

Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch

Wieke M. Tabak, Robert Schreuder, & R. Harald Baayen

1 Introduction

A series of studies (Pinker 1991, 1997; Kim et al. 1991; Pinker 1999; Pinker and Ullman 2002) have argued that differences between irregular and regular verbs are restricted to form. However, there are also studies which suggest that a strict separation of form and meaning may be counterproductive for theories of the mental lexicon. Bybee and Slobin (1982) observed that under time pressure, irregular past tense forms such as *sat* were produced upon presentation of *seat*, instead of the correct regular past tense form *seated*. Apparently, high-frequency irregular past tense forms served as attractors for the past-tense of semantically related regulars. Furthermore, Ramscar (2002) reported that when subjects were asked to say the past tense for pseudowords, the semantic context co-determined whether a regular or an irregular past tense was produced. This result also suggests that regularity interacts with semantics.

Baayen and Moscoso del Prado Martín (2005) addressed the question of the extent to which regular and irregular verbs might differ with respect to distributional semantic variables. They found that when regulars and irregulars are matched for frequency, irregular verbs tend to have more meanings than regular verbs. Regular and irregular verbs also tended to have different aspectual properties, as witnessed by the non-uniform distribution of auxiliary verbs in Dutch. Regulars favored *hebben*, 'have', while irregulars favor *zijn*, 'be', the auxiliary marking telicity (Randall, van Hout, Weissenborn, and Baayen 2003; Lieber and Baayen 1997). In English, the two kinds of verbs were non-uniformly distributed over Levin (1993)'s argument structure alternation classes. In addition, irregulars clustered more tightly in contextual space than regular verbs, and they had greater resonance strength according

to the association norms of Nelson, McEvoy, and Schreiber (1998). This led Baayen and Moscoso del Prado Martín to the hypothesis that irregulars have greater semantic densities than regulars.

Baayen and Moscoso del Prado Martín also assessed the processing consequences of these different distributional properties by means of the subjective frequency ratings and word naming latencies in the database of simple monosyllabic English words compiled by Spieler and Balota (1998); Balota, Cortese, and Pilotti (1999). For instance, an analysis of covariance of the word naming latencies revealed a main effect of regularity, after having partialled out the effect of frequency, with irregulars eliciting slightly longer naming latencies than regulars. Interestingly, this effect emerged for simple uninflected present-tense forms, which by themselves are not irregular. Longer naming latencies for irregulars were also observed for the present tense forms used in the study by Jaeger, Lockwood, Kemmerrer, Van Valin, and Murphy (1996). These authors also observed different patterns of brain activity using fMRI in a naming study in which subjects had to say the past-tense form upon presentation of the present-tense form. In the light of the findings of Baayen and Moscoso del Prado Martín these different patterns of brain activity may also have arisen, at least in part, as a consequence of the different semantic and syntactic properties of irregular verbs.

The aim of the present paper is to study the distributional properties of regular and irregular verbs and their consequences for lexical processing in greater depth for Dutch. In section 2, we describe the lexical statistics of Dutch verbs, and investigate to what extent distributional variables may predict regularity as well as a verb's historical age. Section 3 reports the results of a large regression study using visual lexical decision as a first step towards understanding the predictivity of the distributional variables for lexical processing in Dutch. In this study, we orthogonally contrasted number (singular versus plural), tense (present versus past), and regularity (regular versus irregular), while including a large number of covariates in the analysis. We also investigated the potential effect of the sex of the subject, as Ullman, Estabrooke, Steinhauer, Brovotto, Pancheva, Ozawa, Mordecai, and Maki (2002) reported sex-related processing differences for regulars and irregulars.

2 Lexical statistics of Dutch simplex verbs

2.1 Data selection

We selected all verbs from the Dutch section of the CELEX lexical database (Baayen, Piepenbrock, and Gulikers 1995) that are not characterized as complex. For each verb, we registered the form of its stem (which is identical to the present first person singular), the form of the singular past tense, and the form of its past participle. For each verb form, we determined its wordform frequency in CELEX. We discarded those verbs for which one or more of these three inflectional variants had zero frequency in CELEX. In this way, we obtained a dataset with 1061 verbs; 177 irregular verbs and 884 regular verbs.

For each verb form, we calculated its orthographic length, its mean bigram frequency, and its neighborhood density. We defined a wordform's mean bigram frequency as the mean of the logarithms of the frequencies of all consecutive letter pairs in the word, where we included the initial and final space as letters. A word's neighborhood density estimates the number of words in the lexicon that are orthographically (or phonologically) very similar to that word. This similarity is generally approximated by means of a Hamming distance of 1, i.e., by counting the number of words with the same length but differing in one and only one letter, cf. Coltheart, Davelaar, Jonasson, and Besner (1977).

For each verb, we extracted from CELEX the lemma frequency, the summed frequencies of all inflectional variants. In addition, we calculated its morphological family size, the number of complex words in which the verb stem occurs as a constituent (Schreuder and Baayen 1997), the auxiliary selected by the verb for the present perfect, as well as its number of verb synsets in EuroWordNet (Vossen, Bloksma, and Boersma 1999). A synset (synonym set) in WordNet represents one underlying lexical concept. Examples of synsets are provided in Table 1. A synset may consist of a single verb, but generally may comprise both simplex verbs, complex verbs and phrasal verbs. Moreover, a given verb may occur in more than one synset, in which case that verb has several distinguishable meanings. In our analyses, we use the number of synsets as an estimate of a verb's number of meanings.

Another variable that Baayen and Moscoso del Prado Martín (2005) observed to be predictive for regularity is the number of argument structure alternation classes in which a verb participates, based on the alternation classes listed in Levin (1993). In the absence of a similar classification for Dutch, we gauged the syntactic properties of a verb through a count of the number

Table 1: The synsets of *lopen*, 'walk', in the Dutch EuroWordNet.

-
- 1 doorgaan (*go on*), aanhouden (*keep on*), continueren (*continue*), doorlopen (*pass through*), lopen (*walk*), voortduren (*continue*)
 - 2 draaien (*turn*), lopen (*walk*)
 - 3 leiden (*lead*), lopen (*walk*), lopend voeren (*lead along*)
 - 4 lopen (*walk*), gaan (*go*), treden (*tread*)
 - 5 verlopen (*elapse*), gaan (*go*), lopen (*walk*), marcheren (*march*)
 - 6 vloeien (*flow*), lopen (*walk*)
 - 7 volgen (*follow*), doorlopen (*keep on walking*), lopen (*walk*)
 - 8 zitten (*sit*), liggen (*lie*), lopen (*walk*), staan (*stand*)
-

of argument structures and complementation patterns in a dictionary derived data resource compiled at the Max Planck Institute for Psycholinguistics, Nijmegen. Table 2 lists examples for the verb *werken*, 'to work'.

We also kept track of whether a verb had both irregular and regular inflectional variants. We will refer to such verbs as verbs with doublets. An example of such a verb is *jagen* ('to hunt'), which has both a regular past tense form, *jaagde*, and an irregular past tense form, *joeg*. Finally, we registered whether a verb belonged to the native stratum or the non-native stratum of the Dutch lexicon.

