Semantics and (Ir)regular Inflection in Morphological Processing

ISBN: 978-90-76203-35-5
Cover illustration: "Mama aan het werk" by Thijmen Soer
Cover design: GVO drukkers en vormgevers | Ponsen & Looijen, Ede
Printed and bound by GVO drukkers en vormgevers | Ponsen & Looijen, Ede

Semantics and (Ir)regular Inflection in Morphological Processing

een wetenschappelijke proeve
op het gebied van de Letteren

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen
op dinsdag 12 oktober 2010
om 13.30 uur precies

door
Wieke Marije Tabak
geboren op 29 juni 1979 te Emmen

Promotores: Prof. dr. R. Schreuder
Prof. dr. R. H. Baayen (University of Alberta)

Manuscriptcommissie: Prof. dr. R. W. N. M. van Hout
Prof. dr. U. Frauenfelder (University of Geneva)
Dr. M. T. C. Ernestus

Felix, qui potuit rerum cognoscere causas
(Happy is he who understands the causes of things)
Vergilius, Georgiche, II, 489

Contents

Introduction

Chapter 1

Words are stored in the mental lexicon. How words are stored and how they are retrieved from memory are central questions in the field of psycholinguistics. This thesis is concerned with the storage and retrieval of regular and irregular verbs. The dominant view in psycholinguistics is that the form differences between regular and irregular verbs are not systematically related to any semantic difference. Consider, for example, the following verbs in Dutch:

present	past	English
slijpen	*slepen*	to sharpen
grijpen	*grepen*	to grab
kijken	*keken*	to look
wijzen	*wezen*	to point
wijten	*weten*	to blame
mijden	*meden*	to avoid

All these irregular verbs share the same stem vowel *ij* in the present tense and the same stem vowel *e* in the past tense. However, there is no link between form elements and meaning.

Now consider the next set of verbs:

present	past	English
steken	*staken*	to sting
prikken	*prikten*	to prick
porren	*porden*	to jab

These verbs do share similarities in meaning, they are nearly synonyms, but they do not have correspondences in their past tense form: *steken* is an irregular verb while *prikken* and *porren* are regular verbs with different past tense suffixes.

Observations like this led generative linguists to claim that form and meaning are strictly separated in our grammar.

Another general assumption in generative linguistics, that seems to follow the above mentioned assumption, is the modular treatment of regular and irregular forms. In this view, regular forms are created by rules and irregular forms are retrieved from the mental lexicon. Pinker (1991, 1997, 1999) popularized this generative model of morphological processing. He claims that regular and irregular past tense forms are neuro-anatomically localised differently in the brain. Pinker's hypothesis is supported by neurocognitive evidence from e.g., Jaeger, Lockwood, Kemmerer, Van Valin and Murphy (1996) and Beretta, Campbell, Carr, Huang, Schmitt, Christianson and Cao (2003) These studies report differences in activation and localisation for regular and irregular past tense forms.

In 2002, Pinker and Ullman relax the claim that storage in the mental lexicon would be restricted to irregulars, stating that the words and rules model 'does not posit that regular forms are never stored, only that they do not have to be'. They argue that, while irregular verbs always show effects of their frequency, frequency effects would vanish for regular verbs with lower-frequency. Alegre and Gordon (1999) proposed a frequency threshold of 6 per million for English below which inflectional frequency effects would be absent.

However, both assumptions (a strict separation between form and meaning, irregular verbs being retrieved form memory and regulars verbs being created by rules) do not seem to be empirically justifiable. Psycholinguistic research has shown that storage in the mental lexicon is not exclusive for irregular forms (see, e.g. Taft, 1979 and Baayen, Schreuder, De Jong and Krott, 2002). The idea that regularity and irregularity are directly mapped on rule-governed and memory-based processes in the brain is perhaps too simplistic.

Furthermore, a large-scale regression analysis of the English Lexicon Project (Baayen, Wurm and Aycock, 2007) revealed frequency effects all the way down, with attenuation of the effect for higher-frequency words instead of lower-frequency words. This indicates that it is lack of statistical power that is at issue, not the possibility that lower-frequency words would leave no traces in lexical memory.

One of the aims of this dissertation is to invalidate the assumption that meaning does not play a role in regularity and irregularity. There are several reasons for this. In English, irregular plural nouns, for example, are structured by semantics, as the following examples show:

(a) man-men , woman-women , child-children
(b) die-dice , penny-pence
(c) ox-oxen , goose-geese , louce-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), pikie(s), herring(s), flounder(s)

Notice that hunted and fished animals either do not have a plural suffix (as in d) or do not have to bear a plural suffix (as in e; see also Quirk, Greenbaum, Leech and Svartvik, 1985). This example shows that there is a relation between form and meaning in the formation of plural suffixes: the suffix type (or the absence of a plural suffix) is determined by the noun type.

Bertram, Laine and Karvinen (1999) and Bertram, Schreuder and Baayen (2000) investigated the role of properties of derivational suffixes in lexical acces to words. They showed that, for respectively Finnish and Dutch, properties like suffix productivity and suffix homonymy (suffix ambiguity in serving more than one semantic function) affected processing: words with productive and non-homonymous suffixes were more likely to induce morpheme-based processing (see also Hagiwara, Sugioka, Ito, Kawamura and Shiota, 1999, for Japanese).

The fact that semantics play a role in past tense formation has already been shown by Bybee and Slobin (1982). They observed that high-frequency irregular past tense forms could serve as attractors for the past-tense formation of semantically related regulars. Ramscar (2002) showed how a semantic interpretation, that was made in context, strongly influences how a pseudo-word is inflected. If context leads one into the interpretation of *to drink*, a pseudo-word like *prink* is likely to be inflected in the past tense as *prank*. On the other hand, speakers favor the regular form *prinked* if the context leads to the interpretation of *to blink*. In other words, the productivity of irregularity is partly determined by semantics.

Baayen and Moscoso del Prado Martín (2005) concluded that regular and irregular verbs differ on a wide range of semantical dimensions and that these differences are reflected in behavioral measures like response latencies in visual lexical decision and word naming studies. They found that, when matched for frequency, irregular verbs tend to have more meanings and more synonyms than regular verbs. Regular and irregular verbs also tend to have different aspectual properties, as is witnessed by the non-uniform distribution of auxiliary verbs in

Dutch and German. Regulars favor *hebben*, ’have’, while irregulars favor *zijn*, ’be’, the auxiliary marking telicity. Regular and irregular verbs are also non-uniformly distributed over Levin’s verb argument alternation classes (Levin, 1993). Another dimension on which regular and irregular verbs differ is the information complexity of a verb’s paradigm, as estimated by the inflectional entropy measure in Dutch (Moscoso del Prado Martín, Kostic and Baayen, 2004). Irregular verbs tend to have somewhat reduced inflectional entropies compared to regulars.

If inflection is indeed dependent on semantics, this is problematic for the dual mechanism model, proposed by Pinker and colleagues, which is purely form-based and independent of semantics. Instead, we would have to turn to approaches in which inflected forms exist side by side and are accessed directly from the mental lexicon. In psycholinguistics, Stemberger (Stemberger, 2002, 2004; Stemberger and Middleton, 2003) advocate this approach. In their research on speech errors, they found evidence for a selection process between different inflectional variants. Another approach, that is very much in line with the approach of Stemberger and colleagues, is word and paradigm morphology (see e.g. Matthews, 1974; Aronoff, 1994; Beard 1995 and Blevins, 2003).

Word and Paradigm morphology focuses on the paradigmatic organization of inflected forms and takes the word itself, whether it is simplex or complex, to be the basic meaningful unit of a language. Word and Paradigm morphology is based on two central notions. The first notion is that lexical knowledge is ultimately exemplar-based. The second notion is that paradigmatic relations between the exemplars in the mental lexicon reflect morphological structure. On the basis of these two notions, one would expect that all words are stored in the mental lexicon, irrespective of whether they are simplex or complex, and irrespective of whether they are regular or irregular.

Aims and outline of the dissertation

The main aim of this thesis is to come to a better understanding of how regular and irregular verb forms are produced and comprehended and how semantics interacts with (ir)regularity. The guiding hypothesis in this investigation is that there is no strict separation between regular and irregular verbs and that inflection is dependent on semantics. We expect this to be reflected in processing.

The research described in this dissertation is different from previous research in several ways. First of all, it focuses mainly on Dutch, whereas most studies on regularity and semantics focus on English. Dutch is an interesting language in this topic of research because it is morphologically richer than English, in terms of its lexical properties (cf., Moscoso del Prado Martin, 2003). In chapter 5 the focus lies on Spanish, a language that is, in turn, morphologically richer than Dutch. Furthermore, we made use of a range of different experimental paradigms and conducted a lexical statistical research on corpus data as well. This allows us to thoughroughly investigate the differences between regular and irregular verbs and how semantics interacts with (ir)regularity in different modalities of the language.

This thesis is organized as follows:

Chapter 2 presents the data of a lexical statistical study and a visual lexical decision experiment. In this chapter I want to find out how different distributional properties can predict a verb's regularity as well as its historical age. The focus lies on the question which (semantic) predictors can differentiate between regular and irregular verbs and how the influence of these predictors is reflected in online lexical processing.

Chapter 3 investigates the consequences of the different distributional properties of irregular and regular verbs for lexical processing further. In this chapter the focus shifts from visual recognition to the production of verb forms. First, a series of simple word naming studies are presented and second, a series of cross-tense word naming studies. In this chapter, the question is addressed whether regular inflected words in languages such as Dutch and English are accessed indirectly through their stems in a sparse lexicon or whether they are available directly in a redundant lexicon and retrieved from memory in normal lexical processing. Word naming experiments tend to be less sensitive to predictors related to semantics. Participants are provided with written language input and do not necessarily need to activate any semantic information concerning this input in order to be able to pronounce the right output. Therefore, this chapter is less likely to give an anwer to the second part of our research question: whether regular and irregular verbs are influenced by semantics.

Chapter 4 is concerned with a series of picture naming experiments. This chapter concentrates on the production of verb forms in the absence of linguistic input. The picture naming task engages the full production process, from conceptualization to articulation. The experiments in this chapter elicit participant responses that are closer to normal conceptualization in everyday speech than the experiments in chapter 2 and 3 do. In picture naming experiments, the production of verbs is more likely to be semantically driven than in the word naming experiments. Therefore, this chapter can shed light on the debate about the relation between regularity on the one hand and memory and semantics on the other.

Chapter 5 shifts the focus to the production and comprehension in another language. In this chapter I present a picture naming experiment and a visual lexical decision experiment in Spanish, a language that uses a different lexicalization pattern to express verbal meaning than Dutch. The main question in this chapter is whether native speakers of Spanish are sensitive to certain semantic properties of a verb, even if these properties are not lexicalized in Spanish, and whether this would have an effect on processing.

Chapter 6 summarizes the research presented in the previous chapters and concludes with a discussion about the implications of the findings of this thesis on the debate about regularity and memory. Furthermore, it outlines topics for further investigation.

References

Alegre, M. and Gordon, P.: 1999, Frequency effects and the representational status of regular inflections, *Journal of Memory and Language* **40**, 41–61.

Aronoff, M.: 1994, *Morphology by itself: stems and inflectional classes*, The MIT Press, Cambridge, Mass.

Baayen, R. H. and Moscoso del Prado Martín, F.: 2005, Semantic density and past-tense formation in three germanic languages, *Language* **81**, 666–698.

Baayen, R. H., Schreuder, R., De Jong, N. H. and Krott, A.: 2002, Dutch inflection: the rules that prove the exception, *in* S. Nooteboom, F. Weerman and F. Wijnen (eds), *Storage and Computation in the Language Faculty*, Kluwer Academic Publishers, Dordrecht, pp. 61–92.

Baayen, R. H., Wurm, L. H. and Aycock, J.: 2007, Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities, *Mental Lexicon* **2**, 419–463.

Beard, R.: 1995, *Lexeme-morpheme base morphology: a general theory of inflection and word formation*, State University of New York Press, Albany, NY.

Beretta, A., Campbell, C., Carr, T., Huang, J., Schmitt, L. M., Christianson, K. and 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., Laine, M. and Karvinen, K.: 1997, Slow, average and fast recognition of morphologically complex words: evidence from Finnish, *Submitted*.

Bertram, R., Schreuder, R. and Baayen, R. H.: 2000, The balance of storage and computation in morphological processing: the role of word formation type, affixal homonymy, and productivity, *Journal of Experimental Psychology: Learning, Memory and Cognition* **26**, 489–511.

Blevins, J. P.: 2003, Stems and paradigms, *Language* **79**, 737–767.

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

Hagiwara, H., Sugioka, Y., Ito, T., Kawamura, M. and Shiota, J.: 1999, Neurolinguistic evidence for rule-based nominal suffixation, *Language* **75**, 739–763.

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

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

Matthews, P. H.: 1974, *Morphology. An Introduction to the Theory of Word Structure*, Cambridge University Press, London.

Moscoso del Prado Martín, F., Kostić, A. and Baayen, R. H.: 2004, Putting the bits together: An information theoretical perspective on morphological processing, *Cognition* **94**, 1–18.

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

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. and Ullman, M.: 2002, The past and future of the past tense, *Trends in the Cognitive Sciences* **6**(11), 456–462.

Quirk, R., Greenbaum, S., Leech, G. and 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.

Stemberger, J. and Middleton, C.: 2003, Vowel dominance and morphological processing, *Language and Cognitive Processes* **18**(4), 369–404.

Stemberger, J. P.: 2002, Overtensing and the effect of regularity, *Cognitive Science* **26**, 737–766.

Stemberger, J. P.: 2004, Phonological priming and irregular past, *Journal of Memory and Language* **50**, 82–95.

Taft, M.: 1979, Recognition of affixed words and the word frequency effect, *Memory and Cognition* **7**, 263–272.

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

Chapter 2

This chapter has been published as a book chapter: Wieke Tabak, Robert Schreuder and R. Harald Baayen (2005). In: S. Kepser and M. Reis (eds.), *Linguistic Evidence – Emperical, Theoretical, and Computational Perspectives*. Mouton de Gruyter, Berlin, pp. 529-555

Abstract

This study provides an analysis of the lexical statistics of regularity for Dutch verbs, combined with an experimental study of the consequences of the different distributional properties of irregular and regular verbs for lexical processing. We show that it is possible to predict whether a verb is (ir)regular from not only frequency, but also neighborhood density, inflectional entropy, family size, and number of synsets. Analyses of covariance revealed that these measures were predictive for the error rate and the response latencies in visual lexical decision. Notably, the greater number of synsets characterizing irregulars led to shorter response latencies for irregular past plurals, in line with the semantic density hypothesis of Baayen and Moscoso del Prado Martín (2005). Furthermore, a higher information complexity as measured by inflectional entropy led to shorter response latencies, especially so for irregular verbs. Compared to men, women turned out to be more sensitive to inflectional entropy, and they responded faster to plurals. This suggests that women not only have superior verbal memory (Kimura, 1999), but are also more proficient in morphological parsing.

Introduction

A series of studies (Pinker, 1991; Pinker, 1997; Kim, Pinker, Prince and Prasada, 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 favor *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 Levins argument structure alternation classes (Levin, 1993). 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) and 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 2 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, Brovetto, Pancheva, Ozawa, Mordecai and Maki (2002) reported sex-related processing differences for regulars and irregulars.

Lexical statistics of Dutch simplex verbs

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

Table 2.1: 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*)

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 (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 2.1. 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

Table 2.2: Examples of argument structures and complementation patterns for the Ducth 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*)

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 of argument structures and complementation patterns in a dictionary derived data resource compiled at the Max Planck Institute for Psycholinguistics, Nijmegen. Table 2.2 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 Kostic, Markovic and Baucal (2003), Moscoso, Kostic 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 *et al*, 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, 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 *et al* (1991),

Table 2.3: 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

Pinker (1999) and 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 emperical 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, 2004).

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 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 2.3 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*.

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 2.3 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 2.1 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 2.1 $X^2_{(1)} = 9.76, p = 0.0018$), its auxiliaries (last panel on the second row, $X^2_{(1)} = 17.67, p = 0.0001$), its noun-verb frequency ratio (lower left panel, $X^2_{(1)} = 7.73, p = 0.0054$), and its spoken-written frequency ratio (lower middle panel, $X^2_{(1)} = 4.05, p = 0.0441$). 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.

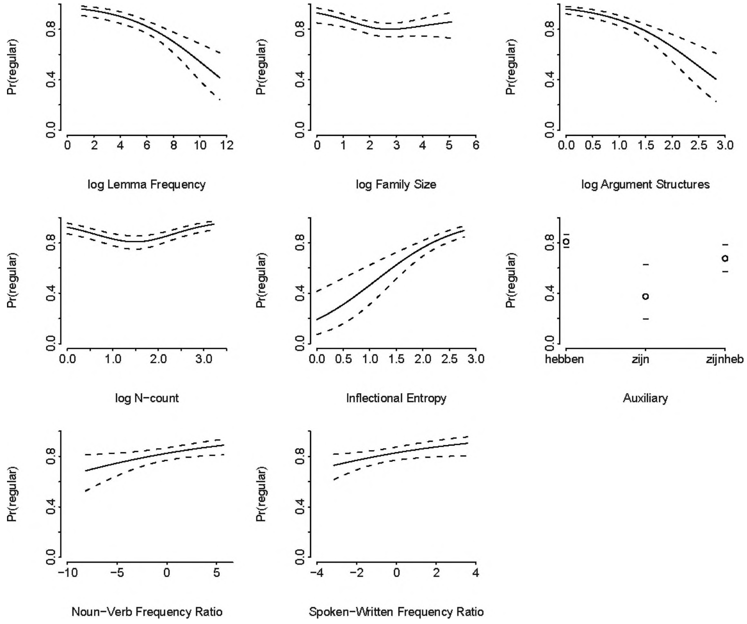

Figure 2.1: Partial effects of the predictors in logistic regression models for the data set of 677 simple native verbs, with 95% confidence intervals (dashed lines).

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 (Vries, 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.

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 2.2 shows how these predictors affect the probabilities, and Table 2.4 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 2.1 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 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 2.2 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 2.2 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

Table 2.4: 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	$< .0001$

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 panel of Figure 2.3 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.

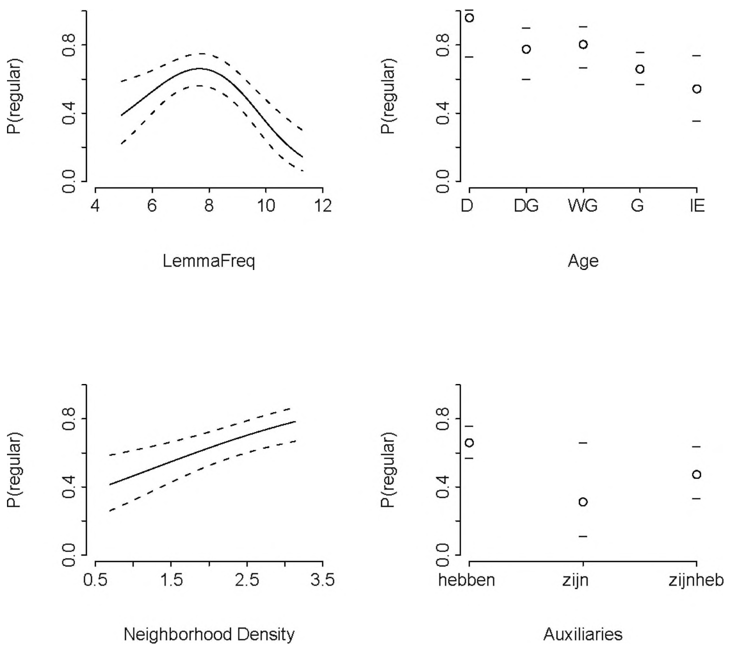

Figure 2.2: Logistic regression model for the probability of regularity in the experimental data set of 286 verbs with equal numbers of regulars and irregulars, and with approximate between-group matching for lemma frequency.

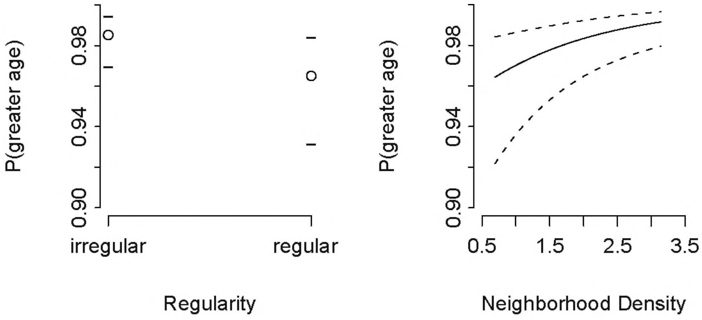

Figure 2.3: Partial effect with 95% confidence intervals according to a proportional odds model for etymological age as a function of regularity (left) and neighborhood density (right).

In contrast to lemma frequency, neighborhood density is a predictor for age. The upper trellis graph of Figure 2.4 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 2.2. 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.

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.

Method

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

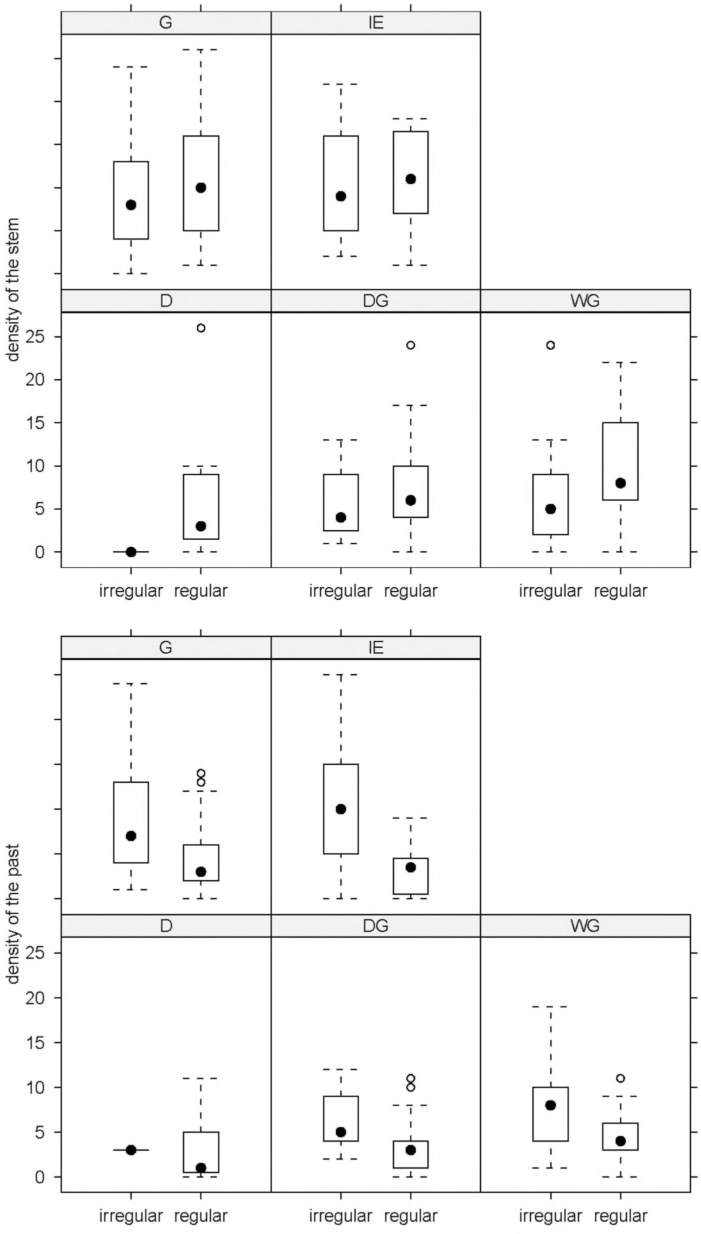

Figure 2.4: Trellis graphs for the distribution of neighborhood density as a function of regularity and age. The upper panel shows the neighborhood density of the stem, and the lower panel the neighborhood density of the singular past tense form. D: Dutch, DG: Dutch-German, WG: West-Germanic, G: Germanic, IE: Indo-European

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).

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.

Table 2.5: 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

We therefore applied logarithmic transformations to reduce the potential adverse effects of outliers (see, e.g., Chatterjee, Hadi, Price, 2000).

Table 2.5 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 2.5, the splines are indicated by primes. To understand the nonlinearities, we present the partial effects of the covariates in Figure 2.5. 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

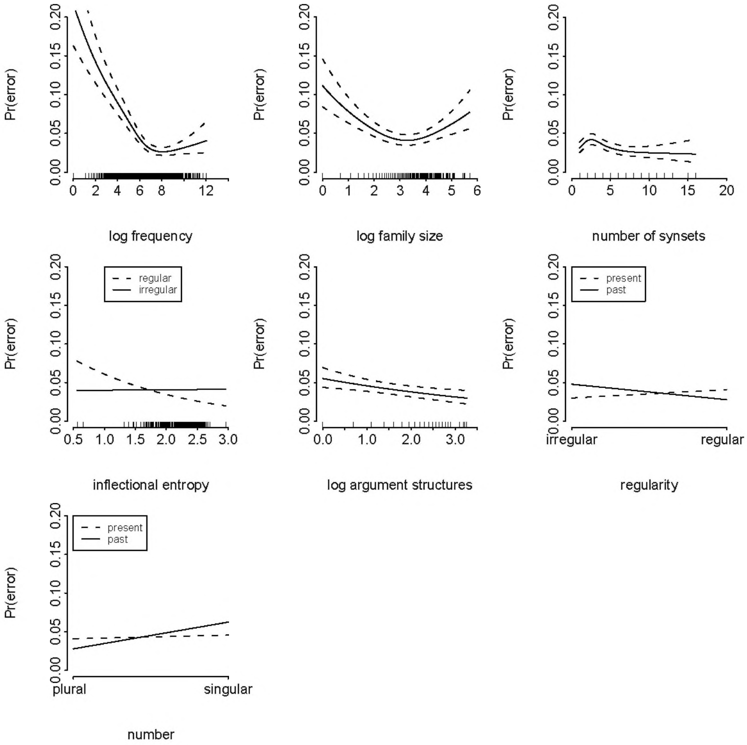

Figure 2.5: Partial effects of the covariates in the logistic regression model fit to the accuracy data of the visual lexical decision experiment. 95% confidence intervals are shown.

verbs (dashed line). The central panel in Figure 2.5 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 been 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 2.5 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 multilevel 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 2.6. 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.

The next step in our analysis was to allow interactions into the model, in combination with our covariates. Table 2.7 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 non-linearities and interactions. Figure 2.6 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

Table 2.6: 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

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

Table 2.7: 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

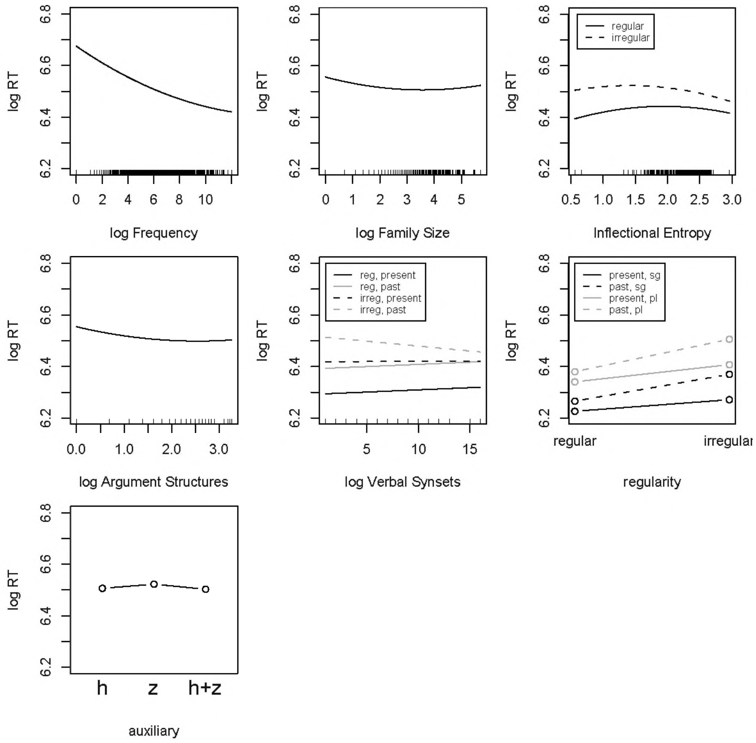

Figure 2.6: Partial effects of the covariates in the multilevel regression model fit to the response latencies in the visual lexical decision experiment.

