



Lexical Semantics and Irregular Inflection

Citation

Huang, Yi Tang, and Steven Pinker. 2010. Lexical semantics and irregular inflection. *Language and Cognitive Processes* 25(10): 1411-1461.

Published Version

doi:10.1080/01690961003589476

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:10021556>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Lexical Semantics and Irregular Inflection

Yi Ting Huang
The University of North Carolina at Chapel Hill

Steven Pinker
Harvard University

Acknowledgements: Supported by NIH grant HD-18381. We thank Jeff Birk for his assistance in programming, Jesse Snedeker and members of the Harvard Laboratory for Developmental Studies for feedback, and three anonymous referees for invaluable comments on an earlier version. Portions of this work have been presented at the 27th annual meeting of the Cognitive Science Society in Stresa, Italy. Authors' address: Dept. of Psychology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; Dept. of Psychology, Harvard University, Cambridge MA 02138. Email addresses: ythuang@email.unc.edu; pinker@wjh.harvard.edu.

Abstract

Whether a word has an irregular inflection does not depend on its sound alone: compare *lie-lay* (recline) and *lie-lied* (prevaricate). Theories of morphology, particularly connectionist and symbolic models, disagree on which nonphonological factors are responsible. We test four possibilities: (1) Lexical effects, in which two lemmas differ in whether they specify an irregular form; (2) Semantic effects, in which the semantic features of a word become associated with regular or irregular forms; (3) Morphological structure effects, in which a word with a headless structure (e.g., a verb derived from a noun) blocks access to a stored irregular form; (4) Compositionality effects, in which the stored combination of an irregular word's meaning (e.g., the verb's inherent aspect) with the meaning of the inflection (e.g., pastness) doesn't readily transfer to new senses with different combinations of such meanings. In four experiments, speakers were presented with existing and novel verbs and asked to rate their past-tense forms, semantic similarities, grammatical structure, and aspectual similarities. We found (1) an interaction between semantic and phonological similarity, coinciding with reported strategies of analogizing to known verbs and implicating lexical effects; (2) weak and inconsistent effects of semantic similarity; (3) robust effects of morphological structure, and (4) robust effects of aspectual compositionality. Results are consistent with theories of language that invoke lexical entries and morphological structure, and which differentiate the mode of storage of regular and irregular verbs. They also suggest how psycholinguistic processes have shaped vocabulary structure over history.

Lexical Semantics and Irregular Inflection

The psychological and historical factors that determine whether a word has a regular or an irregular inflectional form has important implications for theories of the mental representations underlying language. Of the kinds of influencing this process, the role of sound is not in dispute. Most irregular verbs fall into clusters with similar sounds (e.g., *sing-sang*, *ring-rang*, *drink-drank*; *find-found*, *wind-wound*, *bind-bound*), and when people are asked to provide past-tense forms for meaningless words such as *plip* or *spling*, they show high agreement in guessing *plipped* and *splung* (Bybee & Moder, 1983; Prasada & Pinker, 1993). Yet it is also clear that sound cannot be the *only* determinant of irregularity, because words with the same sound can have different past-tense forms, e.g., *lie-lay* (recline) vs. *lie-lied* (prevaricate), *hang-hung* (suspend) vs. *hang-hanged* (execute). The words in these pairs clearly have different meanings, and this suggests that meaning, like sound, can affect a verb's inflected form. There are, however, many *ways* in which a word's meaning could determine its inflected forms, with different implications for the architecture of language processing. This paper aims to distinguish among these possibilities, both theoretically and empirically.

The effects of lexical semantics on irregular inflection became a subject of theoretical interest when Pinker and Prince (1988) noted that pairs such as *lied* and *lay* cannot be generated by any model of language whose input consists only of a phonological representation of a word. These include Rumelhart and McClelland's (1986) influential connectionist model of past-tense formation and many of the models developed in its wake. Pinker and Prince used this shortcoming as part of their argument for the psychological reality of linguistic structures eschewed in most connectionist models, such as lexical entries and morphological representations. Yet while many papers have since debated the nature and explanation of the phenomenon, there remains a lack of consensus on this issue (Baayen & Moscoso del Prado Martin, 2005; Berent, Pinker, & Shimron, 1999; Berent, Pinker, & Shimron, 2002; Clahsen, 1999; Daugherty, MacDonald, Petersen, & Seidenberg, 1993; Egedi & Sproat, 1991; Gordon & Miozzo, 2008; Harris, 1992; Joanisse & Seidenberg, 1999; Kim, Marcus, Pinker, Hollander, & Coppola, 1994; Kim, Pinker, Prince, & Prasada, 1991; MacWhinney & Leinbach, 1991; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; McClelland & Patterson, 2002; Nakisa & Hahn, 1996; Pinker, 1999; Pinker & Prince, 1994; Pinker & Ullman, 2002b; Ramscar, 2002, 2003; Shirai, 1997; Tabak, Schreuder, & Baayen, 2005; Tyler et al., 2002, 2005). We suggest that this is largely a consequence of a lack of precision in what is meant by "semantics" and its role in linguistic computation. Thus we distinguish between four possibilities.

1. *Semantic association.* Connectionist researchers propose that the effect of meaning on inflection is similar to the effect of sound, namely associations between a distributed set of input features the inflected output form. Thus a straightforward way to get connectionist models to distinguish homophones is to augment their input array, which consists of units for phonological features, with additional units for semantic features. For example, MacWhinney and Leinbach (1991) devised a model with units for features such as "*was liquid added or removed?*" and "*was there a high-pitched sound?*" Their units could distinguish the meanings of all the homophones in the training set, allowing the model to strengthen associative connections between the semantic units specific to a word and the output units for its appropriate past-tense form, and to strengthen inhibitory connections to the competing past-tense form.

MacWhinney and Leinbach's model embodied the most common representation for lexical semantics in connectionism: a large array of context-sensitive sensorimotor "microfeatures" (see e.g., Hinton, McClelland, & Rumelhart, 1986; Smolensky, 1988). Such representations take advantage of the architecture of connectionist pattern associators, where the large number of units and their ability to be associated during training compensates for their lack of a mechanism representing hierarchical and syntactic structure. Such models also have the advantage of automatically generalizing by similarity, because the overlap of features among input items implicitly defines subclasses with similar outputs (Hinton, McClelland, & Rumelhart, 1986). In the case of phonology, this architecture enables such models to easily learn input-output pairs that fall into a subclass: having been taught *fling-flung* and *string-strung*, they have an advantage in learning *cling-clung*. It also allows them to generalize a subregularity to new inputs: given *spling*, the models can guess *splung*.¹ Numerous studies have confirmed that children and adults use these phonological similarities to learn and generalize irregular patterns (Berent, Pinker, & Shimron, 1999; Bybee & Moder, 1983; Bybee, 1982; Clahsen, 1999; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Marcus et al., 1992; Prasada & Pinker, 1993; Ullman, 1999). For this reason, Pinker and Prince (1988) and Pinker and Ullman (2002) have suggested that even symbolic models of language should incorporate an associative component to the memory system that stores words' phonological representations.

Pinker and Prince noted, however, that such similarity effects are hard to find when it comes to the *semantics* of irregular verbs. At least at first glance, families of verbs with similar irregular forms are heterogeneous in meaning (e.g., *sing*, *ring*, and *drink*), and conversely, near-synonyms can have distinct kinds of past-tense forms (compare, e.g., *hit-hit*, *strike-struck*, and *slap-slapped*). This led them to several alternative hypotheses on the nonphonological determinants of irregularity.

2. *Lemma distinctness*. In traditional linguistics and many psycholinguistic models (e.g., Dell, Burger, & Svec, 1997; Levelt, 1988; Roelofs, 1992; Levelt, Roelofs, & Mayer, 1998; Janssen, Roelofs, & Levelt, 2002; Marslen-Wilson, 1987; Orsolini & Marslen-Wilson, 1997), a lexical item has a unique entry or address in memory. This abstract entry, often called a lemma, holds together the different pieces of idiosyncratic information that define a distinct word—its sound, meaning, and inflectional class—but cannot be reduced to any one of them. In principle, a pair of words which vary with respect to their sounds and meanings may also adopt different irregular forms. Critically, however the meaning difference would be an epiphenomenon of the distinctness of the two lemmas: the irregular form is linked to the lemma, which is also linked to the word's semantic representation, but the irregular form need not be linked to the semantic representation directly. Thus, in contrast to the associative effects that have been demonstrated for phonology, lemma effects do not lead to semantically cohesive subclasses or generalizations to semantically similar forms. Examples of this difference may be seen in pairs such as *lie-lay* and *lie-lied*, *ring-rang* and *wring-wrung*, and *meet-met* and *mete-meted*, in which the members have different past-tense forms, but do not belong to families of other verbs with similar behavior.²

¹ Though multilayer connectionist models with distributed representations can be trained to inhibit such generalization, it is the default tendency (Marcus, 2001).

² Although homophonous lemmas are often distinguished by their orthography, as in *meet/mete*, the example of *lie/lie* shows that they need not be.

Lemma distinctness would explain regular-irregular homophones by the following mechanism. Upon hearing a new verb that is homophonous with an existing verb, learners create a separate lemma for it in their mental lexicons, and append an irregular form to it (if they hear one) or allow it to be inflected with a regular suffix (if they don't). On this account, the semantic representations of the two verbs play no role in the speaker's decision as to how to inflect it. As a result, as the language develops over history, there is no tendency for semantically similar verbs to have similar past-tense forms. This phenomenon may be seen in pairs such as *lie-lay* and *lie-lied*, *ring-rang* and *wring-wrung*, and *meet-met* and *mete-meted*, in which the members have different past-tense forms, but do not belong to families of semantically similar verbs with similar past-tense forms.³ On this hypothesis, then, the effects of lexical semantics are very different from those of phonology, where learners *do* attend to a verb's phonological representation when deciding how to inflect it, resulting in a tendency to generalize the past-tense form of an existing verb to a new, phonologically similar one, and the development over history of clusters of phonologically similar verbs with similar past-tense forms. (In contrast, on the Semantic Association hypothesis, the effects of lexical semantics would be closely parallel to those of phonology, with gradients of generalization in both cases.)

3. *Morphological headlessness*. Pinker and Prince argued that a larger class of regular-irregular homophones arise as effects of differences in their grammatical structure, in particular when one of them is *exocentric* or *headless*. They noted that in most regular-irregular homophone pairs, the regular member is a denominal or deadjectival verb or participle (Kiparsky, 1982):

- (1) She grandstanded/**grandstood* during the debate [play to the *grandstand*].
 Geoffrion got high-sticked/**high-stuck* [hit with a high *stick*].
 Powell ringed/**rang* the city with artillery [formed a *ring* around].
 Vera costed/**cost* out the equipment requirements [ascertained the *cost* of].
 Mongo spitted/**spat* the pig [put on a *spit*].
 Vernon braked/**broke* for the moose [applied the *brakes*].
 Swans are dark-meated/**dark/met* fowl [having dark *meat*].
 Martina two-setted/**two-set* Chris [beat in two *sets*].
 Ortiz singled in the first and flied/**flew* in the fourth [hit a *fly*].
 After you've meaned/**meant* both columns, you can do the t-test [computed the mean].
 Mom was flying home in a box, to be waked/**woke* and buried [given a *wake*].
 Most snow or sugar snap peas need to be stringed/**strung* [have the *string* removed].
- (Kiparsky, 1982; Pinker, 1999: chap. 6)

According to the headlessness theory, this effect falls out of the mechanism governing the way that complex words are assembled from morphemes, namely the right-hand head rule (di Sciullo & Williams, 1987; Lieber, 1980; Selkirk, 1982). In English, a complex word ordinarily inherits its features from its rightmost morpheme, its head. For example, suppose the novel verb *to unring* is coined for the saying, *Once a bell is rung, it cannot be unringed*. The head of the novel verb *unring* is *ring*. As a consequence, *unring* is a verb (it inherits the "verb" category information from *ring*), it refers to the act of ringing, namely reversing its effects (because it

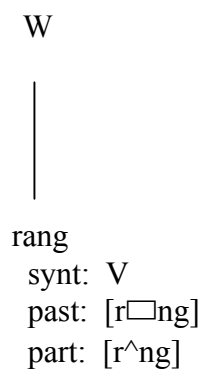
³ Although homophonous lemmas are often distinguished by their orthography, as in *meet/mete*, the example of *lie/lie* shows that they need not be.

inherits the semantic features of *ring*), and it has the irregular participle form *unrung* (because it inherits the participle form *rung* which is stored with the root *ring*).

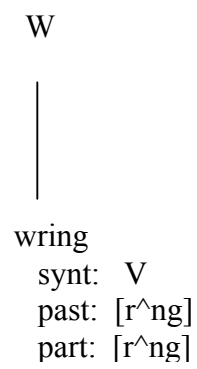
But in a small family of exceptions—headless or exocentric words—this mechanism is disabled. For example, some words have a different grammatical category from that of their rightmost morpheme. Denominal verbs, in particular, are verbs based on nouns (e.g., *to ring the city*, based on the noun *a ring*). In this regard, they contrast with conventional endocentric verbs like *unringt*, which are based on verbs. For that to be possible, the right-hand inheritance mechanism must be disabled. As a result, any irregular form stored in the lexicon cannot be applied to the word as a whole, and instead the regular suffix is applied by default (Marcus et al., 1995; Pinker, 1999).

The diagrams in (2) illustrate in simplified form the morphological structures for the kinds of verbs discussed so far. 2(a) and (b) show a pair of irregular words, *rang* and *wrung*, which belong to distinct lemmas: each word node annotated with its syntactic category and its irregular past and participle forms. 2(c) shows a complex derived verb (*unring*) headed by an irregular verb (c). The action of the right-hand-head rule is indicated by the dashed arrow, which symbolizes the copying of the syntactic category, semantic representation (not shown), and irregular forms to the node representing the whole complex word. 2(d) shows a derived verb (*ring*, in the sense of “form a ring around”) which is headed by a noun (*a ring*). The fact that the N category label cannot be copied from the root to the complex word (because the root is a noun but the complex word must be a verb) means that the right-hand-head-rule must be disabled (indicated by the “X”), making any sound associated with *ring* (such as *rang* or *wrung*) unavailable. The structures in 2(c-d) are what represents a speaker’s intuition that one word contains, or is based on, another word, such as the sense that the verb in *to ring the city* is based on the noun *a ring*. They will apply (or, if necessary, avoid applying) the right-hand-head rule when they coin a new word themselves, or when they hear it only in one form (say, the present tense) and have to generate another one (say, the past). They may also consult the structure and rule when reflecting on the word in the course of judging the grammaticality of various possible usages. On the other hand, if a speaker lacks any sense that a verb is based on a noun—say, they have no intuition that the baseball term *to fly* means “to hit a fly”—then they will represent the verb with a structure like 2a or 2b, and will stick with the irregular, as in *He flew out in the third*, unless they hear evidence to the contrary from other speakers.