In addition, we calculated the inflectional and derivational entropy measures developed by Kostić, Marković, and Baucal (2003); Moscoso del Prado Martín, Kostić, and Baayen (2004). 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 (Unlike the study of Moscoso del Prado Martín, Kostić, and Baayen (2004), we calculated the inflectional entropies strictly from the frequencies of the verbal inflected forms and discarded inflectional variants with zero frequencies). The derivational entropy is a measure for the informational complexity of the morphological family. It can be viewed as an alternative to the family size count,

Table 2: Examples of argument structures and complementation patterns for the Dutch verb *werken*, 'to work'

- 1 op het land werken (*to work the soil*)
 - 2 voor een examen werken (*to study for an exam*)
 - 3 onder iemand werken (*to work under someone*)
 - 4 de funderingen werken (*the foundation is settling*)
 - 5 iets in de hand werken (*to pave the way for something*)
 - 6 iemand eruit werken (*to get rid of someone*)
-

an alternative in which the family members are not only counted but also weighted for their token frequencies.

In response to the studies of Pinker and Prince (1988); Kim, Pinker, Prince, and Prasada (1991); Pinker (1999); Ramscar (2002), we included the ratio of the nominal and verbal frequencies (henceforth, the noun-verb ratio) into our analyses. Pinker and Prince (1988) claim that verbs that are derived from nouns by conversion are regular, irrespective of their phonological or semantic properties. Ramscar (2002), however, obtained empirical evidence contradicting this generalization. We therefore included the noun-verb ratio as a measure for the likelihood of conversion and potentially the likelihood of regularity.

Finally, we considered the ratio of the verb's frequency in spoken and written Dutch (henceforth, the spoken-written frequency ratio). We included this measure in order to investigate potential effects of register and age of acquisition (Baayen 2005).

Although we selected only those verbs from CELEX that are not marked as complex, there are verbs that nevertheless have some morphological structure. An example from the set of irregular verbs is *beginnen*, 'to begin'. There is no independent verb *ginnen* in Dutch, but its past participle reveals the presence of the prefix. For truly simplex verbs, Dutch participles take the prefix *ge-*, while complex verbs are formed without this prefix. The past participle of *beginnen* is *begonnen* and not *gebegonnen*, which shows that it is at least formally complex. Within the set of regular verbs, there were two kinds of partially complex verbs: verbs like *verkennen*, 'to explore', and verbs such

Table 3: Counts of verbs broken down by regularity, stratum, and complexity.

Stratum	Regularity	Complexity		
		iterative	complex	simplex
native	regular	111	119	527
native	irregular	0	27	150
non-native	regular	0	125	2
non-native	irregular	0	0	0

as *hinkelen* 'to play hopscotch'. Verbs such as *verkennen* were not marked as complex because the combination of the prefix *ver-* and the base verb *kennen*, 'to know', is not fully compositional. Verbs such as *hinkelen* have a final semi-suffix expressing iteration. Table 3 lists the counts of verbs broken down by regularity, stratum, and complexity. The two non-native simplex verbs in this table are *borduren*, 'to embroider', and *polijsten*, 'to polish'. The primary stress on the second syllable marks these verbs as non-native. All other non-native verbs are complex in the sense that they contained the suffix *-eer*.

2.2 Predicting regularity

The question that arises at this point is to what extent it might be possible to predict the regularity of a verb from the numerical predictors described above. Since it is clear from Table 3 that the non-native verbs and the complex verbs in our dataset are (nearly) all regular, we restricted ourselves to the 677 native simplex verbs. In order to gauge the extent to which it is possible to predict whether a verb is regular from its distributional properties, we fitted a logistic regression model to this data set, using a stepwise model selection procedure. For each predictor in the model we investigated whether it was linear or non-linear, using restricted cubic splines. For technical details, the reader is referred to Harrell (2001).

Figure 1 shows how the probability of being regular can be predicted from the verb's lemma frequency (upper left panel, $X^2_{(1)} = 12.74, p = 0.0004$), its family size (upper middle panel, $X^2_{(2)} = 11.99, p = 0.0025$, non-linear: $X^2_{(1)} = 9.31, p = 0.0023$), its argument structures (upper right panel, $X^2_{(1)} = 10.63, p = 0.0011$), its neighborhood density (first panel on the second row, $X^2_{(2)} = 17.98, p = 0.0001$, non-linear: $X^2_{(1)} = 10.49, p = 0.0012$), its inflectional entropy (central panel of Figure 1 $X^2_{(1)} = 9.76, p = 0.0018$), its auxil-

aries (last panel on the second row, $X_{(1)}^2 = 17.67, p = 0.0001$), its noun-verb frequency ratio (lower left panel, $X_{(1)}^2 = 7.73, p = 0.0054$), and its spoken-written frequency ratio (lower middle panel, $X_{(1)}^2 = 4.05, p = 0.0441$). By itself, the number of verbal synsets is also predictive for regularity, but as this variable is highly correlated with the count of argument structures and complementation patterns, it does not emerge with an independent contribution within this regression model. Finally, we note that very similar results are obtained with a logistic regression model with the complete dataset of 1061 verbs.

Note that the likelihood that a verb is regular or irregular can be predicted not only from its frequency, but also from other distributional variables. What this shows is that morphological regularity is not just a matter of form. Morphological (ir)regularity has consequences for syntax (the argument structure alternation classes), semantics (number of synsets, auxiliary selection), stem productivity (family size), register (spoken-written frequency ratio), and the information complexity of the inflectional paradigm (inflectional entropy). These findings falsify the claim advanced by Pinker and colleagues that morphological irregularity would be restricted to form.

2.3 Historical age

Thus far, we have not taken into consideration that there are differences in the historical age of the verbs that we have been studying. It is possible that the more frequent verbs in the language are also the verbs that have a longer history of use through the ages. In order to gain insight into how historical age relates to regularity and the quantitative variables studied above, we consulted an etymological dictionary of Dutch (De Vries and De Tollenaere 1991) for a subset of the data consisting of equal numbers of regular and irregular verbs approximately matched for lemma frequency.

We selected this subset for three reasons. First, the size of the new subset, 286 words, is just about the maximum that we could present in the visual lexical decision experiment reported below — the 286 words have to be paired with 286 pseudowords. Second, we thought it desirable to study the same number of regular and irregular verbs, giving equal numeric weight to both. Third, by matching approximately for lemma frequency at the level of the group means for regulars and irregulars, we attempted to bring this important predictor under some experimental control. For doublets, we only included those irregular verbs in this data set for which the frequency of the irregular past tense form exceeded the frequency of the regular past tense form.

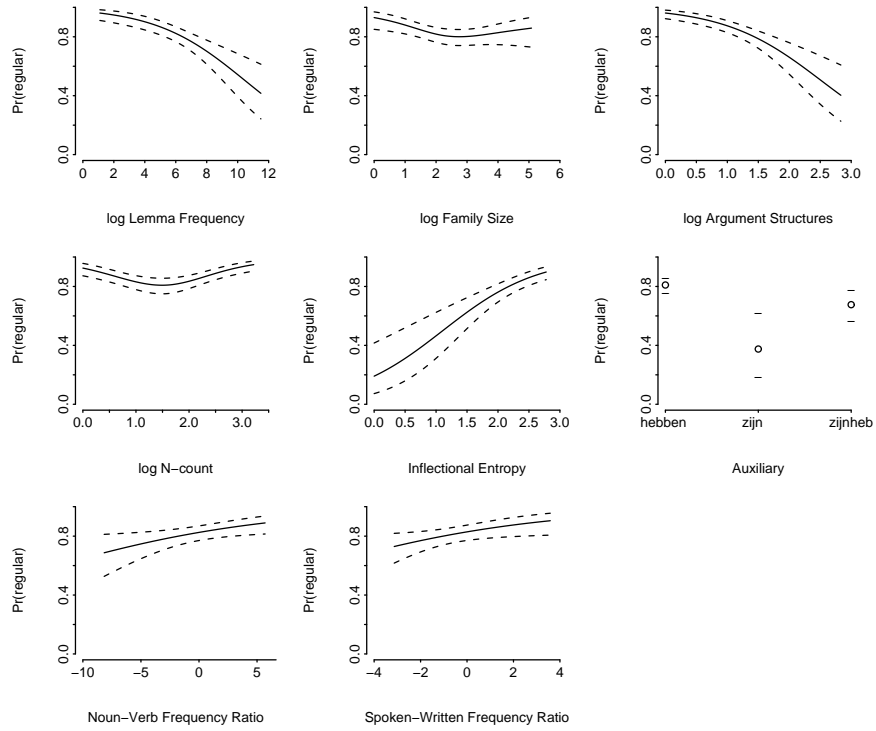


Figure 1: Partial effects of the predictors in logistic regression models for the data set of 677 simple native verbs, with 95% confidence intervals (dashed lines).

