



# Jan Kochanowski University Press

This is a contribution from *Token: A Journal of English Linguistics*  
Volume 14/2022.

Edited by John G. Newman and Marina Dossena.

© 2022 Jan Kochanowski University Press.

# Frequency, selection and random chance in the (ir)regularization of English verbs<sup>1</sup>

Beata Łukaszewicz  
*University of Warsaw*

## ABSTRACT

This paper offers a few remarks on the role of frequency, selection and random chance in the development of (ir)regular preterite forms in English verbs. It considers the possibility of quantifying the evolutionary dynamics of the process, as proposed in recent large-scale studies. Such studies view language as a dynamical system or a mutating living organism, affected by competing random and natural selective forces. They shed light on stochasticity- and selection-driven aspects of the evolution of the linguistic system, and are thus complementary with respect to frequency-oriented linguistic approaches (such as usage-based/exemplar models).

Keywords: regularization, irregularization, analogical levelling, English preterite, ablaut, linguistic change, linguistic innovation.

## 1. Introduction

Frequency has long been considered to be a prominent factor in language change. High- and low-frequency words (or even whole expressions) are known not to be equally affected by various diachronic processes. Reductive sound change tends to affect high-frequency words before low-frequency words (as in *have* > *'ve*, *will* > *'ll*, *do not* > *don't*, etc.; e.g. Mańczak 1969 et seq.;

---

<sup>1</sup> I would like to dedicate this article to Professor Jerzy Wełna in appreciation of his contribution to the field of the history of the English language. Working for many years in the same department, it has been a privilege and a great pleasure to witness his immense knowledge of intricacies and unsolved riddles of historical linguistics, which he has so enthusiastically shared with generations of students. Thank you!

Hooper 1976); the extremely high frequency of certain constructions may lead to their grammaticalization (as in the case of the English construction *going to*; Bybee 2003, 2006: 719-721); analogical levelling or regularization tend to affect low-frequency words before high-frequency words (e.g. Bybee 2006: 728-729; Lieberman et al. 2007).

One of the best documented cases of analogical levelling is the levelling of English strong verbs to the 'weak' verb pattern, which has occurred throughout the history of English and manifests itself also in the present-day dialectal variability of the language (Campbell 1998: 92). (There have also been numerous studies of strong-verb paradigm levelling in other Germanic languages; see e.g. Nübling 2000; Dammel et al. 2010; Knooihuizen – Strik 2014; De Smet – Van de Velde 2020; Nowak 2020.) However, although many sources point to frequency-driven regularization as the main direction of the evolutionary change, irregularization has also been attested (e.g. Hare – Elman 1995; Lieberman et al. 2007; Anderwald 2013; Newberry et al. 2017). The mechanisms underlying these processes are not fully understood.

In exemplar-based models of language production and perception (e.g. Bybee 2002, 2006), both morphologically regular and irregular words (as well as phrases) are assumed to be stored in the mental lexicon; the strength of mental representations is assumed to vary depending on frequency. Such models predict that the high token frequency of irregular verbs will make them immune to regularization, as frequent usage makes exceptional forms easy to learn, store in memory, and transmit flawlessly from one generation of speakers to another. Conversely, low-frequency irregular verbs are predicted to be more difficult to acquire, more liable to memory decay, and thus more likely to be attracted by the dominating regular (high type frequency) pattern. In short, 'among English irregular verbs the low-frequency verbs are more likely to regularize (*weep, weeped*) while the high-frequency verbs maintain their irregularity (*keep, kept*)' (Bybee 2006: 715).

However, not only the regular pattern can attract linguistic innovation. Some productivity in irregularization has also been attested, both historically and in psycholinguistic experiments. For instance, in the experiment reported in Bybee and Slobin (1982), both adult and child participants sometimes supplied innovative irregular past tense forms for the verbs they heard, in addition to the elicited regularization. To illustrate, *swum* was the elicited past tense form of *swim* in 25% of adult responses and in 7% of third-graders' responses; *shrunk* appeared as the past tense form of *shrink* in 25% of adult responses and in 60% of the children's responses (see also the summary in Bybee – Moder 1983: 254). The productivity of the *string/*

