I’m Crept Out!: Factors Motivating Lexical Decision for Multiple Past Tense Word Forms of the Same Lexeme

Rebecca Gips

Abstract

In English, verb forms can take on the past tense in different ways. Adding a past-tense *-ed* suffix is considered a regular form of changing a verb to the past tense in a process known as concatenation. Concatenation can be defined as the process of two morphemes being put in a sequential order to form words, where morphemes are the smallest meaningful use of language. However, irregular forms of inflecting past tense through concatenation also exist, for example with *-t*, rather than *-ed*. In this way, speakers must decide which form they will use in their utterances. This paper examines the factors that influence people's choices between word forms, with a focus on phonological and semantic context. Through the exploration of corpora data, it was found that semantic context has a greater influence on word form choice than phonological factors, but additional factors such as utterance register were also found to be influential.

DOI: 10.5195/pur.2025.136

1: Introduction

Irregular forms of the past tense for many verbs exist, in addition to their regular past tense forms. A lexeme is a unit of lexical meaning related to other words through inflection. According to the Google Books Ngram Viewer, for the lexeme *creep*, *crept* (the irregular form) is the more historically used past tense form, as compared to *creeped* (the regular form). However, speakers may apply regularization processes if the technically correct, irregular form is unknown to them, creating a situation where both forms are used by speakers.

Consider the following:

(1) Two forms of concatenation of *creep*
She *crept* into the room to avoid suspicion.
She *creeped* into the room to avoid suspicion.

According to grammaticality judgements, most native speakers would consider using either of the above forms to be grammatical, but might choose one form over the other unknowingly. This shouldn't be a problem, as many words have close synonyms where both forms are used.

However, consider the counterexample:

(2) Two forms of concatenation of *creep*
The interaction creeped me out.
*The interaction crept me out.

Now, by grammaticality judgements, it is obvious that there are cases where only one past tense form of the verb works in certain contexts. Therefore, some sentences allow either form to occur, while some sentences become ungrammatical when choosing one of the forms. So, what is it about the context that makes a speaker choose one form over another? In this paper, the motivations for choosing between two competing concatenative past-tense inflections will be explored.

Section 2 of this paper first explores the possibilities for why there are multiple word forms of the same lexeme. Regularization and replacement are first explained, and then the idea that there may be semantic and phonological contexts for each word form is introduced. Next, section 3 presents the two research questions and posits predictions and justifications based on research regarding phonologically conditioned morphology. Next, section 4 describes the methodology used in the study to collect data. The data is then presented and analyzed in section 5, with section 6 synthesizing the data together and addressing the limitations of the study, as well as presenting ideas for future research on this topic.

DOI: 10.5195/pur.2025.136

2: Background

When two forms compete for the same meaning, several things can happen. First, there is the possibility of regularization, in which a speaker derives a form using morphological processes when they are unfamiliar with the irregular form. In this way, it is possible for both forms to exist simultaneously, although one may be in the process of being replaced. Second, there is the possibility that both word forms exist because they are used in different contexts, notably: semantic (meaning-based) and phonological (sound-based) contexts. In this section, I review each of these possibilities.

2.1 Regularization

In terms of concatenation, regularization processes are often applied, even when an irregular form already exists. This can happen when a speaker is unfamiliar with the irregular form, but is in a communicative situation where they need to use the past tense form. For example, in the past tense of *burn, burnt* is the irregular form. When a regular *-ed* process is applied, however, *burned* is created. Recently, this regularized form became more popular than the irregular form (Gray et al., 2018). When speakers are given the choice between two forms, it creates an alternation where the morpheme's phonological realization differs based on the context it appears in. It is increasingly common to see speakers pick the regularized form. Regularization gives rise to the idea of productivity, because people are able to apply regular forms to words with lower token frequency in order to produce "new" words. Although these words may already have an irregular form, speakers who are not familiar with them may apply regular morphological processes, such as the past-tense *-ed,* to derive what they are looking for.