Using the etymological dictionary, we estimated historical age on the basis of the cognates listed for other Indo-European languages. In what follows, we will refer to this estimate as 'etymological age'. Verbs that are Dutch innovations were classified as 'Dutch'. Words common to Dutch and German but not attested in other languages were classified as 'Dutch-German'. Words with as oldest cognates words in other West-Germanic languages were classified as 'West-Germanic'. If a verb had cognates in a non-West-Germanic Germanic language, it was labeled as 'Germanic'. Non-Germanic verbs were categorized as 'Indo-European'. There were a few cases where we did not agree with the etymological dictionary. For instance, we did not see any correspondence between the Dutch verb *gieten*, 'to pour', and its supposed root, Latin *fundo*. In such cases, we were conservative and assigned it to the greatest uncontested age (Germanic for *gieten*). Other cases in which we did not agree with the etymological dictionary involved verbs for which we had supplementary knowledge about their occurrence in other Indo-European languages. An example of such a case is the Dutch verb *spoelen*, 'to rinse'. According to our dictionary, this verb would be a Dutch innovation. However, we happened to know that the Swedish verb for '(to) rinse' is (*att*) *spola*, which has nearly the same phonological form. We therefore classified *spoelen* as 'Germanic' instead of 'Dutch'.

In what follows, we first include etymological age as a predictor for whether a verb is regular or irregular. We then reverse the direction of the prediction and ask what factors predict etymological age.

In the stepwise logistic regression analysis for the probability of regularity, we included etymological age as an ordered factor with the levels

Dutch < Dutch-German < West-Germanic < Germanic < Indo-European.

This analysis revealed that lemma frequency, age, neighborhood density, and auxiliary selection were significant predictors for regularity. Figure 2 shows how these predictors affect the probabilities, and Table 4 summarizes the Wald statistics.

The upper left panel shows a convex curve for the partial effect of lemma frequency observed for our selection of 286 verbs. Note that this curve differs from the curve in Figure 1 representing the partial effect of lemma frequency in the larger data set of 677 simplex native verbs. Nevertheless, the two curves are very similar from a log lemma frequency of 8 onwards. The difference arises due to the matching for lemma frequency that we imposed on the experimental data set. There are many low-frequency regulars and

Table 4: Wald statistics for the predictors in the logistic regression model for regularity fit to the data set of 286 simple verbs.

Factor	Chi-Square	d.f.	P
Lemma Frequency	20.62	2	0.0001
Nonlinear	17.52	1	< .0001
Age	11.49	4	0.0216
Neighborhood density	11.24	1	0.0008
Auxiliaries	8.70	2	0.0129
TOTAL	40.07	9	< .0001

few low-frequency irregulars. By imposing the constraint of equal numbers of regulars and irregulars, the sample no longer reflects the probabilities of regulars and irregulars in the population. In other words, the results shown in Figure 2 are conditional on the selection criteria for the experimental data set. The lower panels visualize the effects of neighborhood density and of the auxiliary selected by the verb for the present perfect. The classification by auxiliary results in three classes of verbs that differ significantly with respect to the number of synsets ($F(2, 282) = 7.64, p = 0.0006$). The verbs that select only *hebben* have the lowest mean number of synsets (3.5), the verbs selecting both auxiliaries have the highest mean (5.1), and the verbs selecting only *zijn* have an intermediate of 4.1. If auxiliary is not included as a predictor, the number of synsets is significant with irregulars having more synsets than regulars.

The upper right panel of Figure 2 is the most interesting, as it shows that the likelihood of regularity decreases with increasing etymological age, as expected, independently of lemma frequency, neighborhood density, number of synsets and auxiliary selection.

In order to examine the possibility that etymological age might itself be predictable, we fit a proportional odds model (Harrell 2001) to the experimental data set, with age as dependent variable. The counts of verbs for each of the levels of age were 8 (Dutch), 28 (Dutch-German), 43 (West-Germanic), 174 (Germanic), and 33 (Indo-European). As predictors, we started out with the lemma frequency of the verb, Regularity, the number of neighbors, inflectional entropy, family size, number of synsets, and length in letters. A stepwise model selection procedure revealed that the only relevant predictors were, as expected, Regularity ($\hat{\beta} = -0.86, Z = -3.52, p = 0.0004$), and, surprisingly, neighborhood density ($\hat{\beta} = 0.06, Z = 3.08, p = 0.0021$). The left

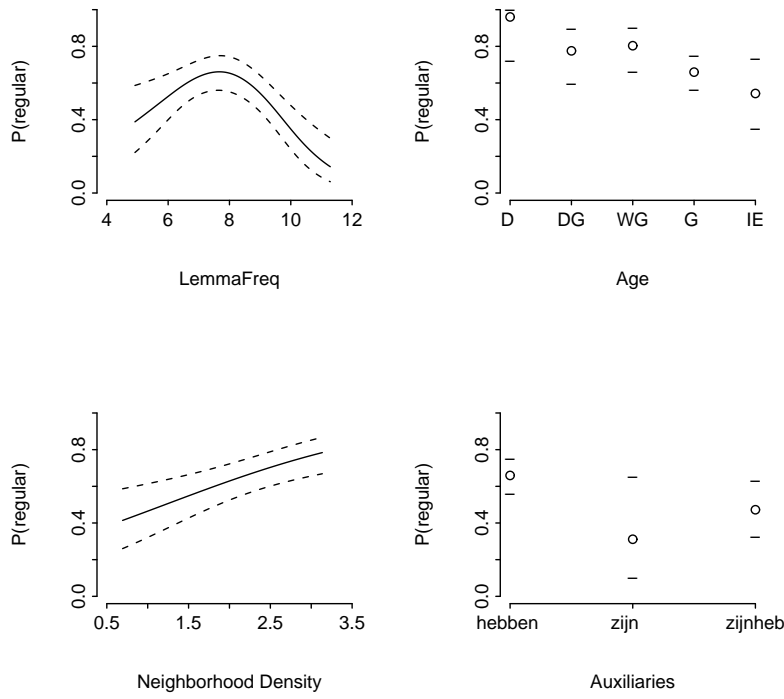


Figure 2: Logistic regression model for the probability of regularity in the experimental data set of 286 verbs with equal numbers of regulars and irregulars, and with approximate between-group matching for lemma frequency.