*strung* class to which the above verbs were attracted can be seen not only from the synchronic but also from the diachronic perspective. The *string/strung* class has been viewed as the most productive class of strong verbs in the history of English, attracting a large proportion of new members: two thirds of the contemporary members of the class were not strong verbs in Old English (Jespersen 1942; for discussion see Bybee – Moder 1983). Bybee and Moder (1983) investigated why some irregular subclasses become productive and attract new members, while others gradually decay as their members become regularized in the historical process. Their experimental data (nonce words eliciting past tense forms) pointed to some phonological attributes of the *string/strung* class that can be linked to its productivity. Phonological features, such as the final consonants as well as the initial consonants and clusters, and to a minor degree also the base vowel identity, were found to define the class in terms of a family resemblance to a prototypical member of the class. Historical change has also been modelled using connectionist nets, relying on the fact that more frequent patterns or those that share phonological regularities with a number of others are learned more quickly and with lower error (Hare – Elman 1995). Relatedly, the relationship between frequency of occurrence and response accuracy (productivity) has also been observed in first language acquisition studies. In the seminal ‘wug-test’ study of Berko (1958), the productivity score (i.e. % correct responses) for the regular preterite inflectional pattern differed enormously depending on whether a real word or a synthetic item (nonce word) was the input; cf. *melted*: 72% (preschoolers)/74% (first graders) versus *motted*: 32% (preschoolers)/33% (first graders). While the role of frequency and stochasticity in the productivity of the English morphological pattern has been widely acknowledged, the questions whether analogy- or rule-based mechanisms are at play, and whether regulars and irregulars are acquired and computed using the same mental mechanisms are highly debatable. The acquisition of the English past tense morphology has been modelled, among others, within the connectionist framework, attributing all kinds of generalizations to analogy (Rumelhart – McClelland 1986; Plunkett – Marchman 1993), the dual mechanism model (Prasada – Pinker 1993), in which regular inflection is carried out by a general rule (or rules) while irregular inflection is based on analogy, and the multiple-rule model (Albright – Hayes 2003), where regulars and irregulars are both computed using stochastic rules characterized by various degrees of reliability.

In this paper, I address the role of frequency, selection, and random chance in the process of (ir)regularization of English preterite forms,

focusing on the large-scale quantification of the (ir)regularization process. In §2, I outline the basic differences between Old English and Modern English conjugation of preterite and past participle forms, and sketch out various (ir)regularization paths of English verbs, as known from linguistic studies. In §3, I discuss the results of recent studies based on huge digital corpora, which attempt to quantify the long-term evolutionary dynamics of (ir)regularization or, using rigorous methods from evolutionary biology, test whether the process depends on selective forces or random chance. §4 contains concluding remarks.

## 2. (Ir)regularization paths: A comparison of Old English and Modern English verbs

Old English possessed an elaborate system of conjugation, with two broad classes of verbs: strong and weak, which differed in how their preterite and past participle were formed (e.g. Mitchell – Robinson 2001). The strong verbs exhibited a change in the root vowel (known as ‘ablaut’ or ‘gradation’), the weak verbs formed their preterite (pret.) and past participle (past ptc.) by adding a dental suffix. Seven classes of Old English strong verbs are distinguished, each exhibiting a different gradation series. In some strong verbs, four different vowels appeared in the so-called ‘principal parts’ of the verb: infinitive (inf.), preterite singular (pret. sg.), preterite plural (pret. pl.), and past participle (past ptc.); e.g. *crēopan* ‘creep’ (inf.), *crēap* (pret. sg.), *crupon* (pret.pl.), *cropen* (past ptc.), a class II verb, or *feohtan* ‘fight’ – *feaht* – *fuhton* – *fohten*, a class III verb. In others, only three or two different vowels can be found; e.g. *faran* ‘go’ – *fōr* – *fōron* – *faren*, a class VI verb (2001: 36-39). Weak verbs are usually divided into three classes, all of which are characterized by the addition of the dental suffix, as in Modern English. The stem vowel usually remains unaltered across the paradigm, although exceptions also exist, e.g. *sēcan* ‘seek’ (inf.) – *sōhte* (pret. sg.) – *sōht* (past ptc.), a class 1 verb (2001: 49). In such cases, the opaque vowel alternations had arisen through phonological factors operating in the prehistoric past. One of such factors was *i*-mutation; cf. *sēcan* < \**sōkjan* (2001: 49).

Superficially, it seems that the present-day irregular and regular verbs are direct descendants of the Old English strong-weak division. However, a closer comparison of Old English and Modern English conjugations shows that although the correspondence is close, it is certainly not one-to-one.