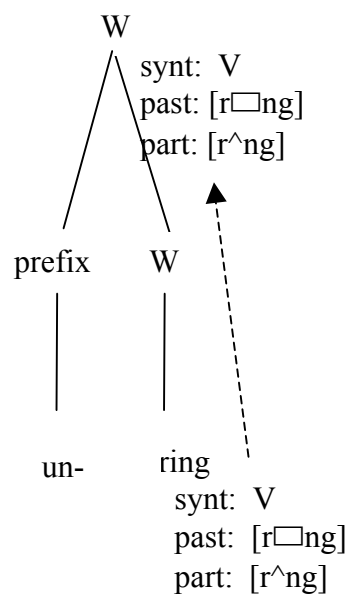
(2)



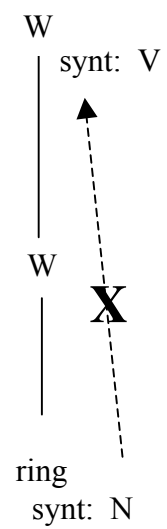
(a)



(b)



(c)



(d)

Crucially, the headlessness effect predicts that regular-irregular homophones may have different meanings but that this difference is not the *cause* of the inflectional difference, but an epiphenomenon of the difference in headedness. Thus an effect of morphological headlessness may be distinguished from an effect of semantic association as follows: in pairs of homophones with equally different meanings, but in which both members are headed by the same verb (such as in cases of polysemy, metaphor, metonymy, and idiom, all of which change meaning without requiring headlessness), the members should not, all things being equal, diverge in their past-tense forms.

4. *Aspectual compositionality*. Bolinger (1980) and Pinker and Prince (1988) call attention to polysemous verbs in which an irregular form is not perceived as categorically ungrammatical (such as in (1)) but coexists with a regularized version in a doublet, as in *dreamt* and *dreamed*. Sometimes the sense of a verb that can appear in such doublets varies subtly from the sense that can appear in a single past-tense form.⁴

- (2)
- a. The picture hung/*hanged on the wall.
The prisoner ?hung/hanged for his crime.
 - b. The shoes fit/*fitted, so I wore them.
Dr. Shoicket fit/fitted me with dentures.
 - c. Just before the alarm went off, I dreamed/dreamt that I forgot my suitcase on the train.
After taking the Ambien, I dreamed/?dreamt happily for hours
 - d. He pleaded/?pled with me to give him the part.
The plea bargain was his best option, so he pleaded/pled guilty.
 - e. The idiot on the cell phone weaved/wove in and out of traffic
Sy ?weaved/wove three baskets during his stint in therapy
 - f. Carolyn always ?shined/shone at ribbon-cutting ceremonies

⁴ We have corroborated these judgments as follows. The *American Heritage Dictionary* (2000) tests many linguistic judgments in the semiannual survey of its Usage Panel of writers, journalists, and academics (Nunberg, 1989), currently chaired by the second author. Data corroborating the distinction in 2(a) are reported in the Fourth Edition of the dictionary, which also reports the distinction in 2(b). Sentences 2(c-g) were included in the 2008 Usage Panel survey ($n = 140$). In every case the ratings displayed the predicted interaction between sense and relative preference for the regular or irregular. An error in the survey question made the responses to *knit* (2g) unusable, but a Google search shows an even split of hits for the quoted strings *knitted his brows* and *knit his brows*, but a fifteenfold advantage for *knit a sweater* over *knitted a sweater*.

Mel shined/?shone the light at the snake.

g. He knitted/knit his brows.

He ?knitted/knit a sweater.

Interestingly, in all these cases, the pairs of senses differ in a similar way: in their inherent lexical aspect, event structure, or *Aktionsart*, namely how the event unfolds in time (Dowty, 1979/1991; Sasse, 2002; Vendler, 1957; Pinker, 2007). Inherent aspect varies along two major dimensions: *durativity* (whether an action is instantaneous or protracted) and *telicity* (whether the action is terminated by a specified change of state or can go on indefinitely). In 2(a), the first sense is durative, the second instantaneous. In 2(b-g), the first member is atelic, without a specific terminating event, whereas the second member is telic, ending with a change of state, usually of the referent of the direct object.⁵

Why might a difference in aspect trigger a difference in irregular inflection? A plausible explanation invokes the interaction between the inherent meaning of the verb and the meaning of the past tense. Since regular forms can be generated or analyzed by an affixation process, their meanings are composed additively of the meaning of the verb stem and the concept of pastness. Irregular forms, in contrast, must be stored in a distinct entry because of their unpredictable sounds. This defines a slot that can accommodate a semantic gestalt which is idiosyncratic to the referent of that verb when it is used in a typical past-tense context, rather than the simple addition of “pastness” to its base meaning.⁶ Now in theory, adding the past tense to a verb’s meaning should not result in an idiosyncratic combination, because tense and aspect are semantically independent (Comrie, 1985; Sasse, 2002): anything that can exist or happen now could have existed or happened the same way in the past. In practice, however, the two can interact. Most telic verbs result in a changed state, so when the event has happened in the past, the state by default persists to the moment of speaking, altering the perceived semantics of the past-tense form. For example, if a person truthfully said *The glass broke* (telic), then at the time he says it the glass is no longer intact. In contrast, if he were to say *A dog barked* or *A man jogged* (atelic), no physical state of affairs need be in place at the time for the statement to be true.

The upshot is that when an irregular verb is extended to a sense that has a different inherent aspect, speakers may perceive a clash between the aspect-specific meaning they have stored with the irregular in its original sense (e.g., a permanently suspended object for *hang*, an extended court case for *plead*, the existence of a woven object for *weave*, a well-fitting appurtenance for *fit*) and the meaning of the verb in its new sense, which need only refer to an event that took place in the past, regardless of its consequences. This may be enough to taint the

⁵ In general, telic senses often go with the transitive member of a transitive-intransitive alternation. Though often it is the more telic sense admits the doublet forms, and the atelic sense that is choosier, the example of *shined/shone* shows that this is not always the case. Similarly, though there is a tendency for the more atelic sense to favor the regular form (as in *dreamed*, *pleaded*, *knitted* and *weaved*), with *fitted* and *hanged* it is the other way around.

⁶ The explanation would be parallel to a long-noticed effect in derivational morphology: productive alternations (such as suffixing *-ity*) tend to be semantically predictable (e.g., *liquid-liquidity*), whereas unproductive alternations (such as *-id* and *-or*) may accumulate specialized meanings (e.g., *liquid-liquor*) (Aronoff, 1976).

application of the irregular form to that new sense, and thus to tilt speakers toward the regular. This is especially likely to happen if the irregular is low in frequency, a situation that tends to allow a regular counterpart to coexist with the irregular (Ullman, 1999). When the regular form is available in the speaker's experience, it can be pressed into service as a way of expressing a meaning that does not sit well with the stored irregular. Note that this effect of aspect would pertain not to the semantics of a word as a whole (as in semantic associative effects) but only to the subset of semantic features that interact with the semantics of the inflection. In the case of the past-tense, this consists of inherent aspect since both tense and aspect pertain to time.

Previous Studies of Semantic Effects on Irregularity. Kim et al. (1991) was the first in a series of behavioral studies aimed at distinguishing the different kinds of semantic effects on the choice of a regular or irregular inflectional form. In several questionnaires, they showed that participants strongly preferred regular past-tense versions of homophones of irregular verbs when those verbs were based on nouns (that is, had a headless morphological structure). This was true both with existing denominals like those in (1), and with novel ones like *to sink* "put something in the sink" and *to drink* "supply someone with drinks." To show that this was a consequence of morphological structure and not global semantic dissimilarity, they also presented extended, metaphorical versions of the verbs, such as *sink* in the context "my hopes sank" and *drink up* in the context "drink up the gossip." In these cases, the original *irregular* form was preferred. To ensure that this difference did not arise from the denominal senses being more semantically extended than the metaphorical senses (relative to the original verb), Kim et al. asked another group of participants to rate the degree of semantic extendedness of both kinds of items. They found that headlessness predicted a significant portion of the variance in the ratings of regular versus irregular forms, whereas semantic extendedness did not. Similar results have been recently found by Bandi-Rao and Murphy (2007) and Gordon & Miozzo (2008). They varied semantic extendedness by testing senses of polysemous irregular verbs that are perceived to be highly extended, such as *His pals will spring him from prison*, and compared them to denominal verbs that are not perceived to be as extended, such as *He costed the job at \$100* ("ascertained the cost of").

While these results suggest that morphological headlessness, not semantic association, is the cause of regularization, that conclusion has been challenged in two studies. Ramsar (2002, 2003) asked speakers for the past-tense forms of a novel verb that was semantically and phonologically similar either to a regular or an irregular verb. Subjects saw a sentence where *frink* meant either "eyelids opening and closing rapidly and uncontrollably" (similar to *blink*) or "consuming vast quantities of vodka and pickled fish" (similar to *drink*). He found that when *frink* was semantically similar to *blink*, most participants preferred the regular *frinked* (73%), but when it was similar to *drink*, most preferred the irregular *frank* (77%). In another study, he attempted a replication of Kim et al.'s regression analyses with certain procedural changes, and found that denominal status no longer predicted the regular and irregular ratings when semantic extendedness was held constant.

Another challenge came from Baayen and Moscoso del Prado Martin (2005), who sorted the monomorphemic verbs listed in the English CELEX corpus into sets of synonyms using the Wordnet database (Miller & Fellbaum, 1991). They found that irregular verbs were more likely to appear in synonym sets than were regular verbs, and that synonym sets tended to contain greater numbers of irregular verbs. They also found that in English, Dutch, and German, irregulars have different classwise semantic properties than regulars: they are more polysemous,

more likely to take the auxiliary *zijn* or *sein* “be” (which correlate with telicity), more likely to fall into several argument-structure subclasses (as in Levin, 1993; Pinker, 1989), and more likely to have a greater number of semantic associates (largely because they are higher in frequency, frequency correlates with polysemy and with number of semantic associates). The authors concluded that irregulars tend to cluster in a denser semantic space than regulars, which is consistent with the idea that people tend to generalize inflectional patterns among verbs with similar meanings. Similar patterns are reported in a statistical analysis of Dutch vocabulary, and in a visual lexical decision experiment, by Tabak, Schreuder, and Baayen (2005). They also found an effect consistent with the morphological structure hypothesis: verbs with homophonous nouns that were higher in frequency (suggesting that the verb was based on the noun) were more likely to be regular, holding other factors statistically constant.

How might we reconcile these discrepancies? We suggest that they arise in part from inconsistent characterizations of “semantics,” which can lead to confounded materials and contradictory interpretations. In Ramscar (2002), each of the novel words was so similar (both in sound and meaning) to an existing verb that participants may have assumed that they were really being asked for the correct past-tense forms of *drink* and *blink*. If so, the study would have been tapping a lemma-based effect, rather than a semantic associative one.⁷

Ramscar’s regression study may have failed to distinguish associative effects from headlessness effects as well. His measure of headlessness consisted of asking subjects to rate the extent to which verbs were “being used in a normal ‘verblike’ way” or “as a verb in relation to a noun.” The latter end of the rating scale was illustrated with an invented example of a denominal verb, specifically, the verb *to fly* being used in the sense “to greet customers in a fly costume.” Since these instructions contrasted headless with “normal” usages, and explained headlessness only via a second-order metalinguistic judgment (as opposed to asking whether a specific verb was based on a specific noun), participants may have based their judgments on semantic familiarity or similarity rather than on headlessness per se. This concern is reinforced by anomalous findings in the study, such as participants indicating that they perceived the verb *to brake* as no more closely related to a noun than the verb *to break*, despite the transparent relation of *to brake* with *brakes*. (See Bandi-Rao & Murphy (2007) and Gordon & Miozzo (2008) for discussions of other methodological problems with this study.)

Similarly, while the studies by Baayen and Moscoso del Prado Martin and by Tabak, Schreuder, and Baayen employ a large and well characterized lexical database and considerable statistical sophistication, these any vocabulary analysis is necessarily a correlational study and hence is limited by the natural confounds inherent to vocabularies that develop historically. In general, irregular verbs may be more polysemous (and have related class-wise semantic properties) because of higher token frequencies in the historical periods that shaped the language,

⁷ This effect may have been enhanced by the fact that “to consume vast quantities of vodka and pickled fish” is probably not a lexicalizable concept. Its meaning is unlike those of any of the English denominal verb types analyzed in Clark & Clark’s (1979) extensive taxonomy. Denominal verbs specify their incorporated objects generically. For example, *to butter* does not require actual butter; one can say *He buttered his toast with cheap margarine* (Carter, 1976; Talmy, 1985; Pinker, 1989; Jackendoff, 2008). Ramscar’s novel verb, in contrast, specifies a conjunction of two precisely specified substances. This semantic unnaturalness may have encouraged participants to consider the nearest legitimate existing verb.

greater familiarity in ways not measured by corpus frequencies, or earlier age of acquisition. Regular verbs may have fewer and more distant regular semantic neighbors because their homophonous forms are more likely to be perceived as denominal, more likely to be phonologically distant, or more likely to have entered the language at a different historical period. In the particular case of Dutch, Tabak et al. noted that a verb's semantic density was confounded with the number of its argument structures (which, as we noted, may be confounded with aspectual differences, especially the transitive-intransitive distinction). Most importantly, many of these are confounded with the age of the word in the language's history, which in turn is confounded with its etymology, which in turn is partly confounded with prosody. As a result, when the authors entered etymological age into their regressions, they no longer reported effects of semantic density, number of argument structures, or relative frequency of the verb and its homophonous nouns. They did find an effect of auxiliary selection, which is confounded with aspectual properties. Thus, while these correlational analyses are valuable, it is desirable to complement them with experimental manipulations of the lexical materials, and with tasks that measure the relevant independent and dependent variables in contemporary speakers.

The experiments in this paper are designed to help resolve these uncertainties. Experiment 1 distinguishes between lemma-based and semantic associative effects in a factorial design that varies semantic and phonological similarity of novel items to families of existing verbs, and independently assesses speakers' analogizing to specific lexical items. Experiment 2 aims to disentangle semantic extendedness from headless morphological structure by replicating the regression methods of Kim and colleagues (1991), Ramscar (2002), Bandi-Rao and Murphy (2007), and Gordon and Miozzo (2008) using a larger sample of verbs extending over a larger range of semantic extendedness, and using controlled and antecedently motivated measures of semantic similarity. Experiment 3 inverts the methodology, and uses the contrast between existing regular and irregular verbs as the dependent variable, and participants' ratings of semantic similarity, headless status, and several other lexical variables as predictors. This is an indirect way of examining possible psycholinguistic effects on the historical developments that shaped the language. Experiment 4 tests for aspectual compositionality by contrasting the aspectual similarity and the global semantic similarity of the regular and irregular senses of doublet verbs.

Experiment 1

This study asks whether the apparent effects of semantic similarity on irregular inflection reported by Ramscar (2002) may instead be lemma effects. Ramscar's design relied on the fact that while most English verbs with the rime *-ink* are irregular (e.g., *drink*, *shrink*, *stink*), the two regular exceptions—*blink* and *wink*—are semantically similar to each other. In his first study, people's ratings of novel verbs reflected this fact: a novel verb (*frink*) was given a regular form when it was semantically similar to *blink* but an irregular form when it was similar to *drink*. But as mentioned, these experimental items were nearly identical to existing verbs in phonology, semantics, and grammar that participants may have directly mapped the item to an existing lexical lemma, rather than being sensitive to semantic overlap. This could have happened because of demand characteristics, in which participants (half of whom were passersby in a shopping mall) assumed they were being asked for the correct forms of the existing verbs, or as a result of lexical priming.

