Semantic density and past-tense formation in three Germanic languages

R. Harald Baayen and Fermín Moscoso del Prado Martín

Interfaculty Research Unit for Language and Speech, University of Nijmegen &
Max Planck Institute for Psycholinguistics, P.O.Box 310, 6500 AH, Nijmegen,
The Netherlands.

Abstract

It is widely believed that the difference between regular and irregular verbs is restricted to form. This study questions this belief. We report a series of lexical statistics showing that irregular verbs have a greater density in semantic space. Irregular verbs tend to have greater semantic neighborhoods containing relatively many other irregulars compared to regulars. 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). Rumel-

Submitted, comments welcomed   *26th March 2004*

hart and McClelland (1986), and since then many others including, e.g, Mac-Whinney and Leinbach (1991), Joanisse and Seidenberg (1999), and Plunkett and Juola (2000), argued that regular and irregular verbs are processed by a single distributed 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 pointed out by Kim, Pinker, Prince, and Prasada (1991), 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).

(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), antilope(s)*
   trout(s), fish(es), carp(s), pike(s), herring(s), flounder(s)

The nouns listed under (a) are words for members of the inner family, those listed under (b) denote things that typically occur in pairs or large numbers and the nouns listed under (c) refer to animals that occur in pairs or groups (cf. Tiersma, 1982, and Baayen, Dijkstra, and Schreuder, 1997). The nouns in (d) all share the property that they have a plural that is identical to the singular. All these nouns are names for animals. The nouns listed under (e) have two plural forms, either a plural in *-s* or a plural that is identical to the singular. Again, all these nouns are names for animals. Which of the 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., Quirk, Greenbaum, Leech, & Svartvik, 1985). The probability that we find only animal names among the (d) nouns for which the plural is grammaticalized to be always completely identical to the singular by chance is vanishingly small. 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 (*whale*). 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 irregularity and semantic similarity 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 2000:134 for semantic constraints on word formation).

In the connectionist literature, Hahn and Nakisa (2000) is an example of a study which addresses a similar phenomenon in the inflection of German

proper nouns, where the distinction between surnames and Christian names may co-determine the choice of the plural suffix.

However, when it comes to the past tense in English, German, or Dutch, there is no obvious semantic difference to guide the choice between regular and irregular. In the connectionist model of Joanisse and Seidenberg (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, Lambon Ralph, Hodges, and McClelland (2001), however, reported that patients with varying degrees of severity of semantic dementia had selective problems with irregular paste tense formation. They trace this specific problem for irregular past tense forms to the reduced phonological support that irregulars receive compared to regulars, 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, perfomance 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 judgement 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.

However, 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 by Bybee and Slobin (1982) and by Ramscar (2002). Bybee and Slobin (see also Bybee, 1985) 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 semantically 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*.

Summing up, irregularity may interact with semantics in inflection, but thus far distributional evidence 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 paper 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 would be 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 will 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 will show that these data sets suffer from subtle confounds of regularity and semantic density. In the general discussion, we will 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 clearcut, near-deterministic differences. Instead, what we expect to find are graded differences for semantic random variables, with significant differences in central

tendency between the distributions in combination with substantial overlap of these distributions.

There are two methodological decisions underlying all analyses to be reported here that we need to make explicit from the very beginning. The first methodological decision is to investigate only monomorphemic verbs, and to leave complex verbs for further research. Monomorphemic verbs are the basic lexical units of the language, and 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.

A 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, Piepenbrock, & Gulikers, 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 rigour is necessary to avoid circularity, with reanalyses potentially reshaping the data to fit the hypothesis derived from the data.

In section 2.1, we begin with an investigation of synonym sets and regularity. In section 2.2, we study the distribution of auxiliaries for regular and irregular verbs in German and Dutch. A related semantic variable, the distribution of regular and irregular verbs in argument structure alternation classes, is examined for English in section 2.3. Section 2.4 addresses the asymmetries in the frequencies of use of the present, past, and past participle forms in German and Dutch. In the final subsection, we use a corpus-based computational measure of distance in context space to show that regulars and irregulars differ in semantic density. This computational measure implicitly captures many of the differences between regulars and irregulars traced by the preceding more discrete semantic variables.

Throughout our discussion, we will 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, we will show that some of our variables capture aspects of regularity even after frequency is partialled out. Other variables have no such additional predictivity. Nevertheless, even such variables turn out to be relevant to understanding lexical processing for words words matched on frequency.

6

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 will 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.

2.1.1 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, potentially blurring differences between regular and irregular verbs, thereby working against our hypothesis. In this way, we obtained a database of 1600 verbs, 146 irregular and 1454 regular. For each verb, we added its lemma frequency in CELEX, cumulating the frequencies of homographs such as *bank*.

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

> *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

> *grow, develop, produce, get, acquire*

7

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.
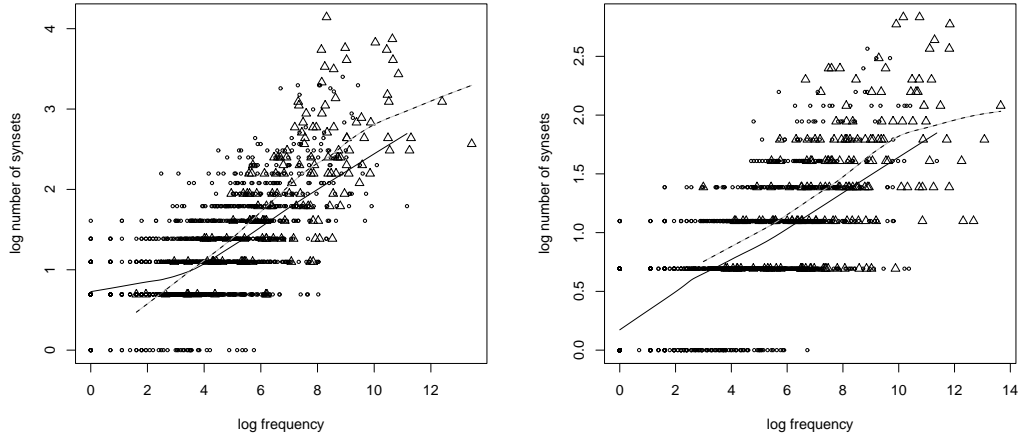

Figure 1. Number of synsets as a function of lemma frequency for English (left panel) and Dutch (right panel) regular (dots) and irregular (triangles) simplex verbs. The lines represent non-parametric regression lines for the regular (solid line) and irregular (dashed line) verbs.

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 non-parametric regression lines (Cleveland, 1979) for the regulars and irregulars respectively. The linear relation between the logarithm of the number of meanings and log frequency was first documented by Koehler (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 > 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 linear model to the data, predicting log number of synsets as a function of log frequency and regularity. The model revealed 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 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.

In order 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. 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 1545 words contained 163 irregular and 1382 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, Bloksma, and Boersma, 1999), which is part of the EuroWordNet distributed by the European Language Resources Association.

The right panel of Figure 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) seems to run parallel to the regression line for the regulars (solid line). A linear 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, two differences can be observed. 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 already 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. 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 form property (having an irregular past tense) with a semantic property (the tendency of having more meanings). This is a first aspect of what we will refer to as the greater semantic density of irregular verbs.

2.1.2 Synset pairs and triplets

Given the observation 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
Observed and expected counts of synsets containing pairs or triplets of simplex regular and irregular verbs for English.

Regular	Irregular	Observed	Expected (Type)	Expected (Token)	p
2	0	455	611.93	123.09	p_r^2
1	1	207	122.89	357.83	$2p_r p_i$
0	2	79	6.17	260.07	p_i^2
3	0	102	144.09	13.00	p_r^3
2	1	58	43.40	56.68	$3p_r^2 p_i$
1	2	27	4.35	82.39	$3p_r p_r^2$
0	3	5	0.15	39.92	p_i^3

Expected (Type): expected counts given type-based probability estimates.
Expected (Token): expected counts given token-based probability estimates.

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 (e) by means of the binomial model ($e = np$), using the probabilities p listed in the last column of Table 1 and the numbers of synset pairs ($n = 741$) and synset triplets ($n = 192$).

