at the token level as a binary conversion status, encoding whether a lemma occurred in its converted or base category in a given context. For noun-to-verb conversion, a “success” (in the binomial sense) was coded when the lemma functioned as a verb, and a “failure” when it functioned as a noun. For verb-to-noun conversion, these definitions were reversed. Under this framework, a higher conversion degree indicates a greater likelihood of the lemma appearing in the target category (e.g., a noun surfacing as a verb), whereas a lower degree reflects a tendency to retain the base category. Unlike ratio-based metrics that aggregate counts prior to analysis, this operationalization preserves information about occurrence frequency. When counts are collapsed into ratios beforehand, a lemma appearing 10 times and one appearing 1000 times contribute equally to the analysis if their proportions are identical, despite the higher-frequency lemma providing far more reliable estimates. By preserving the raw frequency counts, this

approach allows the statistical model to naturally weight lemmas according to their corpus frequency, ensuring that more frequently attested items exert proportionally greater influence on the parameter estimates.

To provide a descriptive baseline, we also derived lemma-level conversion degree estimates by computing the proportion of tokens in the converted category for each lemma. These descriptive summaries indicate considerable variation across lemmas: for noun-to-verb conversion, the mean conversion degree was 0.192 ( $SD = 0.176$  across lemma-level means); for verb-to-noun conversion, the mean was 0.284 ( $SD = 0.210$ ). These figures reflect lemma-level averages across the 1920s–2010s and are reported here for descriptive purposes; all inferential modeling was performed on the binary token-level data.

### 2.2.2. Independent variables

The independent variables operationalized contextual predictability, the core construct in our hypotheses. To test the universal constraint hypothesis (H1), we computed lexical surprisal from two complementary sources: 5-gram models (capturing local co-occurrence) and GPT-2 (capturing broad contextual dependencies). To test the independent constraints (H2a) and modulation hypotheses (H2b), we further distinguished syntactic surprisal (structural slot expectation) from pure lexical surprisal (lexical unpredictability after partialing out syntactic influence).

Surprisal quantifies the unexpectedness of a linguistic unit given its preceding context. For a language model defining a probability distribution over the vocabulary, the surprisal  $S$  of a target unit  $w_t$  given its context  $w_1 \dots w_{t-1}$  is defined as its negative log probability (Hale, 2001; Levy, 2008):

$$S(w_t) = -\log_2 P(w_t | w_1 \dots w_{t-1}).$$

We employed three modeling architectures, each designed to isolate a distinct source of predictive information: 5-gram models for immediate lexical co-occurrence, GPT-2 for broader contextual dependencies, and POS-gram models for abstract grammatical structure. This approach allows us to examine how different dimensions of predictability constrain conversion.

To quantify constraints imposed by immediate lexical context, we computed surprisal estimates using 5-gram language models. Previous studies demonstrate that count-based n-gram models can serve as useful predictors of human processing difficulty, particularly for local dependencies and fixed expressions (e.g., Shain et al., 2020, 2024; Smith & Levy, 2013). Following Shain (2019), we trained decade-specific models on the full COHA corpus using the KenLM toolkit (Heafield, 2011). Given data sparsity in historical subcorpora, we employed modified Kneser–Ney smoothing (Ney, Essen, & Kneser, 1994) with discount fallback enabled to ensure robust probability estimates for lower-order sequences. These estimates thus reflect local lexical probability specific to each decade, conditioned on the preceding four-word window.

To capture long-distance dependencies and semantic constraints beyond the n-gram window, we computed surprisal using the pretrained GPT-2 Small architecture (117M parameters; Radford et al., 2019). Recent work has demonstrated strong alignment between

autoregressive transformer representations and measures of human sentence processing (e.g., Hao, Mendelsohn, Sterneck, Martinez, & Frank, 2020; Schrimpf et al., 2021; Shain, Meister, Pimentel, Cotterell, & Levy, 2024). Furthermore, our selection of the Small architecture follows Shain et al. (2024), who reported that GPT-2 Small outperforms larger models (e.g., GPT-3) in predicting human subjective surprisal. They attribute this result to data contamination in larger models, where training on test sets artificially inflates probability estimates for surprising words. Because GPT-2 uses Byte-Pair Encoding, which segments words into subword units, we derived word-level estimates by summing the surprisals of constituent tokens. Although GPT-2 is trained on modern text (WebText) and does not capture diachronic shifts, this limitation is partially mitigated by its capacity to model stable, long-range dependencies. As a modern baseline for broad contextual predictability, GPT-2 complements our decade-specific *n*-gram models, allowing us to jointly examine dynamic local norms and stable global dependencies.

To isolate the predictability of the grammatical slot independent of the specific lexical item, we estimated syntactic surprisal using decade-specific 5-gram POS models (hereafter POS-gram models). These models were trained on the mapped universal dependencies sequences described in Section 2.1.3. As Frank and Bod (2011: 4) noted, POS-gram models abstract away from lexical content, ensuring that the resulting surprisal values reflect structural constraints alone. A critical methodological step was the application of a forced replacement strategy: for each target word instance, we replaced its original POS tag with the target conversion category (verb for noun-to-verb conversion; noun for verb-to-noun conversion) before computing syntactic probability. While the probability is computed for the abstract grammatical category, we assign the resulting surprisal value to the specific lemma occupying that slot. This procedure yields an instance-specific measure of syntactic predictability for each token. By decoupling structural expectation from lexical identity, this metric captures the syntactic constraints imposed by the grammatical context.

To quantify the relationship between these three predictive proxies, we calculated Pearson correlation coefficients using decade-standardized values (consistent with the standardization procedure detailed in Section 2.3.1). Local 5-gram surprisal and broad-context GPT-2 surprisal exhibited moderate positive correlations ( $r = .20$  for noun-to-verb;  $r = .43$  for verb-to-noun; both  $p < .001$ ), indicating that they capture related but distinct sources of lexical probability. In contrast, syntactic (POS-gram) surprisal exhibited consistently weak correlations with both lexical measures across datasets ( $|r| \leq .16$ ). This pattern of weak correlations supports treating syntactic and lexical surprisal as empirically distinct predictors.