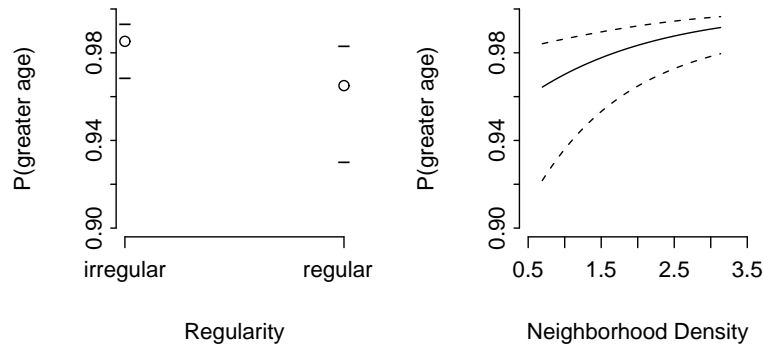


Figure 3: Partial effect with 95% confidence intervals according to a proportional odds model for etymological age as a function of regularity (left) and neighborhood density (right).

panel of Figure 3 shows the effect of regularity: the probability of greater age is smaller for regulars than for irregulars. The right panel shows that the probability of greater age increases with neighborhood density: verbs with a greater similarity neighborhood tend to be older. (The residuals stratified by age for regularity and density were centered around zero, indicating that the presupposition of proportionality of the proportional odds model was met)

Interestingly, lemma frequency is not a significant predictor ($p > 0.5$) in this model. Regular verbs as a group are less frequent than irregular verbs as a group, regular verbs as a group tend to be younger verbs than irregular verbs, and nevertheless, lemma frequency does not predict age. This lack of predictivity is probably the result of changing frequencies for the actions and states denoted by the verbs. In prehistoric times, hunting was a frequent activity, comparable to present-day shopping. At present, hunting is an infrequent sport requiring a license. The Dutch verb for hunting, *jagen*, still bears the hallmark of its age in its irregular past tense form, but its frequency is now relatively low.

In contrast to lemma frequency, neighborhood density is a predictor for

age. The upper trellis graph of Figure 4 illustrates this predictivity. Moving from the lower left panel to the upper right panel, etymological age increases from 'Dutch' to 'Indo-European', and in parallel, the means and ranges for both the regulars and irregulars move upward on the scale of neighborhood density. In these analyses, we calculated the neighborhood density for the singular present tense forms, which are identical to the stems. What we observe, then, is that stems with many neighbors tend to occupy a position in phonological space that is shared with many other words. These stems therefore will tend to be phonologically more regular. Thus, the older verb stems 'live' in phonologically attractive neighborhoods, which may have contributed to their continuing existence. Note that the densities tend to be higher for regulars compared to irregulars, which is in accordance with the lower left panel of Figure 2. Apparently, morphological regularity goes hand in hand with greater phonological regularity.

The lower trellis graph shows the effect of neighborhood density when we calculate the number of neighbors for the singular past tense forms instead of for the stems. Note, first of all, that as we move through etymological time, the density increases, just as in the upper trellis graph. However, it is now the irregulars that have the higher density, instead of the regulars. This is a direct consequence of the presence of the past tense suffix in the regulars, which restricts their neighborhoods to other regular past tense forms. The singular past tense forms of the irregulars are themselves simple phonological words, just as the present tense stems of both regulars and irregulars, and they tend to have neighborhoods that are nearly, but not quite as large ($F(1, 284) = 4.57, p = 0.03$) as those of regular stems. These quantitative observations reinforce the point made by Burzio (2002), namely, that irregular past tense forms are phonologically more regular than regular past tense forms.

This completes our survey of the lexical statistics of regularity in Dutch.

3 A visual lexical decision experiment

Our survey of the lexical statistics of regularity and irregularity for Dutch simplex verbs revealed systematic semantic differences between regular and irregular verbs. The question to be addressed in what follows is whether these differences are reflected in on-line lexical processing. In the present study, we have made use of visual lexical decision as a first step towards addressing this issue.

Materials The materials consisted of the abovementioned selection of 143 irregular and 143 regular native simplex verbs, complemented with 286 pseu-

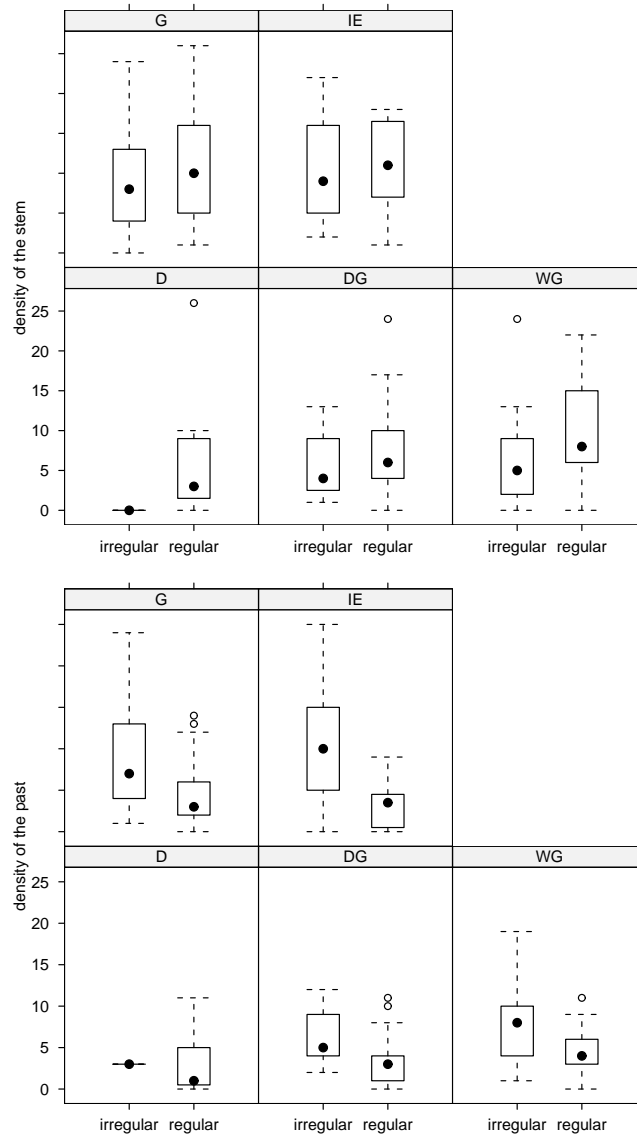


Figure 4: Trellis graphs for the distribution of neighborhood density as a function of regularity and age. The upper panel shows the neighborhood density of the stem, and the lower panel the neighborhood density of the singular past tense form. D: Dutch, DG: Dutch-German, WG: West-Germanic, G: Germanic, IE: Indo-European

dowords that were phonotactically legal and that had the same distribution of first phonemes, of CV-structure, and of the number of syllables as the target words, pairwise matched to the words. For each verb, we selected (or constructed, in the case of the pseudo-verbs) four inflectional variants: the present singular (e.g., *loop*, 'I walk'), the present plural/infinitive (e.g., *lopen*, 'we/you/they/to walk'), the past singular (e.g., *liep*, 'I/you/he walked') and the past plural (e.g., *liepen*, 'we/you/they walked'). We then constructed four master lists, one for each inflectional variant, each containing the items for one sub-experiment. We randomized each master list (with 286 words and 286 pseudowords) ten times, and added twenty practice items to the beginning of each list.

Subjects Twenty women and twenty men, all students at the University of Nijmegen, participated in the experiment. They all had normal or corrected-to-normal vision. Ten women and ten men first participated in the sub-experiment with the present plural forms. One week later, the same subjects performed the sub-experiment with the corresponding past plural forms. Ten other men and ten other women first performed the sub-experiment with the present singular forms, and, a week later, performed the sub-experiment with the past singular forms. Subjects received 5 Euro for each sub-experiment.