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 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 Albright and Hayes, 2003).

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 and Pinker, 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; Pinker, 1997; Pinker, 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) and Baayen, Schreuder, De Jong and Krott (2002), see also Taft (1979) and Sereno and Jongman (1997) for regular inflection in English and New, Brysbaert, Segui, Ferrand 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 and 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 *et al.* (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 frequency 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, 1999). 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 *et al.* (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

Albright, A. and Hayes, B.: 2003, Rules vs. analogy in English past tenses: A computational/experimental study, *Cognition* **90**, 119–161.

Baayen, R.: 2005, Data mining at the intersection of psychology and linguistics, *in* A. Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*, Erlbaum, Hillsdale, N.J., pp. 69–83.

Baayen, R. H. and Moscoso del Prado Martín, F.: 2005, Semantic density and past-tense formation in three germanic languages, *Language* **81**, 666–698.

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

Baayen, R. H., Schreuder, R., De Jong, N. H. and Krott, A.: 2002, Dutch inflection: the rules that prove the exception, *in* S. Nooteboom, F. Weerman and F. Wijnen (eds), *Storage and Computation in the Language Faculty*, Kluwer Academic Publishers, Dordrecht, pp. 61–92.

Balota, D., Cortese, M. and Pilotti, M.: 1999, Visual lexical decision latencies for 2906 words, *[On-line], Available: http://www.artsci.wustl.edu/∼dbalota/lexical_decision.html*.

Belsley, D. A.: 1991, *Conditioning Diagnostics: Collinearity and Weak Data in Regression*, Wiley, New York.

Belsley, D. A., Kuh, E. and Welsch, R. E.: 1980, *Regression Diagnostics. Identifying Influential Data and sources of Collinearity*, Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Burzio, L.: 2002, Missing players: Phonology and the past-tense debate, *Lingua* **112**, 157–199.

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

Chatterjee, S., Hadi, A. and Price, B.: 2000, *Regression analysis by example*, John Wiley & Sons, New York.

Coltheart, M., Davelaar, E., Jonasson, J. T. and Besner, D.: 1977, Access to the internal lexicon, *in* S. Dornick (ed.), *Attention and performance*, Vol. VI, Erlbaum, Hillsdale, New Jersey, pp. 535–556.

De Vries, J. and De Tollenaere, F.: 1991, *Etymologisch woordenboek*, Het Spectrum, Utrecht.

Harrell, F.: 2001, *Regression modeling strategies*, Springer, Berlin.

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

Joanisse, M. F. and Seidenberg, M. S.: 1999, Impairments in verb morphology after brain injury: a connectionist model, *Proceedings of the National Academy of Sciences* **96**, 7592–7597.

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

Kimura, D.: 2000, *Sex and cognition*, The MIT press, Cambridge, MA.

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

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

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

Moscoso del Prado Martín, F., Kostić, A. and Baayen, R. H.: 2004, Putting the bits together: An information theoretical perspective on morphological processing, *Cognition* **94**, 1–18.

Nelson, D. L., McEvoy, C. L. and Schreiber, T. A.: 1998, The University of South Florida word association, rhyme, and word fragment norms, *[On-line], Available: http://www.usf.eduFreeAssociation*.

New, B., Brysbaert, M., Segui, F. L. and Rastle, K.: 2004, The processing of singular and plural nouns in French and English, *Journal of Memory and Language* **51**, 568–585.

Patterson, K., Lambon Ralph, M. A., Hodges, J. R. and McClelland, J. L.: 2001, Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge, *Neuropsycologia* **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. and Prince, A.: 1988, On language and connectionism, *Cognition* **28**, 73–193.

Pinker, S. and Ullman, M.: 2002, The past and future of the past tense, *Trends in the Cognitive Sciences* **6**(11), 456–462.

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. and Baayen, R. H.: 2003, Acquiring unaccusativity: a cross-linguistic look, *in* A. Alexiadou, E. Anagnostopoulou and M. Everaert (eds), *The unaccusativity puzzle*, Oxford University Press, Oxford, pp. 332–353.

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

Schreuder, R., De Jong, N. H., Krott, A. and Baayen, R. H.: 1999, Rules and rote: beyond the linguistic either-or fallacy, *Behavioral and Brain Sciences* **22**, 1038–1039.

Sereno, J. and Jongman, A.: 1997, Processing of English inflectional morphology, *Memory and Cognition* **25**, 425–437.

Spieler, D. H. and Balota, D. A.: 1998, Naming latencies for 2820 words, *[On-line], Available: http://www.artsci.wustl.edu/∼dbalota/naming.html*.

Taft, M.: 1979, Recognition of affixed words and the word frequency effect, *Memory and Cognition* **7**, 263–272.

Ullman, M.: 2001, The declarative/procedural model of lexicon and grammar, *Journal of Psycholinguistic Research* **30**, 37–69.

Ullman, M. T., Estabrooke, I. V., Steinhauer, K., Brovetto, C., Pancheva, R., Ozawa, K., Mordecai, K. and Maki, P.: 2002, Sex differences in the neurocognition of language, *Brain and Language* **83**, 141–143.

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

Is inflection computational? Evidence for a redundant lexicon from word naming.

Chapter 3

This chapter has has been submitted to Quaterly Journal of Experimental Psychology as Wieke Tabak, Robert Schreuder, and R. Harald Baayen. Is inflection computational? Evidence for a redundant lexicon from word naming.

Abstract

This paper reports the results of two word naming experiments. The first experiment is a straight forward visual word reading task. The second experiment is a cross tense naming task (also called morphonaming) in which participants have to produce a past tense form when presented with a present tense form, and a present tense form when presented with a past tense form. We found some evidence in favor of a sparse lexicon in which inflected forms are derived by rules from their simple bases. In simple word naming, we found no independent effect of the frequency of inflectional forms. In cross tens word naming, measures tapping into the paradigmatic properties of inflected words were present only for irregular verbs. However, we also found evidence against a sparse lexicon and in favor of a redundant lexicon in which all inflected forms are retrieved form memory. In our straight forward word naming experiment, we found an effect of inflectional entropy and this effect was present for both regular and irregular verbs. This effect of inflectional entropy was modulated by neighborhood density, which reflects paradigmatic aspects of lexical competition. In our cross tense word naming task, we found a facilitatory frequency effect for regular inflected forms (and irregular inflected forms), when the lemma frequency of the base was controlled for. Considered jointly, we think that memory plays a much more pervasive role than

anticipated by theories positing a sparse lexicon.

Introduction

How do speakers produce past-tense forms such as *walked*? In linguistics, two very different answers to this question have been worked out. The dominant view, that can be traced back to Bloch (1947) and Chomsky and Halle (1968), argues that the past-tense form is created from its corresponding present-tense form. The weaver model of speech production (Levelt, Roelofs and Meyer, 1999) implements this view. Following conceptualization, the lemma of WALK is accessed. The lemma provides access to the wordform of the singular. When a diacritic feature for past-tense is active, the past-tense suffix *-ed* is also activated, and added to the singular form. Processes of syllabification and phoneme selection and alignment then converge on the spoken form *walked*. The work by Pinker (Pinker, 1991, 1997, 1999) has popularized this generative model of morphological processing, but various connectionist models have been developed from the same generative premises (e.g. Rumelhart and McClelland, 1986; MacWhinney and Leinbach, 1991). The experimental tasks used most often by researchers working in this tradition is a word naming task in which subjects are presented with a present-tense form and asked to say aloud its past-tense form. This task, and the underlying theory, reflect how pedagogical grammars typically proceed from simpler forms to more complex forms.

A very different approach, one that has largely escaped attention from the psycholinguistics community, argues that inflected forms are not ‘derived’ from more simple forms by morphological rules. Instead, different inflected forms are taken to exist side by side, organized into inflectional paradigms (Matthews, 1974; Blevins, 2003). The past tense form *walked*, according to ‘Word and Paradigm morphology’, is retrieved from memory, instead of being assembled from *walk* and *-ed*.

If inflected forms exist side by side, then one would expect inflectional variants to enter into a selection process. Evidence for such a selection process emerged from research on speech errors by Stemberger (2002, 2004) and Stemberger and Middleton (2003). One of their observations is that the likelihood of producing the correct form of a verb exhibiting vowel alternation, such as *freeze*/*froze*, is co-determined by preceding vowels. For instance, when *cream* precedes *freeze* verb in a past-tense context, the overregularization error (*freezed*) is, although rare, more likely to occur compared to an unrelated control condition. This suggests that it is indeed the case that the two stems are considered side by side in the production process. Normally, we do not see this highly efficient selection process at work, but under experimental magnification, it becomes visible, in conditions where it is being confused by preceding occurrences of the wrong stem vowel.

Further evidence for the availability of full inflectional forms in their paradigms comes from experiments using the picture naming paradigm (Baayen, Levelt, Schreuder and Ernestus, 2008). This study investigated the production of regular Dutch plurals and singulars. The expected facilitation from lemma frequency (the sum of the frequencies of a word's inflectional variants) was observed. In addition, the information carried by the word's paradigm was also predictive. Information was estimated with the help of Shannon's entropy. For a given noun with singular and plural probabilities p_{sg} and p_{pl} (estimated by the corresponding inflectional frequencies, normed by the word's lemma frequency), entropy is defined as $H = -(p_{sg} \log_2 p_{sg} + p_{pl} \log_2 p_{pl})$. As expected, this inflectional entropy was positively correlated with the picture naming latencies: It took more time to retrieve forms from information rich, complex, number paradigms than from information-poor, simple, number paradigms. Effects of entropy have also been reported in comprehension studies, including studies on Serbian, a language with a much more complex system of nominal inflection than Dutch (see, e.g., Milin, Filipović Durdević, and Moscoso del Prado Martín, 2009; Milin, Kuperman, Kostić and Baayen, 2009). This body of results suggest not only that paradigmatic structure is relevant for understanding the processing of inflected words, but also that long-term memory is highly sensitive to the probabilities of the individual inflected variants.

Frequency effects have a long history, beginning with Taft and Forster (1976), of being used as diagnostics for whole-word and constituent-driven processing. Proponents of the generative view of inflectional processing have been very uncomfortable with the whole-word frequency effects for inflected forms. For instance, Pinker (1991) argued that frequency effects would be present only for irregular inflections, and should not be observable for regular inflections, the idea being that regular inflections are produced and understood by rule, and therefore do not leave traces in memory. In later work, e.g., Pinker and Ullman (2002) it is acknowledged that there may be frequency effects for regular inflections, but that whole-word representations for regular inflected forms are not used in normal language use. Such whole-word representations would be useful, and come into play, only under the specific task requirements of experimental paradigms such as lexical decision.

More recent work in the generative framework, however, modifies this strong claim in two ways. First, frequency effects for regular verbs in English are allowed to arise when regular verbs have rimes that are phonologically similar to the rimes of irregular verbs (Ullman, 2001; Hartshorne and Ullman, 2006). Second, Ullman, Estabrooke, Steinhauer, Brovetto, Pancheva, Ozawa, Mordecai and Maki (2002) (see also Walenski and Ullman, 2005) observed frequency effects for regular inflected verbs in English for females but not for males, a finding that is in line with the superior verbal memory of females (see for a review Kimura, 2000). Hartshorne and Ullman (2006), furthermore, discuss evidence suggesting that males may depend more on rule-driven processing in procedural memory compared to females. The possibility that these differences between males and females may be more a matter of degree than an absolute difference, with males having a somewhat 'deficient' declarative memory, is supported by auditory comprehension studies by Balling and Baayen (2008, 2009), who also observed stronger whole-word frequency effects for complex words for females, in line with the observations of Ullman and colleagues. However, the frequency effect for the Danish males, although weaker, still reached full significance.

The aim of the present paper is to explore the possibility that regular inflections do leave traces in lexical memory, that these traces affect lexical processing with a task that does not have a decision component, and that the strength of these traces may vary between men and women. Experiment 1 is a straightforward visual word naming task, Experiment 2 is a cross-tense word naming task in which subjects are presented with present-tense forms for which they have to say aloud the past-tense forms, and past-tense forms for which they have to produce the present-tense forms. Equal numbers of females and males participated in the experiments, so that potential processing differences between the sexes could be investigated. A key variable for both studies is the probability of the present and past tense forms in their paradigms. If regular inflected forms leave traces in lexical memory, then these probabilities should emerge as significant predictors across regulars and irregulars. If inflected forms are not derived from their stems, but retrieved from their embedding paradigms, then cross-tense naming should reveal itself as a task generating experimental artefacts. Before discussing the two experiments in detail, we first introduce the full set of our predictors.

Predictors

In both experiments, we implemented orthogonal contrasts for the factors *Regularity* (regular versus irregular), *Number* (singular versus plural) and *Tense* for verb forms in the first person (I/we), as shown in Table 3.1. To obtain insight in the possible differences between males and females, we registered the Sex of our participants (male versus female).

Verb	**Number**	**Tense**	**Regularity**
loop	singular	present tense	irregular
liep	singular	past tense	irregular
loop-en	plural	present tense	irregular
liep-en	plural	past tense	irregular
lach	singular	present tense	regular
lach-te	singular	past tense	regular
lach-en	plural	present tense	regular
lach-ten	plural	past tense	regular

Table 3.1: Examples of the first person forms of a regular verb ('lopen' *walk*) and an irregular verb ('lachen' *laugh*) in Dutch.

Various predictors were introduced as controls. Response latencies in an experiment constitute a time series, with a latency at trial t entering into significant correlations with latencies at prior trials. Following De Vaan, Schreuder and Baayen (2007) and Balling and Baayen (2008), we took interdependencies of the current trial with the preceding four trials into account by decorrelating the vectors of response latencies at the four preceding trials with principal components analyses, and adding the first three principal components (henceforth *PC1, PC2, PC3*) as predictors in our regression equations.

Another control variable was the position of a word in the experiment, measured by its rank (trial number) in the experimental list (henceforth *Trial*). Counterbalancing controls for the adverse effects of habituation or fatigue, but it does not take the noise associated with these longitudinal effects out of the model. By including this variable, potential longitudinal effects of habituation or fatigue can be brought into the model, and removed from the error term. To control for voicekey artefacts (Kessler, Treiman and Mullennix, 2002), we included the factor *Voice* (with levels voiced and voiceless) and *Manner* of the initial phoneme (with levels plosive, fricative, nasal, approximant, vowel, lateral, and r).

We used the number of letters of the word as shown on the computer screen as a measure of its length. To anticipate results, this measure, nor length in phonemes, reached significance in our analyses.

We also explored several measures gauging phonological similarities. Following

Ullman (2001), we determined whether a regular verb had a rhyme that is also attested with irregular verbs. Instead of introducing this dimension in the form of a dichotomy, we used the number of irregular verbs that rhymed with a given target verb (*RhymeCount*) as a covariate, as Tabak, Schreuder and Baayen (2005) observed (for visual lexical decision) that the number of rhyming irregulars was a significant predictor of processing time. We note here that the models of Albright and Hayes (2003) and Keuleers (2008) also predict that the continuous measure should be a superior predictor. If the results observed by Ullman generalize to Dutch, we should observe stronger frequency effects for regulars with many rhyming irregulars.

In addition, we considered several measures of neighborhood density. The N-count measure of Coltheart (1977) never reached significance for this data set, and therefore will not be discussed below. Instead, positional neighborhood counts emerged as potentially relevant. Positional neighborhood counts were first studied by Bien (2007) and Baayen (2007). The intuition underlying positional counts is that the number of neighbors that mismatch at the initial position might be more influential than the number of neighbors that mismatches at the final position of the word, due to the simple fact that pronunciation has to start with the initial segment. In a cued response production paradigm, Bien observed an inhibitory effect of the count of initial neighbors, *N1*. We also considered the neighbors that differed at the second position, *N2*. The counts of remaining neighbors at later positions were collapsed into a third measure, *Nrest*. Of special interest is the question whether these neighborhood measures will enter into significant interactions with regularity.

In Dutch, syllable-final obstruents may alternate between voiced and voiceless (a phenomenon similar to the alternation in English *wife–wives*, but more regular and more widespread throughout the lexicon). Such alternations are known to affect auditory comprehension (Ernestus and Baayen, 2007). We therefore incuded a factor specifying whether the verb's stem-final obstruent undergoes voicing alternation in order to further explore the predictivity of this factor (henceforth *Alternating*).

As a morphological predictor we included the verb's morphological family size (*FamilySize*), the number of complex words for which the verb stem is a constituent (Schreuder and Baayen, 1997; De Jong, Schreuder and Baayen, 2000). In our experiments, family sizes ranged from 0 (*moeten* 'must') to 298 (*dienen* 'to serve'). It is unclear whether we should expect the *Family Size* predictor to reach significance in the word naming task, as Baayen, Feldman and Schreuder (2006) did not observe its effect in the naming latencies of the English Lexicon Project (Balota, Cortese, Sergent Marschall, Spieler and Yap, 2004), whereas they did observe it for the corresponding lexical decision latencies. For analogous reasons, it is unclear whether an effect might be observed for the verb's number of meanings, as gauged by the Number of Synsets (Baayen and Moscoso del Prado Martiín, 2005) in which the verb is listed in the verb section of the Dutch EuroWordNet (Vossen, Bloksma and Boersma, 1999), the Dutch version of the English WordNet developed by Miller and colleagues (Miller, 1990). To anticipate results, the synset measure never reached significance in our naming studies, and will therefore not be discussed further. The verb's *Inflectional Entropy, H*, by contrast, did reach significance.

As frequency measures we included, first of all, the verb's lemma frequency *LemmaFreq*. We addressed the role of inflectional frequencies indirectly, through two measures: the verb's inflectional entropy on the one hand, and the proportionality of the frequencies of the present and past tense forms on the other hand. An effect of inflectional entropy has been observed in lexical decision for both English (Baayen, Feldman and Schreuder, 2006) and Dutch (Tabak, Schreuder and Baayen, 2005), but given that this measure failed to reach significance for the naming latencies in the English Lexicon Project, it is unclear whether it will emerge as significant for the present naming experiments.

A measure that we investigated, following Baayen, Levelt, Schreuder and Ernestus (2008), is the *Inflectional Paradigmatic Probability*, the token frequency of the inflectional form to be read aloud divided by the the sum of this token frequency and the token frequency of the corresponding form in the other tense. Given a verb with inflectional token frequencies 10 for present-tense singular and 20 for past-tense singular, the *Inflectional Paradigmatic Probability* is 10/30 when reading aloud the present tense form, and 20/30 when reading aloud the past-tense form. We calculated these probabilities separately for singular tense pairs and for plural tense pairs. The *Inflectional Paradigmatic Probability* estimates the ease with which an inflectional form can be accessed in its inflectional paradigm.

Of special interest is whether the *Inflectional Paradigmatic Probability* will enter into interactions with *Regularity* and *Sex*. If Ullman's results generalize to Dutch, then one would expect a stronger effect of the *Inflectional Paradigmatic Probability* for females for both regulars and irregulars, and for males an effect only for irregulars.

Experiment 1: Same-Tense Word Naming

Method

Materials The materials consisted of 143 irregular and 143 regular native (Germanic) monomorphemic verbs, matched in the mean for Lemma Frequency in CELEX (Baayen, Piepenbrock and Gulikers, 1991), which were also studied by Tabak *et al.* (2005) with the visual lexical decision task.

We constructed four master lists, each with 286 word forms. The first list contained present singulars, the second list present plurals, the third past singulars, and the fourth past plurals. Each master list was used for a separate sub-experiment. For a given sub-experiment, the words in its master list were presented in ten different random orders. Each sub-experiment was preceded by twenty practice items.

Participants Twenty women and twenty men, all students at the university of Nijmegen, participated in the same-tense word naming experiment. Ten women and ten men participated in the two experiments with the plural forms. Half of these subjects did the sub-experiment with the present plurals first, the other half did the sub-experiment with the past plurals first. The two sub-experiments were carried out one week apart. The remaining ten women and ten men similarly participated in the sub-experiments with the singular verb forms. Thus, Tense was a within-subject factor, and Number a between-subject factor. All participants had normal or corrected-to-normal vision. Participants received 5 Euro for each of the two sub-experiments in which they took part.

Procedure Participants were tested individually in a noise-attenuated experimental room. They were asked to say aloud the presented form as quickly and accurately as possible. Each visual stimulus was preceded by a fixation mark in the middle of the screen for 500 ms. After 50 ms, the stimulus appeared 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 they remained on the screen for 1500 ms. A new trial was initiated 500 ms afterwards. There was a short break following the practice items, and three further short breaks during the experiment itself. The total duration of an experimental session was approximately 35 minutes.

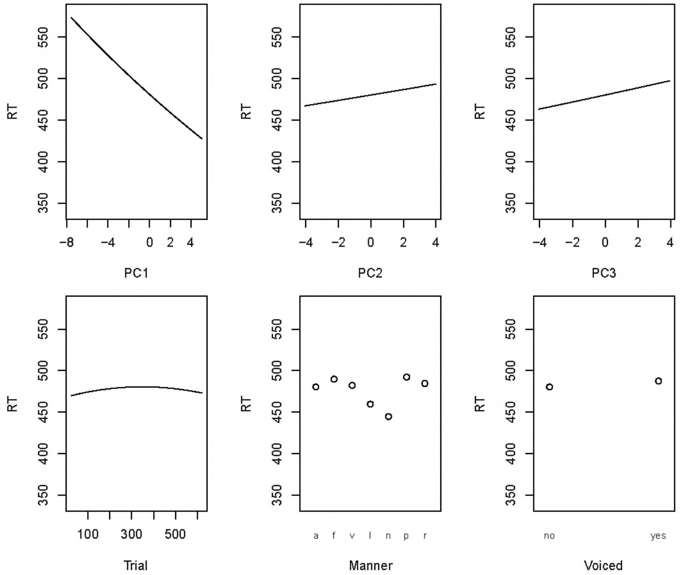

Figure 3.1: Partial effects of the control predictors for the visual naming latencies of Experiment 1.

Results

We removed data points with responses where the voicekey was triggered by vocalizations other than those of the onset of the target word (1.8% of the observations). The distribution of the remaining latencies was highly skewed. We reduced this skewness by means of a logarithmic transformation. We fitted a linear mixed-effects model (Baayen, Davidson and Bates, 2008) to the data, using a backwards stepwise variable selection procedure, with participant and verb as crossed random effects. Factors were modeled with contrast coding. The residuals of this model showed marked deviations from normality. We therefore removed data points with absolute standardized residuals exceeding 2.5 (2.4% of the trials), and refitted the model. The qualitative pattern in this trimmed model is the same, but the estimates of the coefficients and their standard deviations are more precise. We then examined whether by-subject random slopes for covariates were justified by likelihood ratio tests, which turned out to be the case for *Lemma Frequency, Inflectional Paradigmatic Probability, N1, N2*, and *Nrest* (all $p < 0.0001$).

	Estimate	std err	t value	p(t)	p(MCMC)	95% lower	MCMC mean	95% upper
Intercept	6.3085	0.0761	82.9056	0.0000	0.0001	6.2280	6.3321	6.4283
Manner: fricative	0.0193	0.0108	1.7970	0.0723	0.0598	-0.0004	0.0194	0.0402
Manner: vowel	0.0039	0.0320	0.1230	0.9021	0.9108	-0.0579	0.0037	0.0639
Manner: lateral	-0.0443	0.0153	-2.8968	0.0038	0.0020	-0.0723	-0.0443	-0.0159
Manner: nasal	-0.0773	0.0164	-4.7149	0.0000	0.0001	-0.1069	-0.0774	-0.0465
Manner: plosive	0.0245	0.0109	2.2522	0.0243	0.0200	0.0038	0.0244	0.0450
Manner: r	0.0087	0.0171	0.5113	0.6091	0.5840	-0.0215	0.0090	0.0429
Voiced: yes	0.0148	0.0064	2.3242	0.0201	0.0124	0.0028	0.0146	0.0264
PC1	-0.0233	0.0011	-21.9414	0.0000	0.0001	-0.0256	-0.0235	-0.0214
PC2	0.0068	0.0013	5.2207	0.0000	0.0001	0.0042	0.0068	0.0094
PC3	0.0089	0.0014	6.4236	0.0000	0.0001	0.0062	0.0089	0.0117
Tense: present	-0.0340	0.0027	-12.6539	0.0000	0.0001	-0.0393	-0.0341	-0.0288
Number: singular	-0.2064	0.0951	-2.1703	0.0300	0.0006	-0.3151	-0.2047	-0.1008
LemmaFreq	-0.0071	0.0021	-3.2985	0.0010	0.0002	-0.0112	-0.0071	-0.0031
FamilySize (linear)	-0.0261	0.0125	-2.0865	0.0369	0.0282	-0.0493	-0.0260	-0.0032
FamilySize (quadratic)	0.0037	0.0020	1.8791	0.0603	0.0482	-0.0001	0.0037	0.0072
Sex: female	-0.2695	0.0951	-2.8345	0.0046	0.0001	-0.3725	-0.2667	-0.1648
N1	-0.0392	0.0176	-2.2266	0.0260	0.0254	-0.0706	-0.0377	-0.0031
N2	-0.0034	0.0044	-0.7623	0.0000	0.0004	-0.0263	-0.0181	-0.0098
Nrest	-0.0006	0.0008	-0.7249	0.4685	0.4544	-0.0023	-0.0006	0.0011
IPP	-0.0087	0.0101	-0.8634	0.0034	0.0028	-0.0480	-0.0289	-0.0110
Regularity: regular	0.0263	0.0079	3.3331	0.0009	0.0006	-0.0411	-0.0267	-0.0115
H	-0.0191	0.0123	-1.5514	0.1208	0.1230	-0.0421	-0.0185	0.0048
Alternating: yes	-0.0037	0.0062	-0.5884	0.5563	0.5288	-0.0156	-0.0038	0.0078
Trial (linear)	0.0001	0.0000	6.3910	0.0000	0.0001	0.0001	0.0001	0.0002
Trial (quadratic)	-0.0000	0.0000	-6.1652	0.0000	0.0001	0.0000	0.0000	0.0000
IPP * Regularity: irregular	-0.0204	0.0058	-3.5439	0.0004	0.0006	0.0087	0.0205	0.0315
H * N1	0.0161	0.0074	2.1829	0.0291	0.0318	0.0016	0.0156	0.0298
Number: singular * Alternating: yes	0.0195	0.0040	4.8383	0.0000	0.0001	0.0116	0.0195	0.0273
Number: singular * Sex: female	0.3062	0.1344	2.2778	0.0227	0.0001	0.1568	0.3033	0.4514
Number: singular * N1	0.0231	0.0022	10.5856	0.0000	0.0001	0.0188	0.0231	0.0273
Number: singular * Nrest	-0.0021	0.0006	-3.5315	0.0004	0.0008	-0.0033	-0.0021	-0.0010
N2 * Regularity: regular	-0.0142	0.0049	-2.9213	0.0035	0.0030	0.0051	0.0144	0.0239

Table 3.2: Regression model fitted to the same-tense naming latencies of Experiment 1. IPP: *Inflectional Paradigmatic Probability*

The resulting model is summarized by Table 4.1, which lists the coefficients of the fixed-effect factors and covariates, Table 3.3, which lists the standard deviations of the random effects, Figure 3.1, which graphs the effects of the control predictors, and Figure 3.2, which visualizes the effects of the key predictors.

First consider the effects of the control variables in Figure 3.1. Naming latencies entered into significant positive correlations (r approximately 0.2) with the naming latencies elicited for the preceding trials (cf. De Vaan *et al.*, 2007; Balling and Baayen, 2008): the longer the latencies to the preceding trials, the longer the response latency to the target. All three principal components obtained by orthogonalizing the vectors of response latencies at preceding trials (*PC1–PC3*) emerged as significant predictors. Inspection of the respective loadings shows that longer preceding latencies predict longer latencies to the target, as expected. The upper left panel of Figure 3.1 shows that the effect of PC1 is substantial, with a range of some 150 ms. This effect size exceeds the effects of all other predictors, highlighting the importance of bringing this source of experimental noise into the statistical analysis.

The lower left panel indicates that as participants went through the experiment they first responded increasingly slowly, but towards the end of the experiment responded more quickly. The remaining two lower panels illustrate the effects of the two control variables for the differential sensitivity of the voice key to different kinds of initial phonemes: *Manner* and *Voice*. Voiced initial segments triggered the voicekey before voiceless segments. Plosives were picked up by the voice key much later than notably nasals and the l: the voicekey is triggered by the burst following the initial silence of the plosive, leading to artefactually elongated latencies. By bringing *Manner* and *Voice* into the model, these imprecisions in the measurements are brought under statistical control.