### 2.2.3. *Control variables*

To isolate the effects of contextual predictability on conversion degree, we included three control variables capturing distinct psycholinguistic constraints. First, we controlled for baseline lexical accessibility using unigram surprisal (the negative log probability of the base form). This metric serves as a proxy for lexical entrenchment and blocking effects, as highly entrenched base forms tend to be more resistant to conversion (Bybee, 2007, 2010). Second, we included word length, measured as the number of orthographic characters. This variable

controls for Zipfian economy effects (Zipf, 1949): shorter words tend to be more polysemous and productive, making them more susceptible to conversion. Third, we included sentence position, measured as the log-transformed ordinal index of the target word within the sentence. This control accounts for informational accumulation: words appearing later in a sentence are naturally more predictable due to the buildup of prior constraints (Smith & Levy, 2013). Controlling for sentence position ensures that our surprisal effects reflect contextual predictability specific to conversion rather than the general increase in predictability associated with sentence position.

### 2.3. Statistical analysis

#### 2.3.1. Standardization and orthogonalization

All continuous predictors were standardized to account for diachronic shifts in baseline distributions. Historical corpora often exhibit significant drifts, such as vocabulary expansion, which can inflate raw surprisal values in later decades. Furthermore, because GPT-2 is trained on contemporary internet text, it may assign inflated surprisal scores to valid historical expressions due to diachronic distance rather than genuine processing difficulty.

We applied decade-wise standardization to surprisal estimates (5-gram, GPT-2, and POS-gram), unigram frequency, and sentence position. Transforming these variables into *z*-scores within each decade ensures that values represent relative cognitive prominence within the contemporaneous linguistic environment rather than absolute historical shifts. Because word length is a static orthographic property, we standardized it globally across the entire dataset.

To test whether lexical and syntactic constraints operate independently or interact, we disentangled their contributions through residualization. Within each decade, we regressed 5-gram surprisal (standardized) on POS-gram surprisal (standardized) and extracted the residuals. These residuals constitute our measure of pure lexical surprisal, representing lexical unpredictability after accounting for structural constraints. Performing this operation separately by decade respects the dynamic relationship between structure and lexicon, reflecting the assumption that linguistic expectations are calibrated to period-specific norms. By allowing the regression slope to vary across decades, this metric captures the degree to which a word is unexpected given the structural-lexical dependencies characteristic of that period.

To validate this approach against potential statistical artifacts, we conducted two robustness checks (see Supplementary Appendix A2). First, we compared our main model against a pooled residualization model applying a fixed regression slope across the entire century. Second, we tested an alternative model using globally standardized frequency to ensure that decade-wise standardization did not obscure the effects of absolute lexical entrenchment. Both checks yielded results qualitatively identical to the main analysis, confirming that our findings reflect genuine linguistic patterns rather than methodological artifacts.

#### 2.3.2. Model specification

To test our hypotheses, we fitted generalized linear mixed models (GLMMs) with a binomial link function using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R (version 4.5.2; R Core Team, 2025). These models estimate the log-odds of a lemma

occurring in the converted category (a noun functioning as a verb for noun-to-verb conversion; a verb functioning as a noun for verb-to-noun conversion) versus retaining its base form. Given the longitudinal structure of our dataset, where the same lemmas are observed across 10 decades, the data violate the independence assumption of standard regression. To address this, we included random intercepts for each lemma, accounting for within-lemma dependencies and baseline differences in conversion propensity.

To evaluate the universal constraint hypothesis (H1), we constructed a baseline model testing whether contextual predictability facilitates conversion. We implemented a  $2 \times 2$  factorial design, fitting the model separately to both datasets (noun-to-verb and verb-to-noun) using two surprisal measures (5-gram for local context; GPT-2 for broad context). The model estimated the log-odds of conversion for lemma  $j$  in decade  $t$  as a function of raw lexical surprisal ( $S_{raw}$ ), controlling for time ( $T$ ) and a vector of control covariates  $C$  (unigram frequency, word length, and sentence position):

$$\text{logit}(p_{jt}) = \beta_0 + \beta_1 S_{raw, jt} + \beta_2 T_t + \gamma C_{jt} + u_j.$$

Here,  $S_{raw, jt}$  represents standardized surprisal (5-gram or GPT-2), and  $u_j$  represents the random intercept for lemma  $j$ . The coefficient of interest is  $\beta_1$ . Because surprisal is inversely related to predictability, a significant negative coefficient would be consistent with H1: as predictability decreases (surprisal increases), the probability of conversion decreases.

To examine the distinct contributions of syntactic and lexical constraints (H2a) and their potential interaction (H2b), we extended the baseline model by replacing raw surprisal with the residualized predictors: syntactic surprisal ( $S_{syn, jt}$ ) and pure lexical surprisal ( $S_{lex, jt}$ ), along with their interaction:

$$\text{logit}(p_{jt}) = \beta_0 + \beta_1 S_{syn, jt} + \beta_2 S_{lex, jt} + \beta_3 (S_{syn, jt} \times S_{lex, jt}) + \beta_4 T_t + \gamma C_{jt} + u_j.$$

Here,  $S_{syn, jt}$  represents standardized POS-gram surprisal (structural slot expectation), and  $S_{lex, jt}$  represents residualized lexical surprisal (lexical unpredictability after partialing out syntactic influence). The main effects ( $\beta_1$  and  $\beta_2$ ) test H2a: a significant negative  $\beta_1$  would be consistent with the interpretation that structurally unexpected slots suppress conversion regardless of lexical content, while a significant negative  $\beta_2$  would be consistent with the interpretation that lexical unpredictability suppresses conversion independent of syntactic constraints. The interaction coefficient ( $\beta_3$ ) tests H2b: a significant interaction would suggest that the effect of lexical predictability varies with the syntactic environment. The direction of this interaction distinguishes between two competing mechanisms: if high syntactic surprisal attenuates lexical sensitivity, this would be consistent with the resource competition account, whereby structural processing depletes shared cognitive resources; if high syntactic surprisal amplifies lexical sensitivity, this would be consistent with the UID-based compensatory account, whereby the system compensates for high structural load by constraining lexical selection.