Over time, diachronic processes can lead to the specialization of one of the forms, meaning that it gets crystallized to a specific context. Therefore, this crystalized form takes on a separate meaning from the original past tense form. In this way, it is possible for "different variants [to] find distinct niches, with most variants being specialized and the apparent elsewhere variant sweeping up the residue" (Aronoff, 2019). Therefore, although applying regular processes was likely intended to inflect for forms that speakers were unfamiliar with, they can eventually derive forms with different meanings. For example, speakers can apply the past tense *-ed* to lexemes that already have an irregular *-t* ending, but this application of the past tense *-ed* may form a semantically distant word. In this way, this crystallization of one form can be what decides between two forms. This possibility will be further discussed in sections 5 and 6.

2.2 Replacement

Inflection done by base modification often makes a distinction between the basic past tense and past participle. For example, in the past tense for *sing*, there is *sang* (the basic past tense), and *sung* (the past participle). This means that grammatical context is the sole determinant of which past tense form is used. However, it is also possible for processes solely including concatenation to be semantically context-dependent, rather than required by grammatical agreement or government. In fact, in some examples with two forms of concatenation, one form is prescriptively grammatical for both the past tense and past participle. For example, although both *creeped* and *crept* are used interchangeably in some contexts, as shown in section 1, *crept* is the actual past tense *and* past participle (Merriam Webster, n.d.). Because the other form may be created by speakers who are unfamiliar with the technically correct form, it may also be used in both contexts (past and past participle). Now, speakers find themselves in a situation where there are two forms of a word that can be used in either context. This demonstrates further that the semantic or phonological context may be what has an influence on the lexical decision, rather than grammatical agreement.

Lastly, this informal variation of past-tense inflection may be *replacing* the grammatical form. It is shown that "semantic factors… predict the rate of lexical replacement of content words" (Vejdemo & Hörberg, 2016). Multiple forms expressing the same meaning are possible in a language, as one form may be in the process of *replacing* the other. If two forms of a past tense exist and both forms can be used in either context (past tense or past participle), it is possible that the replacement has not yet finished, and therefore, both forms still continue to be used. This phenomenon of two variations of the same lexical item existing at once is important to explore because it gives insight into how speakers adapt language to better suit their usage, rather than relying on prescriptive grammatical rules. Although this longitudinal aspect of lexical decision falls outside the scope of this synchronic study, it is still relevant to explain why there may be two forms of the same lexeme that both have the same usage.

2.3: Semantic and Phonological Context

Although it may be possible that two forms having the same meaning can exist in a language because one simply has not yet been replaced, it is also possible that each form exists in a certain context and therefore will *not* be replaced. Two different contexts, the semantic and phonological, will be investigated in this paper. Firstly, it has been shown that lexical selection can be affected by semantic context, meaning that the word or words immediately surrounding the target word have an effect on which lexical item is selected (Rahman, 2009). Semantic context can include specific set phrases which only the target word can appear in. For example, semantic priming experiments have shown that showing semantically related words can *prime* participants to activate semantically related words. Response times to identify if the target is a word or a nonword are faster when the target word is related than if it is unrelated. If the prime is "nurse", participants would be faster to respond that "doctor" is a word, than something unrelated, such as "dog" (Schendan 2012).

Furthermore, phonological context, or the phonemes that precede and follow the target word, can also have an effect on which word is chosen. The past tense suffix in English may simply have two forms, *-t* and *-ed*, meaning that they are allomorphs (different forms of the same morpheme) of each other. It is well documented that allomorphy can be conditioned by the phonological environment (Nevins et al., 2011, 1). The reason that speakers may choose one suffix over the other is because of phonological allomorphy, where "it is evident that the historical reason for the existence of the morphophonological rule and thus for the allomorphy is to facilitate pronunciation" (Haspelmath & Sims, 2010, 24). This means that one of the allomorphs may make more sense in a certain phonological context, such as having similar voicing or other similar phonological features, making them phonological allomorphs (Haspelmath & Sims, 2010, 25). Therefore, word choice, by way of the chosen allomorph, can be affected by phonological context. There is large ongoing discussion within the field of phonology over allomorph selection, particularly about which of the allomorphs is the underlying representation, or *default*, and which one is the phonologically marked version, or *optimizer* (Nevins et al., 2011, 21). If the alternations discussed in this paper are found to be selected based on the phonological environment, then an appropriate next step would be to determine which of the allomorphs are the default and the optimizer.

