How Close Is Close Enough?*

Nikola Datkova
University of Pennsylvania, University of Cambridge

nd546@cam.ac.uk
ndatkova@sas.upenn.edu
no-moore.com

Abstract.

Understanding how the human mind categorises lexical items based on phonological similarity has long been a challenge in linguistic research. Previous studies suggest that phonological distance metrics, such as the Levenshtein Distance, can predict perceived similarity and influence past-tense formation preferences (i.e., following regular or irregular patterns) in the case of English verbs. However, additional research (rightly) argues that there is a lot of other factors—both intra- (e.g., syllable structure) and extra-linguistic (e.g., context)—influencing categorisation. That said, the academic community has neglected to ask the question of: How much predictive power do pure phonological distance metrics have? This study uses experimental data to determine the maximum phonological distance at which the human mind still considers two items as belonging to the same category. The task of forming past tenses for nonce verbs increasingly distant from existing lexemes was used to observe when participants would switch from regular to irregular past-tense formation. The study had trouble replicating results from previous work and patterns for Levenshtein Distance, although present, were insignificant. Other phenomena such as the Uniqueness Effect and high volatility of data with short-distance changes were observed.

*I'm grateful to Prof John Williams (University of Cambridge) for putting a lot of time into overseeing my work as I was making my way through experimental linguistics. Many thanks to Prof Charles Yang (University of Pennsylvania) for recommending relevant readings, Prof Meredith Tamminga (University of Pennsylvania) for giving me a much-needed crash course into experimental linguistics, Prof David Embick (University of Pennsylvania) for making me reconsider my life choices, and Prof Bert Vaux (University of Cambridge) and his former students Ollie and Sammy for looking over and verifying my research methods. Last but by no means least, I would also like to thank Dan for spending way more time than I deserve consulting the materials with me and for taking the time to record the audio prompts (twice!).

Contents

I Introduction & Literature Review

John likes to gleed. Look, there he's gleeding. Everyday he gleeds. Also yesterday he...? Many native speakers would respond 'gled', despite the standard rule that dictates suffixing the regular morpheme *-ed* to form the past tense of the vast majority of verbs in the English language. The intuitive leap to reproducing irregular paradigms on newly encountered, made-up verbs highlights the brain's capacity to recognise significant similarities with established lexical items and use the analogy with existing lexical entries to override endorsed generative structures.

Previous Literature. When conjugating verbs, the brain has essentially two (i.e., two immediately obvious) options: either modify the whole shape of the word in analogy with an existing lexeme (e.g., *gleed* → *gled* in analogy with *bleed* → *bled*) or default to the generative past-tense inflection (e.g., *gleed* → *gleded* by suffixing *-ed*). In the effort to explain the split production of past tense, two main schools of thought emerged—the single- and dual-route models—each analysing the framework through a distinct lens.

The single-route (also known as the 'connectionist') model (e.g., Rumelhart and McClelland, 1986) proposes that the past tense is produced analogically across both regular and irregular examples. Evidence supporting this approach incorporates the learning curve of a neural network (trained purely by exposure to instances of both regular and irregular verbs) copying that of a child learner as well as the network's correct generalisations of patterns to previously unseen verbs (Rumelhart and McClelland, 1986).

The dual-route model (e.g., Prasada and Pinker, 1993), on the other hand, argues that while irregular paradigms may be reproduced by analogy, there must exist a separate mechanism for handling regular verbs through a rule-based process. This model is supported by various pieces of evidence, for instance from neurology: it has been shown that patients with damage to the brain area responsible for memory cannot produce correct irregular past tense yet form suitable regular past tense, while those with damage to the area responsible for grammar produce irregular past tense correctly without being able to form accurate regular past tense (Pinker and Ullman, 2002).

The dual-route model has also been demonstrated in linguistic experiments. For instance, it has been shown that participants systematically reproduce irregular paradigms on nonce words similar to existing irregular lexical entries but default to the generative morphology when the nonce words no longer resembled these entries (Albright and Hayes, 2003). This particularly influential study not only supports the dual-route model in linguistic processing but also naturally invites further conversation regarding the roles of phonological similarity in general.

That phonological distance is relevant to how humans perceive words in relation to one another has been demonstrated through various studies. For example, when tasked to remember and subsequently recall how far from one another two lexemes appeared on a screen, participants tended to recall two phonologically similar lexemes closer compared to phonologically dissimilar pairs, suggesting that the human brain also understands this similarity in terms of space (Tuena et al., 2023).

(The value of phonological neighbourhoods also helps explain several real-life conditions. One such example is monolingual dyslexia, a reading disorder that only shows in one of the speaker's multiple languages. A commonly described case of monolingual dyslexia occurs in English (e.g., Wydell and Butterworth, 1999), a language where the spelling does not reliably represent sound. Consider the example of *though*, *through*, and *thought*, lexemes that all look—but do not sound—similar. This misrepresentation of phonological neighbourhoods in orthography causes delays for English readers but not readers of other languages with more consistently mapping writing systems (e.g., Caravolas et al., 2013).)

The human mind's sensitivity to phonological neighbourhoods is also reflected in language production. In picture-naming tasks, participants took longer to name items with meagre phonological neighbourhoods compared to those with dense phonological neighbourhoods, implying that neighbouring items get co-activated and promote the retrieval of one another (Vitevitch, 2002). Likewise, children with specific speech disorders tend to stutter more frequently on lexemes with few phonological relatives while those from rich phonological neighbourhoods were generally more fluent (Anderson, 2007).

It is hence clear that we, humans, like to relate words to one another based on how they sound. The question remains: how much? An attentive reader may have noticed that many of the above-mentioned studies on how the human mind categorises lexemes already assume 'phonological similarity', but few inquire about what 'phonological similarity' actually is. This is largely due to the fact that linguists have developed various, often very sensible metrics for measuring phonological relatedness.

One such metric is the Levenshtein Distance (Levenshtein, 1965). The Levenshtein Distance is a popular, relatively straightforward algorithm for computing distances between strings. Simply put, it measures the minimum number of steps required to get from string A to string B. For example, the distance from *can* to *cat* would be 1, the necessary step being $n \rightarrow t$. The distance from *can* to *cats* would be 2, the necessary steps being $n \rightarrow t$ and $\varnothing \rightarrow s$. For values more appropriate to scale, the number may also be divided by the initial string's length.

To fully appreciate the [evaluative power] of the Levenshtein Distance, however, it is important to understand where sound per se comes from. Every sound that can possibly come out of the human mouth is a result of 25 separate features that can engage with the sound as it travels through the articulatory system. These include 5 places of articulation (of which typically one and occasionally two are active) and 20 manners of articulation (of which many may be engaged, though different languages have different restrictions on what combinations they allow).

That said, every sound can be represented as a vector of 25 separate values (see Appendix B to view the vectors representing the English consonant inventory). This way, the Levenshtein Distance can be computed based on these individual values rather than the final sound their combination produces. Such an approach appreciates that, for example, the words *pick* and *pig* are mutually closer to each other than any of them is with the word *pin*, as the final voiceless and voiced velar plosives (the /k/ and /g/ sounds respectively have more features in common than any of them shares with the voiced alveolar nasal, /n/).

Of course most modern linguists would also recognise that while classification is certainly something that the human mind needs and tends to put novel items through, there are many factors beyond phonological distance that contribute to how speakers group multiple distinct lexemes together. In other words, while the human mind likes to relate words to one another based on how they sound, it is not the sole factor controlling the categorisation of lexical items. There are several other intra- as well as extra-linguistic influences to this process.

A telling example of an intra-linguistic phenomenon that reflects classification is the number of syllables. In Russian, the last vowel of a stem deletes when a vowel-initial suffix merges with an eligible stem. The applicability of this phonological rule is generally thought to be restricted lexically (i.e., some, typically loaned, words simply will not accept this deletion by definition). However, observations of this process on nonce words show that the native mind prefers to apply this word on di- but not mono-syllabic lexemes. That is, Russian speakers subconsciously classify words to those where the rule applies and those where it does not based on the number of syllables (Gouskova and Becker, 2013).

Another influence comes from the syllable structure. When English speakers were tested on the production of past tense for nonce verbs, they showed a strong preference for irregular forms when presented with nonce verbs with complex onsets. This effect was even stronger in words with complex onsets consisting of the voiceless alveolar fricative (the /s/ sound) followed by two additional consonants. This preference occurred even though this cluster is not by far exclusive to English irregular verbs (e.g., *strive*, *spring*, *strike*) but also commonly occurs in regular verbs (e.g., *stroke*, *scratch*, *scream*) as well as other word classes such as nouns (e.g., *string*, *screen*, *scrap*) (Bybee and Moder, 1983). This shows that speakers categorised novel verbs as regular or irregular based on their syllable complexity.

An example of an extra-linguistic influence on classification is context. When presented with a novel verb similar to two existing lexical entries, English speakers choose the past tense form based on context. For instance, when presented with the novel verb *frink* in the context of an eye-related health condition, English-speaking participants choose *frinked* to be the past tense form in the analogy with *blink*. However, when presented with the same novel verb in the context of therapy for terminally ill patients at the Moscow hospital, the participants leaned towards *frank* instead, in analogy with *drink* (Ramscar, 2002).

Objectives. The above overview outlines the general understanding that phonological neighbourhoods matter in classifying novel lexemes, that this phonological distance is well-measurable, and that it is not the sole predictor of how the human mind categorises lexical entries. However, how big of a say does the phonological distance metric actually have in the process of categorisation? How much power do the phonological distance metrics per se hold? In other words, can it be predicted, using phonological distance metrics only, at what point are two words still considered 'similar'?

The present study aims to explore the brain's sensitivity to phonological distance as a standalone metric. It asks the question: Is phonological distance by itself a good enough predictor of perceived similarity? This study observes the notion of the English past tense to observe the categorisation patterns of the human mind. It searches for a distance threshold where novel verbs stop being categorised as similar to one another, stop triggering the analogically derived irregular past-tense forms, and shift towards the regular, generative forms instead.

II Methodology

Experiment Design Overview. To investigate these inquiries, this study selects 7 verbs (the selection process of which is described in the upcoming section) and creates 4 incrementally altered variants of each (using the Levenshtein Distance, with attention to individual features of each phoneme rather than the whole sound). Changes will be made to consonants in the coda, avoiding any transformations of vowels, as vowels are a key cue to past tense formation in the English language. Given the resources allocated to this study, it was possible to test modifications to either onset or coda but not both. Therefore, this study chose to focus on modifications in the coda to mitigate the effect of rhyming. This approach yields 28 nonce verbs, ready to be presented to the native speakers of English to determine their intuitive choice of past tense (analogical or generative). The generally recommended ratio of 1:1 was applied to the number of experimental and filler items.