Prior to model fitting, we assessed multicollinearity using variance inflation factors (VIFs). As expected from the orthogonalization procedure, VIF values for all predictors, including interaction terms, were approximately 1.0, and all control variables fell

well below the conventional threshold of 5.0 (Shadish, Cook, & Campbell, 2002). These results confirm that multicollinearity was not a concern for the interpretation of individual predictors.

### 3. Results

#### 3.1. *The universal constraint of contextual predictability*

Our first analysis evaluated the universal constraint hypothesis (H1), which posits that a higher conversion degree is associated with lower contextual surprisal (and thus higher contextual predictability). We fitted the baseline model across four configurations: two conversion directions (noun-to-verb and verb-to-noun) crossed with two surprisal measures (5-gram for local context and GPT-2 for broad context). Table 2 summarizes the fixed effects across all four model iterations.

Consistent with H1, the coefficients for contextual surprisal were negative and statistically significant across all model configurations (all  $p < .001$ ). Regardless of whether predictive information derives from the immediate local window (5-gram) or broad context (GPT-2), and regardless of conversion direction, words in highly surprising contexts are consistently less likely to occur in the converted category. Based on the GPT-2 estimates, a one-standard-deviation increase in contextual surprisal is associated with a decrease in conversion log-odds of 0.175 for noun-to-verb and 0.196 for verb-to-noun conversion. These results indicate that contextual surprisal serves as a universal constraint on conversion. In diachronic terms, the historical rise in conversion degree is accompanied by a concurrent increase in contextual predictability, suggesting that the linguistic environment has evolved to provide stronger scaffolding for grammatical flexibility.

Notably, the 5-gram model, constrained to the immediately preceding four words, already yielded robust inhibitory effects ( $\beta = -0.122$  and  $-0.144$ ). The GPT-2 model, capable of leveraging broader context and richer representations acquired from large-scale pretraining, produced stronger coefficients ( $\beta = -0.175$  and  $-0.196$ ), representing approximately 36.1–43.4% increases in effect magnitude. This pattern suggests that while local co-occurrence patterns impose a substantial constraint on conversion, access to broader contextual dependencies and richer semantic representations is associated with stronger inhibitory effects.

Regarding control variables, unigram surprisal exhibited divergent patterns across the two conversion directions. For noun-to-verb conversion, the positive coefficients ( $\beta_{ngram} = 0.233$ ,  $\beta_{gpt} = 0.256$ ) indicate that lower-frequency nouns (corresponding to higher unigram surprisal) are more likely to undergo verbalization. For verb-to-noun conversion, the coefficients were negative ( $\beta_{ngram} = -0.259$ ,  $\beta_{gpt} = -0.271$ ), indicating that higher-frequency verbs are more likely to undergo nominalization. These opposing directions suggest that frequency exerts category-specific effects on conversion degree.

Sentence position showed a consistent negative effect across all models ( $\beta = -0.084$  to  $-0.147$ , all  $p < .001$ ). This pattern suggests that conversion is sensitive to incremental processing dynamics: as a sentence unfolds, prior context progressively narrows the set of plausible continuations. In later positions, these stronger expectations favor standard, prototypical

Table 2  
Summary of mixed-effects logistic regression models evaluating the universal constraint hypothesis

| Predictor            | Noun-to-verb conversion       |                               | Verb-to-noun conversion       |                               |
|----------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
|                      | 5-gram                        | GPT-2                         | 5-gram                        | GPT-2                         |
| Intercept            | −1.982***<br>[−2.089, −1.876] | −1.984***<br>[−2.085, −1.883] | −1.266***<br>[−1.407, −1.124] | −1.267***<br>[−1.408, −1.125] |
| Contextual surprisal | −0.122***<br>[−0.130, −0.114] | −0.175***<br>[−0.186, −0.164] | −0.144***<br>[−0.149, −0.138] | −0.196***<br>[−0.206, −0.187] |
| Unigram surprisal    | 0.233***<br>[0.221, 0.245]    | 0.256***<br>[0.244, 0.268]    | −0.259***<br>[−0.275, −0.243] | −0.271***<br>[−0.287, −0.255] |
| Word length          | 0.014<br>[−0.092, 0.121]      | 0.015<br>[−0.086, 0.116]      | 0.319***<br>[0.178, 0.460]    | 0.326***<br>[0.185, 0.468]    |
| Sentence position    | −0.147***<br>[−0.154, −0.140] | −0.084***<br>[−0.088, −0.079] | −0.108***<br>[−0.113, −0.103] | −0.096***<br>[−0.101, −0.091] |
| Time (decade)        | 0.095***<br>[0.093, 0.097]    | 0.098***<br>[0.096, 0.100]    | −0.004***<br>[−0.006, −0.002] | −0.001<br>[−0.003, 0.001]     |

*Note.* Values represent estimated fixed-effect coefficients (log-odds,  $\beta$ ) from the GLMM analysis, with 95% confidence intervals reported in brackets. Confidence intervals were computed using the Wald approximation. Significance levels: \*\*\* $p < .001$ . All continuous predictors were standardized using z-score transformation prior to model fitting. Unigram surprisal is the negative log probability of the word form; higher values correspond to lower frequency.