As discussed, there are multiple reasons why there might be two forms of the same lexeme that have the same meaning. Firstly, speakers may be applying regularization processes to lexemes where they are unfamiliar with the irregular form of the word that already exists. From this, both terms can begin to be used interchangeably, and it is possible that the regularized version may eventually replace the irregular version, although this possibility would have to be explored longitudinally. On the contrary, it is possible that both forms appear in separate contexts and therefore would not be subject to lexical replacement, as both are individually necessary. It could be that one lexeme has crystallized to a specific context, taking

on its own separate semantic meaning. It is also possible that there are specific phonological contexts that demand a specific word choice over another.

DOI: 10.5195/pur.2025.136

3: Research Question and Hypothesis

The principal questions that guide the analysis in this study are:

1. Are there specific semantic and/or phonological contexts that trigger a certain past tense form of a verb to appear in an utterance?
2. Do these contexts map across different verbs, or are they lexically specific?

The main hypothesis was that the context of the surrounding words would have an effect on the lexical item chosen, and therefore, there isn't complete replacement. In this way, there must be some sort of context that triggers each form of the verb, or triggers one form if one of the allomorphs is considered to be the default. Analysis of semantic context will focus on phrasal constraint; do these lexical items always occur in a set phrase, or are they more free to occur alongside many lexical items? The prediction was that one form will be used in a more generalized context (the default), and one will have an extremely limited application, and on the extreme end have a new context-specific meaning due to the effects of crystallization (the optimizer). The phonological context was also analyzed. For the purpose of this study, the phonemes immediately preceding and following the lexical item in focus will be taken and analyzed for the phonological context. A secondary prediction was that phonological context will have a larger impact than semantic context due to the well-known concept of phonologically conditioned morphology as described in section 2.3.

4: Methodology

The data in this study was collected from the "Corpus of Contemporary American English", or COCA (Davies 2008-). The words at the focus of the study were *creep (*creeped, crept)*, learn* (learned, learnt)*, burn* (burned, burnt), and *kneel* (kneeled, knelt). 20 instances of each past tense form were analyzed. With four lexemes of two word forms each, 80 contexts were analyzed for each lexeme, and 320 different semantic and phonological contexts were analyzed in total.

For data collection, each word form was given a column in a spreadsheet. The columns to the left and right of the word form contained the preceding and following semantic and phonological context (for a total of four tables per lexeme). The broad IPA transcription for the phonological context was done by hand by the author. An example of a data set for one of the lexemes is shown in Figures 1 and 2.

she will definitely be	creeped	out in case you	a voice has	crept	in
at least a LITTLE	creeped	out by seeing how	which had	crept	higher after summer's starvation
I was	creeped	out	change has	crept	through the house I
you want to be	creeped	out	The sun	crept	downward
and am continually	creeped	out the show wants	The black ooze	crept	ashore on the waves
While some were	creeped	out by the news of	on the road he	crept	#
to get a little	creeped	out	I	crept	up to one of
who put it on	creeped	me out	This storm, however	crept	along the Eastern Seaboard
feel slightly less	creeped	out and that their iPhone,	The doubt finally	crept	in
person you'll be	creeped	out	insecurities about care	crept	in
you how much it	creeped	me out watching and	the prices have	crept	upward for years
the only one whose	creeped	out by yet another	She	crept	onto the damask chair
The only cake that	creeped	me out was the	Salt Lake City where I	crept	onto a chair
I get mega	creeped	out by bodybuilders	The temperature	crept	up in the late
Prank calling has never	creeped	me out more after	after them and	crept	home to some
If her mom's	creeped	out by this news	tiny voice	crept	into my brain
She is	creeped	out and is considering	what happens after, has	crept	into fiction
She is	creeped	out and is considering	Slowly but surely materialism	crept	into the church
he has always	creeped	me out to the max	a car	crept	up behind us
bad romance video really	creeped	me out	This omission has	crept	into the

Figure 1: Semantic context for "creep" (creeped/crept)