Modern English has a highly regular conjugation pattern whereby the pret. and past ptc. are expressed by suffixing *-ed* to the stem, as in *love/loved/loved* (with the suffix exhibiting phonologically governed allomorphy: /d/ ~ /t/ ~ /ɪd/). It also has a set of irregular verbs, within which certain subpatterns are discernible. Gleason (1955: 102-103) presents a (non-exhaustive) list of 52 irregular subclasses: two with more than ten members each, ten containing at least three verbs each, six with two verbs each and thirty-four containing only a single verb each.

Numerous Modern English irregular verbs are remnants of the Old English strong conjugation. In the case of such verbs, it is immediately apparent that in the course of history stem levelling must have taken place, as the Old English pret. sg. and pret. pl. forms correspond to a single Modern English pret. form. For example, ModE *choose/chose/chosen* descends from the OE *céosan* 'choose' (inf.) – *céas* (pret. sg.) – *curon* (pret. pl) – *coren* (past ptc.). The pret. sg. and the pret. pl. forms are illustrated in (1).

- (1) *Oxford English Dictionary (OED)* entry: 'choose, *v.*' (c. 893)

a1000 *Ps.* (Spelm.) cxviii[i]. 173 *Bebodu ðine ic ceas.*

OE *Genesis* 1803 *Him þa wic curon.*

Some other Modern English irregular verbs show the suffixation combined with a special pret. or past ptc. form of the stem, e.g. *seek/sought/sought*. As mentioned above, *sēcan* 'seek' (inf.) – *sōhte* (pret. sg.) – *sōht* (past ptc.) was a weak class 1 verb in Old English and 'had the same irregularity [in the stem vowel] even then' (Mitchell – Robinson 2001: 48). In Chaucer's time, this kind of irregularity was still present not only in *seke(n) – soughte – sought*, but also e.g. in *werke(n) – wroghte – wroght*, which corresponds to the now fully regular *work/worked/worked*; cf. Burnley (1983: 27). (The irregular *wrought* has been lexicalized.)

Throughout the centuries, many English verbs have developed the regular suffixation pattern; e.g. the ModE *help/helped/helped* used to be a strong verb in Old English. In Chaucer's time the regularization process was still incomplete, i.e. both weak and strong variants were attested, as exemplified in (2) (the variable prefix *y-/i-* is omitted in past ptc. forms); see Burnley (1983: 28). For a quantitative study of the shift from the strong to the weak verbal category in the period between Old English and the year 1500, the reader is referenced to Krygier (1994).

## (2) Burnley (1983: 28)

*crepe(n)* (inf.) – *crepte* (pret.) – *crept* ~ *cropen* (past ptc.)  
*wepe(n)* (inf.) – *wepte* ~ *weepe* (pret.) – *wept* ~ *wepe(n)*, *wope(n)* (past ptc.)  
*helpe(n)* (inf.) – *heelpe* (pret.) – *holpe(n)* (past ptc.)  
*fare(n)* (inf.) – *ferde* (pret.) – *fare(n)* (past ptc.)  
*hange(n)* ~ *honge(n)* (inf.) – *he(e)ng* ~ *honged* (pret.) – *hanged* ~ *honged*  
 (past ptc.)

Interestingly, the first two examples in (2), *crepte* (pret.) and *wepte* (pret.), which survive into Modern English as part of the irregular pret./past ptc. subpattern in (3a), can be considered as examples of (temporary) ‘regularization’ from a diachronic perspective. They not only exhibit suffixation, but are also fully analogous to other examples in which the stem allomorphy was caused through a transparent historical phonological process; e.g. in *keep/kept/kept* the shortening of /e:/ resulted in Late Old English ~ Early Middle English /e/ in the environment of the following consonantal cluster (Welna 1978: 64-65); the long /e:/ changed later to /i:/ (as part of the Great Vowel Shift starting in the 15<sup>th</sup> c. (1978: 88-89). The /i:/ ~ /e/ alternation is also characteristic of some Modern English verbs whose stems end in *-t/-d* (3b).

## (3) Gleason (1955: 102)

a. Pret. = Past Ptc. = /-t/ suffixation plus /i:/ > /e/ change in the stem  
*creep*; *deal*; *feel*; *keep*; *leap*; *mean*; *sleep*; *sweep*; *weep*

b. Pret. = Past Ptc. = /i:/ > /e/ change in the stem  
*bleed*; *breed*; *feed*; *lead*; *meet*; *plead*; *read*; *speed*

Regularization is not the only possible direction of change. Irregularization has also been attested, although, as often underscored, it tends to be much less common (e.g. Lieberman et al. 2007; Anderwald 2013; Anderson 2015). In the words of Lieberman et al. (2007: 713): “[r]egular verbs become irregular much more rarely: for every ‘sneak’ that ‘snuck’ there are many more ‘flew’ that ‘flied’ out”. Still, there is a number of originally regular verbs (e.g. *dive*, *plead*, *sneak*) which have developed irregularity in some Modern English dialects (*dove*, *pled*, *snuck*). This happened sometimes through proportional analogy, i.e. after the patterns of other verbs (*drive/drove*, *lead/led*), and

sometimes, as in *sneak/snuck*, with no obvious relationship to previously existing patterns (Anderwald 2013; Anderson 2015).