In this study, we designed novel verbs varying in phonological and semantic similarity to existing irregular verbs to assess whether there is an effect of semantic associative similarity that corresponds to the well-documented effects of phonological similarity. We compare it to an alternative in which semantic similarity only comes into play when both semantic and phonological similarity are simultaneously high, because this combination reminds speakers of the lemma of an existing verb. Crucially, on this alternative, the predicted interaction should correlate with participants' reporting that they explicitly thought of an existing irregular verb similar to the one they are rating and that they based their rating on an analogy to it.

Method

Participants. Seventy-two native English-speaking Harvard undergraduates were given course credit for their participation.

Materials. Test sentences contained novel verbs that orthogonally varied in phonological and semantic similarity to existing irregular verbs. The forms of the novel verbs were taken from Prasada and Pinker (1993), and were based on clusters of existing verbs with similar irregular past-tense forms. The meanings were adapted from specific verbs within each of those clusters: *swing*, *sink*, and *cling* from the *ing-ung* ablaut class; *lead* and *read* from the vowel-laxing cluster, *blow* and *throw* from the *ow-ew* cluster; and *bear* from the E-O ablaut class (Pinker & Prince, 1988).

Phonological triplets were created using operations specified by Bybee and Moder (1983) and Prasada and Pinker (1993). The prototype in each triplet (e.g., *froe*) was based on a family of similar irregulars (in this case, *throw*, *blow*, *grow*, and so on), and shared both a rime and a typical onset cluster with the verbs in the family. The spelling of these items was altered to ensure that the intended pronunciation was clear to the participants, in this case, that they would not think that the verb rhymed with *brow* and *prow*. The two other verbs differed from the prototype in the initial or final consonant (e.g., *poe*) or in both consonants (e.g., *joam*). The resulting triplets were *spling-splung*, *fring-frung*, *trisp-trusp*; *sprink-sprunk*, *frink-frunk*, *blip-blup*; *cleed-cled*, *cleef-clef*, *gleef-glef*; *clow-clew*, *zow-zew*, *goav-goov*; *plare-plore*, *jare-jore*, *flape-flop*; *froe-frew*, *voe-voo*, *joam-joom*; *preed-pred*, *preek-prek*, *keeb-keb*; and *skring-skrung*, *ning-nung*, *nist-nust*.

Semantic similarity (High, Moderate, Low) was varied by altering the semantic domain of an existing irregular verb and the goal of its agent. High-similarity meanings were near-synonyms of the verb, such as *bear* (see 3a). Moderate-similarity meanings, such as 3b, were in the same semantic domain but differed in the agent's goal, generally involving an opposite effect. Low-similarity meanings, such as 3c, differed in both domain and goal.

Though we generated the stimulus materials using these *a priori* manipulations, the independent variable in our analyses is based on empirically determined assessments of phonological and semantic similarity, elicited from a separate group of 24 participants. For phonological judgments, participants were asked to rate the similarity between the sounds of the target (e.g., *throw*) and a novel verb (e.g., *froe*) on a 7-point scale, where 1 was defined as "very dissimilar" and 7 as "very similar." For semantic judgments, they were asked to rate how associated the meaning of the target (e.g., *to block*, a semantic alteration of *throw*) was to the relevant known verb (in this case, *throw*) on a 7-point scale, where 1 was defined as "not at all associated" and 7 as "very associated." To ensure that a single rater was presented with only one

pairing of target verb and a novel or known verb, the items were divided into six counterbalanced lists of 16 each, and each participant was given a questionnaire with one of these lists. As it turned out, the order of the mean ratings of semantic similarity always matched our prior assignment of High, Medium, and Low similarity. The order of the mean ratings of phonological similarity differed for three of the eight verbs (*blow*, *cling*, and *throw*), where the items we had constructed as medium-similarity were judged as the most similar. These items were reassigned accordingly.

In the main experiment, each verb was introduced with a context sentence that revealed its meaning. This was followed by two test sentences, one presenting the verb in a regular past-tense form, the other in an irregular form. Novel verbs were underlined in the context and test sentences.

- (3) a. **HIGH SEMANTIC SIMILARITY:** Mules and horses have been bred to plare heavy burdens.
 IRREGULAR: This was a necessity when settlers moved west and the animals plore the weight of their belongings for months on end.
 REGULAR: This was a necessity when settlers moved west and the animals plared the weight of their belongings for months on end.
- b. **MODERATE SEMANTIC SIMILARITY:** After moving, it took Sally a long time to plare all the boxes of books and clothes. Every item had to be put in the right spot on the right shelf.
 IRREGULAR: After several weeks of work, Sally finally plore the last box.
 REGULAR: After several weeks of work, Sally finally plared the last box.
- c. **LOW SEMANTIC SIMILARITY:** The annual Pie-Off is a contest to see who can plare the most pies in a fixed period of time.
 IRREGULAR: Last year's winner ended up in the hospital after he plore twenty pies in ten minutes.
 REGULAR: Last year's winner ended up in the hospital after he plared twenty pies in ten minutes.

The phonological and semantic similarities were crossed to create nine types of verbs, which were crossed with the eight verb families to yield 72 items. Since we could not present a given verb sound more than once to a single participant, and did not want to make the independent variables obvious by presenting items that directly contrasted in semantic or phonological similarity, we divided the items into three counterbalanced lists of 24, each containing eight items apiece in three of the nine possible conditions. That is, while every participant rated items with high, medium, and low phonological similarity, and items with high, low, and medium semantic similarity, no participant saw a pair of items varying only in phonological similarity or only in semantic similarity. A complete list of the materials and ratings for all experiments may be obtained from the first author.

Procedure. Participants were asked to rate the acceptability of the irregular and regular forms on a 7-point scale, where 1 was defined as “very unnatural” and 7 as “very natural.” The scale was explained using irregular verbs that did not appear in the study: *He came/comed home to Boston* as an example of a strongly irregular verb, *She dreamt/dreamed that she was falling out of a plane* as an illustration of a verb with an acceptable regular alternative. Participants were told that these examples show that it was possible to give a high rating to one verb form

and a low rating to the other, a high rating to both, or low ratings to both. Participants were told to “pay attention to the way the new verb is used in the example” and to base their judgments on “how natural the verbs feels to you rather than on your guess as to what is ‘proper’ or ‘correct’.”

After participants completed the ratings, they were asked to go back and indicate how they made their judgments. They were presented with four possibilities (4) and asked to pick as many as were applicable, or to indicate their own strategy if it was not on the list:

- (4) a. The novel word reminded of a specific word I already knew, so I simply borrowed the past-tense form from that verb. If so, please indicate which verb you had in mind.
- b. The meaning of the novel word made one form seem better than the other.
- c. The sound of the novel word made one form seem better than the other.
- d. I didn’t really think of any particular strategy or reason for my choice: one of the past-tense forms just seemed better than the other
- e. Other, please indicate.

Strategies were elicited only after the rating task so that our listing of the strategies would not artifactually encourage participants to use them. Though several options were provided to avoid further biasing the participant, we were interested only in the first justification, since it tested the possibility that the combination of high phonological and semantic similarity would lead participants to analogize a specific known verb to the test item. We assessed the reliability of the reports in a split-half test: for each participant and each condition he or she participated in, we compared the most frequent strategy the participant reported (e.g., “known word”) for the even-numbered and the odd-numbered items. These were the same in 81% of the comparisons, suggesting that the preferred strategy, even though it was recalled after the fact, was consistent across items within a condition. .

Results

We examined the tendency to irregularize a verb by subtracting the mean rating of its regular form from the mean rating of its irregular form. For each trial, this difference score could range from -6 (very unnatural irregular /very natural regular) to 6 (very natural irregular /very unnatural regular) and were averaged over participants for each item. These difference scores were entered into a two-way Analysis of Variance with Phonological Similarity, Semantic Similarity, and List as fixed effects. Because each participant was presented with only three of the nine conditions in the factorial design, the random effect in all the analyses is Items.

The mean preferences are presented in Figure 1. The effects of Phonological Similarity are seen in the differences between the three lines: as in Bybee and Moder (1983) and Prasada and Pinker (1993), participants showed a strong tendency to prefer irregular forms for items that were more phonologically similar to existing irregular verbs, $F(2, 10) = 58.70, p < .001$.⁸

INSERT FIGURE 1 HERE

The ANOVA also revealed a main effect of Semantic Similarity, $F(2, 20) = 5.85, p < .05$. However, visual inspection suggests that it is confined to the condition of High Phonological

⁸ The ten degrees of freedom in the denominator reflect the way that the eight items in each condition are embedded in the three lists, namely, three items in each of two lists and two in the other.

Similarity, and even then only to the difference between High and Moderate Semantic Similarity. This is confirmed by a significant interaction between Phonological Similarity and Semantic Similarity, $F(4, 20) = 7.44, p < .01$, and tests of simple main effects for Semantic Similarity that are significant only with High Phonological Similarity, comprising significant differences between the conditions for High Similarity (3.8) and Low (0.8, $t(7) = 8.01, p < .01$) and Moderate Similarity (0.5, $t(7) = 3.83, p < .01$), but no difference between the latter two conditions ($t(7) = 0.44, p > .60$).

The pattern of ratings is consistent with the hypothesis that semantic similarity effects on generalization of irregular patterns are lemma effects rather than associative effects: semantic similarity matters only when it makes a novel item so similar to an existing irregular verb that people treat it as the equivalent of that verb. To verify this explanation, we compared the pattern of ratings to participants' reported strategies, in particular, how often they based a response on the exact existing irregular word we had used to construct the item. We coded the tendency to analogize as 1 if the participant reported basing his or her response on a similar word, and provided the exact word we used to construct the item, 0 otherwise. Figure 2 indicates that participants were much more likely to use an analogy to the target verb when the novel verb simultaneously had High Semantic Similarity and High Phonological Similarity to known verbs.

INSERT FIGURE 2 HERE

Finally, we tested whether analogy to an existing lemma accounts for the interaction between semantic and phonological similarity seen in the past-tense judgments. We again averaged over participant ratings for each item and performed a multiple regression on the irregular-regular difference scores, with Phonological Similarity (three levels), Semantic Similarity (three levels), their interaction, and reports of analogy to a known word as regressors (see Table 1). With the use of analogy statistically held constant, there remains no effect of Semantic Similarity, and interaction between Semantic and Phonological Similarity disappears, whereas the effect of Phonological Similarity survives. In contrast, the use of the analogy strategy is significant when the other factors are held constant. The same tests can be done as an Analysis of Covariance with the two kinds of similarity as fixed effects and reports of analogy as the covariate (see Figure 3). While there remains a strong main effect of Phonological Similarity, there is no effect of Semantic Similarity, and the interaction between Phonological and Semantic Similarity essentially disappears. This suggests that semantic similarity has an effect on past-tense judgments only to the extent that it (in conjunction with phonological similarity) calls to mind an existing irregular verb.

INSERT TABLE 1 AND FIGURE 3 HERE

Discussion

The results of the experiment suggest that phonological similarity and semantic similarity have different effects on past-tense judgments. Replicating Bybee and Moder (1983), Prasada and Pinker (1993), and others, we find a robust generalization gradient for the phonological similarity of a test word to existing irregular verbs, which exists across all levels of semantic similarity. However, the effects of semantic similarity are significant only when phonological similarity is simultaneously high. By itself this interaction between phonological and semantic similarity cannot rule out effects of semantic association, since it is possible that nonlinearities in an associative model of inflection, in conjunction with different statistical clustering of

phonological and semantic properties of items in its training set, could duplicate such an interaction.

Nevertheless, these findings constrain any account of semantic effects on inflectional morphological in two critical ways. First, they demonstrate that while generalization based on phonology varies in a monotonic relationship to the number of shared features, generalization based on semantics clearly does not. Second, they show that semantic effects are largely, if not entirely, accounted for by the tendency of known words to enter people's consciousness and be used by them explicitly as a lexical analogy. These analogical mechanisms may be similar to the meta-cognitive reasoning processes that support generalization of facts across domains of knowledge (Gentner & Markman, 1997; Goldstone & Sakamoto, 2003). Altogether, these findings are most consistent with the hypothesis that lemma distinctness is the major determinant of regular-irregular homophony and suggest that the apparent semantic generalizations of irregular patterns in Ramscar's (2002) *drink/blink* experiment are lemma effects rather than semantic effects.

Experiment 2

This experiment pits semantic similarity against morphological structure. We assess whether the regularization of denominal headless verbs (as in *ringed the city*) is a by-product of the fact that denominals tend to be more semantically dissimilar from a homophonous irregular (e.g., *rang the bell*) than are extended senses (e.g., *rang the alarm on government corruption*). It may be the case that people associate an irregular form with the phonological and semantic features of an irregular verb stem, and when they are confronted with a new sense that is sufficiently dissimilar from an old one, the semantic associations would be diluted, and the more statistically pervasive regular pattern would win out. If this is true, headlessness itself would play no role.

Adapting the design from Kim and colleagues (1991), we presented participants with a prototypical sense of an irregular verb (e.g. *throw the ball*), as well as with a novel headless version derived from a noun (e.g. *throw the couch* = *cover with a throw* [a kind of blanket]), and also with two less prototypical, extended senses (e.g. *throw a party*, *throw the game*). As mentioned, three sets of studies have used such a design, with Ramscar's (2002) conclusions differing from those of Kim et al. (1991), Bandi-Rao and Murphy (2007), and Gordon and Miozzo (2008). There are several possible reasons for the discrepancies. One is the small number of items used in previous studies, ranging from 2 to 13. A second could be participants' failure to analyze some of the existing verbs as headless. For example, even if the experimenters had classified it that way the verb *to lie* as a denominal based on the noun *a lie*, participants may not have thought of it that way. A third possibility could be a lack of clarity on participants' part as to what they were supposed to be rating. If they were unsure of what it meant for a verb to be based on a noun, their judgments could have been contaminated by global similarity; conversely, if they were unclear about semantic similarity, the syntactic difference between a denominal and an extended-sense verb could have contaminated their ratings. To address these limitations, we used a larger sample of verbs (23), tested two extended senses rather than one, and provided explicit, antecedently motivated criteria to manipulate and measure semantic extendedness.

If inflectional morphology is determined by semantic similarity, the degree of overlap in meaning between a novel sense and the prototypical sense should predict peoples' preference for

a regular or irregular form, when headedness is held constant. If it is determined by headedness, then the regression analyses should reveal the opposite pattern.

Method

Participants. Twenty native English-speaking Harvard undergraduates completed Part 1 (past-tense judgments) and Part 2 (semantic judgments) for course credit.

Materials. Participants saw 92 sentences that used four variations of each of 23 polysemous irregular verbs. The Prototypical sense of a word corresponded to its first definition in the second edition of the unabridged *Oxford English Dictionary* (5a).⁹ The two Extended senses were taken from the two subsequent definitions (5b and 5c). Though the dictionary entries helped us select the items, the actual values of the extendedness variable used in the analyses came from participants' ratings, as explained below. Finally, each item appeared in a denominal sense that was based on a homophonous noun (5d). All items were introduced by a context sentence and were followed by two test sentences, which presented the verb in a regular and an irregular past-tense form. For half the items, participants saw the irregular form first; for the other half, the regular form. The verb to be rated was underlined in the context and test sentences.