There are two options for estimating the probabilities of regulars (p_r) and irregulars (p_i) figuring in the probabilities p. We can estimate these probabilities on the basis of the type counts of regulars and irregulars. This gives us the estimates $p_r = 0.909$ 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. The expected counts (np) for the two counts 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 will 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 ($X^2_{(2)} = 957.5, p < 0.0001$). There are fewer synset pairs with only regular verbs than expected, and there are more synsets involving irregular verbs. While only some seven 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 ($X^2_{(3)} = 296.4, p < 0.0001$). While no synset triplets with three irregulars are expected, 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
Observed and expected counts of synsets containing pairs or triplets of simplex regular and irregular verbs for Dutch.

Regular	Irregular	Observed	Expected (Type)	Expected (Token)	p
2	0	252	286.45	25.36	p_r^2
1	1	84	67.57	139.85	$2p_r p_i$
0	2	22	3.98	192.79	p_i^2
3	0	65	74.43	1.96	p_r^3
2	1	26	26.34	16.22	$3p_r^2 p_i$
1	2	11	3.11	44.72	$3p_r p_r^2$
0	3	2	0.12	41.10	p_i^3

Expected (Type): expected counts given type-based probability estimates.
Expected (Token): expected counts given token-based probability estimates.

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 ($X^2_{(2)} = 89.6, p < 0.0001$ for the synset pairs, $X^2_{(3)} = 50.1, p < 0.0001$ 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.

Summing up, our analyses of the composition of synset pairs and triplets of simplex verbs shows 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, the synsets in which they appear tend to contain other 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 only considered 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.

2.1.3   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 and 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 simple truism, we can derive the non-trivial 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 32 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 12 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 circles). 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 non-parametric 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.

Table 3

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

class size	number of classes	average mean log number of synsets
1	12	2.332456
2	8	2.422791
3	2	1.778933
4	1	2.514352
5	1	2.310333
6	1	2.034079
7	2	1.746377
8	1	1.356990
10	2	1.591421
17	1	1.815977
38	1	1.710856

Interestingly, the class with the greatest size and the lowest mean log number of synsets is the set of verbs with no vocalic alternation, i.e., 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.
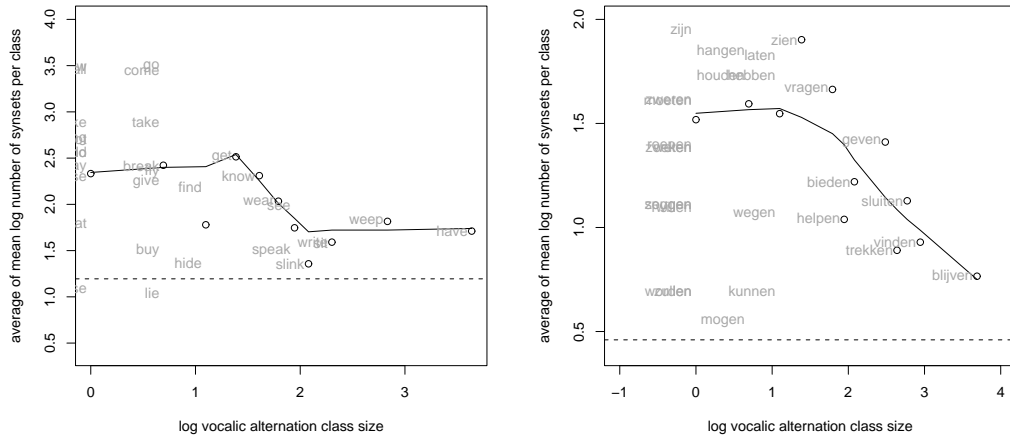

Figure 2. Correlation between log vocalic alternation class size (horizontal axis) and the average of the mean log number of synsets in the classes of this class size (vertical axis) for English (left panel) and Dutch (right panel) irregular verbs. Averages are represented by circles, the verbs represent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

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, and the right panel of Figure 2 provides the corresponding scatterplot, a non-parametric 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 $F(1, 10) = 10.25, p = 0.0095$). 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.

Summing up, 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.

14

Table 4

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

class size	number of classes	average mean log number of synsets
1	14	1.5180629
2	4	1.5941297
3	5	1.5468993
4	1	1.9022177
6	1	1.6634081
7	1	1.0389141
8	1	1.2196631
12	1	1.4105786
14	1	0.8903825
16	1	1.1282097
19	1	0.9293848
40	1	0.7657510

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 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.

However, in German and Dutch, a verb's auxiliary is a formal marker of as-

pectual properties (see, e.g., Randall, Van Hout, Weissenborn, and Baayen, 2003, and Lieber and Baayen, 1997). Especially telicity, the event having an (inferable) endpoint, emerges as an important determinant of the use of the marked auxiliary, 'to be' (*sein* in German, and *zijn* in Dutch). Hence, we can use verbal auxiliary selection in German and Dutch as an objective measure for ascertaining whether regular and irregular verbs differ in their aspectual semantics. The prediction is that irregular verbs, being predominant among the verb types in early acquisition and hence closer to the prototype of Shirai and Anderson (1995), should be 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).

We first consider Dutch. The CELEX lexical database specifies the auxiliary (*zijn*, *hebben*, or both *hebben* and *zijn*) for 1530 of the 1545 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 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 non-uniform distribution is not expected under chance conditions ($\chi^2_{(2)} = 51.43, p < 0.0001$).

Figure 3 summarizes the differences in the token frequencies of the regular and irregular verbs, broken down by auxiliary. Not surprisingly, the irregulars are more frequent than the regulars. Within the set of irregulars, there are no significant differences in mean frequency between the three types of verbs ($F(2, 159) = 1.28, p = 0.2796$), although the variance seems to be somewhat larger for the verbs selecting only the auxiliary *zijn*. Within the set of regulars, there are significant differences in frequency ($F(2, 1365) = 9.92, p < 0.0001$): The verbs taking both auxiliaries are used more frequently (p-values < 0.01) than the other two groups, which do not differ among themselves ($p > 0.2$). This pattern in the data suggests that the differences between regulars and irregulars in the type-based analysis summarized in Table 5 are largely independent of token frequency.

The second half of Table 5 replicates the same pattern for German. From the CELEX lexical database, we extracted all 1296 verbs listed as monomorphemic.

Table 5

 Counts (and percentages) of regular and irregular verbs with the auxiliary *hebben*, the auxiliary *zijn*, and with both auxiliaries, for Dutch and German.

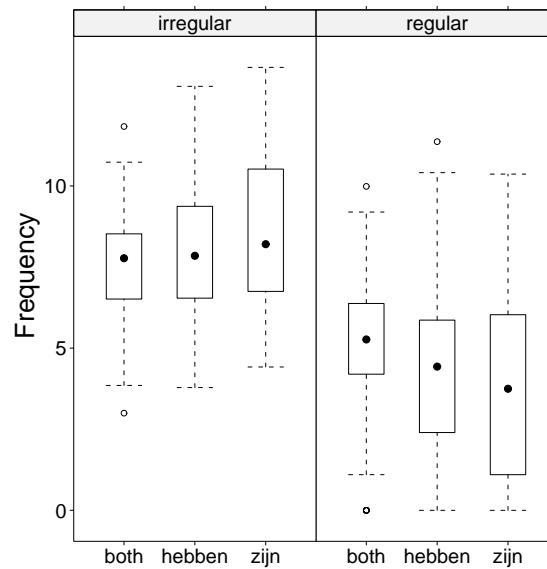|	Dutch			
Auxiliary	Irregular		Regular	
hebben	109	(67.28)	1199	(87.65)
zijn	15	(9.26)	34	(2.49)
both	38	(23.46)	135	(9.87)
	German			
Auxiliary	Irregular		Regular	
haben	124	(71.67)	971	(88.43)
sein	20	(11.56)	36	(3.27)
both	29	(16.76)	91	(8.29)

Figure 3. Boxplots for lemma frequency broken down by regularity and auxiliary for Dutch.

For 1271 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 non-uniform distribution of auxiliary use for irregulars and regulars is statistically significant ($\chi^2_{(2)} = 39.51, p < 0.0001$). An analyses of the token frequencies of the verb classes did not reveal any differences within the sets of irregulars and regulars, which is probably due to the small size of the corpus underlying the German frequency counts.

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 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. However, for English, Levin (1993) provides a detailed partitioning of verbs into argument structure alternation classes. 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 57 main alternation classes. In what follows, we consider the distribution of the 1053 simplex verbs (929 regulars and 124 irregulars) in our database that appear in Levin's alternation classes, and we inspect how the irregulars and regulars are distributed over these classes.