Next consider Figure 3.2, which illustrates the effects of the predictors of theoretical interest. The upper panels present the three neighborhood measures. The counts of neighborhoods for the first and second segment were log-transformed to avoid outlier effects from isolated words with very high counts. The effect of N1 changed sign between singulars, for which it was inhibitory, and plurals, for which it was facilitatory. Inhibition from initial neighbors was observed by Bien (2005) in a cued-response production task. The upper left panel also shows that plurals elicited longer latencies than singulars. The central upper panel visualizes the interaction of regularity by the count of neighbors at the second position (N2). For irregulars, but not for regulars, this count was facilitatory. For

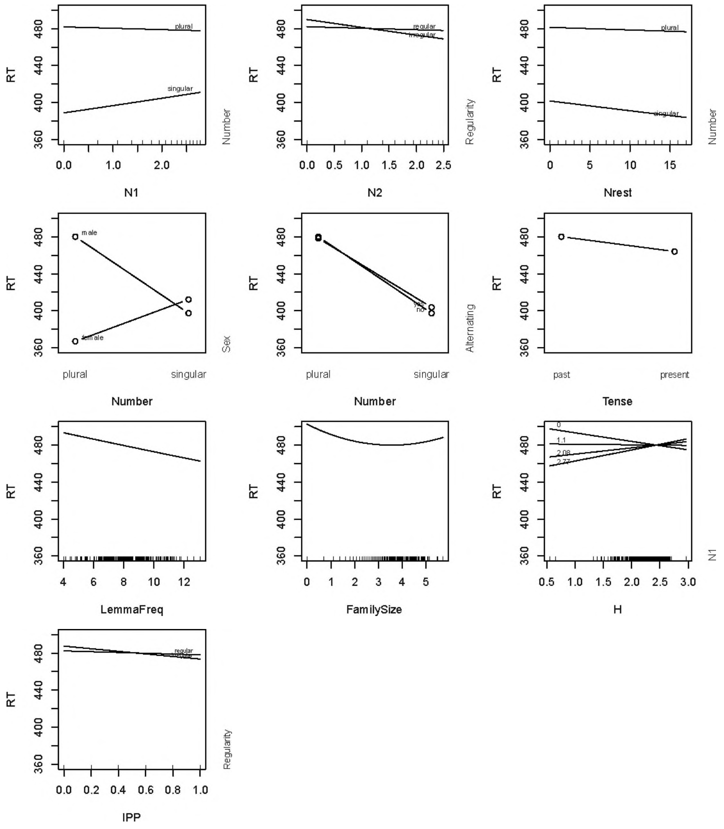

Figure 3.2: Partial effects of the critical predictors for the same-tense naming latencies of Experiment 1. Each graph is adjusted for the reference levels of the other factors in the model, and for the median of the other covariates. H: *Inflectional Entropy; IPP:* Inflectional Paradigmatic Probability.

Groups	Name	Std.Dev.
Verb	Intercept	0.0402
Subj	Intercept	0.2108
Subj	Lemma Frequency	0.0067
Subj	Inflectional Frequency Ratio	0.0578
Subject	N1	0.0275
Subject	N2	0.0092
Subject	Nrest	0.0025
Residual		0.1354

Table 3.3: Standard deviations for the random intercepts, random slopes, and residual error estimated by the regression model fitted to the same-tense naming latencies of Experiment 1.

most verbs, the second position is the position of the vowel, the segment that for irregulars is different between the past and present forms. For irregulars, a large N2 implies a large number of words supporting the vowel in the word form that is to be pronounced. This support affords faster processing and shorter naming latencies. The effect of the neighborhood density at later positions in the word (Nrest) was facilitatory for singulars, and irrelevant for plurals.

In summary, for singulars, we observe inhibition from initial neighborhood density, that reverses into facilitation at later positions, albeit at the second position only for irregulars. For plurals, facilitation is present from the first position, persisting for irregulars at the second position, and disappearing at later positions. Thus, the inhibition from initially diverging neighbors for singulars is the exception to overall facilitation from neighborhood density. As 75% of the singulars are monosyllabic (only regular past tense singulars are bisyllabic), our working hypothesis is that it is the planning of the first segment for articulation for unmarked monosyllables that is disturbed by larger numbers with the same rhyme but a different (simple) onset. We leave this issue for further research.

An interaction of Sex by Number reached significance. This interaction is summarized in the first panel in the second row of Figure 3.2 (by-subject random contrasts for number were not justified by likelihood ratio testing, indicating we are confronted here with a true group difference). Both sexes responded approximately equally fast to singulars, but plurals elicited longer latencies from males than from females. This is the only interaction with Sex that reached significance in this experiment. The elongated latencies of males in response to plural verbs is at odds with the hypothesis of Hartshorne and Ullman (2006) that males would be more proficient in processing complex words, in contradistinction to the advantage of females in retrieving words from memory. Our data testify to an overall advantage of females in verbal skills.

It might be argued that the males are slower in processing plurals because they use rules parsing the visual input into stem and suffix, and subsequently production rules to prepare the plural forms for articulation given stem and suffix. Females, by contrast, would simply retrieve plurals from memory, allowing them to respond faster. Although this explanation is in line with the assumption of parallel dual route models that the parsing route is slower compared to the retrieval route (see, e.g., Hay, 2001), we will argue below that this explanation is challenged by the presence of an effect of *Inflectional Entropy* that, moreover, is the same for females and males.

The central panel of Figure 3.2 illustrates the small but significant effect of a verb having a stem-final alternating obstruent. If so, singulars, but not plurals, are at a small disadvantage. This is probably due to the orthography presenting a voiced final segment to the participant, who then has to articulate the voiceless counterpart. The next panel shows that the marked past-tense forms elicited longer latencies than the unmarked present-tense forms, as expected.

Turning to the third row of panels in Figure 3.2, we observe a healthy effect of *Lemma Frequency*, with a span of some 50 ms. Inclusion of by-subject random slopes for frequency proved necessary ($p < 0.0001$), but there was no evidence for a group difference between females (supposedly showing a stronger frequency effect), and males (supposedly with the weaker frequency effect).

A concave effect of morphological Family Size also emerged. Initial facilitation followed by inhibition for the largest family size counts was also observed for English monomorphemic words by Baayen *et al.* (2006) for visual lexical decision. No by-subject random slopes were required, and no group difference between females and males could be detected. As family size effects are generally interpreted as reflecting semantic processing (De Jong, Schreuder and Baayen, 2000; Moscoso del Prado Martín, Kostić and Baayen, 2004; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder and Baayen, 2004), the presence of an effect of *Family Size* indicates that participants were not paying attention to word form only, but also accessed the verbs' meanings.

The next panel shows the interaction of *Inflectional Entropy* (H) by the initial neighborhood count (H1). For large initial neighborhoods, inflectional entropy is inhibitory. As initial neighborhood size decreases, the effect of inflectional entropy eventually reverses and becomes facilitatory. No further interactions with regularity or number reached significance. The *Inflectional Entropy* measure quantifies the amount of information carried by the verb's inflectional paradigm. The facilitation observed in the absence of initial orthographic competitors is in line with the facilitation observed in other studies (see, e.g., Baayen *et. al.*, 2006). As the number of initial competitors increases, an informationally rich paradigm seems to increase the complexity of selecting the target form from its neighbors. The presence of an effect of *Inflectional Entropy* shows that the probabilities of inflected forms, irrespective of regularity, contribute to the information carried by the paradigms in which they participate, and provides support for the theory of 'Word and Paradigm Morphology'.

The final panel of Figure 3.2 shows that the *Inflectional Paradigmatic Probability* had a small facilitatory effect for irregular verbs. That is, if a present tense of an irregular verb is to be read aloud, then its naming latency will be shorter the more often this present-tense form is used compared to its corresponding irregular past-tense form. Similarly, if an irregular past tense form is to be named, this will be accomplished more quickly the more the frequency of this past-tense form exceeds that of the corresponding present-tense form.

The absence of an effect of the *Inflectional Paradigmatic Probability* for regular verbs can be understood as evidence supporting the generative dual mechanism model, since according to this model regular inflectional forms are not stored in memory. In fact, in the absence of interactions with *Sex*, the data do not even support the more graded version of this model as developed by Ullman and collaborators, in which females, but not males, remember higher-frequency regular inflected forms. Given that a significant effect of inflectional entropy is present that does not interact with regularity, it cannot be argued that the brain would not be sensitive to the probabilities of regular inflected forms.

Returning to the substantial delay (some 100 ms) for males compared to females when naming plurals, the conclusion must be that Dutch males must have some general lexical processing disadvantage compared to females. We have seen no evidence in the present data that males would have inferior lexical memory compared to females. All that the present data allow us to say is that males apparently are less proficient in using and integrating information from lexical memory.

Experiment 2: Cross-Tense Word Naming

Experiment 2 investigates the production of verbal inflections in Dutch using cross-tense naming. In cross-tense naming, participants are shown a verb form on the computer screen, and are requested to say aloud the corresponding inflected form in the opposite tense. When presented with a present-tense form, they have to produce the past-tense form, and when shown a past-tense form, they have to produce the corresponding present-tense form. Cross-tense naming, restricted to naming the past from the present, has been used to investigate the role of regularity in the production of past-tense verbs (Jaeger, Lockwood, Kemmerer, Van Valin and Murphy, 1996; Ullman, Bergida and O'Craven, 1997; Ullman *et. al.*, 2002). Present-to-past cross-tense naming reflects both the practice in pedagogical grammars to proceed from the unmarked to the marked forms, as well as the central claim of the dual mechanism model and other generative models of morphology, according to which a past tense form such as *walked* is produced from the present-tense form *walk* by means of a rule adding the suffix *-ed*. We explored the cross-tense naming task in two directions, not only from the present to the past, but also from the past to the present.

Method

Materials

The materials were identical to those used in Experiments 1. We constructed four master lists, each with 286 word forms. The first list contained present singulars, the second list present plurals, the third past singulars, and the fourth past plurals. Each master list was used for a separate sub-experiment. For a given sub-experiment, the words in its master list were presented in ten different random orders. Each sub-experiment was preceded by twenty practice items.

Participants Twenty women and twenty men (other than those participating in the same-tense word naming experiment), students at the university of Nijmegen, participated in the cross-tense naming experiment. As for same-tense word naming, *Tense* was a within and *Number* a between-subject factor. All participants had normal or corrected-to-normal vision. Participants received 5 Euro for each of the two sub-experiments in which they took part.

Procedure Participants were tested individually in a noise-attenuated experimental

Groups	Name	Standard Deviation
Verb	Intercept (Sex: male)	0.0490
Verb	Sex: female	0.0203
Subj	Intercept	0.1966
Subj	Nrest	0.0211
Subj	N1	0.0174
Subj	Inflectional Paradigmatic Probability	0.1065
Subj	Lemma Frequency	0.0031
Residual		0.2151

Table 3.4: Standard deviations for the random intercepts, random slopes, and residual error estimated by the regression model fitted to the cross-tense naming latencies of Experiment 2.

room. We asked the participants to say aloud the past tense form when presented with the present, and the present tense form when presented with the past tense form. Each visual stimulus was preceded by a fixation mark in the middle of the screen for 500 ms. After 50 ms, the stimulus appeared 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 they remained on the screen for 1500 ms. A new trial was initiated 500 ms afterwards. There was a short break following the practice items, and three further short breaks during the experiment itself. The total duration of an experimental session was approximately 35 minutes.

Results

For the analysis of the response latencies, we first removed error responses (voice key errors, 2.5%, as well as incorrect responses, 9.7%, in all 12.2%) from the data set. As in the preceding experiment, the distribution of the latencies was highly skewed, with outliers in both tails. We reduced this skewness by means of a logarithmic transform. As for same-tense naming, model fitting was followed by removal of standardized residuals with an absolute value exceeding 2.5. In what follows, this trimmed model is discussed.

We analyzed the response latencies with a linear mixed-effects model with random intercepts for Subject and Verb (Table 3.4). By-subject random slopes

	Estimate	std err	t value	p(t)	p(MCMC)	95% lower	MCMC mean	95% upper
Intercept	7.2257	0.0767	94.1809	0.0000	0.0001	7.0861	7.2208	7.3533
Manner: fricative	-0.0043	0.0123	-0.3496	0.7266	0.7212	-0.0262	-0.0043	0.0190
Manner: vowel	-0.0473	0.0365	-1.2945	0.1955	0.1822	-0.1127	-0.0465	0.0233
Manner: lateral	-0.1035	0.0174	-5.9584	0.0000	0.0001	-0.1354	-0.1036	-0.0714
Manner: nasal	-0.0513	0.0184	-2.7917	0.0052	0.0052	-0.0862	-0.0514	-0.0164
Manner: plosive	-0.0117	0.0124	-0.9418	0.3463	0.3310	-0.0352	-0.0116	0.0108
Manner: r	-0.0616	0.0195	-3.1642	0.0016	0.0008	-0.1005	-0.0613	-0.0268
Voiced: yes	0.0353	0.0073	4.8712	0.0000	0.0001	0.0220	0.0353	0.0490
PC2	0.0195	0.0019	10.3033	0.0000	0.0001	0.0157	0.0194	0.0231
PC3	-0.0109	0.0020	-5.3673	0.0000	0.0001	-0.0150	-0.0110	-0.0069
Number: singular	-0.1156	0.0633	-1.8254	0.0680	0.0188	-0.2106	-0.1145	-0.0196
H	-0.1064	0.0167	-6.3637	0.0000	0.0001	-0.1375	-0.1061	-0.0757
Regularity: regular	-0.2756	0.0496	-5.5581	0.0000	0.0001	-0.3687	-0.2749	-0.1813
LemmaFreq (linear)	-0.0977	0.0132	-7.4055	0.0000	0.0001	-0.1211	-0.0973	-0.0720
LemmaFreq (quadratic)	0.0055	0.0008	6.9512	0.0000	0.0001	0.0040	0.0055	0.0070
Tense: past	-0.0438	0.0146	-2.9975	0.0027	0.0036	-0.0717	-0.0432	-0.0143
N1	0.0230	0.0044	5.2407	0.0000	0.0001	0.0144	0.0231	0.0314
IPP	-0.0554	0.0214	-2.5918	0.0096	0.0086	-0.0939	-0.0546	-0.0149
Alternating: yes	0.0384	0.0070	5.5112	0.0000	0.0001	0.0255	0.0387	0.0513
Trial (linear)	-0.0002	0.0000	-4.5199	0.0000	0.0001	-0.0003	-0.0002	-0.0001
Trial (quadratic)	0.0000	0.0000	3.8486	0.0001	0.0004	0.0000	0.0000	0.0000
RhymeCount	-0.0245	0.0072	-3.4269	0.0006	0.0006	-0.0381	-0.0244	-0.0110
H * Regularity: regular	0.0942	0.0211	4.4618	0.0000	0.0001	0.0545	0.0941	0.1336
Tense: past * IPP	0.1183	0.0225	5.2547	0.0000	0.0001	0.0732	0.1171	0.1614
Regularity: regular * Tense: past	-0.0693	0.0139	-5.0030	0.0000	0.0001	-0.0968	-0.0693	-0.0423
Regularity: regular * IPP	0.0573	0.0149	3.8403	0.0001	0.0002	0.0272	0.0571	0.0858
RhymeCount * Tense: past	-0.0151	0.0046	-3.2720	0.0011	0.0012	-0.0241	-0.0151	-0.0061
RhymeCount * Regularity: regular	0.0479	0.0089	5.3803	0.0000	0.0001	0.0313	0.0477	0.0648
Tense: past * Regularity: regular * IPP	-0.0863	0.0222	-3.8789	0.0001	0.0001	-0.1293	-0.0863	-0.0425

Table 3.5: Regression model fitted to the cross-tense naming latencies of Experiment 2. IPP: Inflectional Paradigmatic Probability.

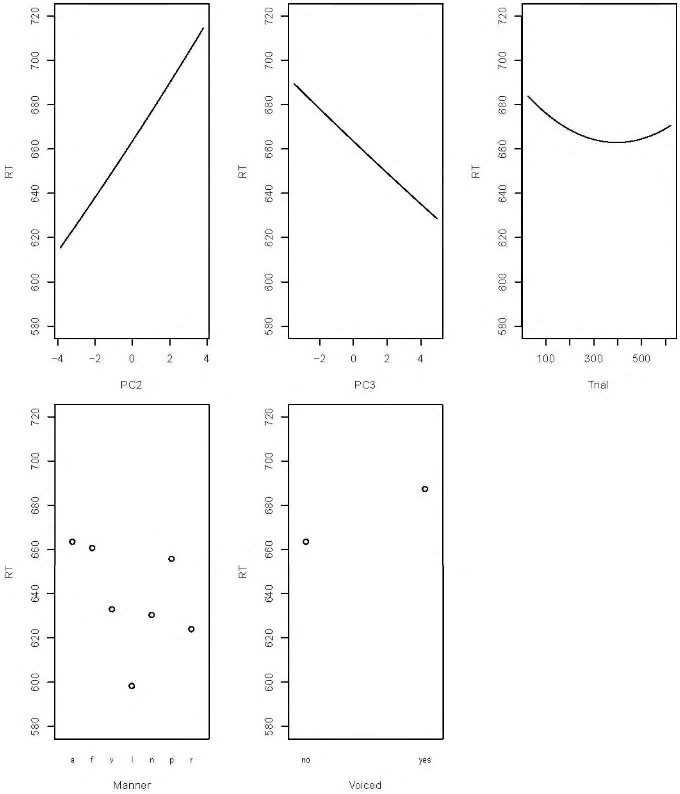

Figure 3.3: Partial effects of the control predictors for the cross-tense naming latencies of Experiment 2.

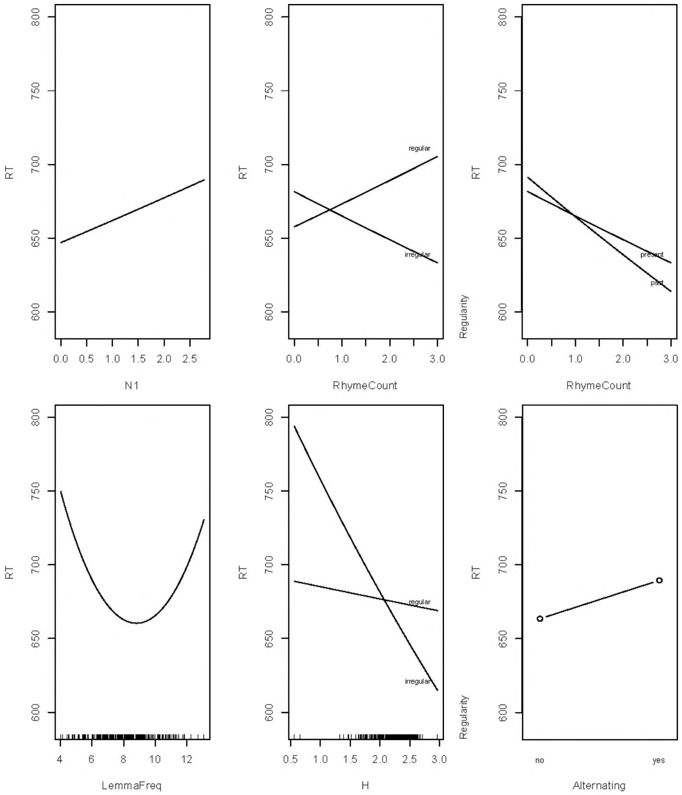

Figure 3.4: Partial effects of the critical predictors for the cross-tense naming latencies of Experiment 2. Graphs are adjusted for the reference level of the other factors, and for the median of the other covariates in the model.

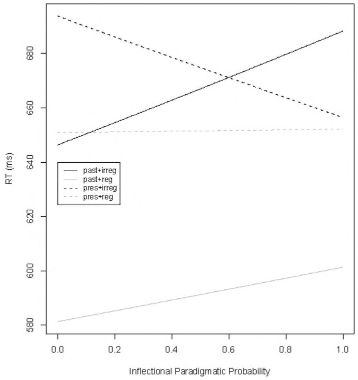

Figure 3.5: The three-way interaction of *Inflectional Paradigmatic Probability* by *Tense* by *Regularity* in Experiment 2. Intercepts are adjusted for the medians of the other covariates. Past: present-to-past naming, Pres: past-to-present naming.

reached significance for the initial neighborhood count (*N1*) and the neighborhood count at later positions in the word (*Nrest*), for *Lemma Frequency*, and for *Inflectional Paradigmatic Probability*. Furthermore, by-verb random contrast were required for *Sex*: the standard deviation of the by-verb intercepts for males was twice as large as that for females.

Table 5.5 and Figures 3.3, 3.4 and 3.5 present the effects of fixed-effect factors and covariates. First consider the effects of the controls, shown in Figure 3.3. As in the same-tense naming experiment, longer latencies at preceding trials predicted longer current latencies, as shown by the principal components *PC2* and *PC3* orthogonalizing the vectors of the four preceding naming latencies. The upper right panel indicates that in the course of the experiment, subjects first habituated to the task, but fatigued towards the end of the experiment. There were substantial differences in the sensitivity of the voice key to the manner and voicing of the initial segment of the word, as shown in the lower panels of Figures 3.3.

Turning to the linguistic predictors, the upper left panel of Figure 3.4 illustrates the longer latencies elicited by verbs with large initial neighborhoods. The next two panels summarize the effect of *RhymeCount*, the count of irregular verbs that share the rhyme of the target verb. Whereas in same-tense naming this predictor did not reach significance, it emerged in cross-tense naming in interaction with *Regularity* and *Tense*. The first of these panels shows the effect of *RhymeCount* for regular and irregular verbs for the reference level of *Tense*, present-tense naming.

For regular verbs, similarity to many irregulars led to longer response latencies. Withstanding the pull of irregularization takes time. By contrast, irregular verbs with large rhyme counts elicited shorter naming latencies. Being part of a larger gang of similar irregulars speeded retrieval of the form to be named. The interaction with *Tense* shown in the upper right panel indicates that when the past-tense form was to be named, the slope for *RhymeCount* was more negative. Combining the two interactions, we have that the facilitation for irregulars is even stronger in past-tense naming, and that the inhibition for regulars is more muted in past-tense naming. Since in this experiment, subjects were presented with the past-tense form when having to say aloud the present-tense form, the confusion for regulars caused by large numbers of rhyming irregulars is most prominent when subjects start from the past tense, the tense in which irregularity manifests itself, as expected. The effect of *RhymeCount* did not vary significantly between the sexes.

The bottom left panel of Figure 3.4 graphs the U-shaped effect of *Lemma Frequency*. For higher-frequency verbs, the usually observed facilitation from frequency reverses into inhibition. This U-shaped effect was stable across *Tense* and *Regularity*. Given that in same-tense naming the same predictor, for the same set of verbs, is facilitatory across the full range of frequencies, the unexpected inhibition must be induced by the task of cross-tense naming. Apparently, subjects optimize their response speed for the lemma frequencies that are most likely in the experiment. The more a frequency is unexpected, the more exceptional it is, and the longer the latencies are, irrespective of whether the unexpected frequency is very high or very low.

We examined this effect in further detail by adding as a predictor to the model the form frequency of the inflections, orthogonalized with lemma frequency by entering into the model the residuals of a linear model regressing form frequency on lemma frequency. The residualized measure ($r = 0.58$ with the original measure) captures experience with the inflected form to the extent that this cannot be predicted from lemma frequency. In the model with (residualized) form frequency as additional predictor, the effect of lemma frequency remained U-shaped, but an additional effect of form frequency emerged as well. This effect was stronger for irregulars (contrast in slope: $\hat{\beta} = -0.01, t = -2.17$) than for regulars, and reached significance also for the regulars ($\hat{\beta} = -0.006, t = -1.93, p = 0.03$, one-tailed test). A reanalysis of Experiment 1 did not reveal an effect for residualized form frequency. We therefore conclude that the cross-tense naming task invites subjects to optimize their performance, firstly, by betting a verb is a probable verb (resulting in the

U-shaped effect of lemma frequency), and secondly, by utilizing information in lexical memory about their familiarity with the inflected form (the facilitatory effect of residualized form frequency).

The next panel shows that an effect of inflectional entropy (*H*) was present for irregular, but not for regular verbs. No interaction of inflectional entropy by neighborhood density could be detected, nor any evidence for differentiation between the sexes. Whereas in same-tense naming the effect of inflectional entropy is small and generalizes across regulars and irregulars, its effect in cross-tense naming is restricted to irregulars, for which it is huge.

Verbs with alternating final obstruents elicited longer naming latencies than verbs with stable final obstruents, as can be seen in the lower right panel of Figure 3.4. There was a non-significant trend for this effect to be stronger for singulars.

Figure 3.5 visualizes the three-way interaction of the *Inflectional Paradigmatic Probability* with *Tense* and *Regularity*. There was no effect of *Inflectional Paradigmatic Probability* for regulars. Irregulars (black) elicited longer latencies than regulars (gray). When naming an irregular past tense given the present-tense form, a greater proportional frequency of the past-tense form led to longer naming latencies (black solid line). When naming a present tense form from an irregular past-tense form, a greater proportional frequency of the present-tense form afforded shorter latencies (dashed black line).

In order to understand the inhibitory effect of *Inflectional Paradigmatic Probability* in present-to-past naming for irregulars, consider that in this sub-experiment, in which subjects produce a past-tense form given a present-tense form, 75% of the forms they either read or say aloud have the unmarked present-tense stem. Half of the verbs are regular, they contribute 50% of the unmarked stems. The irregular forms read are also unmarked stems that participate in the person-number agreement system in the same way as the stems of regular verbs, and contribute 25%. In other words, the irregular past-tense stems are a 25% minority. Our hypothesis is that subjects optimize their task performance (speeded naming) for the unmarked stem across reading (regulars and irregulars) and production (regulars only). As a consequence, processing is delayed for the irregulars (as witnessed by the more than 40 ms difference in the intercepts of the solid lines in Figure 3.5). Furthermore, the greater the probability of the irregular past-tense form in its paradigm, the more this form works against the optimization strategy favoring regular stems, leading to longer naming latencies.

Past-to-present naming is more difficult than present-to-past naming, as witnessed by the elongated latencies for regulars (compare the gray dashed line in Figure 3.5 with the gray solid line). Optimizing performance for the unmarked stem (again, only 25% of the words in the experimental list contain an irregular stem form) now works to the advantage of irregulars. Since it is the unmarked stem that is to be pronounced, a greater proportional frequency of the present-tense form now results in facilitation.

In summary, the effect of *Inflectional Paradigmatic Probability* does not reflect normal paradigmatic processing, but points to a task artefact: Subjects optimize their performance by betting on having to produce the unmarked stem. We note here that in a given sub-experiment, subjects only performed cross-tense naming in one direction, from past to present, or from present to past. Hence, the present effects cannot be attributed to confusion as to the required direction of naming.

What does Experiment 2 contribute to our understanding of the controversy between theories with a sparse lexicon versus theories with a redundant lexicon? Advocates of the sparse lexicon might argue that Experiment 2 supports their theory, since paradigmatic effects (*Inflectional Entropy, Inflectional Paradigmatic Probability*) are present only for irregulars, exactly the forms that have to be listed even in the sparse lexicon. The absence of an effect of the latter measure in Experiment 1 for regulars also argues in favor of a sparse lexicon.

However, there are several deep problems with this line of reasoning. The effect of inflectional entropy in Experiment 1 provides support for the redundant lexicon. Furthermore, cross-tense naming is a much less natural task to perform than same-tense naming. Firstly, the task is more difficult to perform, as witnessed by an increase in mean naming latency from 429 ms in Experiment 1 to 653 ms in Experiment 2, combined with an error rate of 9.7% of wrong responses in Experiment 2, compared to error-free performance in Experiment 1. Secondly, subjects developed task strategies, as witnessed by the U-shaped effect of *Lemma Frequency* that was absent in Experiment 1, the privileging of the unmarked stem emerging from the interaction of the *Inflectional Paradigmatic Probability* with *Tense* for the irregulars in Experiment 2, and the huge effect of *Inflectional Entropy* for irregulars in Experiment 2. From debriefing the subjects participating in Experiment 2, we know that the task is felt to be difficult and even frustrating in the sense that it is impossible not to become confused and not to commit errors. Finally, the frequency of the inflected form was a highly significant predictor in Experiment 2, a finding that is totally at odds with the hypothesis of a sparse lexicon. We conclude that it is unlikely that the cross-tense naming paradigm would provide a good approximation to the comprehension and production of regular and irregular verbs.

