SEMANTIC DENSITY AND PAST-TENSE FORMATION IN THREE GERMANIC LANGUAGES

R. Harald Baayen
Radboud University Nijmegen & Max Planck Institute for Psycholinguistics

Fermín Moscoso del Prado Martín
Medical Research Council—Cognition and Brain Sciences Unit (UK)

It is widely believed that the difference between regular and irregular verbs is restricted to form. This study questions that belief. We report a series of lexical statistics showing that irregular verbs cluster in denser regions in semantic space. Compared to regular verbs, irregular verbs tend to have more semantic neighbors that in turn have relatively many other semantic neighbors that are morphologically irregular. We show that this greater semantic density for irregulars is reflected in association norms, familiarity ratings, visual lexical-decision latencies, and word-naming latencies. Meta-analyses of the materials of two neuroimaging studies show that in these studies, regularity is confounded with differences in semantic density. Our results challenge the hypothesis of the supposed formal encapsulation of rules of inflection and support lines of research in which sensitivity to probability is recognized as intrinsic to human language.*

1. Introduction. The irregular verbs of Germanic languages such as English, German, and Dutch occupy a central position in the study of morphological processing, especially after the seminal paper by Rumelhart and McClelland (1986). Rumelhart and McClelland, and since then many others including MacWhinney and Leinbach (1991), Joanisse and Seidenberg (1999), and Plunkett and Juola (1999), argued that regular and irregular verbs are processed by a single subsymbolic processing system that is highly sensitive to the distributional form properties of regulars and irregulars in the language. Conversely, Pinker (1991, 1997, 1999), Clahsen (1999), Ullman, Bergida, and O'Craven (1997), and many others have argued for a dual-route architecture in which irregular verbs are subserved by an associative memory system and in which regulars are processed by a separate symbolic rule system. The latter researchers tend to view these two systems as competitive, such that processing proceeds either in the associative memory, or in the rule system, but never simultaneously.

Several intermediate positions have been defended as well. Baayen, Dijkstra, and Schreuder (1997) argued for a symbolic model in which the two routes operate in parallel, thus allowing synergy and temporal optimization of morphological processing. A different approach was developed by Moscoso del Prado Martín (2003) with a model consisting of a memory system with distributed representations for form, a memory system with distributed representations for meaning (derived by means of separate networks), and a subsymbolic 'rule' system for mapping the form representations onto the meaning representations. This is a dual system in that it separates processes and representations, just as the abovementioned dual-route models. It differs fundamentally from these models in that all processes and representations are fully distributed, and all representations are derived bottom-up from large corpora and databases.

Within this arena of very different research paradigms addressing processing and representation in the mental lexicon, there is one area where there seems to be consensus, namely, that the difference between regular and irregular verbs is restricted to the level of phonological (and orthographic) form. As Kim and colleagues (1991) pointed out,

* The authors are indebted to Laurie Feldman, Mirjam Ernestus, Jóhanna Barðdal, Rachel Kemps, Rob Schreuder, Peter Bosch, Ulli Dressler, Werner Abraham, Beth Levin, Brian Joseph, Norma Mendoza-Denton, and two referees for their comments and discussion. This research was made possible by a PIONIER grant of the Dutch national science foundation (NWO) to the first author.

the past tense does not seem to depend directly on recurring semantic distinctions. Similarity in meaning does not seem to go hand in hand with similarity in form. Although *slap*, *hit*, and *strike* are similar in meaning, they have quite different past-tense forms. Conversely, the class of irregular verbs *sting*, *sing*, *drink*, *shrink*, *swing*, *sling*, *spring*, *stink*, and *ring* share the same vocalic alternation, but have no obvious semantic features in common. Thus, it would seem that the form system operates independently from the semantic system, in line with strong modularity assumptions about human cognition.

The strict modular separation of form and semantics espoused by the dual-route tradition is not a starting point for most connectionist researchers. For them, the question of whether form and meaning interact is an empirical question that requires detailed examination on a case by case basis.

As an example of interaction of form and meaning in the grammar of English, consider the following list of Germanic irregular noun plurals (excluding pairs with voicing alternation in the stem-final obstruent, such as *wife-wives*, where the plural contains the regular plural suffix).

(1) a. man-men, woman-women, child-children, brother-brethren
b. die-dice, tooth-teeth, foot-feet, penny-pence
c. ox-oxen, goose-geese, louse-lice, mouse-mice
d. grouse, snipe, sheep, moose, deer, plaice, salmon, cod, hake
e. reindeer(s), elk(s), swine(s), antelope(s), trout(s), fish(es), carp(s), pike(s), herring(s), flounder(s)

The nouns listed in 1a are words for people. Those in 1a, b, and c have plurals that are used relatively frequently compared to their singulars (cf. Tiersma 1982, Anshen & Aronoff 1988, Baayen et al. 1997). The nouns in 1c, d, and e all refer to animals. The nouns in 1d share the property that they have a plural that is identical to the singular. The nouns listed in 1e have two plural forms, either a plural in *-s* or a plural that is identical to the singular. Which of their two plural forms is used varies; the zero plural forms are more likely to be used when the animals are considered in mass as food or game.

It is clear that the Germanic irregular nouns of English, although formally and etymologically highly heterogeneous, pattern along lines of semantic similarity (see e.g. the description in Quirk et al. 1985). The probability that we find only animal names among the 1d nouns for which the plural is always completely identical to the singular by chance is vanishingly small, at least from a synchronic point of view. So what this example shows is that irregulars can cluster along dimensions of semantic similarity. For the irregular verbs of English, there are no obvious semantic clusters, but for the irregular nouns of English, there are. It is important to keep in mind that this observation does not imply that simple deterministic semantic rules might be formulated for the English nouns, rules such as 'names for fish have a plural that is identical to the singular'. There are just too many exceptions to such a putative rule, from small (*sardine*) to big (*marlin*). The point we want to make is that there are various subsets of irregular nouns that can be characterized by the coincidence of a formal property (e.g. no overt suffix) and a semantic trait (e.g. reference to an animal). In other words, formal irregularities and semantic similarities can go hand in hand. In derivation, semantics may similarly play a role, in fact, semantic similarity may even be a constitutive part of a productive rule (see e.g. Bauer 2001:134 for semantic constraints on word formation).

In the connectionist literature, Hahn & Nakisa 2000 is an example of a study that addresses a similar phenomenon in the inflection of German proper nouns, where the distinction between surnames and Christian names may codetermine the choice of the plural suffix.

When it comes to the past tense in English, German, or Dutch, however, there is no obvious semantic difference to guide the choice between regular and irregular. In Joanisse and Seidenberg's connectionist model (1999:7593), one therefore finds semantic units, but these units serve no systematic purpose with respect to the formation of past-tense forms. Joanisse and Seidenberg point out that their semantic representations do not capture semantic similarities between verbs. Semantic similarities are taken to be crucial for other phenomena, but to be irrelevant for the past tense. Here, for once, they find themselves in remarkable agreement with Pinker and Ullman (2002). Patterson and colleagues (2001), however, reported that patients with varying degrees of severity of semantic dementia had selective problems with irregular past-tense formation. They trace this specific problem for irregular past-tense forms to the reduced phonological support that irregulars receive compared to regulars (which share the past-tense suffix), as in the computational model of Joanisse and Seidenberg. In this model, irregulars come to depend more on their semantic connectivity, and when this connectivity degrades, performance is affected specifically for irregulars. Patterson and colleagues also observed that a patient's degree of deficit on irregular verbs correlated with performance on a synonym judgment task. Interestingly, McClelland and Patterson (2002) also suggested that semantic effects during acquistion might leave their mark on the adult system, but they provided no distributional evidence to substantiate this intuition.

But there are several indications in the literature that if a regular verb is conceptualized in the semantic neighborhood of an irregular, it may be inflected irregularly. Evidence supporting the possibility of such local semantic attraction is discussed in Bybee & Slobin 1982 and Ramscar 2002. Bybee and Slobin (see also Bybee 1985 and Harris 1993) report an experiment in which participants were asked to name under considerable time pressure the past tense of verbs presented in the present tense. Interestingly, a substantial number of errors involved irregular past-tense verbs that were incorrectly supplied for regular verbs that were similar in meaning. Thus, the stimulus *seat* was often responded to with the past-tense form *sat* instead of *seated*, and *search* similarly sometimes elicited *sought* as past tense instead of *searched*. These data suggest that participants arrived in appropriate semantic domains in lexical memory but then were attracted into the similarity space of the irregular near-synonym.

Ramscar (2002) reports a series of experiments addressing the question of whether semantic similarity to an existing verb affects the past-tense form produced for a nonce verb. He shows that when participants are familiarized with a nonce verb such as *frink* or *sprink* in a context biasing for the meaning of *drink*, they are more likely to respond with an irregular past-tense form (*frank* or *sprank*) than when the context biases for the meaning of *wink* or *blink*. For the latter context, participants were more likely to produce the regular past-tense forms *frinked* or *sprinked*.

In summary, irregularity may interact with semantics in inflection, but thus far, distributional evidence for English is restricted to pluralization of nouns. There is also experimental evidence suggesting that irregular verbs can be semantic attractors, but it is unclear whether this is the result of idiosyncratic analogy or whether it is grounded in the language system itself.

The aim of the present article is to argue that indeed the local attraction effects are grounded in subtle systematic distributional differences in semantic density between

regulars and irregulars. To avoid any misunderstanding, we are NOT claiming that there are deterministic semantic 'rules' governing the past tense. What we do claim is that there is a conspiracy of subtle probabilistic (graded) semantic distributional properties that lead to irregulars having somewhat different semantic properties compared to regulars. More specifically, we argue that irregulars tend to entertain more lexical relations and tend to be more similar to each other in semantic space than is the case for regulars. This greater semantic density may have contributed to the resistance of irregulars through time to regularization. The greater semantic density of irregulars also questions the current interpretation of the neurolinguistic experimental data fueling the debate about single- versus dual-route models.

In what follows, we first present a series of lexical statistical surveys that document a number of semantic dimensions along which regulars and irregulars reveal graded differences. We then proceed to show how these differences affect association norms, subjective familiarity ratings, and response latencies in visual lexical-decision and word naming. Finally, we inspect the data sets of two neuroimaging studies that have been used to argue for the dual-route model. We show that these data sets suffer from subtle confounds of regularity and semantic density, and we discuss the consequences of our findings for linguistic theory.

2. LEXICAL STATISTICS. The aim of this section is to ascertain whether regular and irregular verbs might differ systematically along dimensions of semantic similarity. At the outset, it should be kept in mind that it is unrealistic to expect to find clear-cut, near-deterministic differences—such differences would have been noticed long ago and would not need lexical statistics in order to be brought to light. Instead, what we expect to find are graded differences for semantic random variables, differences that are visible in the means of distributions that at the same time overlap substantially.

There are two methodological decisions underlying all analyses to be reported here that we need to make explicit from the very beginning. The first decision is to investigate only monomorphemic verbs, and to leave complex verbs for later research. Monomorphemic verbs are the basic lexical units of the language, and they provide an excellent starting point for our study while avoiding the many complications that arise with respect to compositionality and semantic transparency when semantic similarity has to be considered for morphologically complex words. The second methodological decision is that we always rigidly adhere to the analyses in our data sources, even when we would prefer alternative analyses. For instance, if the CELEX lexical database (Baayen et al. 1995) lists a given word as a simplex verb and the corresponding noun as obtained from that verb by conversion, we stick with this analysis even though for particular items one might be inclined to change the direction of the conversion. This methodological rigor is necessary to avoid circularity, with reanalyses potentially reshaping the data to fit preconceived hypotheses.

This section begins with an investigation of synonym sets and regularity. We then study the distribution of auxiliaries for regular and irregular verbs in German and Dutch as well as a related semantic variable, the distribution of regular and irregular verbs in Levin's argument-structure alternation classes (Levin 1993). Finally, we address probabilistic asymmetries in the frequencies of use of regular and irregular inflectional paradigms. Throughout our discussion, we address the relation of our semantic variables to frequency of occurrence. It is well known that irregular verbs tend to have higher frequencies than do regular verbs. Most of the semantic variables that we have studied are correlated with frequency. Interestingly, our variables capture aspects of regularity over and above frequency (i.e. even after frequency is partialled out).

2.1 SYNONYM SETS IN ENGLISH AND DUTCH. WordNet (Miller 1990, Fellbaum 1998) is an electronic lexical database for English, structured to reflect lexical organization in human memory. In this database, nouns, verbs, adjectives, and adverbs are organized into synonym sets, to which we refer as synsets. Each synset represents one underlying lexical concept. The synsets themselves are linked by various lexical relations. In this section, we investigate whether regular verbs and irregular verbs might differ in their distribution over the synsets in WordNet.

SYNSETS, FREQUENCY, AND REGULARITY. We selected all verbs marked as monomorphemic in the English lemma lexicon in the CELEX lexical database for analysis. From this set of verbs, we excluded the defective verbs *ought*, *might*, *may*, *must*, *should*, *would*, and *shall*, which have no proper past-tense form, as well as the verbs *wit* and *wend*. The verbs *lump*, *port*, *spoil*, and *heave* have ambiguous analyses in CELEX, and were excluded as well. Verbs with both regular and irregular forms (e.g. *lie*, *hang*, *shine*) were classified as irregular. In this way, we obtained a database of 1,600 verbs, 146 irregular and 1,454 regular. For each verb, we added its lemma frequency in CELEX, cumulating the frequencies of homographs like *bank*.