Many of these main classes contain just a few words, which makes it impossible to straightforwardly apply a chi-squared test, due to a great many cells in the contingency table with zero or very low frequencies. We therefore ran analyses over two subsets of the data. We first selected those alternation classes with at least 10 verbs and with either more than 5 regular verbs or more than 5 irregular verbs. A chi-squared test applied to the resulting 40 alternation classes revealed significant non-uniformity ($\chi^2_{(39)} = 128.97, p < 0.0001$), with the proportion of regulars ranging from 0.56 to 1.00. In a second analysis, we selected those alternation classes for which there are more than five regulars

and more than five irregulars, which resulted in a total of 23 classes, with proportions of regulars ranging from 0.56 for class 48 to 0.899 for class 43. Table 6 lists the simplex verbs in these two classes. A chi-squared test again shows that the regulars and irregulars are not uniformly distributed over the alternation classes ($\chi^2_{(22)} = 102.15, p < 0.0001$). These two analyses show that there is a robust difference between the regulars and the irregulars across a wide range of argument structure alternation classes.

Table 6
Levin's alternation classes 43 and 48. Irregular verbs are shown in upper case letters..

Simplex verbs in Levin alternation class 43 ('sparkle' verbs, 'bang' verbs, 'reek' verbs, '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 alternation class 48 ('appear' verbs, 'assert' verbs, 'die' verbs, 'transpire' verbs):

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

Leaving the precise interpretation of these differences to further research, we proceed to study the non-uniform distribution of regular and irregular verbs over the alternation classes from the perspective of the individual verbs. The simplest measure that comes to mind is the proportion of regulars in a verb's alternation class. However, as a given verb may occur in several alternation classes, we calculated, for each verb, the average of the proportion of regular verbs in its alternation classes. Figure 4 summarizes by means of boxplots the distributions, for regular and irregular verbs, of these means. It is clear that the irregular verbs tend to participate in alternation classes in which there are slightly more irregular verbs than is the case for the alternation classes in which the regulars tend to participate. The difference in central tendency (the mean proportion of regulars equal to 0.827 for the regular verbs and 0.808 for the irregular verbs) is significant both parametrically ($t(203.5) = 4.60, p < 0.0001$) and non-parametrically ($W = 74749.5, p < 0.0001$). We conclude that irregular verbs tend to occur in verb argument alternation space with more

19

other irregulars than is the case for regular verbs.

A technical note is in order here. In the counts on which these tests and Figure 4 are based, we have excluded a given individual word from the calculations of the proportions of regulars in its alternation classes. Including an irregular verb would give rise to greater proportions of irregulars in the alternation classes of that verb, and including a regular verb would similarly give rise to greater proportions of regulars in its alternation classes. In this way, the difference between regulars and irregulars would be inflated. We have therefore opted for a more conservative testing procedure. In section 4.1, we will use the more intuitive measure based on the proportions of regulars in the complete argument structure alternation classes.

Finally, it turns out that the proportion of regulars in a verb's alternation class is correlated with the (log) frequency of that verb ($r = -0.23, p < 0.0001$, $r_s = -0.33, p < 0.0001$). In an analysis of covariance, with the proportion of regulars as dependent variable, regularity has no additional explanatory power if frequency is entered into the model before regularity. However, we will see below that this measure may nevertheless be relevant for the processing of sets of words that are matched for frequency.
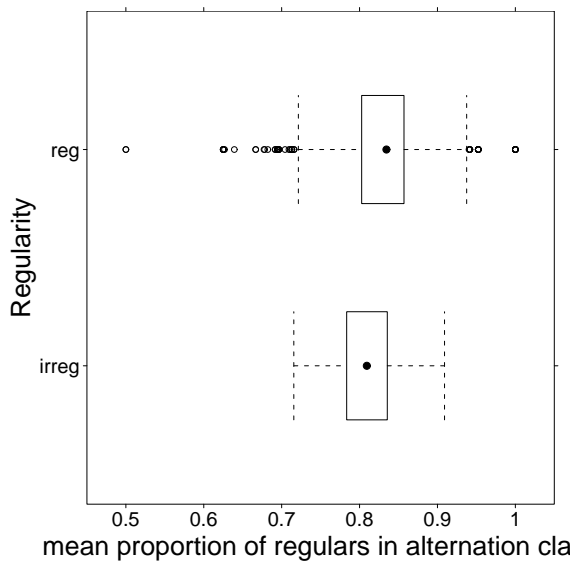

Figure 4. Boxplots for the mean proportion of regulars in a verb's alternation classes, regular (reg) and irregular (irreg) verbs.

2.4 Inflectional paradigms in German and Dutch

Thus far, we have treated regular and irregular verbs as atomic units. However, 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.

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:

	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 question sentences and topicalized sentences. 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 7 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 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 7.

Table 7

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.

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

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

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 times that 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, i.e., 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 than the unmarked, regular present tense form than the corresponding regular forms.

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 differ-

ence 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 7 shows.

Summing up, 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 to structure the temporal and aspectual flow of the unfolding discourse.

We also checked whether the pattern observed for Dutch generalizes to German. The lower half of Table 7 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 do 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.

Summing up, regular and irregular verbs differ not only with respect to the frequency of the lexeme (in the sense of Aronoff, 1994), with irregular lexemes having higher frequencies than regular lexemes. They also differ with respect to the quantitative structure of their inflectional paradigms.

2.5   Distances in context space

In the preceding analyses, we have used a range of simple measures to trace semantic differences between regular and irregular verbs. Current work on computational semantics such as latent semantic analysis (LSA, Landauer & Dumais, 1997) and hyperspace analogue to language (HAL, Lund & Burgess, 1996) provide vectors in a high-dimensional context space for words that can be used to trace systematic differences in the semantic representations of regular and irregular words.

Before going into further details, we would like to make explicit that we do not think that context-based semantic vectors would represent the full semantics of words. To the contrary, context-based semantic vectors are in many

respect highly impoverished representations. For instance, sensory-motor aspects of meaning are nearly impossible to capture on the basis of co-occurrence information in written language. In spite of their obvious limitations, semantic vectors are useful for capturing at least some important aspects of word meaning.

In what follows, we estimate semantic distances between verbs using a textual co-occurrence technique that makes use of the recurrent artificial neural network technology of Moscoso del Prado and Sahlgren (2002). Moscoso del Prado and Sahlgren trained a simple recurrent network (Elman, 1990) to predict the next word in a sequence of words from a real corpus, where each word is represented by a pattern of a few random active bits selected from a large binary vector. Once the network had been trained on this predictive task, they ran a larger corpus through the network while accumulating the activation values of the network's hidden unit in the immediate context of each word. The final representation for a given word type consists of the patterns of activation in the hidden layer of the network summed across all the contexts in which that word type appears, with the context of a particular token being the weighted sum of the patterns of activation in the hidden layer in the three preceding and three following time steps. In their experiments, Moscoso del Prado and Sahlgren showed that, after training on 20 million words from the British National Corpus (`http://www.hcu.ox.ac.uk/BNC/`), the vectors contained reliable information about word meaning, as well as information about a word's morphological and syntactic properties. (These vectors formed the semantic representations of the network model of Moscoso del Prado Martín, 2003, mentioned in the introduction.)

Using this technique, we obtained 150-element vectors for the citation forms of 1281 English simplex verbs (130 irregulars, 1151 regulars) and calculated the normalized cosine distance (Schone & Jurafsky, 2001) between each pair of verbs, increasing each value by 2.0 in order to obtain positive distances for all word pairs. The resulting 1281 by 1281 distance matrix was subjected to multidimensional scaling, using Kruskal's non-metric multidimensional scaling to obtain an optimal representation of distances in two-dimensional space that improves considerably on the representation obtained by classical multidimensional scaling (see, e.g., Cox & Cox, 2001). Figure 5 plots the regular verbs (left panel) and the irregular verbs (right panel) in the resulting plane. (We use two separate plots for expository purposes only.) Note that the irregular verbs are less spread out on the first dimension and asymmetrically positioned on the second dimension compared to the regular verbs.