Most irregular verbs in Modern English are survivals from Old English; new verbs coined or borrowed today follow the weak conjugation, e.g. *google – googled – googled* (Lieberman et al. 2007: 713). However, sporadic irregularity is also observed in new words. For example, Mitchell and Robinson (2001: 35) note that when *strive* was borrowed from French in the 13<sup>th</sup> c., it followed the pattern of *drive* because the two rhymed; hence ModE *strive – strove – striven*. At the same time, *jive*, which is a comparably new verb, is conjugated *jived – jived*, not *jive – jove – jiven*. (OED traces the origin of the word *jive* to the 1928 title of L. Armstrong's gramophone record *Don't jive me*.) Some variability in the conjugation of *strive* may be observed among native speakers of English; e.g. Campbell (1998: 92) notes that '*strive/strove/striven* for many speakers has changed to *strive/strived/strived*'. (However, according to Google Ngrams, accessed 30-01-2022, the irregular *strove* still prevails.)

Apart from (ir)regularization, there are well-known shifts of irregular verbs, which have not resulted in their regularization. The suppletive form *went* replaced the earlier pret. *ēode*, which had also been a suppletive form with respect to the infinitival *gān*. The intricacies of this 'suppletion for suppletion' process are succinctly described in Weřna (2001). Some English strong verbs exhibit 'partial levelling' – a diachronic shift from one strong pattern to another (Campbell 1998: 92). Bybee and Moder (1983: 252) speak of a historical trend for members of the class with three forms (*sing/sang/sung*) to lose their separate past-tense forms, and thus become members of the *string/strung* class. They point out that many speakers of American English do not use such forms as *sprang*, *shrank*, and *stank*. (In general, ablaut patterns with the same vowel in the pret. and past ptc. forms have been preferred throughout the history of English as well as other Germanic languages, e.g. German and Dutch, but not Swedish; Dammel et al. 2010.)

### 3. Quantifying the (ir)regularization: Frequency, selection, and random chance

Below I consider two recent large-scale quantitative studies of the evolution of Modern English (ir)regular verbs. Lieberman et al. (2007) treat language as a dynamical system and study how the rate of regularization depends on the frequency of word usage. Newberry et al. (2017) assume that languages



mutate like genes, and using methods from evolutionary biology, study to what extent language evolution depends on natural selection or is governed by random chance.

As is well known, regular and irregular verbs in Modern English are not distributed evenly across the lexicon. Regular verbs are very frequent in terms of total numbers – thus contributing to the high type frequency of the morphological pattern (the infinitival stem plus the *-ed* suffix). Irregular verbs constitute a much smaller set, only less than 3% of modern verbs are irregular. However, they tend to occupy the highest frequency bins in the lexicon: ‘the ten most common verbs are all irregular (be, have, do, go, say, can, will, see, take, get)’; Lieberman et al. (2007: 713).

Lieberman et al. (2007) investigated the dynamics of evolution of the regular pret./past ptc. pattern over the past 1,200 years, compiling data from Old English through Middle English to Modern English on the basis of a large collection of grammar textbooks. They created a list of 177 Old English irregular verbs that remain part of the language to this day. ‘Of these 177 Old English irregulars, 145 remained irregular in Middle English and 98 are still irregular in Modern English’ (p. 713).