- (5) a. **PROTOTYPE SENSE:** My lifelong dream is to throw the first pitch at a Cubs game. I am getting in practice—last year,
 IRREGULAR: I threw the first pitch for my brother's little league game.
 REGULAR: I threwed the first pitch for my brother's little league game.
- b. **EXTENDED SENSE:** I'm famous for throwing lively parties on weekends. Last Saturday, the party
 IRREGULAR: I threw got so loud that the police came to break it up.
 REGULAR: I threwed got so loud that the police came to break it up.
- c. **EXTENDED SENSE:** There are severe penalties in college basketball for throwing games to benefit gamblers. Charlie was banned from the game for life when
 IRREGULAR: He threw a key game last month.
 REGULAR: He threwed a key game last month.
- d. **DENOMINAL SENSE:** The couch was so ugly that I needed some way of hiding it with a throw. I found what I needed at IKEA, and as soon as I got home,
 IRREGULAR: I threw the couch, concealing the fabric perfectly.
 REGULAR: I threwed the couch, concealing the fabric perfectly.

Procedure. Participants first read sentences like those in (5) and rated them on a scale from 1 ("very unnatural") to 7 ("very natural"), which was explained with the instructions used in Experiment 1. In a second questionnaire, participants were presented with the same materials but were asked to rate the semantic similarities among the different senses. They first read the context sentence introducing the prototypical sense (5a) and then rated the semantic similarity of the extended (5b and 5c) and denominal senses (5d) with respect to it. Presentation senses was

⁹ Though the OED orders its listed senses historically, the order is highly correlated with that of other dictionaries such as the *American Heritage Dictionary* ($r = .87$ for our materials), which order their senses mainly according to frequency (personal communication from Joe Pickett, Editor of the *American Heritage Dictionary*, July 14, 2008).

randomized across verbs such that an equal number of each type appeared first over the course of the experiment. Each item was judged on a scale of 1 (“very dissimilar”) to 7 (“very similar”).

Since the purpose of the experiment was to distinguish the ways in which semantics might affect inflection, it was critical to obtain a measure of semantic similarity that specifically taps the semantic representation of the verb, rather than other lexical factors such as its derivation, syntax, morphology, or history. That is, we did not want participants to rate *sink the beer mugs* (“put in the sink”) as different from *the ship sank* simply because the first is unfamiliar, based on a noun, or different in its past-tense form; we wanted the rating to focus on its referring to a different kind of event. Since denominal verbs necessarily differ in meaning from a homophonous irregular, complete statistical independence of semantics and headlessness is impossible, but a suitable measure should yield enough unconfounded variance for it to be possible to differentiate their effects in regression analyses. Thus the instructions directed participants’ attention to the real-world referents of the verb sense, rather than to the verb as a lexical item. The participants were told that the same verb can be used in slightly different senses and that senses share a similar meaning “to the extent that they refer to the same kinds of things and actions in the world.”

To focus participants’ attention on the linguistically relevant things and actions in the world, we provided a list of diverse semantic features such as the actor’s intentions, the action’s manner, the nature of the object affected, and the resulting change of state (see Pinker, 1989). To ensure that the features were maximally relevant to theories that try to explain irregular inflection by semantic association, they were taken from MacWhinney and Leinbach’s (1991) connectionist model of the past tense, which successfully distinguished the common regular-irregular homophones in English. The features were presented as 20 questions about the action denoted by the verb, such as whether it began abruptly, resulted in a sound, involved a flexible object, involved use of the hands, was voluntary, referred to motion in a vertical direction, or referred to liquid being added or removed. Participants were told that usages should be rated as similar “if the two senses shared most or all of their aspects of meaning like these,” and dissimilar “if one of the senses had some of these aspects of meaning, and the other one had a different set.” The feature list was included only to draw participants’ attention to the real-world referents of the verb pairs; participants were not asked to rate the applicability each feature to each verb.

Results

As in Experiment 1, we calculated the tendency to irregularize a verb as the difference between the mean ratings of its regular and irregular forms. These scores were averaged within each item and were analyzed in a one-way analysis of variance with the four levels of semantic extendedness. The mean irregularity scores were 4.8 (SD 1.0) for the prototype, 4.2 (SD 1.4) for the first extended sense, 4.6 (SD 1.0) for the second extended sense, and -1.6 (SD 2.6) for the denominal (headless) sense, $F(3, 66) = 116.67, p < .001$. Planned comparisons revealed that while the denominal sense was judged to be more regular than the other three senses ($p < .001$), there were no differences among this latter group (all p ’s $> .5$).

The mean ratings of semantic overlap with the prototype were 4.3 (SD 1.2) for the first extended sense, 3.2 (SD 1.2) for the second extended sense, and 2.1 (SD 1.0) for the denominal sense, $F(2, 44) = 30.30, p < .001$. Planned comparisons revealed significant differences between

all adjacent senses (all p 's < .001). The convergence between the ordering of the dictionary senses and the ordering of the ratings is evidence for the construct validity of the ratings.

A multiple regression with items as the random effect was preformed on the irregularity scores of the 69 test items (23 verbs, presented in two extended senses and one denominal sense; the prototypes were excluded). The first predictor was the mean semantic similarity rating. The second was headlessness, with 0 for the extended senses and 1 for the denominal verbs (see Table 2). As expected, these two variables were correlated ($r = .55$, $p < .001$), reflecting the fact that denominal verbs are semantically nonprototypical. Fortunately, the correlation is much less than 1, allowing effects of unconfounded variance to be tested. The results of the regression revealed that headlessness uniquely accounted for 45% of the variance of in the irregularity scores, whereas semantic similarity uniquely accounted for virtually none (0.4%). The confounded effects of the two variables accounted for an additional 26%.

INSERT TABLE 2 HERE

To ensure that the lack of an effect of semantic similarity was not due to a restriction in range to the difference between the two extended senses, we redid the analysis with the prototypes included, and their semantic similarity score set to the maximum value 7. The results were similar: denominal status uniquely accounted for a significant proportion of the variance (43%) and semantic similarity accounted for almost none (0.5%). The confounded effects of the two variables accounted for an additional 30%.

Another possibility to be ruled out is that the effects of semantic similarity on regularization are nonlinear, having a detectable effect only with more extreme values of extendedness. This could systematically favor denominal items, since their semantic similarity to the prototype was on average less than those of the extended-sense items. Thus we redid the analysis with subsets of extended-sense and denominal items that are matched on their average semantic similarity ratings. If we pick pairs of verbs whose similarity ratings differ by less than a point, we end up with nine pairs with identical mean similarity ratings of 2.1 (SD = 0.68 for the extended senses, SD = 0.74 for the denominals). The mean past-tense judgment score was 4.7 (SD = 0.88) for the extended senses and -1.4 (SD = 1.6) for the denominals. Again, while denominal status accounted for a large and significant proportion of the variance in past-tense judgments (85%), semantic similarity accounted for essentially none (0.4%). The confounded variance accounted for an additional 2.6%. The same pattern was obtained when other matched subsets were chosen, e.g., 11 pairs each differing by no more than 1.4 rating points (yielding mean semantic similarity ratings of 2.3 for extended senses and 2.2 for denominals) or 13 pairs differing by no more than 1.5 points (mean ratings of 2.4 and 2.2, respectively).

Finally, since some of the denominals, but none of the extended senses, had been invented for the experiment, we wanted to ensure that the results are not due to novelty as opposed to denominal status. We redid the analysis excluding the six denominal senses that lacked entries in the dictionary: *split*, *make*, *sink*, *set*, *fit*, *throw*, and *fit*. Headlessness uniquely accounted for 44% of the variance in the past-tense difference ratings, while semantic similarity uniquely accounted for less than 0.1% (the confounded effects accounted for 35%). These results suggest no qualitative weakening in the effects of grammatical structure when novel denominals were excluded.

Discussion

This experiment disentangled the confounded effects of grammatical structure and semantic prototypicality among verb senses by constructing materials which controlled for variation among existing natural senses (selected *a priori* using a dictionary and confirmed *a posteriori* using ratings), and by providing a careful definition of semantic similarity motivated directly by the semantic association theory. Replicating the findings of Kim et al. (1991) and Gordon and Miozzo (2008), who used novel metaphorical and idiomatic senses, and Bandi-Rao and Murphy (2007), who used more conventional ones, the experiment found that morphological structure uniquely accounted almost half of the variance in the preference for past tense forms, while semantic similarity uniquely accounted for virtually none. The results suggest that the tendency to regularize a verb with a nonprototypical sense is related to the morphological structure of the verb rather than to its semantic similarity to a prototypical sense.

Experiment 3

This goal of this experiment is to converge on the results of Experiment 2 with a different design and dependent variable. Rather than manipulating semantics and morphology and then eliciting contemporary speakers' judgments of the acceptability of regular and irregular forms, we use the regular and irregular status of verbs as designated by a dictionary as the datum to be explained (on the assumption that they reflect the judgments of past speakers) and elicit contemporary speakers' judgments of semantics and morphology. In a sense, we are asking how these psychological factors have shaped the stock of regular versus irregular verbs over the course of the history of the language. The study thus complements Tabak, Schreuder, and Baayen's (2005) large-scale regression analysis of irregularity in Dutch, which analyzed the verb vocabulary in the CELEX database near-exhaustively and employed lexical statistical measures as its independent variables.

It also addresses a general problem for studies that use existing verbs as stimulus materials. It has long been known that people are capable of memorizing regular forms (Baayen & Schreuder, 2002; Caramazza, Laudanna, & Romani, 1988, Alegre & Gordon, 1999), especially when their stems are homophonous with or similar to irregulars (Ullman, 1999). There are historical circumstances that can lead an irregular-sounding verb to have a regular form even if it does not have a denominal structure (for example, its pronunciation may have converged with that of another verb, or it may have originally been based on a noun but the connection may have become opaque to current speakers). In those cases, speakers may memorize the regular form without mentally analyzing the verb as denominal; for them, it will be regular because of a lemma effect rather than a morphological effect. Gordon and Miozzo (2008) suggest that this may explain the discrepancy between the regression analyses from Ramscar (2002) and Kim et al. (1991). This difference appears to arise from just three anomalous items (*lie*, *broadcast*, and *brake*) which Ramscar's raters did not rate as denominal but which they strongly preferred as regular. When these items were excluded, Ramscar's data *did* show a correlation between regularity and headless morphology, consistent with the other studies. A similar lemma effect may explain why Ramscar's British participants did not accept the American regular past-tense *flied out* (a baseball term) even when the denominal analysis ("to make a fly") was explained to them. This morphological structure may be unintuitive to them (and to some contemporary American speakers) but may have been transparent to the early coiners of baseball terms who first used *flied*. Americans, but not Britons, have memorized that regular lemma since that era.

These problems remind us to pay attention to the historical processes that gave rise to a given verb when testing for possible morphological and lemma effects.

In this study, we combine the study of historical and psychological factors influencing inflectional morphology by focusing on clusters of phonologically-yoked families of regular and irregular verbs. In particular, within each rime class, pairs of verbs were classified as “friends” if they had the same irregular past-tense form (e.g. *throw*, *blow*, *grow*) or if they are both regular (e.g. *row*, *snow*, *glow*), and as “enemies” if one has an irregular form and the other a regular form. For example, within the *-ow* rime class, we find pairs of irregular friends such as *throw* and *blow*, pairs of regular friends such as *row* and *snow*, and pairs of enemies such as *throw* and *snow*. By holding phonological similarity constant, we can examine how speakers’ ratings of the key variables, together with lexical statistical measures predict irregularity. We considered ten independent variables.

Variables 1 and 2. Of primary interest was whether **semantic similarity to regular verbs** and **semantic similarity to irregular verbs** has affected the tendency of a verb to be regular or irregular. This can be assessed by determining whether verbs with the same past tense form (i.e., both of them irregular, or both of them regular) are more semantically similar than verbs with different past-tense forms (one regular, the other irregular), holding phonological similarity constant. As Ramscar noted, this seems to be true of the semantically similar pair *wink* and *blink*, both of which are regular, despite the strong competing irregular alternation in *sink-sank*, *drink-drank*, and so on. If this pattern is a general one, we might expect irregulars like *sink*, *drink*, and *stink* to be rated as semantically different from regular *wink* and *blink* but similar to one another, and so on for other phonological families.

Variable 3. We also asked whether the **perceived morphological structure** of the verbs (the degree to which a verb is sensed to be based on a noun) is correlated with its having a regular past-tense form, again holding phonology constant. For example, the verb *to pit* has the past tense *pitted*, which defies the associative pull of *hit-hit*, *sit-sit*, and so on. Perhaps this is because it is a headless verb based on a noun: according to the *OED*, the two main senses of the verb are “to create pits in” and “to throw two animals in a pit” (the source of the metaphoric extension “to set up a confrontation”). To see if this is a recurring pattern in English vocabulary, we need independent assessments of the morphological structure and the regular/irregular status of a large number of verbs. Thus we gathered ratings from a sample of speakers on the degree to which a verb is sensed to be based on a homophonous noun or vice-versa. Crucially, this measure assesses not just the existence of a noun-verb pair but the perceived direction of the derivation: whether today’s speakers sense that the verb is based on the noun, or the noun based on the verb. This is important because the mere existence of a homophonous noun should not be enough for speakers to treat the verb as a denominal (see discussion above).

Variables 4 and 5. If denominal status turns out to be significantly correlated with regularity, one can then test whether the effects are better predicted by contemporary participants’ judgments or by the words’ historical derivation. The latter may be captured by whether the dictionary lists a **primary noun entry** and lists the related verb under it. If the dictionary is the better predictor, it would suggest that the regularization-through-headlessness effect can leave a mark on the language in lemma information even when the headlessness is not transparent in contemporary speakers’ analyses of a complex word. This could happen if a noun root was transparent to speakers in the past who regularized the word, and they bequeathed it to

today's speakers as a regular verb, stored with the verb's lemma and reproduced from memory, regardless of whether today's would have generated a regular form themselves if they had never heard it used by others in the past tense..

As with the headedness judgments, it is important to determine whether it is the mere existence of a related noun that makes a verb regular, or whether the noun has to be perceived as the basis of the verb. This can be assessed by comparing the effects of a verb's being listed under a primary noun entry with the effects of the verb having a **noun entry** at all, collapsing across whether the noun or verb is listed first.

Variable 6. Another relevant historical variable is a verb's **native versus borrowed status**. Pinker and Prince (1988), Prasada and Pinker (1993), Marcus et al. (1995) and Pinker (1999) suggest that an irregular form must be perceived as a basic verb root, eligible for storage as an irreducible lexeme associated with a lemma. In contrast, a transparent foreign borrowing is treated in a similar way to denominals, namely as lacking a root in the language and instead having being converted from some other kind of sound. The Native or foreign status of a word may be perceived by contemporary speakers from the context in which it is introduced and used, by its spelling, or by its morphological and phonological composition (see Pinker, 1989, pp. 118-123, for a review of the cues to nativeness and their effects on linguistic phenomena).