Following the informed consent form, a sample of 100 participants was presented with randomised audio prompts presenting each of the experiment and filler items in a sentence of the following structure: *'John likes to* [nonce verb]. *Look, there he is* [nonce verb]*ing. He*[nonce verb]*s ever day. Even yesterday he* [blank].*'* The audio prompts were recorded by a male native speaker of American English. The participants were instructed to type in their intuitive past tense form for each missing item without paying too much attention to the spelling and instead focus on the sound of their responses.

This design was estimated to yield at least 80 valid responses, which could be subsequently plotted as a function of phonological similarity and inclination to regularity. The hope was that this analysis would reveal a critical point where the pattern of responses shifts, indicating a potential threshold in the participants' linguistic intuition or perception related to the constructed verbs' phonological properties and their alignment with regular verb forms. Alternatively, if the shift turned out to be more gradual, the findings would be subject to further analysis in an attempt to understand the nuances.

Preparation of Materials. The preparation of the experimental items for this study consisted of three major steps: (i) narrowing down the pool of real-world irregular verbs to those eligible to serve as a basis for the experimental items, (ii) selecting suitable items based on their syllable structure, and (iii) modification of the selected items in compliance with the English phonotactics (i.e., making sure that the sound clusters of the experimental items are easily pronounceable to an English speaker).

Narrowing Down the Pool of Eligible Items. The nonce verbs used in the present study are inspired by real irregular verbs from the English language. However, irregular verbs in the English language also follow a number of patterns, and not all of these patterns can be particularly helpful for the purposes of this study. In order to determine the most comprehensive selection of items to represent the English irregular landscape, the present study considers analyses proposed in existing literature.

One such analysis can be found in *The Cambridge Grammar of the English Language* (Huddleston and Pullum, 2002, p. 1599–1608). This work, based on a detailed qualitative study of patterns across the English irregular verbs, recognises four classes of irregular verbs outlines in Table 1.

Class 1:	secondary *-ed* formation
Class 2:	vowel alternations
Class 3:	past participles formed with the *-en* suffix
Class 4:	other formations

Table 1: Irregular verb classes according to Huddleston and Pullum, 2002, p. 1599–1608.

These classes are additionally divided into 25 different subclasses listed in Table 2. Luckily enough, not all of these subclasses are of interest to this study. This study specifically focuses on those subclasses that contain a sufficient number of verbs exhibiting observable changes when turned into their past tense form, which is essential for examining the speakers' tendencies towards generative versus analogical strategies in forming the past tense.

That said, this study excludes the following subclasses: (i) classes focusing solely on devoicing or reduction of the final suffix (1A, 1C, 1E), (ii) classes with miscellaneous nucleus modifications (2D, 2E, 3D, 3E, 3F, 3G, 4B), (iii) classes of anomalies with no apparent patterns (3H, 4E), and (iv) classes with too few instances (1G, 2C).

Group (i) was excluded due to high barrier to observation. Devoicing of the final suffix is not unique to irregular verbs; the regular suffix also gets devoiced when preceded by a voiceless consonant (e.g., *liked*, *stopped*, *watched*). Group (ii) was excluded due to the impossibility of telling the past form apart from the null entry. Groups (iii) and (iv) containing mostly exceptions and inconsistent alternations were excluded due to an insufficient number of unique entries for the human mind to form analogies based on them

To control for the potential effect of the number of syllables, all disyllable irregular verbs were also excluded from this analysis (there weren't many). After filtering out these categories, 10 classes (bolded) remain for examination. Verbs and classes (3A, 3B) with no change to the simple past tense form (incl. verbs that do not inflect by tense at all, such as *set* → *set*, and verbs that only inflect irregularly in the participle form, such as *show* → *showed* → *shown*) were also excluded from the set. For the purposes of this study, the remaining eligible classes (1B, 1D, 1F, 2A, 2B, 3C, 4A,

SUBCLASS	EXAMPLE		
1A: devoicing of the suffix	*burn*	*burnt*	*burnt*
1B: vowel shortening	*sleep*	*slept*	*slept*
1C: consonant reduction	*spread*	*spread*	*spread*
1D: vowel shortening with devoicing of the suffix	*mean*	*meant*	*meant*
1E: consonant reduction with devoicing of the suffix	*build*	*built*	*built*
1F: vowel shortening with [final] consonant reduction	*bleed*	*bled*	*bled*
1G: devoicing in the base with vowel shortening	*leave*	*left*	*left*
2A: /ɪ/ ∼ /æ/ ∼ /ʌ/	*drink*	*drank*	*drunk*
2B: /ɪ/ ∼ /ʌ/ ∼ /ʌ/	*win*	*won*	*won*
2C: /aɪ/ ∼ /aʊ/ ∼ /aʊ/	*find*	*found*	*found*
2D: miscellaneous vowel alternations	*get*	*got*	*got*
2E: vowel change in preterite, past participle identical with base	*come*	*came*	*come*
3A: regular preterite, *-en* added to base	*show*	*showed*	*shown*
3B: regular preterite, vowel change in past participle	*swell*	*swelled*	*swollen*
3C: /aɪ/ ∼ /oʊ/ ∼ /ɪ/	*ride*	*rode*	*ridden*
3D: vowel shortening in preterite and past participle	*bite*	*bit*	*bitten*
3E: vowel change in preterite, *-en* added to base	*eat*	*ate*	*eaten*
3F: vowel change in preterite, *-en* added to preterite	*choose*	*chose*	*chosen*
3G: vowel change in both preterite and past participle	*fly*	*flew*	*flown*
3H: suppletive preterites	*go*	*went*	gone
4A: vowel shortening and *-d* suffix	*flee*	*fled*	*fled*
4B: vowel change and *-d* suffix	*hear*	*heard*	*heard*
4C: loss of consonant	*make*	*made*	*made*
4D: forms with *-ought* and *-aught*	*seek*	*sought*	*sought*
4E: preterites of the modals	*can*	*could*	

Table 2: Detailed subclasses of English irregular verbs identified in Huddleston & Pullum, 2017.

4D) can be further organised into the following five distinct patterns outlined in Table 3. Given

PATTERN 1:	vowel shortening
PATTERN 2:	/ɪ/ ∼ /æ/
PATTERN 3:	/ɪ/ ∼ /ʌ/
PATTERN 4:	/aɪ/ ∼ /aʊ/
PATTERN 5:	forms with *-ought* and *-aught*

Table 3: Patterns of irregular verbs in English described for the purposes of this study.

that this study aims to explore the brain's ability to differentiate varying degrees of phonological similarity, it is best that verbs selected to represent the experimental patterns are as dissimilar as possible from all other verbs in the English lexicon (especially from more commonly used and regular entries). This way, the observation of the similarity threshold between analogical and generative past tense will yield clearer insights into the processing and categorisation of linguistic structures.

That said, this study considered every irregular verb in the English language that subscribes to the above-mentioned patterns and, with the assistance of a few native speakers' intuitions, determined its similarity to other English verbs that emerges from modifying any segment of this verb. More details on the identified similarities can be found in Appendix A. To determine these similarities, every consonant phoneme in each of the target verbs was permuted with every other phoneme in the English phonemic inventory, as described in Appendix B.

Selecting Suitable Items Based on Syllable Structure. After narrowing down the pool of irregular verbs of interest, it was time to select the representative verbs that would consequently be gradually modified and included in the experiment. It was previously shown (Bybee and Moder, 1983) that the syllable structure of the entire word (i.e., the complexity of its onset and coda) has a significant priming effect on the past tense formation strategy speakers lean towards. The English irregular verbs split into five types (outlined in Table 4) based on the syllable being modified. Irregular verbs previously identified as eligible to be tested in this study were divided into these five types (see Appendix C for the complete split). For each type outlined in Table 4, an eligible

Type 1:	CVC	*(simple onset and coda)*
Type 2:	CCVC	*(complex onset, simple coda)*
Type 3:	CCCVC	*(super complex onset, simple coda)*
Type 4:	CVCC	*(simple onset, complex coda)*
Type 5:	CCVCC	*(complex onset and coda)*

Table 4: Categories of English irregular verbs based on the complexity of onset and coda.

verb with the least amount of similar entries (based on Appendix A) after the modification of the various consonants in the coda (based on Appendix B) was selected. (Note that every available entry in Type 4 and Type 5 has a number of similar entries after the modifications in the onset, which makes these types particularly hard to test).

Ultimately, five representative base verbs were selected, all outlined in Table 5. **Modification**

Type 1:	shoot
Type 2:	dream
Type 3:	spring
Type 4:	sink
Type 5:	drink

Table 5: Final selection of experimental items.

of Selected Items. Each of the seven consonants in the coda positions (3 consonants in simple codas and 4 consonants in complex codas) was modified to four different degrees based on the Levenshtein Distance. Given that only one phoneme was being modified at a time, this distance was calculated based on the number of changes in the phonemic features of each individual phoneme rather than the phonemes as a whole.

This can be demonstrated on a simple example. Consider the lexeme *shoot*. This lexeme has a simple coda, meaning that only the phoneme /t/ is to be modified. For the first stage of modification, the best candidates for replacing see are phonemes relatively closely related to /t/ (i.e., phonemes whose vector is identical to that of /t/ bar a few changes; see a collection of the vectors of the English consonants in Appendix B). One suitable phoneme is /s/. The phonemes

/s/ and /t/ can both be represented by their own 25-dimensional vectors (see Appendix B). These matrices are identical except for two values, one representing continuity (airflow throughout the continuity of the sound) and the other representing stridency (distinguishing mostly sibilants such as /s/ and /z/ from non-sibilants such as /t/ or /p/). /s/ has both of these features while in /t/ they are absent:

$$/\mathrm{s}/ = \begin{bmatrix} \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix}$$

$$/\mathrm{t}/ = \begin{bmatrix} \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \end{bmatrix}$$

In order to get from /t/ to /s/, one needs to exchange these two features. Therefore, the minimum number of steps for the /t/ → /s/ operation is 2. Given that the word *shoot* (as well as the intended *shoos*) has three sounds (/ʃ/ + /uː/ + /t/ = /ʃuːs/), the total length of this string is 75 (3 sounds × 25 features). In this example, a modification to two of these 75 features was made, resulting in the Levenshtein Distance of $\frac{2}{75} = \frac{\text{changes}}{\text{length of string}} = 0.02\overline{6}$.

Although the Levenshtein Distance is typically normalised for the length of a complete string, this study normalises it for the 'length' of the phoneme vector, which is always constant at 25. This means that every feature that changes in this phoneme increases the Levenshtein Distance by a constant value of 0.04. Changing all features in the phoneme would result in the Levenshtein Distance of 1 (note that this never happens, as consonants in English (and any other language) tend to share many features with each other). Nonetheless, these numbers make it a little easier to see the differences relative to each other, disregarding the length of the experimental item in general. Note that this method also does not account for the costs of this change such as changes to surrounding phonemes that might come up as a result of assimilation.