Table 8 summarizes the mean and median distances within the set of regular verbs, the distances within the set of irregular verbs, and the distances between regulars and irregulars. Pairwise t-tests suggest these differences are all significant ($p < 0.0001$ for all comparisons), but given the huge numbers of distances
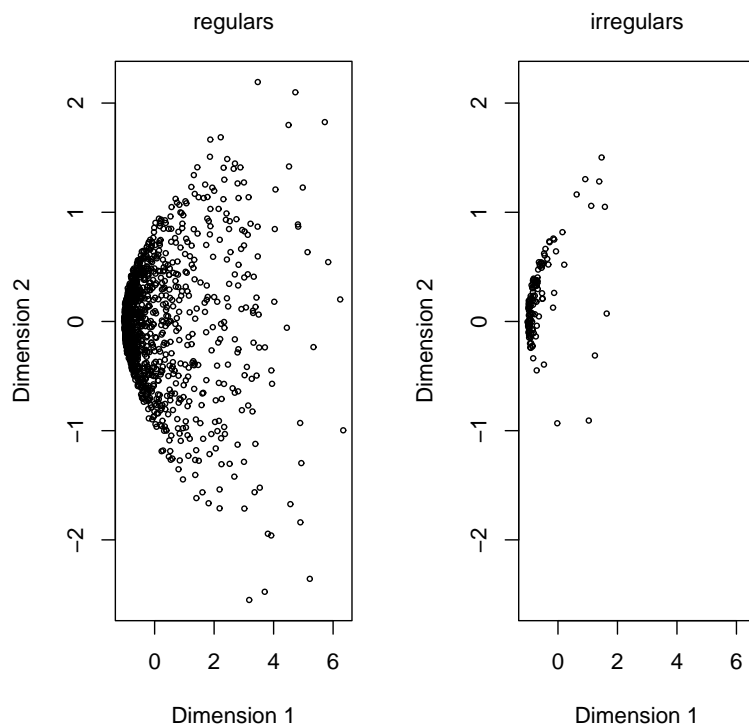
Figure 5. Distances between simplex verbs in co-occurrence space using Kruskal's non-metric multidimensional scaling (Cox & Cox, 2001). Regulars and irregulars are plotted in different panels for expository reasons only.

Table 8

 Mean and median of semantic distances within regulars, within irregulars, and between regulars and irregulars.

	regular		irregular	
	mean	median	mean	median
regular	2.32	2.13	2.11	1.89
irregular	2.11	1.89	1.69	1.55

(there are more than a million distances for the within-regular comparisons) and the skewed nature of these distance distributions, these t-tests are not very informative. We therefore studied the robustness of these differences by randomly sampling 200 distances from each of the three sets of distances and using Wilcoxon tests for the three comparisons instead of t-tests to avoid potential side-effects due to violation of the normality assumptions. Table 9 summarizes the results of 10000 Wilcoxon tests on samples of 200 distances for each comparison, listing the proportion of tests in which the Wilcoxon test was significant at the 5% and the 1% levels. The comparisons of just 200 within-regular distances with 200 within-irregular distances are always signif-

icant at both alpha levels, and the same observation holds for comparisons of within-irregular and regular-irregular comparisons. For within-regular and regular-irregular comparisons, the proportions of significant Wilcoxon tests lies far above chance level.

Table 9

Proportion of test runs (out of 10,000) in which a Wilcoxon test revealed a significant differences for 200 distances x and 200 distances y, at the 5% and 1% significance levels.

distances x	distances y	$\alpha = 0.05$	$\alpha = 0.01$
within-regular	within-irregular	1.0000	1.0000
within-regular	regular-irregular	0.9162	0.7667
within-irregular	regular-irregular	1.0000	1.0000

In order to gain some insight into the extent to which distance in context space might covary with frequency, we selected the 360 verbs with frequency greater than 400 (per 18 million) in our data set, of which 101 are irregular and 259 regular. A Wilcoxon test revealed that the average value of these irregulars on the second dimension shown in Figure 5 differed significantly from the average value of the regulars on this dimension ($W = 10612, p = 0.0054$). This allows us to conclude that the abovementioned asymmetry on this dimension for the irregulars is, at least to some non-negligible extent, independent of frequency of occurrence.

What this essay in computational semantics shows is that there are demonstrable differences in the contextual distributional properties of regular and irregular verbs. These contextual distances are implicit measures of semantic density, complementing explicit measures such as number of meanings, auxiliary selection, and asymmetries in the inflectional paradigm. It is important to keep in mind that the technique by means of which distances in syntactic-semantic space were obtained has no form information whatsoever at its disposal. The greater contextual similarity observed for irregular verbs therefore provides additional support for our hypothesis that irregular verbs cluster in semantic space to a greater extent than do regulars.

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 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, e.g., Ernestus & Baayen, 2003; Baayen, 2003; Albright, 2002, for linguistic studies, and, e.g., 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 lexemes does 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 et al. 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). Let $f(w, v)$ denote the forward strength of cue word w and target word v, and let $b(w, v)$ denote the backward strength. Denoting the number of associates of the cue word by $N(w)$, the resonance strength $r(w)$ is defined as

$$r(w) = \sum_{i=1}^{N(w)} f(w, i) b(i, w) \tag{1}$$

This measure 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) that just failed to reach significance under a two-tailed hypothesis ($t(191.9) = -1.95, p = 0.0531$). Note, however, that the difference is in the predicted direction, with irregulars revealing a higher mean connectivity than regulars, and a one-tailed test is more appropriate. Finally, we analysed 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.

It might be objected 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 linear regression analysis. To do so, however, a transformation of the highly skewed dependent variable is required to bring it in line with the normality assumptions of linear regression. 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 1000 and by taking the logarithm of the result. We analysed this rescaled resonance strength using linear regression with log surface frequency, number of synsets in Wordnet, log morphological family size, derivational entropy and regularity as predictors. 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
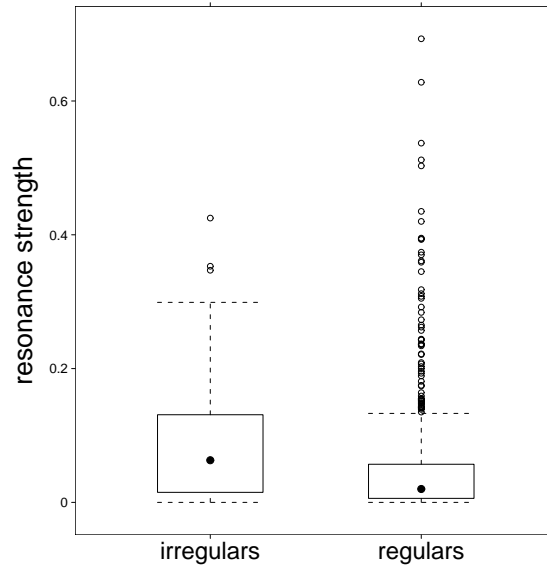
Figure 6. Boxplot for the distributions of resonance strength for irregular and regular verbs in the University of South Florida association norms (Nelson, McEvoy, & Schreiber, 1998).

connectivity in the mental lexicon that is overtly expressed in the morphology (see Schreuder & Baayen, 1997; Bertram, Baayen, & Schreuder, 2000; De Jong, Schreuder, & Baayen, 2000). Derivational entropy (Moscoso del Prado, Kostić, & Baayen, in press) is an information-theoretic analogue of family size that estimates the amount of information carried by the morphological family.

A stepwise linear regression showed resonance strength to be positively correlated with surface frequency ($t(633) = 3.641, p = 0.0003$) and with family size ($t(633) = 2.465, p = 0.0140$), and to be negatively correlated with number of synsets ($t(633) = -2.14, p = 0.0330$) and derivational entropy ($t(633) = -3.01, p = 0.0027$). Crucially, regularity emerged as a significant predictor, even in a sequential analysis of variance in which the contributions of all other factors are partialled out first ($F(1, 633) = 9.7935, p = 0.0018$). This analysis 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.

The next question that we considered was whether the higher semantic density of irregular verbs might also affect on-line behavioral measures such as response latencies. We therefore examined the databases compiled by Balota and his co-workers (Balota, Cortese, & Pilotti, 1999, Spieler & Balota, 1998, see also Balota, Cortese & Pilotti, 1999a, Spieler & Balota, 1997, and Balota & Spieler, 1998) which report reaction times for visual lexical decision and word naming in English for two age groups (young versus old participants), as well as subjective familiarity ratings. From these databases, we extracted the 854 words that according to CELEX are monomorphemic verbs that have not arisen due to conversion from a noun. (These databases do not provide any word category information themselves.) Of these 854 verbs, we selected the 832 verbs for which information is available for reaction times in both visual lexical decision and word naming. In this way, we obtained reaction times and familiarity ratings for 127 irregular and 705 regular verbs. All these data concern the form of the verb without overt inflections, i.e., the form shared by the infinitive and the present-tense form with the exception of the third person singular. For ease of reference, we will refer to these uninflected forms of the verb as the verb stem. The joint data extracted from the databases made available by the seminal work of Balota and his colleagues will be refered to as the Balota data.