Procedure Subjects were tested individually in a quiet room. They received standard lexical decision instructions, specifying that they had to decide as quickly and as accurately as possible whether a presented letter string was a Dutch word or not. If it was a word, they had to push the right one of two response keys, otherwise the left one. For left-handed subjects, the order of the response buttons was reversed.

Each trial consisted of the presentation of a fixation mark (asterisk) in the middle of the screen, followed after 50 ms by the stimulus centered at the same position. Stimuli were presented on Nec Multisync color monitors in white upper-case letters (font: triplex; font size: 36) on a dark background and remained on the screen until a subject pressed one of the two response buttons, or disappeared after a time period of 1500 ms if no response was given. A new trial was initiated 500 ms after responding or time-out. There was a short break following the practice items, and three further short breaks during the experiment itself. The total duration of the experimental session was approximately 35 minutes.

Results We combined the results of the four sub-experiments into one data set, with Subject both as a within-Tense factor (a given subject saw both a present and a past tense form) and as a between-Number factor (a given subject saw either singulars or plurals, but never both).

Table 5: Estimates of the coefficients in the logistic regression model fit to the accuracy data of the visual lexical decision experiment.

	Coef	S.E.	Wald Z	P
Intercept	0.9964	0.29027	3.43	0.0006
Form Frequency	-0.2551	0.03689	-6.91	0.0000
Form Frequency'	-0.4169	0.13185	-3.16	0.0016
Form Frequency''	2.7174	0.59458	4.57	0.0000
Family Size'	-0.3885	0.05516	-7.04	0.0000
Family Size''	0.4575	0.07340	6.23	0.0000
Number of Synsets	0.3282	0.07982	4.11	0.0000
Number of Synsets''	-4.7781	1.06611	-4.48	0.0000
Number of Synsets'''	7.4880	1.70304	4.40	0.0000
Inflectional Entropy (H_i)	-0.5919	0.12000	-4.93	0.0000
Argument structures	-0.1960	0.06461	-3.03	0.0024
Regularity=regular	-1.9228	0.36831	-5.22	0.0000
Tense=present	-0.4816	0.10659	-4.52	0.0000
Number=singular	0.8405	0.07668	10.96	0.0000
Tense=present * Number=singular	-0.7277	0.12407	-5.87	0.0000
Regularity=regular * Tense=present	0.8765	0.11283	7.77	0.0000
Regularity=regular * H_i	0.6075	0.16760	3.62	0.0003

For each trial in the experiment, we recorded whether the response (word or pseudoword, as indicated by the button-presses) was correct. The overall error rate (incorrect responses and time-outs) was 7%. We analyzed the accuracy measure by means of a logistic analysis of covariance (Harrell 2001), with as predictors the factors Regularity (regular versus irregular), Sex (male versus female), Tense (present versus past), Number (singular versus plural), and the covariates form frequency (the string frequency of the inflected form presented in the experiment as listed in CELEX), inflectional entropy (H_i), morphological family size, argument structures, auxiliary, and the number of synsets in WordNet. The distributions of form frequency and morphological family size are highly skewed with long right tails. We therefore applied logarithmic transformations to reduce the potential adverse effects of outliers (see, e.g., (Chatterjee, Hadi, and Price 2000)).

Table 5 summarizes the model resulting from a stepwise selection procedure. For the covariates frequency, family size, and number of synsets, significant non-linearities were present, which we modeled by means of restricted cubic splines. In Table 5, the splines are indicated by primes. To understand the nonlinearities, we present the partial effects of the covariates

in Figure 5. The upper left panel shows that a higher frequency corresponds with a lower probability of an error, with a floor effect for the higher frequencies. The upper middle panel shows a U-shaped curve for the family size measure. Apparently, low and very large family sizes elicited more error responses, while medium sized families led to fewer error responses. As can be seen in the upper right panel, the probability of an error tended to decrease slightly with a larger number of synsets. The left panel on the second row illustrates the effect of inflectional entropy and its interactions with regularity. A high inflectional entropy led to fewer errors, but only for the regular verbs (dashed line). The central panel in Figure 5 shows that verbs participating in many argument structure alternation classes were less error-prone. The right panel on the second row plots the effect of the interaction between regularity and tense on the error rate. The probability of an error decreases for the regular past forms, compared to the regular present forms and the irregular past forms. As can be seen in the first panel on the lowest row the past singular forms elicited more errors than the past plural forms. There is no such effect for the forms in the present tense.

With respect to the factorial contrasts, the coefficients listed in Table 5 indicate that in general, regulars elicited fewer errors than irregulars ($\hat{\beta} = -1.9228$), and especially so in the case of past tense regulars (in which case the coefficient is adjusted to $-1.9228 - 0.8765 = -2.7993$). Furthermore, past tense forms elicited fewer errors ($\hat{\beta} = -0.4816$, especially in the singular (with as coefficient $-0.4816 - 0.7277 = -1.2093$).

For the analysis of the response latencies, we first removed the data points with error responses from the data set. The distribution of the latencies was highly skewed, with outliers in both tails. We reduced the skewness by means of a logarithmic transform and by removing, after visual inspection of the distribution, the individual datapoints with log latencies less than 5.8 and greater than 7.3, which amounts to 7.2% of the (correct) data.

As a first step in the analysis, we considered a simple main effects multi-level model with subject as main grouping factor. In addition to the predictors regularity, tense and number, we also included the status of the preceding item (word versus pseudoword) and the correctness of the preceding trial (correct versus incorrect) as control factors. T-tests on the coefficients revealed significant effects for all factors except regularity, as shown in Table 6. If the preceding item was a word, reaction times were shorter, if the preceding item elicited an error, reaction times were longer. Past inflections elicited longer reaction times than present inflections, and singulars elicited shorter reaction times than plurals.

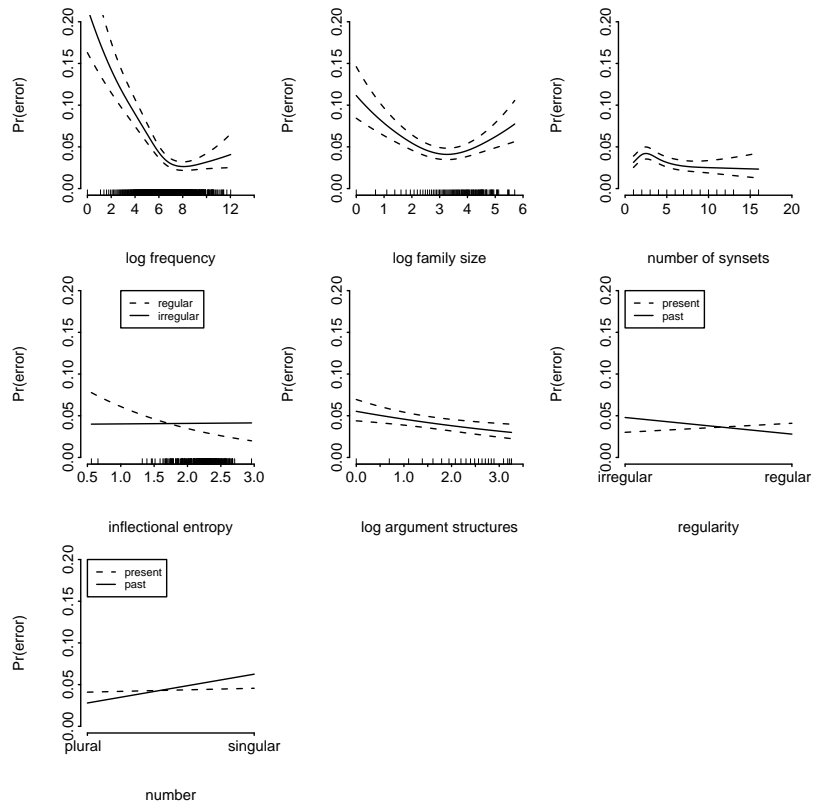


Figure 5: Partial effects of the covariates in the logistic regression model fit to the accuracy data of the visual lexical decision experiment. 95% confidence intervals are shown.