[dɛfɪnɪtli bi]	creeped	[aʊt]	[vɔɪs hæd]	crept	[ɪn]
[ʌ lɪrl]	creeped	[aʊt]	[wɪʧ hæd]	crept	[haɪjɹ]
[aɪ wʌz]	creeped	[aʊt]	[ʧeɪnʤ hæz]	crept	[θɹu]
[tu bi]	creeped	[aʊt]	[ðʌ sʌn]	crept	[dɑʊnwəɹd]
[kʌntɪnjuəli]	creeped	[aʊt]	[uz]	crept	[ʌʃoɹ]
[wʌɹ]	creeped	[aʊt]	[ɹod hi]	crept	#
[ɪtrl]	creeped	[aʊt]	[aɪ]	crept	[ʌp tu]
[ɔn]	creeped	[mi aʊt]	[hɑʊɛvɹ]	crept	[ʌloŋ]
[lɛs]	creeped	[aʊt]	[faɪnʌli]	crept	[ɪn]
[jol bi]	creeped	[aʊt]	[ʌbɑʊt kɐ]	crept	[ɪn]
[mʌʧ ɪt]	creeped	[mi aʊt]	[pɹaɪsɛz hæv]	crept	[ʌpwə-d]
[huz]	creeped	[aʊt]	[ʃi]	crept	[ɔntu]
[kɛɪk ðæt]	creeped	[mi aʊt]	[wɛ- aɪ]	crept	[ɔntu]
[mɛgə]	creeped	[aʊt]	[tɛmpɛɹʌtʃɹ]	crept	[ʌp ɪn]
[hæz nɛvɹ]	creeped	[mi aʊt]	[ænd]	crept	[hom]
[mɔmz]	creeped	[aʊt]	[vois]	crept	[ɪntu]
[ʃi ɪz]	creeped	[aʊt]	[hæz]	crept	[ɪntu]
[ʃi ɪz]	creeped	[aʊt]	[mʌtiuʌlɪzm]	crept	[ɪntu]
[ɔlweɪz]	creeped	[mi aʊt]	[ʌ kɔɹ]	crept	[ʌp]
[ɹili]	creeped	[mi aʊt]	[hæz]	crept	[ɪntu]

Figure 2: Phonological context for "creep" (creeped/crept)

5: Data and Analysis

5.1 Learn and Burn

The first discussion will discuss the data from *learn* and *burn*. These two lexemes will be discussed together because they demonstrate extremely similar findings. The regularized *-ed* forms are both more frequent than the irregular *-t* forms, with *learned* (106055 instances) being about 60 times more common than *learnt* (1744 instances). *Burned* (24871 instances) is also more frequent than *burnt* (5390 instances), but only by about 4.6 times. In terms of phonological context, all four word forms appear before and after voiced and voiceless phonemes, meaning this context does not play a role. Briefly, voicing is a feature of consonants that specifies whether the vocal folds are vibrating or not. For example, /f/ is unvoiced, while /v/ is voiced. Additionally, the semantic context also appears to not play a large role. Therefore, the prediction that the semantic or phonological context would have an influence on the chosen word form was not borne. This leaves the possibility that one of the forms is replacing the other, but as previously mentioned, these longitudinal possibilities cannot be explored due to the scope of the available data.

The only noteworthy finding for these two lexemes comes from the instances of the word forms appearing as an *adjective* rather than the past tense of the verb. This phenomenon occurs twice for *learn* and nine times for *burn*, although interestingly, this word class change happens for the regularized *-ed* for learn (e.g., *learned societies*) and the irregular *-t* for burn (e.g., *burnt food*). *Learnt* and *burned* never occur as adjectives, only as past tense verbs. Further research would have to be done to determine why this change in word class is not applied regularly to one of the forms (e.g., the *-ed* suffix for both forms) of the past tense morpheme. Since the data shows that it is not semantic or phonological context, and we can determine frequency does not play a role (since this adjectival formation appears on one form that is more frequent than its counterpart and one less frequent form), there must be some other determining factor.

5.2 Creep

The next lexeme for analysis is *creep*. *Crept* (3891 instances) is about 6.8 times more frequent than *creeped* (568 instances). The crucial finding for this data is that 100% of the instances of *creeped* occur in the phrase “creeped out”. Some occur with an intervening word such as “me” (in phrases such as “creeped me out”), but this word form still exhibits phrasal constraint. *Crept* occurs 18/20 times before a prepositional phrase (“crept through…”, “crept along…”). This indicates that these two forms have taken on separate meanings from each other. Only *crept* indicates the past tense of movement, which is what the expected past tense form for *creep* would show. However, *creeped* has become collocated with “out”, crystallizing to a different semantic meaning, relating more so to feelings or emotions than movement, expressing a spooked or scared attitude. Because the data for this alternation can be explained by semantic context, the phonological context was found to be irrelevant and will not be discussed.