From the CELEX lexical database, we added to the Balota data the surface frequencies of the verb stems, that is, the summed frequencies of all the forms listed in the CELEX wordform lexicon sharing the same orthographic form, collapsing over word category and inflectional variants. As in the study of Baayen, Feldman, and Schreuder (2004), we included a range of other distributional properties in our study, both form-related measures such as neighborhood density (the N-count of Coltheart, Davelaar, Jonasson, & Besner, 1977), word length in letters, the frequency of the initial bigram, and mean bigram frequency, as well as semantic variables: morphological family size, derivational entropy, inflectional entropy (the amount of information carried by the inflectional paradigm, see Moscoso del Prado, Kostić & Baayen, in press, for further details, and also Kostić, Marković, & Baucal, 2003), and the number of synsets in which a word is listed in WordNet.

The question of whether verb regularity has predictive value for response latencies and subjective familiarity ratings over and above covariates such as frequency, family size, inflectional entropy, and word length, is complicated by the fact that the regularity of a verb is itself partly predictable from such covariates. In order to obtain better insight into this predictability, we fitted a logistic regression model to the data with the log odds ratio of regularity versus irregularity as the dependent variable, and with the lexical measures

described above as predictors. A stepwise model selection procedure resulted in a model with frequency, morphological family size, inflectional entropy, and number of synsets as predictors. The latter two predictors had significant non-linear components (fitted with restricted cubic splines with 5 knots), as shown in Table 10.

Figure 7 plots the partial effects of these predictors on the probability of a verb being regular, i.e., the effect of one variable when the other variables are held constant. The upper left panel shows that as log frequency increases, the probability of being regular decreases, slowly at first, but more rapidly for higher frequencies. The upper right panel shows that there is a small non-linear effect for morphological family size as well. Verbs with higher family sizes are slightly more likely to be regular than irregular. This suggests that regularity favors morphological productivity. The lower left panel illustrates the relation between inflectional entropy and the probability of being regular. Verbs with a log inflectional entropy up to 1.75 are almost certainly regular, verbs with higher inflectional entropies show a marked decrease in the probability of regularity. The lower right panel presents the nonlinear partial effect of the number of synsets. Except for the smallest counts of synsets, the likelihood of regularity decreases as a verb has more meanings. The other measures listed above had no significant additional contribution. The proportion of variance explained by this model is $R^2 = 0.39$. When frequency is removed as predictor, the proportion of variance explained reduces to 0.30, which shows that although frequency has a large predictive value, it is not a predominant predictor explaining nearly all the variance by itself.

Having clarified the dependency of regularity on frequency, family size, number of synsets, and inflectional entropy, we consider the question of whether regularity is relevant as a predictor for the visual lexical decision latencies in the Balota data. It turns out that when a regression model is fitted to these latencies, regularity is not a significant predictor. However, as allmost all measures are highly correlated, it is possible that the predictivity of regularity is masked by collinearity. We therefore orthogonalized the data matrix by means of a principal components decomposition, as described in detail in Baayen, Feldman, and Schreuder (2004). Informally, orthogonalization can be viewed as a means to combine variables that are highly correlated into new predictors that are uncorrelated. In the orthogonalization of the present data matrix of verbs, six explanatory principal components were observed. The first principal component emerged as a component capturing word meaning: Variables with high loadings on this component were surface frequency, number of synsets, family size, and derivational entropy. Form variables such as word length, mean bigram frequency, and neighborhood density had loadings close to zero on this principal component, but emerged with high loadings on the second principal component. Word frequency, family size, derivational entropy, and number of synsets had low loadings on this second component. Together,
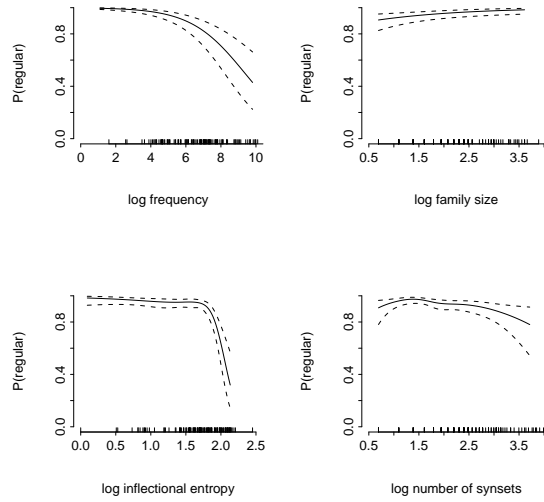
Figure 7. Partial effects of frequency, family size, inflectional entropy, and number of synsets as predictors of the probability that a verb is regular. The rug at the bottom of each panel provides the density for the irregular verbs.

Table 10
Frequency, family size, inflectional entropy, and number of synsets as predictors of the probability that a verb is regular. See also Figure 7.

Factor	Chi-Square	degrees of freedom	p-value
CelS	47.90	1	<0.0001
Vf	8.01	1	0.0047
Ient	58.70	4	<0.0001
Nonlinear	31.58	3	<0.0001
NsyS	14.90	4	0.0049
Nonlinear	9.87	3	0.0197
TOTAL NONLINEAR	38.25	6	<0.0001
TOTAL	122.53	10	<0.0001

the first two principal components account for 50% of the variance in the data matrix and separate the primary contributions of form and meaning in the distributional space of our predictors.

When we modeled the visual lexical decision latencies of the Balota data as a function of the principal components and regularity, significant effects were observed for principal components 1–4 and 6 (including nonlinear components for principal components 1, 2, and 6) and for Age, and in addition a significant interaction between regularity and the first principal component ($F(4, 1595) = 2.94, p = 0.0196$), the component capturing the contribution

of semantic variables. This interaction is shown in the upper left panel of Figure 8. The black line represents the regular verbs, and the grey line the irregular verbs. The first principal component is shown on the horizontal axis. A smaller value for this component represents the combined larger values of frequency, family size, and number of synsets. What Figure 8 shows is that the irregulars have a processing advantage in visual lexical decision compared to the regulars for the higher ranges of values of the first principal component, i.e., when frequency is low, when the family size is small, and when there are few synsets. Conversely, for lower values of PC1, Figure 8 reveals neutralization of the processing difference between regulars and irregulars.

An analysis of the subjective frequency ratings (which are available only for the young subject group) yielded similar results in an analysis without orthogonalization, in which case no effect of regularity is measurable. After orthogonalization, however, a stepwise regression analysis resulted in a model in which all six principal components were significant (with nonlinear components for the principal components 1,2, and 6), and in which regularity was significant as main effect ($F12, 1608) = 6.41, p < 0.0001$) as well as in interaction with principal components 1–5 (for all interactions, $F(11, 1608) = 6.49, p < 0.0001$). The R^2 was 0.599 and was adjusted to 0.586 in a bootstrap validation of the model, so we are not overfitting the data. The upper right panel of Figure 8 shows the interaction of PC1 and regularity for the subjective frequency ratings. Note that the general pattern is the mirror image of the pattern for lexical decision. This is a consequence of the intrinsic negative correlation of ratings and response latencies. Words that elicit short RTs elicit large ratings, while words that elicit long RTs generally receive low ratings. Note that the downward bulge for the interval $[0, 2]$ of PC1 in lexical decision is mirrored by a small upward bulge for the ratings. The interaction observed for PC1 and regularity in visual lexical decision therefore receives independent support from a similar interaction in subjective frequency rating. Our hypothesis is that the greater semantic density of irregulars is responsible for this interaction. For high-frequency verbs with many meanings and large morphological families, there is no detectable additional benefit from having a denser semantic neighborhood, but for lower-frequency verbs with smaller morphological families and fewer meanings, this benifit is visible in the principal components regression.