Evidence that the native/borrowed distinction may be relevant to irregularity comes from historical linguists such as Curme (1935/1983) and Jespersen (1938/1982), who note that borrowings tend to be regular, even when their sounds should make them irregular. For example, despite the pattern in *thief-thieves*, *shelf-shelves*, and *life-lives*, borrowed nouns with such sounds tend to be regular, such as the French-derived *chiefs* and *gulfs*, and the German-derived *fifes*. The same factor explains the regularity of *mongooses* (**mongeese*), from Marathi, and of *talismans* (**talismen*), from the Arabic *tilasm*.¹⁰ Similarly, most of the thousands of French and Latin loan words in English are regular, even when they are phonologically similar to native irregulars, such as *derided* (**derode*) and *succumbed* (**succamb*). Marcus et al. (1995), Berent, Pinker, and Shimron (1999), and Shaoul (1993), testing German-, Hebrew-, and French-speaking participants respectively, invented novel nouns and presented them to participants either as simple verb roots or as borrowings from another language. They found that the participants were more likely to prefer irregular plurals for the native forms and regular plurals for the foreign ones. Tabak et al. (2005) also found that native status in Dutch was correlated with irregularity. In this analysis, we attempt a parallel test among existing English verbs by comparing foreign and native verbs, holding sound constant.

Variable 7. Also potentially relevant is a verb's **date of entry** into the language. According to historical linguists (Curme, 1935/1983; Jespersen, 1938/1982; Johnson, 1986; Levin, 1964; Pyles & Algeo, 1982), many irregular patterns are fossils of morphological or phonological rules that operated in earlier historical periods and whose outputs were memorized in lemmas once the rules became too opaque for speakers to master as productive operations. Note that this variable is partly confounded with native versus borrowed status because borrowed words tend to be more recent (hence Tabak et al. combined the two in a single variable).

¹⁰ These are the plurals sanctioned by dictionaries, and actual usage patterns are consistent with them. Google searches in mid-2008 indicate that *mongooses* outnumbers *mongeese* 179,000 to 47,600, and *talismans* outnumbers *talismen* 1,370,000 to 38,300.

Variables 8 through 10. Finally, we estimated **frequency** of past, present, and total usages of these verbs using the Francis and Kucera (1982) statistics from the Brown corpus.

Using a multiple regression, we can test the effects of semantic similarity and morphological derivation on the regular/irregular status of current English vocabulary while holding constant the other confounding variables. In addition, we can examine other hypotheses about the interaction between synchronic and diachronic effects on irregularity. If irregular forms are more memory-dependent than regular ones, one would expect to replicate the finding of Marcus et al. (1995) and Tabak et al. (2005) that irregulars have higher token frequencies than regulars (see also Lieberman et al., 2007). If headless verbs have been regularized by past speakers, regular verbs should be more likely to be listed in the dictionary after a related noun entry, and more likely to have a foreign root. If regularization is a default operation in Modern English, whereas irregulars are fossils of archaic rules, then we would expect later entry dates for regular than irregular verbs. Finally, if denominal status on average reflects both the actual historical derivation of a verb from an earlier noun (as sensed by past speakers) and contemporary speakers' perceptions of word structure, participants' judgments of verb- versus noun-headedness should correlate with actual etymology listed in the dictionary.

Method

Participants. Ninety-six adults made semantic similarity judgments of the verbs (Part 1), and forty adults made grammatical headedness judgments (Part 2). All were English-speaking Harvard undergraduates who were given course credit for their participation.

Materials. The materials comprised 62 irregular verbs and 57 regular verbs falling into 14 families of rimes, as classified in the Appendix to Pinker and Prince (1988). No doublets were included. The regular members were chosen from a search of all English verbs that rhymed with the irregular verbs, using an on-line rhyming dictionary (rhymezone.com).

In the semantic rating task, verbs within each rime family were exhaustively combined to create three pair types:

- (6) a. **IRREGULAR FRIENDS:** Both irregular, e.g. *throwing-blowing*
- b. **ENEMIES:** One irregular, one regular, e.g. *throwing-snowing*
- c. **REGULAR FRIENDS:** Both regular, e.g. *snowing-glowing*

This yielded 588 verb pairs, which were randomly assigned to lists of 98 items, each presented to a different group of participants. In order to circumvent the ambiguity between verb and nonverb senses (e.g. *snow* as a noun rather than as a verb), we presented all the verbs in the progressive, as shown in (6).

Procedure. In the semantic rating task, participants were presented with pairs of regular and irregular verbs and were asked to rate the similarity in meaning between the verbs in each pair on a scale from 1 to 7. Semantic similarity was explained in the same way as in Experiment 2. The mean ratings for pairs of verbs ranged from 2.1 to 5.0 for irregular friends, 1.7 to 4.8 for regular friends, and 1.9 to 4.5 for enemies, indicating that participants responded to variation among the pairs in each group. Reliability of the measure was assessed by computing intraclass correlations for each of the six lists. The average across the lists was .83, range .73 to .91, all p 's < .001. This indicates that the different participants were responding in a consistent pattern to the 98 pairs that each of them was rating.

In the headedness rating task, verbs were presented individually and participants were asked to rate whether the use of the word as a noun or as a verb felt more basic, using a scale of 1 (“The verb is more basic”) to 7 (“The noun is more basic”). The scale was explained using two examples, one where the verb was clearly more basic (*He reached for his socks* versus *He had a long reach*), and another where the noun clearly more basic (*He bought some butter* versus *He buttered his bread*). Participants were told that while items could be used as either nouns or verbs, they should give their “gut feeling as to which of these is more fundamental or basic.”

Using the *OED*, we recorded a verb’s date of first recorded usage, and its language of origin, which we dichotomized into Native (Germanic root, including Anglo-Saxon, Scandinavian, German, and Dutch) and foreign (non-Germanic root). We also recorded whether an item possessed a noun entry in addition to its verb entry, and if so, whether the first entry listed was the verb or the noun sense. Finally, using Francis & Kucera (1982), we obtained three measures of the verb’s frequency: Present (summing the frequencies of the bare present-tense form, the gerund, and the present participle forms, as in *throw*, *throws*, and *throwing*); Past (summing the preterite and participle, e.g., *threw* and *thrown*), and the Total occurrences (summing the present and past counts).

Results

We report three sets of analyses: first, t-tests ascertaining whether irregular and regular verbs significantly differed on the variables of interest (Table 3); simple correlations of these variables to each other and to the irregular/regular status of the verbs (Table 4); and third, logistic regressions that explored how well each variable predicts regular versus irregular status in the presence of the other variables (Table 5).

INSERT TABLES 3 AND 4 HERE

Date and Language of origin (variables 6 and 7). On average, irregular verbs are 251 years older than phonology-matched regular verbs (1070 versus 1321). Irregular verbs are also significantly more likely than regular verbs to have native roots (87% versus 65%). Table 4 shows the two variables are correlated with each other ($r = .53$), as one would expect given the history of the English language, in which many Latinate words infiltrated a Germanic substrate following the Norman invasion and during the Renaissance. Though both are correlated with Regularity ($r = .44$ and $r = .26$ respectively), Date of Entry has a clearer effect, continuing to correlate with it when language of origin (Borrowed) is partialled out (partial $r = .37$, $t(116) = 4.28$, $p < .001$), while the reverse is not true (partial $r = .01$, $t(116) < 1$).

Frequency (variables 8-10). Not surprisingly, irregular verbs had significantly higher log frequencies in the present tense (1.1 versus 0.4), past tense (0.82 versus 0.33), and overall (1.3 versus 0.6). The difference is consistent with the hypothesis that irregular verbs are more memory-dependent than regular ones and require higher token frequencies to sustain its form (Marcus et al., 1995).

Semantic similarity (variables 1 and 2). Irregular verbs were rated as significantly more similar in meaning to other irregular verbs in the same family than to the regular verbs in that family (3.0 vs. 2.7). In contrast, regular verbs were not rated as more semantically similar to other regular verbs in their family than to irregular verbs in that family (2.8 vs. 2.7). Correlation coefficients confirm that the irregular and regular similarity judgments were independent ($r = .11$) and that only the ratings of similarity to irregular verbs correlated with irregularity ($r = .24$).

The difference is consistent with earlier findings suggesting that irregular forms (presumably because they are more dependent on memory) are likely to attract new members by analogy, and that regular forms (presumably because they are more likely to be assembled in real time) show little or no such tendency (Marcus et al., 1992; Ullman, 1999). The results are related to those of Baayen and Moscoso del Prado Martin (2005), who found a greater density in semantic space of irregular forms compared to regular ones, though the present results show in addition that irregulars in specific past-tense clusters (and not just irregulars in general) show measurable semantic similarity. This constitutes evidence (provisional on the ruling out of confounds) that semantic similarity can be one of the dimensions encouraging analogies.

Headedness: Dictionary data (variables 4 and 5). Regular verbs were significantly more likely than irregular ones to have a noun as their primary entry (46% versus 14%). They were also slightly more likely to have a noun as an entry at all (72% versus 61%), but this difference was not significant. The two variables were, of course, moderately correlated ($r = .30$) since a verb lacking a noun entry cannot have a noun as a first entry. But only Primary Noun Entry significantly correlated with Regularity ($r = .34$). This is what one would expect if it is the direction of derivation, rather than the mere presence of a noun form, that causes regularity.

Headedness: Judgment data (variable 3). We found that irregular verbs were significantly less likely than regular verbs to be judged as having a noun as their basic form (2.5 vs. 3.6). In the simple correlations, these judgments were correlated with irregularity ($r = .35$), and not surprisingly, with having a noun entry in the dictionary (.46) and with having a noun as the primary entry ($r = .68$). The similar-sized effects, and the high intercorrelation between Primary entry and Headedness judgment, make it difficult to determine which is the better predictor. Each is significant or marginally significant when the other is partialled out (both $p < .10$), when Past-tense frequency is partialled out (both $p < .01$), and when Date of entry is partialled out ($p < .10$ for Primary entry; $p < .05$ for Headedness judgment).

Comparisons of Headedness judgments with Semantic Similarity to Irregulars revealed that Headedness predicted irregularity while partialing out Similarity ($r = .32$, $t(117) = 3.65$, $p < .0004$), and vice-versa ($r = -.19$, $t(117) = -2.11$, $p < .05$). When Primary Noun entry rather than Headlessness was used as the measure of grammatical derivation, a similar pattern emerged, but with only a marginal effect of semantic similarity: the partial correlation of irregularity with Primary Noun Entry was $r = .30$ ($t(117) = 3.38$, $p < .001$), and the partial correlation of irregularity with Similarity to Irregulars was $r = -.17$ ($t(117) = -1.90$, $p < .06$).

Finally, binary logistic regressions were performed to see whether frequency, language of origin, headedness, and semantic similarity independently predict whether a verb is irregular or regular when their numerous intercorrelations are considered simultaneously (see Table 5). We first narrowed the set of predictors by focusing on conceptually related and intercorrelated subsets and choosing the member with the highest correlation with regularity/irregularity. This resulted in six initial variables: a verb's Past-tense Frequency ($r = -.35$), its Date of Entry ($r = .44$), whether it was Borrowed ($r = .26$), participants' judgments of Headedness ($r = .35$), ratings of semantic Similarity to Irregular verbs ($r = .24$), and ratings of semantic Similarity to Regular verbs ($r = .09$). Three of the variables predicted irregularity independently of their confounded variance with the others. Compared to regular verbs, irregular verbs were higher in past-tense frequency, entered the language earlier, and were more semantically similar to other irregular verbs. Irregular verbs were also less likely to be rated as being headed by a noun, though this

difference was only marginally significant. We found no significant effect on irregularity of a verb's language of origin, nor of its rated semantic similarity to regular verbs.

INSERT TABLE 5 HERE

Because Date of Entry is correlated with other predictors, it may have overshadowed some of them. Table 4 confirms that a verb's entry into the English language is strongly correlated with whether it has a Germanic or a borrowed root ($r = .53$, $p < .001$). In addition, noun-headed verbs entered the language significantly later than verb-headed items; this is true both for irregular verbs (1254 vs. 1038; $t(60) = 2.91$, $p < .01$), and for regular verbs (1459 vs. 1205; $t(55) = 3.58$, $p < .01$). Presumably this is due to the gradual impoverishment of overt morphology beginning in the Middle English period. In time, more verbs were zero-converted from nouns rather than derived by overt affixation and vowel changes, thus yielding more noun-verb homophones.

Because of the enormous effect of the Norman invasion and other changes that marked the transition from Old to Middle English, we divided the sample into the 62 verbs that were first recorded during the Old English period (prior to 1200 AD) and the 57 that were first recorded during the Middle English period (following 1200AD). Among the Old English verbs, only Past-tense frequency predicted irregularity independently of their confounded variance with the others. Among the Middle English verbs, two of the six variables independently predicted irregularity: irregular verbs were less likely to be rated as being headed by a noun root, and as being more semantically similar to other irregular verbs. Irregular verbs were more likely than regular verbs to have originated from a Germanic root, but the effect was only marginally significant. Past-tense frequency and rated semantic similarity to regular verbs did not significantly predict irregular status.

Discussion

This experiment confirms that in comparison with regular verbs, phonologically similar irregular verbs on average are older, are more likely to have Germanic roots, have higher frequencies, are less likely to be headed by a noun, and are semantically more similar to other irregular verbs. Of these variables, a verb's date of entry into the language is completely robust, persisting as a significant predictor in every analysis. Three of the other variables (headless judgment, semantic similarity to irregulars, frequency) are reasonably robust, persisting as significant predictors when many of the other variables were taken into account. The effect of language of origin is suggestive but not proven: it is a significant predictor in isolation but only marginally significant when other variables are simultaneously controlled.

The overall pattern of results is consistent with those reported for Dutch by Tabak et al. (2005). They reported that higher frequency verbs were more likely to be irregular, that non-native verbs tend to be regular (100%, in their case), that verbs with a higher ratio of noun to verb frequencies (an indicator of denominal or headless status) are more likely to be regular, and that older verbs were less likely to be regular. And as mentioned, the finding that irregular verbs are more semantically similar than regular ones is consistent with the findings of Baayen and Moscoso del Prado Martin (2005). These results also offer quantitative suggestions about the historical and psychological forces that shaped the irregular-regular balance in English that are consistent with accounts by historical linguists and extended by Marcus et al. (1995), Prasada and Pinker (1993), Pinker (1999), and Lieberman et al. (2007). Speakers' judgments of a verb's

morphological structure (specifically, whether a verb has a verbal head or is based on a noun) has contributed to the preponderance of regular verbs in English, especially since the radical changes initiated in the Middle English period: new verbs derived from nouns, or borrowed from other languages, were likely to be regular, even if they sounded like irregular forms.

Contrary to the claims of MacWhinney and Leinbach (1991) and Ramscar (2003), the effect of morphological structure cannot be reduced to semantic extendedness. The data do, however, provide support for the suggestion that semantic similarity encourages generalization of irregular patterns (Baayen & Moscoso del Prado Martin, 2005). Presumably this effect arises from the greater likelihood of remembering irregular forms, and generalizing their patterns to new forms, when the forms overlap in their semantic representation; such an effect has long been known to hold for phonological overlap. Interestingly, a verb's semantic similarity to *regular* verbs did not predict whether it was likely to be regular or irregular. This parallels the non-effects of phonological similarity for children's overregularization errors (Marcus et al., 1992) and for adults' acceptability judgments for regulars (Ullman, 1993), and suggests that regular inflected forms do not form clusters of overlapping forms in memory that support memory and generalization.

Semantic similarity ratings and neighborhood measures collapse many semantic features, and the mere existence of an effect does not illuminate why it came about. In Experiment 4, we test one such feature, for which there is an *a priori* reason to expect a semantic effect on inflectional morphology.