With the above-described process in mind, all selected lexemes were modified to four various extents, depending on the availability of consonants and string segments in accordance with the rules of the English phonotactics (i.e., pronounceable to English native speakers). This is also why the differences between the modifications of individual items are not identical. All final experimental nonce words are outlined in Table 6 (with their respective Levenshtein Distance in parentheses). For every experimental nonce word outlined in Table 6, the data set contained an additional filler item to eliminate priming. All filler items can be found in Appendix D.

ʃuː**t**	ʃuː**s**	($\Delta = 0.08$)	ʃuː**n**	($\Delta = 0.2$)	ʃuː**b**	($\Delta = 0.24$)	ʃuː**g**	($\Delta = 0.32$)
dɹiː**m**	dɹiː**b**	($\Delta = 0.16$)	dɹiː**n**	($\Delta = 0.2$)	dɹiː**f**	($\Delta = 0.32$)	dɹiː**s**	($\Delta = 0.44$)
spɹɪ**ŋ**	spɹɪ**g**	($\Delta = 0.12$)	spɹɪ**m**	($\Delta = 0.28$)	spɹɪ**f**	($\Delta = 0.48$)	spɹɪ**ʃ**	($\Delta = 0.52$)
sɪn**k**	sɪn**p**	($\Delta = 0.16$)	sɪn**θ**	($\Delta = 0.32$)	sɪn**z**	($\Delta = 0.4$)	sɪn**d͡ʒ**	($\Delta = 0.52$)
sɪ**n**k	sɪ**l**k	($\Delta = 0.12$)	sɪ**m**k	($\Delta = 0.2$)	sɪ**s**k	($\Delta = 0.24$)	sɪ**f**k	($\Delta = 0.44$)
drin**k**	drin**p**	($\Delta = 0.16$)	drin**t**	($\Delta = 0.2$)	drin**z**	($\Delta = 0.4$)	drin**d͡ʒ**	($\Delta = 0.52$)
dri**n**k	dri**l**k	($\Delta = 0.12$)	dri**m**k	($\Delta = 0.2$)	dri**s**k	($\Delta = 0.24$)	dri**f**k	($\Delta = 0.44$)

Table 6: Final list of experimental items.

Execution and Participants. The interface of the experiment was prepared in PCIbex (see Appendix E for the full code) and designed to be executed online. The design included: an introduction to the study, informed consent, a practice round, a randomised stream of fifty-six rounds of experimental and filler items embedded in the testing context (*'John likes to* [NONCE VERB]. *Look, there he is* [NONCE VERB]*ing. He* [NONCE VERB]*s ever day. Even yesterday he* [BLANK]*.'*), and a final redirection to a confirmation site securing the participants' reward once their responses have been recorded.

The participants were recruited using the platform Prolific and paid the recommended rate of £9 per hour. The sample included 100 native speakers of American and British English. Participants were tested in batches of 20 (rather than the full group at once) at various times of day to promote adequate variability within the group that might otherwise be obstructed due to time differences. There was no time pressure, although the participants were advised to enter the first, most intuitive form that came to mind.

All recorded responses were initially stored in PCIbex Farm and subsequently exported as a CSV file to the local disk.

III Data Analysis and Results

Out of 100 tested participants, 93 successfully completed the study. All the data consisting of 5 208 items (93 responses × 56 items) has been cleansed, devoid of noise (incl. fillers), and prepared for further analysis. The 2 604 relevant items (93 responses × 28 experimental items) have been split into 28 sets depending on the respective experimental items they belonged to. This way, all 93 participants' past forms of each of the 28 individual experimental items were clearly visible.

The Strict Analysis. In every set, each unique entry received a unique label. At this stage, non-contributory data, such as responses where participants evidently misheard the presented items and any missing responses, was excluded. This is why in the upcoming figures, not all sets of responses add up to 93 (or even the same number). The remaining data was systematically aggregated and are presented in separate tables for each experimental item. Tables 7 through 34 show the variety of recorded answers. Individual items are ordered by frequency, with the target item (i.e., the item that follows the exemplar verb's original pattern) being highlighted in blue.

Note that forms of experimental items featuring the short near-close near-front unrounded vowel (the /ɪ/ sound, as in *drink*) accepted as irregular were both those with a near-open front unrounded vowel (the /æ/ sound, as in *drank*) and those with an open-mid back unrounded vowel (the /ʌ/ sound, as in *drunk*). This is because despite the efforts of prescriptive grammars the ɪ → ʌ remains a frequently applied vowel alternation in everyday speech (e.g., Anderwald, 2011).

ENTRY	#	%
shoosed	53	79.10
shussed	6	8.96
shosed	2	2.99
no change	2	2.99
shased	1	1.49
shose	1	1.49
shas	1	1.49
shossed	1	1.49
shoss	0	0

Table 7: *shoos*

ENTRY	#	%
shooned	65	77.38
shone	6	7.14
no change	5	5.95
shunned	2	2.25
shan	2	2.25
shoont	2	2.25
shoned	1	1.12
shaned	1	1.12
shon	0	0

Table 8: *shoon*

ENTRY	#	%
shoobed	64	77.11
shubbed	5	6.02
shobed	4	4.82
no change	2	2.41
shobbed	2	2.41
shobe	2	2.41
shabe	1	1.2
shabed	1	1.2
shubt	1	1.2
shob	1	1.2

Table 9: *shoob*

ENTRY	#	%
shooged	45	71.43
shugged	8	12.7
shoge	3	4.76
shug	3	4.76
shoged	2	3.17
shage	1	1.59
shot	1	1.59
shog	0	

Table 10: *shoog*

ENTRY	#	%
dreabed	55	66.27
dreamt	6	7.23
drebbed	5	6.02
no change	3	3.61
drobe	3	3.61
dreamed	2	2.41
drubbed	2	2.41
dreebt	2	2.41
drebt	1	1.2
drabed	1	1.2
dribed	1	1.2
drabbed	1	1.2
drobed	1	1.2

Table 11: *dreeb*

ENTRY	#	%
dreaned	55	63.22
dreamt	8	9.2
dreamed	7	8.05
no change	3	3.45
drenned	2	2.3
drun	2	2.3
droned	2	2.3
dreant	1	1.15
draned	1	1.15
dran	1	1.15
dranpt	1	1.15
drone	1	1.15
drane	1	1.15
drined	1	1.15
dreenated	1	1.15

Table 12: *drean*

ENTRY	#	%
dreafed	54	64.29
dreft	6	7.14
driffed	4	4.76
drifed	3	3.57
drofe	3	3.57
no change	3	3.57
dreafted	2	2.38
dreffed	2	2.38
drafed	2	2.38
drofed	2	2.38
droof	1	1.19
drafe	1	1.19
drufed	1	1.19

Table 13: *dreaf*

ENTRY	#	%
dreased	66	79.52
no change	3	3.61
drose	3	3.61
drused	2	2.41
drised	2	2.41
drosed	2	2.41
drased	1	1.2
druss	1	1.2
drissed	1	1.2
drase	1	1.2
dressed	1	1.2

Table 14: *dreas*

ENTRY	#	%
sprigged	53	62.35
spriged	10	11.76
sprug	7	8.24
spragged	5	5.88
no change	3	3.53
sprag	3	3.53
sprugged	2	2.35
sproge	1	1.18
spruged	1	1.18

Table 15: *sprig*

ENTRY	#	%
sprimmed	34	41.98
sprimed	22	27.16
spram	6	7.41
sprung	4	4.94
sprummed	3	3.7
sprum	3	3.7
sprim	3	3.7
sprummed	1	1.23
sprammed	1	1.23
sprome	1	1.23
spremed	1	1.23
spromed	1	1.23
sprom	1	1.23

Table 16: *sprim*

ENTRY	#	%
sprished	72	81.82
no change	4	4.55
sprashed	4	4.55
sprosh	3	3.41
sproshed	2	2.27
sprash	2	2.27
sprushed	1	1.14
sprush	0	0

Table 17: *sprish*

ENTRY	#	%
spriffed	58	70.73
sprifted	6	7.32
sprifed	4	4.88
spraffed	4	4.88
no change	3	3.66
sprift	2	2.44
sprave	1	1.22
spruffed	1	1.22
spruff	1	1.22
spruft	1	1.22
sprove	1	1.22
spraff	0	0

Table 18: *spriff*

ENTRY	#	%
sinped	74	87.06
sinp	3	3.53
sunped	2	2.35
sanp	1	1.18
sunp	1	1.18
samped	1	1.18
sinpted	1	1.18
semped	1	1.18
simpered	1	1.18

Table 19: *sinp*

ENTRY	#	%
sinthed	54	88.52
santh	4	6.56
no change	2	3.28
sonthed	1	1.64
sunth	0	0

Table 20: *sinth*

ENTRY	#	%
sinzed	42	87.5
no change	2	4.17
sanzed	2	4.17
sunzed	2	4.17
sanz	0	0
sunz	0	0

Table 21: *sinz*

ENTRY	#	%
sinched	81	91.01
no change	3	3.37
sanched	2	2.25
sunch	2	2.25
sanch	1	1.12

Table 22: *sinch*

ENTRY	#	%
silked	80	90.91
salk	3	3.41
sulk	2	2.27
no change	2	2.27
sulked	1	1.14

Table 23: *silk*

ENTRY	#	%
sinked	23	28.75
simked	19	23.75
sunk	18	22.5
sank	14	17.5
no change	1	1.25
samk	1	1.25
sumk	1	1.25
samked	1	1.25
simkt	1	1.25
somk	1	1.25

Table 24: *simk*

ENTRY	#	%
sisked	7	89.68
no change	3	3.45
susked	2	2.3
sask	2	2.3
susk	1	1.15
sosk	1	1.15

Table 25: *sisk*

ENTRY	#	%
sifked	65	86.67
no change	3	0.04
sifkt	2	2.67
safk	2	2.67
sufk	1	1.33
sove	1	1.33
safked	1	1.33

Table 26: *sifk*

ENTRY	#	%
drimped	67	80.72
drumpt	4	9.64
dramped	4	9.64
no change	3	3.61
drump	1	1.2
drank	1	1.2
drempt	1	1.2
drumped	1	1.2
drimpt	1	1.2

Table 27: *drinp*

ENTRY	#	%
drinted	62	72.09
drant	9	10.47
drunt	7	8.14
no change	5	5.81
drank	3	3.49

Table 28: *drint*

ENTRY	#	%
drinzed	67	82.72
dranzed	4	4.94
dranz	3	3.7
drunz	3	3.7
no change	2	2.47
drinzt	1	1.23
drinzinated	1	1.23

Table 29: *drinz*

ENTRY	#	%
dringed	73	82.02
drange	5	5.62
drunged	5	5.62
no change	3	3.37
dranged	2	2.25
dronged	1	1.12
drunge	0	0

Table 30: *dringe*

ENTRY	#	%
drilked	68	80
dralk	4	4.71
drulk	3	3.53
drank	3	3.53
no change	2	2.35
drinked	2	2.35
drunk	1	1.18
drulked	1	1.18
drolk	1	1.18

Table 31: *drilk*

ENTRY	#	%
drimked	25	31.25
drank	23	28.75
dramk	12	15
drunk	9	11.25
drinked	5	6.25
no change	3	3.75
drumk	2	2.5
drimkt	1	1.25

Table 32: *drimk*

ENTRY	#	%
drisked	65	75.58
drask	5	5.81
no change	4	4.65
drusked	4	4.65
drusk	3	3.49
drosk	2	2.33
draskt	1	1.16
drosked	1	1.16
dresk	1	1.16

Table 33: *drisk*

ENTRY	#	#
drifked	62	84.93
no change	3	4.11
drafk	2	2.74
drifkt	2	2.74
drufked	1	1.37
drofkt	1	1.37
drafkt	1	1.37
drifkted	1	1.37
drufk	0	0

Table 34: *drifk*

The data was subsequently grouped into larger categories based on the past-tense-formation patterns relevant to this study. The following six most prominent patterns were identified: **(i) regular**, **(ii) irregular expected**, **(iii) irregular other**, **(iv) double past (vowel change + regular suffix)**, **(v) no change**, **(vi) past tense of the original item**, and **(vii) other**.