The lower panels of Figure 8 highlight two other interactions in the rating data. The lower left panel clarifies the interaction between PC2 and regularity. A high value of PC2 correlates with a greater length and a higher mean bigram frequency, and a lower neighborhood density. Apparently, longer irregulars are rated lower than longer regulars, other things being equal. The lower right panel shows that PC3 is relevant only for the irregulars. PC3 correlates most strongly with inflectional entropy. Therefore, inflectional entropy seems to be especially important for irregular verbs, with a greater inflectional en-
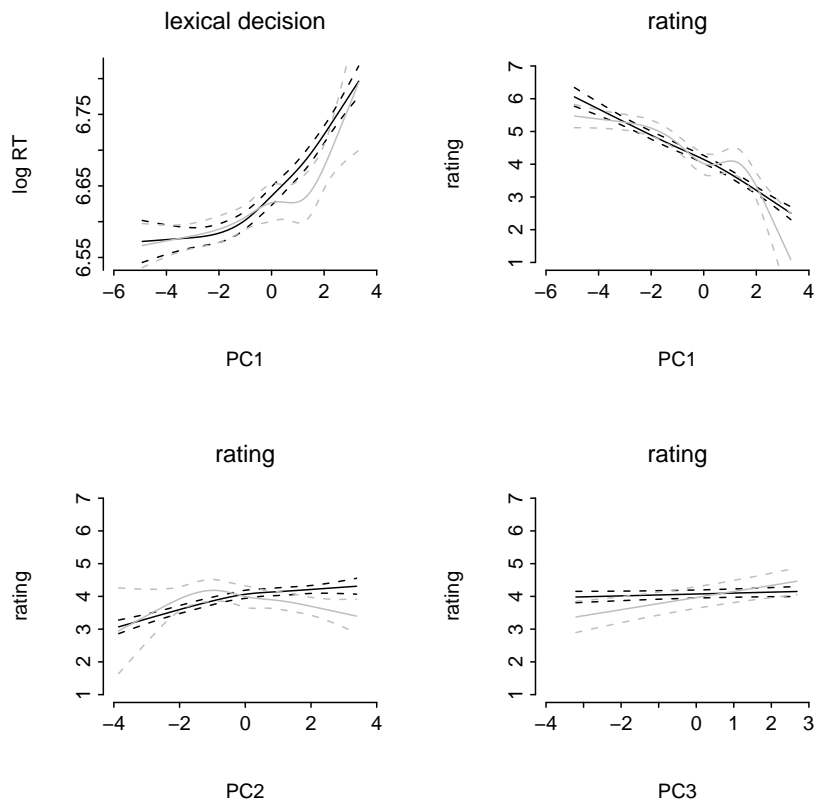
Figure 8. Interaction of regularity by principal components in visual lexical decision (upper left panel) and subjective familiarity ratings (remaining panels). Black lines represent regulars, grey lines represent irregulars. Dotted lines show the 95% confidence intervals.

tropy leading to higher ratings. The positive correlation of inflectional entropy with the ratings is mirrored by the negative correlation between inflectional entropy and visual lexical decision latencies. Unlike for the ratings, however, inflectional entropy is facilitatory for both regulars and irregulars in visual lexical decision. We think that, in general, a greater entropy, i.e., a greater information load of the inflectional paradigm, provides a source of evidence in lexical decision that an item is indeed a word, and likewise makes a word feel more familiar.

The analysis of the visual lexical decision latencies and subjective frequency ratings revealed subtle effects of regularity that became detectable only after orthogonalization. For the word naming latencies, the effect of regularity is much easier to observe. Table 11 lists the predictors that emerged from a stepwise regression analysis without orthogonalization. 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

34

Table 11

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

Factor	d.f.	F	P
age	1, 1655	24371.90	<.0001
inflectional entropy	1, 1655	4.34	0.0375
neighborhood density	4, 1655	44.12	<.0001
nonlinear	3, 1655	5.57	0.0008
frequency	1, 1655	172.06	<.0001
regularity	1, 1655	7.14	0.0076

density (a greater neighborhood density led to shorter naming latencies; this effect had a non-linear 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. Below, we will return to this naming disadvantage for irregulars. For this model, $R^2 = 0.937$. This value remained the same to three decimal digits in bootstrap validation, indicating that we are not overfitting the data. Note that for the verbs studied in the present paper, inflectional entropy is inhibitory in naming but facilitatory in lexical decision. Our hypothesis is that in naming, a greater inflectional entropy is indicative of greater competition within the inflectional paradigm for the inflectional form to be selected for pronunciation.

It is important to keep in mind that the processing differences observed for regular and irregular verbs concern 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, past participle) on the processing of the present tense form.

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, as well as 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 A PET study on English irregulars

Jaeger, Lockwoord, Kemmerrer, Van Valin, and Murphy (1996) asked participants to read aloud several lists of words. One of these lists contained only regulars verb stems, another of these lists contained 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 the word lists of Jaeger et al. (1996) 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 from the Balota database as well as data on the probability of regulars in the verb alternation classes discussed in section 2.3. This left us with 76 verbs (out of 92 in the original study). It turned out that their regular and irregular verbs were perfectly matched on all but two variables, the inflectional entropy measure of Moscoso del Prado, Kostić and Baayen (2003) (see also Kostić, Marković, & Baucal, 2003), and the proportion of regulars in a verb's argument structure alternation classes.

Inflectional entropy quantifies (in bits) the amount of information carried by a word's inflectional paradigm. The higher the inflectional entropy, the greater the information load of the inflectional paradigm, and the greater its paradigmatic uncertainty. This measure assesses the probability distribution of forms within a word's inflectional paradigm from an information theoretic perspective, complementary to the descriptive and comparative approach used in section 2.4 to study differences in the frequency distributions within the inflectional paradigms of Dutch and German. The mean Inflectional Entropy for the regular verbs in the Jaeger et al. study was 1.60, for the irregular verbs the Inflectional Entropy was 1.78, a difference that is statistically significant $(t(75.871) = -3.17, p = 0.0022)$.

Recall that Figure 4 in section 2.3 showed that irregulars tend to participate in alternation classes with more irregulars than is the case for regulars. In the dataset of 96 verbs of the Jaeger et al. (1996) study, a similar pattern is visible. Two of their irregular verbs and eleven of their regular verbs are not mentioned in Levin's alternation classes. For the remaining verbs, the
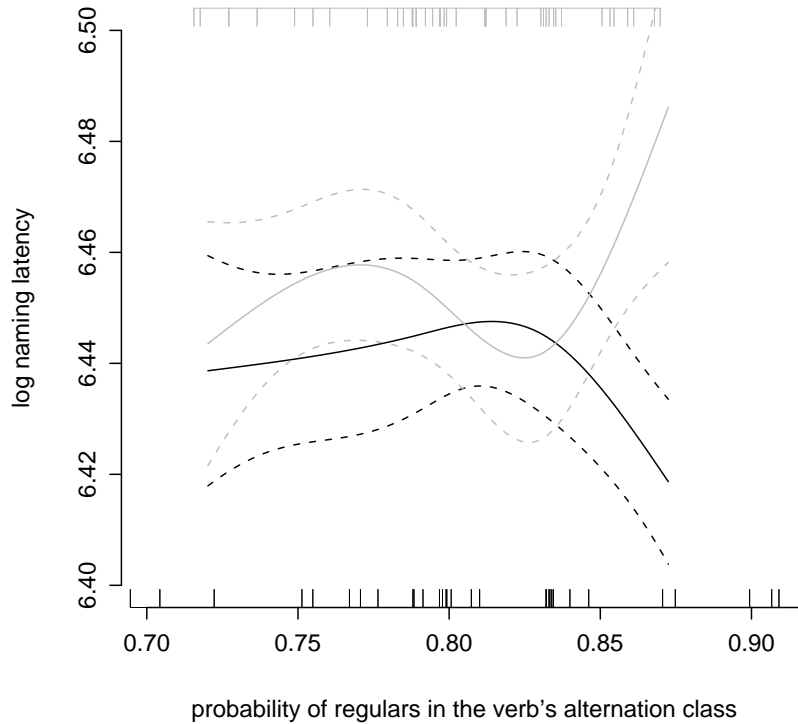
Figure 9. Interaction of regularity by the probability of regulars in the argument alternation classes, with 95% confidence intervals. The black lines represent regulars, the grey lines, irregulars. The rug on the horizontal axis shows the density of the regulars, the rug at the top of the plot shows the density of the irregulars.

difference in a verb's average proportion of regulars in the alternation classes approaches significance $(t(68.75) = 1.949, p = 0.05542)$ in a two-tailed test.