Table 6: Estimated coefficients of the factors in a multilevel model with simple main effects only fit to the response latencies of the visual lexical decision experiment.

	Value	Std.Error	DF	t-value	p-value
(Intercept)	6.4938	0.0254	20308	255.2100	0.0000
Previous=word	-0.0096	0.0029	20308	-3.3413	0.0008
PrevCorrect=incorrect	0.0526	0.0066	20308	7.9759	0.0000
Regularity=regular	-0.0002	0.0029	20308	-0.0869	0.9307
Tense=present	-0.0708	0.0043	20308	-16.5588	0.0000
Number=singular	-0.1274	0.0372	20308	-3.4248	0.0006

The next step in our analysis was to allow interactions into the model, in combination with our covariates. Table 7 lists the significant predictors and their t -values and p -values in the final model obtained with a stepwise variable selection procedure. As in the error analysis, we observed various nonlinearities and interactions. Figure 6 plots the partial effects of the predictors. The upper left panel graphs the facilitatory effect of frequency, which, as in the error analysis, levels off for the higher frequencies. The upper middle panel shows a shallow u-shaped curve, indicating that a facilitatory effect of family size is obtained primarily in the range of the low to medium family sizes. In shape, its effect is similar to that observed in the error analysis. The upper right panel illustrates the interaction of inflectional entropy with regularity. For both regulars (solid line) and irregulars (dashed line), inflectional entropy first increases and then decreases, but for the irregulars, the increase is less marked and the decrease more pronounced. The first panel on the second row reveals a facilitatory effect of the count of argument structures and complementation patterns on the response latencies. The middle panel on the second row plots the linear facilitatory effect of the number of synsets, which is present only for irregular verbs in the past tense. (the three-way interaction of Tense by Regularity by Number of Synsets). It remains noteworthy that the effect of number of synsets is restricted to the irregular verbs. Apparently, it is when an irregular verb is encountered in its irregular form that comprehension is facilitated by a high semantic density.

We also observed interactions of Regularity by Number ($F(1, 20361) = 210.60, p = 0.0001$) and of Regularity by Tense ($F(1, 20291) = 50.53, p < 0.0001$). As can be seen in the third panel on the second row, in which the interactions of regularity with tense and number are plotted, regulars were responded to faster than irregulars and even more so in the past tense and for the plural forms. In the lowest left panel the effect of the auxiliaries are plotted. Verbs with the auxiliary "zijn" elicited the longest responses.

A final issue is the possibility that for regular verbs, form frequency effects are restricted to those verbs that rhyme with irregular verbs, as reported for English by Ullman (2001). We therefore investigated whether there might be a similar dissociation for the regular verbs in our experiment. We therefore constructed a new factor that indicated whether a regular verb had one or more rhyming irregular verbs. In a first analysis, we observed a main effect for this factor ($F(1, 10195) = 3.16, p = 0.0016$) as well as interactions with form frequency ($F(1, 10195) = -2.83, p = 0.0046$) and the linear term of inflectional entropy ($F(1, 10195) = -2.40, p = 0.0162$). Form frequency was significantly facilitatory for all regular verbs, but more so for regular

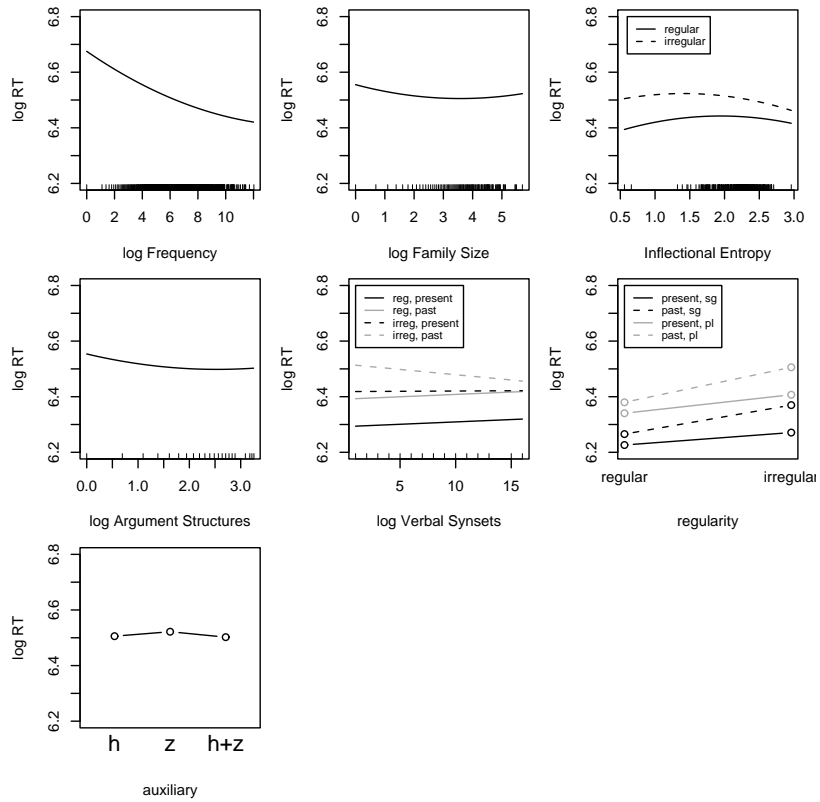


Figure 6: Partial effects of the covariates in the multilevel regression model fit to the response latencies in the visual lexical decision experiment.

Table 7: Estimated coefficients of the predictors in a multilevel model fit to the response latencies of the visual lexical decision experiment.

	Value	t-value	p-value
Form Freq	-0.0344	-9.8198	0.0000
Form Frequency ²	0.0011	3.7692	0.0002
Inflectional Entropy (H_i)	0.0714	2.3362	0.0195
Inflectional Entropy ² (H_i^2)	-0.0253	-3.4130	0.0006
Family Size	-0.0279	-3.9638	0.0001
Family Size ²	0.0039	3.4886	0.0005
Synsets	-0.0038	-3.5102	0.0004
Auxiliary=zijn	0.0161	2.4573	0.0140
Aux=zijn,hebben	-0.0035	-0.9431	0.3457
Argument structures	-0.0437	-5.7277	0.0000
I(Argument Structures ²)	0.0086	3.2773	0.0010
Regularity=regular:Tense=present	0.0594	6.1545	0.0000
Regularity=regular:Number=singular	0.0217	3.8859	0.0001
Regularity=regular:H	0.0269	2.6746	0.0075
Regularity=regular:nSynV	0.0055	3.5713	0.0004
Regularity=irregular:Tense=present:nSynV	0.0040	3.3784	0.0007

verbs with rhyming irregulars. For such verbs the inflectional entropy was less inhibitory. In other words, the behavior of regular verbs with rhyming irregulars, in terms of the inflectional entropy, was more similar to that of the irregular verbs. In a second analysis, we added the lemma frequency and the difference between the lemma frequency and the form frequency as predictors to the model. We used this frequency difference in an attempt to disentangle the overall frequency of the verb and the frequency of a specific inflectional variant without overly increasing the collinearity in the data. Like the effect of form frequency, the lemma frequency was significantly facilitatory for all verbs, and again especially so for the rhyming regulars. The difference between the lemma frequency and the surface frequency was inhibitory and did not interact with whether the verb had rhyming irregulars. This suggests that having a rhyming regular interacts with a verbs overall frequency rather than with the specific frequency of a given inflectional form. What these analyses show is, first of all, that frequency effects are modulated by whether a verb has a rhyming irregular as observed by Ullman. At the same time, these analyses also show that in Dutch verbs without rhyming irregulars are subject to solid frequency and entropy effects, contrary to what Ullman reports for English. Our findings suggest that there are form-based attractor effects that

cut across the sets of regular and irregular verbs (see also ?).