Experiment 4

Why might a word's semantics affect whether it has regular or irregular inflection? In the case of a word's phonology, the effect is no mystery, because the process of inflection implements a phonological operation (suffixation or phonological modifications). During the history of a language, the interaction between the phonological properties of the stem and the phonological changes effected by the inflectional process can bring about additional changes to ease pronunciation or perception (Pinker & Prince, 1988; Pinker, 1999). Two examples are the vowel-shortening in forms like *sleep-slept* and *feel-felt*, which originated as an adjustment to the lengthening of the syllable by the addition of a suffix, and the devoicing and degemination found in irregulars whose stems end in *t* or *d* (*hit-hit*, *cut-cut*, *bend-bent*), which was an adjustment to the creation of a coda with two similar or identical consonants.

In this experiment, we test whether an analogous kind of stem-suffix chemistry occurs in the *semantic* operation of the past tense, and whether it can explain the existence of semantic associative effects in inflection. As discussed in the Introduction, because tense and aspect both pertain to time, the concept of pastness inserted by the tense-marking process could interact with the aspectual properties of the activity inhering in the verb stem. This would manifest itself when the combination of meaning of an irregular verb and pastness are stored in memory, which would itself be a consequence of the requirement that the unpredictable past-tense form be stored in memory. A new sense of an irregular verb that preserves that semantic gestalt (i.e., by expressing an event with the same telicity or durativity as the original sense) will naturally inherit the irregular form. In contrast, a new sense that is incompatible with that semantic gestalt because it has a different inherent aspect and would clash with the meaning stored for that verb in the past. They may incline speakers to revert to a regular form. Examples might include *knitted his brows* (since past-tense *knit* ordinarily implies the existence of a knitted product), *hanged the prisoner*

(since *hung* ordinarily implies an enduring state of a suspended object rather than the instantaneous change in a killing), and *pleaded with me* (since *pled* in the legal sense creates the ongoing state of a formal plea).

This study will focus on doublets (verbs with both regular and irregular past-tense forms), since they are matched for most relevant lexical variables, and may reflect a historical change in progress. We test whether the members of a doublet differ in their global semantics, in their inherent aspect, or both. Adapting the materials and procedures of Experiment 2, we first elicited judgments of the regular and irregular forms of different senses of doublet verbs, in order to compare them with verbs that are also polysemous but have a single past-tense form. Next, we elicited ratings of global semantic similarity, as in the other experiments. Finally, we had participants rate the extent to which a verb's sense exemplified each of several aspectual event classes (*Aktionsarten*). If polysemous verbs differ in regularity because of an aspectual clash among their senses, the senses of a doublet should show less aspectual similarity than the senses of a verb that is consistently irregular, controlling global semantic similarity. If they differ only because of global semantic dissimilarity, then the reverse pattern should emerge.

Methods

Participants. One hundred thirty-two adults participated, divided into four sets of 34 who completed each of four tasks. All were English-speaking Harvard undergraduates who were given course credit for their participation.

Materials. Participants saw 46 sentences, which varied the meanings of 16 verbs which have both regular and irregular past-tense or participle forms in standard English (*burn-burned/burnt*, *dive-dived/dove*, *dream-dreamed/dreamt*, *hang-hanged/hung*, *heave-heaved/hove*, *kneel-kneeled/knelt*, *knit-knitted/knit*, *leap-leaped/leapt*, *light-lighted/lit*, *plead-pleaded/pled*, *prove-proved/proven* [participle], *sew-sewed/sewn* [participle], *shine-shined/shone*, *sneak-sneaked/snuck*, *strive-strived/strove*, *weave-weaved/wove*). The doublets were taken from Ullman (1993), and sets of sentences were constructed with two, three, or four senses of each verb. The prototypical sense was the one in the first entry in the *OED* (e.g., 7a); the extended senses were from subsequent entries (7b-d). As in preceding experiments, the dictionary was consulted only to select the senses for the experimental materials; the actual semantic extendedness values were assessed empirically. All verbs were presented in both regular and irregular past tense or participle forms; for half the items, the irregular was presented first; for the other half, the regular was presented first. The verbs were underlined.

(7) a. PROTOTYPICAL SENSE

IRREGULAR: Last Christmas, the children hung their stockings by the fireplace with hopes of receiving gifts from Santa.

REGULAR: Last Christmas, the children hanged their stockings by the fireplace with hopes of receiving gifts from Santa.

b. EXTENDED SENSE

IRREGULAR: The first painting I sold hung on the wall of my parent's home.

REGULAR: The first painting I sold hanged on the wall of my parent's home.

c. EXTENDED SENSE

IRREGULAR: The whole town showed up to watch when the sheriff hung the thief.

REGULAR: The whole town showed up to watch when the sheriff hanged the thief.

d. EXTENDED SENSE

IRREGULAR: Last night, my friends and I **hung** out at our favorite bar hours after they closed the doors.

REGULAR: Last night, my friends and I **hanged** out at our favorite bar hours after they closed the doors.

The non-doublet verbs used for comparison were taken from the materials of Experiment 2, and consisted of 69 sentences comprising one prototypical and two extended senses of each of 23 irregular verbs. We excluded denominal forms, because the previous experiments established that these are consistently regularized.

Procedure. The first task measured the relative acceptability of the regular and irregular past-tense forms of each sense of the doublet verbs. Participants read sentences introducing a sense of the verb (e.g., *hang* as “execute,” or *hang* as “suspend”), and were asked to rate the acceptability of the regular (*hanged*) and irregular (*hung*) forms. This scale was explained using the instructions from Experiments 1 and 2.

The second task assessed global semantic similarity. Participants read sentences introducing the prototypical sense of the verb (7a) and were asked to rate the semantic similarity to each of extended senses (7b-d). The scale was explained using the instructions for the semantic ratings taken from Experiment 2 and 3.¹¹

The third and fourth tasks assessed aspectual similarity among the senses of doublet and non-doublet verbs, respectively. Participants were told that each event described in a sentence can unfold over time in four possible ways, with the following examples:

- (8)
- a. **UNCHANGING STATE**: These are situations that are static or unchanging, with no distinguishable phases, e.g. *The flag is red*.
 - b. **MOMENT-OF-CHANGE**: These sentences refer to an instantaneous moment at which something changes, e.g. *I found the key*.
 - c. **ACTIVITY-CULMINATING-IN-CHANGE**: These sentences refer to processes with a clear end point. They embrace two distinct phases—an action taking place and its termination, e.g. *She wrote a letter*.
 - d. **ACTIVITY-WITHOUT-ENDPOINT**: These sentences refer to processes without a specific end point, e.g. *He’s running*.

In traditional theories of aspect (Vendler, 1957), these correspond respectively to a “state,” an “achievement” (a nondurative, telic event), an “accomplishment” (a durative, telic event) and an “activity” (a durative, atelic event). Participants were asked to rate each sentence four times in terms of how appropriately it fit into each category, using a scale from 1 (“not appropriate”) to 7 (“very appropriate”). To avoid any bias from past-tense morphology or semantics, all sentences were presented in the future tense, e.g. *The first painting I will sell will **hang** on the wall of my parent’s home*.

¹¹ Among the twenty features in the MacWhinney-Leinbach set, three pertain to aspectual distinctions, and another ten mention moving or changing, which favor durative or telic aspect. In principle, this could confound our measures of semantic similarity and aspectual similarity, reducing the power of the regression to detect unconfounded effects. In practice, the two measures were uncorrelated ($r = -.10$), suggesting that participants did not attend to the instructions on a feature-by-feature basis but assessed semantic similarity more globally.

Results

We present three sets of analyses of the ratings of 76 senses (30 from doublet verbs, 46 from non-doublet verbs). The first consists of simple correlations (Table 6). Doublet status was coded 0 if the irregular verb did not possess a standard regular form and 1 if it did. The other three variables were continuous, and were calculated as a difference between the variable for an extended sense of the verb and the measure for its prototypical sense. Regularity Difference was calculated by first taking the difference between judgments of the irregular and the regular form of each sense, and then subtracting the difference for an extended sense from the difference for the prototypical sense. The absolute value of this difference-of-differences represents the extent to which the past tense of an extended sense differs from that of its prototypical sense: a high absolute score indicates a greater difference.¹²

Since inherent aspect, as we are manipulating it, is a nominal scale with four categories, it cannot be summarized in a single variable. Moreover, we are estimating it with ratings rather than our own judgments, and the participants were not unanimous or univocal in their judgments (e.g., a sense might sometimes be categorized as an accomplishment, sometimes as an achievement, and sometimes as both). Since we are interested only in the *similarity* in perceived inherent aspect between two senses, we calculated the correlation coefficient between the four aspectual ratings of an extended sense and those of its prototypical sense. This captures the degree to which the pattern of mean ratings across the four aspect classes for the extended sense differed from the pattern of mean ratings for the prototypical sense. A high correlation indicates that the two senses have a similar profile across the four aspect classes, and thus are conceived as aspectually similar. Reliability was assessed by computing intraclass correlations, which were $r = .93$ ($p < .001$) for the doublet items and $r = .81$, ($p < .001$) for the non-doublet items. As with ratings of semantic similarity (Experiment 3), the different participants were responding in a highly similar pattern across the items.

Finally, variations in Semantic Similarity were captured by the mean ratings of the similarity between extended and prototypical senses.

In Table 6, the .35 correlation between Doublet Status and Regularity Difference is not surprising, since by definition a doublet has an acceptable regular form. However, this correlation is not logically necessary (and empirically is nowhere near 1.0) because Regularity Difference captures not the acceptability of a regular variant across the board but specifically the extent to which the past tense form of an extended sense differs from that of a prototypical sense. The imperfect correlation suggests that the reason that some verbs have both regular and irregular forms is not primarily to disambiguate polysemous senses, with one sense allocated to the regular form and the other to the irregular (Harris, 1992; Shirai, 1997), but because of other factors. For example, since doublet verbs are lower in frequency than non-doublet verbs (Ullman, 1993), it has been suggested that they arise when the irregular past-tense of a low-frequency word is imperfectly memorized, leading some speakers to occasionally default to a regular (Marcus et al., 1992; Ullman, 1993; Pinker, 1999). In that case their usages can then become standardized as an alternative, but basically synonymous, past-tense form. Another

¹² The signed counterpart to this measure would represent the direction of the change: positive if the extended sense was more regular, negative if it was more irregular. However, as mentioned, aspectual compositionality effects do not seem to favor regular or irregular forms; see note 6.

possibility is dialect mixture, as in the difference between British *dived* and *learnt* and American *dove* and *learned*.

INSERT TABLE 6 HERE

The Aspectual Correlation between an extended sense and a prototypical sense significantly predicted both the existence of a doublet for the verb (regardless of sense) and the irregular-regular judgments for those senses. This is consistent with the hypothesis that aspectual differences between the prototypical sense of an irregular verb and a novel sense can tilt speakers toward using a regular form for the novel sense. Semantic Similarity, in contrast, showed small and nonsignificant correlations with doublethood and with regularity difference. This pattern was confirmed in a series of t-tests. Not surprisingly, with doublet verbs there was a greater difference in irregularity judgments between prototypical and extended senses than there was with non-doublet verbs (1.6 vs. 0.7; $t(74) = 3.23, p < .01$). More interestingly, the extended senses of doublets were more aspectually distinct from their prototypes than the extended senses of non-doublets: the correlations across the four ratings was $r = .23$ for the doublets and $r = .55$ for the non-doublets, $t(74) = 3.33, p < .01$. In contrast, there were no effects of global semantic similarity: the difference between the extended and prototypical sense was no different for doublet and nondoublet verbs (3.9 vs. 3.7; $t(74) = 0.64, p > .50$).

Finally, a multiple regression was performed on the Regularity Difference of the 76 test items (see Table 7). The first predictor was the Aspectual Correlation between an extended sense and a prototypical sense; the second was the Semantic Similarity between the two senses. These two variables were not significantly correlated with each other ($r = .14, p > .30$). The analyses showed that Aspectual Correlation uniquely accounted for a significant 5.1% of the variance of the irregularity-regularity differences. In contrast, semantic differences uniquely accounted for less than 1.0%.

INSERT TABLE 7 HERE

Discussion

Experiment 4 confirms that to the extent that differences in lexical semantics correlate with a difference in regular and irregular inflection, the key ingredient is a difference in the inherent aspect of the verb (its Aktionsart or event structure), rather than global differences in all its semantic features. Focusing on doublets to achieve maximal control, we found that global semantic similarity did not predict either the existence of a standard inflectional alternative or the acceptability of such an alternative. However, we found that when a verb had an extended sense that differed in inherent aspect from its prototypical sense, it was more likely to have a second past-tense form. Moreover, the more an extended sense differed in aspect from the prototypical sense, the more speakers liked the alternative form. This is consistent with the compositionality explanation for *why* semantics should affect inflection: when an inflectional form has to be stored because of its irregular idiosyncrasies, it creates a slot in memory for any semantic gestalt formed by the combination of the inherent semantics of the stem and the semantics of the affix. In the case of verbs, the affix encodes pastness, so the lexical semantic property that would most readily interact with it would be inherent aspect, which also pertains to time.

General Discussion

The experiments reported here offer the following degrees of support for the four possible kinds of effects of lexical semantics on irregular inflection.

Lexical (or lemma) effects, in which the unique combination of a sound and meaning defining a word may have an irregular form linked to it, were supported by findings in Experiment 1. Generalization of an irregular pattern to a new word is far more likely when the item is highly similar both in sound *and* meaning to an existing irregular item, above and beyond their summed effects. By itself, this pattern could merely reflect a nonlinear interaction between a phonological generalization gradient and an additional contribution from high semantic similarity. But crucially, those are the circumstances in which speakers reported that they had a specific word in mind and based their generalizations on an analogy to that word. Most important, such reports are highly correlated with speakers' tendency to generalize an irregular pattern, and can fully account for the apparent effects of semantic similarity.

Semantic associative effects, in which speakers generalize inflectional patterns because of a global overlap of semantic features between a new item and previously learned items, showed inconsistent effects. Experiment 1 showed few or no signs of a semantic generalization gradient. Experiment 2 showed that irregular past-tense forms of nonprototypical senses of irregular verbs were slightly less acceptable than with the prototypical senses, but the difference was not significant. And Experiment 3 showed that existing English irregular verbs are more semantically similar to one another than they are to regular verbs. However, Experiments 2 and 4 showed no associative semantic effects when other semantic factors (headlessness and temporal composition, respectively) were held constant. We tentatively conclude that there is a small tendency for irregular patterns to be generalized to semantically similar words (as also suggested by Baayen & Moscoso del Prado Martin, 2005), though the possibility remains that even these are caused by tense-aspect compositionality effects.

Regardless of the strength and nature of associative semantic effects, Experiments 2 and 3 establish that they cannot account for the effect of morphological structure. This is the effect, in which the headless structure of a complex verb prevents an irregular form stored with a root from being passed on to the word as a whole. In these experiments, when the two are simultaneously varied and their unconfounded effects individually tested, headlessness showed a strong unconfounded effect, whereas semantic similarity did not. The dictionary analyses in Experiment 3 suggest that this psychological phenomenon may have left its mark on the stock of regular verbs in contemporary English, at least since the Middle English period.