To gain further insight in the potential relevance of these differences, we studied this data set in further detail. A stepwise linear regression model revealed significant main effects of age (younger participants have shorter naming latencies, $t(141) = -61.01, p < 0.0001$), inflectional entropy (a greater inflectional entropy led to longer response latencies, $t(141) = 3.42, p = 0.0008$), length (longer words elicited longer response latencies, $t(141) = 2.75, p = 0.0066$), and a nonlinear interaction (involving a restricted cubic spline with four knots) of regularity by the probability of regulars in the verb's alternation classes $(F(3, 141) = 5.06, p = 0.0023)$. This interaction is shown in Figure 9, and points to a processing advantage for regulars and a processing disadvantage for irregulars when there are very few irregulars in the argument alternation classes. The R^2 for this model was 0.962. In bootstrap validation, this R^2 was adjusted to 0.957, which indicates the model is reliable and does not use too many parameters to fit the data.

37

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 to, on the one hand, 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 by 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 et al. 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. However, what our present data show is that a pattern very similar to what Jaeger et al. 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 the light of the dual route mechanism proposed by Pinker (1991) and subsequent studies such 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 such as *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 ($F(1, 146) = 6.7656, p = 0.0103$) 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, e.g., Taft, 1979, and Baayen, Dijkstra, & Schreuder, 1997).

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 in the study of 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 uncertainty (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 reported for the PET data of Jaeger and colleagues.

Beretta, Campbell, Carr, Huang, Schmitt, Christianson, and Cao (2003) studied regular and irregular nouns and verbs in German using event-related fMRI. They report, first, that irregulars showed greater activation overall than did regulars, and, second, that regulars revealed greater lateralization to the left hemisphere compared to 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 12 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 12 regulars and 12 irregulars, pairwise matched for frequency. Following Marcus, Brinkman, Clahsen, Wiese, and Pinker (1995), they contrasted nouns pluralizing in -s as regulars with nouns in -er as irregulars.
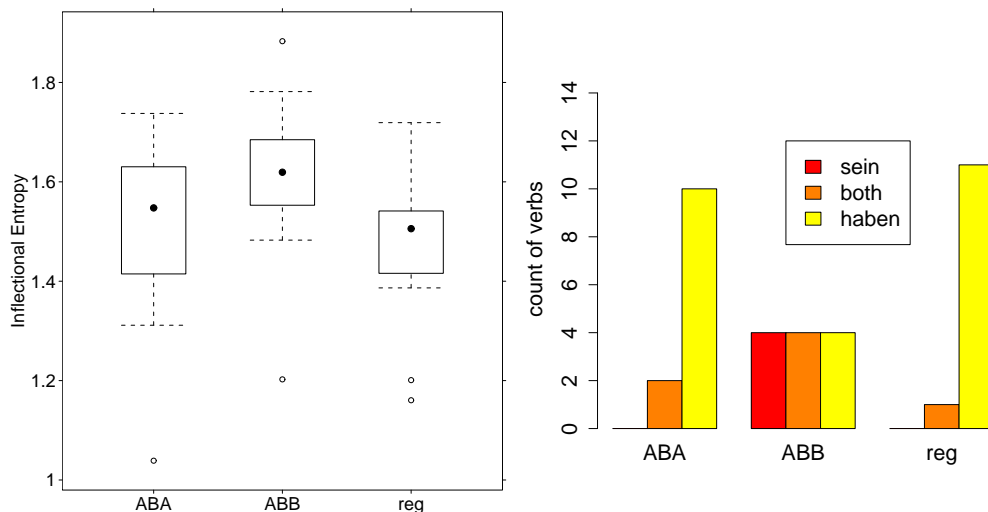

Figure 10. Boxplot of Inflectional Entropy by verb class (left panel), and barplot for auxiliary by verb class (right panel) for the materials of Beretta et al. (2003).

First consider the verbs in this study. The left panel of Figure 10 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 of Jaeger et al. (1996). Recall that the irregular verb stems of 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 et al. (2003) for irregulars is in line with their greater inflectional information load, which may well require more selection resources.

The right panel of Figure 10 shows that the regular and irregular verb classes are also not matched with respect to auxiliary choice. 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 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 finding of Beretta et al. (2003) that brain activity for irregulars is less restricted to the left hemisphere than it is for regulars.
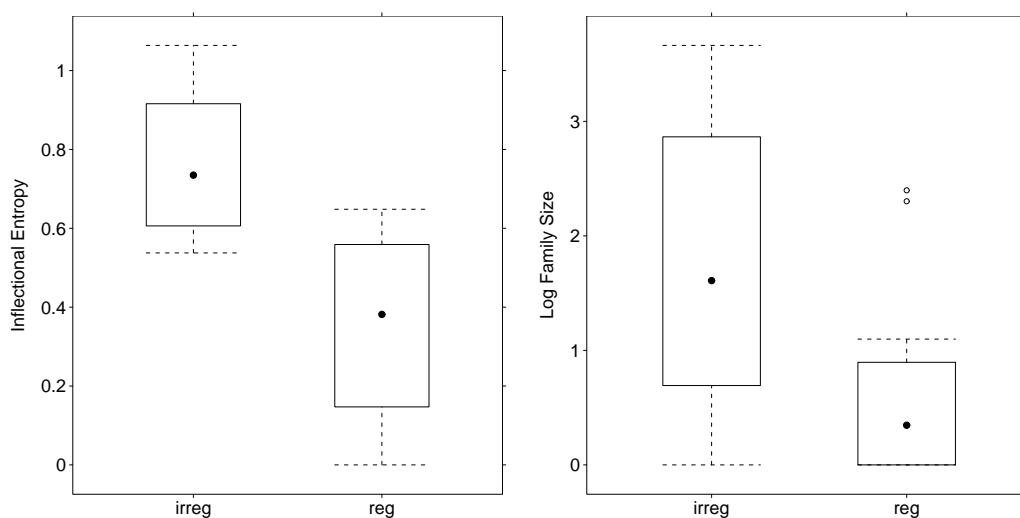

Figure 11. Boxplots of Inflectional Entropy by noun class (left panel), and of Log Family Size by noun class (right panel) for the materials of Beretta et al. (2003). Reg: regular noun; Irreg: irregular noun.

Finally, consider the nouns in this ER-fMRI study. The boxplots in Figure 11 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 of this study, have higher inflectional entropy and larger morphological families. The former is

correlated with a higher processing load, 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 et al. (2003) 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 investigate the possibility that words with a greater semantic density elicit more activation in the brain across a wider range of areas.

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. They 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 those by 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 of Jaeger et al. (1996) and Beretta et al. (2003). On the other hand, these differences are in line with the spirit of connectionist approaches to morphological processing, and were in fact predicted by McClelland and Patterson (2000) to exist. 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 (pages 214–215). 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 the information complexity of 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 by Ullman et al. (1997).

However, the goal of this study 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 with respect 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 (2002) has documented attractor effects within phonological similarity sets for regular verbs in Italian, contradicting the claims of the dual route model, although he couches his findings within a highly complex system of rules. As evidenced by the studies in Bod, Jannedy, and Hay (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.

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, non-systematic 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.

The greater semantic density of irregulars may be the synchronic reflex of diachronic age. The irregular verbs tend to be the older verbs of the language. Through time, they may have acquired a greater range of meanings than the younger words, which typically have more specialized and less rich semantics. Moreover, through time, the greater semantic density of irregulars may have protected them against regularization. In both diachrony and synchrony, the greater semantic density of irregulars may well be functional, perhaps along the lines envisioned by Joanisse and Seidenberg (1999) and Patterson et al. (2001).

The diagnostics that we have 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. However, the combined evidence provided by a series of measures in three Germanic languages, all of which reveal a consistent pattern and which, when pitted against behavioral data, reveal a surprising degree of statistical robustness, suggests that we are on the right track. 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 research will lead to a much more profound understanding of the interaction between semantics and regularity. This is uncharted area, and we have only scratched its surface.

References

Albright, A., Hayes, B., 2001. Rules vs. analogy in English past tenses: A computational/experimental study. Manuscript UCLA.

Aronoff, M., 1994. Morphology by itself. The MIT Press, Cambridge, Mass.

Baayen, R., 2003. Probabilistic approaches to morphology. In: Bod, R., Hay, J., Jannedy, S. (Eds.), Probability theory in linguistics. The MIT Press, pp. 229–287.

Baayen, R., Feldman, L., Schreuder, R., 2004. A multivariate study of simple word recognition. Submitted.

Baayen, R. H., Dijkstra, T., Schreuder, R., 1997. Singulars and plurals in

Dutch: Evidence for a parallel dual route model. Journal of Memory and Language 36, 94–117.

Baayen, R. H., Piepenbrock, R., Gulikers, L., 1995. The CELEX lexical database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Baayen, R. H., Sproat, R., 1996. Estimating lexical priors for low-frequency morphologically ambiguous forms. Computational Linguistics 22, 155–166.