From WordNet, we extracted the verb synsets, the sets of verbs considered to be (roughly) synonymous. Examples of verb synsets are given in 2.

(2) breathe, take a breath, respire
choke
hyperventilate
aspirate
burp, bubble, belch, eruct
force out
hiccup, hiccough
sigh
exhale, expire, breathe out
hold

Note that a synset may consist of a single verb, and that a synset may also comprise both simplex verbs, complex verbs, and phrasal verbs such as *take a breath* or *breathe out*. Verbs may occur in more than one synset. For instance, *grow* is listed in synsets such as 3.

(3) grow, develop, produce, get, acquire
mature, maturate, grow
originate, arise, rise, develop, spring up, grow

For each of the verbs in our database, we counted the number of different synsets in which that verb appears. This count provides an estimate of the number of meanings for a given verb.

The left panel of Figure 1 plots number of synsets as a function of lemma frequency in the bi-logarithmic plane. The dots represent the regular verbs, the triangles the irregular verbs. The solid and dashed lines are nonparametric regression lines (i.e. exploratory regression lines that need not be linear; see e.g. Cleveland 1979) for the regulars and irregulars respectively. We used the natural logarithms of the frequency and synset counts because both counts have highly skewed distributions. Without the log transformation, most of the data points would be lumped together in the lower left corner of the plot. Not only would it be difficult to see the pattern in the data, but many statistical techniques would not be applicable because they require variables to have more or less

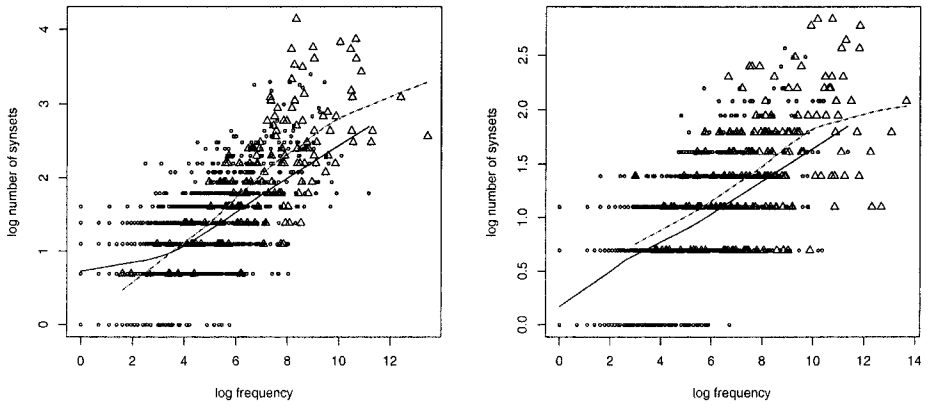

FIGURE 1. Number of synsets as a function of lemma frequency for English (left panel) and Dutch (right panel) regular (dots) and irregular (triangles) simplex verbs. The lines represent nonparametric regression lines for the regular (solid line) and irregular (dashed line) verbs.

symmetrical distributions. (Technically, the logarithmic transform amounts to exchanging a linear model for a power model.) Linear relations in the bi-logarithmic plane are often found in lexical statistics. The linear relation between the logarithm of the number of meanings and log frequency was first documented by Köhler (1986) for German.

What is of interest to us here is that the dashed regression line representing the irregular verbs lies above the solid regression line representing the regular verbs for verbs with log frequency greater than 4. This suggests that, with the exception of the lowest-frequency irregular verbs, irregular verbs might appear in more synsets than regular verbs when matched for frequency of occurrence. To test this possibility, we fitted a multiple regression model to the data, predicting log number of synsets from log frequency and regularity. The model revealed significant main effects of log frequency ($F(1,1596) = 1487.2$, $p < 0.0001$) and regularity ($F(1,1596) = 58.8$, $p < 0.0001$) as well as a significant frequency by regularity interaction ($F(1,1596) = 29.92$, $p < 0.0001$). In order to make sure that these results do not arise due to violations of normality and linearity assumptions, we also ran this analysis using ranks. This second analysis revealed the same pattern of results (all $p < 0.001$).

These data show that number of meanings, measured in terms of number of synsets in WordNet, is not uniformly distributed across regulars and irregulars. Irregulars tend to have more meanings than regulars, especially so for the higher-frequency verbs.

To ascertain whether this pattern of results is replicable, we now turn to another Germanic language, Dutch. We selected the simplex verbs from the Dutch section of the CELEX lexical database. In contrast to the English section of CELEX, the Dutch section does not distinguish between simple verbs and verbs derived from nouns through conversion. To avoid including verbs derived from nouns, only those verbs were included for which the frequency of a nominal counterpart, if available, is less than the frequency of the verb. The resulting list of 1,545 words contained 163 irregular and 1,382 regular verbs. For each verb, we recorded its lemma frequency from the CELEX frequency counts for Dutch and the number of synsets in the Dutch WordNet (Vossen et al. 1999), which is part of the EuroWordNet distributed by the European Language Resources Association.

The right panel of Fig. 1 plots log number of synsets against log frequency for the irregular (triangles) and regular (dots) verbs. Note that the regression line for the

Dutch irregulars (dashed line) runs parallel to the regression line for the regulars (solid line). A multiple regression model fit to the data revealed main effects of frequency ($F(1,1541) = 1627.0$, $p < 0.0001$) and regularity ($F(1,1541) = 18.34$, $p < 0.0001$) and the absence of an interaction ($F < 1$). Analyses using ranks yield similar results ($p = 0.0002$ for both main effects).

Comparing the results for English and Dutch, one can observe two differences. First, the range of log number of synsets is larger in English than in Dutch. This is probably due to the English WordNet being more comprehensive than the Dutch WordNet, with more fine-grained semantic distinctions and a broader coverage of the lexicon. Second, we find an interaction in English that is absent in Dutch: in English, the two regression lines intersect; in Dutch, they run in parallel. This difference may have several sources. First, the English frequency counts in CELEX are based on a smaller corpus (18 million words) than those for Dutch (42 million words). Second, as noted above, the English WordNet is substantially more comprehensive than the Dutch Wordnet. Third, the list of simplex Dutch verbs consists almost exclusively of Germanic verbs, as latinate verbs in Dutch contain a verbalizing suffix. In contrast, the English list of simplex verbs contains non-Germanic verbs (e.g. *convolute*, *efface*, *infringe*) for which the question of irregularity in inflection does not arise. In English, the non-Germanic verbs are all regular, with perhaps *strive*, as a borrowing from French, as the lone exception (even if ultimately of Germanic origin in French) (Brian Joseph, p.c.). Consequently, there might be an imbalance of the lexical strata in the English data that is absent in the Dutch data.

What is common across the two languages, then, is that irregular verbs tend to have more meanings than regular verbs across a wide range of lemma frequencies. In other words, irregular verbs combine a property pertaining to their morphological form (having an irregular past tense) with a property pertaining to their semantics (the tendency to have more meanings). This is a first aspect of what we refer to as the greater semantic density of irregular verbs.

SYNSET PAIRS AND TRIPLETS. Given that irregulars tend to have more meanings than regulars, we can proceed to ask whether irregulars might also appear in semantic neighborhoods in which there are relatively more other irregular verbs than would be the case for regulars. In order to explore this possibility, we consider the make-up of the synsets in some more detail.

Table 1 lists the number of synsets containing two simplex verbs (first three rows) as well as the number of synsets containing three simplex verbs (last four rows), broken down by the number of irregular and regular simplex verbs in the synset. (This classification of synsets ignores complex verbs and phrasal verbs in the synsets.) For both the synset pairs and the synset triplets, the highest numbers are observed for the synsets consisting of only regular verbs. Conversely, the lowest numbers are found for the synsets containing only irregulars. In order to ascertain to what extent the observed counts are reason for surprise, we need to estimate the expected counts. We can obtain estimates of these expected counts as follows.

Suppose we have an estimate of the probability p_r of observing a regular verb in a synset. Then the probability of observing an irregular verb, p_i, is $1 - p_r$. Given these probabilities, the probability of observing two regular verbs in a synset pair is p_r^2, the probability of observing two irregular verbs in a synset pair is p_i^2, and the probability of observing a regular and and irregular verb is $p_i p_r + p_r p_i = 2p_r p_i$. These probabilities are shown in the rightmost column of Table 1, which also lists the corresponding probabilities for the synset triplets.

SYNSETS WITH TWO SIMPLEX VERBS					
REGULAR	IRREGULAR	OBSERVED COUNT	EXPECTED COUNT (TYPE)[a]	EXPECTED COUNT (TOKEN)[b]	p
2	0	455	611.93	123.09	p_r^2
1	1	207	122.89	357.83	$2p_rp_i$
0	2	79	6.17	260.07	p_i^2
SYNSETS WITH THREE SIMPLEX VERBS					
REGULAR	IRREGULAR	OBSERVED COUNT	EXPECTED COUNT (TYPE)[a]	EXPECTED COUNT (TOKEN)[b]	p
3	0	102	144.09	13.00	p_r^3
2	1	58	43.40	56.68	$3p_r^2p_i$
1	2	27	4.35	82.39	$3p_rp_i^2$
0	3	5	0.15	39.92	p_i^3

TABLE 1. Observed and expected counts (according to a binomial model) of synsets containing pairs or triplets of simplex regular and irregular verbs for English.

[a] expected counts given type-based probability estimates

[b] expected counts given token-based probability estimates

There are two options for estimating the probabilities of regulars (p_r) and irregulars (p_i). We can estimate these probabilities on the basis of the type counts of regulars and irregulars in our database. This gives us the estimate $p_r = 0.909$ for the regulars and the estimate $p_i = 0.091$ for the irregulars. When we base our estimates on the token frequencies of the verbs in our database, we obtain the estimates $p_i = 0.592$ for the irregulars and $p_r = 0.408$ for the regulars. For instance, based on the type count, the probability of observing two irregular verbs in a synset with two simplex verbs is $0.091^2 = 0.0083$. When we base our probability p_i on the token frequencies, we obtain a much larger probability, $0.592^2 = 0.3505$.

The probabilities p listed in Table 1 specify, for each kind of synset, the probability of observing that kind of synset. We view this probability as a success probability in a binomial model with $n = 741$ trials (the total number of synset pairs) in the case of the synset pairs, and $n = 192$ trials in the case of the synset triplets. The expected counts e according to the binomial model are given by $e = np$. For instance, the expected count for synset pairs with two irregular verbs, given the type-based estimate of p, is $0.091^2 \times 741 = 6.17$. When we base our probability p_i on the token frequencies, we obtain a much larger expected count (260.07), a direct consequence of the substantially higher value of the token-based p_i. Note that the expected count for tokens can be smaller than that for types: The two counts are independent of each other, and depend on the magnitudes of the respective probabilities.

The expected counts for the two probabilities differ substantially, with the type-based estimates providing a fit that is both qualitatively and quantitatively superior (the mean absolute difference between the observed and expected counts for the token-based estimates is roughly twice that of the type-based estimates, both for the pairs and for the triplets). Hence, we restrict our discussion to the type-based estimates.

First consider the synsets with pairs of simplex verbs. A chi-squared test on the observed and expected values shows that the distribution of the regular and irregular verbs over the synsets departs significantly from what one would expect under chance conditions ($\chi^2(2) = 957.5, p < 0.0001$). There are fewer synset pairs with only regular verbs than expected, and there are more synsets involving irregular verbs. While only some six synset pairs with just irregulars are predicted from the proportion of irregular and regular types in our data set, we actually observe 79 such synset pairs. For the synset triplets, similar observations can be made ($\chi^2(3) = 296.4, p < 0.0001$). While

no synset triplets with three irregulars are expected in the English data, five are observed. Synset triplets with two irregulars and one regular are likewise overrepresented compared to the expected counts. Thus, we can conclude that English irregulars appear with other irregulars in synsets to a degree that cannot be attributed to chance.

Table 2 shows that the pattern observed for English is also present in our Dutch data, even though the Dutch WordNet is much smaller than the English WordNet. Again, there are more synsets with irregulars than one would expect under chance conditions ($\chi^2(2) = 89.6$, $p < 0.0001$ for the synset pairs, $\chi^2(3) = 50.1$, $p < 0.0001$ for the synset triplets). And just as in English, the token-based probability model is qualitatively and quantitatively inferior to the type-based probability model.

SYNSETS WITH TWO SIMPLEX VERBS					
REGULAR	IRREGULAR	OBSERVED COUNT	EXPECTED COUNT (TYPE)[a]	EXPECTED COUNT (TOKEN)[b]	p
2	0	252	286.45	25.36	p_r^2
1	1	84	67.57	139.85	$2p_rp_i$
0	2	22	3.98	192.79	p_i^2
SYNSETS WITH THREE SIMPLEX VERBS					
REGULAR	IRREGULAR	OBSERVED COUNT	EXPECTED COUNT (TYPE)[a]	EXPECTED COUNT (TOKEN)[b]	p
3	0	65	74.43	1.96	p_r^3
2	1	26	26.34	16.22	$3p_r^2p_i$
1	2	11	3.11	44.72	$3p_rp_i^2$
0	3	2	0.12	41.10	p_i^3

TABLE 2. Observed and expected counts (according to a binomial model) of synsets containing pairs or triplets of simplex regular and irregular verbs for Dutch.