Table 3  
Summary of mixed-effects logistic regression models disentangling syntactic and lexical mechanisms

| Predictor                          | Noun-to-verb conversion<br>Coeff. | Verb-to-noun conversion<br>Coeff. |
|------------------------------------|-----------------------------------|-----------------------------------|
| <b>Intercept</b>                   | −1.985*** [−2.089, −1.881]        | −1.266*** [−1.407, −1.125]        |
| <b>Mechanism decomposition</b>     |                                   |                                   |
| Syntactic surprisal                | −0.171*** [−0.177, −0.165]        | −0.081*** [−0.087, −0.075]        |
| Lexical surprisal (residual)       | −0.132*** [−0.140, −0.125]        | −0.144*** [−0.150, −0.138]        |
| Interaction<br>(syntactic×lexical) | 0.032*** [0.029, 0.036]           | 0.004* [0.001, 0.007]             |
| <b>Control variables</b>           |                                   |                                   |
| Unigram surprisal                  | 0.196*** [0.184, 0.207]           | −0.263*** [−0.279, −0.247]        |
| Word length                        | 0.005 [−0.099, 0.109]             | 0.312*** [0.171, 0.454]           |
| Sentence position                  | −0.168*** [−0.175, −0.161]        | −0.118*** [−0.123, −0.113]        |
| Time (decade)                      | 0.087*** [0.085, 0.089]           | −0.004*** [−0.006, −0.002]        |

*Note.* Values represent estimated fixed-effect coefficients (log-odds,  $\beta$ ) from the GLMM analysis, with 95% confidence intervals reported in brackets. Confidence intervals were computed using the Wald approximation. Significance levels: \* $p < .05$ , \*\*\* $p < .001$ . All continuous predictors were z-score standardized prior to model fitting. Lexical surprisal (residual) represents pure lexical unpredictability after partialing out syntactic surprisal. Unigram surprisal is the negative log probability of the word form; higher values correspond to lower frequency.

lexical items, thereby competing with the noncanonical category assignment characteristic of conversion.

Word length revealed an asymmetry across conversion directions. It showed no significant effect on noun-to-verb conversion, but a positive effect on verb-to-noun conversion, indicating that longer verbs exhibit a higher degree of nominalization. Finally, time revealed a distinct asymmetry in historical trends. Noun-to-verb conversion degree increased significantly over the past century ( $\beta = 0.095, p < .001$  for 5-gram;  $\beta = 0.098, p < .001$  for GPT-2), while verb-to-noun conversion remained relatively stable ( $\beta = -0.004, p < .05$  for 5-gram;  $\beta = -0.001, p > .05$  for GPT-2). Although the 5-gram model yielded a formally significant coefficient, its magnitude is negligible, indicating no meaningful diachronic trend in verb-to-noun conversion degree. This asymmetry aligns with established findings on directional preferences in English conversion (Bauer, 1983; Biese, 1941; Marchand, 1969; Wang & Shao, 2025a), and provides diachronic evidence that noun-to-verb conversion degree has shown a consistent upward trend over the past century.

3.2. *Disentangling mechanisms: Independence and modulation*

To examine whether lexical and syntactic surprisal exert independent effects on conversion (H2a), we fitted a model in which lexical surprisal was residualized against syntactic surprisal. As shown in Table 3, both syntactic surprisal and residualized lexical surprisal exerted significant negative effects on conversion degree across both directions (all  $p < .001$ ). This finding suggests that the predictability constraint operates along two independent dimensions:

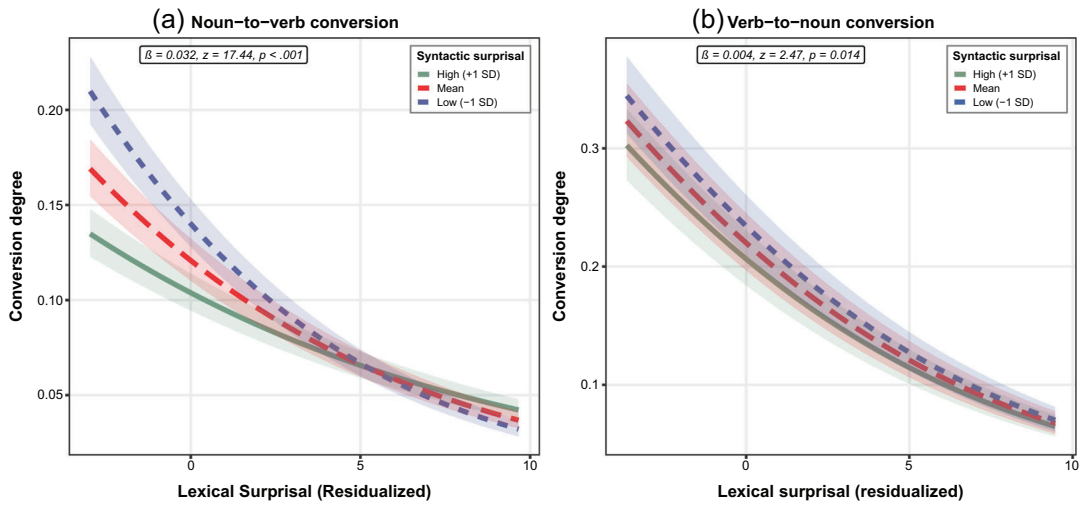


Fig. 1. Interaction effects of lexical and syntactic surprisal on conversion degree for (A) noun-to-verb and (B) verb-to-noun conversion. Curves represent the predicted probability of conversion (y-axis) derived from the fixed effects of the final GLMMs, with all other predictors held at their means. The x-axis represents residualized lexical surprisal. Separate lines represent the relationship at different levels of syntactic surprisal (Low =  $-1$  SD, Mean = 0, High =  $+1$  SD). Shaded areas indicate 95% confidence intervals. The curvature reflects the inverse-logit transformation characteristic of binomial regression.

conversion is inhibited not only when the syntactic slot is unexpected but also, independently, when the specific lexical item is unpredictable.