Balota, D., Cortese, M., Pilotti, M., 1999a. Item-level analyses of lexical decision performance: Results from a mega-study. In: Abstracts of the 40th Annual Meeting of the Psychonomics Society. Psychonomic Society, Los Angeles, CA, p. 44.

Balota, D., Cortese, M., Pilotti, M., 1999b. Visual lexical decision latencies for 2906 words. [On-line].

Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D., 2003. Visual word recognition for single-syllable words. Submitted.

Balota, D. A., Spieler, D. H., 1998. The utility of item level analyses in model evaluation: A reply to Seidenberg and Plaut (1998). Psychological Science 9, 238–241.

Bauer, L., 2001. Morphological Productivity. Cambridge University Press, Cambridge.

Beretta, A., Campbell, C., Carr, T., Huang, J., Schmitt, L. M., Christianson, K., Cao, Y., 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, R., Baayen, R. H., Schreuder, R., 2000. Effects of family size for complex words. Journal of Memory and Language 42, 390–405.

Bod, R., Hay, J., Jannedy, S. e., 2003. Probability theory in linguistics. The MIT Press, Cambridge, MA.

Boersma, P., Hayes, B., 2001. Empirical tests of the gradual learning algorithm. Linguistic Inquiry 32, 45–86.

Bybee, J. L., 1985. Morphology: A study of the Relation between Meaning and Form. Benjamins, Amsterdam.

Bybee, J. L., Slobin, D. I., 1982. Rules and schemas in the development and use of the english past tense. Language 58, 265–289.

Carroll, J. 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, H., 1999. Lexical entries and rules of language: a multi-disciplinary study of German inflection. Behavioral and Brain Sciences 22, 991–1060.

Cleveland, W. S., 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74, 829–836.

Coltheart, M., Davelaar, E., Jonasson, J. T., Besner, D., 1977. Access to the internal lexicon. In: Dornick, S. (Ed.), Attention and performance. Vol. VI. Erlbaum, Hillsdale, New Jersey.

Cox, T., Cox, M., 2001. Multidimensional scaling. Chapman & Hall, New

York.

De Jong, N. H., Schreuder, R., Baayen, R. H., 2000. The morphological family size effect and morphology. Language and Cognitive Processes 15, 329–365.

Elman, J. L., 1990. Finding structure in time. Cognitive Science 14, 179–211.

Ernestus, M., Baayen, R. H., 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. Language 79, in press.

Feldman, L., Prostko, B., 2001. Graded aspects of morphological processing: Task and processing time. Brain and Language 81, 1–16.

Fellbaum, C. e., 1998. WordNet: An electronic database. The MIT Press, Cambridge, MA.

Hahn, U., Nakisa, R., 2000. German inflection: Single route or dual route? Cognitive Psychology 41, 313–360.

Jaeger, J. J., Lockwood, A. H., Kemmerer, D. L., Van Valin, R. D., Murphy, B. W., 1996. A positron emission tomographic study of regular and irregular verb morphology in English. Language 72, 451–497.

Joanisse, M., Seidenberg, M., 1999. Impairments in verb morphology following brain injury: a connectionist model. Proceedings of the National Academy of Sciences, USA 96, 7592–7597.

Kim, J., Pinker, S., Prince, A., Prasada, S., 1991. Why no mere mortal has ever flown out to center field. Cognitive Science 15, 173–218.

Köhler, R., 1986. Zur linguistischen Synergetik: Struktur und Dynamik der Lexik. Brockmeyer, Bochum.

Kostić, A., Marković, T., Baucal, A., 2003. Inflectional morphology and word meaning: orthogonal or co-implicative domains? In: Baayen, R. H., Schreuder, R. (Eds.), Morphological structure in language processing. Mouton de Gruyter, Berlin, pp. 1–44.

Landauer, T., Dumais, S., 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological Review 104 (2), 211–240.

Levin, B., 1993. English Verb Classes and Alternations. A preliminary Investigation. The University of Chicago Press, Chicago.

Lieber, R., Baayen, R. H., 1997. A semantic principle for auxiliary selection in Dutch. Natural Language and Linguistic Theory 15, 789–845.

Lund, K., Burgess, C., 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. Behaviour Research Methods, Instruments, and Computers 28 (2), 203–208.

MacWhinney, B., Leinbach, J., 1991. Implementations are not conceptualizations: revising the verb learning model. Cognition 40, 121–157.

Marcus, G. F., Brinkman, U., Clahsen, H., Wiese, R., Pinker, S., 1995. German inflection: The exception that proves the rule. Cognitive Psychology 29, 189–256.

McClelland, J., Patterson, K., 2002. Rules or connections in past-tense inflections: What does the evidence rule out? TRENDS in Cognitive Sciences 6, 465–472.

McClelland, J. L., Rumelhart, D. E., 1986. On Learning the Past Tenses of

English Verbs. Vol. 2. The MIT Press, Cambridge, Mass., pp. 216–271.

Miller, G. A., 1990. Wordnet: An on-line lexical database. International Journal of Lexicography 3, 235–312.

Moscoso del Prado Martín, F., 2003. Paradigmatic Effects in Morphological Processing: Computational and cross-linguistic experimental studies. MPI Series in Psycholinguistics. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Moscoso del Prado Martín, F., Kostić, A., Baayen, R. H., in press. Putting the bits together: An information theoretical perspective on morphological processing. Cognition.

Moscoso del Prado Martín, F., Sahlgren, M., 2002. An integration of vector-based semantic analysis and simple recurrent networks for the automatic acquisition of lexical representations from unlabeled corpora. In: Lenci, A., Montemagni, S., Pirrelli, V. (Eds.), Proceedings of the LREC'2002 Workshop on Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data. European Linguistic Resources Association, Paris.

Nakisa, R., Plunkett, K., Hahn, U., 2000. A cross-linguistic comparison of single and dual-route models of inflectional morphology. In: Broeder, P., Murre, J. (Eds.), Models of language acquisition: Inductive and Deductive Approaches. Oxford University Press, Oxford, pp. 201–221.

Nelson, D. L., McEvoy, C. L., Schreiber, T. A., 1998. The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Patterson, K., Lambon Ralph, M., Hodges, J., McClelland, J., 2001. Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. Neuropsychologia 39, 709–724.

Pinker, S., 1991. Rules of language. Science 153, 530–535.

Pinker, S., 1997. Words and rules in the human brain. Nature 387, 547–548.

Pinker, S., 1999. Words and Rules: The Ingredients of Language. Weidenfeld and Nicolson, London.

Pinker, S., Ullman, M., 2002. Combination and structure, not gradedness, is the issue. TRENDS in Cognitive Sciences 6, 472–474.

Plunkett, K., Juola, P., 2000. A connectionist model of English past tense and plural morphology. Cognitive Science.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., 1985. A comprehensive grammar of the English language. Longman, London.

Ramscar, M., 2002. The role of meaning in inflection: Why the past tense doesn't require a rule. Cognitive Psychology 45, 45–94.

Randall, J., Van Hout, A., Weissenborn, J., Baayen, R., 2003. Acquiring unaccusativity: a cross-linguistic look. In: Alexiadou, A., Anagnostopoulou, E., Everaert, M. (Eds.), The unaccusativity puzzle. Oxford University Press, Oxford, p. in press.

Schone, P., Jurafsky, D., 2001. Knowledge free induction of inflectional morphologies. In: Proceedings of the North American Chapter of the Association

for Computational Linguistics NAACL-2001.

Schreuder, R., Baayen, R. H., 1997. How complex simplex words can be. Journal of Memory and Language 37, 118–139.

Shirai, Y., Anderson, R., 1995. The acquisition of tense-aspect morphology: a prototype account. Language 71, 743–762.

Spieler, D. H., Balota, D. A., 1997. Bringing computational models of word naming down to the item level. Psychological Science 6, 411–416.

Spieler, D. H., Balota, D. A., 1998. Naming latencies for 2820 words. [On-line].

Tiersma, P. M., 1982. Local and General Markedness. Language 58, 832–849.

Ullman, M., Bergida, R., O'Craven, K. M., 1997. Distinct fMRI activation patterns for regular and irrgular past tense. NeuroImage 5, 549.

Vossen, P., Bloksma, L., Boersma, P., 1999. The Dutch WordNet (CD-ROM). European Language Resources Association (ELRA), Luxembourg.