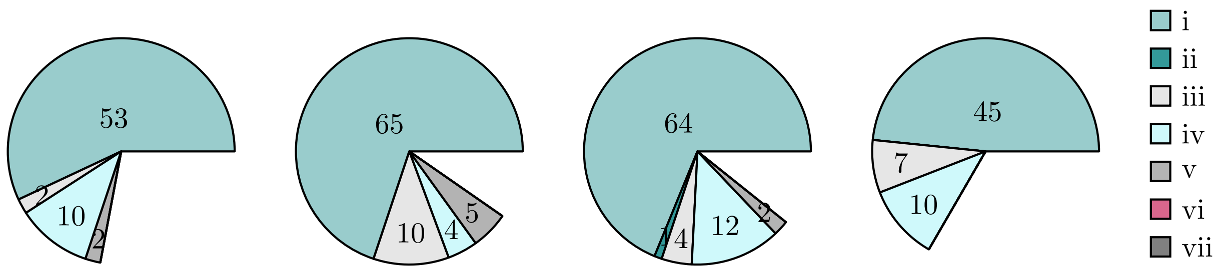

Figure 1: *shoos*, *shoon*, *shoob*, *shoog*

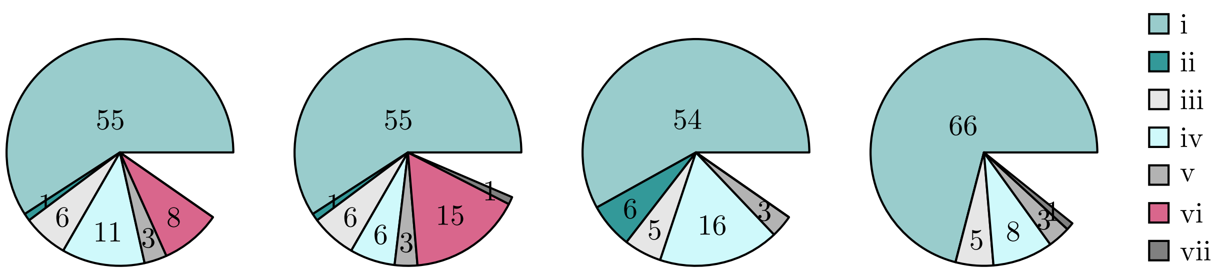

Figure 2: *dreeb*, *drean*, *dreaf*, *dreas*

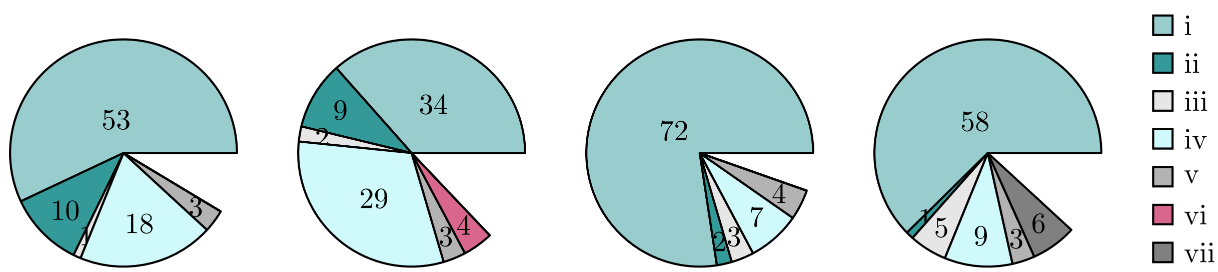

Figure 3: *sprig*, *sprim*, *sprish*, *spriff*

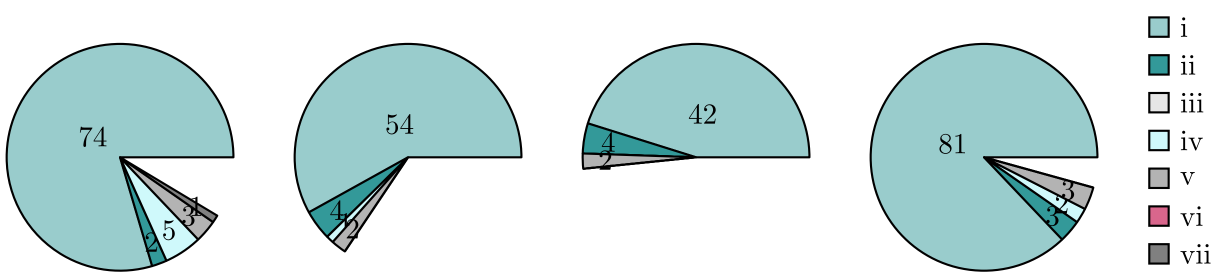

Figure 4: *sinp*, *sinth*, *sinz*, *sinch*

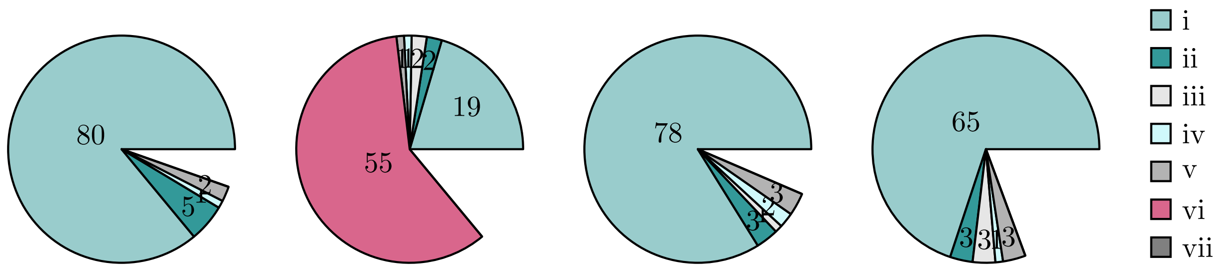

Figure 5: *silk*, *simk*, *sisk*, *sifk*

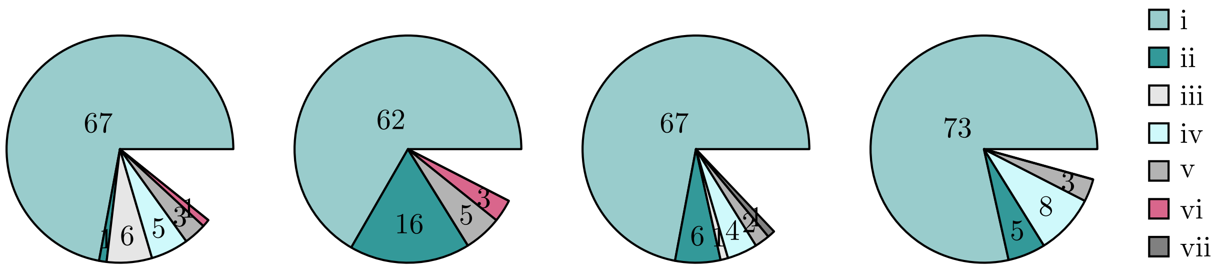

Figure 6: *drinp*, *drint*, *drinz*, *dringe*

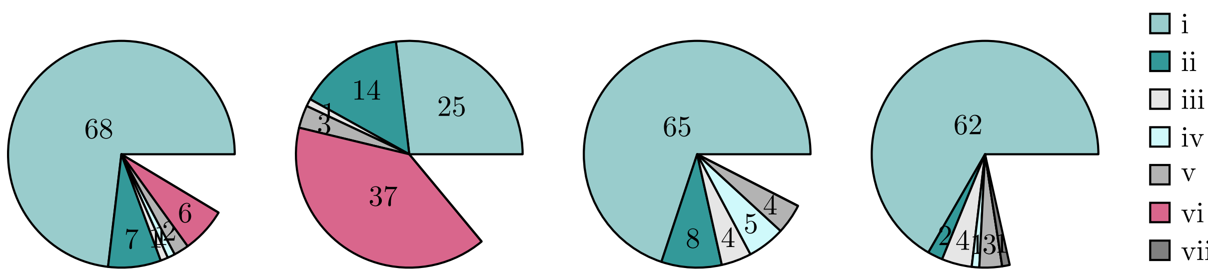

Figure 7: *drilk*, *drimk*, *drisk*, *drifk*

From Figures 1 through 7 it is also immediately obvious that this experiment had significant difficulties prompting the participants to produce any irregular past tense at all—with a few exceptions, the overwhelming majority of responses to the experimental items were regular. Out of the 2 604 valid responses, only 116 (4.45%) were actually irregular by analogy with the intended existing lexical entry. (Just for comparison, there were on average more irrelevant responses than expected irregular responses, cf. 359 (13.79%).) In fact, the only instances where the regular form did not constitute the greatest proportion of responses was when the participants mistook the experimental item for the original irregular verb (in *sprim*, *simk*, and *drimk*).

One possible contributing factor to this result is the fact that the participants took on average 13 minutes and 19 seconds to complete the experiment (see Figure 8 for more detail on the overall completion times), with over a quarter of the responses (red •) being submitted within around or significantly under 10 minutes. However, the assigned audio prompts alone add up to 8 minutes and 12 seconds of listening time, leaving it unclear whether or not the time the participants spent on the experiment was sufficient for careful consideration of the tasks.

On a pace this rushed, the responses may have been skewed by the Chinese Room effect. The Chinese Room is a thought experiment that prompts one to consider a situation where an English speaker who does not speak Chinese gets seated in a room with a thick book of instructions. Over time, this person may become very fast and efficient at transcribing Chinese, but most would argue that reapplying prescriptive rules does not equal the linguistic intuitions a true speaker would demonstrate. Similarly, in this case, the outcomes might have been influenced by a generally advertised rule instead of the intended gut intuition of the participants.