To test the modulation hypothesis (H2b), we examined the interaction between syntactic and lexical surprisal. As shown in Table 3, the interaction term was significantly positive for both noun-to-verb ( $\beta = 0.032$ ,  $p < .001$ ) and verb-to-noun conversion ( $\beta = 0.004$ ,  $p < .05$ ). Given the negative main effects, the positive interaction indicates an attenuating pattern: as syntactic surprisal increases, the inhibitory effect of lexical surprisal on conversion becomes weaker; conversely, when syntactic surprisal is low, the inhibitory effect of lexical surprisal is amplified.

This modulation effect is visualized in Fig. 1. In the noun-to-verb panel (A), the slope of lexical surprisal is steepest when syntactic surprisal is low (blue dashed line,  $-1$  SD) and becomes noticeably flatter when syntactic surprisal is high (green solid line,  $+1$  SD). The verb-to-noun panel (B) shows a similar but weaker pattern, consistent with the smaller interaction coefficient. These results are consistent with the resource competition account: when syntactic processing is relatively easy (low surprisal), cognitive resources are abundant, allowing the system to strictly monitor lexical selection—thus lexical predictability serves as a strong filter (steep slope). When syntactic processing becomes demanding (high surprisal), structural processing consumes shared cognitive resources, attenuating the system's sensitivity to lexical predictability (flatter slope). Notably, unigram surprisal, sentence position, and time remained significant predictors in the same directions as in the baseline model, while word length retained its asymmetric pattern across conversion directions.

#### 4. Discussion

The present study examined the mechanisms underlying the historical increase in English conversion degree over the past century. Drawing on the framework of efficient communication (Gibson et al., 2019; Kemp et al., 2018; Regier et al., 2015) and the UID hypothesis (Jaeger, 2010; Levy & Jaeger, 2007), we tested whether these theories can account for the seemingly paradoxical rise of conversion: a trend where speakers increasingly favor ambiguous, zero-marked forms to minimize articulatory effort (speaker's economy), seemingly at the expense of interpretive clarity required by listeners (auditor's economy). Our findings reveal a pattern of context-dependent optimization: the historical increase in conversion degree is not an unconstrained drift but is systematically licensed by the information-theoretic properties of the communicative context.

At a macro level, we identified contextual predictability as a universal constraint on conversion. The increased degree of English conversion observed over the past century (Shao et al., 2024; Wang & Shao, 2025b) is accompanied by a concurrent increase in contextual predictability, regardless of conversion direction or the scope of contextual information (local vs. broad). This finding offers a potential resolution to the communicative paradox: contextual predictability serves as the compensatory mechanism that reconciles the tension between the speaker's and auditor's economy. As English has historically shed overt inflectional markers (Allen, 2016), the burden of disambiguation has shifted from word-internal morphology to word-external context. The ambiguity inherent in conversion is thus functionally efficient: it allows for code compression precisely where context provides sufficient scaffolding to resolve the intended meaning.

This interpretation aligns with Piantadosi, Tily, and Gibson (2012), who argued that ambiguity is a functional property of efficient communication systems rather than a design flaw: when context is sufficiently informative, reusing an existing form avoids redundancy. Our macro-level findings are consistent with this view. However, Piantadosi et al. (2012) further predict that forms selected for reuse will be those that minimize processing cost, namely, forms that are short, high-frequency, and phonotactically well-formed. At this more granular level, our results diverge. The selection of which forms undergo conversion is not simply a matter of preferring the shortest or most frequent candidates. Rather, communicative pressure interacts with category-specific frequency profiles and morphological competition to shape which forms undergo conversion.

Regarding frequency, we observed opposing effects across the two conversion directions: higher-frequency verbs are more likely to undergo nominalization, while higher-frequency nouns are less likely to undergo verbalization. This divergence reflects the distinct functional profiles of the two lexical categories. For nouns, high frequency is associated with strong referential stability. High-frequency nouns designate conceptually stable, well-bounded referential categories, making them resistant to functional redeployment (Divjak & Caldwell-Harris, 2015; Langacker, 1987). Low-frequency nouns, by contrast, occupy a more marginal status in the lexical network and are more susceptible to recruitment for novel lexical category assignments (Stefanowitsch & Flach, 2017).

For verbs, the relationship runs in the opposite direction because entrenchment operates differently. High-frequency verbs tend toward schematicity through repeated use: their semantic content becomes increasingly dependent on constructional context rather than lexical specificity (Bybee, 2007). This schematic character facilitates integration into nominalization constructions, enabling transparent event readings (e.g., *a walk*, *a cut*) (Grimshaw, 1992). Low-frequency verbs, by contrast, retain richer argument structures that typically favor overt derivational marking over zero-derivation (e.g., *assassination*, not *\*an assassinate*). Thus, frequency does not exert a uniform efficiency pressure on conversion; its direction is determined by the distinct entrenchment profiles of nouns and verbs.

Regarding word length, Piantadosi et al.'s (2012) prediction that shorter forms are preferentially reused is likewise not borne out uniformly. In the present study, word length had no significant effect on noun-to-verb conversion, while for verb-to-noun conversion, longer verbs were more likely to undergo nominalization. Two mechanisms may account for this pattern. First, short words tend to carry higher lexical surprisal ( $r = -.13$  for noun-to-verb conversion;  $r = -.06$  for verb-to-noun conversion), reflecting the fact that short forms are typically more polysemous and thus contextually harder to predict. The categorial ambiguity inherent in conversion amplifies this cost (i.e., whether the form functions as a noun or a verb), rendering short words less suitable as candidates for zero-marking. Second, for verb-to-noun conversion specifically, short high-frequency verbs tend to have well-established suffixal nominalizations that suppress zero-derived nominal use (e.g., *permit* vs. *permission*), whereas longer verbs more often lack competing suffixal nominalizations, leaving zero-derivation as the primary available strategy. The positive effect of word length in verb-to-noun conversion thus reflects the joint action of reduced lexical surprisal and reduced morphological competition among longer forms.