We note here that the blocking of the materials in four sub-experiments for each combination of *Tense* and *Regularity* may have made it particularly difficult to observe the role of normal paradigmatic effects for regulars. In dual (or multiple) route models in which rote and rule operate in parallel (Baayen, Dijkstra and Schreuder, 1997; Kuperman, Schreuder, Bertram and Baayen, 2009), the repeated presentation of forms with the same number and tense may have shifted the normal balance of storage and computation to rule-driven processing, especially in Experiment 1, where no effect of residualized inflectional frequency was present.

General Discussion

The question addressed in this study is whether regular inflected words in languages such as Dutch and English crucially depend on rule-based processes operating on a sparse lexicon, or whether regular inflected words are available in a redundant lexicon and retrieved from memory in normal lexical processing.

There is evidence in our data that supports a sparse lexicon with rules deriving inflected forms from their simple bases. In the same-tense naming latencies of Experiment 1, no independent effect of the frequency of the inflectional forms was present. Furthermore, in the cross-tense naming data (Experiment 2), measures tapping into the paradigmatic properties of inflected words (*Inflectional Entropy*, and the *Inflectional Paradigmatic Probability*) were present only for irregulars and not for regulars. This would suggest that only irregular inflections are available in memory, and that regular inflections are generated and parsed on-line.

There is also evidence in our data supporting a redundant lexicon. An effect of inflectional entropy was present in same-tense naming (Experiment 1) for both regulars and irregulars. In Experiment 1, the effect of inflectional entropy was modulated by measures of neighborhood density, suggesting that the inflectional entropy measure reflects paradigmatic aspects of lexical competition. Furthermore, in Experiment 2, response latencies decreased significantly for more frequent regular inflectional forms when the lemma frequency of the base was controlled for. This indicates that the brain is sensitive to the probabilities of individual inflected forms, irrespective of their regularity.

Importantly, the cross-tense naming paradigm emerged as encumbered with strong task artefacts. Error rates were high, latencies were some 200 ms slower, lemma frequency showed a U-shaped functional relation to the naming latencies, and an inflectional frequency effect emerged even for regulars that was absent in the simple and straightforward task of same-tense naming. Furthermore, subjects optimized their performance for the naming of the present-tense stem forms, as witnessed by the interaction of the Inflectional Paradigmatic Probability by Tense for irregulars. Higher probabilities afforded shorter naming latencies when subjects named the present-tense form given the past-tense form, but longer naming latencies when the past-tense form had to be produced from the present-tense form. This inhibitory effect of the likelihood of the past-tense form must be a task effect, as one would expect facilitation in the production of the past-tense form the more frequent it is.

Experiments 1 and 2 made use of blocked designs, which probably reduced error rates, but may have provided subjects with excellent opportunities to optimize their response strategies.

Considering the combined evidence, theories allowing redundancy into the lexicon seem more appropriate. From a methodological perspective, it is extremely worrisome that the evidence for a dual mechanism model of the kind proposed by Pinker (1999) may be based on a task that is riddled with strategic, meta-linguistic effects. Neuroimaging studies such as Jaeger *et. al.* (1996) and Beretta (2003) reporting dissociations between regulars and irregulars may likewise be monitoring cognitive strategies for complying with a difficult and unnatural task, rather than the processes that characterize normal language comprehension and speech production.

To what extent are the results for the naming of Dutch regular inflections characteristic for normal lexical processing? All we can say here is that the simple word naming task probably approximates normal processing much better than cross-tense naming with presentation of only verbs, either in isolation, or in standardized sentence frames. A troubling methodological concern is that in order to obtain a proper understanding of the balance of storage and computation in speech production, large regression studies monitoring natural language use for unblocked, varied data will be essential. Developing experimental paradigms meeting these requirements presents a major challenge for future research.

Will the present results convince theorists working within the dual mechanism framework that a redundant lexicon plays a role in normal language processing? Probably not. The early version of this model (Pinker, 1999) argued that frequency effects would be restricted to irregulars. Pinker and Ullman (2002) relax this claim, stating that the words and rules model 'does not posit that regular forms are never stored, only that they do not have to be'. They then proceed to argue that inflectional frequency may contribute to the processing of regulars under various special circumstances, which include task demands, speaker-specific factors (such as the sex of the speaker), whether the regular verb resembles irregular verbs, etc. When all these special circumstances do not apply, then frequency effects for regulars need not be present. Given the availability of inflectional rules, frequency-dependent processing is not necessary, and increasingly unlikely the lower the frequency of the inflected form.

It is indeed plausible that there are circumstances under which frequency effects emerge more strongly. We have seen an example in the present study: a frequency effect for regular inflections emerged in the difficult cross-tense naming task, but not in the simple same-tense naming task. Speaker-specific properties may also play a role. The mixed models fitted to the latencies of Experiments 1 and 2 both included by-subject random intercepts for lemma frequency, as well as for various other covariates. Although studies such as Ullman *et al.* (2002) had led us to anticipate a group difference between males and females, there was no evidence for such a group difference in our data. Furthermore, although we did observe an effect of the similarity of a regular verb to irregular verbs (through the *RhymeCount* measure, but only in cross-tense naming, a task apparently inducing this effect), similarity did not interact with frequency in our data. Finally, the claim of Pinker and Ullman that frequency effects would vanish for lower-frequency words is misguided. Alegre and Gordon (1999) (cited by Pinker and Ullman) proposed a frequency threshold of 6 per million for English below which inflectional frequency effects would be absent. A large-scale regression analysis of the English Lexicon Project (Baayen, Wurm and Aycock, 2007), however, revealed frequency effects all the way down, with attenuation of the effect for higher-frequency words instead of lower-frequency words. It is lack of statistical power that is at issue, not the possibility that lower-frequency words would leave no traces in lexical memory.

Instead of positing that regular forms do not have to be stored (Pinker and Ullman), we take the opposite view, that regular forms must be stored in order to obtain optimal exemplar-driven generalization (see, e.g., for evidence from a machine-learning perspective Daelemans, Van den Bosch, Zavrel, 1999; Keuleers, Sandra, Daelemans, Gillis, Durieux, and Martens, 2007; Keuleers, 2008). From the perspective of word and paradigm morphology, moreover, the paradigmatic effects observed in the present data sets are expected. By contrast, the dual mechanism model, which is entirely syntagmatic in nature, is forced to assume that the observed paradigmatic effects are due to task demands.

In summary, do speakers typically produce regular inflected words such a *walked* by deriving them on-line from their stems by a regular rule of inflection attaching *-ed*? We think this is unlikely. Memory plays a much more pervasive role than anticipated by theories positing sparse lexicons, and may well involve both procedural memory for sequences of formatives in the case of regular past-tense forms and declarative memory for stems in the case of irregular past-tense forms. From this perspective, lexical processing of regulars and irregulars can be qualitatively different and nevertheless driven to a considerable extent by past experience.

References

Albright, A. and Hayes, B.: 2003, Rules vs. analogy in English past tenses: A computational/experimental study, *Cognition* **90**, 119–161.

Alegre, M. and Gordon, P.: 1999, Frequency effects and the representational status of regular inflections, *Journal of Memory and Language* **40**, 41–61.

Baayen, R. H.: 2007, Storage and computation in the mental lexicon, *in* G. Jarema and G. Libben (eds), *The Mental Lexicon: Core Perspectives*, Elsevier, pp. 81–104.

Baayen, R. H. and Moscoso del Prado Martín, F.: 2005, Semantic density and past-tense formation in three germanic languages, *Language* **81**, 666–698.

Baayen, R. H., Davidson, D. J. and Bates, D. M.: 2008, Mixed-effects modeling with crossed random effects for subjects and items, *Journal of Memory and Language* **59**, 390–412.

Baayen, R. H., Dijkstra, T. and Schreuder, R.: 1997, Singulars and plurals in Dutch: Evidence for a parallel dual route model, *Journal of Memory and Language* **37**, 94–117.

Baayen, R. H., Feldman, L. B. and Schreuder, R.: 2006, Morphological influences on the recognition of monosyllabic monomorphemic words, *Journal of Memory and Language* **55**, 290–313.

Baayen, R. H., Levelt, W., Schreuder, R. and Ernestus, M.: 2008, Paradigmatic structure in speech production, *in* M. Elliott, J. Kirby, O. Sawada, E. Staraki and S. Yoon (eds), *Proceedings Chicago Linguistics Society 43, Volume 1: The Main Session*, Chicago, pp. 1–29.

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

Baayen, R. H., Wurm, L. H. and Aycock, J.: 2007, Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities, *Mental Lexicon* **2**, 419–463.

Balling, L. and Baayen, R.: 2008, Morphological effects in auditory word recognition: Evidence from Danish, *Language and Cognitive Processes* **23 (7/8)**, 1159–1190.

Balling, L. and Baayen, R.: Manuscript submitted for publication, Evidence for morphologically complex words in a redundant lexicon.

Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D. and Yap, M.: 2004, Visual word recognition for single-syllable words, *Journal of Experimental Psychology:General* **133**, 283–316.

Beretta, A., Campbell, C., Carr, T., Huang, J., Schmitt, L. M., Christianson, K. and 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.

Bien, H.: 2007, *On the production of morphologically complex words. With special attention to frequency effects*, PhD thesis, Max Planck Institute for Psycholinguistics, Nijmegen.

Blevins, J. P.: 2003, Stems and paradigms, *Language* **79**, 737–767.

Bloch, B.: 1947, English verb inflection, *Language* **23**, 399–418.

Chomsky, N. and Halle, M.: 1968, *The sound pattern of English*, Harper and Row, New York.

Coltheart, M., Davelaar, E., Jonasson, J. T. and Besner, D.: 1977, Access to the internal lexicon, *in* S. Dornick (ed.), *Attention and performance*, Vol. VI, Erlbaum, Hillsdale, New Jersey, pp. 535–556.

Daelemans, W., Van den Bosch, A. and Zavrel, J.: 1999, Forgetting exceptions is harmful in language learning, *Machine learning, Special issue on natural language learning* **34**, 11–41.

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

De Vaan, L., Schreuder, R. and Baayen, R. H.: 2007, Regular morphologically complex neologisms leave detectable traces in the mental lexicon, *The Mental Lexicon* **2**(1), 1–24.

Ernestus, M. and Baayen, R.: 2007, Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch, *Language and Cognitive Processes* **22**, 1–24.

Hartshorne, J. and Ullman, M.: 2006, Why girls say’holded’more than boys, *Developmental Science* **9**, 21–32.

Hay, J.: 2001, Lexical Frequency in Morphology: Is Everything Relative?, *Linguistics* **39**, 1041–1070.

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

Kessler, B., Treiman, R. and Mullennix, J.: 2002, Phonetic biases in voice key response time measurements, *Journal of Memory and Language* **47**, 145–171.

Keuleers, E.: 2008, *Memory-based learning of inflectional morphology*, PhD thesis, University of Antwerp, Antwerp, Belgium.

Keuleers, E., Sandra, D., Daelemans, W., Gillis, S., Durieux, G. and Martens, E.: 2007, Dutch plural inflection: The exception that proves the analogy, *Cognitive Psychology* **54**, 283–318.

Kimura, D.: 2000, *Sex and cognition*, The MIT press, Cambridge, MA.

Kuperman, V., Schreuder, R., Bertram, R. and Baayen, R. H.: 2009, Reading of polymorphemic Dutch compounds: Towards a multiple route model of lexical processing, *Journal of Experimental Psychology: Human Perception and Performance* **35**, 876–895.

Levelt, W. J. M., Roelofs and Meyer, A.: 1999, A theory of lexical access in speech production, *Behavioral and Brain Sciences* **22**, 1–38.

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

Matthews, P. H.: 1974, *Morphology. An Introduction to the Theory of Word Structure*, Cambridge University Press, London.

Milin, P., Filipović Durdević, D. and Moscoso del Prado Martín, F.: 2009a, The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from serbian, *Journal of Memory and Language* pp. 50–64.

Milin, P., Kuperman, V., Kostić, A. and Baayen, R.: 2009b, Paradigms bit by bit: An information-theoretic approach to the processing of paradigmatic structure in inflection and derivation, *in* J. Blevins and J. Blevins (eds), *Analogy in Grammar: Form and Acquisition*, Oxford: Oxford University Press, pp. 214–252.

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

Moscoso del Prado Martín, F., Kostić, A. and Baayen, R. H.: 2004, Putting the bits together: An information theoretical perspective on morphological processing, *Cognition* **94**, 1–18.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R. and Baayen, R. H.: 2004, Morphological family size in a morphologically rich language:

The case of Finnish compared to Dutch and Hebrew, *Journal of Experimental Psychology: Learning, Memory and Cognition* **30**, 1271–1278.

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

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

Pinker, S. and Ullman, M.: 2002, The past and future of the past tense, *Trends in the Cognitive Sciences* **6**(11), 456–462.

Rumelhart, D. E. and McClelland, J. L.: 1986, On learning the past tenses of English verbs, *in* J. L. McClelland and D. E. Rumelhart (eds), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models*, The MIT Press, Cambridge, Mass., pp. 216–271.

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

Stemberger, J. and Middleton, C.: 2003, Vowel dominance and morphological processing, *Language and Cognitive Processes* **18**(4), 369–404.

Stemberger, J. P.: 2002, Overtensing and the effect of regularity, *Cognitive Science* **26**, 737–766.

Stemberger, J. P.: 2004, Phonological priming and irregular past, *Journal of Memory and Language* **50**, 82–95.

Tabak, W., Schreuder, R. and Baayen, R. H.: 2005, Lexical statistics and lexical processing: Semantic density, information complexity, sex, and irregularity in Dutch, *in* S. Kepser and M. Reis (eds), *Linguistic evidence — Empirical, Theoretical, and Computational Perspectives*, Mouton de Gruyter, Berlin, pp. 529–555.

Taft, M. and Forster, K. I.: 1976, Lexical storage and retrieval of polymorphemic and polysyllabic words, *Journal of Verbal Learning and Verbal Behavior* **15**, 607–620.

Ullman, M.: 2001, The declarative/procedural model of lexicon and grammar, *Journal of Psycholinguistic Research* **30**, 37–69.

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

Ullman, M. T., Estabrooke, I. V., Steinhauer, K., Brovetto, C., Pancheva, R., Ozawa, K., Mordecai, K. and Maki, P.: 2002, Sex differences in the neurocognition of language, *Brain and Language* **83**, 141–143.

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

Walenski, M. and Ullman, M.: 2005, The science of language, *The Linguistic Review* **22**, 327–346.

Producing inflected verbs: A picture naming study

Chapter 4

This chapter has been published as Wieke Tabak, Robert Schreuder, and R. Harald Baayen (2010). Producing inflected verbs: A picture naming study. *The Mental Lexicon* 5(1), 22-46.

Abstract

Four picture naming experiments addressing the production of regular and irregular past-tense forms in Dutch are reported. Effects of inflectional entropy as well as effects of the frequency of the past-tense inflected form across regulars and irregulars support models with a redundant lexicon while challenging the dual mechanism model (Pinker, 1997). The evidence supports the hypothesis of Stemberger (2004) and the general approach of Word and Paradigm morphology (Blevins, 2003) according to which inflected forms are not derived from the present-tense stem, but accessed independently.

Introduction

Ever since Bloch (1947) and Chomsky and Halle (1968), inflected forms have been understood as being created on-line from their base words by inflectional rules in language comprehension and speech production. This conceptualization reflects how pedagogical grammars proceed from simpler to more complex forms. However, Stemberger and colleagues (Stemberger, 2002; Stemberger and Middleton, 2003; Stemberger, 2004) suggested that inflected forms are not derived in this generative sense. Their evidence, based on speech errors, suggests that in English present and past tense forms are both activated and enter into a competition process that is modulated both by the long-term probabilities of the stem vowels as well as by short-term influences exerted by vowels in preceding words.

If inflection is not driven by computational rules, this has far-reaching consequences for the dual mechanism model proposed by, e.g., Pinker and Prince (1988), Pinker (1991), Pinker and Prince (1994), Pinker (1999) and Pinker and Ullman (2002) for inflection. According to the version of this model formulated in Ullman, Estabrooke, Steinhauer, Brovetto, Pancheva, Ozawa, Mordecai and Maki (2002) and Ullman (2004), irregular verbs are stored in declarative memory, while regular verbs are processed by a procedural memory encoding rules of inflection that parse inflected forms into their constituents in comprehension, or that put together inflected forms from their constituents in production. As pointed out by Stemberger (2004), a model in which present-tense and past-tense forms are in competition is also problematic for many connectionist models adopting the assumption that marked forms are derived from less marked forms (e.g. Rumelhart and McClelland, 1986; MacWhinney and Leinbach, 1991; Plunkett and Marchman, 1993; Plunkett and Juola, 1999, but see the triangle model of Seidenberg and Gonnerman, 2000, for a non-derivational connectionist approach).

On the other hand, the approach advocated by Stemberger is very much in the spirit of a line of research in linguistic morphology that has not received much attention from psychologists, word and paradigm morphology (see, e.g. Matthews, 1974; Aronoff, 1994; Beard, 1995; Blevins, 2003). Word and paradigm morphology takes the word, rather than sub-word formatives, to be the basic meaningful unit of a language, and focuses on the paradigmatic organization of inflected words. Recent experimental work has provided experimental evidence for the relevance of paradigmatic organization for lexical processing. Across a series of experiments (Baayen, Feldman and Schreuder, 2006; Milin, Filipović Durdević, and Moscoso del

Prado Martín, 2009; Milin, Kuperman, Kostić and Baayen, 2009; Tabak, Schreuder and Baayen, 2005), entropy measures gauging the amount of information carried by a paradigm have been found to be robust predictors of response latencies in visual comprehension studies. Rule-based models with a sparse lexicon listing only the irreducably irregulars are severely challenged by these data. If experience with inflected forms would leave no trace in declarative memory, as the dual mechanism model would have us believe, then entropy measures calculated from the probability of a word's inflected forms should not be predictive for lexical processing, contrary to what is actually observed.

Thus far, relatively little is known about the role of paradigmatic structure in speech production. Baayen, Levelt, Schreuder and Ernestus (2008) reported for Dutch an inhibitory effect of inflectional entropy in a picture naming task in which subjects were presented with line drawings with one or with two exemplars of objects, and were requested to say aloud the singular or the plural form. When the singular and the plural form have more similar frequencies, the entropy is large, and response latencies become large as well. When the two forms are more dissimilar in frequency, response times are shorter.

Tabak, Schreuder and Baayen (submitted) observed effects of inflectional entropy for word naming in Dutch and English, in interaction with measures of neighborhood density. When subjects were requested to name the past-tense form when presented with the present-tense form, or the present-tense form when shown the past-tense form (cross-tense naming), the paradigmatic entropy entered into a negative correlation with the naming latencies, but significantly so only for irregular verbs.

For the study of speech production, naming paradigms are suboptimal because they have a substantial comprehension component. This component is so clearly present that simple word naming is often used as a task to gauge visual comprehension, complementing the visual lexical decision task. Cross tense word naming seeks to avoid giving away what the participant has to say by asking the subject to change the tense of the stimulus. However, participants are still provided with language input, and are required to perform a meta-linguistic task which, as we found out, leads to the development of task-specific strategies (Tabak, Schreuder and Baayen, submitted).

A task that is not contaminated by the reading process is the picture naming task. This task engages the full production process, from conceptualization to articulation. The goal of the present study is to ascertain to what extent the probabilities of inflectional forms can be detected when this ecologically more valid task is used. Effects of the probabilities of regular inflected forms (estimated by their relative frequency) as well as effects of the paradigmatic entropy (estimated from the distribution of relative frequencies in a word's inflectional paradigm) would support the non-derivational view of inflection proposed by Stemberger (2004) in psychology and by Word and Paradigm morphology in linguistics.

The following four experiments compare picture naming for present tense forms (experiments 1 and 3) and past tense forms (experiments 2 and 4), at the same time contrasting picture naming without (experiments 1 and 2) and with (experiments 3 and 4) prior familiarization with the intended picture names. We anticipated that in naming without prior familiarization, processes related to the interpretation of the picture and the retrieval of an appropriate picture name might dominate, while in naming with prior familiarization, processes related to word form encoding might be more prominent.

Experiment 1: Unprepared present-tense picture naming

Method

Materials

The materials consisted of photographs of a young woman enacting verbs for position and motion. Photographs were commissioned for all picturable verbs available in the set of 286 Dutch verbs studied by Tabak, Schreuder and Baayen (2005) using visual lexical decision and Tabak *et. al.* (submitted) using word naming. Photographer and actress were instructed to minimize variability between pictures by using the same background and a highly restricted set of ancillary objects. Examples are shown in Figure 4.1. A total of 170 photographs was obtained, of which 85 depicted regular verbs, and 85 irregular verbs. Verbs ranged in log *Lemma Frequency* (the frequency of the verb across all its inflected variants) from 4.5 to 11.3 in a 42 million word corpus (median 8.0), and ranged in *Length* (in phonemes) from 2 to 6 (median 4). The irregular verbs were of slightly higher frequency than the regular verbs 8.2 versus 7.8, ($t(153.9) = 1.9076, p = 0.06$). The complexity of the pictures, henceforth *Picture Complexity*, evaluated in terms of the size (in kilobytes) of their jpg files, differed between the regulars (mean 93.6) and irregulars (mean 71.9, $t(153.4) = -6.06, p < 0.0001$). This difference mirrors the

Figure 4.1: Stimuli for *walking* (left) and *sitting* (right).

verbal report we received from the photographer and actress that the regular verbs were more difficult to depict than the irregular verbs.

Participants

Ten women and seven men, all students at the University of Nijmegen, participated in this experiment. They all had normal or corrected-to-normal vision. Subjects received 5 Euro for each experiment.

Procedure

Participants were tested individually in a noise attenuated experimental booth. They received standard picture naming instructions, specifying that they had to name the presented picture as quickly and accurately as possible. Naming latencies were registered with a Sennheiser microphone placed at a distance of 20 cm from the participant.

Each picture was preceded by *Vandaag ...* ('Today ...') displayed on the computer screen. We asked the participants to complete the prompt with a clause consisting of the verb form followed by the third person pronoun, e.g., *loopt ze* ('she walks') for the photograph depicting walking. This prompt was presented in the center of the screen for 1000 ms. After 50 ms, the photograph was shown, in portrait mode, using the full vertical dimension of the computer screen. Photographs remained on the screen for 3000 ms. A new trial was initiated 500 ms afterwards. Each participant was presented with the pictures in a different random order.

There were three short breaks during the experiment. The total duration of an experimental session was approximately 45 minutes.

Results

We removed data points with responses where the voicekey was triggered by vocalizations other than those of the onset of a real word (1.1% of the observations). The distribution of the remaining latencies was highly skewed. We reduced this skewness by means of a logarithmic transformation, which outperformed the inverse transform. We fitted a linear mixed-effects model to the data with participant and verb as crossed random effects. Factors were modeled with contrast coding. The residuals of this model showed marked deviations from normality. We therefore removed data points with absolute standardized residuals exceeding 2.5, and refitted the model. The qualitative pattern in this trimmed model is the same, but the estimates of the coefficients and their standard deviations are more precise. The same procedure was followed in the analyses of the other three experiments discussed below. For all analyses, we used a stepwise variable elimination procedure to obtain the most parsimoneous model providing a close fit to the data. Predictors that did not reach significance were removed from the model specification. Many two-way interactions, and occasionally three-way interactions, were examined. Again, only those were retained that reached significance. For the present experiments, in which many new or not well-established predictors are considered, we reasoned that a conservative strategy allowing the model becoming clogged with superfluous main effects and interactions would be counterproductive. In the tables and figures presented below, only those predictors and interactions are presented that reached significance. In other words, tables and figures represent the minimally adequate models that we fitted to the data. Tables (and figures) should therefore not be consulted as specifying for the full set of predictors and their interactions whether they were significant.

As predictors, we considered several measures in addition to *Lemma Frequency*, *Picture Complexity*, and *Regularity*. First, two control variables were included: *Trial* (the index of the item in the experimental list), and the response latency to the preceding trial (*Previous RT*). The former measure allows us to explicitly account for effects of habituation or fatigue. The second measure is necessary to account for potential dependencies between the successive trials in the experiment (see, e.g., De Vaan, Schreuder and Baayen, 2007; Balling and Baayen, 2008; Tabak *et. al.*, submitted). Counterbalancing nullifies the adverse effects of non-independencies in the sequence of trials. By including these two control measures, we remove at least part of the variance associated with this non-independence from the error term, and thereby obtain more precise models.

Supplementing *Lemma Frequency*, we included as a predictor the frequency of the present-tense form. As this frequency is highly correlated with *Lemma Frequency* ($r = 0.92$), we orthogonalized it by replacing it by the residuals of a linear model regressing form frequency on lemma frequency. The residualized measure was positively correlated with the original frequency ($r = 0.35, p < 0.0001$) and represents the inflectional form's frequency in so far as that frequency cannot be predicted from lemma frequency.

A related predictor was the verb's *inflectional entropy* (Baayen, Feldman and Schreuder, 2006; Milin, Kuperman, Kostic and Baayen, 2009; Tabak *et. al.*, submitted), the amount of information carried by the verb's paradigm. It was estimated as

$$H_i = \sum_j p_j \log_2 p_j, \tag{4.1}$$

with j ranging over the different phonologically distinct inflectional variants of the verb, and p_j representing the relative frequency of that inflectional variant in its paradigm.

In addition to the factor *Regularity*, distinguishing between regular and irregular verbs, we included the factor *Sex*, distinguishing between female and male speakers, as Ullman *et. al.* (2002) reported for English that females but not males would have representations in declarative memory for regular past-tense inflections.

As subjects received no initial instructions on what names to use for the pictures, the names actually produced included many that were not targeted. We refer to trials for which the targeted form was produced as “correct” trials (61% of responses) and to the untargeted forms as “incorrect” (39% of the responses). As a measure of uncertainty about how to name the picture, we calculated the entropy of the frequency distribution of names elicited by a given picture,

$$H_p = \sum_k p_k \log_2 p_k, \tag{4.2}$$

with k ranging over the different names produced, and p_k the relative frequency with which name k was produced in the experiment. In what follows, we refer to this entropy measure as *Picture Entropy*. For all unintended, “incorrect” responses, lemma frequency, present-tense form frequency, inflectional entropy, and length in phonemes were determined, so that distributional measures were always tied to the form actually produced. In our statistical analysis, we build on the robustness of linear mixed-effects models with respect to unequal numbers of observations (see, e.g., Baayen, Davidson and Bates, 2008): words with fewer replicates contribute less weight to the model's estimates. By including the untargeted responses, we minimize data loss.

A predictor that we also considered specifies whether a verb has a stem-final obstruent that alternates between voiced (in, e.g., the infinitive: *schrijven*, 'to write') and voiceless (in, e.g., the singular present-tense forms: *schrijft*, 'you, he, she writes'). For experimental studies addressing this alternation, the reader is referred to Ernestus and Baayen (2003), Ernestus and Baayen (2004), Ernestus and Baayen (2007). Tabak *et. al.* (submitted) observed that whether a verb has an alternating obstruent may affect response latencies in word naming. We therefore included the factor *Alternating*, with levels 'alternating' and 'non-alternating', as a predictor in our analyses.

In addition, across all four experiments, we considered whether measures of neighborhood density (the N-count measure, as well as the positional neighborhood measures explored by Bien, Levelt and Baayen (2005)) might help explain the variance in naming latencies. However, none of these measures ever reached or approached significance, not as simple main effects, nor as participants in interactions. In what follows, they will therefore not be discussed further. Similarly, the length (in phonemes) of the target word never reached significance, and is not mentioned in the analyses.

The model fitted to the data of Experiment 1 is summarized by Table 4.1, which lists the coefficients of the fixed-effect factors and covariates, and Figure 5.1. The standard deviation for the by-subject random intercepts was 0.088, that for the by-item random intercepts was 0.158, and the standard deviation for the by-observation noise was 0.241.