[a] expected counts with type-based probability estimates

[b] expected counts with token-based probability estimates

In summary, our analyses of the composition of synset pairs and triplets of simplex verbs show that in both English and Dutch, irregulars enter into more semantic relations among themselves than one would expect under chance conditions. Irregulars not only tend to appear in more synsets, but the synsets in which they appear tend to contain a higher than chance number of irregulars as well. Note that this pattern cannot be attributed simply to the difference in frequency between regulars and irregulars, as the token-based probability models make entirely the wrong predictions. We conclude that irregulars as a group are characterized by a slightly higher semantic density than are regulars, in two ways: irregulars have more meanings, and their semantic neighborhoods tend to contain more irregulars. In these analyses, we have considered only synomym sets, but we suspect that similar differences in density might be observed for sets of antonyms. In other words, that *come* and *go* are both irregular may not be a complete coincidence. We leave the study of such additional semantic relations to further research.

NUMBER OF SYNSETS AND CLASS SIZE. Thus far, we have treated the distinction between regular and irregular verbs as an absolute distinction. However, it is well known that there are different subsets of irregular verbs, some containing only one member (e.g. the suppletive pair *go-went*), others being instantiated by many verbs (e.g. the verbs sharing the alternation exemplified by the pair *weep-wept*); see e.g. Bybee & Slobin 1982. There is a simple correlation between the size of an irregular subclass and (sub)regularity, namely, that a greater class size corresponds to a more general subregularity, and that a smaller class size corresponds to greater idiosyncrasy. From this truism,

we can derive the nontrivial prediction that the greater subclasses should be more similar to the regulars in terms of the number of synsets in which they appear.

To test this prediction, we classified the English irregular verbs according to the vocalic alternation between the present-tense and past-tense forms. Verbs with no vowel change (such as *build* and *have*) were brought together in a no-alternation class. In this way, we obtained thirty-two classes of irregular verbs, ranging in class size from 1 to 38. For each class, we calculated the mean of the log-transformed number of synsets of the class members. This logarithmic transformation is necessitated by the leftward skewing of the distribution of the number of synsets and allows us to obtain a mean that is not unduly influenced by outliers. The distribution of class sizes itself has a Zipfian shape, with class size 1 being represented by twelve classes, and class size 38 by only one. Table 3 lists class size, number of classes with a given class size, and the average of the mean log-transformed number of synsets for the classes of that class size. The left panel of Figure 2 plots average mean log number of synsets against log class size (using asterisks). This figure also plots the individual classes, labeled by their most frequent member. The dashed horizontal line represents the mean log number of synsets for the regular verbs. The nonparametric regression line (the solid line) suggests a decrease in the number of synsets with increasing class size. This suggestion is supported by a Spearman correlation of $r_s = -0.72$ ($p = 0.0168$) as well as by a linear regression using weighted least squares ($F(1,9) = 11.998$, $p = 0.0071$), with as weights the number of classes of a given class size.

CLASS SIZE	NUMBER OF CLASSES	AVERAGE MEAN LOG NUMBER OF SYNSETS
1	12	2.33
2	8	2.42
3	2	1.78
4	1	2.51
5	1	2.31
6	1	2.03
7	2	1.75
8	1	1.36
10	2	1.59
17	1	1.82
38	1	1.71

TABLE 3. Vocalic alternation class size, number of classes, and average of the mean log number of synsets in the classes of the specified class size for English irregular verbs.

Interestingly, the class with the greatest size and the lowest mean log number of synsets is the set of verbs with no vocalic alternation, verbs such as *have*, *show*, and *put*. These verbs approximate regulars in that their stem remains unchanged (*put*), or in that their past form is regular (*show*; only the past participle is irregular, *shown*), or in that the past-tense form ends in an alveolar stop (*have*) reminiscent of the stop of the regular past-tense suffix. In other words, the class of irregulars that is most similar to the regulars (both in class size and in the stability of the stem in the present and past-tense forms) is also more similar to the regulars in terms of semantic density.

For the simplex verbs in our database for Dutch, we carried out a similar analysis. Table 4 lists class size; number of classes with this class size, and the average mean log number of synsets; the right panel of Fig. 2 provides the corresponding scatterplot, a nonparametric regression (solid line), and the average for the Dutch regular verbs (dashed line). As in English, we observe a negative correlation ($r_s = -0.75$, $p = 0.0047$; a weighted least-squares analysis yields $F(1,10) = 10.25$, $p = 0.0095$).

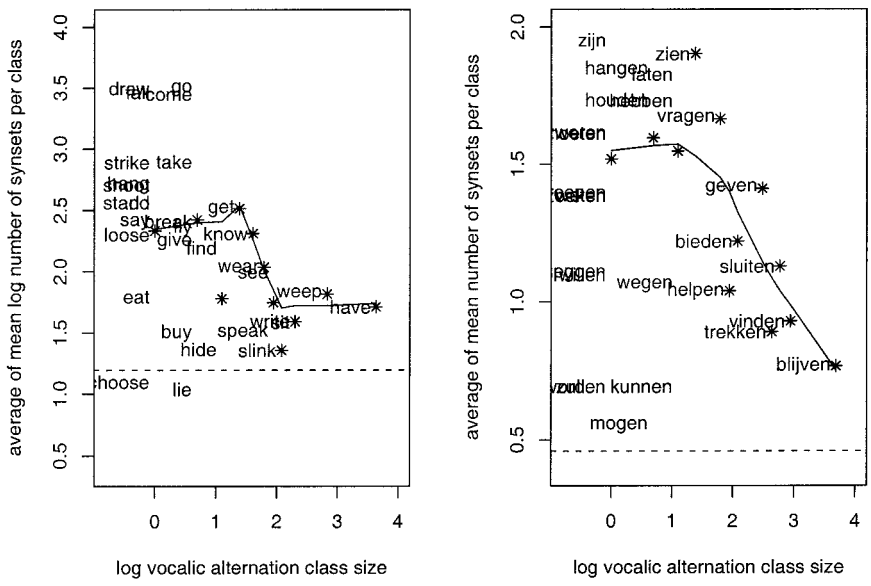

FIGURE 2. Correlation between log vocalic alternation class size (horizontal axis) and the average of the mean log number of synsets in classes of this size (vertical axis) for English (left panel) and Dutch (right panel) irregular verbs. Averages represented by asterisks; verbs represent highest-frequency verb for each of the individual vocalic alternation classes. Dashed line represents mean log number of synsets for regular verbs.

A greater class size implies a more prominent subregularity, which in turn emerges as linked to a smaller semantic density approaching the low semantic density (0.4648) of the regular verbs.

In summary, our analyses of the distribution of regular and irregular verbs over synsets in English and Dutch has revealed that irregulars have a greater semantic density, and that the degree of semantic density itself is positively correlated with the degree of irregularity measured by means of class size.

2.2 AUXILIARIES IN GERMAN AND DUTCH. We have seen that regular and irregular verbs in Dutch differ in semantic density. We have also seen that within the class of

CLASS SIZE	NUMBER OF CLASSES	AVERAGE MEAN LOG NUMBER OF SYNSETS
1	14	1.52
2	4	1.59
3	5	1.55
4	1	1.90
6	1	1.66
7	1	1.04
8	1	1.22
12	1	1.41
14	1	0.89
16	1	1.13
19	1	0.93
40	1	0.77

TABLE 4. Vocalic alternation class size, number of classes, and average of the mean log number of synsets in the classes of the specified class size for Dutch irregular verbs.

irregular verbs, irregularity is itself graded, with the larger vocalic alternation classes being more similar to the regulars. In this section, we return to the simple qualitative distinction between regulars and irregulars, in order to address yet another way in which regulars and irregulars might differ, namely in their aspectual semantics.

Shirai and Anderson (1995) suggest that in acquisition, English aspectual semantics develop from a prototype of events that are punctual, that have an endpoint, and that have a clear resulting state. Given that the proportion of irregular verbs in one's vocabulary decreases during the acquisition process as more and more (low-frequency) regular verbs are mastered, one would expect that irregular verbs as such are clustered more closely around the prototype, while regular verbs should show a wider scatter away from the prototype. Unfortunately, English does not lend itself easily to a statistical test of this possibility, because the semantic distinctions at issue are not formally marked in the grammar. In German and Dutch, however, the auxiliary required by a verb for (present and past) perfect formation is a formal marker of its aspectual properties (see e.g. Randall et al. 2003 and Lieber & Baayen 1997). Telicity, the event having an (inferable) endpoint, emerges as an especially important determinant of the use of the marked auxiliary, 'to be' (*sein* in German, and *zijn* in Dutch). Hence, we can use auxiliary selection for the present and past perfect in German and Dutch as an objective measure for ascertaining whether regular and irregular verbs differ in their basic aspectual semantics. We expect to find that irregular verbs, because they are predominant among the verb types in early acquisition and hence closer to Shirai and Anderson's prototype (1995), are characterized by a higher proportion of verbs selecting *sein* and *zijn*. Conversely, regular verbs should show a preference for the auxiliary 'have' (*haben* in German, *hebben* in Dutch). In order to avoid misunderstanding, we note that the aspectual semantics of the verb in the sentence are, of course, further codetermined by the sentential context (e.g. by whether the auxiliary is present, and by the definiteness of the verb's arguments; cf. Verkuyl 1989).

We first consider Dutch. The CELEX lexical database specifies the auxiliary (*zijn*, *hebben*, or both *hebben* and *zijn*) for 1,530 of the 1,545 verbs in our database. (Verbs that appear with both auxiliaries have different aspectual readings depending on which auxiliary is used. For instance, the verb *lopen* used with the auxiliary *hebben* has the reading of having engaged in the activity of walking. When used in combination with the auxiliary *zijn* and a prepositional phrase specifying a locative endpoint, the verb has a telic interpretation.)

The upper part of Table 5 lists the counts of regular and irregular verbs broken down by the auxiliary they select. Most verbs select *hebben*, a minority selects *zijn*, and a slightly bigger minority allows both *hebben* and *zijn*. Interestingly, Table 5 shows that

	DUTCH			
AUXILIARY	IRREGULAR		REGULAR	
hebben	109	(67.28)	1199	(87.65)
zijn	15	(9.26)	34	(2.49)
both	38	(23.46)	135	(9.87)
	GERMAN			
AUXILIARY	IRREGULAR		REGULAR	
haben	124	(71.67)	971	(88.43)
sein	20	(11.56)	36	(3.27)
both	29	(16.76)	91	(8.29)

TABLE 5. Counts (and percentages) of regular and irregular verbs with (present and past) perfect auxiliary *hebben*, auxiliary *zijn*, and with both auxiliaries, for Dutch and German.

irregulars have a relative preference for the auxiliary *zijn*, either exclusively or in combination with *have*. For instance, 9.3% of the irregular verbs allow only *zijn* as auxiliary, while only 2.5% of the regular verbs do so. A chi-squared test confirms that this nonuniform distribution is not expected under chance conditions ($\chi^2(2) = 51.43$, $p < 0.0001$).

The second half of Table 5 replicates the same pattern for German. From the CELEX lexical database, we extracted all 1,296 verbs listed as monomorphemic. For 1,271 of these verbs, the selected auxiliary is available. As in Dutch, some verbs allow both *sein* and *haben* as auxiliary, again with a concomitant change in aspect. The nonuniform distribution of auxiliary use for irregulars and regulars is statistically significant ($\chi^2(2) = 39.51$, $p < 0.0001$).

The key result of this analysis, both for Dutch and for German, is that the direction of the difference in the type-based counts is exactly as expected given the Shirai and Anderson (1995) study, with the irregulars revealing a graded preference for the marked auxiliary (*sein*, *zijn*). This result allows us to conclude that in addition to the quantitative difference in semantic density as revealed by the synset data, there is an additional qualitative difference in semantic density, with irregulars clustering more closely around the telicity of the semantic prototype.

2.3. VERB ALTERNATION CLASSES IN ENGLISH. Given the differences that we have observed between the regulars and irregulars with respect to the use of the auxiliaries in German and Dutch, and given the findings of Shirai and Anderson (1995), the prediction follows that the argument structures of regulars and irregulars should differ as well. For instance, as irregular verbs show a slight preference for the present and past perfect auxiliary *sein* in German and the auxiliary *zijn* in Dutch, and since these auxiliaries are typically required for intransitive verbs, we may expect that irregular verbs will also show a relative preference for argument structures with an internal argument only.

We have no detailed data on argument structure available for Dutch or German. For English, however, Levin 1993 provides a comprehensive classification of verbs according to their argument-structure alternations. An alternation class is a set of verbs that occur with the same patterns of arguments and adjuncts. An example of an alternation class is the set of verbs described as 'verbs of caring for a specific body part' (Levin 1993:34): *brush*, *floss*, *shave*, *wash*. These verbs can occur with their direct object, as in *I flossed my teeth*, but also without their direct object, as in *I flossed*. Levin lists many such detailed alternation classes, which are grouped together in 49 main alternation classes that are divided into 192 subclasses. Table 6 provides two examples. We note, as many before us, that verbs sharing an alternation class tend to share aspects of meaning in addition to syntactic properties. In what follows, we consider the 1,617 simplex verbs (1,507 regulars and 110 irregulars) in Levin's alternation classes that are listed in CELEX as verbs with nonzero frequency.