For each of the 177 verbs, frequencies were obtained on the basis of the CELEX corpus (containing 17.9 million words). The frequency range was divided into six logarithmically spaced bins (from  $10^{-6}$  to 1). The findings on the spacing of irregular verbs within each bin are summarised in Table I below. The verbs that were identified in Lieberman et al.’s study as those which had regularized are indicated in bold type. It can be readily observed that the lower the frequency bin, the larger the fraction of the verbs that have undergone regularization. There are only irregular verbs in the two highest frequency bins, thus the proportion of regularization is 0% in these bins. In the third frequency bin,  $10^{-3}$  to  $10^{-2}$ , 33 of 37 (90% of) verbs are still irregular in Modern English, which gives the regularization score of 10%. In absolute terms, the largest number of irregular verbs is found in the fourth frequency bin,  $10^{-4}$  to  $10^{-3}$ . According to Lieberman et al., in this bin, 65 irregular Old English verbs have left 57 in Middle English and 37 in Modern English, which in relative terms yields the regularization rate of 43%. The fifth and the sixth frequency bins,  $10^{-5}$  to  $10^{-4}$  and  $10^{-6}$  to  $10^{-5}$ , show still higher overall regularization scores: 72% and 91%, respectively. In terms of raw numbers, Lieberman et al. note that in the fifth frequency bin, 50 irregulars of Old English have left 29 in Middle English and 14 in Modern English, whereas in the sixth frequency bin, 12 irregulars of Old English have left 9 in Middle English and only 1 in Modern English. The quantitative analysis

allows them to conclude that the regularization process applies gradually, at a certain frequency-dependent rate which is stable across time: ‘an irregular verb that is 100 times less frequent is regularized 10 times as fast’ (p. 714).

Table 1. The frequency bins and regularization percentages within each bin for 177 Old English irregular verbs; see Lieberman et al. (2007: 714)

Frequency	Verbs	Regularization %
$10^{-1}-1$	be, have	0
$10^{-2}-10^{-1}$	come, do, find, get, give, go, know, say, see, take, think	0
$10^{-3}-10^{-2}$	begin, break, bring, buy, choose, draw, drink, drive, eat, fall, fight, forget, grow, hang, <b>help</b> , hold, leave, let, lie, lose, <b>reach</b> , rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, <b>walk</b> , win, <b>work</b> , write	10
$10^{-4}-10^{-3}$	arise, <b>bake</b> , bear, beat, bind, bite, blow, <b>bow</b> , burn, burst, <b>carve</b> , <b>chew</b> , <b>climb</b> , cling, creep, <b>dare</b> , dig, <b>drag</b> , flee, <b>float</b> , <b>flow</b> , fly, <b>fold</b> , freeze, grind, leap, lend, <b>lock</b> , <b>melt</b> , <b>reckon</b> , ride, <b>rush</b> , <b>shape</b> , shine, shoot, shrink, <b>sigh</b> , sing, sink, slide, <b>slip</b> , <b>smoke</b> , spin, spring, <b>starve</b> , steal, <b>step</b> , <b>stretch</b> , strike, <b>stroke</b> , <b>suck</b> , <b>swallow</b> , swear, sweep, swim, swing, tear, wake, <b>wash</b> , weave, weep, <b>weigh</b> , wind, yell, <b>yield</b>	43
$10^{-5}-10^{-4}$	<b>bark</b> , <b>bellow</b> , bid, <b>blend</b> , <b>braid</b> , <b>brew</b> , <b>cleave</b> , <b>cringe</b> , <b>crow</b> , dive, <b>drip</b> , <b>fare</b> , <b>fret</b> , <b>glide</b> , <b>gnaw</b> , <b>grip</b> , heave, <b>knead</b> , <b>low</b> , <b>milk</b> , <b>mourn</b> , <b>mow</b> , <b>prescribe</b> , <b>redden</b> , <b>reek</b> , <b>row</b> , <b>scrape</b> , <b>seethe</b> , shear, shed, <b>shove</b> , slay, slit, <b>smite</b> , sow, <b>span</b> , <b>spurn</b> , sting, stink, strew, stride, swell, <b>tread</b> , <b>uproot</b> , <b>wade</b> , <b>warp</b> , <b>wax</b> , <b>wield</b> , wring, <b>writhe</b>	72
$10^{-6}-10^{-5}$	<b>bide</b> , <b>chide</b> , <b>delve</b> , <b>flay</b> , <b>hew</b> , <b>rue</b> , <b>shrive</b> , slink, <b>snip</b> , <b>spew</b> , <b>sup</b> , <b>wreak</b>	91

With such an elegant expression of the relationship between the lexical frequency of an irregular verb and its propensity for change, it might be tempting to predict ‘the future of the past tense’ (Lieberman et al. 2007: 715). Indeed, Lieberman et al. (2007: 715) assert that ‘if the current trends continue, only 83 of the 177 verbs studied will be irregular in 2500’. They even speculate that the next verb to undergo regularization is likely to be *wed*, which ranks at the very bottom of irregular verbs in Modern English. They further observe that the form *wed* is already being replaced in many contexts by *wedded*.