Panel A shows that as subjects proceeded through the experiment, they responded more slowly. Panel B illustrates that more complex pictures elicited longer latencies, as expected. As shown in panel C, subjects who responded more slowly to a preceding picture tended to take longer for responding to the current trial, exactly as observed in previous studies examining the non-independence of adjacent trials in other chronometric tasks (De Vaan *et. al.*, 2007; Balling and Baayen, 2008; Tabak *et. al.*, submitted).

The interaction of *Correctness* by *Sex* shown in panel D indicates that males

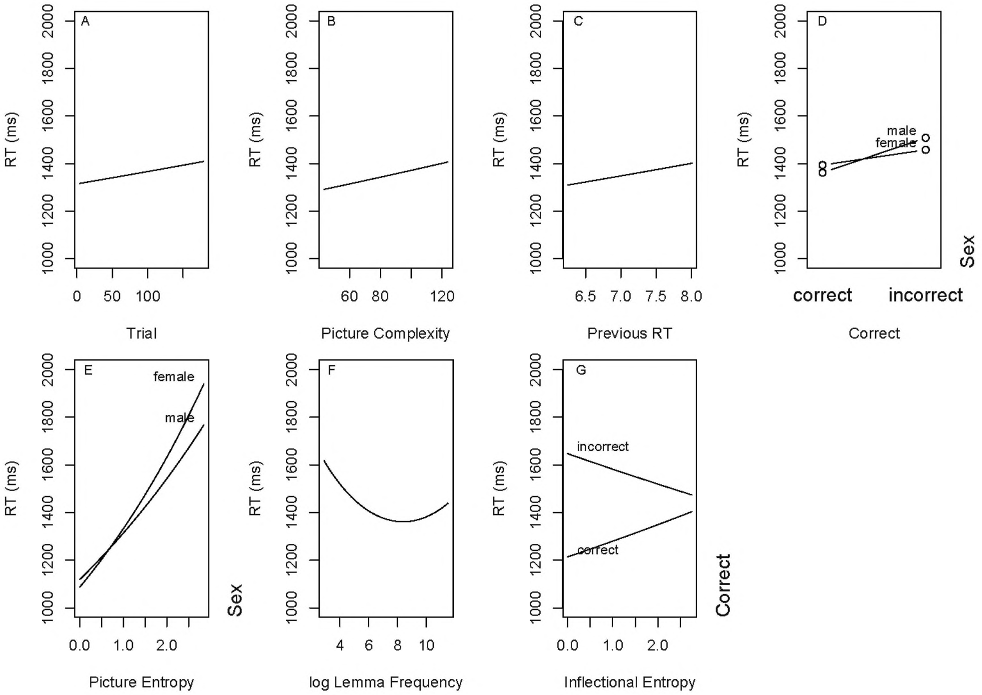

Figure 4.2: Partial effects of the predictors for the unprepared present-tense naming latencies in Experiment 1.

	Estimate	lower HPD95	upper HPD95	p (MCMC)
Intercept	6.8938	6.5028	7.2891	0.0001
Trial	0.0004	0.0002	0.0006	0.0004
Previous RT	0.0384	0.0087	0.0744	0.0152
Picture Entropy	0.1603	0.1249	0.2009	0.0001
Picture Complexity	0.0011	0.0001	0.0020	0.0190
Correct=incorrect	0.3047	0.0958	0.4457	0.0030
Sex: female	-0.0292	-0.1239	0.0666	0.5262
Inflectional Entropy	0.0526	-0.0178	0.0868	0.1992
Lemma Frequency (linear)	-0.0959	-0.1683	-0.0271	0.0118
Lemma Frequency (quadratic)	0.0057	0.0015	0.0103	0.0132
Correct=incorrect : Sex=female	-0.0573	-0.1048	-0.0045	0.0296
Picture Entropy : Sex=female	0.0429	0.0076	0.0732	0.0138

Table 4.1: Coefficients in the mixed-effects model fit to the picture naming latencies of Experiment 1 (unprepared present-tense naming). Upper/lower HPD95: 95 percent credible intervals based on 10,000 Markov chain Monte Carlo samples from the posterior distribution of the parameters.

required more time than females when responding with an untargeted, “incorrect” picture name. Panel E illustrates a second difference between the sexes: The effect of *Picture Entropy*, the uncertainty (or amount of information) carried by the picture was greater for females than for males.

Panel F reports a U-shaped effect for *Lemma Frequency*. Such a U-shaped frequency effect has been reported previously for language production by Bien *et. al.* (2005) and Tabak *et. al.* (submitted).

Panel G illustrates the effect of *Inflectional Entropy*. For targeted names (‘correct’), accessing informationally more complex paradigms required more time, replicating Baayen *et. al.* (2008). For untargeted verbs, the pattern reversed, such that alternative names were selected more quickly as the amount of information carried by their inflectional paradigms increased. This suggests that information-rich paradigms may constitute attractors competing with the target verbs for selection in production.

Experiment 2: Unprepared past-tense picture naming

Method

Materials

The materials were the same as those of Experiment 1.

Participants

Ten women and seven men completed Experiment 2. None of them had participated in Experiment 1. They all had normal or corrected-to-normal vision.

Procedure

The procedure was identical to that of Experiment 1, except that subjects were now asked to produce simple past-tense forms. Therefore, the prompt used in Experiment 1 (*vandaag*, 'today') was replaced by a prompt for past-tense forms, *gisteren*, 'yesterday'. We asked participants to complete the phrase with the appropriate verb form, followed by the pronoun *ze*, 'she', as in *gisteren liep ze*, 'yesterday she walked'.

Results

Table 5.5 and Figure 5.2 present the results of a mixed-effects model fitted to the naming latencies of Experiment 2. As for Experiment 1, untargeted verbs (41% of the trials) were included in the analysis, with predictors such as lemma frequency, past tense frequency, and inflectional entropy recalculated for these forms. Standard deviations for the by-subject random intercepts, the by-item random intercepts, and the by-observation noise were 0.117, 0.140, and 0.260 respectively.

As in Experiment 1, naming latencies increased as subjects proceeded through the experiment, as shown in panel A of Figure 5.2. Panel B replicates the interaction of *Picture Entropy* by *Sex*. Again, the inhibitory effect of *Picture Entropy* was stronger for females than for males. Panel C indicates that *Lemma Frequency* had a U-shaped effect, exactly as in Experiment 1.

Verb forms with a greater past-tense inflectional frequency elicited shorter naming latencies, as illustrated in panel D. Furthermore, the larger the number of irregular verbs with the same rhyme (in the past tense) as the verb produced, the shorter the naming latency was (panel E; this predictor was also examined for Experiment 1, but was not significant).

	Estimate	lower HPD95	upper HPD95	p (MCMC)
Intercept	7.3860	7.1978	7.6560	0.0001
Trial	0.0007	0.0005	0.0009	0.0001
Sex=female	-0.0731	-0.1897	0.0472	0.2222
RhymeCount	-0.0408	-0.0610	-0.0178	0.0004
Past Tense Frequency	-0.0311	-0.0488	-0.0148	0.0010
Picture Entropy	0.1576	0.1209	0.1907	0.0001
Lemma Frequency (linear)	-0.0758	-0.1398	-0.0336	0.0022
Lemma Frequency (quadratic)	0.0049	0.0022	0.0090	0.0018
Picture Entropy : Sex=female	0.0466	0.0146	0.0771	0.0024

Table 4.2: Coefficients in the mixed-effects model fit to the picture naming latencies of Experiment 2 (unprepared past-tense naming)

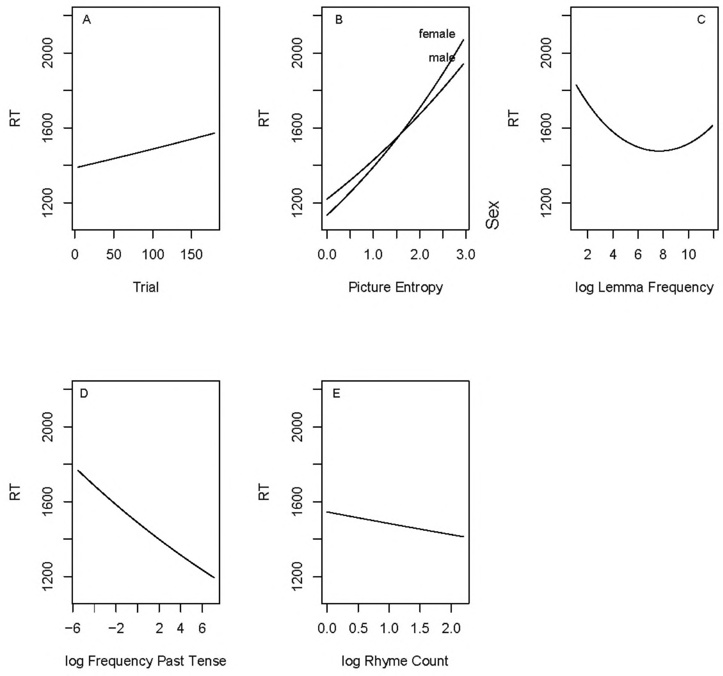

Figure 4.3: Partial effects of the predictors for the unprepared past-tense naming latencies in Experiment 2.

Experiment 3: Prepared present-tense picture naming

Materials The materials were the same as in Experiment 1.

Participants The participants were the same as those participating in Experiment 1.

Procedure After completing Experiment 1, participants were familiarized with the intended, targeted picture names by taking them through a picture book that printed the infinitive below each photograph. All that we asked the participants to do is silently read the printed words, no overt response was elicited. Following this familiarization phase, we carried out a second naming experiment in which participants were asked to use the targeted picture names. We refer to this second version of the experiment as *prepared naming*.

Results

The statistical analysis of the data proceeded along the same lines as for Experiments 1 and 2. The familiarization phase reduced the percentage of untargeted responses from 39% (in Experiment 1) to 16%. Untargeted, incorrect responses were removed from the data set. Table 4.3 and Figure 4.4 summarize the results of a mixed-effects model fitted to the correct, targeted responses. The standard deviation of the by-subject random intercepts was 0.118. Random slopes for (centralized) *Lemma Frequency* also reached significance (standard deviation: 0.009) in a likelihood ratio test ($\chi_2^{(2)} = 6.46, p = 0.039$). The standard deviation of the by-item random intercepts was 0.118, and that of the by-observation noise was 0.217.

Table 4.3 presents the coefficients of the fitted model, and Figure 4.4 illustrates the partial effects of the predictors. *Trial* again emerged as inhibitory (panel A). As expected given Experiment 1, naming latencies increased with *Picture Complexity* (panel B). *Picture Entropy* was likewise inhibitory, this time without an interaction with *Sex* (panel C). *Lemma Frequency* emerged as inhibitory (panel D). There was significant by-subject variation in the slope of this frequency effect, as mentioned above in the discussion of the random-effects structure of the model. Inspection of the by-subject slopes revealed that all estimated slopes were positive.

The inflectional frequency of the present-tense form, which did not reach significance in Experiment 1, emerged in a significant interaction with *Sex*. As can be seen in panel E, words with greater inflectional frequency were responded to faster by males, but slower by females.

Words with final obstruents alternating with respect to their voice specification

	Estimate	lower HPD95	upper HPD95	p (MCMC)
Intercept	6.8398	6.7356	6.9504	0.0001
Trial	0.0003	0.0001	0.0005	0.0010
Picture Entropy	0.1784	0.1428	0.2121	0.0001
Alternating=TRUE	0.0766	0.0381	0.1172	0.0002
Picture Complexity	0.0011	0.0004	0.0018	0.0038
Lemma Frequency (centered)	0.0306	0.0174	0.0442	0.0001
RhymeCount	-0.0059	-0.0104	-0.0016	0.0084
Present Tense Frequency	-0.0170	-0.0525	0.0164	0.3036
Sex=female	-0.0605	-0.1735	0.0464	0.2566
Present Tense Frequency : Sex=female	0.0518	0.0218	0.0820	0.0014

Table 4.3: Coefficients in the mixed-effects model fit to the picture naming latencies of Experiment 3 (prepared present-tense naming)

elicited longer naming latencies (panel F). Panel G depicts the facilitation from the *RhymeCount* measure: verbs that rhyme in the present tense with many other irregular verbs elicited shorter latencies.

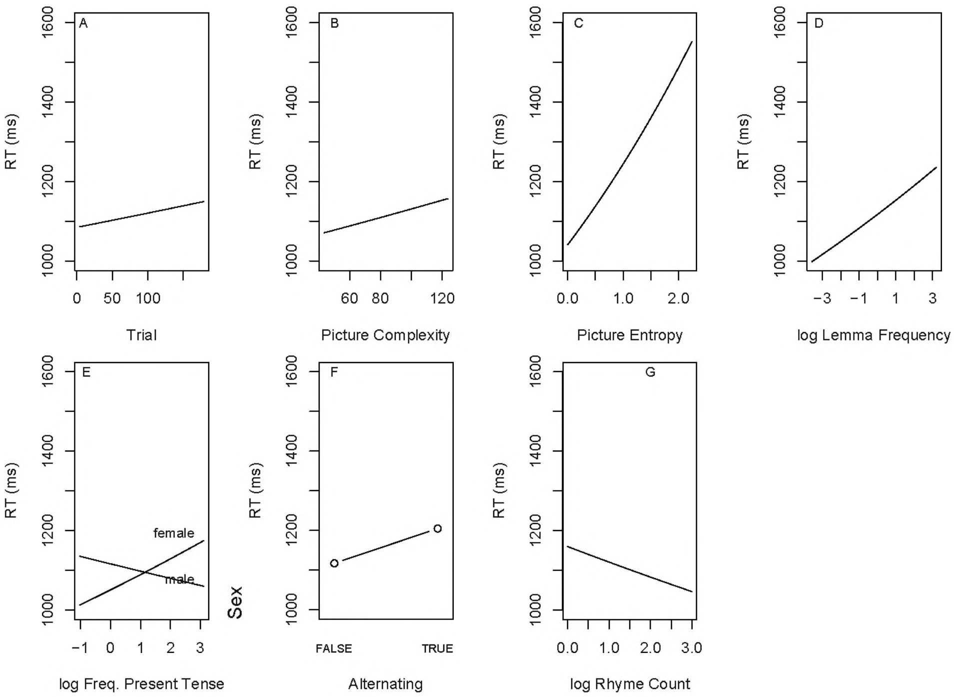

Figure 4.4: Partial effects of the predictors for the prepared present-tense naming latencies in Experiment 3. *Lemma Frequency* was centered.

Experiment 4: Prepared past-tense picture naming

Method

Materials

The materials were identical to those used in Experiments 1–3.

Participants The participants were the same as those participating in Experiment 2.

Procedure After completing Experiment 2, participants were familiarized with the targeted picture names by taking them through a picture book that printed the infinitive below each photograph. Following this familiarization phase, we carried out a second, prepared naming experiment in which subjects were requested to use the targeted picture names.

Results

We analyzed the data in the same way as for Experiment 3, focusing on the targeted, correct responses, and excluding the untargeted responses (15.1%) from consideration. Table 4.4 and Figure 4.4 provide an overview of the model. Standard deviations for the random intercepts for subject, item, and the by-observation noise were 0.079, 0.124, and 0.248 respectively.

The inhibitory effect of *Trial* observed for the preceding three experiments was again present in Experiment 4 (panel A). For regulars, but not for irregulars, longer latencies at preceding trials predicted longer latencies at the current trial (*Previous RT*, panel B). *Picture Entropy* (panel C) was again inhibitory, as in all three preceding experiments. No interaction with *Sex* could be observed. Panel D shows the U-shaped effect of *Lemma Frequency* familiar from Experiments 1 and 2. The frequency of the past-tense inflected form was facilitatory (panel E). There was no trace of interactions with *Sex* or *Regularity*. Panel F illustrates the interaction of *Inflectional Entropy* by *Sex*. For males, but not for females, a greater *Inflectional Entropy* afforded significantly shorter naming latencies. Finally, verbs with alternating final obstruents elicited longer naming latencies (panel G), but this difference was modulated by *Previous RT*, such that for larger preceding response latencies the effect of *Alternating* diminished.

	Estimate	lower HPD95	upper HPD95	p (MCMC)
Intercept	7.1075	6.6139	7.5584	0.0001
Trial	0.0005	0.0003	0.0007	0.0001
Regularity=regular	-0.5060	-1.0148	-0.0129	0.0388
Sex=female	-0.2079	-0.3627	-0.0606	0.0062
Picture Entropy	0.1597	0.1250	0.1958	0.0001
Lemma Frequency (linear)	-0.1163	-0.1865	-0.0501	0.0022
Lemma Frequency (quadratic)	0.0092	0.0047	0.0137	0.0002
Alternating=TRUE	0.6652	0.1433	1.2079	0.0124
Previous RT	0.0519	-0.0021	0.1081	0.0630
Inflectional Entropy	-0.0768	-0.1502	-0.0067	0.0290
Past Tense Frequency	-0.0238	-0.0469	0.0004	0.0520
Previous RT : Alternating=TRUE	-0.0809	-0.1614	-0.0124	0.0278
Inflectional Entropy: Sex=female	0.0636	0.0082	0.1252	0.0318
Previous RT: Regularity=regular	0.0770	0.0093	0.1479	0.0214

Table 4.4: Coefficients in the mixed-effects model fit to the picture naming latencies of Experiment 4 (prepared past-tense naming)

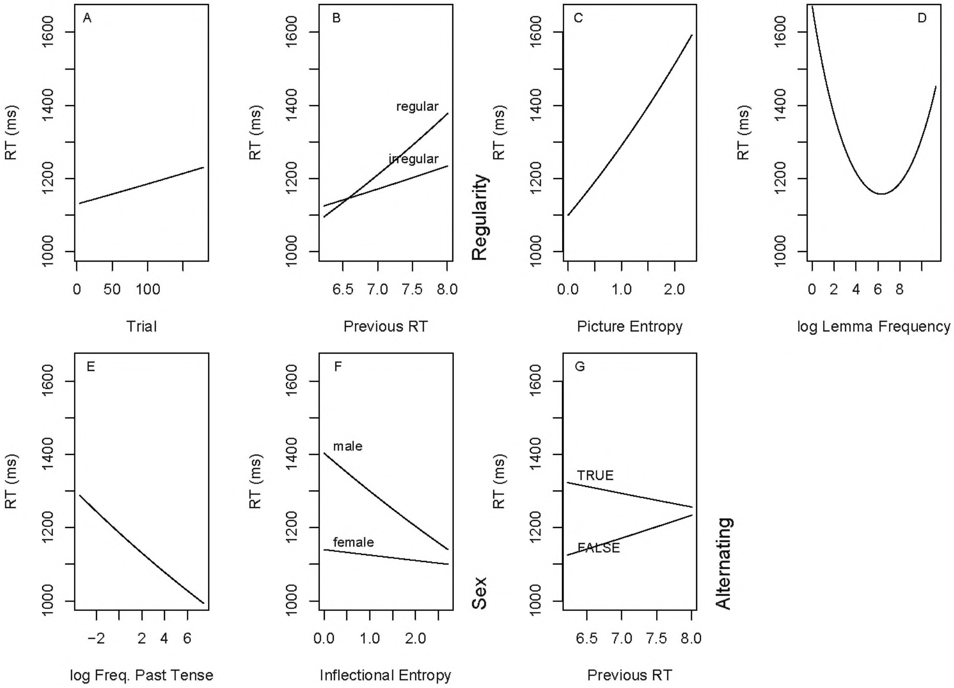

Figure 4.5: Partial effects of the predictors for the prepared past-tense naming latencies in Experiment 4.

General Discussion

Table 4.5 presents a synopsis of the effects observed. The main patterns in this table are supported by an analysis of all four experiments jointly, summarized in Table 4.6. This overall analysis offers the advantage of increased power and the possibility to evaluate for which predictors and interactions elimination from the individual model specifications was too conservative, as well as for which predictors support was restricted to just a single data set. The disadvantage of the overall analysis is that the data set is somewhat unbalanced, as the data elicited for prepared naming excluded untargeted responses.

Consistent across all four experiments are the increasing naming latencies as the experiment proceeds (*Trial*), and the elongated latencies that come with more ambiguous pictures (*Picture Entropy*).

Table 4.5 shows that in present-tense naming, but not in past-tense naming, more complex pictures (*Picture Complexity*) elicited significantly longer latencies, according to the individual analyses. Table 4.6 suggests that there is an inhibitory effect for past-tense naming as well, but with a coefficient that is half that for present-tense naming. The interaction of *Picture Complexity* by *Tense* is probably due to reduced morphological processing load for present-tense compared to past-tense naming, and a concomitant reduction in the noise masking the early stages of picture interpretation.

Consistent with earlier studies, *Previous RT*, when it emerged in the individual analyses (Experiments 1 and 4), had a positive slope, indicating that there is local coherence in the speed with which subjects respond in chronometric tasks. The joint analysis supports this temporal coherence effect across all four experiments.

The analyses of the individual experiments revealed a significant effect of *Alternating* only for prepared naming. The joint analysis supports delayed latencies for words with alternating obstruents across all four experiments, in interaction with *Previous RT* (as in Experiment 4). This suggests that words with more variable morphophonology are at a disadvantage in speech production. This disadvantage is strongest when subjects are naming the pictures quickly, as indexed by *Previous RT*, and decreases where they go through the experiment more slowly. Apparently, selecting the correct phonological form slows processing only when the choice between alternatives has to be made rapidly.

When the count of rhyming irregulars was significant (Experiments 2 and 3), it was facilitating. In Experiment 2 (present-tense naming), we counted the number of irregular verbs sharing the present-tense rhyme, for Experiment 3 (past-tense naming), the *RhymeCount* is based on the number of irregular verbs sharing the same rhyme in the past tense. This effect, which in the joint analysis received full support across all four experiments, is probably best interpreted as a facilitatory neighborhood density effect.

The effect of neighborhood density in speech production is somewhat unclear. Vitevitch (2002) observed facilitation for English, but Vitevitch and Stamer (2006) were confronted with inhibition for Spanish. Bien, Levelt and Baayen (2005) found an inverse U-shaped effect of neighborhood density in the production of compounds in Dutch, and Tabak *et. al.* (submitted) observed inhibition for neighbors differing only in their first phoneme, but facilitation for neighbors differing at later positions. In the present experiments, neighborhood density measures consistently failed to reach significance. One possible reason is that the picture naming task is dominated by conceptual processing and the retrieval of the picture name. The effect of neighborhood density would then presumably arise during later stages of phonological encoding. Another possible reason is that measures of neighborhood similarity based on a single mismatching segment may be too coarse. By contrast, the *RhymeCount* measure effectively defines neighbors on the basis of a mismatching higher unit, the onset of the stem syllable. If this interpretation is correct, it supports the hypothesis of Vitevitch (2002) that in speech production neighborhood similarity is facilitatory.

In three out of four experiments, *Lemma Frequency* emerged with a U-shaped effect. In the context of experiments 1, 2, and 4, the linear inhibitory effect of *Lemma Frequency* in Experiment 3 (prepared present-tense naming) is exceptional. However, the joint model fitted to all four experiments jointly suggests that the coefficient for the quadratic term of *Lemma Frequency* for Experiment 3 does not differ significantly from the corresponding coefficients for the other Experiments. This suggests that the true *Lemma Frequency* effect in Experiment 3 is also U-shaped, and warns against overinterpreting the *Lemma Frequency* effect in Experiment 3 as special.

Following Tabak *et. al.* (submitted), who observed the same U-shaped pattern for word naming, we hypothesize that this U-shaped effect arises as a consequence of subjects optimizing their performance for verbs with the most likely, ‘central’, lemma frequencies. The more exceptional a lemma frequency is in the experiment, irrespective of whether it is large or small, the longer the response latencies are. In other words, if this hypothesis is on the right track, we are observing the brain’s response not to a given lemma frequency as such, but to the probability

Predictor	Exp. 1 present unprepared	Exp. 2 past unprepared	Exp. 3 present prepared	Exp. 4 past prepared
Trial	+	+	+	+
Picture Entropy	+ *Sex*	+ *Sex*	+	+
Picture Complexity	+ *Sex*		+	
Alternating			+	+− *Previous RT*
Previous RT	+			+ *Regularity, Alternating*
Rhyme Count		−	−	
Correctness	+− *Sex*			
Lemma Frequency	U	U	+	U
Form Frequency		−	+ *Sex*	−
Inflectional Entropy	− *Correct*			+− *Sex*
Regularity				− *Previous RT*

Table 4.5: Overview of predictors by experiment. +: positive slope, −: negative slope, U: U-shaped effect; interactions with *Sex*, *Regularity*, and *Correctness* are indicated where present.

of that lemma frequency. Such a higher-level response would be consistent with task-specific optimization, with optimization for general lexical processing in speech production, or with both.

In order to evaluate the effect of *Inflectional Entropy*, we first note that it has been found to be facilitatory in comprehension (Baayen, Feldman and Schreuder, 2006) but inhibitory in picture naming (Baayen, Levelt, Schreuder and Ernestus, 2008). In the naming experiments reported by Tabak *et. al.* (submitted), the effect of *Inflectional Entropy* was modulated by neighborhood density. No such modulation could be observed in the present picture naming experiments. In our picture naming experiments, the evidence for a role for *Inflectional Entropy* is somewhat fragmented across the individual experiments. The joint analysis suggests facilitation for untargeted responses, and inhibition for regular verbs. A three-way interaction in the joint analysis of *Sex* by *Inflectional Entropy* by *Preparedness*, just missing significance ($t = 1.98$), provides modest support for the interaction with *Sex* observed in Experiment 4, with stronger inhibition for females. The main trend emerging from the present data is that in picture naming *Inflectional Entropy* is inhibitory, replicating the picture naming study of plurals reported by Baayen *et. al.*(2008). The reversal of this effect into facilitation for untargeted responses suggests that words with more complex inflectional paradigms serve as more powerful attractors during lemma selection.

	Estimate	lower HPD95	upper HPD95	p(MCMC)
Intercept	6.8780	6.6843	7.0950	0.0000
Lemma Frequency (linear)	-0.0522	-0.0859	-0.0259	0.0001
Lemma Frequency (quadratic)	0.0031	0.0015	0.0055	0.0005
Picture Entropy	0.0787	0.0657	0.0992	0.0000
Sex=female	-0.0655	-0.1324	-0.0030	0.0476
Previous RT	0.0515	0.0325	0.0714	0.0000
Picture Complexity	0.0011	0.0003	0.0019	0.0075
Inflectional Entropy	0.0224	-0.0094	0.0509	0.1857
Correct=incorrect	0.2541	0.1346	0.3628	0.0000
Trial	0.0005	0.0004	0.0006	0.0000
Alternating=TRUE	0.4598	0.1863	0.7429	0.0010
Form Frequency	-0.0165	-0.0279	-0.0062	0.0024
Present=TRUE	-0.1654	-0.2377	-0.0923	0.0000
Regularity=regular	-0.1358	-0.2449	-0.0198	0.0215
Prepared Naming=TRUE	-0.0259	-0.0636	0.0141	0.2048
Rhyme Count	-0.0200	-0.0306	-0.0103	0.0000
Picture Entropy : Sex=female	0.0340	0.0183	0.0490	0.0000
Inflectional Entropy : Correct=incorrect	-0.0789	-0.1284	-0.0234	0.0052
Alternating=TRUE : Prepared Naming=TRUE	-0.0367	-0.0620	-0.0068	0.0138
Picture Complexity : Present=TRUE	0.0011	0.0006	0.0015	0.0000
Picture Complexity : Prepared=TRUE	-0.0007	-0.0011	-0.0002	0.0046
Present=TRUE : Form Frequency	0.0264	0.0100	0.0429	0.0018
Inflectional Entropy : Regularity=regular	0.0539	0.0020	0.1048	0.0438
Previous RT : Alternating	-0.0599	-0.0982	-0.0222	0.0016

Table 4.6: Coefficients of a linear mixed model fit to the picture naming latencies of Experiments 1–4 jointly. P-values are based on 50,000 Markov chain Monte Carlo samples from the posterior distribution of the parameters. The standard deviations estimated for the random effects were 0.134 for Verb and 0.095 for Subject. The standard deviation of the residual error was 0.257.

Given the results reported by Ullman *et. al.* (2002) for English, we explored potential differences in lexical processing between males and females. Ullman and colleagues hypothesized that the superior verbal cognitive skills of females (see for a review Kimura, 2000) allow frequency effects for regular inflected forms to arise in females, but not in males. The present data offer tentative support for modest differences in verbal processing between females and males. However, support for a difference in sensitivity to *Form Frequency* is quite weak, as it is only in Experiment 3 that a frequency-related sex difference reached significance.