The asymmetry in diachronic trends likewise reflects the interplay between efficiency pressure and morphological competition. As Jespersen and Haislund (1949) note, there is a stark structural imbalance in English word formation: verbs can be nominalized through at least 13 productive suffixes (e.g., *-ation*, *-ment*, *-ion*), while nouns can be verbalized through only three primary suffixes (*-en*, *-fy*, *-ize*). Marchand (1969) characterizes this as a fundamental lack of “competitive types” for nominal verbalization. The restricted and often register-specific character of these few verbal suffixes (e.g., *-ize* being technical, *-en* being largely restricted to adjective bases) leaves noun-to-verb conversion relatively uncontested, allowing it to expand historically. Verb-to-noun conversion, by contrast, faces a rich inventory of suffixal rivals that has historically constrained potential diachronic expansion of zero-derived nominals. Even as contextual predictability licenses greater ambiguity, the specific trajectory of conversion is, therefore, shaped by the architecture of the morphological system.

Taken together, these findings reveal a more complex picture than a uniform efficiency pressure on form properties would predict. Piantadosi et al.'s (2012) efficiency account operates at the level of the lexicon as a whole, predicting that systems globally favor reuse of short and frequent forms. Our results show that at the level of individual word formation processes, this global pressure is modulated by category-specific constraints. The direction and magnitude of conversion are shaped by the interaction of three factors: contextual predictability, which licenses zero-marking; category-specific entrenchment, which determines whether frequency

facilitates or resists conversion; and morphological competition, which determines whether zero-derivation is the available strategy in the first place. The relevant unit of analysis for predicting form reuse is, therefore, not the lexical item in isolation but the item situated within its morphological and categorical context.

Beyond the global constraint of contextual predictability, our fine-grained analysis distinguished the roles of structural and lexical expectations. This distinction resonates with a methodological concern in psycholinguistic modeling: the necessity of separating different types of surprisal to determine their individual contributions to processing difficulty. Roark et al. (2009) proposed methods for deriving separate lexical and syntactic surprisal measures, arguing that conflating these components obscures their distinct effects. Demberg and Keller (2008) provided empirical motivation for this separation: they showed that lexicalized surprisal fails to predict reading times due to high correlations with word frequency and transitional probability, whereas unlexicalized (syntactic) surprisal, being orthogonal to these lexical factors, emerged as a significant predictor. Our residualization approach addresses this collinearity problem, enabling the isolation of pure lexical unpredictability from syntactic expectations. Applying this decomposition to conversion, we found that both syntactic and lexical surprisal exert independent inhibitory effects. This aligns with van Schijndel and Schuler (2015), who showed that hierarchical syntactic structure remains a significant predictor of processing difficulty even against strong n-gram baselines, suggesting that the language processor maintains separable expectations at multiple levels of linguistic representation. Since conversion involves a tension between a word's established category and its contextual usage, both the structural slot (is a verb expected here?) and the specific lexical choice (is this particular word expected?) independently constrain the likelihood of a category shift.

Crucially, we observed that syntactic and lexical constraints do not operate in isolation. The significant positive interaction indicates that the inhibitory effect of lexical surprisal was stronger when syntactic surprisal was low and attenuated when syntactic surprisal was high. This pattern aligns with the resource competition account rather than the UID-based compensatory account. We interpret this finding within a capacity-limited processing framework: when syntactic surprisal is low, structural processing proceeds smoothly, leaving cognitive resources available for fine-grained lexical evaluation. Under such conditions, the system can strictly monitor whether a specific lexical item is appropriate, resulting in a steep inhibitory slope for lexical surprisal. Conversely, when syntactic surprisal is high, structural processing demands a larger share of limited cognitive resources (Fedorenko, Gibson, & Rohde, 2006; Futrell & Hahn, 2025). In such high-load contexts, the system may adopt "good-enough" processing strategies (Ferreira, Bailey, & Ferraro, 2002), relaxing its evaluation of lexical appropriateness. This yields an attenuated inhibitory effect, not because lexical predictability ceases to matter, but because the system lacks the surplus capacity to fully exploit this information.

This resource-modulated pattern resonates with recent theoretical advances seeking to unify expectation-based and memory-based accounts of language processing. Hahn et al. (2022) proposed a resource-rational framework in which processing difficulty reflects predictions derived from imperfect memory representations of context. A central insight of their

model is that the utility of predictive cues depends on the fidelity of context representation: when cognitive resources are constrained, fine-grained statistical information is less effectively exploited, and the system falls back on coarser priors. Although their empirical focus was on comprehension of nested recursive structures, the underlying principle, that resource limitations modulate the efficacy of predictive information, applies more broadly. Our finding that the constraining effect of lexical surprisal diminishes under high syntactic load provides converging evidence from word formation: the exploitation of lexical predictability is not absolute but contingent on the processing demands imposed by structural uncertainty.

Several limitations warrant consideration. First, our study focuses exclusively on English. While extending the analysis to other languages would offer valuable typological contrasts, such comparisons require diachronic corpora with full-text access comparable to COHA to derive valid contextual surprisal estimates. Constructing a multilingual dataset that ensures such strict comparability in genre balance and time depth falls outside the scope of the current study. Whether the reliance on contextual predictability to license category shifts holds universally across typologically diverse languages remains an open question. We hypothesize that while the regulatory role of contextual predictability is likely universal, its specific manifestation may vary. For isolating languages (e.g., Mandarin Chinese), we predict an even stronger reliance on contextual predictability, as context constitutes the primary cue for disambiguation. For morphologically rich languages (e.g., German, Italian), where explicit affixation is the norm, the effect of contextual predictability may be modulated by morphological cues. Future cross-linguistic work is needed to determine whether the resource competition pattern observed in English reflects a general property of human cognition or is specific to languages with limited inflectional morphology.