4 General Discussion

In this study, we combined a survey of the distributional properties of regularity for verbs with an experimental study addressing the predictivity of these distributional properties for lexical processing in reading.

Our lexical statistical survey revealed that the regularity of a verb can be predicted (using a logistic regression model) not only from lemma frequency (irregulars tend to be more frequent), but also from family size (regulars tend to have more small and large and less intermediate family sizes), neighborhood density (regulars tend to live in denser similarity neighborhoods), argument structures (regulars tend to have less argument structures than irregulars), auxiliaries (the auxiliary *zijn* was underrepresented for the regulars), inflectional entropy (regulars tend to have greater entropies), the noun-verb frequency ratio (regulars tend to have a higher noun frequency) and the spoken-written frequency ratio (regulars tend to have a higher spoken frequency). All these additional variables are predictive over and above lemma frequency. What this analysis shows is that regularity is part of a conspiracy of factors relating to both form and meaning.

It might be argued that irregular verbs have a greater semantic density than regular verbs because they have been in use in the language community for longer periods of time. In other words, a greater historical age might have allowed the irregulars to develop a greater range of meanings. To investigate this possibility, we estimated historical age by means of an etymological dictionary. A proportional odds regression analysis revealed two predictors for historical age: regularity, and neighborhood density. Irregular verbs tended to be overrepresented for the older words, and the probability of greater age increased with neighborhood density, i.e., with phonological regularity. Interestingly, frequency is not predictive for historical age: Frequency reflects object and concept frequency, and apparently these frequencies have changed substantially through the ages even for simplex verbs. Since the variance explained by the model, $R^2 = 0.072$, is low, and since in addition to frequency, semantic variables such as the number of synsets and morphological family size are not predictive for historical age, we conclude that it is unlikely that historical age is the driving force behind the greater semantic density of irregulars. Phonological regularity, as measured by neighborhood density, is probably more important here. This leads to a functional explanation of the greater semantic density, namely, that the greater entanglement of irregular

verbs with other parts of the grammar makes it easier to remember the irregular forms.

In our visual lexical decision study, we investigated the predictivity of the distributional measures for lexical processing in single word reading, including as factors regularity, number, and tense. The error analysis revealed a decrease in the errors for regulars as compared to irregulars, and the analyses of the response latencies likewise pointed to a processing advantage for regulars. In both analyses, this advantage was most prominent for past-tense forms. This finding challenges strictly cascaded dual route models (e.g., Pinker (1991, 1997)) in which lexical retrieval and rule-based processing are sequentially ordered. Such cascaded theories predict, contrary to fact, that regulars should be more difficult to process than irregulars, because regulars would require decomposition into stem and affix in addition to lexical look-up, and therefore should elicit longer instead of shorter latencies.

As predicted by the semantic density hypothesis, the greater semantic density of irregulars, as measured by means of the number of synsets, led to a processing advantage for irregular verbs in the past tense form. Apparently, the retrieval from lexical memory of the meanings of morphologically complex irregular past tense forms is facilitated by greater semantic connectivity. This finding is in line with previous findings (e.g., Bybee and Slobin (1982); Ramscar (2002); Patterson, Lambon Ralph, Hodges, and McClelland (2001); Joanisse and Seidenberg (1999)) suggesting that a strict separation of form and meaning in the mental lexicon is a simplification of the true complexity of lexical organization in the brain.

For the subset of regular verbs, we observed that frequency effects were stronger for those verbs that rhyme with irregular verbs, as observed for English by Ullman. Contrary for what Ullman reports for English, regular verbs that do not rhyme with any irregular verb also revealed solid facilitatory frequency effects.

Form frequency effects are expected in any theoretical framework for the singular present-tense forms as well as for the singular past-tense forms of irregular verbs. The form frequency effects that we observed for both regular and irregular plurals constitute counterevidence to theories of the mental lexicon (e.g., Pinker (1991, 1997, 1999)) in which storage in lexical memory is restricted to irregular forms. However, the present frequency effects are in line with the frequency effects for regular inflected forms reported for other Dutch inflectional suffixes by Schreuder, De Jong, Krott, and Baayen (1999); Baayen, Schreuder, De Jong, and Krott (2002), see also Taft (1979); Sereno and Jongman (1997) for regular inflection in English and New, Brysbaert,

Segui, and Rastle (2004) for regular inflection in French and English. These frequency effects show that regular inflected words leave traces in the mental lexicon, although it should be kept in mind that this knowledge may not be very helpful for lexical processing in spontaneous language use, where sentential context and pragmatic setting may provide much more accurate probabilistic biases for guiding lexical retrieval.

Further evidence for memory traces for regular and irregular forms in lexical memory is provided by the inflectional entropy measure. In the error analysis, a higher inflectional entropy led to fewer errors for irregular verbs. For the response latencies, a higher inflectional entropy correlated (non-linearly) with shorter latencies, and this effect was more pronounced for the irregulars than for the regulars. The predictivity of inflectional entropy in the present experiment (which is not significantly correlated with either lemma frequency or form frequency, $p > 0.2$ in both cases) provides further support for the hypothesis that the mental lexicon is sensitive to the information load of the inflectional paradigms of individual words.

In multiple regression, it is important to check for potential problems due to collinearity. The collinearity of covariates quantifies the extent to which these covariates are independent, technically distinguishable, predictors. A measure of collinearity is the condition number (Belsley, Kuh, and Welsch 1980; Belsley 1991), which equals 26.0 for the predictors form frequency, family size, inflectional entropy, argument structures, auxiliaries and number of synsets in the analysis of the response latencies. This high a condition number points to substantial, potentially harmful, collinearity. It turns out that this high collinearity is caused by the inflectional entropy. Inflectional entropy enters into low pairwise correlations with the other covariates, but is apparently highly correlated with linear combinations of the other covariates. We therefore carried out two supplementary analyses, one excluding inflectional entropy (in which case the condition number drops from 26.0 to 10.9), and one including inflectional entropy and excluding the other covariates. Both analyses revealed the same pattern of results, which allows us to conclude that the results of the full model are not distorted by collinearity. The same holds for the accuracy data and for the logistic regression model we used to predict regularity.