In Experiment 3, females revealed an inhibitory effect of the frequency of the present-tense form, whereas for males the slope of form frequency did not differ significantly from zero. This finding is consistent with the hypothesis that females, but not males, store regular inflected forms. However, given that in the past-tense naming experiments the effect of form frequency was significant for both sexes, and not modulated by an interaction with *Sex*, nor by an interaction with *Regularity* ($t = 0.18$ for the three-way interaction of *Sex* by *Regularity* by *Tense* in the joint analysis), it cannot be argued that males do not remember regular inflected forms. In fact, frequency effects may be weaker for males than for females, but nevertheless significantly present (see for evidence from Danish Balling and Baayen, 2008; Balling and Baayen, 2009).

Unfortunately, this frequential differences between females and males visible in the isolated analysis of Experiment 3 did not reach significance in the joint analysis ($t = 1.79$ for the interaction of *Sex* by *Tense* by *Form Frequency*). This might be due to an effect specific to prepared past-tense naming being washed out by the other three experiments. But it is equally likely that the effect in Experiment 3 is a false positive. The only robust pattern that emerges from our data, with full support from the joint analysis, is that for females the inhibitory effect of *Picture Entropy* was stronger. Females may have considered more alternatives for naming than males, consistent with the superior female verbal memory.

Conclusions

The general question addressed in this study is whether past-tense inflected forms are generated on-line from the present-tense stem during speech production. Most linguistic theories, especially those in the generative tradition (Pinker, 1999; Pinker and Ullman, 2002) claim that inflected forms are “derived” in this generative sense. Only a minority of linguists have considered non-derivational models (Bybee, 1985; Matthews, 1974; Blevins, 2003).

According to Pinker's dual mechanism model, the past-tense suffixation rule is applied only if a search for an irregular past-tense form fails. This ordering of suffixation after lexical search ensures that irregular verbs are not inflected regularly. Given this ordering, one would expect regular past-tense forms to require longer processing latencies than irregular past-tense forms. However, no clear disadvantage emerged for regular verbs. The joint analysis actually suggests that for all but the largest inflectional entropies, regular verbs were named faster than irregular verbs, instead of more slowly.

A further problem for the dual mechanism model is that in its original conceptualization (e.g., Pinker, 1991), it predicted that regular past-tense forms leave no traces in lexical memory, and hence that effects of frequency of use should not be observable for regular past-tense forms. However, frequency effects for past tense forms emerged in past-tense picture naming, irrespective of regularity. Effects of inflectional form frequency for regular inflected forms were also observed by Tabak *et. al.* (submitted) for Dutch and English in word naming tasks.

Later versions of the dual mechanism model have relaxed the claim that frequency effects should be limited to irregular past-tense forms (Pinker, 1999; Pinker and Ullman, 2002). For instance, for regular past-tense forms that are similar in form to irregular past-tense forms, a frequency effect is now posited, the idea being that storing the regular form in memory would protect this form against irregularization. However, our data do not support the hypothesis that a form frequency effect would be restricted just to regulars that are phonologically similar to irregular verbs.

Following the joint analysis of all four experiments, we observed a facilitatory frequency effect of the inflected form in unprepared and prepared past-tense naming. The slope of this frequency effect is the same for regulars and irregulars. Independently, we observed an effect of the count of rhyming irregulars, which also did not vary across regular and irregular verbs. This suggests that, at least in Dutch, the effect of form frequency in past-tense picture naming is not crucially dependent on phonological similarity to irregulars. In addition, the inhibitory effect of *Inflectional Entropy* (see also Tabak *et. al.*, submitted) provides further evidence against a sparse lexicon, as it provides evidence, albeit indirect evidence, for information in long-term memory of the likelihoods of individual inflected forms.

Finally, consider the effect of the voicing alternation (whether the stem-final obstruent alternates with respect to the feature voice) in prepared picture naming. This effect was not modulated by regularity, and presented itself for present-tense and paste-tense naming alike. According to derivational theories of this alternation, an underlying form is posited with a voiced obstruent. This voiced obstruent is visible in the past-tense forms of regular verbs, but becomes voiceless in all other forms probed in our experiments. Classical derivational theory (see, e.g., Shane, 1973) would lead one to expect an interaction of *Alternating* by *Regularity* in past-tense naming, with the forms with a voiceless final obstruent revealing the longest naming latencies (due to the application of a rule of devoicing). No such interaction was present, however — the effect of *Alternating* was observed even for present-tense naming, in which all final obstruents were realized as voiceless. While derivational theories incorrectly predict no difference for the present tense, the non-derivational theory of Ernestus and Baayen (2003); Ernestus, Baayen and Ling (2004); Ernestus and Baayen (2007) and Ernestus and Baayen (2006) can accomodate this finding as reflecting uncertainty about the stem-final voicing across lexical paradigms. Note that *Alternating* actually represents a second layer of irregularity orthogonal to the standard distinction between regular and irregular verbs based on vocalic alternation of the stem. Just as irregularity gives rise to longer picture naming latencies, voice alternation is time-costly.

Considered jointly, these findings challenge the dual mechanism model, while fitting well with Word and Paradigm Morphology in linguistics and non-derivational approaches to inflection in psychology (Stemberger, 2002; Stemberger and Middleton, 2003; Stemberger, 2004).

The present experiments also shed some light on the consequences for lexical processing of differences in semantic density for regular and irregular verbs. According to the dual mechanism model, differences between regular and irregular verbs are restricted to phonological form. However, recent studies (Ramscar, 2002; Patterson, Lambon Ralph, Hodges and McClelland, 2001) suggest that semantic and contextual factors are also relevant for understanding how regular and irregular verbs are processed. Furthermore, a lexical statistical survey (Baayen and Moscoso, 2005) revealed that regulars and irregulars differ in semantic density: Irregulars tend to have denser semantic networks than regulars. For instance, when regulars and irregulars are matched for frequency, irregular verbs tend to have more meanings than regular verbs. Regular and irregular verbs also tend to have different aspectual properties, as witnessed by the non-uniform distribution of auxiliary verbs in Dutch and German. Regulars favor *hebben*, 'have', while irregulars favor *zijn*, 'be', the auxiliary marking telicity. Regular and irregular verbs are also non-uniformly distributed over Levin's verb argument alternation classes

(Levin, 1993). In the analyses of the present picture naming experiments, we examined various predictors gauging differences in semantic density, such as the count of synonyms in WordNet (Miller 1990). Whereas Tabak *et. al.* (2005) observed this measure to be predictive for visual comprehension, it failed to reach significance for the picture naming task.

Nevertheless, the present study does provide subtle distributional evidence that regulars and irregulars differ in semantic density. The artists who made the photographs used in the picture naming experiment reported that regular verbs were especially difficult to enact. This informal observation is supported by two observations. First, the mean size of the JPG files was larger for the regulars than for the irregulars ($t(165.5) = -6.4157, p < 0.0001$). In order to depict regular verbs, more complex postures and more ancillary attributes were required. Second, irregulars were characterized by substantially smaller Picture Entropy ($t(165.5) = -4.7281, p < 0.0001$). Apparently, the artists were much more succesful in creating unambiguous pictures for irregular verbs compared to regular verbs. In hindsight, this is not surprising, as irregular verbs in Dutch tend to denote primary positions movements and actions of the body, e.g., *zitten* (sit), *staan* (stand), *liggen* (lie), *lopen* (walk), *duiken* (dive), *zwemmen* (swim), *slaan* (hit), *kijken* (look) (cf. Baayen, 2007). Given this unequal distribution of regulars and irregulars across the Picture Complexity and Picture Entropy measures, and given that these two measures are invariably inhibitory across our picture naming experiments, we conclude that, at least in our data, irregular verbs had a processing advantage compared to regulars at the conceptual and semantic levels of processing.

Our experiments also call attention to the role of episodic memory in the picture naming task. In the present experiments, the costs of interpreting the picture, gauged by the compressed file size of the photograph (*Picture Complexity*) was smaller in prepared present-tense and prepared past-tense naming than for unprepared naming, unsurprisingly. Nevertheless, memories for picture names established during the familiarization with the pictures and their intended names may mask interesting differences that characterize more normal processing circumstances that do not depend on prior familiarization. For the present data, the joint analysis suggests that preparation attenuated the processing costs of *Alternating*. Furthermore, given that the *Picture Complexity* for regular verbs was greater than for irregular verbs, preparation may have been disproportionally facilitating for regular verbs. Thus, for the investigation of speech production, unprepared and prepared naming both have advantages and disadvantages to offer. Unprepared picture naming may be more revealing for processes preceding form selection, while prepared picture naming may be more sensitive to form selection and subsequent lexical processing.

The role of *Regularity* as factorial predictor representing what according to Pinker (1997) would be a fundamental organizing principle of human language, is surprisingly small. In the joint analysis, regular verbs are assigned a smaller intercept than irregulars, as well as a greater slope for *Inflectional Entropy*. *Regularity* does not interact with any of the other predictors. On the one hand, this very modest role for *Regularity* may indicate a lack of sensitivity of the experimental paradigm of picture naming. On the other hand, it may also indicate that with the current array of item-bound predictors we have succeeded in capturing those differences between regulars and irregulars that are crucially involved in speech production, rendering *Regularity* as a dichotomous factor largely superfluous.

References

Aronoff, M.: 1994, *Morphology by itself: stems and inflectional classes*, The MIT Press, Cambridge, Mass.

Baayen, R. H.: 2007, Storage and computation in the mental lexicon, *in* G. Jarema and G. Libben (eds), *The Mental Lexicon: Core Perspectives*, Elsevier, pp. 81–104.

Baayen, R. H. and Moscoso del Prado Martín, F.: 2005, Semantic density and past-tense formation in three germanic languages, *Language* **81**, 666–698.

Baayen, R. H., Davidson, D. J. and Bates, D. M.: 2008, Mixed-effects modeling with crossed random effects for subjects and items, *Journal of Memory and Language* **59**, 390–412.

Baayen, R. H., Feldman, L. B. and Schreuder, R.: 2006, Morphological influences on the recognition of monosyllabic monomorphemic words, *Journal of Memory and Language* **55**, 290–313.

Baayen, R. H., Levelt, W., Schreuder, R. and Ernestus, M.: 2008, Paradigmatic structure in speech production, *in* M. Elliott, J. Kirby, O. Sawada, E. Staraki and S. Yoon (eds), *Proceedings Chicago Linguistics Society 43, Volume 1: The Main Session*, Chicago, pp. 1–29.

Balling, L. and Baayen, R.: 2008, Morphological effects in auditory word recognition: Evidence from Danish, *Language and Cognitive Processes* **23 (7/8)**, 1159–1190.

Balling, L. and Baayen, R.: Manuscript submitted for publication, Evidence for morphologically complex words in a redundant lexicon.

Beard, R.: 1995, *Lexeme-morpheme base morphology: a general theory of inflection and word formation*, State University of New York Press, Albany, NY.

Bien, H., Levelt, W. and Baayen, R.: 2005, Frequency effects in compound production, *PNAS* **102**, 17876–17881.

Blevins, J. P.: 2003, Stems and paradigms, *Language* **79**, 737–767.

Bloch, B.: 1947, English verb inflection, *Language* **23**, 399–418.

Bybee, J. L.: 1985, *Morphology: A study of the relation between meaning and form*, Benjamins, Amsterdam.

Chomsky, N. and Halle, M.: 1968, *The sound pattern of English*, Harper and Row, New York.

De Vaan, L., Schreuder, R. and Baayen, R. H.: 2007, Regular morphologically complex neologisms leave detectable traces in the mental lexicon, *The Mental Lexicon* **2**(1), 1–24.

Ernestus, M. and Baayen, R.: 2007, Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch, *Language and Cognitive Processes* **22**, 1–24.

Ernestus, M. and Baayen, R.: 2010, The functionality of incomplete neutralization in Dutch: The case of past-tense formation, *in* L. Goldstein, D. Whalen and C. Best (eds), *Laboratory Phonology 8*, Mouton de Gruyter, Berlin, pp. 27–49.

Ernestus, M. and Baayen, R. H.: 2003, Predicting the unpredictable: Interpreting neutralized segments in Dutch, *Language* **79(1)**, 5–38.

Ernestus, M. and Baayen, R. H.: 2004, Analogical effects in regular past tense production in Dutch, *Linguistics* **42**, 873–903.

Kimura, D.: 2000, *Sex and cognition*, The MIT press, Cambridge, MA.

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

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

Matthews, P. H.: 1974, *Morphology. An Introduction to the Theory of Word Structure*, Cambridge University Press, London.

Milin, P., Filipović Durdević, D. and Moscoso del Prado Martín, F.: 2009a, The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from serbian, *Journal of Memory and Language* pp. 50–64.

Milin, P., Kuperman, V., Kostić, A. and Baayen, R.: 2009b, Paradigms bit by bit: An information-theoretic approach to the processing of paradigmatic structure in inflection and derivation, *in* J. Blevins and J. Blevins (eds), *Analogy in Grammar: Form and Acquisition*, Oxford: Oxford University Press, pp. 214–252.

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

Patterson, K., Lambon Ralph, M. A., Hodges, J. R. and McClelland, J. L.: 2001, Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge, *Neuropsycologia* **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. and Prince, A.: 1988, On language and connectionism, *Cognition* **28**, 73–193.

Pinker, S. and Prince, A.: 1994, Regular and irregular morphology and the psychological status of rules of grammar, *in* S. Lima, R. Corrigan and G. Iverson (eds), *The Reality of Linguistic Rules*, John Benjamins, Amsterdam.

Pinker, S. and Ullman, M.: 2002, The past and future of the past tense, *Trends in the Cognitive Sciences* **6**(11), 456–462.

Plunkett, K. and Juola, P.: 1999, A connectionist model of English past tense and plural morphology, *Cognitive Science* **23**(4), 463–490.

Plunkett, K. and Marchman, V.: 1993, From rote learning to system building: acquiring verb morphology in children and connectionist nets, *Cognition* **48**, 21–69.

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

Rumelhart, D. E. and McClelland, J. L.: 1986, On learning the past tenses of English verbs, *in* J. L. McClelland and D. E. Rumelhart (eds), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models*, The MIT Press, Cambridge, Mass., pp. 216–271.

Seidenberg, M. S. and Gonnerman, L. M.: 2000, Explaining derivational morphology as the convergence of codes, *Trends in Cognitive Sciences* **4**(9), 353–361.

Shane, S. A.: 1973, *Generative Phonology*, Prentice-Hall, New Jersey.

Stemberger, J. and Middleton, C.: 2003, Vowel dominance and morphological processing, *Language and Cognitive Processes* **18**(4), 369–404.

Stemberger, J. P.: 2002, Overtensing and the effect of regularity, *Cognitive Science* **26**, 737–766.

Stemberger, J. P.: 2004, Phonological priming and irregular past, *Journal of Memory and Language* **50**, 82–95.

Tabak, W., Schreuder, R. and Baayen, R. H.: 2005, Lexical statistics and lexical processing: Semantic density, information complexity, sex, and irregularity in

Dutch, *in* S. Kepser and M. Reis (eds), *Linguistic evidence — Empirical, Theoretical, and Computational Perspectives*, Mouton de Gruyter, Berlin, pp. 529–555.

Tabak, W., Schreuder, R. and Baayen, R. H.: 2009, Is inflection derivational? evidence for a redundant lexicon from word naming, *submitted*.

Ullman, M. T.: 2004, Contributions of memory circuits to language: the declarative/procedural model, *Cognition* **92**, 231–270.

Ullman, M. T., Estabrooke, I. V., Steinhauer, K., Brovetto, C., Pancheva, R., Ozawa, K., Mordecai, K. and Maki, P.: 2002, Sex differences in the neurocognition of language, *Brain and Language* **83**, 141–143.

Vitevitch, M. S.: 2002, The influence of phonological similarity neighborhoods on speech production, *Journal of Experimental Psychology: Learning, Memory and Cognition* **28**(4), 735–747.

Vitevitch, M. S.: 2006, The curious case of competition in spanish speech production, *Language and Cognitive Processes* **21**(6), 760–770.

A cross-language perspective on processing in Spanish and Dutch

Chapter 5

This chapter has been submitted to Morphology as Wieke Tabak, Hanke van Buren, Robert Schreuder, and R. Harald Baayen. A cross-language perspective on processing in Spanish and Dutch

Four picture naming experiments addressing the production of regular and irregular past-tense forms in Dutch are reported. Effects of inflectional entropy as well as effects of the frequency of the past-tense inflected form across regulars and irregulars support models with a redundant lexicon while challenging the dual mechanism model (Pinker, 1997). The evidence supports the hypothesis of Stemberger (2004) and the general approach of Word and Paradigm morphology (Blevins, 2003) according to which inflected forms are not derived from the present-tense stem, but accessed independently.

Abstract

Spanish is a language with a different lexicalization pattern than Dutch. In this study, we report a picture naming experiment and a visual lexical decision experiment in Spanish. Semantic variables related to Dutch, such as Dutch derivational entropy, Dutch morphological family size and Dutch picture entropy, had an inhibitory effect on the response latencies in Spanish. This suggest that native speakers of Spanish are sensitive to the semantic properties of a verb, even if these properties are not lexicalized in Spanish. The *verbalizer* of Bierwisch and Schreuder (1992) is an adaptation of the model by Levelt (1989) and might account for these semantic properties.

Introduction

How are semantics and regularity related? This is a central question in psycholinguistics (Patterson, Lambon Ralph, Hodges and McClelland, 2001; Bird, Lambon Ralph, Seidenberg, McClelland and Patterson, 2003; Braber, Patterson, Ellis and Lambon Ralph, 2005; LambonRalph, Braber, McClelland and Patterson, 2005; Tyler, Stamatakis, Jones, Bright, Acres and Marslen-Wilson, 2004) In more detail, the question is whether irregular past tense forms, in morphological processing, depend more on semantic similarities than regular past tense forms. Baayen (2007) and Baayen and Moscoso del Prado Martín (2005) explored this possibility and presented distributional evidence that showed that irregular verbs have denser semantic neighborhoods and a greater degree of embodiment than regulars.

The study of Tabak, Schreuder and Baayen (2010) adresses the question whether past tense forms are generated online from their present tense stems during speech production, in Dutch. They found evidence that supports the hypothesis of Stemberger (2004) and the general approach of Word and Paradigm morphology (Blevins, 2003) according to which inflected forms are not derived by rule, but are accessed independently. In this study, Tabak and colleagues studied Dutch native speakers naming daily activities that could be verbalized with one verb in Dutch. However, the same daily activities cannot be verbalized with one single verb in every language, as languages use different lexicalization patterns. Spanish is an example of a language that uses a different lexicalization pattern to express meaning than Dutch. In what follows, we will briefly discuss the differences between Dutch and Spanish in this respect.

In English and in Dutch is possible to use sentences like the following (English example from Talmy, 1985, p.62-64):

(1) Ik *rende* de trap af
I *ran* down the stairs

(2) Het servet werd *van* de tafel *geblazen*
The napkin *blew off* the table

In these two languages (and also in Chinese and all other Indo-European languages except (post-Latin) Romance), both the way of moving (Manner) and the cause of the movement (Cause) are included in the verb. However, in Spanish it is impossible to express this the way English and Dutch do. Translating the abovementioned examples in Spanish results in the following sentence:

(1') Baje la escalera corriendo
I went down the stairs, running

(2') La servilleta se volió por el viento
The napkin flew from the table, by the wind

On the other hand, in Spanish (and other Romance languages, as well as Semitic and Polynesian languages) the motion (Motion) and direction (Path) are expressed by the verb. This is illustrated by the following examples (Talmy, 1985, p.69]):

(3) a. La botella *entró* a la cueva (flotando)
the bottle moved-in-to the cave (floating)
De fles dreef de grot in
'The bottle floated into the cave'

b. La botella *salió* de la cueva (flotando)
the bottle moved-out from the cave (floating)
De fles dreef de grot uit
'The bottle floated out of the cave'

c. La botella *se fue* de la orilla (flotando)
the bottle moved-away from the bank (floating)
De fles dreef weg van de waterkant
'The bottle floated away from the bank'

d. La botella *volvió* a la orilla (flotando)
the bottle moved-back to the bank (floating)
De fles dreef terug naar de waterkant
'The bottle floated back to the bank'

English and Dutch make use of a satellite (*in* 'in',*uit* 'out', *weg* 'away' and *terug* 'back'), while in Spanish this is part of the verb ('entrar', 'salir', 'irse' and 'volver').

The explained differences in lexicalization patterns, between Dutch and Spanish make it interesting to compare the way these languages are processed in production. In this way possible interactions with regularity and semantics can be explored. We decided to run a picture naming experiment, with the same pictures of common daily activities as used in the study by Tabak *et. al.* (2010), with Spanish native speakers as participants. We investigated the independent contributions to the response latencies of various lexical predictors, which will be discussed in further detail in the next section.

Our aim was to investigate whether the Spanish native speakers are influenced by semantic details of the verb that can be lexicalized in Dutch with one single verb, but that cannot be expressed fully in one single verb in Spanish. We hypothesize that there would be a processing disadvantage for the Spanish native speakers compared to the Dutch native speakers, due to the differences between languages. Spanish native speakers probably do perceive the semantic details about the verb, even though their language does not lexicalize these details and this might be time-costly.

We will also present the same verbs in their written form in Spanish by means of a visual lexical decision experiment. We conducted this experiment in order to distinguish conceptual processes, which we expect to uncover in the picture naming experiment, from more lexical processes. The order of presented experiments might seem odd, going from conceptualization to linguistic form. However, we choose to present it this way because, to our surprise, some of the task-specific predictors in the picture naming experiment, also had an effect in our visual lexical decision experiment and it would be hard to explain this effect without the results of the picture naming experiment.

Experiment 1: Picture naming by Spanish native speakers

Method

Subjects. Twenty three subjects (twenty females and three males) participated in the experiment. They were all students at the University of La Laguna, Spain, and native speakers of Spanish. They all had normal or corrected-to-normal vision. Subjects received course credits for participation in the experiment.

Materials. Experiment 1 was inspired by the picture naming experiment by Tabak *et. al.* (2010) that aimed to obtain more insight into the processing of Dutch verbs in the mental lexicon. From a set of 286 Dutch verbs used in a prior word naming experiment (Tabak, Schreuder and Baayen, 2005), they selected 170 picturable verbs. For each verb, a photograph was made of an actress acting out the verb. Across pictures, the same actress, objects and background were used to keep all pictures as comparable as possible.

The same pictures were used for this experiment to examine the way Spanish verbs are processed. This had as a disadvantage that it was not possible to control for measures like regularity, word length or frequency of the Spanish verbs. However, we were interested to see how speakers of Spanish would respond to especially these pictures because of the different lexicalization patterns between the two languages.

The 170 Dutch target verbs were translated into Spanish. This caused some problems due to the differences between Dutch and Spanish. Dutch verbs expressing Motion as well as Manner or Cause can only be translated with a Spanish verb expressing Motion. As a result, some pictures in the experiment representing different Dutch verbs had to be translated with the same Spanish verb. For example, the Dutch verbs ‘grijpen’ (‘to grab’), ‘pakken’ (‘to take’) and ‘vangen’ (‘to catch’) all refer to getting hold of something in a different manner. In Spanish these verbs are all translated with ‘coger’. In the Dutch experiment by Tabak *et. al.* (2010) every picture represented a different Dutch verb. In this experiment however, because of this difference between Dutch and Spanish, it was not possible that every picture represented a different Spanish verb. Nevertheless, we decided to retain as many pictures as possible in the experiment. Subjects were told before the experiment that the same verb could be shown more than once in different ways. In total there were 34 Dutch verbs that were translated with 17 Spanish verbs. All the other Dutch verbs had their own Spanish equivalent.

Another problem with translating the Dutch verbs into Spanish was the fact that some Dutch verbs have no Spanish equivalent existing of only one word. In some cases this has to do with the conflation of Motion and Manner in Dutch which does not exist in Spanish. For example, the Dutch verb ‘staren’ (‘to stare’) can only be translated into Spanish with ‘mirar fijamente’ (‘to look in a fixed way’). In other cases there were some translation problems that had to do with the tendency in Spanish to use periphrastic constructions An example of this is ‘dar de comer’ (literally, ‘to give food’) as translation of the Dutch verb ‘voeren’ (‘to feed’). In total there were 16 verbs that required a translation with more than one word.

Furthermore, some Dutch verbs could only be translated with a reflexive verb in Spanish. This is because in Dutch and Spanish aspect and causation are lexicalized differently. Whereas the default aspect of a Dutch causative verb is a state, in Spanish this is putting-into-a-state, and therefore sometimes an object is needed. However, in some pictures there was no object. This does not cause any problems for Dutch, but for Spanish it does. For example, the Dutch verb ‘schrikken’ (’to be frightened’) is normally translated into Spanish with the agentive verb ‘asustar’. However, this means frightening someone else and since the girl in the picture is frightened herself, the reflexive morpheme ‘-se’ must be added to make it inchoative: ‘asustarse’. There were 24 verbs that needed an additional reflexive morpheme to match the respective pictures.

Finally, a couple of pictures were excluded from the Spanish experiment because they were to difficult (or even impossible) to interpret for the Spanish native speakers, like the picture of ‘schaven’ (‘to slice’) in which a cheeseslicer is used, which is a completely unknown object in Spain. As a consequence, our experiment comprised 162 pictures representing 145 different Spanish verbs. All translations were checked by a native speaker of Spanish. We made ten lists with the 162 pictures in a different random order.

Procedure. Each subject was tested individually in a darkened and quiet room. Before the experiment started the subjects had to read aloud 20 Spanish verbs presented on the screen of a laptop in order to test if the microphone properly registered their voice. The verbs were different from the verbs included in the experiment. The microphone was positioned approximately 10 cm in front of each subject. After the microphone had been adjusted, subjects received the instruction for the first part of the experiment on paper. They were asked to name the presented pictures with a verb in its infinitival form. They were asked to do this as quickly and as accurately as possible and to avoid coughs, false starts and hesitations (e.g., ‘uhmm’).

Every trial started with a fixation point in the form of an asterisk in the middle of the screen of the laptop. After 1000 ms the asterisk disappeared and 50 ms later the picture was shown at the same position during 3000 ms. The subjects had to name the verbal action or state within these 3000 ms. A new trial was initiated 500 ms later. There were three short breaks during this first part of the experiment.

Henceforth we refer to this part of the experiment as *unprepared naming*, as the subjects had no prior experience with any of the pictures and the words we hoped to elicit.

After the first part of the experiment the subjects saw the pictures again, but this time with the Spanish target verb written underneath. By pressing a button of the mouse, subjects could determine the pace with which they went through the pictures, although they were asked not to spend too much time on each picture.

Finally, during the second part of the experiment, the subjects had to name the 162 pictures again. This part included again three short breaks. Henceforth, we refer to this part of the experiment as *prepared naming*, as the subjects now had seen the pictures matched with the target words. The whole experiment took approximately 45 minutes.

Results and discussion

A logarithmic transformation (using the natural logarithm) was applied to the response latencies because of the skewness of the distribution. Furthermore, all response latencies less than 500 ms (1.2% of the data) and greater than 3000 ms (the time-outs, 12.1% of the data) were removed, as well as invalid responses, like coughs, false starts and hesitations (10.1% of the data). The valid responses given by the subjects can be divided in responses we were trying to elicit (*targeted responses*) and alternatives (*untargeted responses*). In the unprepared part of the experiment, subjects produced more untargeted verbs than targeted verbs (1572 vs. 1174 respectively). In the prepared part of the experiment, subjects were familiar with the targeted verbs and produced a lot more of the intended responses (2230, against 729 alternatives).

A linear mixed-effects regression analysis was fitted to the data set using a stepwise variable selection procedure, with the reaction times (RTs) as the dependent variable and Subject and Verb as crossed random effects. Finally, the model was refitted after removing outliers exceeding 2.5 standard deviations (1.1% of the data). Table 5.1 lists the parameters for the random effects in the model.

Note that the by-verb standard devation is bigger than the by-subject standard devation, which is normally the other way round. This reflects the substantial variability in the difficulty of the pictures. Furthermore, the by-verb and by-subject standard deviations are considerably smaller than the residual standard deviation.

p-values were based on the usual *t*-tests with as degrees of freedom the

Groups	Name	Std.Dev.
Verb	(Intercept)	0.0922
Subject	(Intercept)	0.0678
Residual		0.2402

Table 5.1: Estimated standard deviations of the random effects in the linear mixed-effects regression model for the response latencies in Experiment 1.

upper bound for mixed-effects models (number of observations - number of fixed parameters) (cf. Baayen, Davidson and Bates, 2008). Markov Chain Monte Carlo sampling of the posterior distribution of the parameters supported all significance levels.