While some verbs are listed in only one alternation class, there are other verbs that appear in two or more alternation classes, with a maximum of ten. Table 7 presents counts of regular and irregular verbs broken down by the number of alternation classes in which they are listed. The left panel of Figure 3 plots the proportions of irregular verbs broken down by the number of alternation classes in which they occur using black bars, and the corresponding proportions for regular verbs using white bars. Roughly one third of the irregular verbs are listed in only one alternation class, while nearly two thirds of the regular verbs are listed in one alternation class only. The proportions of

Simplex verbs in **Levin's alternation class 43** ('sparkle', 'bang', 'reek', 'belch' verbs):

blink burn flare flash glare glint glow scintillate SHINE twinkle bang BEAT bellow blare boom buzz chime chink chug clack clank clap clash click CLING clink clump crack creak crunch cry fizz groan growl gurgle hiss hoot howl jangle jingle knock lilt murmur pop purr rap RING roar roll rumble rustle scream screech SING sizzle snap splash squawk squeal squelch STRIKE thrum thud thump tick ting tinkle toll toot trill wail wheeze whine smell STINK BLEED dribble drip drool gush pour puff seep SHED spew spill spout spurt squirt

Simplex verbs in **Levin's alternation class 48** ('appear', 'assert', 'die', 'transpire' verbs):

BREAK BURST COME flow GROW gush pop RISE SHOW spill STEAL turn declare offer shape SHOW suggest occur

TABLE 6. Levin's alternation classes 43 and 48. Irregular verbs shown in uppercase.

n	1	2	3	4	5	6	7	10
irregular	38	24	22	11	9	4	2	0
regular	989	340	133	27	13	4	0	1

TABLE 7. Counts of regular and irregular verbs broken down by the number of alternation classes *n* in which they are listed in Levin 1993.

the regulars and irregulars that are attested in two alternation classes are very similar, and as we move to the right in the barplot, the proportions of irregular verbs exceed those of the regular verbs. The only exception to this pattern is the regular verb *to roll*, the only verb to occur in ten alternation classes. A Fisher exact test of independence applied to the counts of Table 7 shows that this pattern is indeed remarkable, and unlikely to be due to chance ($p \ll 0.0001$). The pattern in the left panel of Fig. 3 bears witness to an isomorphy between morphological irregularity and syntactic irregularity. Irregular verbs are the verbs with the morphologically more complex (because less predictable) past-tense form, and they are also syntactically more complex in the sense that their arguments appear in a wider range of constructions in the sentence.

A complementary analysis takes as point of departure the size of the alternation class, that is, its productivity. Is there a relation between the productivity of an alternation class and the probability of regular verbs (estimated on the basis of the monomorphemic verbs) in that class? One would expect that more productive classes contain higher proportions of regular verbs. The right panel of Fig. 3 shows that this is indeed the

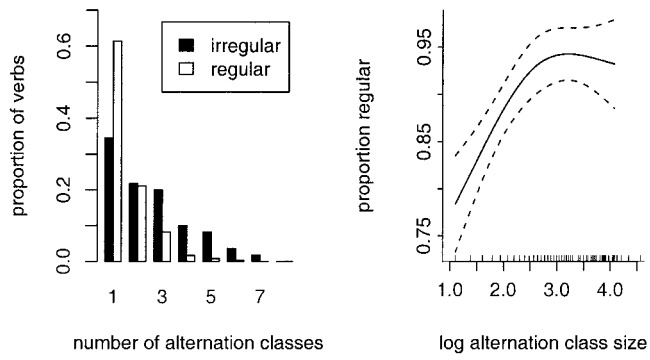

FIGURE 3. Regularity and argument-structure alternation classes. Left panel shows proportions of regular and irregular verbs broken down by number of alternation classes in which they appear. Right panel shows proportion of regular verbs in an alternation class as a function of the productivity (size) of class.

case. As the size of the alternation class increases, the proportion of regular verbs in that class increases as well, at least for the smaller alternation classes. The curve was obtained by fitting an ordinary least-squares regression model to the data with a parametric technique that does not impose linearity a priori (we used a restricted cubic spline; see Harrell 2001). Class productivity emerges as a significant predictor ($F(2,189) = 12.23$, $p < 0.0001$), but the (bootstrap-validated) R^2 is small (0.077), which indicates that the alternation class size offers only a modest contribution to predicting regularity.

In summary, irregular verbs occur relatively often in many alternation classes. Conversely, unproductive alternation classes tend to comprise more irregular verbs. This is another example of how irregulars form denser clusters, this time in syntactic space.

2.4. INFLECTIONAL PARADIGMS IN GERMAN AND DUTCH. Thus far, we have treated regular and irregular verbs as atomic units. But verbs have inflectional paradigms, which constitute the domain on which the regularity and irregularity of verbs is defined. Note that this domain is itself semantic in nature, as witnessed by the opposition between, for instance, the present tense and the past tense. This raises the question of whether regular and irregular verbs might reveal differential preferences for the present, the past, and possibly the past participle.

INTRAPARADIGMATIC ASYMMETRIES. It is known that high-frequency and low-frequency words may differ in the extent to which they make use of the inflectional possibilities offered by the language (see Baayen & Sproat 1996). We therefore inspected the following three forms of the inflectional paradigm of Dutch.

(4)

	IRREGULAR	REGULAR
PRESENT	loop	wandel
PAST	liep	wandelde
PAST PARTICIPLE	gelopen	gewandeld
gloss	'walk'	'stroll'

The form labeled PRESENT is identical to the stem and is used for the first person singular, as well as for the second person singular in interrogatives. We use the label PAST for the first, second, and third person singular in the past tense. To trace potential differences in preferences between regulars and irregulars for these three forms, we investigated, for each of the six pairs of forms, which form has the higher frequency, broken down by regularity. For instance, we calculated the proportion of regular verbs for which the present-tense form is more frequent than the past-tense form and compared this proportion with the corresponding proportion for the irregular verbs. Under chance conditions, the proportions of verbs for which the present is more frequent than the past should not differ significantly for the two regularity conditions.

Table 8 lists the counts of verbs for which the frequency of one inflectional form exceeds that of one of the others, for all six pairs of two by two comparisons, for Dutch and German. (The few occasions on which the two forms have equal frequency are not listed.) First consider the upper half of this table, which concerns Dutch.

The first three rows summarize comparisons for which there is no significant difference between regulars and irregulars. The proportions of verbs for which the present-tense form is of higher frequency than that of the participle or the past-tense form is very similar for regular and irregular verbs. The same holds for the proportions of verbs for which the participle is more frequent than the past-tense form. When we reverse the direction of the frequency comparison, the pattern changes such that the proportions

			DUTCH					
COMPARISON	IRREG	REG	nIRREG	nREG	pIRREG	pREG	p	$\chi^2(1)$
pres>part	31	229	163	1382	19.02	16.57	0.497	0.462
pres>past	7	104	163	1382	4.29	7.53	0.177	1.824
part>past	37	343	163	1382	22.70	24.82	0.618	0.248
pres<part	127	681	163	1382	77.91	49.28	0.000	46.790
pres<past	156	887	163	1382	95.71	64.18	0.000	64.624
part<past	125	667	163	1382	76.69	48.26	0.000	46.017
			GERMAN					
COMPARISON	IRREG	REG	nIRREG	nREG	pIRREG	pREG	p	$\chi^2(1)$
pres>part	24	83	187	1109	12.83	7.48	0.021	5.361
pres>past	8	32	187	1109	4.28	2.89	0.430	0.624
part>past	52	207	187	1109	27.81	18.67	0.005	7.801
pres<part	148	422	187	1109	79.14	38.05	0.000	108.007
pres<past	159	628	187	1109	85.03	56.63	0.000	52.928
part<past	123	496	187	1109	65.78	44.72	0.000	27.582

TABLE 8. Statistics for the number of verbs for which the frequency of one inflected form (present, participle, past) exceeds that of the other form, for Dutch and German.

Irreg, Reg: counts; nIrreg, nReg: totals; pIrreg, pReg: percentages; p, $\chi^2(1)$: statistics of proportions tests.

for the irregulars are always higher than those of the regulars. This is shown in the second set of three rows in the upper half of Table 8.

To understand this pattern, first consider the comparisons involving the present-tense form. For both regular and irregular verbs, the present-tense form is fully regular, by definition, as it consists simply of the stem. The proportion of verbs for which this regular form is used more frequently than the past-tense or participial forms is similar for both regular and irregular verbs. When we consider the reverse comparisons, that is, when we compare how often a past tense or past participle is more frequent than the present-tense form, we find that the irregular forms are more often of higher frequency.

Now consider the comparisons involving the past-tense forms and the past participles. For irregular verbs, the past participle is more regular than the past-tense form. The past-tense participle always contains the prefix *ge-*, irrespective of whether a simplex verb is regular or irregular. In addition, past-tense participles always contain a suffix, which for some irregular verbs can be identical to that used for regular verbs. Given that the past participle is more regular than the past-tense form, we expect to find no significant difference between regular and irregular verbs when we consider the proportions of verbs for which the more regular participle has a higher frequency than the less regular past-tense form. Conversely, when we compare the frequencies of the less regular past-tense forms with those of the more regular past participles, we should observe higher proportions for the irregulars. This is exactly what Table 8 shows.

In summary, within their paradigms, less regular forms are used more intensively than more regular forms. This higher frequency may protect the less regular forms against regularization. Note that the differences in the intensity of the use of the present, the past, and the past participle between regulars and irregulars point to semantic differences: not to differences that merely relate to coindexing of verbal arguments in the syntax, but to differences that relate to how the verb is used with respect to tense (present versus past) and aspect (the present and past perfect). In other words, how often we use a past-tense form in, for instance, a narrative is codetermined, surprisingly, by the morphological regularity of the verb.

We also checked whether the pattern observed for Dutch generalizes to German. The lower half of Table 8 shows that a similar pattern is obtained, with small differences for the first three comparisons and large differences for the second three comparisons. In contrast to Dutch, two of the first three comparisons are also significant, although the magnitude of the chi-squared values provide less reason for surprise. This difference may be due to the smaller size of the corpus underlying the German counts (6 million) compared to that of Dutch (42 million). Alternatively, it may be due to a genuine difference between Dutch and German verbal use. We leave this issue for further research.

INFLECTIONAL ENTROPY. The reader has just seen that there are differences in how regular and irregular verbs make use of the possibilities offered by their inflectional paradigms. An overall measure for gauging differences in how frequently the inflectional forms in a verb's paradigm are used is the inflectional entropy, a measure developed in Moscoso del Prado Martín et al. 2004. In information theory, entropy is a measure for the amount of information (Shannon & Weaver 1949). When applied to the probability distribution of the distinct forms with nonzero frequency in the inflectional paradigm of a verb, the entropy is a measure of the complexity of that verb's inflectional paradigm. The inflectional entropy increases with the number of distinct forms in the paradigm. The inflectional entropy also increases the more the different inflectional forms are used to the same extent.

Figure 4 visualizes the distributions of the inflectional entropies of regular and irregular verbs in Dutch and English. We used the same set of verbs as in §2.1 for the analysis of the number of synsets. Fig. 4 shows that the inflectional entropies in Dutch are somewhat larger than those in English, a consequence of the greater number of inflec-

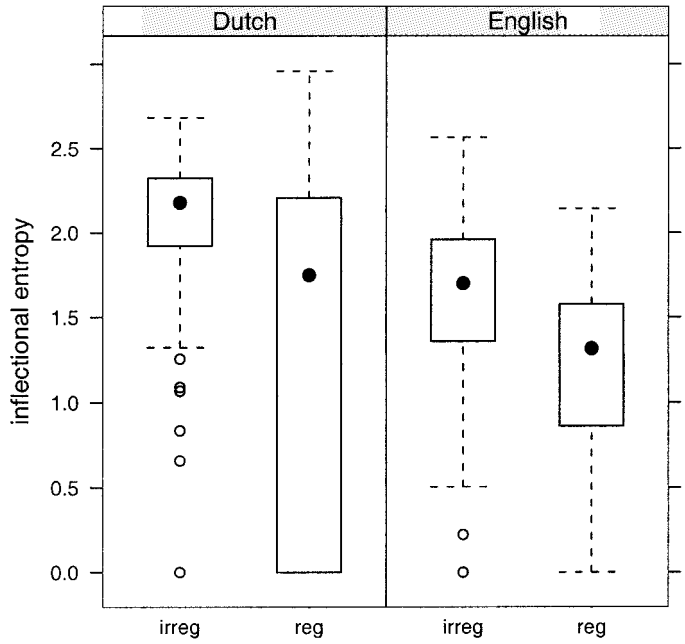

FIGURE 4. Boxplots for inflectional entropy in Dutch and English for regular (reg) and irregular (irreg) verbs. Black disks represent the medians, the boxes the interquartile range; whiskers extend to 1.5 times the interquartile range. Individual outliers are represented by circles.

tional forms in the verbal paradigm of Dutch. In both languages, furthermore, the inflectional entropy tends to be smaller for regular verbs than for irregular verbs. This shows that the frequencies with which the inflectional variants of an irregular verb are used are more similar than those of a regular verb. These differences are statistically highly significant and not confounded with frequency, as shown by an analysis of covariance in which inflectional entropy is modeled as a function of frequency, language, and regularity (all p-values $\ll 0.0001$). This provides further confirmation of the subtly different ways in which regular and irregular verbs are used.