Second, the present study examines the dynamics of conversion degree rather than the emergence of conversion. Our dataset is restricted to lexical items already attested as conversion words in CELEX, meaning the analysis speaks to what contextual conditions modulate the degree to which an existing conversion word is utilized over time, not to what conditions determine whether a word enters into conversion in the first place. The generalizability of our findings to conversion emergence, therefore, remains an open question. By analogy, demonstrating that contextual predictability sustains and accelerates an ongoing process does not entail that it also triggers its initiation. The force sustaining motion need not be identical to the force that sets it in motion.

This focus is deliberate. The study is phenomenon-driven: its primary aim is to account for the seemingly paradoxical diachronic increase in noun-to-verb conversion degree rather than to model the conditions under which new conversion words arise. A study of emergence would additionally face considerable methodological challenges. Many core English conversion words were already established by Middle English (Biese, 1941), a period characterized by substantial grammatical restructuring, for which available corpus resources are too sparse and genre-restricted to support surprisal-based modeling. Reliable estimation of contextual predictability requires large, balanced, and continuously sampled text. These are conditions that diachronic corpora only begin to meet from the 20th century onward. Whether the predictability constraints identified here also govern the emergence of conversion remains an open and worthwhile question for future research.

Third, a related boundary condition concerns the relative frequency of the base versus the converted form. Our framework assumes that conversion is a “high-risk” strategy because the converted form is typically the secondary, less frequent function of the lemma. However, in cases where the converted form becomes the dominant use (i.e., occurring more frequently than the base form), the logic of our prediction inverts: the converted form becomes the default interpretation, and it is the original base form that would arguably require greater contextual scaffolding. Empirically, however, such dominance reversals are the exception rather than the rule. Our data indicate that only 3.4% of noun-to-verb and 9.0% of verb-to-noun lemmas maintained consistent converted-form dominance throughout the century. This confirms that for the vast majority of the lexicon, conversion remains directionally asymmetric, with the base form serving as the high-frequency anchor. These rare, high-dominance items likely represent a distinct, advanced stage of lexicalization, where the converted form has achieved functional independence from its base. How communicative pressures operate at this endpoint of the lexicalization trajectory remains an open question for future studies.

Despite these limitations, this study provides a novel information-theoretic account of language change. The evolution of English conversion is not a random drift, nor is it simply a matter of globally preferring short and frequent forms. Rather, it reflects a dynamic optimization process in which communicative efficiency is constrained by multiple interacting factors: the information-theoretic properties of the communicative context, cognitive resource limitations that modulate the interplay between syntactic and lexical processing, category-specific entrenchment, and the architecture of the morphological system. Viewed through this lens, the lexicon achieves greater expressivity and economy not through uniform optimization pressure but through the structured interaction of contextual, cognitive, categorical, and morphological factors.

## Acknowledgments

This work is supported by the Fundamental Research Funds for the Central Universities of China.

## Conflicts of interest

The authors declare no conflicts of interest.

## Data and code availability statement

The raw corpus data are from the Corpus of Historical American English (COHA), a proprietary dataset available by subscription (<https://www.english-corpora.org/coha/>). Due to licensing restrictions, raw texts cannot be shared. However, a complete replication package, including all Python scripts for surprisal extraction, R scripts for statistical modeling and

visualization, and processed datasets, is available on the Open Science Framework (OSF) at: [https://osf.io/ab26m/overview?view\\_only=94cd0c15d05c4c81a39d28f9c669f8df](https://osf.io/ab26m/overview?view_only=94cd0c15d05c4c81a39d28f9c669f8df).

## Ethics approval statement

No ethics approval was required for this study, as it involved no human participants and relied solely on publicly available corpus data.

## References

- 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>
- Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). CELEX2 (p. 287744 KB). Linguistic Data Consortium. <https://doi.org/10.35111/GS6S-GM48>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauer, L. (1983). *English word-formation*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139165846>
- 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 Tiedeakatemia.
- Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89(1), 1–47. <https://doi.org/10.1037/0033-295X.89.1.1>
- Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University Press.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>
- Davies, M. (2010). *The Corpus of Historical American English (COHA)*. Retrieved from <https://www.english-corpora.org/coha/>
- 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>
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193–210. <https://doi.org/10.1016/j.cognition.2008.07.008>
- Divjak, D., & Caldwell-Harris, C. L. (2015). 3. Frequency and entrenchment. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 53–75). DE GRUYTER. <https://doi.org/10.1515/9783110292022-004>
- Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54(4), 541–553. <https://doi.org/10.1016/j.jml.2005.12.006>
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15. <https://doi.org/10.1111/1467-8721.00158>
- Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, 22(6), 829–834. <https://doi.org/10.1177/0956797611409589>
- Futrell, R., & Hahn, M. (2025). Linguistic structure from a bottleneck on sequential information processing. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-025-02336-w>
- 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>
- Grimshaw, J. (1992). *Argument structure*. MIT Press.