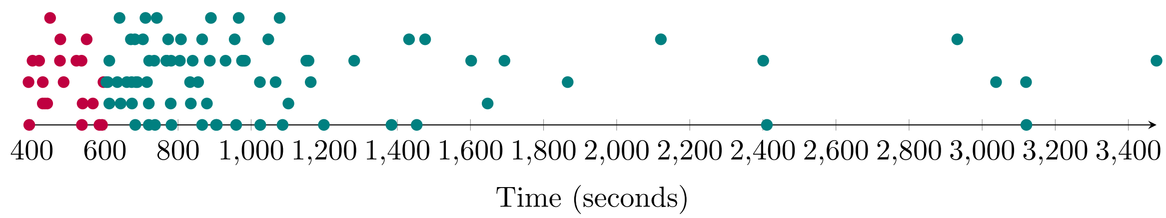

Figure 8: Time taken by participants to complete the experiment

A More Lenient Analysis. An attentive reader may want to remind the author that the participants were instructed to not worry about the spelling of their responses, and this study thus has to be a little more liberal in what it considers a response of interest. With that in mind, the above data was re-grouped to allow entries that would have been previously overlooked and dismissed to receive a more detailed selection of labels. More specifically, these include: (i) the double past of the expected form (i.e., the predicted vowel modification and an extra *-ed* suffixed), and (ii) possible typos (e.g., single rather than double consonant, lenient vowel reading). Tables 35 through 62 show the former highlighted in green and the latter highlighted in pink.

Figures 9 through 15 show how adding each of the above increases the number of target-like responses. Notice that this analysis mostly does not change but rather magnifies the results already showcased by the strict analysis. In two cases, the lenient analysis results in an off-chart peak after the addition of misspellings. This is the case of the $n \to m$ change in front of the voiceless velar plosive (the /k/ sound), where speakers often assumed that the instance they heard was the original base item, probably as a result of assimilation.

Figures 16 and 17 show the slopes formed by the extremes of each analysis separately. Note that not only do the values shift from the majority of data points in the 0-5 range by the strict analysis to the majority of data points in the 5-10 range by the lenient analysis, but also the slope of the lenient analysis is quite a bit steeper than that of the strict one (cf. $m = -8.172$ and $m = -3.364$ respectively).

ENTRY	#	%
shoosed	53	79.10
shussed	6	8.96
shosed	2	2.99
no change	2	2.99
shased	1	1.49
shose	1	1.49
shas	1	1.49
shossed	1	1.49
shoss	0	0

Table 35: *shoos*

ENTRY	#	%
shooned	65	77.38
shone	6	7.14
no change	5	5.95
shunned	2	2.25
shan	2	2.25
shoont	2	2.25
shoned	1	1.12
shaned	1	1.12
shon	0	0

Table 36: *shoon*

ENTRY	#	%
shoobed	64	77.11
shubbed	5	6.02
shobed	4	4.82
no change	2	2.41
shobbed	2	2.41
shobe	2	2.41
shabe	1	1.2
shabed	1	1.2
shubt	1	1.2
shob	1	1.2

Table 37: *shoob*

ENTRY	#	%
shooged	45	71.43
shugged	8	12.7
shoge	3	4.76
shug	3	4.76
shoged	2	3.17
shage	1	1.59
shot	1	1.59
shog	0	

Table 38: *shoog*

ENTRY	#	%
dreabed	55	66.27
dreamt	6	7.23
drebbed	5	6.02
no change	3	3.61
drobe	3	3.61
dreamed	2	2.41
drubbed	2	2.41
dreebt	2	2.41
drebt	1	1.2
drabed	1	1.2
dribed	1	1.2
drabbed	1	1.2
drobed	1	1.2

Table 39: *dreeb*

ENTRY	#	%
dreaned	55	63.22
dreamt	8	9.2
dreamed	7	8.05
no change	3	3.45
drenned	2	2.3
drun	2	2.3
droned	2	2.3
dreant	1	1.15
draned	1	1.15
dran	1	1.15
dranpt	1	1.15
drone	1	1.15
drane	1	1.15
drined	1	1.15
dreenated	1	1.15

Table 40: *drean*

ENTRY	#	%
dreafed	54	64.29
dreft	6	7.14
driffed	4	4.76
drifed	3	3.57
drofe	3	3.57
no change	3	3.57
dreafted	2	2.38
dreffed	2	2.38
drafed	2	2.38
drofed	2	2.38
droof	1	1.19
drafe	1	1.19
drufed	1	1.19

Table 41: *dreaf*

ENTRY	#	%
dreased	66	79.52
no change	3	3.61
drose	3	3.61
drused	2	2.41
drised	2	2.41
drosed	2	2.41
drased	1	1.2
druss	1	1.2
drissed	1	1.2
drase	1	1.2
dressed	1	1.2
dresst	0	0

Table 42: *dreas*

ENTRY	#	%
sprigged	53	62.35
spriged	10	11.76
sprug	7	8.24
spragged	5	5.88
no change	3	3.53
sprag	3	3.53
sprugged	2	2.35
sproge	1	1.18
spruged	1	1.18

Table 43: *sprig*

ENTRY	#	%
sprimmed	34	41.98
sprimed	22	27.16
spram	6	7.41
sprung	4	4.94
sprummed	4	4.94
sprum	3	3.7
sprim	3	3.7
sprammed	1	1.23
sprome	1	1.23
spremed	1	1.23
spromed	1	1.23
sprom	1	1.23

Table 44: *sprim*

ENTRY	#	%
sprished	72	81.82
no change	4	4.55
sprashed	4	4.55
sprosh	3	3.41
sproshed	2	2.27
sprash	2	2.27
sprushed	1	1.14
sprush	0	0

Table 45: *sprish*

ENTRY	#	%
spriffed	58	70.73
sprifted	6	7.32
sprifed	4	4.88
spraffed	4	4.88
no change	3	3.66
sprift	2	2.44
sprave	1	1.22
spruffed	1	1.22
spruff	1	1.22
spruft	1	1.22
sprove	1	1.22
spraff	0	0

Table 46: *spriff*

ENTRY	#	%
sinped	74	87.06
sinp	3	3.53
sunped	2	2.35
sanp	1	1.18
sunp	1	1.18
samped	1	1.18
sinpted	1	1.18
semped	1	1.18
simpered	1	1.18

Table 47: *sinp*

ENTRY	#	%
sinthed	54	88.52
santh	4	6.56
no change	2	3.28
sonthed	1	1.64
sunth	0	0

Table 48: *sinth*

ENTRY	#	%
sinzed	42	87.5
no change	2	4.17
sanzed	2	4.17
sunzed	2	4.17
sanz	0	0
sunz	0	0

Table 49: *sinz*

ENTRY	#	%
sinched	81	91.01
no change	3	3.37
sanched	2	2.25
sunch	2	2.25
sanch	1	1.12

Table 50: *sinch*

ENTRY	#	%
silked	80	90.91
salk	3	3.41
sulk	2	2.27
no change	2	2.27
sulked	1	1.14

Table 51: *silk*

ENTRY	#	%
sinked	23	28.75
simked	19	23.75
sunk	18	22.5
sank	14	17.5
no change	1	1.25
samk	1	1.25
sumk	1	1.25
samked	1	1.25
simkt	1	1.25
somk	1	1.25

Table 52: *simk*

ENTRY	#	%
sisked	7	89.68
no change	3	3.45
susked	2	2.3
sask	2	2.3
susk	1	1.15
sosk	1	1.15

Table 53: *sisk*

ENTRY	#	%
sifked	65	86.67
no change	3	0.04
sifkt	2	2.67
safk	2	2.67
sufk	1	1.33
sove	1	1.33
safked	1	1.33

Table 54: *sifk*

ENTRY	#	%
drimped	67	80.72
drumpt	4	9.64
dramped	4	9.64
no change	3	3.61
drump	1	1.2
drank	1	1.2
drempt	1	1.2
drumped	1	1.2
drimpt	1	1.2

Table 55: *drinp*

ENTRY	#	%
drinted	62	72.09
drant	9	10.47
drunt	7	8.14
no change	5	5.81
drank	3	3.49

Table 56: *drint*

ENTRY	#	%
drinzed	67	82.72
dranzed	4	4.94
dranz	3	3.7
drunz	3	3.7
no change	2	2.47
drinzt	1	1.23
drinzinated	1	1.23

Table 57: *drinz*

ENTRY	#	%
dringed	73	82.02
drange	5	5.62
drunged	5	5.62
no change	3	3.37
dranged	2	2.25
dronged	1	1.12
drunge	0	0

Table 58: *dringe*

ENTRY	#	%
drilked	68	80
dralk	4	4.71
drulk	3	3.53
drank	3	3.53
no change	2	2.35
drinked	2	2.35
drunk	1	1.18
drulked	1	1.18
drolk	1	1.18

Table 59: *drilk*

ENTRY	#	%
drimked	25	31.25
drank	23	28.75
dramk	12	15
drunk	9	11.25
drinked	5	6.25
no change	3	3.75
drumk	2	2.5
drimkt	1	1.25

Table 60: *drimk*

ENTRY	#	%
drisked	65	75.58
drask	5	5.81
no change	4	4.65
drusked	4	4.65
drusk	3	3.49
drosk	2	2.33
draskt	1	1.16
drosked	1	1.16
dresk	1	1.16

Table 61: *drisk*

ENTRY	#	#
drifked	62	84.93
no change	3	4.11
drafk	2	2.74
drifkt	2	2.74
drufked	1	1.37
drofkt	1	1.37
drafkt	1	1.37
drifkted	1	1.37
drufk	0	0