2.5. A MODEL FOR (IR)REGULARITY. We have seen that there are several, often related, semantic dimensions on which regular and irregular verbs pattern in subtly different ways. To what extent do measures such as a verb's number of synsets, its inflectional entropy, and the number of alternation classes in which it occurs contribute independently to explaining whether a verb is regular or irregular?

In order to address this question, we fitted a logistic regression model to the data for the 1,617 simple verbs for which we have information about their argument-structure alternation classes. Logistic regression is a technique that allows us to assess the probability that a verb is regular (or irregular) from the values of a set of predictors. It allows us to ascertain whether a predictor makes an independent and significant contribution to explaining regularity.

Logistic regression does not estimate probabilities directly from the predictors; it provides these estimates indirectly through the log of the odds ratio (the log of the ratio of successes and failures). In the simplest case, the relation between a predictor and the log odds ratio is linear. After back-transforming the log odds ratios to probabilities, the relation is no longer linear, but curved upwards or downwards. The relation between a predictor and the log odds ratio may also be nonlinear, in which case the probability curve takes a more complex form. We did not impose a priori restrictions on the functional shape of nonlinear relations (by using restricted cubic splines) and retained nonlinear relations in the model only when they were statistically significant. Significance is evaluated with the Wald statistic, which follows a chi-squared distribution. For technical details, the reader is referred to Harrell 2001.

Figure 5 summarizes the partial effects of six statistically significant predictors for regularity in English. The partial effect of a predictor is its effect when the other variables in the model are held constant; it is its independent contribution. The six predictors for the present data set of 1,617 simple verbs are log word frequency, inflectional entropy, log number of alternation classes, log number of synsets, and two variables that we have not yet discussed, the morphological family size and the noun-to-verb ratio.

The morphological family size of a verb is the type count of morphologically complex words in which that verb occurs as a constituent. This measure captures the semantic connectivity in the mental lexicon that is overtly expressed in the morphology (see Schreuder & Baayen 1997, Bertram et al. 2000, De Jong et al. 2000).

The noun-to-verb ratio is the log of the ratio of a form's frequency as a noun to that form's frequency as a verb. (We added 1 to both frequencies in order to avoid arithmetic problems with zero frequencies.) When the noun-to-verb ratio is zero, the verbal and nominal frequencies are identical. When it is less than zero, the verb is more frequent than the noun. When it is greater than zero, the noun is more frequent than the verb. We included this variable because it has been argued that denominal verbs must be regular (Kim et al. 1991, but see Ramscar 2002 for counterevidence). Since we cannot predict regularity from conversion when conversion is determined by regularity, we

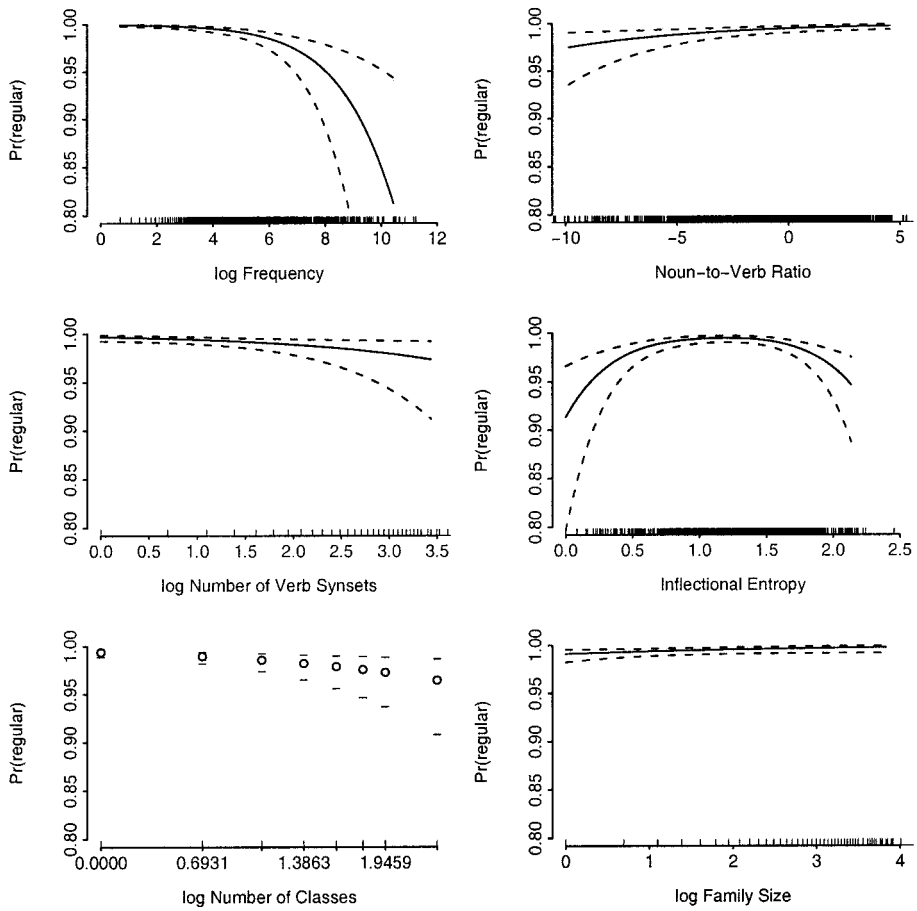

FIGURE 5. Partial effects of six predictors in a logistic regression model for the probability of an English verb being regular. Rugs on the horizontal axis highlight distribution of each predictor.

make use of the noun-to-verb ratio as a measure of the likelihood of conversion from noun to verb (a high positive noun-to-verb ratio) or from verb to noun (a large negative noun-to-verb ratio). The prediction is that the likelihood of being regular increases with the noun-to-verb ratio.

The upper left panel of Fig. 5 shows the partial effect of frequency. As frequency increases, the likelihood that a verb is regular decreases, as expected ($\chi^2(1) = 26.36$, $p < 0.0001$). The upper right panel shows that the probability that a verb is regular increases slightly with increasing noun-to-verb ratio, as predicted ($\chi^2(1) = 8.94$, $p = 0.0028$). The first panel on the second row shows that the probability of being regular decreases slightly with increasing number of synsets ($\chi^2(1) = 5.77$, $p = 0.0163$), and a similar slight increase is also apparent in the lower left panel for the productivity of the verb's argument-structure alternation class ($\chi^2(1) = 10.76$, $p = 0.0010$). A much larger, nonlinear effect emerges for inflectional entropy ($\chi^2(2) = 33.28$, $p < 0.0001$). Regular verbs tend to have medium-sized inflectional entropies. Finally, there is a marginal effect for log morphological family size ($\chi^2(1) = 3.34$, $p = 0.0676$), which suggests that morphological families might be slightly larger for regular verbs. The

bootstrap-validated R^2 (0.45) and the bootstrap-validated Somers's *Dxy* rank correlation (0.82) are close to the original values (0.46 and 0.83), which indicates that we are not overfitting the data. This allows us to conclude that each of our semantic variables is a significant predictor even when all of the other variables are taken into account at the same time. For a comparable logistic regression model for regularity in Dutch, the reader is referred to Tabak et al. 2005.

3. BEHAVIORAL DATA. Our lexical statistical survey revealed statistically significant semantic differences between regular and irregular verbs. Irregular verbs tend to have more meanings, and their semantic neighborhoods tend to contain more irregular verbs. These differences in what we refer to as semantic density, however, are graded differences. Although the distributions of any of the measures that we have examined differ significantly in central tendency, we are always dealing with substantially overlapping distributions. This raises the question of to what extent the subtle graded differences in semantics that we have observed might be relevant for storage and computation in the mental lexicon. Although probability has been shown to play an important role in various domains of morphological structure (see Ernestus & Baayen 2003, Baayen 2003, Albright & Hayes 2003 for linguistic studies, and Feldman & Prostko 2001 for a processing study), it remains an empirical question whether the attested subtle semantic differences between regular and irregular verbs would show up in behavioral data.

In what follows, we discuss evidence from association norms, visual lexical-decision and word-naming latencies, as well as familiarity ratings suggesting that the differences in semantic density between regular and irregular verbs do indeed have processing consequences.

3.1. ASSOCIATION NORMS. The observed higher semantic density of irregular verbs predicts that they should elicit more and more tightly linked responses in a free association task. We therefore inspected the University of South Florida association norms made available by Nelson, McEvoy, and Schreiber (1998). They provide a large database with a great many measures obtained from a large series of free association experiments. Nelson and colleagues presented words to participants and asked them to write down, for each word, the first word that they could think of that was related in meaning or associated to that word. A given participant was required to list a single associate for each word. Appendix A of their study contains 5,019 normed words and their 72,176 responses. From this appendix, we extracted the cue set size, the mean connectivity, and the resonance strength for the verbs in our database for which association norms are available. This resulted in a database of 692 verbs, 122 irregulars and 570 regulars.

Cue set size is an index of how many near neighbors a word has. It is calculated by counting the number of different target responses given for a cue word by two or more participants. The mean connectivity of a cue word is calculated by comparing the associates produced by separate groups of participants and counting the number of connections among the associates in the set, normalized by the size of the set. This index is described as capturing the density and in some sense the level of organization among the strongest associates of the cue.

The resonance strength of a word is based on forward strength (the cue-to-target strength, i.e. the proportion of participants who provide a given target in the presence of the cue word) and the backward strength (the target-to-cue strength, i.e. the forward strength with cue word and target word reversed). The resonance strength captures the extent to which a cue word has reciprocal associates.

Cue set size revealed a small but significant difference ($t(171.365) = -2.18$, $p = 0.0305$) between regulars (mean 14.50) and irregulars (mean 15.66). As expected, irregulars have slightly higher cue set sizes. While cue set size roughly follows a normal distribution, mean connectivity has a skewed distribution. In order to reduce this skewness, we logarithmically transformed mean connectivity, adding one in order to include the items with a zero score. A *t*-test revealed a difference between the regulars (mean 0.86) and the irregulars (mean 0.91) in the expected direction ($t(191.9) = -1.95$, $p = 0.0266$, one-tailed test), with irregulars revealing a higher mean connectivity than regulars. Finally, we analyzed the resonance strength of regulars and irregulars. Resonance strength is a highly skewed variable, so we used a Wilcoxon test, which revealed a highly significant difference between the irregulars and regulars ($W = 23309$, $p < 0.0001$). Figure 6 provides a graphical illustration of the greater resonance strength characterizing irregular verbs.

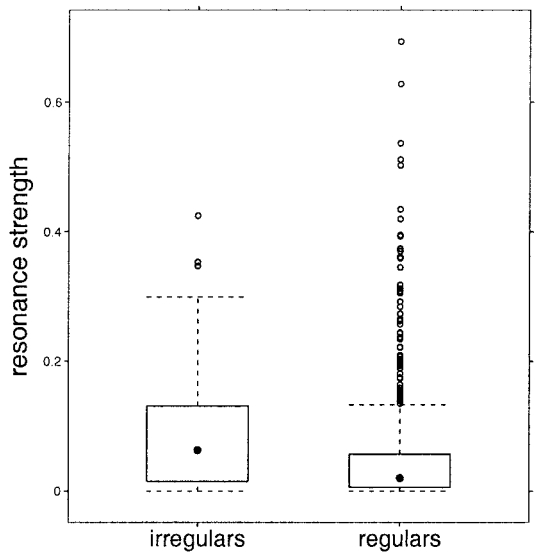

FIGURE 6. Boxplot for distributions of resonance strength, irregular and regular verbs, University of South Florida association norms (Nelson et al. 1998).

The reader might object that the irregulars have a greater resonance strength because they are more frequent, and that regularity as such is not a predictor of resonance strength in its own right. To address this issue, we carried out a multiple regression analysis. To do so, however, a transformation of the highly skewed dependent variable is required. As the skewness persisted after various simple transformations such as the log transformation, we decided to remove the words with zero resonance strength (leaving 639 words for further analysis, 118 irregulars and 521 regulars) and to transform resonance strength for these 639 words by multiplying it by 1,000 and by taking the logarithm of the result. We analyzed this rescaled resonance strength using multiple regression with log form frequency and regularity as predictors.

This analysis showed that form frequency was a (nonlinear) predictor of resonance strength ($F(2,635) = 15.74$, $p = < 0.0001$; the nonlinear component was modeled with a restricted cubic spline with 3 knots, $F(1,635) = 22.57$, $p < 0.0001$). In addition, regularity emerged as significant ($F(1,635) = 5.20$, $p = 0.0229$). This shows that the

robustness of regularity as a predictor for resonance strength is not due to a confound with frequency. On the contrary, regularity is supported as a genuine predictor of resonance strength in its own right.

Considered jointly, all three measures obtained from the University of South Florida word-association norms support the hypothesis that the higher semantic density of irregular verbs affects lexical organization in memory.

3.2. FAMILIARITY RATINGS. Regularity is also predictive in another off-line task, familiarity rating. In this task, participants are asked to provide their subjective estimate, often on a seven-point scale, of how frequently they think a word is used. Familiarity ratings are available for a large number of English simplex words in the databases compiled by Balota and his colleagues (Spieler & Balota 1998, Balota et al. 1999b; see also Spieler & Balota 1997, Balota & Spieler 1998, Balota et al. 1999a, and Balota et al. 2004), which also bring together reaction times for visual lexical-decision and word naming in English for two age groups (young versus old participants). From this database, we selected the 617 monomorphemic verbs that according to the CELEX lexical database have not arisen due to conversion from a noun, and that also appear in the argument-structure alternation classes of Levin 1993. Of these verbs, 511 are regular and 106 irregular.