- Hahn, M., Degen, J., & Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient trade-off of memory and surprisal. *Psychological Review*, 128(4), 726–756. <https://doi.org/10.1037/rev0000269>
- Hahn, M., Mathew, R., & Degen, J. (2021). Morpheme ordering across languages reflects optimization for processing efficiency. *Open Mind*, 5, 208–232. [https://doi.org/10.1162/opmi\\_a\\_00051](https://doi.org/10.1162/opmi_a_00051)
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies 2001 - NAACL*. <https://doi.org/10.3115/1073336.1073357>
- Hao, Y., Mendelsohn, S., Sterneck, R., Martinez, R., & Frank, R. (2020). Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. In E. Chersoni, C. Jacobs, Y. Oseki, L. Prévot, & E. Santus (Eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 75–86). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.cmcl-1.10>
- Heafield, K. (2011). KenLM: Faster and smaller language model queries. In C. Callison-Burch, P. Koehn, C. Monz, & O. F. Zaidan (Eds.), *Proceedings of the 6th Workshop on Statistical Machine Translation* (pp. 187–197). Association for Computational Linguistics. Retrieved from <https://aclanthology.org/W11-2123/>
- Jaeger, T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. <https://doi.org/10.1016/j.cogpsych.2010.02.002>
- Jespersen, O., & Haislund, N. (1949). A modern English grammar on historical principles. Part VII. Retrieved from <http://archive.org/details/a-modern-english-grammar-on-historical-principles-part-vii>
- Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, 4(1), 109–128. <https://doi.org/10.1146/annurev-linguistics-011817-045406>
- 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). <https://doi.org/10.18653/v1/W16-2015>
- Langacker, R. W. (1987). Nouns and verbs. *Language*, 63(1), 53–96. <https://doi.org/10.2307/415384>
- Leech, G., & Smith, N. (2000). *Manual to accompany the British National Corpus (Version 2) with improved word-class tagging*. Retrieved from [https://uclrel.lancs.ac.uk/bnc2/bnc2postag\\_manual.htm](https://uclrel.lancs.ac.uk/bnc2/bnc2postag_manual.htm)
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hofmann (Eds.), *Advances in neural information processing systems 19* (pp. 849–856). MIT Press. <https://doi.org/10.7551/mitpress/7503.003.0111>
- Mahowald, K., Dautriche, I., Gibson, E., & Piantadosi, S. T. (2018). Word forms are structured for efficient use. *Cognitive Science*, 42(8), 3116–3134. <https://doi.org/10.1111/cogs.12689>
- Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach* (2nd Revised & Enlarged edition). C. H. Beck'sche Verlagsbuchhandlung.
- Meister, C., Pimentel, T., Haller, P., Jäger, L., Cotterell, R., & Levy, R. (2021). Revisiting the uniform information density hypothesis. In M.-F. Moens, X. Huang, L. Specia, & S. W. Yih (Eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 963–980). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.emnlp-main.74>
- 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>
- Mollica, F., Bacon, G., Zaslavsky, N., Xu, Y., Regier, T., & Kemp, C. (2021). The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, 118(49), e2025993118. <https://doi.org/10.1073/pnas.2025993118>
- Ney, H., Essen, U., & Kneser, R. (1994). On structuring probabilistic dependences in stochastic language modelling. *Computer Speech & Language*, 8(1), 1–38. <https://doi.org/10.1006/csla.1994.1001>
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526–3529. <https://doi.org/10.1073/pnas.1012551108>

- 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. Retrieved from <https://www.penguin.co.uk/books/16256/the-language-instinct-by-steven-pinker/9780141980775>
- Plag, I. (2003). *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. Retrieved from <https://www.R-project.org/>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*.
- Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (1st ed., pp. 237–263). Wiley. <https://doi.org/10.1002/9781118346136.ch11>
- Rezaii, N., Mahowald, K., Ryskin, R., Dickerson, B., & Gibson, E. (2022). A syntax–lexicon trade-off in language production. *Proceedings of the National Academy of Sciences*, 119(25), e2120203119. <https://doi.org/10.1073/pnas.2120203119>
- Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In P. Koehn & R. Mihalcea (Eds.), *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 324–333). Association for Computational Linguistics. Retrieved from <https://aclanthology.org/D09-1034/>
- Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45), e2105646118. <https://doi.org/10.1073/pnas.2105646118>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.
- Shain, C. (2019). A large-scale study of the effects of word frequency and predictability in naturalistic reading. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4086–4094). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1413>
- Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, 138, 107307. <https://doi.org/10.1016/j.neuropsychologia.2019.107307>
- Shain, C., Meister, C., Pimentel, T., Cotterell, R., & Levy, R. (2024). Large-scale evidence for logarithmic effects of word predictability on reading time. *Proceedings of the National Academy of Sciences*, 121(10), e2307876121. <https://doi.org/10.1073/pnas.2307876121>
- 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>
- Slaats, S., Meyer, A. S., & Martin, A. E. (2024). Lexical surprisal shapes the time course of syntactic structure building. *Neurobiology of Language*, 5(4), 942–980. [https://doi.org/10.1162/nol\\_a\\_00155](https://doi.org/10.1162/nol_a_00155)
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319. <https://doi.org/10.1016/j.cognition.2013.02.013>
- Stefanowitsch, A., & Flach, S. (2017). The corpus-based perspective on entrenchment. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 101–127). American Psychological Association. <https://doi.org/10.1037/15969-006>
- van Schijndel, M., & Schuler, W. (2015). Hierarchic syntax improves reading time prediction. In R. Mihalcea, J. Chai, & A. Sarkar (Eds.), *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1597–1605). Association for Computational Linguistics. <https://doi.org/10.3115/v1/N15-1183>

- Wang, G., & Shao, B. (2025a). Exploring the effect of semantic diversity on boundary permeability in verb/noun heterosemy using deep contextualized word embedding. *Linguistics Vanguard*, 11(1), 525–534. <https://doi.org/10.1515/lingvan-2025-0121>
- Wang, G., & Shao, B. (2025b). *Less is more: Diachronic evidence for efficient communication in English noun-to-verb conversion over the past century*. School of International Studies, Zhejiang University.
- 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>
- Xu, Y., Regier, T., & Malt, B. C. (2016). Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, 40(8), 2081–2094. <https://doi.org/10.1111/cogs.12312>
- Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31), 7937–7942. <https://doi.org/10.1073/pnas.1800521115>
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.

### Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information