Table 62: *drifk*

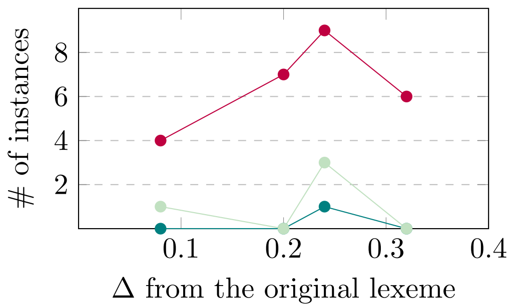

Figure 9: *shoo_*

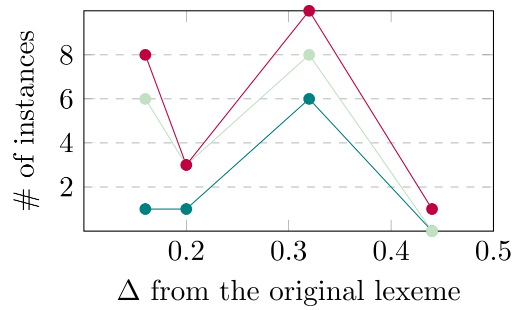

Figure 10: *drea_*

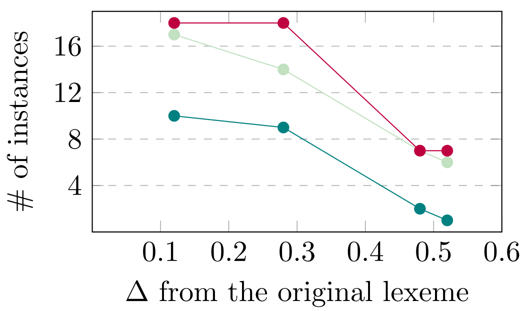

Figure 11: *spri_*

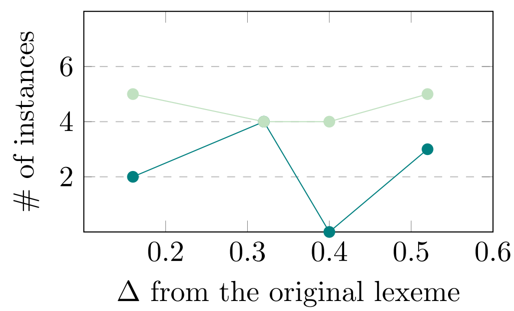

Figure 12: *sin_*

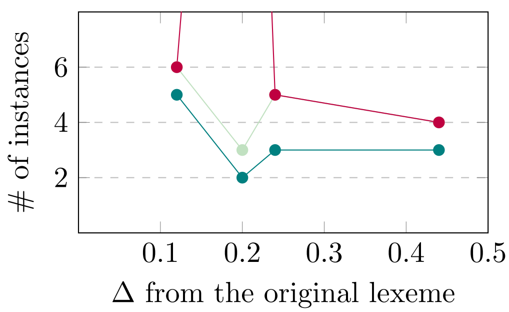

Figure 13: *si_k*

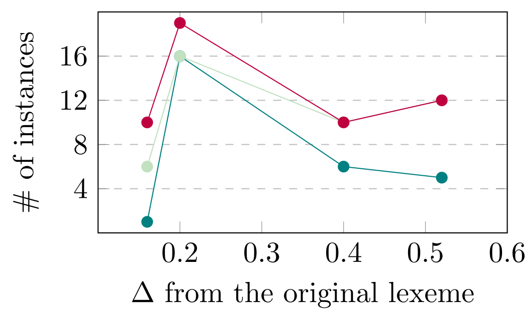

Figure 14: *drin_*

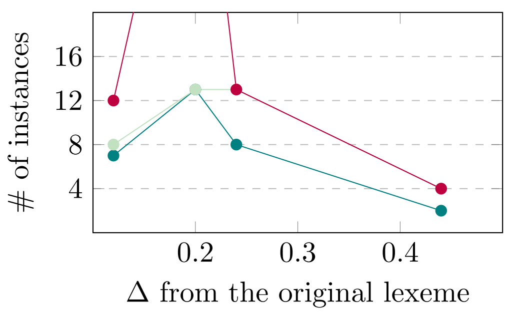

Figure 15: *dri_k*

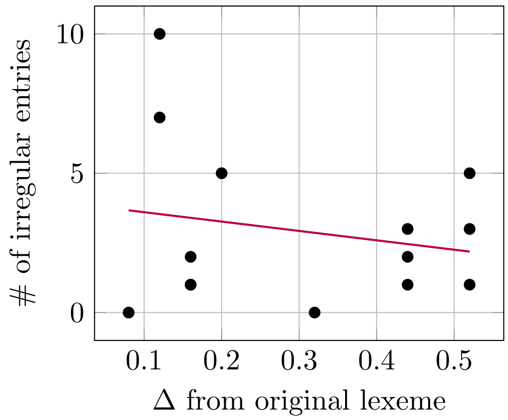

Figure 16: Decreasing preference for regular forms (the strict analysis).

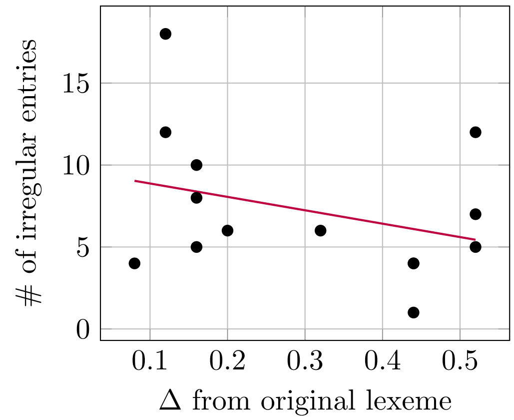

Figure 17: Decreasing preference for irregular forms (the lenient analysis).

IV Discussion

Despite the difficulties this study had with obtaining any irregular data from the participants at all, there nonetheless emerged several noteworthy discoveries—both in response to the existing literature as well as novel contributions.

Reproducing Previous Findings. The first relevant observation to mention is that this study did manage to reproduce results from previous literature on the past tense where increased distance from the initial lexeme resulted in more frequent use of regular past-tense forms. Figure 18 shows the number of regular entries for the extremes (i.e., least and most modified) of each experimental item and models an increasing trend line ($y = 14.70x + 59.88$).

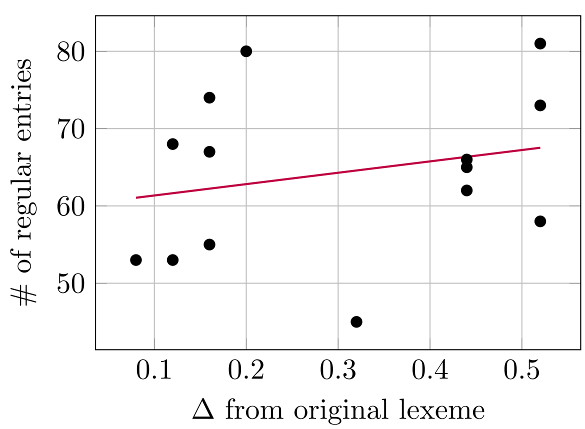

Figure 18: Increased preference for regular past-tense forms.

However, as Tables 7 to 34 show, the remaining, non-regular entries were not at all the expected irregular forms derived from the base lexeme. In other words, although participants did indeed prefer to use regular past-tense forms for modifications further away from the base lexeme, it is not necessarily due to analogising the closer modifications with the base lexeme per se. This is in direct contradiction with either results or fundamental assumptions of several previous findings (e.g., Albright and Hayes, 2003, Ramscar, 2002). This brings up an obvious question: If speakers do not follow the expected analogies, what is then happening?

One possibility is the previously proposed hypothesis that the complexity of the onset is a strong predictor of the rate of irregular intuitions (Bybee and Moder, 1983). However, the exact contrary can be observed at two points of the above data. First, the series with the most complex onset as a whole (Figure 3) does not show consistent outcomes. In other words, while *sprig* yields a relatively high irregular rate, *spriff* does not at all. This shows that the complexity of the onset itself does not suffice.

Second, even though the series with the most complex onset shows on average quite a few instances of the expected irregular pattern for the standards of the present study (6.55%), the *drin_* series (Figure 6) with a less complex onset had even more (8.26%). Series with the simplest possible onset, the *sin_* and *si_k* series (Figures 4 and 5 respectively) showed dangerously similar rates as well (4.59% and 3.94%). This means that not only is the complexity of the onset not predictive by itself, but it also has no significant predictive power in comparison to other types of onsets.

Another possibility proposed in previous research suggests that codaic nasals significantly increase the probability of irregular past-tense formation (Bybee and Moder, 1983) but the present data does not reflect this at all. Two pieces of evidence support the contrary. First, experimental items ending in a nasal (*shoon* and *drean*) would have yielded higher rates of irregular outcomes than their counterparts (see Figures 1 and 2). Given that these items had few to no irregular responses overall, it is clear that this prediction was not supported by the data.

Second, if nasals were that significant in predicting the irregular past-tense formation, one would expect significant differences between two sets based on the same existing lexeme one of which modifies the codaic nasal while the other makes a change to the other sound in the complex coda. In the present data, such pairs can be observed twice, in the *sin_* and *si_k* series (Figures 4 and 5) as well as in the *drin_* and *dri_k* series (Figures 6 and 7). Instances which make a modification to the nasal show similar outcomes to those that modify the other sound, showing that the nasal might not be that crucial after all.

Original Observations. The Uniqueness Effect. While the irregular entries were rather sporadic and not quite as pronounced as expected, there are several other intriguing phenomena in the collected data worthy of further investigation. For one, notice how neither series of experimental items with a long vowel (Figures 1 and 2) demonstrated any pattern with regard to the expected irregular inflection at all. No *shoon* became *shon* by the analogy with *shoot* → *shot* and no *drean* became *dreant* by the analogy with *dream* → *dreamt*.

Although no result is a fair result, there is one exception to the general absence of irregular past-tense inflection in the instance of *dreaf*. Relatively many *dreaf*s turned into *dreaft*s (see Figures 2 and 10). Rather than attributing this phenomenon to chance, the present study considered other existing lexical entries that might prompt analogical past-tense production in this case. A very good candidate was *leave* → *left*, which likewise shortens the vowel and adds a devoiced suffix to achieve its past-tense form and, although by devoicing the original final consonant, arrives at the same, relatively unique consonant cluster, *-ft*, in the final version.

Recall that earlier in the methodology section, this study refused to include *leave* in the pool of verbs that should be tested due to its uniqueness. It was thought to be too secluded to offer enough input for any English-speaking mind to form analogies based on it (see Yang, 2006 for more on pattern extraction). As it turns out, the uniqueness of an existing lexical entry may actually be what promotes connections, to the extent that it may even override phonological distance—even in items with complex onsets with a generally recognised significant priming effect (Bybee and Moder, 1983).

Proximity Effect on Variance. Perhaps the most interesting observation comes from the short-vowel items (Figures 3 through 7 and 11 through 15). Figures 19 and 20 model each of these 20 experimental items (5 base verbs × 4 modifications each) and show how the number of target responses (i.e., past-tense forms created by analogy with the irregular pattern of the base verb) changes with distance from the base item.

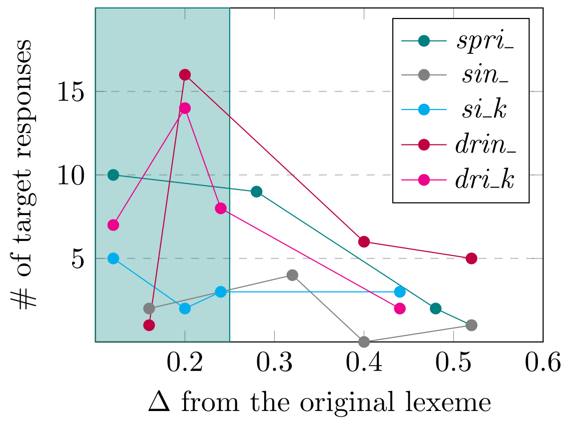

Figure 19: Decreasing variance of responses (the strict analysis).