5.3 Kneel

The last lexeme to be analyzed will be *kneel*. *Knelt* (3591 instances) was about 8.5 times more frequent than *kneeled* (423 instances). This lexeme is similar to “learn” and “burn” in that neither the semantic nor phonological contexts have much of an impact on the choice of lexeme. For semantic context, both *knelt* and *kneeled* occur before a prepositional phrase (19/20 and 14/20 times, respectively), so this context cannot be used to predict which form will be used. Furthermore, they most often occur preceding or following voiced phonemes. Kneeled follows a voiced phoneme 19/20 times and knelt 19/20 times. Kneeled precedes a voiced phoneme 18/20 times, and knelt does 20/20 times. Therefore, the voicing of phonemes cannot be used to predict which will be chosen because both occur in this environment. In this

way, the prediction that semantic or phonological context will have an impact on word form choice was also not borne for this lexeme.

However, a very compelling finding was discovered with this lexeme. Interestingly, it was found that the genre of the utterance affects this lexeme more so than the others. *Kneel* was the only lexeme where the COCA didn't report the genre for every instance of the word form as "BLOG" (words from 1.8 million webpages in 20 English-speaking countries). For this set of data, the first 100 instances of the word form were analyzed in order to explore the genre of utterance more thoroughly. For *kneeled*, of the first 100 instances, 19 are labeled as "FIC" (fiction), 9 are "MOV" (movie), and 4 are "TV". There is also a small portion of these instances (11/100) that come from "NEWS" or "MAG" (magazine). An extreme majority of the remaining instances of *kneeled*, as well as almost all of the instances of *knelt,* occur in religious and fan fiction contexts (labeled as either "WEB" or "BLOG"). Obtaining an exact number of instances for appearing in these contexts is difficult because it is not always apparent the difference between a fictional story and an individual's blog post. Additionally, although the immediate context of the instance remains on the COCA, some of the original websites are no longer online, making it impossible to know the larger subject of the publication.

Despite these limitations, it is still clear that these fictional and religious contexts make up the majority of the remaining instances. The majority of the instances occurring in a religious or fictional context gives rise to the idea that register of utterance might also have an impact on the choice of lexeme. If it is a nonfiction context, *kneeled* will be the choice over *knelt*, but other synonyms, such as "bent" or "crouched", seem to be more likely to be used in this context. *Kneeled* and *knelt* are somewhat restricted to fiction environments based on this data.

6: Discussion

The data demonstrates that none of the lexemes are influenced by the phonological context, but they *can* be influenced by semantic context if one of the word forms has crystallized to a different meaning. *Creeped/crept* was the only alternation that was influenced by semantic context, as *creeped* occurs only in a phrase with "out", demonstrating phrasal constraint. This can demonstrate that the grammaticality judgements clash with the corpora data. For example, in section 1, it was described that "She *creeped* into the room to avoid suspicion" was grammatical. However, the corpora data would indicate that it must be *crept* that is used in this context. This answers the first research question, which asked whether phonological and semantic contexts can trigger one word form over another. In this case, the semantic context triggers the use of one form.

Due to the fact that the influence of context is lexically specific, the answer to the second question, which asked if these contexts mapped across lexical items, is no. One of the alternations having a form that crystallized to a different meaning does not indicate that all the lexemes will have this same result. Further research can be done on why this process only happened on one of the lexemes, and perhaps determine a pattern for other items that crystallization might happen to.

Furthermore, the hypothesis that there will not be complete replacement also depends on the lexeme in question. *Creeped* will not replace *crept* because they are used in different semantic contexts. However, further research will need to be done to determine whether full replacement will occur with "learn", "burn", and "kneel" because both forms of the past tense occur in the same semantic and phonological contexts. This could be done in a longitudinal study, observing the frequency of these word forms over the next decade, for example, to see if one form becomes increasingly used while the other loses relevance as the regularized form replaces it as the default.