Finally, stem-affix compositionality effects (in the case of tense, the interaction between the aspect class of the verb and the concept of pastness contributed by the suffix) appear to have a significant effect in the generation or acceptance of regular alternatives to irregular verbs. Experiment 4 showed that people shy away from an irregular form for a new sense of a polysemous verb not because the new sense is globally different from the prototype but because it implies a different aftermath of the event, one that is inconsistent with the aftermath stored with the verb's past-tense entry. This raises the possibility that semantic effects on inflection are at least partly predictable in their content, and are not just fortuitous associations to the semantic properties of the words that fall together because of a shared irregular pattern.

The results have several sets of implications for theories of language. First, by showing that morphological structure (specifically, headlessness) is one of the determinants of irregularity and regularity rather than an epiphenomenon of global semantic dissimilarity, it reinforces the

need for a structured information-inheritance mechanism in word formation (di Sciullo & Williams, 1987; Selkirk, 1982), analogous to the compositionality mechanisms that are indispensable in syntax. Together with the demonstration that irregularity is affected by lemma distinctness rather than simply semantic similarity, the results support models of language that invoke lexical entries and morphological structure as intermediate levels of representation, and suggest limitations in theories that rely on direct mappings from meaning to sound, as in many connectionist, functionalist, and statistical models of language.

The results bear on the controversy between connectionist and symbolic models of language in two other ways. The existence of an associative pull of the semantic properties of irregular forms, but not regular forms, reinforce a similar asymmetry for phonological similarity that has been shown in intact adults (Ullman, 1999), children (Marcus et al., 1992), and neurological patients (Pinker & Ullman, 2003; Pinker & Ullman, 2002a, 2002b; Ullman et al., 1997; Ullman et al., in press). Second, the compositionality effect demonstrated in Experiment 4 was a direct prediction of the hypothesis that irregular forms (but not regular forms) are mandatorily stored in memory. These findings do not rule out connectionist models of language in general, since morphological structure can be built in to the models in various ways (see Smolensky & Legendre, 2006), and since certain combinations of architecture and training set can result in dissociations between the memory-dependence of regular and irregular forms. But they do provide constraints on models of inflection from all approaches, pointing to the importance of differences in memory-dependence among different kinds of linguistic forms.

The data support the models of word production that do posit a level of representation between sound and meaning, such as localist word nodes (Dell, 1986; Dell, Burger, & Svec, 1997), lemmas (Levelt, 1989, Roelofs, Meyer, & Levelt, 1998), or lexemes (Caramazza, 1997). These models naturally accommodate the listing of distinct irregular forms without requiring them to be associated directly to either phonological or semantic representations. For example, in the model of Roelofs et al. (1997), an irregular lexeme may be linked to the word's lemma and marked by diacritics for the relevant inflectional feature, such as tense. During speech production, the irregular lexeme is inserted into a single slot in a morphological structure, whereas regularly inflected verbs would occupy separate slots for the stem and suffix (Janssen, Roelofs, & Levelt, 2002). The current results support three postulates of this model: the existence of a lemma level, the representation of morphological structure, and a qualitative distinction between regular and irregular forms. They also suggest several refinements. The effects of aspectual compositionality suggest the need for models with semantic representations that are specific to irregular lexemes, rather than completely separated from them at the lemma level. The effects of headless morphological structures suggest models where information flow from the morphological structure to the lexeme selection process.

A final set of implications concerns the interaction between psychology and history in the shaping of a language. Most psycholinguists have treated the structure of a language as an independent variable, providing an input to children as they learn the language and to adults as they parse or judge it. But the constant change over a language's history shows that speakers actively affect, and not just are affected by, their language. Traditional linguistic analyses of the diachronic shaping of regular and irregular inflectional systems have partially acknowledged this by hypothesizing a role for frequency and similarity (e.g., Curme, 1935/1983; Jespersen, 1938/1982; Kurylowicz, 1964). This has been confirmed by recent quantitative studies which

suggest that lower-frequency irregular verbs tend to become regular over time, and that clusters of similar irregulars tend to attract new members (Bybee, 1985; Lieberman et al., 2007; Marcus, et al., 1995). The studies in this paper suggest that a larger set of psycholinguistic factors is at work at well. We envision that future studies of inflection can shed light on the interaction between the psychology of speakers and the historical contexts of language communities which shape a language over time. If so, languages can increasingly be analyzed as effects, not just causes, of psycholinguistic phenomena.

References

- American Heritage Dictionary of the English Language* (2000). Fourth edition. Boston, MA: Houghton-Mifflin.
- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Baayen, R. H., & Moscoso del Prado Martin, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language*, 81(3), 666-698.
- Bandi-Rao, S., & Murphy, G. L. (2007). The role of meaning in past-tense inflection: Evidence from polysemy and denominal derivation. *Cognition*, 104, 150-162.
- Baayen, R. H., & Schreuder, R. (2002). The balance of storage and computation in the mental lexicon: The case of morphological processing in language comprehension. In S. Nootebohm, F. Weerman & F. Wijnen (Eds.), *Storage and computation in the language faculty*. Dordrecht: Kluwer.
- Berent, I., Pinker, S., & Shimron, J. (1999). Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition*, 72, 1-44.
- Berent, I., Pinker, S., & Shimron, J. (2002). The nature of regularity and irregularity: Evidence from Hebrew nominal inflection. *Journal of Psycholinguistic Research*, 31(5), 459-502.
- Bolinger, D. (1980). *Language: The loaded weapon*. New York: Longman.
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Philadelphia: Benjamins.
- Bybee, J. L., & Moder, C. L. (1983). Morphological classes as natural categories. *Language*, 59, 251-170.
- Bybee, J. L. S., D. I. (1982). Rules and schemes in the development and use of the English past tense. *Language*, 58, 265-289.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177-208.
- Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28, 297-332.
- Carlson, G., Keyser, S. J., & Roeper, T. (1977). *Dring, drang, drung* (Unpublished manuscript). Rochester, NY: University of Rochester.
- Carter, R. J. (1976). Some constraints on possible words. *Semantikos*, 1, 27-66.
- Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991-1013.
- Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 55, 767-811.
- Comrie, B. (1985). *Tense*. New York: Cambridge University Press.
- Curme, G. O. (1935/1983). *A grammar of the English language* (Vol. Vol. 1.). Essex, CT: Verbatim.

- Daugherty, K. G., MacDonald, M. C., Petersen, A. S., & Seidenberg, M. S. (1993). *Why no mere mortal has ever flown out to center field but people often say they do*. Paper presented at the Fifteenth Annual Conference of the Cognitive Science Society.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104(1), 123-147.
- di Sciullo, A. M., & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.
- Dowty, D. R. (1979/1991). *Word meaning and Montague Grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*. Boston: Kluwer.
- Egedi, D. M., & Sproat, R. W. (1991). *Connectionist networks and natural language morphology* (Unpublished manuscript). Murray Hill, NJ: Linguistics Research Department, Lucent Technologies.
- Francis, N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Gordon, P., & Miozzo, M. (2008). Can word formation be *understood* or *understanded* by semantics alone? *Cognitive Psychology*, 56, 30-72.
- Harris, C. L. (1992). *Understanding English past-tense formation: The shared meaning hypothesis*. Paper presented at the Fourteenth Annual Conference of the Cognitive Science Society.
- Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In J. M. D. Rumelhart, and the PDP Research Group (Ed.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
- Janssen, D. P., Roelofs, A., & Levelt, W. J. M. (2002). Inflectional frames in language production. *Language and Cognitive Processes*, 17, 209-236.
- Jespersen, O. (1938/1982). *Growth and structure of the English language*. Chicago: University of Chicago Press.
- Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Science*, 96, 7592-7597.
- Johnson, K. (1986). Fragmentation of strong verb ablaut in Old English. *Ohio State University Working Papers in Linguistics*, 34, 108-122.
- Kim, J. J., Marcus, G. F., Pinker, S., Hollander, M., & Coppola, M. (1994). Sensitivity of children's inflection to morphological structure. *Journal of Child Language*, 21, 173-209.
- Kim, J. J., Pinker, S., Prince, A., & Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science*, 15, 173-218.
- Kiparsky, P. (1982). Lexical phonology and morphology. In I. S. Yang (Ed.), *Linguistics in the morning calm* (pp. 3-91). Seoul: Hansin.
- Kurylowicz, J. (1964). *The inflectional categories of Indo-European*. Heidelberg: C. Winter.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

- Levelt, W.J.M., Roelofs, A., Meyer, A.S. (1998). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Levin, S. R. (1964). A reclassification of the Old English strong verbs. *Language*, 40, 156-161.
- Lieber, R. (1980). *On the organization of the lexicon*. Unpublished Ph.D. dissertation, MIT, Cambridge.
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449, 713-716.
- MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, 40, 121-157.
- Marcus, G. F. (2001). *The algebraic mind: Reflections on connectionism and cognitive science*. Cambridge, MA: MIT Press.
- Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, 29, 189-256.
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, 57.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71-102.
- McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, 6, 465-472.
- Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition*, 41, 197-229.
- Nakisa, R. C., & Hahn, U. (1996). *Where defaults don't help: The case of the German plural system*. Paper presented at the 18th Annual Conference of the Cognitive Science Society.
- Nunberg, G. (1989). What the usage panel thinks. In C. Ricks & L. Michaels (Eds.), *The state of the language*. Berkeley: University of California Press.
- Orsolini, M., & Marslen-Wilson, W. D. (1997). Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes*, 12, 1-47.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, Mass.: MIT Press.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: HarperCollins.
- Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. London: Penguin.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition*, 28, 73-193.

- Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S. D. Lima, R. L. Corrigan & G. K. Iverson (Eds.), *The reality of linguistic rules*. Philadelphia: John Benjamins.
- Pinker, S., & Ullman, M. (2003). Beyond one model per phenomenon. *Trends in Cognitive Science*, 7(3), 108-109.
- Pinker, S., & Ullman, M. T. (2002a). Combination and structure, not gradedness, is the issue. *Trends in Cognitive Sciences*, 6(11), 472-474.
- Pinker, S., & Ullman, M. T. (2002b). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11), 456-463.
- Prasada, S., & Pinker, S. (1993). Generalizations of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8, 1-56.
- Pyles, T., & Algeo, J. (1982). *The origins and development of the English language* (3rd ed.). New York: Harcourt Brace Jovanovich.
- Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, 45, 45-94.
- Ramscar, M. (2003). The past-tense debate: Exocentric forms versus the evidence. *Trends in Cognitive Sciences*, 7, 107-108.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42, 107-142.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2: Psychological and biological models.). Cambridge, MA: MIT Press.
- Sasse, H.-J. (2002). Recent activity in the theory of aspect: Accomplishments, achievements, or just non-progressive state? *Linguistic Typology*, 6, 199-271.
- Selkirk, E. O. (1982). *The syntax of words*. Cambridge, MA: MIT Press.
- Shaoul, C. (1993). *Regularization in the inflection of French nouns* (Unpublished manuscript). Cambridge, MA: Department of Brain & Cognitive Sciences, MIT.
- Shirai, Y. (1997). Is regularization determined by semantics, or grammar, or both? Comments on Kim, Marcus, Pinker, Hollander & Coppola. *Journal of Child Language*, 24, 495-501.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1-74.
- Smolensky, P., & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vol. 1: Cognitive architecture). Cambridge, MA: MIT Press.
- Tabak, W., Schreuder, R. and Baayen, R. H. (2005). Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. In M. Reis and S. Kepser (eds), *Linguistic Evidence*, Mouton, 529-555.

- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description III: Grammatical categories and the lexicon*. New York: Cambridge University Press.
- Tyler, L. K., de Mornay-Davies, P., Anokhina, R., Longworth, C., Randall, B., & Marslen-Wilson, W. D. (2002). Dissociations in processing past tense morphology: Neuropathology and behavioral studies. *Journal of Cognitive Neuroscience*, 14(1), 79-94.
- Tyler, L. K., Marslen-Wilson, W. D., & Stamatakis, E. A. (2005). Differentiating lexical form, meaning, and structure in the neural language system. *Proceedings of the National Academy of Sciences*, 102(23), 8375-8380.
- Ullman, M. T. (1999). Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighborhood effects. *Language and Cognitive Processes*, 14, 47-67.
- Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., et al. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, 9, 289-299.
- Ullman, M. T., Izvorski, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (in press). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*.
- Vendler, Z. (1957). Verbs and times. *Philosophical Review*, 66, 143-160.

Figure 1: Mean past tense acceptability judgments of novel verbs varying in phonological and semantic similarity to English irregular verbs.

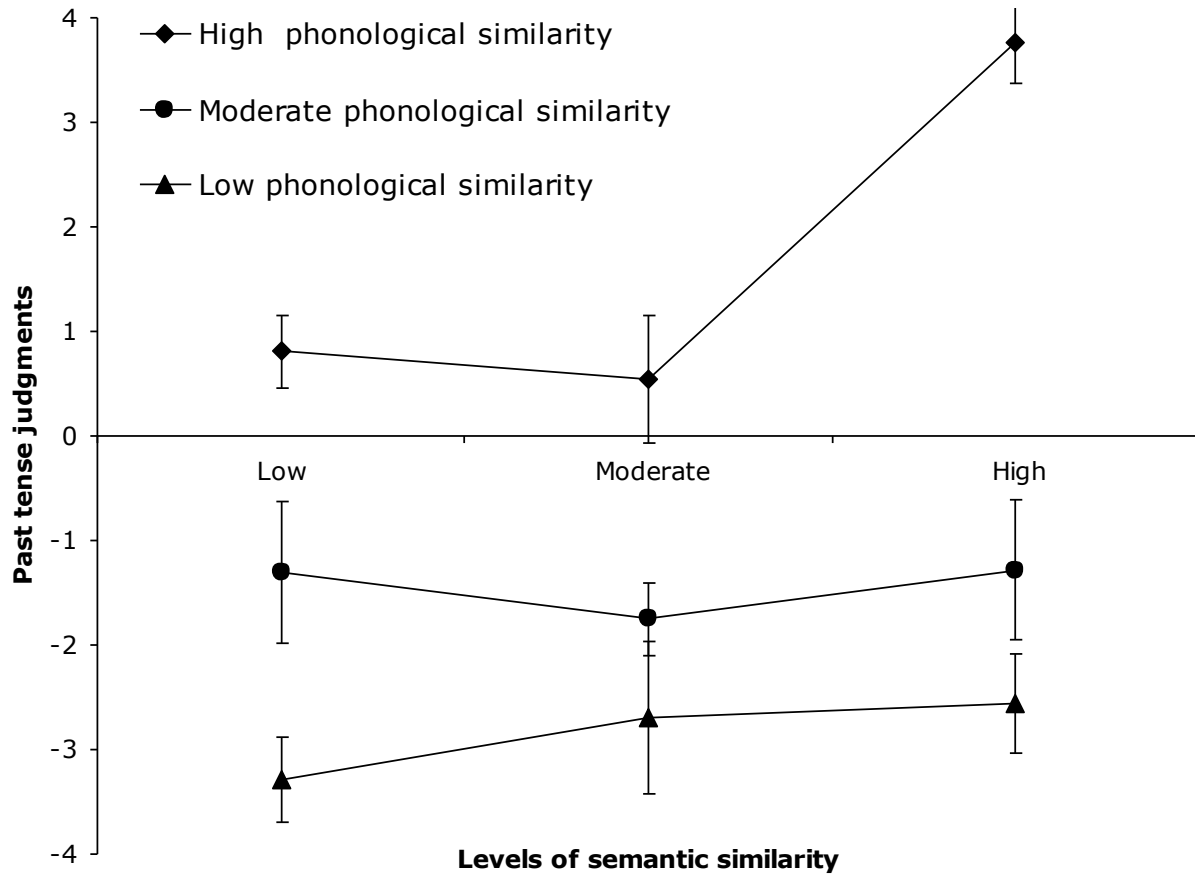


Figure 2: Percentage of novel words analogized to the target irregular verb in Experiment 1