A total of 24.5% of the data points had to be removed (1.2% RTs < 500ms, 12.1% time-outs, 10.1% invalid responses, 1.1% outliers), whereas in the Dutch version of this experiment by Tabak *et. al.* (2010) only 9.7% of the data had to be left out of the analysis. This substantial difference can be traced to the fact that the pictures were based on Dutch verbs and therefore harder to lexicalize for the Spanish subjects than for the Dutch subjects. Furthermore, the Dutch participants were regular subjects of the Max Planck Institute in Nijmegen and hence used to experiments like picture naming, whereas the Spanish subjects were ignorant of this type of task and sometimes even got nervous during the experiment.

Various covariates were studied in the analysis. The frequency variables we took into account were Lemma Frequency and Said Frequency. Lemma Frequency is the sum of the frequencies of the inflectional variants of the targeted verb. Said Frequency is the surface frequency of the verb actually produced by the subject, e.g. the string frequency of the verbs the subjects came up with; targeted or untargeted. In case subjects produced more than one word to describe a picture, we calculated the surface frequency of the main verb of the utterance.

Other covariates we studied were the complexity of the picture gauged by means of the size in bits of its JPG file (JpgSize) and Picture Entropy. This last covariate was calculated as follows: we counted the number of different alternative verbs each picture elicited and calculated Shannon's entropy over the relative frequencies in the experiment of these alternatives (Shannon and Weaver, 1949). Tabak *et. al.* (2010) used this entropy measure as well in their analyses to gauge the uncertainty on the part of the subjects with respect to the required picture name (see also Bates, D'Amico, Jacobsen, Szekely, Andonova, Devescovi, Herron, Ching

Lu, Pechmann, Pléh, Wicha, Federmeier, Gerdjikova, Gutiérrez, Hung, Hsu, Iyer, Kohnert, Mehotcheva, Orozco-Figueroa, Tzeng and Tzeng, 2003).

Various morphological covariates were also included in the analysis. The first one is Morphological Family Size, which is the number of different words in the morphological family (Schreuder and Baayen, 1997). We also looked at the possible influence of Derivational Entropy, which is Shannon's entropy calculated for the probability distribution of the words in the morphological family. This measure gauges the amount of information carried by a word's derivational paradigm and was developed by Moscoso del Prado Martín, Kostic and Baayen (2004). Furthermore, Inflectional Family Size (the number of different inflections in the inflectional family, see Traficante and Burani, 2003) and Inflectional Entropy were studied. Inflectional Entropy is a measure of paradigmatic inflectional complexity based on Shannon's entropy (Moscoso del Prado Martín *et. al.*, 2004). In our data, Inflectional Entropy is highly correlated with Said Frequency. Other morphological covariates in our analysis were the conjugation group the target verb belongs to (verb endings in '-ar', '-er' or '-ir', henceforth Conjugation Group); and whether the target verb was regular or not (Regularity). In addition, we included as predictors whether the target item was a reflexive verb or not (Reflexiveness) and the Length in letters of the verb actually produced by the subject (SaidLength).

Furthermore, to control for voicekey artefacts, some phonological variables were taken into account: Initial Phoneme; Place of Articulation; Manner of Articulation; and Voicing, all with respect to the target verbs.

All these covariates were calculated for this experiment using codings made by hand in the frequency database by Alameda and Cuetos (1995) (based on almost 1 million Spanish word tokens). Due to limitations of time we decided to code only the morphological and inflectional families for targeted verbs. As a result, all covariates could only be calculated for targeted verbs. Only Length in letters of the produced form (SaidLength) and Said Frequency were based on targeted verbs as well as untargeted verbs, since these could be calculated without additional coding. Lemma Frequency and Said Frequency were based on the larger frequency database by Sebastián-Gallés, Martí, Carreiras, and Cuetos, 2000 (based on 5.6 million word tokens) as these measures did not require coding.

Although obviously the Spanish subjects do not speak any Dutch at all, we wanted to find out whether effects from Dutch covariates can be detected as well. This would indicate that not only language-specific effects are important, but also more general processes of picture naming, like conceptual familiarity or accessibility (Bates *et. al.*, 2003; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder and Baayen, 2004). Therefore, Dutch Surface Frequency, Dutch Lemma Frequency, Dutch Picture Entropy, Dutch Morphological Family Size, Dutch Derivational Entropy, Dutch Inflectional Entropy and Dutch Regularity were studied as well (these covariates had already been calculated for the picture naming experiment by Tabak *et. al.* (2010)).

All covariates, except SaidLength, were logarithmically transformed in order to reduce the skewness in their distributions. Control variables in the model were Trial number (Trial), the response latency of the previous item (Previous RT), whether it was unprepared or prepared naming (henceforth Preparation) and whether the targeted verb was produced or not (Targeted). Factors were analyzed by means of contrast coding. Table 5.2 lists the coefficients and their associated statistics for the covariates and the interactions in the final model.

Control variables

All control variables in the model turned out to be significant predictors. The main effect of Trial is shown in the upper left panel of Figure 5.1. Response latencies increased slightly with increasing Trial number. This implies that the subjects slowed down a bit in the course of both parts of the experiment due to fatigue.

Unsurprisingly, the control variable Preparation turned out to be significant: Response latencies were longer during unprepared naming than during prepared naming. This may not inmediately become clear from Table 5.2, but its effect can be seen very clearly in Figure 5.1. Figure 5.1 gives the net effect of Preparation for targeted, unreflexive verbs with all interactions taken into account by fixing other numerical predictors at their median values. The grey line, which represents the unprepared part of the experiment, remains above the black line of the prepared part in every panel of Figure 5.1, indicating longer response latencies. In the prepared part of the experiment subjects were faster because they had just seen the pictures matched with the targeted verbs, which reduced the uncertainty about picture names.

The control variable Targeted, i.e. whether the response is a targeted verb or not, interacted with Preparation. Response latencies were shorter for targeted words than for alternatives in the prepared part of the experiment; this effect was less strong in the unprepared part of the experiment. This indicates that familiarization was effective.

The last control variable was Previous RT. There was a strong inhibitory effect

Table 5.2: Significant effects and interactions in the linear mixed-effects regression model of covariance for the response latencies in Experiment 1. The intercept represents prepared, untargeted, reflexive verbs (5602 degrees of freedom).

	Value	*t*-value	*p*-value
Intercept	6.943	17.945	< 0.0001
Trial	< 0.001	4.348	< 0.0001
Previous RT	0.032	4.639	< 0.0001
Preparation (unprepared)	-0.773	-2.905	0.0037
Targeted (true)	-0.147	-5.958	< 0.0001
Targeted by Preparation (true, unprepared)	0.100	6.593	< 0.0001
Spanish Said Frequency (linear)	-0.063	-5.750	< 0.0001
Spanish Said Frequency (quadratic)	0.007	4.970	< 0.0001
JpgSize	-0.016	-0.533	0.5941
JpgSize by Preparation (unprepared)	0.083	4.173	< 0.0001
Spanish Picture Entropy	0.101	7.918	< 0.0001
Dutch Picture Entropy	0.040	3.462	0.0005
Spanish Derivational Entropy	-0.013	-1.356	0.1752
Spanish Derivational Entropy by Targeted (true)	0.020	2.795	0.0052
Spanish Derivational Entropy by Preparation (unprepared)	0.038	5.766	< 0.0001
Dutch Derivational Entropy	0.013	0.717	0.4734
Dutch Derivational Entropy by Preparation (unprepared)	-0.044	-3.351	0.0008
Spanish Inflectional Family Size	0.023	1.526	0.1271
Spanish Inflectional Family Size by Preparation (unprepared)	-0.053	-5.019	< 0.0001
Reflexiveness (no)	0.162	5.109	< 0.0001
Reflexiveness by Targeted (no, true)	-0.086	-3.524	0.0004
Reflexiveness by Preparation (no, unprepared)	-0.081	-3.451	0.0006

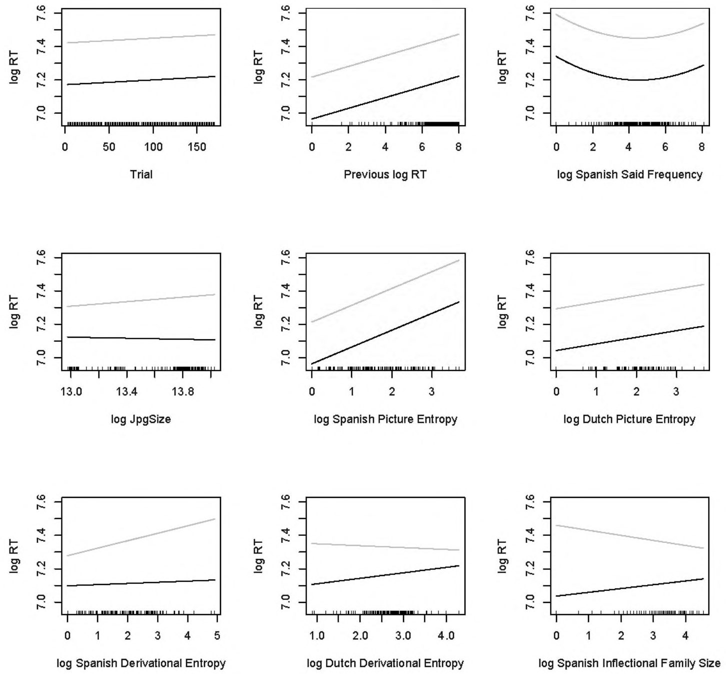

Figure 5.1: Significant partial effects of the predictors for response latencies in picture naming by Spanish native speakers. All curves are adjusted for the medians of all other predictors. The curves shown depict targeted, non-reflexive verbs. The grey line represents the unprepared part of the experiment, the black line the prepared part.

of response latencies to the preceding items in both parts of the experiment (upper central panel of Figure 5.1). This replicates and extends previous research suggesting that word naming latency is affected by the relative difficulty of the preceding stimulus (Tabak *et. al.* (2010); Taylor and Lupker, 2001; Vaan, Schreuder, Baayen, 2007; Wurm, Aycock, Baayen, 2006).

Frequency variables

We included two frequency variables in our analysis: Lemma Frequency and Said Frequency. Although these two variables are correlated, it turned out that Said Frequency was the key predictor. With Said Frequency included in the model, the effect of Lemma Frequency disappeared completely. An analysis with only targeted verbs in the prepared part of the experiment still revealed the same effect of Said Frequency and no effect of Lemma Frequency, not even without Said Frequency in the model. This argues in favor of item-specific learning in the mental lexicon, since there is a frequency effect of a specific form and not an effect of general frequency. This result is not expected in strict decompositional models. This will be discussed in further detail below.

Interestingly, the effect of Said Frequency was nonlinear and U-shaped, both for unprepared picture naming and for prepared picture naming, as the upper right panel of Figure 5.1 shows. The lower frequencies have a facilitatory effect, which levels off and turns into an inhibitory effect for the higher frequencies. Usually, the surface frequency effect is a facilitatory, linear effect (e.g. Allen, McNeal and Kvak, 1992; Forster, 1973; Schilling, Rayner, and Chumbley, 1998). However, a similar nonlinear U-shaped effect has been found in some recent studies as well (Tabak *et. al.* (2010); Tabak, Schreuder and Baayen (submitted); Bien, Levelt and Baayen, 2005).

Tabak *et. al.* (2010) did a similar picture naming experiment as described in this section but with Dutch subjects naming Dutch verbs. They conducted one experiment in which the subjects had to produce the present singular form of the verb, and another one eliciting the past singular form of the verb. In the experiment with the present singular forms each stimulus was preceded by 'Vandaag...' ('Today...') in the middle of the screen for 1000 ms. In the experiment with the past singular forms each stimulus was preceded by 'Gisteren...' ('Yesterday...') in the middle of the screen for 1000 ms. For present singular naming, there was a nonlinear effect of frequency in unprepared naming and a facilitatory effect in prepared naming. In the experiment with past singular naming, frequency showed a U-shaped curve in both unprepared and prepared naming. Interestingly, in our experiment we found no interaction with Preparation; in both unprepared and prepared naming frequency showed a U-shaped curve.

Bien *et. al.* (2005) investigated the role of frequency in compound production. Subjects first learned to associate a compound with a visually marked position on a computer screen. During the experiment they had to produce the associated compound when they saw the position mark. Bien *et. al.* (2005) varied the frequencies of the first and second constituent and of the compound itself. In the overall regression analysis the frequency effect turned out to be nonlinear. They suggest that the inhibition for the higher frequencies "[...] might represent a floor effect, however, because it might be an artifact of modeling nonlinearity with a simple, quadratic polynomial." (p.17880). However, our results indicate an even stronger inhibition than Bien *et. al.* (2005) found. Apparently, this U-shaped frequency effect is a robust and replicable effect.

Interpretational complexity variables

An interaction between the complexity of the picture, gauged by means of the size in bits of its JPG file (the JpgSize), and Preparation was found. JpgSize had a small inhibitory effect in the unprepared part, as can be seen in the middle left panel of Figure 5.1. The greater the visual complexity of the picture, the longer it took the subjects to respond, as expected. In the prepared part of the experiment there was no such effect, probably because the subjects were now familiarized with the pictures.

Spanish Picture Entropy is a measure that gauges the uncertainty on the part of the subjects with respect to the required picture name. It is calculated from the relative frequencies of the alternatives produced in the experiment. The central panel of Figure 5.1 shows that response latencies were longer with increasing Spanish Picture Entropy in both parts of the experiment. This suggests that the greater the uncertainty about how to name a picture in Spanish, the longer it took the subjects to produce a verb. Interestingly, there is no interaction with Preparation which seems to indicate that conceptual ambiguity is constant.

Remarkably, there was also an effect of Dutch Picture Entropy. This variable was calculated based on the results of the similar Dutch picture naming experiment by Tabak *et. al.* (2010) and expresses the uncertainty about how to name the pictures of our experiment in Dutch. Note that this means that Dutch Picture Entropy has been calculated on the basis of another group of subjects than Spanish Picture Entropy. The last panel in the second row of Figure 5.1 shows that response latencies of the Spanish subjects were longer with increasing interpretational uncertainty in Dutch in both parts of the experiment.

The interpretational uncertainty in Dutch is mainly caused by details about Manner. For example, if Dutch subjects see a picture which represents ‘lopen’ (‘to walk’), they have to make a decision between ‘lopen’, ‘wandelen’ (‘taking a walk’), ‘slenteren’ (‘to stroll’), ‘drentelen’ (‘to saunter’), and so on. Although Spanish subjects do not have the possibility to express all these details about Manner in single verbs in Spanish, this does not mean they cannot perceive these details. Apparently, these differences in Manner cause some additional uncertainty for the Spanish subjects at a conceptual level, apart from the effect for Spanish Picture Entropy. This indicates that at a conceptual level in addition to language specific knowledge more general cognitive aspects come into play. We discuss this issue in further detail in the general discussion.

Morphological variables

The effect of Spanish Derivational Entropy is shown in the lower left panel of Figure 5.1. Spanish Derivational Entropy interacted with Preparation: Response latencies were longer for increasing Spanish Derivational Entropy in the unprepared part of the experiment. This effect was no longer significant in the prepared part of the experiment.

In addition, Dutch Derivational Entropy interacted with Preparation as well: In contrast to Spanish Derivational Entropy, this variable was inhibitory in the prepared part of the experiment (lower central panel of Figure 5.1). To better understand the effects of Spanish and Dutch Derivational Entropy it is necessary to take a look at the derivational paradigms in both languages.

To derive verbs from verbs, Spanish uses prefixes like ‘anti-’ (contrast); ‘des-’ (negation); ‘pos(t)-’ (posteriority); ‘pre-’ (anteriority); ‘re-’ (repetition); ‘semi-’ (half, semi); and ‘super-’ (superiority). Similar functions are expressed in Dutch by prefixes such as ‘anti-’; ‘ont-’; ‘na-’; ‘voor-’; ‘re-’ or ‘her-’; ‘semi-’; and ‘super-’. However, in Dutch there is also a possibility to use prepositions to derive phrasal verbs, for example ‘boven’ (above) as in ‘bovenhalen’ (pull up); and ‘op’ (on, upon) as in ‘opvegen’ (sweep up). These prepositions may tell us something about the Path of the verb. In Spanish, Path is usually expressed by a conflation of Motion and Path in the main verb (see also introduction), described with a subordinate construction, for example ‘sacar a la superficie’ for ‘bovenhalen’, or Path is simply omitted, like in ‘barrer’, for ‘opvegen’. However, deriving a verb from a verb using a preposition is not possible in Spanish, since Spanish is a verb-framed language and not a satellite-framed language, as Dutch is.

We can see clearly the difference between verb-framed languages and satellite-framed languages if we take a closer look at the derivational paradigms in Dutch and Spanish: In Dutch there are many derivations, while in Spanish the derivational paradigm is quite limited. Table 5.3 shows the difference in size of the derivational paradigms for Dutch and Spanish with respect to the verb ‘to walk’. In Dutch, this is ‘lopen’; in Spanish this verb can be translated by ‘andar’ as well as by ‘caminar’. The full derivational paradigms of both ‘andar’ and ‘caminar’ are given. Only derived verbs are presented.

As can be seen in Table 5.3, Dutch has many verbs derived from the verb ‘lopen’ (‘to walk’), whereas Spanish has only a few verbs derived from ‘andar’ and ‘caminar’. Some of the meanings of the Dutch derivations are abstract, but many derivations have a concrete meaning, referring to different ways of physically walking.

In the unprepared part of the experiment there was an inhibitory effect of Spanish Derivational Entropy. This indicates that the more uncertainty about the intended verb for the subjects, the longer the response latencies were. This effect is the strongest in the primary language, in this case Spanish, explaining the lack of an effect of Dutch Derivational Entropy in the unprepared part of the experiment.

In the prepared part of the experiment, there was no longer uncertainty about the Spanish target verbs for the subjects. However, still much remains unspecified at a conceptual level. These unspecified details are expressed in the Dutch derivational paradigm. For instance, the derivational paradigm of the verb ‘lopen’ in Table 5.3 shows that Dutch has a greater proliferation of senses of a verb than Spanish. Although expressing these different senses is language-specific for Dutch, it does not mean that the Spanish subjects are not aware of the wide array of conceptual possibilities. The larger the derivational family in Dutch, the bigger the conceptual complexity. This leads to longer response latencies, also in Spanish.

Another morphological predictor in our analysis is Spanish Inflectional Family Size: the number of inflectional variants of the target verbs realized in the corpus of Alameda and Cuetos (1995) (see also Traficante and Burani, 2003) for details about Inflectional Family Size in Italian). This variable interacted with Preparation: in the unprepared part of the experiment response latencies were shorter for verbs with larger inflectional paradigms (lower right panel of Figure 5.1). Apparently large inflectional paradigms contribute to the conceptualization process when there

DUTCH			**SPANISH**	
lopen			**andar**	**caminar**
aanlopen	kaartlopen	scheeflopen	andar(se)	caminar
achteraanlopen	kroeglopen	skilopen	andorrear	descaminar
achterlopen	kromlopen	sprietlopen	desandar(se)	desencaminar
achternalopen	landlopen	steltlopen		encaminar(se)
achteromlopen	langslopen	stormlopen		
achteroplopen	leeglopen	stuklopen		
achteruitlopen	lopen	tegenlopen		
aflopen	loslopen	teruglopen		
belopen	meelopen	toelopen		
bijeenlopen	míslopen	uiteenlopen		
bijlopen	mislópen	uitlopen		
binnenlopen	nalopen	vastlopen		
boegsprietlopen	neerlopen	verlopen		
doodlopen	omhooglopen	vollopen		
dooreenlopen	omlopen	voorbijlopen		
dóórlopen	omverlopen	voorlopen		
doorlópen	onderlopen	vooroplopen		
drooglopen	ontlopen	voortlopen		
gelijklopen	openlopen	vooruitlopen		
groenlopen	oplopen	vrijlopen		
hardlopen	overhooplopen	wachtlopen		
heenlopen	óverlopen	wadlopen		
heetlopen	overlópen	warmlopen		
honklopen	platlopen	wedlopen		
hordenlopen	roedelopen	weglopen		
ineenlopen	rondlopen	zaklopen		
inlopen	samenlopen			

Table 5.3: The derivational paradigms of the verb ‘to walk’ in Dutch (‘lopen’) and Spanish (‘andar’ and ‘caminar’), only with respect to verbs (Dutch derivational paradigm based on the CELEX database (Baayen, Piepenbrock and Gulikers), Spanish derivational paradigm based on the databases by Alameda and Cuetos (1995) and Sebastián-Gallés *et. al.* (2000).

are no episodic memory associations yet between picture and response. This can be explained assuming that such verbs are better remembered due to more paradigmatic structure in lexical memory.

The last significant predictor in our analysis was Reflexiveness. Response latencies were shorter when the target verb was not reflexive and the subjects produced the intended verb. This may be because reflexive verbs were exceptional in the experiment, so producing them lead to longer reponse latencies.

Reflexiveness also interacted with Preparation. Verbs that were not reflexive had a facilitatory effect in the unprepared part of the experiment and an inhibitory effect in the prepared part. This reversal is probably due to the instruction on the one hand and the set-up of the experiment (first unprepared naming, then a demonstration of the targeted verbs matched with the pictures, followed by prepared naming) on the other. Subjects were instructed to use the infinitive of the verb, that is the uninflected form. It appeared that subjects responded differently on whether the reflexive morpheme '-se' was an inflection or not. Most subjects did not often use a reflexive verb during the unprepared part of the experiment (4.7% of the responses, whereas 14.8% of the target verbs was reflexive). However, after this first part of the experiment, the subjects saw the targeted responses, which included various reflexive verbs. As a result, subjects used more reflexive verbs in the prepared part of the experiment (12.3% of the responses), sometimes even in cases where it was not necessary. The longer response latencies for verbs that were not reflexive in this last part of the experiment can probably be explained by the fact that subjects this time consciously thought about whether they were supposed to add the reflexive morpheme '-se' or not.

Experiment 2: Lexical decision by Spanish native speakers

Method

Subjects. Twenty seven subjects (twenty four females and three males) participated in the experiment. They were all students at the University of La Laguna, Spain, and native speakers of Spanish. They all had normal or corrected-to-normal vision. Subjects received course credits for participation in the experiment.

Materials. The materials consisted of the same verbs as the targeted verbs of Experiment 1. In those cases where the translation consisted of more than one word, the main verb was selected. In some cases this main verb corresponded to a verb that already was included in the experiment; for instance the main verb of ‘dar el pecho’ (literally ‘to give the breast’, a translation of the Dutch verb ‘voeden’, which means ‘to (breast)feed’) is ‘dar’, but the verb ‘dar’ (to give) was also included in the picture naming experiment. In cases of overlap, the verb was only included once in the lexical decision experiment. As a consequence, 138 target verbs were included in the experiment. These verbs were complemented with 138 phonotactically legal pseudoverbs, created by changing one or two letters of the target verbs. We made ten lists with the 276 verbs in a different random order and added twenty practice items to the beginning of each list.

Procedure. Each participant was tested individually in a darkened and quiet room. The subjects received a written instruction in which they were told to decide as quickly and accurately as possible whether a presented letter string was a possible Spanish word or not by pressing a button with ‘SÍ’ or ‘NO’. For right-handed subjects the button with ‘SÍ’ was the right one of the two buttons, for left-handed subjects it was the left one. Before the experiment started there were twenty practice items to familiarize the participants with the procedure. After this practice set subjects could ask questions if they had any.

Every trial started with a fixation point in the form of an asterisk in the middle of the screen of the laptop. 50 ms after the asterisk disappeared the letter string was shown at the same position for 1750 ms in black lower-case, 36-point letters on a white background. The subjects had to press a button within these 1750 ms. A new trial was started after responding or time-out. There were two short breaks. In total, the experiment took approximately twenty minutes.

Results and discussion

A logarithmic transformation was applied to reduce the skewness of the distribution of the response latencies. Then, all response latencies greater than 1750 ms (the time-outs, 0.4% of the data) were removed. We fitted a linear mixed-effects regression model to the data set, using a stepwise variable selection procedure, again with the response latencies (RTs) as the dependent variable and Subject and Verb as crossed random effects. Finally, we removed outliers exceeding 2.5 standard deviations (2.2% of the data) and refitted the model. The parameters for the random effects in the model are given in Table 5.4.

Groups	Name	Std.Dev.
Verb	(Intercept)	0.0606
Subject	(Intercept)	0.1395
Residual		0.1802

Table 5.4: Estimated standard deviations of the random effects in the linear mixed-effects regression model for the response latencies in Experiment 2.

As expected, the by-subject standard deviation is bigger than the by-verb standard deviation. Both standard deviations are smaller than the residual standard deviation.

p-values were based again on the usual *t*-tests with as degrees of freedom the upper bound for mixed-effects models (number of observations - number of fixed parameters). All significance levels were supported by Markov Chain Monte Carlo sampling of the posterior distribution of the parameters.

In the analysis we studied various covariates. With respect to frequency, we looked at Lemma Frequency and Surface Frequency. The morphological variables in our analysis were Morphological Family Size, Derivational Entropy, Inflectional Family Size, Inflectional Entropy, Conjugation Group, and Regularity. The form variables we took into account were Reflexiveness and Spanish Word Length. Although this experiment did not include any pictures as in Experiment 1, we also included the interpretational complexity variables JpgSize and Spanish Picture Entropy, as these variables are informative about the verbs at the conceptual level. Furthermore, we looked at the phonological variables Initial Phoneme, Place of Articulation, Manner of Articulation and Voicing, even though the subjects did not have to say the words aloud during this experiment, as Baayen, Feldman and Schreuder (2006) and Balota, Cortese, Sergent and Spieler (2004) found that phonological properties such as the voicing of the first segment codetermined visual lexical decision latencies.

Additionally, the Dutch variables Dutch Lemma Frequency, Dutch Surface Frequency, Dutch Picture Entropy, Dutch Morphological Family Size, Dutch Derivational Entropy, Dutch Inflectional Entropy and Dutch Regularity were studied as well, to trace potential cross-language predictivity (cf. Moscoso del Prado

Table 5.5: Significant effects and interactions in the linear mixed-effects regression model of covariance for the response latencies in Experiment 2. The intercept represents reflexive verbs for which the lexical decision was correct (3490 degrees of freedom).

	Value	*t*-value	*p*-value
Intercept	7.0720	26.818	$<$ 0.0001
Trial (linear)	-0.0003	-1.586	0.1128
Trial (quadratic)	$<$ 0.0001	3.446	0.0006
Previous RT	-0.0378	-1.015	0.3102
Correct (false)	-0.0666	-3.709	0.0002
Spanish Lemma Frequency (linear)	-0.1480	-5.795	$<$ 0.0001
Spanish Lemma Frequency (quadratic)	0.0092	5.294	$<$ 0.0001
Dutch Lemma Frequency	-0.0200	-2.794	0.0052
Spanish Picture Entropy	0.0234	3.099	0.0020
Reflexiveness (no)	0.0070	0.171	0.8642
Spanish Morphological Family Size	0.0413	2.679	0.0074
Spanish Morphological Family Size by Reflexiveness (no)	-0.0569	-3.489	0.0005
Dutch Morphological Family Size	0.0175	1.745	0.0811
Dutch Morphological Family Size by Trial (linear)	-0.0001	-2.825	0.0048
Spanish Inflectional Family Size	0.0564	2.454	0.0142
Spanish Word Length	-0.1355	-3.808	0.0001
Spanish Word Length by Previous RT	0.0232	4.363	$<$ 0.0001

Martín *et. al.*, 2004).

We included Trial, Previous RT and whether the lexicial decision was correct or not (Correct) as control variables in our model. All covariates, except Spanish Word Length, were logarithmically transformed in order to reduce the skewness in their distributions. Factors were analyzed by means of contrast coding. Table 5.5 lists the coefficients and their associated statistics for the covariates and the interactions in the final model.

Control variables

All control variables turned out to be significant predictors. The upper left panel of

Figure 5.2 shows the effect for Trial. Towards the end of the experiment subjects got slower, possibly due to fatigue.