However, it seems that the future of individual verbs is more uncertain than the overall long-term decay of irregular forms, and factors other than frequency may be at play. First, we may observe that verbs such as *wed* do not directly compare to the verbs in Table I, which all either descend from the Old English strong conjugation patterns or, like *seek* – from a weak conjugation, but with an irregularity in the stem caused by a prehistoric phonological pattern. Unlike those verbs, *weddian* was a weak verb in Old English. *OED* lists two possible conjugation patterns in Modern English, *wedded* and *wed*, but also notes that ‘the form *wed* in the past tense is now only dialectal; in the past participle it is common *dialect* but otherwise rare except in poetry’. Thus, if both *wedded* and *wed* exist as competing pret. forms in some varieties of English, the latter must be an instance of irregularization. An interesting question is then what factors might have contributed to the irregularization process. Some enlightenment seems to be offered by studies in language acquisition. The famous *wug* experiment by Berko (1958) revealed that past tense inflection was particularly difficult for children in the case of nonce stems ending in /-t/ or /-d/. For such stems, which were *mot* and *bod* in that experiment, ‘the wrong answers, which were in the majority, were overwhelmingly a repetition of the present stem’. In other words, ‘[t]o the forms ending in /-t/ or /-d/ the children added nothing to form the past’ (p. 166). Similar results were also reported in Bybee and Slobin (1982: 274). Thus, there seems to be a parallel between the patterns in acquisition and the *wedded* > *wed* shifts in adult dialectal speech. We may also observe that the irregular *cut/cut/cut* class to which the verb *wed* seems to be attracted is the most populated class of irregular verbs in Modern English: according to Gleason (1955: 102), it consists of 19 members. (For comparison, the productive *string/strung* class is the second largest irregular class in Gleason’s study, and consists of 14 members.)

Another interesting question arises through comparison of the verbs *wed/wedded* (~*wed*), and *slink/slunk* (~*slinked*). The latter verb also occupies

the bottom frequency bin (as seen in Table I). Unlike *wed*, it was a strong verb in Old English (see Wright – Wright 1914, §498, p. 257). Despite a persistent chance of becoming regular, as judged by its frequency, it has remained an exception, joining the most productive irregular *string/strung* class in the course of history. It seems then that the paths of evolution of individual verbs may stabilize in different ways depending on factors other than frequency alone.

Thus, there are clearly further points of interest, parallel to the question about frequency-driven regularization rates, namely, how fast irregularization proceeds and what mechanisms govern it, as well as what factors may hinder the regularization process in the case of low-frequency verbs. The list of Modern English irregular verbs compiled by Lieberman et al. (and available in their Supplementary Information) contains a substantial number of verbs which used to be weak verbs in Old English – such verbs constitute nearly one third of the list (the total of 249 Modern English irregular verbs – 177 Old English irregular verbs = 72 Old English weak verbs). According to Newberry et al. (2017), irregularization is as common as regularization in Modern English.

Newberry et al. (2017) used numerical methods from population genetics to investigate to what extent the development of (ir)regular verbal forms in American English, e.g. *spilt* > *spilled* was driven by stochasticity or governed by natural selection. They analysed regular versus irregular past-tense variants for 36 verbs (704,081 tokens in total), extracted from the Corpus of Historical American English, comprising over four million words from over 100,000 texts between the years 1810 and 2009. The assumption is that language mutates like genes, and thus both stochasticity (cf. ‘genetic drift’ in population genetics) and selective forces are expected to be at play. Stochasticity in transmission is a random change in frequency from one generation to another, caused solely by accidents of sampling. Selection entails the influence of various forces (e.g. linguistic, cognitive and social). Although the irregular past-tense forms were expected to ‘regularize over time for reasons of economy or cognitive ease’ (2017: 224), Newberry et al. report that random chance rather than selective pressures appeared to be the factor underlying most of the system’s variability. The overall finding was that rare forms were more prone to replacement than common ones; more common words, regardless of whether selective forces were at play or not, were affected by less stochasticity in transmission. The study reports selection as the driving force in several verbs exhibiting variability in the pret. form. Surprisingly, the irregular variant is favoured in four verbs (*lighted* >