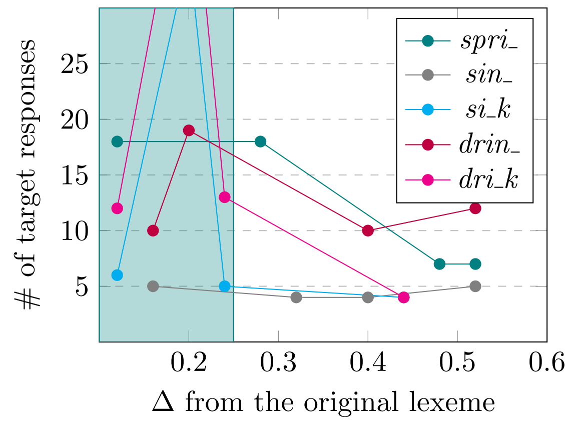

Figure 20: Decreasing variance of responses (the lenient analysis).

Notice that the data starts off very scattered: some lines begin to form a decreasing pattern immediately, some initially increase, some show significant, hardly explicable jumps, and multiple show inconsistent results between the strict and the lenient analysis. This anarchy rules on until around $\Delta = 0.25$, as showed in teal. After this era of chaos, all items follow a stable decline. The relevant metrics confirm this observation: the standard deviation (i.e., the $\sqrt{\text{variance}}$ in data) is significantly larger for the data outside the teal area (1.98× and 2.95× larger for the strict and lenient analyses respectively). Table 63 outlines this for both types of analyses.

		Δ	
		< 0.25	> 0.25
Analysis	Strict	$\sigma = 5.22$	$\sigma = 2.64$
	Lenient	$\sigma = 13.65$	$\sigma = 4.62$

Table 63: Standard deviation from the mean of data inside versus outside the teal area.

This data indicates that up to a Levenshtein Distance of $\Delta \approx 0.25$, the key factor for recognising the initial item is not the distance itself but rather the specific characteristics of the sound being replaced and the nature of the sound it is being replaced by. For instance, replacing a final nasal with a closely similar nasal yields high expected irregular rates while replacing a final stop with a closely similar stop does not. This could suggest that nasals are less sensitive to change than stops, possibly due to the fact that the English language has fewer phonetically distinct nasal consonants (more specifically, $n = 2$) than phonetically distinct stops ($n = 6$). It is only once modifying more distant sounds ($\Delta > 0.25$) that the Levenshtein Distance gains its predictive power.

V Conclusions and Limitations

Observations: Summary. Recall the initial questions of the present study: What predictive power does the Levenshtein Distance have on the categorisation of novel items? Is there a specific Δ at which an item stops resembling the original lexical entry? How phonologically close does an item need to be to a lexical entry to trigger analogical past-tense formation? In other words: How close is close enough?

The journey to the answer to the above questions brought several interesting discoveries. For one, the results of the experiment designed to shed more light on categorisation curiously contradicted several previous findings. These findings proposed that the more complex the syllable structure, the higher the chance of irregular past-tense forms, and that the contribution of certain types of sounds, such as nasals, likewise increased the likelihood of irregularity (Bybee and Moder, 1983). None of these predictions were borne out by this study's data.

The data shows a high variation of results between items with super-complex onsets (Figures 3 and 11) and preserving nasals in complex codas (Figures 4 and 6) did not yield higher rates of the irregular past-tense formation than swapping them with other sounds (Figures 5 and 6), showing that neither of this hypotheses had a strong enough predictive power. What mattered more, though, was vowel length: regardless of their onset and coda complexity, items with short vowels yielded more irregular past-tense forms than their long-vowel counterparts.

Furthermore, this study introduced some novel observations. The first highlight of the analysis is the Uniqueness Effect. Participants showed a strong preference for irregular past-tense formation when the experimental item resembled a lexeme displaying a one-of-a-kind inflection pattern among all verbs in the English lexicon (Figures 2 and 10). This was highly unexpected, as it is generally thought that a larger number of instances is necessary for the human brain to extract and reapply patterns (Yang, 2006). However, based on this finding, it is possible that general theories on pattern reapplication do not hold in resemblance tasks, and unique processes (in this case, /iː/ + fricative $\rightarrow$ /e/ + devoiced fricative) are just as competitive in derivation.

Finally, the answer to the initially posed questions emerged after the close analysis of short-vowel items. Here it became immediately obvious that the Levenshtein Distance only gains its predictive power when the modified item moves away from the initial lexeme. Therefore, while there seems to be no Δ threshold after which a novel lexeme no longer resembles the initial one, there is a threshold ($\Delta \approx 0.25$, roughly equal to the modification of 6 features) for when the characteristics of specific features stop weighing in as much and Δ becomes more predictive of how close two lexemes are perceived.

Learning Reflections and Limitations. However, it is also important to state and restate the multiple limitations this study encountered. More specifically, the author thought it important to comment on unaccounted-for variables (namely the effect of rhyme, differences between specific features of sound, and the frequency of use), the shortcomings of the experiment design (namely item choice and execution format), and the significance of the results considering the bigger picture.

Variables Disregarded. The first disregarded variable to mention is the effect of rhyme. This study chose to make changes to codas in the hope of seeing more pronounced results of the distance metrics. However, it turns out that alternating the coda might have been what prompted the participants to associate words with different lexemes, as they no longer rhymed with the original one. For example, although *shoot* and *shoon* are mutually closer than *shoon* and *spoon*, the latter pair might be associated with each other more often due to the two items sharing the same rhyme. Further insights in response to this limitation might come from a similar study on onsets.

This study used the Levenshtein Distance metric, which assigns the same weight to changes to any particular feature of sound. However, it disregards the fact that not all features are born equal and that some phonological features may have a greater impact on the perception of similarity than others. For example, while /t/ and /θ/ may be equally distant from one another as /t/ and /d/, the continuity feature (responsible for the difference between /t/ and /θ/) might carry greater weight than the voice feature (responsible for the difference between /t/ and /d/), leading towards different outcomes in perception. A greater consideration of the Feature Hierarchy (e.g., Jakobson et al., 1951) might be necessary.

This study also did not pay attention to the frequency of use of the individual items. For example, it considered the word *shrink* a better candidate for the base item selection than *slink* because it has fewer phonological relatives (*drink* versus *blink* and *stink*; see Appendix A), but it did not consider the fact that having fewer relatives might sometimes cause more distraction, as the fewer specific lexemes could be more frequent in use, leading to stronger associations and a greater intervention for the purposes of this experiment. In this case, a larger corpus overview which this study did not have sufficient capacity for might determine more feasible items for testing.

Gaps in Experiment Design. Even though this study tried its best to replicate the procedures from previous projects with similar intentions, maybe the methodologies are simply not suitable for the purpose. In other words, maybe methods that used to work back in the heyday of the past tense are no longer going to do the job. Other researchers in the field, and especially those researching inflection preferences, have encountered similar hardships with online experiments, effectively conveying their research objectives and ensuring that participants remain engaged and attentive throughout the experiment (e.g., Chuprinko et al., in preparation).

The typical resolution to this problem is running an additional, multiple-choice experiment. In this type of experiment, participants do not freely produce the requested inflected form but instead select one of the possibilities of interest (which would in this case be the target irregular form and a regular form of the past tense). This set-up provides a controlled environment that can mitigate participant distraction and highlight specific insights, albeit in exchange for true free-response answers. This trade-off might, however, be a cost we need to pay to adjust to the modern experimental culture and to ensure the clarity of the findings.

The author must also shamefully admit that despite several consultations with multiple native speakers of American English, it never occurred to them that *dream*, which was one of the selected lexemes for the testing of categorisation (see Table 5), is typically regular for the speakers of American English (while being typically irregular for the speakers of British English). Given that the study consisted of a mix of native speakers of both American and British English, the data for this experimental item might have been compromised on this account. Should this experiment ever be reproduced for one reason or another, this experimental item would need further reconsideration.

Significance of Data. Finally, it would be prudent to offer a modest reminder that all the above-described patterns were drawn from very fragile data. Recall that only 4.45% of responses were in the target irregular form. This translates to an average of 4.19 (out of 93) irregular responses per experimental item, which in practice means that if a single participant made a more careless typo or did not manage to accurately represent their preferred pronunciation through the chosen spelling of the newly introduced fabricated verbs, the data could be easily shifted by almost 25%. Such substantial volatility weakens the validity of this data for drawing any definitive conclusions.

Concluding Remarks. The present study investigated how good of a predictor the Levenshtein Distance, the traditional metrics of phonological distance, is to 'similarity'. The conclusion is that the metric only accounts for the preferences for the inflection of items that are phonologically more distanced from the initial lexeme. For items close to the initial lexeme, the hunt for an intra-linguistic metric that would accurately foresee the acceptability of variation between two 'similar' items continues. In other words, if the Levenshtein Distance does not provide a clear definition of phonological similarity for all items, then what does?

Future studies should explore alternative, more nuanced metrics that capture the perception of phonological neighbourhoods. Additionally, incorporating factors such as lexical frequency and Feature Hierarchy could provide a more holistic understanding of the process. Experimental designs that combine well-renowned, comprehensive methods with methods adjusted to the modern experimental culture might also provide new, more adequate results. Through these advancements, the field of linguistics will grow better equipped to understand the complexities of language processing as well as to model and predict linguistically conditioned preferences in general.*

*Conflict of Interests. The author declares no conflict of interest. This study was funded by the Pinkel Fund, which supports undergraduate research in the multidisciplinary field of Cognitive Science at the University of Pennsylvania. The funding covered the combined participant rewards and related fees for using the Prolific platform. Special thanks to Professor Charles Yang (University of Pennsylvania) and Dr Russel Richie (University of Pennsylvania) for considering and approving the application for this project.

References

Albright, A. and Hayes, B. (2003). Rules vs. Analogy in English Past Tenses: a Computational/Experimental Study. *Cognition*, 90:119–161.

Anderson, J. (2007). Phonological Neighborhood and Word Frequency Effects in the Stuttered Disfluencies of Children Who Stutter. *Journal of Speech, Language, and Hearing Research*, 50:229–47.

Anderwald, L. (2011). Norm vs variation in british english irregular verbs: the case of past tense sang vs sung. *English Language and Linguistics*, 15(1):85–112.

Bybee, J. L. and Moder, C. L. (1983). Morphological Classes as Natural Categories. *Language*, 59(2):251–270.

Caravolas, M., Lervåg, A., Defior, S., Málková, G. S., and Hulme, C. (2013). Different Patterns, but Equivalent Predictors, of Growth in Reading in Consistent and Inconsistent Orthographies. *Psychological Science*, 24(8):1398–1407.

Gouskova, M. and Becker, M. (2013). Nonce Words Show that Russian Yer Alternations Are Governed by the Grammar. *Natural Language & Linguistic Theory*, 31:735–765.

Huddleston, R. and Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.

Jakobson, R., Fant, C. G., and Halle, M. (1951). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. MIT Press, Cambridge.

Levenshtein, V. I. (1965). Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Doklady Physics*, 10:707–710.