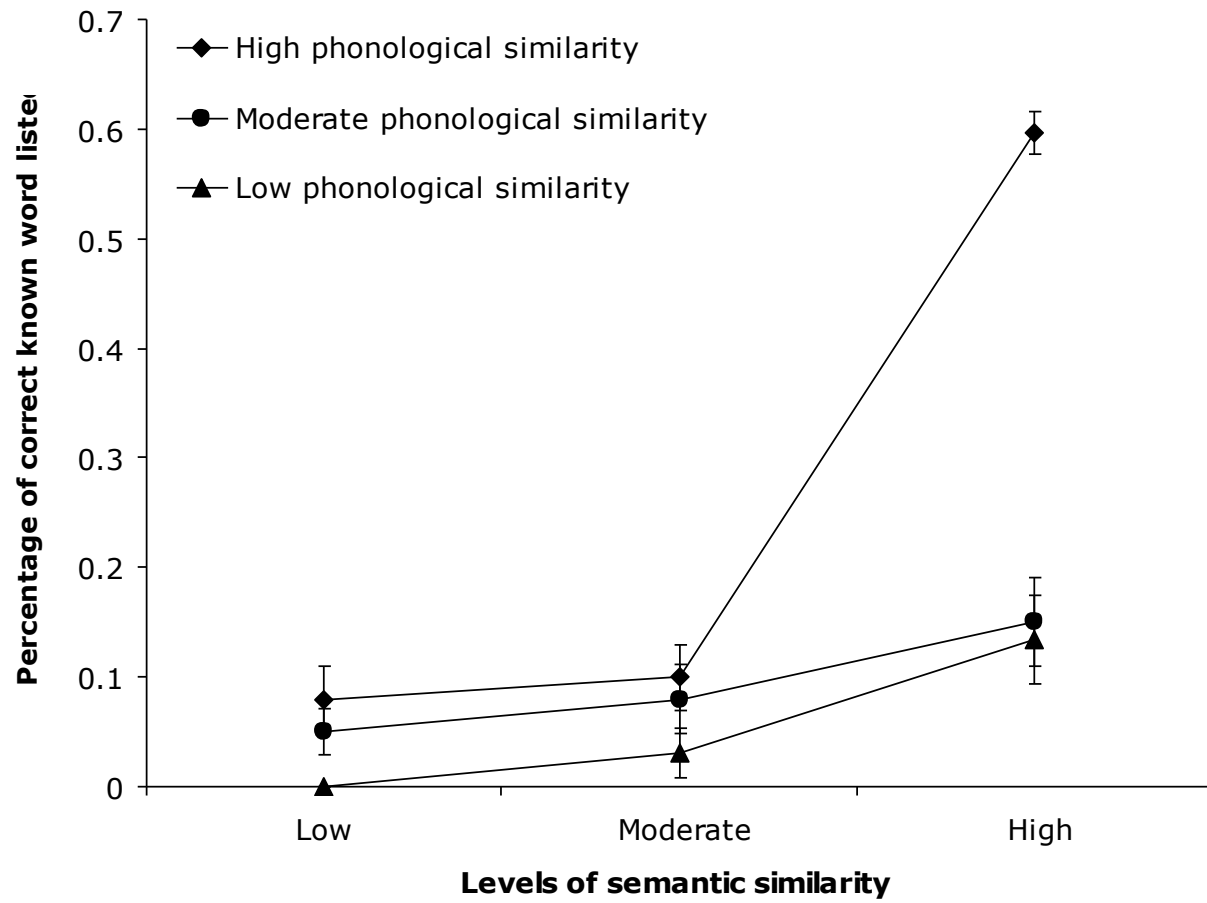


Figure 3: Residual means of past tense acceptability judgments in Experiment 1, with reported analogies to the target verb as the covariate.

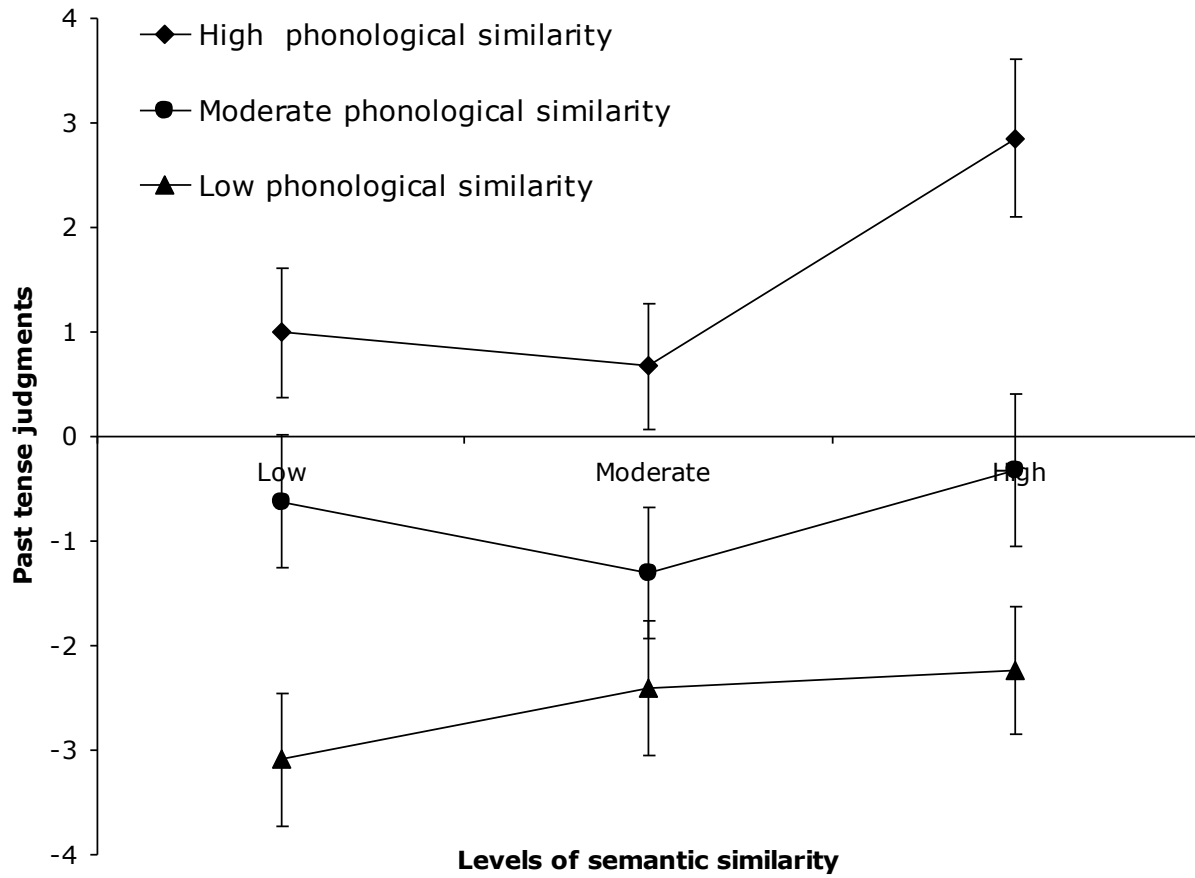


Table 1: Summary of linear regression analysis for variables predicting past tense judgments in Experiment 1 (N=62)

Variable	<i>B</i>	<i>SE B</i>	β
Known word	3.97	1.58	0.33*
Similar sound	2.04	0.65	0.70**
Similar meaning	0.45	0.63	0.14
Sound by meaning interaction	-0.14	0.34	-0.14
R^2		0.64	
F		25.92**	

Note: * $p < .05$. ** $p < .01$.

Table 2: Summary of linear regression analysis for variables predicting past tense judgments in Experiment 2

Variable	No prototypes (n=69)			All verbs (n=92)			Match verbs (n=18)			Existing verbs (n=51)		
	<i>B</i>	<i>SE B</i>	β	<i>B</i>	<i>SE B</i>	β	<i>B</i>	<i>SE B</i>	β	<i>B</i>	<i>SE B</i>	β
Semantic similarity	0.17	0.18	0.07	0.14	0.10	0.09	0.19	0.51	0.04	0.01	0.24	0.01
Headlessness	-5.69	0.55	-0.80**	-5.72	0.48	-0.80**	-6.04	0.32	-0.92**	-6.60	0.69	-0.88**
R^2		0.71			0.73			0.87			0.78	
F		82.77**			119.21**			74.80**			81.00**	

Note: * $p < .05$. ** $p < .01$.

Table 3: Properties of English irregular and regular verbs estimated in Experiment 3

Variable	Past tense form		T-test
	Irregular	Regular	
Variable 1. Similarity to regulars: rating of similarity in meaning to regulars, lower numbers indicate that the verb is less similar to regular verbs in the same family (collapsing over regular and irregular targets)	2.7 (SD 0.5)	2.8 (SD 0.7)	$t(117) = 0.97, p > .30$
Variable 2. Similarity to irregulars: rating of similarity in meaning to irregulars, lower numbers indicate that the verb is less similar to irregular verbs in the same family (collapsing over regular and irregular targets)	3.0 (SD 0.7)	2.7 (SD 0.5)	$t(117) = 2.63, p < .05$
Variable 3. Headedness judgment: rating of whether the verb is based on a noun, lower numbers indicate that the verb is more basic	2.5 (SD 1.2)	3.6 (SD 1.7)	$t(117) = 3.99, p < .001$
Variable 4. Primary noun entry: proportion of verbs with noun as its first entry, 0 if an item's first entry was a verb and 1 if it was a noun	0.14 (SD 0.4)	0.46 (SD 0.5)	$t(117) = 3.92, p < .001$
Variable 5. Noun entry: proportion of verbs with additional noun entry, 0 if an item possessed only verb entries in the dictionary, 1 if it also possessed a noun entry	0.61 (SD 0.5)	0.72 (SD 0.4)	$t(117) = 1.50, p > .20$
Variable 6. Borrowed: proportion of verbs with foreign etymology, 0 if the verb had a Germanic root, 1 if it did not	0.13 (SD 0.3)	0.35 (SD 0.5)	$t(117) = 2.93, p < .01$
Variable 7. Date of entry: year of entry into the English language as listed in the OED	1070 (SD 218)	1321 (SD 293)	$t(117) = 5.33, p < .001$
Variable 8. Past frequency: log number of past form occurrences per million in the Francis and Kucera corpus	0.82 (SD 0.8)	0.33 (SD 0.5)	$t(117) = 3.98, p < .001$
Variable 9. Present frequency: log number of present form occurrences per million in the Francis and Kucera corpus	1.1 (SD 0.7)	0.4 (SD 0.6)	$t(117) = 5.66, p < .001$
Variable 10. Total frequency: log number of total occurrences per million in the Francis and Kucera corpus	1.3 (SD 0.7)	0.60 (SD 0.6)	$t(117) = 5.91, p < .001$

Table 4: Correlation coefficients in Experiment 3.

	Regularity	Similarity to regulars	Similarity to irregulars	Headedness judgment	Primary noun entry	Noun entry	Borrowed	Date of entry	Past frequency	Present frequency	Total frequency
Regularity	1	0.09	-.24**	.35**	.34**	0.11	.26**	.44**	-.35**	-.46**	-.48**
Similarity to regulars	0.09	1	0.11	-0.13	-0.08	0.05	0.15	0.08	0.13	0.14	0.17
Similarity to irregulars	-.24**	0.11	1	-.18*	-.24**	-0.04	0.15	-0.12	0.10	0.10	0.13
Headedness judgment	.35**	-0.13	-.18*	1	.68**	.46**	0.03	.35**	-.34**	-.49**	-.48**
Primary noun entry	.34**	-0.08	-.24**	.68**	1	.30**	0.09	.49**	-.25**	-.31**	-.31**
Noun entry	0.11	0.05	-0.04	.46**	.30**	1	-0.07	0.08	-0.04	-0.15	-0.10
Borrowed	.26**	0.15	0.15	0.03	0.09	-0.07	1	.53**	-0.14	-.21*	-.20*
Date of entry	.44**	0.08	-0.12	.35**	.49**	0.08	.53**	1	-.27**	-.38**	-.36**
Past frequency	-.35**	0.13	0.10	-.34**	-.25**	-0.04	-0.14	-.27**	1	.76**	.89**
Present frequency	-.46**	0.14	0.10	-.49**	-.31**	-0.15	-.21*	-.38**	.76**	1	.95**
Total frequency	-.48**	0.17	0.13	-.48**	-.31**	-0.10	-.20*	-.36**	.89**	.95**	1

Note: * $p < .05$. ** $p < .01$.

Table 5: Summary of logistical regression analysis for variables predicting past tense judgments in Experiment 3

Variable	All verbs (n=119)			Old English verbs (n=62)			Middle English verbs (n=57)		
	<i>B</i>	<i>SE B</i>	<i>e^B</i>	<i>B</i>	<i>SE B</i>	<i>e^B</i>	<i>B</i>	<i>SE B</i>	<i>e^B</i>
Similarity to regulars	0.50	0.38	1.66	0.30	0.57	1.35	0.84	0.58	2.33
Similarity to irregulars	-0.96**	0.43	0.38	-0.99	0.70	0.37	-0.96	0.59	0.38
Headlessness judgment	0.30*	0.17	1.35	0.13	0.27	1.14	0.50**	0.25	1.65
Borrowed	0.92	0.67	2.50	1.49	1.74	4.44	1.08	0.79	2.94
Date of entry	0.01**	0.01	1.00	0.01	0.01	1.00	0.01*	0.01	1.00
Past frequency	-0.83**	0.37	0.43	-1.09**	0.49	0.34	-0.57	0.64	0.56
X^2		44.52			10.97			17.09	
<i>df</i>		6			6			6	
% correctly classified by past tense		78.2			72.6			78.9	

Note: e^B = exponentiated *B*. * $p < .10$. ** $p < .05$.

Table 6: Correlations among properties of verbs tested in Experiment 4.

	Doublet Status	Regularity Difference	Aspectual Correlation	Semantic Similarity
Doublet Status	1	.35*	-.26*	.07
Regularity Difference	.35*	1	-.23*	-.04
Aspectual Correlation	-.26*	-.23*	1	.14
Semantic Similarity	.07	-.04	.14	1

Note: * $p < .05$. ** $p < .01$

Table 7: Summary of linear regression analysis for variables predicting past tense judgments in Experiment 4 (N=76)

Variable	<i>B</i>	<i>SE B</i>	<i>β</i>
Aspectual correlation	-0.05	0.29	-0.21*
Semantic difference	-0.11	0.13	-0.10
<i>R</i> ²		0.06	
<i>F</i>		2.44*	

Note: **p* < .05. ***p* < .01.