*lit, waked > woke, sneaked > snuck, dived > dove*), whereas the regular one only in two (*wove > weaved, smelt > smelled*). As Newberry et al. point out, one possible explanation for the irregularization patterns is the natural propensity for rhyming, as reported in psycholinguistic studies. For example, in Prasada and Pinker's (1993) study, speakers are likely to mimic or invent irregular forms (such as *spling/splung*) which rhyme with existing irregular verbs. Newberry et al. admit that similar selection can also be present in verbs whose dynamics turned out to be dominated by stochastic drift, e.g. *quitted > quit*. Notably, the irregularization trend is also inferred for *wedded > wed*, contradicting the results of Lieberman et al. (2017) which were based on long-term trends. It can additionally be observed that Lieberman et al. did not take into account the synchronic variability of forms, while Newberry et al.'s study is based exclusively on variable past tense forms, i.e. 'lemmas with two past-tense variants that each occurred at least 50 times in the corpus' (p. 223). Given that language is a system in flux, it is also conceivable that, depending on the method, we can make more accurate short-term or long-term predictions, while these predictions are not necessarily expected to be the same.

The large-scale studies focus on stabilization and selection based on existing forms; however they do not address questions which seem crucial from the point of view of linguistic productivity, i.e. in what mental form morphological patterns are learned and how that affects the emergence of novel forms. According to Bybee and Moder (1983: 255), the innovation in forms such as *spling/splung* must be caused by product-oriented rather than source-oriented generalizations, i.e. generalizations based on relations among the pret. forms of different verbs (e.g. *strung, slung, swung, wrung, hung*), rather than those between the base and the derived forms. Thus, the 'propensity for rhyming', mentioned as a potential factor causing irregularization in Newberry et al.'s study, must refer to the generalization that 'the past form for a verb of the *string/strung* type must end in /n/ followed by a nasal or a velar; but the vowel of the base does not necessarily have to be /ɪ/' (Bybee – Moder 1983: 255). Morphological innovations such as *strike/struck* and *sneak/snuck* crucially depend on the availability of mechanisms other than proportional analogy because there are no analogous present – pret. pairs in the English lexicon that could serve as the model for these novel forms. In contrast, source-oriented generalizations can effectively account for the regular English past-tense suffixation (e.g. *walk/walked, google/googled*). Interestingly, Albright and Hayes (2003) argue that irregulars and regulars are not necessarily handled by different mechanisms, but point to

potential advantages of rule- rather than analogy-based mechanisms. Their 'wug test' data on English show that both irregulars and regulars are subject to gradient ratings dependent on the phonological structure of the stem. In their simulations, the analogical model is unable to focus on the relevant fragments of phonological structure, and is also biased towards implausible responses influenced by single exemplars, which bear high resemblance to the test items.

#### **4. Concluding remarks**

The (ir)regularization of English verbs taking place over centuries provides an excellent testing ground for theories of language evolution and mechanisms of change. While frequency has long been considered to be a prominent factor in language change (e.g. in exemplar-based linguistic approaches), it has not been fully understood to what extent factors such as natural selection or random chance are in operation. Large-scale studies based on big datasets and the employment of powerful statistical methods, such as those used in evolutionary biology, allow to address quantitative questions concerning the evolutionary dynamics of language. The quantification methods used in these studies make it possible to estimate the rates of change and to empirically verify the claims made earlier by linguists about the relationship between the propensity for (ir)regularization and lexical frequencies (and/or morpheme shapes), emphasising the role of stochasticity in linguistic innovation. As expected, they also reveal differences in predictions depending on long versus short time scales over which these predictions are made. However, focusing merely on stabilization and selection of existing forms, such studies leave unanswered other questions which seem vital from the point of view of understanding linguistic innovation: in what mental form morphological patterns are learned and how that affects potential paths of linguistic creativity.

#### REFERENCES

##### **Sources**

Google Ngrams

<https://books.google.com/ngrams/>, accessed January 2022.

*Oxford English Dictionary Online (OED online)*

<http://www.oed.com/>, accessed January 2022.