Figure 7 shows six measures that jointly explain 76.5% of the variance in these ratings (75.5% after bootstrap validation). Not surprisingly, frequency is the most important predictor, as can be seen in the upper left panel ($F(2,599) = 587.08$, $p < 0.0001$). The frequency effect is slightly stronger for regular verbs. The upper central panel reveals a nonlinear relation between the noun-to-verb ratio and familiarity ($F(2,599) = 86.92$, $p < 0.0001$; the nonlinear component was modeled with a restricted cubic spline with 3 knots, $F(1,599) = 88.24$, $p < 0.0001$). The greater the frequency of the nominal counterpart of a verb, the more familiar it is judged to be, if it exists. (The initial horizontal part of the curve represents verbs with no conversion alternant.)

When verbs are used more often in spoken than in written language, they are rated higher as well ($F(1,599) = 133.06$, $p < 0.0001$). We note that verbs that are used more often in spoken language are also the verbs that are likely to have an earlier age of acquisition. (See Baayen 2005 for further details on this measure.)

The lower left panel shows that ratings decrease slightly with inflectional entropy ($F(1,599) = 8.48$, $p = 0.0037$). A greater information load of the inflectional paradigm apparently leads to a decrease in perceived familiarity. The lower central panel illustrates a slight increase with derivational entropy ($F(1,599) = 6.66$, $p = 0.0101$), an information-theoretic analogue of the morphological family size measure developed by Moscoso del Prado Martín and colleagues (2004). Derivational entropy is a measure that weights the type count of family members for their individual frequencies of use. The increase in familiarity with an increase in morphological connectivity is consistent with earlier studies (see Schreuder & Baayen 1997, Bertram et al. 2000, De Jong et al. 2000).

The lower right panel of Fig. 7 visualizes the effect of lexical connectivity based on shared argument alternation classes ($F(91,599) = 8.52$, $p = 0.0036$). What we have named the Levin connectivity is the (log) total number of verbs in all argument alternation classes in which a given verb participates. It is a measure of verbal connectivity at the syntactic level. The Levin connectivity is slightly larger for regular verbs (mean 3.46) than for irregular verbs (mean 3.24), a difference that is significant ($p = 0.0388$, Wilcoxon test), as expected given that regular verbs are more typical for productive

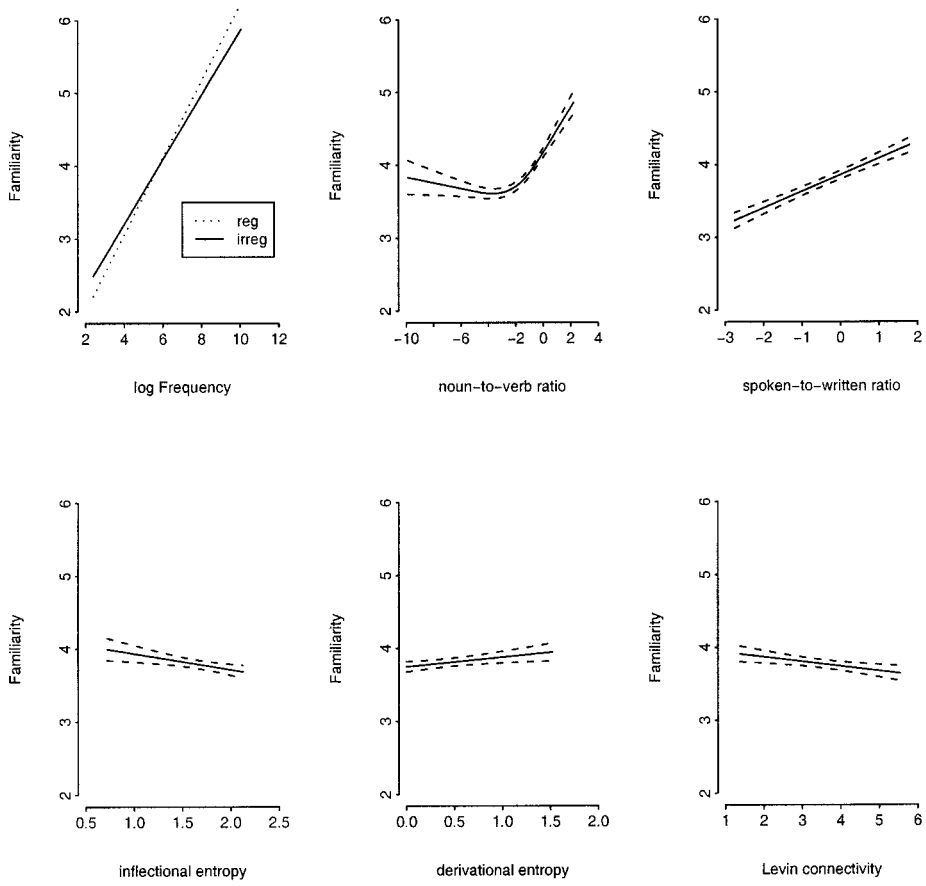

FIGURE 7. Partial effects of six predictors for subjective familiarity ratings for 106 irregular and 511 regular verb stems. Dashed lines represent 95% confidence intervals.

alternation classes. Here, we find that apparently, a greater Levin connectivity leads to a small decrease in perceived familiarity.

The position of the curves in Fig. 7 with respect to the vertical axis is calibrated for regular verbs. For irregular verbs, all curves (with the exception of the dotted line in the upper left panel) have to be shifted up by half a rating unit. In other words, independently of all the other predictors, irregular verbs elicit higher ratings than regular verbs. This suggests to us that our measures for semantic connectivity are in need of considerable refinement before they will be able to capture the full impact of semantic density on familiarity ratings.

We conclude with two observations. First, note that the advantage for irregular verbs is observed here for verbs presented in their unmarked, present-tense form, a form that by itself is fully regular. This shows that it is impossible to rate the familiarity of a verb form without considering its paradigmatic properties. Second, the significance of measures such as inflectional entropy and Levin connectivity shows that the dimensions on which we have observed differences between regular and irregular verbs are relevant to understanding lexical processing and representation.

3.3. WORD-NAMING LATENCIES. From the abovementioned databases compiled by Balota and colleagues, we extracted 832 monomorphemic verbs and their response

latencies in word naming. According to CELEX, these verbs are not derived from nouns. Of these verbs, 705 were regular and 127 irregular. All naming latencies concern the reading aloud of the verb without overt inflections, that is, the form shared by the infinitive and the present-tense form with the exception of the third person singular. For ease of reference, we refer to these uninflected forms of the verb as the verb stem.

Table 9 lists the predictors that emerged from a stepwise multiple regression analysis. Regularity was a significant predictor together with age (younger subjects responded faster), inflectional entropy (a greater inflectional entropy led to longer naming latencies), frequency (a greater frequency led to shorter latencies), and neighborhood density (a greater neighborhood density led to shorter naming latencies; this effect had a nonlinear component). The effect of regularity was quite small but significant ($p < 0.01$): regulars were named 4 ms faster than irregulars, when the other variables were held constant. We return below to this naming disadvantage for irregulars.

PREDICTOR	DEGREES OF FREEDOM	F	p
age	1,1655	24371.90	< 0.0001
inflectional entropy	1,1655	4.34	0.0375
neighborhood density	4,1655	44.12	< 0.0001
nonlinear	3,1655	5.57	0.0008
frequency	1,1655	172.06	< 0.0001
regularity	1,1655	7.14	0.0076

TABLE 9. Analysis of covariance with age, inflectional entropy, neighborhood density, frequency, and regularity as predictors of naming latencies.

It is important to keep in mind that the processing differences observed for regular and irregular verbs involve the uninflected present-tense forms of the verbs. These present-tense forms (or infinitives) are in no way irregular as such. In other words, we are observing a paradigmatic effect, the effect of a formal morphological property of a verb's other inflected forms (past, present and past participle) on the processing of the present-tense form.

3.4. VISUAL LEXICAL-DECISION LATENCIES. Tabak and colleagues (2005) studied the role of regularity in comprehension of Dutch verbs by means of a visual lexical-decision experiment. They used a regression design with 286 verbs, half of which were irregular. The sets of regular and irregular verbs were matched in the mean for their lemma frequency in CELEX. Participants were asked to make lexical decisions for verbs in singular and plural form in the present and past tense. One half of the participants was presented with singular forms, and the other half saw only plural forms. A given subject first performed the task for the present-tense forms, and repeated the task a week later with the past-tense forms. Tabak and colleagues observed significant facilitatory effects of frequency and morphological family size, as well as effects of number and tense in interaction with regularity. We leave the details of these effects aside and call attention to the predictivity of the following semantic variables.

For irregular verbs in the past tense, the log of the number of verbal synsets in the Dutch WordNet emerged as facilitatory. Apparently, the greater semantic density of irregular verbs is advantageous in lexical decision precisely for the irregular forms themselves. This suggests that retrieval of stems from memory involves deeper semantic processing in the case of the irregular past-tense stems. Tabak and colleagues (2005) also observed nonlinear effects of inflectional entropy. For irregular verbs, the effect was mainly facilitatory; for regular verbs, it was mainly inhibitory. In addition, lexical-

decision latencies varied significantly with the auxiliary selected by the verb for the present and past perfect. In the absence of a database with verb argument-structure alternations, a dictionary-based count of argument structures and complementation patterns was used to gauge syntactic connectivity. The logarithm of this count turned out to be predictive as well: verbs with more argument structures elicited shorter response latencies.

What this study shows is that a range of measures that tap into aspects of semantic density are significant predictors for an on-line measure of lexical processing. The general pattern is one in which greater lexical connectivity allows faster responses, and also one in which irregular verbs may benefit from this connectivity over and above their distributional advantage, as witnessed by the interactions involving inflectional entropy and the synsets measure.

4. META-ANALYSES OF TWO NEUROIMAGING STUDIES. We have seen that the greater semantic density of irregular verbs is reflected in free association norms, in subjective familiarity ratings, in word-naming latencies, and in visual lexical-decision latencies. In this section, we consider two neuroimaging studies that have reported different patterns of activation in the brain for regulars and irregulars. These studies attribute the observed differences in activation to the difference between regulars and irregulars at the form level. However, differences in activation might also arise due to differences in semantic density. We therefore inspected the data used in an English PET study and a German ER-fMRI study with respect to the possibility of a confound of regularity and semantic density.

4.1. ENGLISH IRREGULARS. Jaeger and colleagues (1996) asked participants to read aloud several lists of words. One list contained only regular verb stems, another only irregular verb stems. The regular and irregular verbs in the two lists were matched for frequency of occurrence and dispersion using Carroll's (1970) standard frequency index. They asked their participants to read these stems and to say aloud their past-tense forms. They measured naming latencies, accuracy, and brain activity by means of the PET technology. They observed that the irregular verbs elicited longer response latencies than the regular verbs. They also observed that various areas of the brain were differentially activated during past-tense naming.

We studied Jaeger and colleagues' word lists in order to ascertain whether their regular and irregular verbs might differ on any of the semantic variables that we have considered. We therefore selected all verbs in their study for which we have data on the probability of regulars in the verb argument-structure alternation classes discussed previously. This left us with 86 verbs (out of 92 in the original study). It turned out that their regular and irregular verbs were perfectly matched on all but three variables. The irregular verbs had a greater mean inflectional entropy than the regular verbs (1.78 versus 1.58, $t(82.944) = -3.20$, $p = 0.0019$). The irregular verbs also had a smaller proportion of regular verbs in their argument-structure alternation classes (0.73 versus 0.88, $t(59.407) = 4.06$, $p = 0.0001$). Finally, the regular verbs had corresponding nouns with much higher log lemma frequencies (6.2 versus 3.5, $t(82.793) = 4.87$, $p < 0.0001$). These three variables are significant predictors for regularity in a logistic regression model fitted to this data set ($R^2 = 0.554$, Somers's $Dxy = 0.764$). This shows that in the Jaeger and colleagues' study, regularity is seriously confounded with differences in conversion, syntax, and inflection.

To gain further insight into the potential relevance of these differences, we studied this data set in further detail by adding the naming latencies from the Balota data. This

further constrained our dataset and left us with 76 verbs. A stepwise linear regression model revealed significant main effects of age (younger participants had shorter naming latencies, $t(145) = -58.31$, $p < 0.0001$), inflectional entropy (a greater inflectional entropy led to longer response latencies, $t(145) = 2.52$, $p = 0.0129$), length (longer words elicited longer response latencies, $t(145) = 2.35$, $p = 0.0202$), and an interaction of regularity by the probability of regulars in the verb's alternation classes ($t(145) = -2.06$, $p = 0.0415$). Regular verbs were named faster the greater the proportion of regulars in their argument-structure alternation classes.