Ullman, Estabrooke, Steinhauer, Brovotto, Pancheva, Ozawa, Mordecai, and Maki (2002) observed, using several experimental methods including word naming, that form frequency effects were greater for women than for men. In English and Spanish, women showed frequency effects for both regular and irregular inflected forms, while the men in their study revealed fre-

quency effects only for irregulars. In our study, we also observed some hints of an effect of sex. In one error analysis, women turned out to be more sensitive to inflectional entropy. For them, a small entropy led to more errors, and a high entropy to fewer errors, compared to men. This would be in line with the greater verbal memory capacity of women (Kimura 2000). Females also responded more quickly than males, but only for plurals, which would suggest that females have an advantage not only with respect to verbal memory (as witnessed by the sex by entropy interaction), but also with respect to the efficiency of morphological decomposition. However, these effects of the sex of the participant lost significance when interactions of subject by number and subject by frequency were allowed into the model. Apparently, the individual differences between subjects are much larger than this group difference. Therefore, our data provide no support for the hypothesis of Ullman, Estabrooke, Steinhauer, Brovotto, Pancheva, Ozawa, Mordecai, and Maki (2002) that women would show stronger form frequency effects than men.

The general picture that emerged from this study is that the distinction between regular and irregular verbs is not a simple one. Regulars and irregulars differ not only with respect to their formal properties, but also with respect to their semantic properties and the information structure of their inflectional paradigms. These subtle distributional differences affect lexical processing. Both our distributional measures and the task that we have used are simple and crude, but the results obtained invite further study of the fascinating and enigmatic phenomenon of regularity and irregularity in the mental lexicon.

References

- Baayen, R. Harald
2005 Data mining at the intersection of psychology and linguistics. In Anne Cutler, (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*, p. Erlbaum, Hillsdale, N.J.
- Baayen, R. Harald and Fermin Moscoso del Prado Mart'ın
2005 Semantic density and past-tense formation in three germanic languages. *Language*.
- Baayen, R. Harald, Richard Piepenbrock, and Leon Gulikers
1995 *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

- Baayen, R. Harald, Rob Schreuder, Nivja H. De Jong, and Andrea Krott
2002 Dutch inflection: the rules that prove the exception. In S. Nootboom, F. Weerman, and F. Wijnen, (eds.), *Storage and Computation in the Language Faculty*, pp. 61–92. Kluwer Academic Publishers, Dordrecht.
- Balota, David A., Michael J. Cortese, and Maura Pilotti
1999 Visual lexical decision latencies for 2906 words. [On-line], Available: http://www.artsci.wustl.edu/~dbalota/lexical_decision.html.
- Belsley, David A.
1991 *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. Wiley, New York.
- Belsley, David A., Edwin Kuh, and Roy E. Welsch
1980 *Regression Diagnostics. Identifying Influential Data and sources of Collinearity*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.
- Burzio, Luigi
2002 Missing players: Phonology and the past-tense debate. *Lingua*, 112: 157–199.
- Bybee, Joan L. and Dan I. Slobin
1982 Rules and schemas in the development and use of the english past tense. *Language*, 58: 265–289.
- Chatterjee, Samprit, Ali S. Hadi, and Bertram Price
2000 *Regression analysis by example*. John Wiley & Sons, New York.
- Coltheart, Max, Eileen Davelaar, Jon T. Jonasson, and Derek Besner
1977 Access to the internal lexicon. In S. Dornick, (ed.), *Attention and performance*, volume VI, pp. 535–556. Erlbaum, Hillsdale, New Jersey.
- De Vries, J. and F. De Tollenaere
1991 *Etymologisch woordenboek*. Het Spectrum, Utrecht.
- Harrell, Frank E.
2001 *Regression modeling strategies*. Springer, Berlin.
- Jaeger, Jeri J., Allan H. Lockwood, David L. Kemmerer, Robert D. Van Valin, and Brian W. Murphy
1996 A positron emission tomographic study of regular and irregular verb morphology in English. *Language*, 72: 451–497.

- Joanisse, Mark F. and Mark S. Seidenberg
1999 Impairments in verb morphology after brain injury: a connectionist model. *Proceedings of the National Academy of Sciences*, 96: 7592–7597.
- Kim, John J., Steven Pinker, Alan Prince, and Sandeep Prasada
1991 Why no mere mortal has ever fown out to central field. *Cognitive Science*, 15: 173–218.
- Kimura, Doreen
2000 *Sex and cognition*. The MIT press, Cambridge, MA.
- Kostić, Aleksander, Tanja Marković, and Aleksander Baucal
2003 Inflectional morphology and word meaning: orthogonal or co-implicative domains? In R. Harald Baayen and Robert Schreuder, (eds.), *Morphological structure in language processing*, pp. 1–44. Mouton de Gruyter, Berlin.
- Levin, Beth
1993 *English Verb Classes and Alternations. A preliminary Investigation*. The University of Chicago Press, Chicago.
- Lieber, Rochelle and R. Harald Baayen
1997 A semantic principle for auxiliary selection in Dutch. *Natural Language and Linguistic Theory*, 15: 789–845.
- Moscoso del Prado Martínez, Fermin, Aleksander Kostić, and R. Harald Baayen
2004 Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94: 1–18.
- Nelson, Douglas L., Cathy L. McEvoy, and Thomas A. Schreiber
1998 The University of South Florida word association, rhyme, and word fragment norms. [On-line], Available: <http://www.usf.eduFreeAssociation>.
- New, Boris, Marc Brysbaert, Juan Segui, Ludovic Ferrand, and Kathy Rastle
2004 The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, 51: 568–585.
- Patterson, Karalyn, Matthew A. Lambon Ralph, John R. Hodges, and James L. McClelland
2001 Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia*, 39: 709–724.
- Pinker, Steven
1991 Rules of language. *Science*, 153: 530–535.
1997 Words and rules in the human brain. *Nature*, 387: 547–548.

- 1999 *Words and Rules: The Ingredients of Language*. Weidenfeld and Nicolson, London.
- Pinker, Steven and Alan Prince
1988 On language and connectionism. *Cognition*, 28: 73–193.
- Pinker, Steven and Michael Ullman
2002 The past and future of the past tense. *Trends in the Cognitive Sciences*, 6(11): 456–462.
- Ramscar, Michael
2002 The role of meaning in inflection: Why the past tense doesn't require a rule. *Cognitive Psychology*, 45: 45–94.
- Randall, Janet, Angeliek van Hout, Jurgen Weissenborn, and R. Harald Baayen
2003 Acquiring unaccusativity: a cross-linguistic look. In Artemis Alexiadou, Elena Anagnostopoulou, and Martin Everaert, (eds.), *The unaccusativity puzzle*, pp. 332–353. Oxford University Press, Oxford.
- Schreuder, Rob and R. Harald Baayen
1997 How complex simplex words can be. *Journal of Memory and Language*, 37: 118–139.
- Schreuder, Rob, Nivja H. De Jong, Andrea Krott, and R. Harald Baayen
1999 Rules and rote: beyond the linguistic either-or fallacy. *Behavioral and Brain Sciences*, 22: 1038–1039.
- Sereno, Joan and Allard Jongman
1997 Processing of English inflectional morphology. *Memory and Cognition*, 25: 425–437.
- Spieler, Daniel H. and David A. Balota
1998 Naming latencies for 2820 words. [On-line], Available: <http://www.artsci.wustl.edu/~dbalota/naming.html>.
- Taft, Marcus
1979 Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7: 263–272.
- Ullman, Michael
2001 The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, 30: 37–69.
- Ullman, Michael T., Ivy V. Estabrooke, Karsten Steinhauer, Claudia Brovotto, Roumyana Pancheva, Kaori Ozawa, Kristen Mordecai, and Pauline Maki
2002 Sex differences in the neurocognition of language. *Brain and Language*, 83: 141–143.

Vossen, Piek, Laura Bloksma, and Paul Boersma
1999 *The Dutch WordNet (CD-ROM)*. European Language Resources Association (ELRA), Luxembourg.