**Special studies**

- Albright, Adam – Bruce Hayes  
 2003 “Rules vs. analogy in English past tenses: A computational/experimental study”, *Cognition* 90, 119–161.
- Anderson, Stephen R.  
 2015 “Morphological change”, *The Routledge Handbook of Historical Linguistics*, 264-285.
- Anderwald, Lieselotte  
 2013 “Natural language change or prescriptive influence?”, *English World-Wide* 34, 146–176.
- Berko, Jean  
 1958 “The child’s learning of English morphology”, *Word* 14, 150-177.
- Burnley, David  
 1983 *A Guide to Chaucer’s Language*. London: The Macmillan Press Ltd
- Bybee, Joan  
 2002 “Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change”, *Language Variation and Change* 14, 261–90.  
 2003 “Mechanisms of change in grammaticization: The role of frequency”. *The Handbook of Historical Linguistics*. Oxford: Blackwell, 602–623.  
 2006 “From usage to grammar: The mind’s response to repetition”, *Language* 82, 711–733.
- Bybee, Joan L. – Carol L. Moder  
 1983 “Morphological classes as natural categories”, *Language* 59, 251-270.
- Bybee, Joan L. – Dan I. Slobin  
 1982 “Rules and schemas in the development and use of the English past tense”, *Language* 58, 265-289.
- Campbell, Lyle  
 1998 *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Dammel, Antje – Jessica Nowak – Mirjam Schmuck  
 2010 “Strong-verb paradigm leveling in four Germanic languages: a category frequency approach”, *Journal of Germanic Linguistics* 22 (4), 337-359.
- De Smet, Isabeau - Freek Van de Velde  
 2020 “A corpus-based quantitative analysis of twelve centuries of preterite and past participle morphology in Dutch”, *Language Variation and Change* 32 (2), 241-265.
- Gleason, Henry A., Jr.  
 1955 *An Introduction to Descriptive Linguistics*. New York: Holt, Rinehart & Winston.

Hare, Mary - Jeffrey L. Elman

1995 "Learning and morphological change", *Cognition* 56, 61–98.

Hooper, Joan

1976 "Word frequency in lexical diffusion and the source of morphophonological change". *Current progress in historical linguistics*. Amsterdam: North-Holland, 96–105.

Jespersen, Otto

1942 *A Modern English Grammar. On Historical Principles. IV: Morphology*. Copenhagen: Munksgaard.

Knooihuizen, Remco – Oscar Strik

2014 "Relative productivity potentials of Dutch verbal inflection patterns", *Folia Linguistica Historica* 35, 173-200.

Krygier, Marcin

1994 *The Disintegration of the English Strong Verb System*. Frankfurt am Main: Peter Lang.

Lieberman, Erez - Jean-Baptiste Michel - Joe Jackson - Tania Tang - Martin A. Nowak

2007 "Quantifying the evolutionary dynamics of language", *Nature* 449, 713-716.

Mitchell, Bruce – Fred C. Robinson

2001 *A Guide to Old English* (6<sup>th</sup> edn.). Oxford: Blackwell.

Mańczak, Witold

1969 *Le développement phonétique des langues romanes et la fréquence*. (Zeszyty naukowe Uniwersytetu Jagiellońskiego, 205. Prace językoznawcze, fasc. 24), Kraków: Nakładem Uniwersytetu Jagiellońskiego.

Newberry, Mitchell G. - Christopher A. Ahern - Robin Clark - Joshua B. Plotkin

2017 "Detecting evolutionary forces in language change", *Nature* 551, 223–226.

Nowak, Jessica

2020 "Ablaut pattern extension as partial regularization strategy in German and Luxembourgish". In: B. Drinka (ed.) *Historical Linguistics 2017: Selected Papers from the 23rd International Conference on Historical Linguistics*. Amsterdam: John Benjamins, 183-204.

Nübling, Damaris

2000 *Prinzipien der Irregularisierung. Eine kontrastive Untersuchung von zehn Verben in zehn germanischen Sprachen*. Tübingen.

Plunkett, Kim - Virginia Marchman

1993 "From rote learning to system building: Acquiring verb morphology in children and connectionist nets", *Cognition* 48, 21-69.

Prasada, Sandeep – Steven Pinker

1993 "Generalisation of regular and irregular morphological patterns", *Language and Cognitive Processes* 8, 1–56.



Rumelhart, David E. – James L. McClelland

- 1986 “On learning the past tenses of English verbs”. In: D. E. Rumelhart – J. L. McClelland – The PDP Research Group (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 2. Cambridge, MA: MIT Press, 216-271.

Wełna, Jerzy

- 1978 *A Diachronic Grammar of English. Part One: Phonology*. Warszawa: Państwowe Wydawnictwo Naukowe.
- 2001 “Suppletion for suppletion, or the replacement of *ēode* by *went* in English”, *Studia Anglica Posnaniensia* 36, 95 – 110.

Wright, Joseph – Elizabeth M. Wright

- 1914 *Old English Grammar* (2<sup>nd</sup> edn.). London, Edinburgh, Glasgow: Humphrey Milford, Oxford University Press

Address: BEATA ŁUKASZEWICZ, Institute of English Studies, University of Warsaw, ul. Dobra 55, 00-312 Warszawa, Poland.

ORCID code: 0000-0001-5817-0625.