The overall processing disadvantage for irregular verbs compared to regular verbs is some 12 ms ($p < 0.001$ for both the young and old subject groups in the Balota data). The preceding analyses allow us to conclude that this difference can be attributed, at least in part, on the one hand to the higher inflectional entropy for irregulars, and on the other hand to the kind of argument structures that the verbs allow. This finding sheds new light on the far more substantial processing disadvantage for irregular past-tense forms reported in Jaeger et al. 1996. These authors point out that the magnitude of their effect (on the order of 100 ms) might be due to their blocked presentation of regulars and irregulars. The qualitative differences observed in their PET data, however, would argue against this being the only source or even the main source of the differences in the response latencies. Jaeger and colleagues therefore conclude that regulars and irregulars are processed by two qualitatively different mechanisms dedicated to the production of the past-tense form in word naming: rules and rote. Our present data show that a pattern very similar to what Jaeger and colleagues observed in their response latencies can already be observed for the stems of the regular and irregular verbs in their study. This implies that being regular or irregular has processing consequences even when no rules of inflection are involved. This is a surprising result in light of the dual-route mechanism proposed in Pinker 1991 and in such subsequent studies as Clahsen 1999. The dual-route approach, with its strict modular separation of form and meaning, would lead one to expect that retrieval of uninflected forms like *walk* and *give* from the lexicon might reveal effects of form, but not effects of the entropy of a verb's inflectional paradigm, nor effects of whether inflected forms in that paradigm are regular or irregular, nor effects originating from its argument-structure alternation neighborhood.

Note, incidentally, that the main effect of inflectional entropy across regulars and irregulars shows that in simple word naming the processing system must be sensitive to the token frequencies of regular inflectional forms (on which the calculation of the inflectional entropy measure is based). This argues against the dual-route model's assumption that regular inflected forms would leave no trace in memory, while supporting the frequency effects for regular inflections that have been observed in lexical decision (see Taft 1979 and Baayen et al. 1997; New et al. 2004).

In conclusion, our post-hoc analyses suggest the possibility that when subjects had to name the past-tense form when presented with a verb in the present tense by Jaeger and colleagues, the regular verbs might have had an advantage over irregular verbs at the level of argument structure, while irregular verbs might have been at a disadvantage due to the greater complexity (inflectional entropy) in their inflectional paradigms. These advantages and disadvantages may help explain, at least in part, the different patterns of activation in the brain Jaeger and colleagues reported for their PET data.

4.2. GERMAN IRREGULARS. Beretta and colleagues (2003) studied regular and irregular nouns and verbs in German using event-related fMRI. They reported, first, that

irregulars showed greater activation overall than did regulars, and, second, that regulars revealed greater lateralization to the left hemisphere than irregulars. The authors interpret their results as supporting Pinker's dual-mechanism account, but they express some astonishment at the greater activation observed for irregulars. They ascribe the more extensive bilateral activation for irregulars to more attentional resources supposedly needed for irregulars and to left-lateralization being supposedly specific for the linguistic on-line computational processes used by regulars.

The force of these conclusions hinges on the nature of the materials used in this study. We therefore investigated their materials in some more detail, which combine nouns and verbs. Their set of verbs contains twelve triplets matched for frequency. Each triplet consists of a regular verb, an irregular verb with vocalic alternation only in the past tense (henceforth ABA irregulars), and an irregular verb with vocalic alternation in both past tense and participle (henceforth ABB irregulars). The nouns in their study are twelve regulars and twelve irregulars, pairwise matched for frequency. Following Marcus et al. 1995, they contrasted nouns pluralizing in *-s* as regulars with nouns in *-er* as irregulars.

First consider the verbs in this study. The left panel of Figure 8 shows that inflectional entropy is not uniformly distributed over the three sets. The verbs of the ABB class of irregulars, for which vocalic alternation characterizes both past tense and participle, have a higher inflectional entropy than the regulars ($t(21.897) = 2.145, p = 0.04329$). This is reminiscent of what we found for the English data in Jaeger et al. 1996. Recall that the irregular verb stems in that study elicited longer response latencies than matched regular verb stems, independently of the presence of morphological structure. In this light, the greater activation observed by Beretta and colleagues (2003) for irregulars is in line with their greater inflectional information load, which may well require more computationally demanding selection processes.

The right panel of Fig. 8 shows that the regular and irregular verb classes are also not matched with respect to the auxiliary appropriate for the present and past perfect. Note that verbs that select *sein* as auxiliary, either exclusively or in alternation with *haben*, are overrepresented in the ABB class of irregulars ($p = 0.0082$, Fisher exact

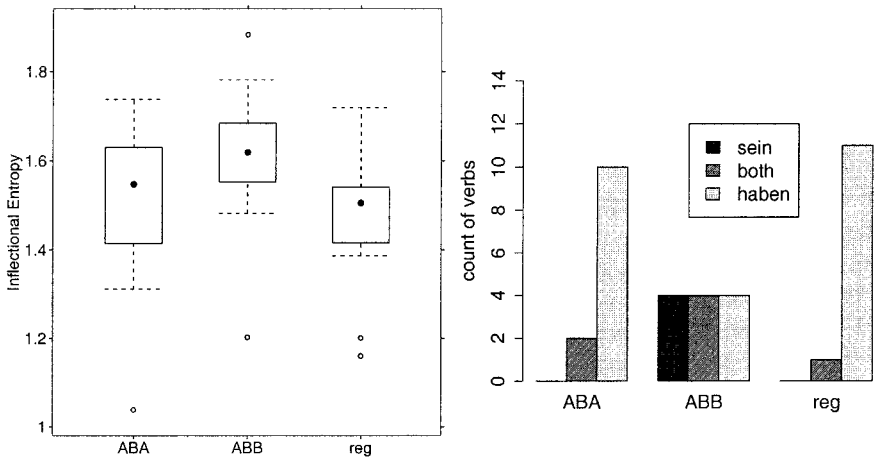

FIGURE 8. Boxplot of inflectional entropy by verb class (left panel), and barplot for present and past perfect auxiliary by verb class (right panel) for materials of Beretta et al. 2003.

test of independence). Since verbs allowing or selecting *sein* as auxiliary tend to be verbs of motion, it is to be expected that the irregulars should activate other regions in the brain compared to regulars. This may have contributed to the Beretta and colleagues' study finding that brain activity for irregulars is less restricted to the left hemisphere than it is for regulars.

Finally, consider the nouns in this ER-fMRI study. The boxplots in Figure 9 show remarkable differences between the two sets of nouns, both with respect to inflectional entropy (left panel, $t(20.709) = -4.94, p < 0.0001$) and with respect to morphological family size (right panel, $t(19.468) = -2.436, p = 0.0246$). The nouns that pluralize in *-er*, the irregulars in this study, have higher inflectional entropy and larger morphological families. The former is indicative of a greater complexity of the inflectional paradigm, and the latter shows that the *-er* nouns are more tightly integrated in the lexical network of semantic relations between morphologically related words. Given these properties of the noun materials, the greater activation observed by Beretta and colleagues for irregulars (without differentiation between nouns and verbs) comes as no surprise. Instead of postulating that irregulars would require more attentional resources (something one would expect for regulars rather than for irregulars within the framework of the dual-route mechanism), our data suggest that a more promising line of inquiry is to investigate the possibility that words with a greater semantic density elicit more activation in the brain across a wider range of areas.

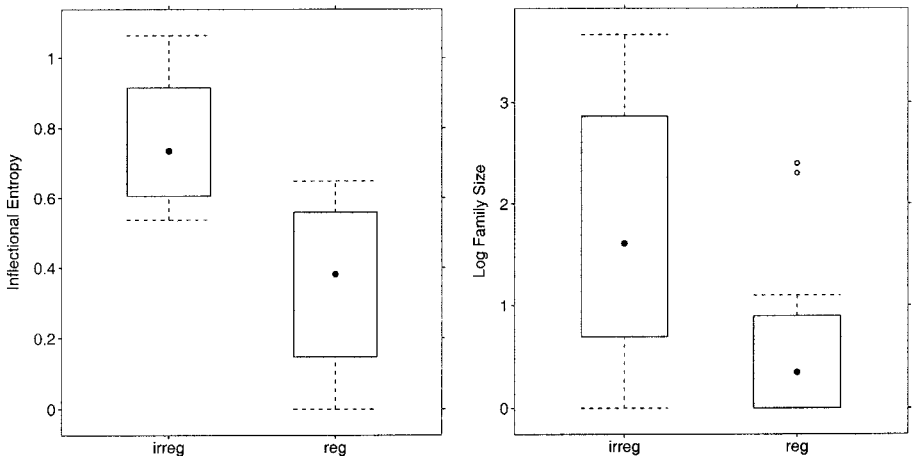

FIGURE 9. Boxplots of inflectional entropy by noun class (left panel), and of log family size by noun class (right panel) for materials of Beretta et al. 2003. Reg: regular noun; irreg: irregular noun.

5. CONCLUDING REMARKS. It is widely believed that there are no semantic dimensions along which regular and irregular verbs might differ systematically. While the formation of noun plurals in English (Quirk et al. 1985) and German (Hahn & Nakisa 2000) is characterized by subtle interactions of regularity and semantics, no such linguistic evidence is reported in the literature for verbs. Kim et al. 1991 argued for the English past tense that irregularity would be restricted to form, and that meaning would be irrelevant. Kim and colleagues interpreted the absence of semantic generalizations for the formation of the past tense as evidence supporting a modular organization of the lexicon in which form and meaning are strictly separated. Whereas various researchers in the connectionist research paradigm have questioned the necessity of such a strict

separation of form and meaning (e.g. Joanisse & Seidenberg 1999), and while some have suspected that there might be more subtle, distributional, semantic differences between regulars and irregulars (McClelland & Patterson 2002), no clear evidence for such differences has been reported.

The present study shows, by means of a series of lexical statistical surveys, that there are indeed various graded differences in semantics between regulars and irregulars. We have also shown that such differences affect both off-line and on-line behavioral measures, independently of frequency of occurrence. Finally, inspection of the data sets used in two neuroimaging studies revealed various potential semantic confounds between regulars and irregulars.

Our results shed new light on the ongoing debate about the advantages and disadvantages of single-route connectionist and dual-route symbolic models of morphological processing. If one accepts the premise that semantics would be irrelevant for the distinction between regulars and irregulars, then the prima facie evidence that has accumulated over the past ten years would seem to favor the symbolic dual-route model. Although there is no unequivocal evidence proving that connectionist accounts are wrong, neuroimaging studies such as Jaeger et al. 1996 and Beretta et al. 2003 reveal patterns of activity in the brain that are not straightforwardly predicted by connectionist single-route theories.

However, the subtle, graded semantic differences between regulars and irregulars that we have documented in the present study are not straightforwardly predicted by the symbolic dual-route theory. These distributional semantic differences and their behavioral consequences challenge a strict modular separation of form and meaning in the brain and question the interpretation of the data in Jaeger et al. 1996 and Beretta et al. 2003. By contrast, these differences are in line with the spirit of connectionist approaches to morphological processing and were in fact predicted to exist by McClelland and Patterson (2002). The different patterns of activation in the brain observed for regulars and irregulars, unexpected or at the very least not straightforwardly predicted from the point of view of single-route modeling, might, at least in part, arise due to subtle but significant graded differences in the way regulars and irregulars are networked semantically in the mental lexicon.

In the dual-system connectionist model of visual lexical-decision developed by Moscoso del Prado Martín (2003), an effect of regularity is observed over and above frequency and inflectional entropy (214–15). As mentioned in the introduction, this model's architecture shares with dual-route models that rules and representations are distinguished, and shares with connectionist models that rules and representations are conceived of as subsymbolic. Although this model is meant to provide a mathematical tool for the statistical study of information complexity in visual lexical-decision, its architecture is compatible with the body of evidence suggesting that declarative and procedural knowledge have their primary locus in different brain regions, as argued in Ullman et al. 1997.

The goal of this study, however, is not to decide between symbolic versus subsymbolic, and single- versus dual-route approaches to linguistic cognition. Far more important is that our results unequivocally support probabilistic approaches to language and its structure. Generative grammar has a long history of antagonism to the possibility that probabilistic pattern matching might be an essential cognitive component of language. Nevertheless, probability theory has recently found its way into optimality theory (Boersma & Hayes 2001), and Albright and Hayes (2003) have documented attractor effects within phonological similarity sets for regular verbs in English, contradicting the claims

of the dual-route model, although they couch their findings within a highly complex system of rules. As evidenced by the studies in Bod et al. 2003, probability is a central concept for all levels of language structure. What we hope to have clarified is that for a proper understanding of the details of the past tense in English, German, and Dutch, the graded, probabilistic notion of semantic density is essential.

For instance, the greater semantic density of irregulars may help explain the semantic attractor effects reported by Bybee and Slobin (1982) and by Ramscar (2002). If form and meaning were truly encapsulated along the lines of the dual-mechanism model, such attractor effects would be difficult to account for and would invite dismissal as experimental artifacts, or as examples of incidental, nonsystematic analogy. Given the existence of significant, graded semantic differences between regulars and irregulars, it seems more profitable to seek to understand the semantic attractor effects as grounded in semantic density. More generally, the present findings provide further support for the network view of morphological structure advocated by Bybee (1985, 1988, 2001). Our findings also show that the presently available data resources, combined with statistically informed data mining, make it possible to develop detailed quantitative models for tracing the consequences of lexical connectivity for morphological processing and representation.