Pinker, S. and Ullman, M. T. (2002). The Past and Future of the Past Tense. *Trends in Cognitive Sciences*, 6(11):456–463.

Prasada, S. and Pinker, S. (1993). Generalisation of Regular and Irregular Morphological Patterns. *Language and Cognitive Processes*, 8(1):1–56.

Ramscar, M. (2002). The Role of Meaning in Inflection: Why the Past Tense Does Not Require a Rule. *Cognitive psychology*, 45:45–94.

Rumelhart, D. E. and McClelland, J. L. (1986). On Learning the Past Tenses of English Verbs. In *Parallel Distributed Processing*, volume 2 of *Explorations in the Microstructure of Cognition: Psychological and Biological Models*, chapter 18. The MIT Press.

Tuena, C., Di Lernia, D., Riva, G., Serino, S., and Repetto, C. (2023). Phonological Similarity between Words Is Represented in Declarative Memory as Spatial Distance. *Psychological Research*, 87:2499–2510.

Vitevitch, M. (2002). The Influence of Phonological Similarity Neighborhoods on Speech Production. *Journal of experimental psychology. Learning, memory, and cognition*, 28:735–47.

Wydell, T. N. and Butterworth, B. (1999). A Case Study of an English-Japanese Bilingual with Monolingual Dyslexia. *Cognition*, 70(3):273–305.

Yang, C. D. (2006). *The Infinite Gift*. Scribner, New York.

A Relevant English Irregular Verbs by Similar Verbs

Pattern 1:

VERB	**C**CCVCC	C**C**CVCC	CC**C**VCC	CCCV**C**C	CCCVC**C**
keep	beep, peep, weep, seep, leap, reap, heap	N/A	N/A	keel	N/A
leap	beep, peep, weep, seep, keep, reap, heap	N/A	N/A	leaf, lead, lease, lean, leash, leer, leak	N/A
creep	∅	∅	N/A	cream, creak, crease	N/A
sleep	bleep	steep, sweep	N/A	sleet	N/A
sweep	∅	steep, sleep	N/A	sweet	N/A
weep	beep, peep, leap, seep, keep, reap, heap	N/A	N/A	weave, weed, wheeze, ween	N/A
feel	peel, deal, seal, kneel, reel	N/A	N/A	feed	N/A
mean	ween, dean, lean	N/A	N/A	meet	N/A
deal	peel, feel, seal, kneel, reel	N/A	N/A	ASK	N/A
dream	cream	∅	N/A	∅	∅
kneel	peel, deal, seal, feel, reel	N/A	N/A	ASK	N/A
lean	ween, dean, lean	N/A	N/A	leaf, lead, lease, leap, leash, leer, leak	N/A
bleed	plead	breed	N/A	bleep, bleat, ASK	N/A
feed	bead, weed, seed, need, lead, read, head	N/A	N/A	feel, fear	N/A
meet	beat, heat	N/A	N/A	ASK	N/A
read	bead, weed, seed, need, lead, feed, head	N/A	N/A	reel ASK	N/A
breed	ASK	bleed	N/A	brief ASK	N/A
speed	∅	ASK	N/A	speak ASK	N/A
lead	bead, weed, seed, need, read, feed, head	N/A	N/A	leaf, lean, lease, leap, leash, leer, leak	N/A
slide	glide ASK	∅	N/A	slice, slight, slime	N/A
light	bite, fight, sight, knight, light	N/A	N/A	bite, fight, sight, knight, write	N/A
shoot	boot, moot, toot, suit, loot, root, hoot	N/A	N/A	∅	N/A

Pattern 2:

VERB	**C**CCVCC	C**C**CVCC	CC**C**VCC	CCCV**C**C	CCCVC**C**
drink	shrink	∅	N/A	∅	∅
ring	bing, ping, wing, ding, sing	N/A	N/A	rid, rim, rip, rig	N/A
shrink	drink	∅	N/A	∅	∅
sing	bing, ping, wing, ding, ring	N/A	N/A	sip, sit, sin	N/A
sink	think, link	N/A	N/A	∅	∅
spring	∅	∅	∅	∅	N/A
stink	∅	∅	N/A	∅	∅
swim	∅	skim, slim	N/A	switch, swish	∅

Pattern 3:

VERB	**C**CCVCC	C**C**CVCC	CC**C**VCC	CCCV**C**C	CCCVC**C**
dig	pig, rig	N/A	N/A	dip, dim, diss, dish, ditch	N/A
win	bin, pin, sin	N/A	N/A	wiz, will, wish, whip	N/A
cling	fling, sling	∅	N/A	clip, click	N/A
fling	sling, cling	∅	N/A	flip, flick	N/A
sling	fling, cling	sting	N/A	slick, slit, slim, slip	N/A
slink	blink	stink	N/A	∅	∅
spin	∅	skin	N/A	spit, spill	N/A
stick	∅	∅	N/A	stiff, stitch, sting	N/A
sting	∅	sling, swing	N/A	stiff, stitch, stick	N/A
string	∅	spring	∅	∅	N/A
swing	∅	sting, sling	∅	swim	N/A
wring	bing, ping, wing, ding, sing	N/A	N/A	rid, rim, rip, rig	N/A

Pattern 4:

VERB	**C**CCVCC	C**C**CVCC	CC**C**VCC	CCCV**C**C	CCCVC**C**
ride	guide, bide, chide, side	N/A	N/A	rhyme, rive, write, rise, rile	N/A
drive	thrive	∅	N/A	∅	N/A
smite	∅	slight	N/A	smile	N/A
rise	size	N/A	N/A	rhyme, rive, write, ride, rile	N/A
strive	∅	∅	∅	stripe, strike	N/A
write	bite, fight, sight, knight, light	N/A	N/A	rhyme, rive, ride, rise, rile	N/A

Pattern 5:

VERB	<u>**C**</u>CCVCC	C<u>**C**</u>CVCC	CC<u>**C**</u>VCC	CCCV<u>**C**</u>C	CCCVC<u>**C**</u>
be**seech**	∅	N/A	N/A	besiege	N/A
bring	∅	∅	N/A	brim, bridge, brick	N/A
buy	pie, vie, die, tie, sigh, lie	N/A	N/A	buff, bud, buzz, bus, bum, budge	N/A
catch	match, patch, batch, latch, hatch	N/A	N/A	cab, cap, can, cash	N/A
seek	peak, leak, reek, geek	N/A	N/A	seep, seem, seethe, seed, seize, seat, seal	N/A
teach	beach, leech, reach	N/A	N/A	team, tease, tear	N/A
think	link, sink	N/A	N/A	∅	∅
fight	bite, write, sight, knight, light	N/A	N/A	fine, file, fire	N/A

B English Consonant Inventory

	/b/	/p/	/m/	/w/	/f/	/v/	/θ/	/ð/	/t/	/d/	/s/	/z/	/n/
cons	1	1	1	0	1	1	1	1	1	1	1	1	1
son	0	0	1	1	0	0	0	0	0	0	0	0	1
voice	1	0	1	1	0	1	0	1	0	1	0	1	1
rnd	0	0	0	1	0	0	-1	-1	-1	-1	-1	-1	-1
ant	-1	0	-1	-1	-1	-1	1	1	1	1	1	1	1
dist	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0
high	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1
back	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1
low	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
atr	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
rtr	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
cont	0	0	0	1	1	1	1	1	0	0	1	1	0
delr	0	0	-1	-1	0	0	0	0	0	0	0	0	-1
lat	0	0	0	0	0	0	0	0	0	0	0	0	0
nas	0	0	1	0	0	0	0	0	0	0	0	0	1
strid	0	0	-1	-1	1	1	0	0	0	0	1	1	-1
sg	0	0	0	0	0	0	0	0	0	0	0	0	0
cg	0	0	0	0	0	0	0	0	0	0	0	0	0
trill	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
flap	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
lab	1	-1	1	1	1	1	0	0	0	0	0	0	0
cor	0	0	0	0	0	0	1	1	1	1	1	1	1
dors	0	0	0	1	0	0	0	0	0	0	0	0	0
lar	0	0	0	0	0	0	0	0	0	0	0	0	0
tr	0	0	0	0	0	0	0	0	0	0	0	0	0

	/l/	/ɾ/	/ʃ/	/ʒ/	/t͡ʃ/	/d͡ʒ/	/j/	/ɹ/	/k/	/g/	/ŋ/	/h/	/ʔ/
cons	1	1	1	1	1	1	0	1	1	1	1	0	0
son	1	1	0	0	0	0	1	1	0	0	1	1	1
voice	1	1	0	1	0	1	1	1	0	1	1	0	0
rnd	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1
ant	1	1	0	0	0	0	0	1	-1	-1	-1	-1	-1
dist	0	0	1	1	1	1	1	0	-1	-1	-1	-1	-1
high	-1	-1	-1	-1	-1	-1	1	-1	1	1	1	-1	-1
back	-1	-1	-1	-1	-1	-1	0	-1	1	1	1	-1	-1
low	-1	-1	-1	-1	-1	-1	1	-1	0	0	0	-1	-1
atr	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
rtr	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
cont	1	1	1	1	0	0	1	1	0	0	0	1	0
delr	-1	-1	0	0	1	1	0	-1	0	0	-1	-1	-1
lat	1	0	0	0	0	0	0	0	0	0	0	0	0
nas	0	0	0	0	0	0	0	0	0	0	1	0	0
strid	-1	-1	1	1	1	1	-1	-1	0	0	0	-1	-1
sg	0	0	0	0	0	0	0	0	0	0	0	1	0
cg	0	0	0	0	0	0	0	0	0	0	0	0	1
trill	-1	0	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1
flap	-1	1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1
lab	0	0	0	0	0	0	0	0	0	0	0	0	0
cor	1	1	1	1	1	1	1	1	0	0	0	0	0
dors	0	0	0	0	0	0	1	0	1	1	1	0	0
lar	0	0	0	0	0	0	0	0	0	0	0	1	1
tr	0	0	0	0	0	0	0	0	0	0	0	0	0

C Relevant English Irregular Verbs by Structure

CVC	keep, leap, weep, feel, mean, deal, kneel, lean, feed, meet, read, lead, light, shoot, ring, sing, dig, win, wring, ride, rise, write, buy, catch, seek, teach, fight (beseech, begin)
CCVC	creep, sleep, sweep, dream, bleed, breed, speed, slide, swim, sling, cling, fling, spin, stick, sting, swing, drive, smite, bring
CCCVC	spring, string, strive
CVCC	sink, think
CCVCC	drink, shrink, stink, slink

D Filler Items

siːt͡ʃ	siːf	siːθ	sɪːb
dɹis	dɹið	dɹik	dɹis
spʊl	spʊs	spʊf	spʊk
hæp	hæd	hæz	hæb
læf	lær	læð	læt
tiːk	tiːd	tːg	tiːl
tuːp	tuːk	tuːf	tuːd

E Experiment (Code)

Demonstration link with the full code available upon copying the project.