If there is no full replacement, then there must be some other factor that keeps both word forms in the lexicon. As found in the *kneeled/knelt* alternation, genre of utterance can have an impact. Perhaps another factor that can influence word form choice is regional dialect variation. The COCA encompasses only American English, so perhaps other dialects of English use one form more frequently than the other. To explore this, this study could be replicated using the British National Corpus, or BNC (Davies 2004). Although it may be possible that similar findings are found about the effect of the contexts, a frequency comparison could be done to see if the dialect has an impact on which word form is chosen. A small exploration of the BNC shows that *learned* is only 2.4 times more frequent than *learnt*, as compared to the 60 times more frequent usage in American English. Additionally, *burned* and *burnt* occur in equal frequency. This can suggest that the irregular *-t* suffix may be more commonly used in this dialect, and therefore that the dialect of the speaker can have an effect on which word form is used.

7: Conclusion

By exploring over 300 instances of lexemes that have two choices for the past tense form, it has been demonstrated that phonological context has no impact on the choice of lexeme, while semantic context can have an impact in specific contexts. However, the data also showed that other factors, such as genre, can also impact the choice of the speaker.

This study was limited due to the fact that only four lexemes were analyzed, and more patterns may have been able to be discovered had more data been collected. Expanding this study to more lexemes would allow for the possibility of phonological context having more of an influence, as well as determining more examples of lexemes that have crystallized to specific contexts. In this case, it might be possible to determine a pattern for crystallization. Further research would also allow the discovery of more factors that impact word form choice to be determined, such as the utterance genre. Additionally, studying the word forms that did not exhibit semantic or phonological context effects over a long time period could determine if one of the forms is currently in a state of being replaced. Lastly, a study on what may cause word class change in one form but not the other (for example, *learned* but not *learnt*) may be done.

The implications of this research are that the choice of word form may not have to do with the immediate context of the utterance, and instead that the pragmatics may have more of an impact. Therefore, normal language processes such as regularization and replacement, which are done subconsciously by speakers, can have a rippling effect on language usage and set up interesting phenomena for study.

References

Aronoff, Mark. 2019. "Competitors and Alternants in Linguistic Morphology." *Studies in Morphology* 5:39-66. https://doi.org/10.1007/978-3-030-02550-2_2.

Davies, Mark. 2004. "British National Corpus (from Oxford University Press)." English Corpora. https://www.english-corpora.org/bnc/.

Davies, Mark. 2008. "COCA: "the Corpus of Contemporary American English." English Corpora. https://www.english-corpora.org/coca/.

Google. n.d. "NGrams." Google Books NGram Viewer. https://books.google.com/ngrams/.

Gray, Tyler J., Andrew J. Reagan, Peter S. Dodds, and Christopher M. Danforth. 2018. "English verb regularization in books and tweets." *PLOS ONE* 13 (12). https://doi.org/10.1371/journal.pone.0209651.

Haspelmath, Martin, and Andrea D. Sims. 2010. *Understanding Morphology*. N.p.: Hodder Education.

Merriam Webster. n.d. "What's the past tense of 'creep'? Is it creeped or crept?" Merriam-Webster. Accessed July 17, 2025. https://www.merriam-webster.com/grammar/whats-the-past-tense-of-creep.

Nevins, Andrew, Marc van Oostendorp, Elizabeth Hume, and Keren Rice. 2011. "Phonologically-Conditioned Allomorph Selection." *The Blackwell Companion to Phonology*, (January).

Rahman, Rasha A. 2009. "Semantic context effects in language production: A swinging lexical network proposal and a review." *Language and Cognitive Processes* 24 (5): 713-734. https://doi.org/10.1080/01690960802597250.

Schendan, H. E. 2012. "Implicit Memory." In *Encyclopedia of Human Behavior (Second Edition)*, 400-409. N.p.: Elsevier. https://doi.org/10.1016/B978-0-12-375000-6.00200-7.

Vejdemo, Susanne, and Thomas Hörberg. 2016. "Semantic Factors Predict the Rate of Lexical Replacement of Content Words." *PloS one* 11 (1). https://doi.org/10.1371/journal.pone.0147924.