The reader may feel that one important question is left unanswered. Why is it that irregular verbs have the greater semantic density? One possibility is that their greater semantic density is the synchronic reflex of greater age. The irregular verbs tend to be the older verbs of the language. Through time, they might have acquired a greater range of meanings than the younger words, which would then typically have more specialized and less rich semantics. Tabak and colleagues (2005) investigated this possibility for Dutch through the etymological cognates of 286 verbs. They observed that the age of a verb could be predicted only from regularity, as expected, and, surprisingly, from neighborhood density. Verbs with a greater neighborhood density, that is, verbs with greater phonological regularity, were more likely to have greater age. Frequency and a wide range of measures for semantic density were not predictive. This suggests that there is no simple relation between semantic density and age. A possibility that seems more likely, therefore, is that a greater semantic density makes it easier to remember irregular inflected forms, perhaps along the lines envisioned by Joanisse & Seidenberg 1999 and Patterson et al. 2001 (but see Tyler et al. 2004).

The diagnostics we used to gauge the importance of semantic structure for the distinction between regulars and irregulars are crude and in need of considerable sophistication and refinement. Furthermore, the scope of this article is restricted, limited as it is to just three Germanic languages, to simplex verbs only, and to experimental tasks with words presented in isolation. But the combined evidence of our data, which reveal a consistent pattern and which, when pitted against behavioral data, reveal a surprising degree of statistical robustness, suggests that this is a promising new area of research. Since more and more empirical resources for coming to grips with the meanings of words and their argument structures are becoming available, we are confident that future studies will lead to a much more profound understanding of the interaction between semantics and regularity. This is an uncharted area, and we have only scratched its surface.

REFERENCES

ALBRIGHT, ADAM, and BRUCE HAYES. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90.119–61.

ANSHEN, FRANK, and MARK ARONOFF. 1988. Producing morphologically complex words. *Linguistics* 26.641–55.

BAAYEN, R. HARALD. 2003. Probabilistic approaches to morphology. *Probability theory in linguistics*, ed. by Rens Bod, Jennifer Hay, and Stephanie Jannedy, 229–87. Cambridge, MA: MIT Press.

BAAYEN, R. HARALD. 2005. Data mining at the intersection of psychology and linguistics. *Twenty-first century psycholinguistics: Four cornerstones*, ed. by Anne Cutler. Hillsdale, NJ: Lawrence Erlbaum, to appear.

BAAYEN, R. HARALD, and RICHARD SPROAT. 1996. Estimating lexical priors for low-frequency morphologically ambiguous forms. *Computational Linguistics* 22.155–66.

BAAYEN, R. HARALD; TON DIJKSTRA; and ROBERT SCHREUDER. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 36.94–117.

BAAYEN, R. HARALD; RICHARD PIEPENBROCK; and LEON GULIKERS. 1995. *The CELEX lexical database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium.

BALOTA, DAVID A., and DANIEL H. SPIELER. 1998. The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut (1998). *Psychological Science* 9.238–41.

BALOTA, DAVID A.; MICHAEL J. CORTESE; and MAURA PILOTTI. 1999a. Item-level analyses of lexical decision performance: Results from a mega-study. *Abstracts of the 40th annual meeting of the Psychonomics Society*, 44. Los Angeles, CA: Psychonomic Society.

BALOTA, DAVID A.; MICHAEL CORTESE; and MAURA PILOTTI. 1999b. Visual lexical decision latencies for 2906 words. Online: http://www.artsci.wustl.edu/~dbalota/lexical_decision.html

BALOTA, DAVID A.; MICHAEL J. CORTESE; SUSAN D. SERGENT-MARSHALL; DANIEL H. SPIELER; and MELVIN J. YAP. 2004. Visual word recognition for single-syllable words. *Journal of Experimental Psychology: General* 133.283–316.

BAUER, LAURIE. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

BERETTA, ALAN; CARRIE CAMPBELL; THOMAS H. CARR; JIE HUANG; LOTHAR M. SCHMITT; KIEL CHRISTIANSON; and YUE CAO. 2003. An ER-fMRI investigation of morphological inflection in German reveals that the brain makes a distinction between regular and irregular forms. *Brain and Language* 85.67–92.

BERTRAM, RAYMOND; R. HARALD BAAYEN; and ROBERT SCHREUDER. 2000. Effects of family size for complex words. *Journal of Memory and Language* 42.390–405.

BOD, RENS; JENNIFER B. HAY; and STEFANIE JANNEDY (eds.) 2003. *Probability theory in linguistics*. Cambridge, MA: MIT Press.

BOERSMA, PAUL, and BRUCE HAYES. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32.45–86.

BYBEE, JOAN L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

BYBEE, JOAN L. 1988. Morphology as lexical organization. *Theoretical morphology: Approaches in modern linguistics*, ed. by Michael Hammond and Michael Noonan, 119–41. London: Academic Press.

BYBEE, JOAN L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

BYBEE, JOAN L., and DAN I. SLOBIN. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58.265–89.

CARROLL, JOHN B. 1970. An alternative to Juilland's usage coefficient for lexical frequencies, and a proposal for a standard frequency index (SFI). *Computer Studies in the Humanities and Verbal Behavior* 3.61–65.

CLAHSEN, HARALD. 1999. Lexical entries and rules of language: A multi-disciplinary study of German inflection. *Behavioral and Brain Sciences* 22.991–1060.

CLEVELAND, WILLIAM S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74.829–36.

DE JONG, NIVJA H.; ROBERT SCHREUDER; and R. HARALD BAAYEN. 2000. The morphological family size effect and morphology. *Language and Cognitive Processes* 15.329–65.

ERNESTUS, MIRJAM, and R. HARALD BAAYEN. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79.5–38.

FELDMAN, LAURIE, and B. PROSTKO. 2001. Graded aspects of morphological processing: Task and processing time. *Brain and Language* 81.1–16.
FELLBAUM, CHRISTIANE E. 1998. *WordNet: An electronic database*. Cambridge, MA: MIT Press.
HAHN, ULRIKE, and RAMIN C. NAKISA. 2000. German inflection: Single route or dual route? *Cognitive Psychology* 41.313–60.
HARRELL, FRANK E., JR. 2001. Regression modeling strategies. New York: Springer.
HARRIS, CATHERINE L. 1993. Using old words in new ways: The effect of argument structure, form class and affixation. *Chicago Linguistic Society* 29.139–53.
JAEGER, JERI J.; ALAN H. LOCKWOOD; DAVID L. KEMMERER; ROBERT D. VAN VALIN, JR.; BRIAN W. MURPHY; and HANIF G. KHALAK. 1996. A positron emission tomographic study of regular and irregular verb morphology in English. *Language* 72.451–97.
JOANISSE, MARC F., and MARK S. SEIDENBERG. 1999. Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences, USA* 96.7592–97.
KIM, JOHN J.; STEVEN PINKER; ALAN PRINCE; and SANDEEP PRASADA. 1991. Why no mere mortal has ever flown out to center field. *Cognitive Science* 15.173–218.
KÖHLER, REINHARD. 1986. *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik.* Bochum: Brockmeyer.
LEVIN, BETH. 1993. *English verb classes and alternations: A preliminary investigation.* Chicago: University of Chicago Press.
LIEBER, ROCHELLE, and R. HARALD BAAYEN. 1997. A semantic principle for auxiliary selection in Dutch. *Natural Language and Linguistic Theory* 15.789–845.
MACWHINNEY, BRIAN, and JARED LEINBACH. 1991. Implementations are not conceptualizations: Revising the verb learning model. *Cognition* 40.121–57.
MARCUS, GARY F.; URSULA BRINKMAN; HARALD CLAHSEN; RICHARD WIESE; and STEVEN PINKER. 1995. German inflection: The exception that proves the rule. *Cognitive Psychology* 29.189–256.
MCCLELLAND, JAMES, and KARALYN PATTERSON. 2002. Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences* 6.465–72.
MILLER, GEORGE A. 1990. Wordnet: An on-line lexical database. *International Journal of Lexicography* 3.235–312.
MOSCOSO DEL PRADO MARTÍN, FERMÍN. 2003. *Paradigmatic effects in morphological processing: Computational and cross-linguistic experimental studies*. (MPI series in psycholinguistics.) Nijmegen: Max Planck Institute for Psycholinguistics.
MOSCOSO DEL PRADO MARTÍN, FERMÍN; ALEKSANDER KOSTIĆ; and R. HARALD BAAYEN. 2004. Putting the bits together: An information theoretical perspective on morphological processing. *Cognition* 94.1–18.
NELSON, DOUGLAS L.; CATHY L. MCEVOY; and THOMAS A. SCHREIBER. 1998. The University of South Florida word association, rhyme, and word fragment norms. Online: http://w3.usf.edu/FreeAssociation/.
NEW, BORIS; MARC BRYSBAERT; JUAN SEGUI; LUDOVIC FERRAND; and KATHY RASTLE. 2004. The processing of singular and plural nouns in French and English. *Journal of Memory and Language* 51.568–85.
PATTERSON, KARALYN; MATTHEW LAMBON RALPH; J. HODGES; and JAMES MCCLELLAND. 2001. Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia* 39.709–24.
PINKER, STEVEN. 1991. Rules of language. *Science* 153.530–35.
PINKER, STEVEN. 1997. Words and rules in the human brain. *Nature* 387.547–48.
PINKER, STEVEN. 1999. *Words and rules: The ingredients of language*. London: Weidenfeld and Nicolson.
PINKER, STEVEN, and MICHAEL ULLMAN. 2002. Combination and structure, not gradedness, is the issue. *Trends in Cognitive Sciences* 6.472–74.
PLUNKETT, KIM, and PATRICK JUOLA. 1999. A connectionist model of English past tense and plural morphology. *Cognitive Science* 23.463–90.
QUIRK, RANDOLPH; SIDNEY GREENBAUM; GEOFFREY LEECH; and JAN SVARTVIK. 1985. *A comprehensive grammar of the English language*. London: Longman.
RAMSCAR, MICHAEL. 2002. The role of meaning in inflection: Why the past tense doesn't require a rule. *Cognitive Psychology* 45.45–94.

RANDALL, JANET; ANGELIEK VAN HOUT; JURGEN WEISSENBORN; and R. HARALD BAAYEN. 2003. Acquiring unaccusativity: A cross-linguistic look. *The unaccusativity puzzle*, ed. by Artemis Alexiadou, Elena Anagnostopoulou, and Martin Everaert, 332–53. Oxford: Oxford University Press.

RUMELHART, DAVID E., and JAMES L. MCCLELLAND. 1986. On learning the past tenses of English verbs. *Parallel distributed processing: Explorations in the microstructure of cognition, vol, 2: Foundations*, ed. by James L. McClelland and David E. Rumelhart, 216–71. Cambridge, MA: MIT Press.

SCHREUDER, ROBERT, and R. HARALD BAAYEN. 1997. How complex simplex words can be. *Journal of Memory and Language* 37.118–39.

SHANNON, CLAUDE E., and WARREN WEAVER. 1949. *The mathematical theory of communication*. Urbana: University of Illinois Press.

SHIRAI, YASUHIRO, and ROGER W. ANDERSON. 1995. The acquisition of tense-aspect morphology: A prototype account. *Language* 71.743–62.

SPIELER, DANIEL H., and DAVID A. BALOTA. 1997. Bringing computational models of word naming down to the item level. *Psychological Science* 6.411–16.

SPIELER, DANIEL H., and DAVID A. BALOTA. 1998. Naming latencies for 2820 words. Online: http://www.artsci.wustl.edu/~dbalota/naming.html.

TABAK, WIEKE; ROBERT SCHREUDER; and R. HARALD BAAYEN. 2005. Lexical statistics and lexical processing: Semantic density, information complexity, sex, and irregularity in Dutch. *Linguistic evidence—Empirical, theoretical, and computational perspectives*, ed. by Stephan Kepser and Marga Reis. Berlin: Mouton de Gruyter, to appear.

TAFT, M. 1979. Recognition of affixed words and the word frequency effect. *Memory & Cognition* 7.263–72.

TIERSMA, PETER MEIJES. 1982. Local and general markedness. *Language* 58.832–49.

TYLER, LORRAINE K.; EMMANUEL A. STAMATAKIS; ROY W. JONES; PETER BRIGHT; KADIA ACRES; and WILLIAM MARSLEN-WILSON. 2004. Deficits for semantics and the irregular past tense: A causal relationship? *Journal of Cognitive Neuroscience* 16.1159–72.

ULLMAN, MICHAEL; RUTH BERGIDA; KATHLEEN M. O'CRAVEN. 1997. Distinct fMRI activation patterns for regular and irregular past tense. *NeuroImage* 5.S549.

VERKUYL, HENK. 1989. Aspectual classes and aspectual composition. *Linguistics and Philosophy* 12.39–94.

VOSSEN, PIEK; LAURA BLOKSMA; and PAUL BOERSMA. 1999. *The Dutch WordNet* (CD-ROM). Luxembourg: European Language Resources Association (ELRA).

Baayen
Max Planck Institute for Psycholinguistics
PO Box 310
6500 AH, Nijmegen, The Netherlands
[baayen@mpi.nl]

Moscoso del Prado Martín
Medical Research Council
Cognition and Brain Sciences Unit
15 Chaucer Road
Cambridge CB2 2EF, UK
[fermin.moscoso-del-prado-martin@mrc.cam.ac.uk]

[Received 7 April 2004;
accepted 8 February 2005]