There was no main effect for Previous RT. However, Previous RT interacted with Spanish Word Length. This interaction will be discussed below.

Furthermore, the response latencies of the items for which the lexical decision was correct were longer than the response latencies of items with incorrect decisions. As instructed, subjects tried to react as quickly as possible, but sometimes decided too quickly, thus making errors.

Frequency variables

The upper center panel of Figure 5.2 shows a non-linear, facilitatory effect of Spanish Lemma Frequency, which levels off for the higher frequencies (for a similar floor effect, see Baayen, Feldman and Schreuder, 2006).

Interestingly, we found an additional effect of Dutch Lemma Frequency (upper right panel of Figure 5.2). Response latencies decreased with increasing Dutch Lemma Frequency, even though the subjects had no knowledge of Dutch. A possible explanation could be that this second, independent frequency measure gives additional precision. Another possibility is that frequency reflects familiarity and accessibility of concepts that is shared over languages. This finding is in agreement with Bates *et. al.*, (2003), who found other-language frequency effects for picture naming in seven languages.

Interpretational complexity

The first panel of the second row of Figure 5.2 shows an inhibitory effect of Spanish Picture Entropy. In Experiment 1, Spanish Picture Entropy was a measure that gauged the uncertainty on the part of the subjects with respect to the required picture name. In this experiment, the subjects did not have to produce a name. However, Spanish Picture Entropy tells us something about the number of verbs of which the action or state looks more or less the same. Verbs that differ only slightly in their semantics enter into stronger competition, leading to delayed lexical decision latencies. This result confirms that lexical decision is a task that taps into semantic processing.

Morphological variables

We also found an effect of Spanish Morphological Family Size. We found that verbs that were not reflexive (as were most of the items) were processed faster with

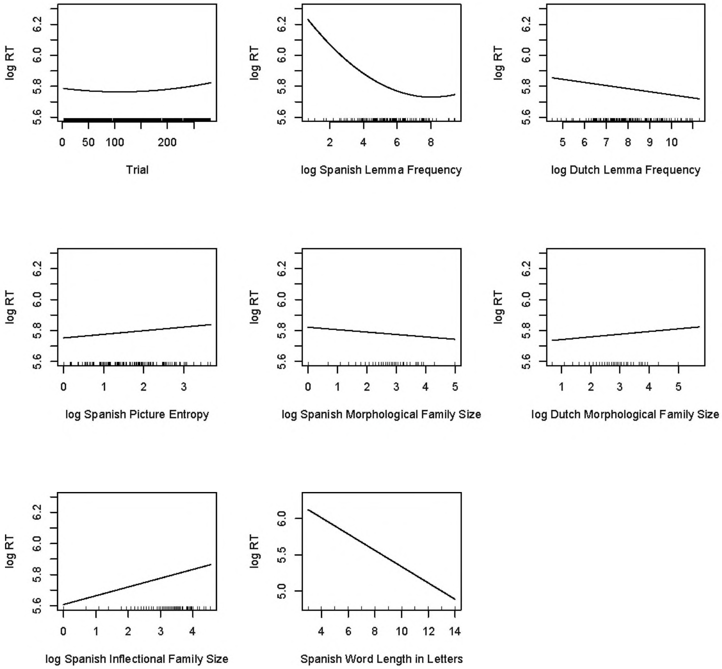

Figure 5.2: Significant partial effects of the predictors for response latencies in lexical decision by Spanish native speakers. All curves are adjusted for the medians of all other predictors. The curves shown depict non-reflexive verbs for which the lexical decision was correct. Note that the Y-axis of Spanish Word Length (last panel) has been extended due to the relative large effect of this variable in comparison with the other variables.

increasing Spanish Morphological Family Size (central panel of Figure 5.2). This interaction with Reflexiveness could be due to the fact that there were relatively few reflexive items in the experiment. Another possibility is that non-reflexive verbs coactivate other morphological families than reflexive verbs. The facilitation of Spanish Morphological Family Size replicates and extends previous reserach by Bertram, Baayen and Schreuder, (1999) and Schreuder and Baayen, (1997), who also found a negative correlation between morphological family size and visual lexical decision latencies. Apparently, words with greater morphological connectivity allow faster decisions due to more lexical coactivation. This facilitatory effect of family size for lexical decision is in contrast with the inhibitory effect of Derivational Entropy we found for unprepared picture naming by Spanish native speakers. However, in picture naming, which is a production task, greater general coactivation leads to greater selection problems for articulation. For lexical decision, the verb only has to be recognized, and then more coactivation helps, apparently.

Remarkably, there was also an effect of Dutch Morphological Family Size, as can be seen in the middle right panel of Figure 5.2. Response latencies increased with increasing Dutch Morphological Family Size, but this effect was slightly attenuated with increasing Trial number. In the part about the picture naming experiment, it was pointed out that the derivational paradigms of Dutch comprise many more members than do Spanish derivational paradigms. We also observed that larger families go hand in hand with increased conceptual complexity. Increased conceptual complexity leads to a greater processing load and to longer response latencies. The reason that this effect is somewhat attenuated with increasing Trial number may be due to saturation of morphological resonance (see e.g. DeJong, Schreuder and Baayen, 2003).

The lower left panel of Figure 5.2 shows that response latencies increased with increasing Spanish Inflectional Family Size. Above, we have seen that there was a facilitatory effect for Spanish Morphological Family Size. Apparently, inflections and derivations have a different effect on response latencies for lexical decision. Whereas semantics help to make a faster lexical decision, more inflections seem to delay the decision. Possibly, semantics help to recognize the concept, but more complex inflectional paradigms delay the recognition due to more competition.

The last significant predictor in our analysis was Spanish Word Length in letters. The last panel of Figure 5.2 shows a strong facilitatory effect for Spanish Word Length (note that the Y-axis of this panel has been extended due to the relative large effect of this variable in comparison with the other variables). This means that the greater the orthographic length of the verb, the faster the subjects could make a lexical decision about it. We do not know what causes this effect.

Spanish Word Length also interacted with Previous RT. Longer response latencies of the previous item enhanced the facilitatory effect of Spanish Word Length.

General discussion

The aim of this study was to obtain a better understanding of the processing of regular and irregular verbs in two languages with different lexicalization patterns. We investigated whether the Spanish native speakers are influenced by semantic details of the verb that can be lexicalized in Dutch with one single verb, but that cannot be expressed in on single verb in Spanish. We conducted two different experiments: picture naming and visual lexical decision.

In both experiments, we studied the influence of two interpretational complexity measures, Spanish and Dutch Picture Entropy, on the response latencies. These variables express the uncertainty about the required picture name in, respectively, Spanish and Dutch. We found an inhibitory effect for Spanish Picture Entropy in both picture naming and lexical decision. Furthermore, we found an additional inhibitory effect for Dutch Picture Entropy in picture naming, even though this group of subjects had no knowledge of Dutch. Most of the uncertainty for Dutch native speakers is caused by details about Manner. Spanish native speakers are not used to express these details about Manner in single verbs, but, apparently, they do perceive them. This causes additional uncertainty, and it is this uncertainty that we think is captured by the inhibitory effect of Dutch Picture Entropy. These results show that at the conceptual level not only language-specific knowledge is used in picture naming, but also more general conceptual knowledge.

Interestingly, the inhibitory effect of Spanish Picture Entropy was found for picture naming as well as for lexical decision. As Picture Entropy is informative about the number of verbs that are semantically close, this effect informs us that conceptual processing co-determines lexical decision latencies.

A significant effect of Morphological Family Size was observed in lexical decision. However, for picture naming the closely related variable Derivational Entropy (cf. Moscoso del Prado Martín *et. al.*, 2004) emerged as a better predictor. Apparently, for conceptualization the amount of information carried by a word's derivational paradigm is more important than the mere size of its paradigm.

For picture naming, increasing Derivational Entropy led to longer response latencies, probably because greater general coactivation creates greater selection problems for articulation. On the other hand, more coactivation helps for lexical decision (see for details Grainger and Jacobs, 1996). This is also what we observed: Spanish Morphological Family Size facilitated lexical decision times.

Furthermore, we found effects for the Dutch variants of Derivational Entropy and Morphological Family Size. We found an inhibitory effect of Dutch Derivational Entropy for picture naming. This indicates, similar to the effect of Dutch Picture Entropy, that at the conceptual level not only language-specific knowledge is important. More general, cognitive knowledge seems to contribute to the conceptualization process as well. The inhibitory effect of Dutch Morphological Family Size in lexical decision by the same group of subjects confirms this hypothesis. The increased conceptual complexity reflected in larger Dutch morphological families is perceived by the Spaniards, leading to a greater processing load and to longer response latencies.

Another morphological variable, Inflectional Family Size, was a reliable predictor as well. For (unprepared) picture naming, Spanish Inflectional Family Size had a facilitatory effect on response latencies. An inhibitory effect was found in the lexical decision experiment. It seems that large inflectional paradigms contribute to the conceptualization process when the pictures are observed for the first time. However, for lexical decision more complex paradigms lead to more competition, resulting in longer response latencies.

We also considered frequency of occurence. For picture naming, we found an effect for surface frequency of the verb produced by the subjects (the Said Frequency). In strict decompositional models, like the models by Pinker (1991, 1997) and Levelt (Levelt, Roelofs and Meyer, 1999), complex words are assumed to be produced through their constituents: The constituents are stored in memory, and complex words are created by applying rules. Therefore, according to these models, an effect for the frequency of the constituents is expected, and not of the word as a whole. Our effects of Said Frequency show the contrary and are more in line with theories in which each forms have their own representation in the mental lexicon, such as Word and Paradigm Morphology (Matthews, 1974; Aronoff, 1994; Beard 1995 and Blevins, 2003)

For lexical decision on the other hand, we did not find an effect for Surface Frequency, instead we observed an effect of Lemma Frequency. This is probably caused by the fact that for lexicial decision the response is also based on familiarity with the concept (Baayen, Feldman and Schreuder, 2006).

Finally, we found an additional effect of a Dutch frequency variable in Spanish behavior: Dutch Lemma Frequency had a facilitatory effect for lexical decision. It could be that this second, independent frequency measure gives additional precision, but it could also be the case that frequency reflects familiarity and accessibility of concepts that is shared across languages. An indication that the latter explanation may be on the right track, is the research by Bates *et. al.* (2003), who found frequency effects of languages that were unknown to the subjects in picture naming in seven languages.

Comparing picture naming and lexical decision we observe primarily differences with respect to the morphological variables. As mentioned above, for picture naming we found effects for Derivational Entropy while for lexical decision we found effects for Morphological Family Size. This could be because for conceptualization more information than just the size of a word's paradigm is important.

Furthermore, in our picture naming experiment, Derivational Entropy turned out to be inhibitory and Inflectional Family Size was facilitatory, whereas in our lexical decision experiment, Morphological Family Size was facilitatory and Inflectional Family Size was inhibitory. The opposite directions of the morphological variables in both tasks confirm the idea that tasks like picture naming and lexical decision are driven by different processes.

How do these results contribute to our understanding of lexical processing? We have seen the influence of the different lexicalization patterns between Spanish and Dutch in our research by the effects for Dutch variables for native speakers of Spanish who had no knowledge of Dutch. The inhibitory effects for Dutch Picture Entropy, Dutch Derivational Entropy and Dutch Morphological Family Size suggest that semantic properties of a verb (details concerning Manner), that cannot be lexicalized in Spanish in a single verb, were perceived by the native speakers of Spanish and induced additional uncertainty for them. Our results suggest there is remarkable similarity in how Dutch and Spanish native speakers perceive at the conceptual level. However, they cannot verbalize these concepts in the same way, because their languages focus on different details.

There are different points of view according to what happens if someone wants to verbalize a certain concept. In the model by Levelt (1989), a speaker's communicative intentions are formatted by the so-called *conceptualizer*, the processing system that gathers all the relevant information and turns it into a preverbal message. This preverbal message must then be verbalized and grammatically organized by the *formulator*. Finally, the *articulator* develops the articulatory output of an utterance. According to this model, language-specific knowledge is available to the conceptualizer.

Bierwisch and Schreuder (1992) propose an adaptation of this model, in which they add a *verbalizer* which transforms a non-linguistic conceptual structure into a representation containing a semantic form, which then can be mapped to a lemma by the formulator. If the verbalizer is considered to be a part of the formulator, then the conceptualizer and the formulator are two completely indepedent modules. In this way, the conceptual system does not need linguistic knowledge at all.

The *verbalizer* of Bierwisch and Schreuder might account for the semantic details that the Spanish native speakers do perceive, although their language does not lexicalize these details in one single verb.

Our results suggest that at the conceptual level more processes than just language-specific ones take place. Conceptual information sometimes has to be adjusted in order to be verbalized in a particular language. Although further research is needed comparing languages that differ in their lexicalization patterns, it seems that looking at online processing of picture naming and lexical decision from a cross-language perspective, as we have done, is a good way to investigate lexical processing in more detail.

References

Alameda, J. R. and Cuetos, F.: 1995, *Diccionario de frequencias de las unidades lingüísticas del castellano*, Servicio de Publicaciones de la Universidad de Oviedo, Oviedo.

Allen, P., McNeal, M. and Kvak, D.: 1992, Perhaps the lexicon is coded as a function of word frequency, *Journal of Memory and Language* **31**, 826–844.

Aronoff, M.: 1994, *Morphology by itself: stems and inflectional classes*, The MIT Press, Cambridge, Mass.

Baayen, R. H.: 2007, Storage and computation in the mental lexicon, *in* G. Jarema and G. Libben (eds), *The Mental Lexicon: Core Perspectives*, Elsevier, pp. 81–104.

Baayen, R. H. and Moscoso del Prado Martín, F.: 2005, Semantic density and past-tense formation in three germanic languages, *Language* **81**, 666–698.

Baayen, R. H., Davidson, D. J. and Bates, D. M.: 2008, Mixed-effects modeling with crossed random effects for subjects and items, *Journal of Memory and Language* **59**, 390–412.

Baayen, R. H., Feldman, L. B. and Schreuder, R.: 2006, Morphological influences on the recognition of monosyllabic monomorphemic words, *Journal of Memory and Language* **55**, 290–313.

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

Baayen, R. H., Wurm, L. H. and Aycock, J.: 2007, Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities, *Mental Lexicon* **2**, 419–463.

Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D. and Yap, M.: 2004, Visual word recognition for single-syllable words, *Journal of Experimental Psychology:General* **133**, 283–316.

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutiérrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A. and Tzeng, O.: 2003, Timed picture naming in seven languages, *Psychonomic Bulletin and Review* **10**, 344–380.

Beard, R.: 1995, *Lexeme-morpheme base morphology: a general theory of inflection and word formation*, State University of New York Press, Albany, NY.

Bertram, R., Baayen, R. H. and Schreuder, R.: 2000, Effects of family size for complex words, *Journal of Memory and Language* **42**, 390–405.

Bien, H., Levelt, W. and Baayen, R.: 2005, Frequency effects in compound production, *PNAS* **102**, 17876–17881.

Bierwisch, M. and Schreuder, R.: 1992, From concepts to lexical items, *Cognition* **42**, 23–60.

Bird, H., Lambon Ralph, M., Seidenberg, M., McClelland, J. and Patterson, K.: 2003, Deficits in phonology and past-tense morphology: What's the connection?, *Journal of Memory and Language* **48**, 502–526.

Blevins, J. P.: 2003, Stems and paradigms, *Language* **79**, 737–767.

Braber, N., Patterson, K., Ellis, K. and Lambon Ralph, M. A.: 2005, The relationship between phonological and morphological deficits in Broca's aphasia: Further evidence from errors in verb inflection, *Brain and Language* **92**, 278–287.

De Jong, N. H., Schreuder, R. and Baayen, R. H.: 2003, Morphological resonance in the mental lexicon, *in* R. H. Baayen and R. Schreuder (eds), *Morphological structure in language processing*, Mouton de Gruyter, Berlin, pp. 65–88.

De Vaan, L., Schreuder, R. and Baayen, R. H.: 2007, Regular morphologically complex neologisms leave detectable traces in the mental lexicon, *The Mental Lexicon* **2**(1), 1–24.

Forster, K. I. and Chambers, S. M.: 1973, Lexical access and naming time, *Journal of Verbal Learning and Verbal Behavior* **12**, 627–635.

Grainger, J. and Jacobs, A. M.: 1996, Orthographic processing in visual word recognition: A multiple read-out model, *Psychological Review* **103**, 518–565.

Lambon Ralph, M. A., Braber, N., McClelland, J. L. and Patterson, K.: 2005, What underlies the neuropsychological pattern of irregular > regular past-tense verb production?, *Brain and Language*.

Levelt, W. J. M.: 1989, *Speaking. From intention to articulation*, The MIT Press, Cambridge, Mass.

Levelt, W. J. M., Roelofs, A. and Meyer, A. S.: 1999, A theory of lexical access in speech production, *Behavioral and Brain Sciences* **22**, 1–38.

Matthews, P. H.: 1974, *Morphology. An Introduction to the Theory of Word Structure*, Cambridge University Press, London.

Moscoso del Prado Martín, F., Kostić, A. and Baayen, R. H.: 2004, Putting the bits together: An information theoretical perspective on morphological processing, *Cognition* **94**, 1–18.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R. and Baayen, R. H.: 2004, Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew, *Journal of Experimental Psychology: Learning, Memory and Cognition* **30**, 1271–1278.

Patterson, K., Lambon Ralph, M. A., Hodges, J. R. and McClelland, J. L.: 2001, Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge, *Neuropsycologia* **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.

Schilling, H. E. H., Rayner, K. and Chumbley, J. I.: 1998, Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences, *Memory & Cognition* **26**, 1270–1281.

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

Sebastián-Gallés, N., Martí, M. A., Carreiras, M. and Cuetos, F.: 2000, *LEXESP: Léxico informatizado del español*, Edicions Universitat de Barcelona, Barcelona.

Shannon, C. E. and Weaver, W.: 1949, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana.

Stemberger, J. P.: 2004, Phonological priming and irregular past, *Journal of Memory and Language* **50**, 82–95.

Tabak, W., Schreuder, R. and Baayen, R. H.: 2005, Lexical statistics and lexical processing: Semantic density, information complexity, sex, and irregularity in Dutch, *in* S. Kepser and M. Reis (eds), *Linguistic evidence — Empirical, Theoretical, and Computational Perspectives*, Mouton de Gruyter, Berlin, pp. 529–555.

Tabak, W., Schreuder, R. and Baayen, R. H.: 2009, Is inflection derivational? evidence for a redundant lexicon from word naming, *submitted*.

Tabak, W., Schreuder, R. and Baayen, R. H.: 2010, Producing inflected verbs: a picture naming study, *The Mental Lexicon*.

Talmy, L.: 1985, Lexicalization patterns: semantic structure in lexical forms, *in*

T. Shopen (ed.), *Language typology and syntactic description: Grammatical categories and the lexicon*, University Press, Cambridge, pp. 57–148.

Taylor, T. E. and Lupker, S. J.: 2001, Sequential effects in naming: A time-criterion account, *Journal of Experimental Psychology: Learning, Memory and Cognition* **27**, 117–138.

Traficante, D. and Burani, C.: 2003, Visual processing of Italian verbs and adjectives: The role of the inflectional family size, *in* R. H. Baayen and R. Schreuder (eds), *Morphological structure in language processing*, Mouton de Gruyter, Berlin, pp. 45–64.

Tyler, L., Stamatakis, E., Jones, R., Bright, P., Acres, K. and Marslen-Wilson, W.: 2004, Deficits for semantics and the irregular past tense: A causal relationship?, *Journal of Cognitive Neuroscience* **16**, 1159–1172.

Summary and Conclusion

Chapter 6

The goal of this dissertation was to gain more insight in the way regular and irregular verbs are produced and comprehended and especially how their processing is influenced by their semantics. We hypothesized that regular and irregular verbs are not processed differentially and that regularity is not independent of semantics. We tested this hypothesis in a series of experiments on Dutch, but we conducted some experiments in Spanish as well.

In this chapter, we provide a summary of the content and implications of the preceding chapters. Furthermore, we draw the general conclusion of our studies taken all together.

Within the field of psycholinguistic research, there are two opposite views on how inflected forms are formed. According to the dual mechanism model (Pinker 1991; Pinker, 1997; Pinker, 1999), regular inflected forms are created from their stem forms by rules, while irregular verbs are retrieved from memory. Other theories, like the approach of Stemberger and colleagues (Stemberger, 2002; Stemberger, 2004; Stemberger & Middleton, 2003), claim that inflected forms exist side by side and that both irregular and regular inflected forms are accessed directly from the mental lexicon. The Word and Paradigm morphology (Matthews, 1974; Aronoff, 1994; Beard, 1995; Blevins, 2003) is in line with this approach and focuses on the paradigmatic organization of inflected forms and takes the word itself (both simple and complex words) to be the basic meaningful unit of a language.

Verbs in production and comprehension

In Chapter 2 we discuss the results of our visual lexical decision experiment, in which we investigated the predictivity of the distributional properties for lexical processing in single word reading. In this study, we orthogonally contrasted number, tense and regularity, while including a large number of covariates in the analysis. Both the error analysis and the analyses of the response latencies showed a processing advantage for regular verbs. In both analyses, this advantage was most prominent for the past tense forms. This effect contradicts theories in which regular verb forms are being created by rules and irregular verb forms are retrieved from memory. Following these theories, one would expect regular past-tense forms to require longer processing latencies than irregular past-tense forms. We found frequency effects and effects of inflectional entropy for both regular and irregular verbs in the present and the past tense forms. This led us to conclude that we have evidence for memory traces in the mental lexicon for irregular and regular verbs, contradicting theories of the mental lexicon in which storage is restricted to irregular forms. In this study, we also observed an effect of gender. Compared to men, women turned out to be more sensitive to inflectional entropy, as was reflected in their response latencies, and they responded faster to plurals. This suggests that women not only have a better verbal memory (Kimura, 1999), but also are more proficient in morphological parsing.

In Chapter 3, we switch from comprehension to production. In this chapter, we report the results of two word naming experiments. The first experiment is a straightforward visual word reading task. The second experiment is a cross-tense naming task (also called morpho-naming) in which participants have to produce a past tense form when presented with a present tense form, and a present tense form when presented with a past tense form. For both studies the crucial variable is the probability of the present and past tense forms in their paradigms. We argued that these probabilities should emerge as significant predictors for both regular and irregular verbs. The results from our two naming studies are not very straightforward. We both found evidence in our data that supports a sparse lexicon in which rules derive inflected forms from their simple bases, and at the same time, evidence that supports a redundant lexicon in which inflected forms are retrieved from memory. In simple word naming, we observed no independent effect of the frequency of inflectional forms. In cross-tense word naming, measures tapping into the paradigmatic properties of inflected words were present only for irregular verbs. These two results are compatible with Pinker’s dual mechanism model. Our most convincing evidence against a sparse lexicon, is the effect of inflectional entropy, which was present in our simple word naming experiment and, most importantly, for both regular and irregular verbs. This effect was modulated by measures of neighborhood density, suggesting that the inflectional entropy measure reflects paradigmatic aspects of lexical competition (as the number of initial competitors increases, an informationally rich paradigm seems to increase the complexity of selecting the target form from its neighbors). Furthermore, in cross-tense word naming, response latencies decreased significantly for more frequent regular inflectional forms when the lemma frequency of the base was controlled for. This indicates that the brain is sensitive to the probabilities of individual inflected forms, irrespective of their regularity.

In Chapter 4 we address the production of regular and irregular inflected forms using the picture naming paradigm. This is a task that engages the full production process, from conceptualization to articulation. We conducted picture naming experiments in both the present and the past tense. Following Pinker's dual mechanism model, one would expect regular past-tense forms to require longer processing latencies than irregular past-tense forms. However, we did not find an effect of regularity in any of our past-tense picture naming experiments. Although we did find an effect of regularity in the other past-tense picture naming experiment, this was into the opposite direction: regulars were named faster than irregulars. Furthermore, we observed frequency effects for past tense forms, irrespective of regularity. These frequency effects were not restricted to regulars that are phonologically similar to irregular verbs. In addition, we found an effect of inflectional frequency in two out of our four picture naming experiments, which provides evidence for information about individual inflected forms in long-term memory.

The results of our visual lexical decision experiment and our picture naming experiments are compatible with our main hypothesis that regular and irregular verb forms exist side by side and that they are accessed directly from the mental lexicon. In both experiments, we did not observe a processing advantage for irregular verbs and we presented evidence for frequency effects for both regular and irregular verbs. In all these experiments and in our word naming experiments, we found effects of inflectional frequency. These effects were irrespective of regularity. Considered jointly, these findings provide further evidence against a sparse lexicon.

The role of semantics

In Chapter 2, we report a lexical statistical analysis of the distributional properties of regular and irregular verbs. This survey revealed systematic semantic differences between regular and irregular verbs. First of all, it revealed that, with a logistic regression model, the regularity of a verb can be predicted not only by lemma frequency but also by several other linguistic variables, such as family size, neighborhood density, argument structures, auxiliaries, inflectional entropy, the noun-verb frequency ratio and the spoken-written frequency ratio. Secondly, the analysis revealed that historical age is a good predictor of regularity. Irregular verbs are more likely to be old words than regular verbs and this effect increases with increasing phonological regularity (as measured by neighborhood density).

In the second part of Chapter 2, we discuss the results of our visual lexical decision experiments. Analyses of covariance revealed that the above-mentioned distributional properties of regular and irregular verbs were predictive for the error rate and the response latencies in visual lexical decision. A larger semantic density of a verb, as measured by the number of synsets and characterizing irregulars, led to shorter response latencies for irregular past plurals.

In Chapter 3, we report the results of our word naming studies. Measures related to semantics, such as number of synonym sets and argument structures, are no longer significant predictors, possibly because they are less relevant for word naming (which is known to be less sensitive to meaning than lexical decision). The paradigmatic effects still exist (although the effect of family size disappears in cross-tense naming, possibly because of the amount of noise in this artificial task). However, the predictors related to form appear to be predominant in our word naming studies.

In Chapter 4, we studied picture naming. Measures related to semantics showed up again in our analyses (contrary to our word naming experiments). Picture entropy and picture complexity measures appeared to have an inhibitory effect on the response latencies. These measures did not interact with regularity, providing further evidence against a sparse lexicon.

In Chapter 5 we switch from Dutch to Spanish, a language with a different lexicalization pattern than Dutch. We present the results of two different experiments: picture naming and visual lexical decision, using exactly the same materials and procedures for these two experiments as for the corresponding experiments in Dutch, discussed in Chapters 2 and 4. Our main research question for this study was whether native speakers of Spanish are influenced by semantic properties of a verb that can be lexicalized in Dutch with one single verb, but that cannot be fully expressed in one single verb in Spanish. We hypothesized that this difference between the two languages could lead to processing costs for the Spanish native speakers. In both experiments, we found effects of variables related to Dutch on the Spanish response latencies. These variables were either based on lexical properties of the verbs, such as Dutch derivational entropy and Dutch morphological family size, or obtained from the experiments (see Chapter 4), such as Dutch picture entropy, which captures the uncertainty about the verb displayed in the picture for the Dutch participants. The effects were inhibitory and suggest that, interestingly, native speakers of Spanish indeed are sensitive to the semantic properties of a verb, even though these semantic details cannot be lexicalized within one single verb in Spanish, leading to additional uncertainty for the native speakers of Spanish, compared to the native speakers of Dutch. The *verbalizer* of Bierwisch and Schreuder (1992), which is an adaptation of the model by Levelt (1989), might account for these semantic details. We conclude that the interpretation of the picture and probably also the conceptualization of the item are similar for native speakers of Spanish and Dutch, but that both languages have language-specific strategies when it comes to lexicalization.

We can conclude that our lexical statistical survey (Chapter 2) provided a clear answer when it comes to effects of semantic differentiation between regular and irregular verbs. The survey revealed systematic semantic differences between regular and irregular verbs. These effects were reflected in our visual lexical decision experiment and in some extent in our picture naming experiments in Dutch and Spanish, but not in our word naming experiments.

Concluding remarks

The aim of this thesis was to come to a better understanding of how regular and irregular verbs are produced and comprehended, with a special focus on the interaction of semantics with (ir)regularity. The most dominant view in psycholinguistic research on morphological processing is that form and meaning are separated by different modules in the mental lexicon. Within this view, differences between regular and irregular verbs are not systematically related to any semantic difference. Regular forms would be created by rules, and irregular forms would be retrieved directly from the mental lexicon. Pinker (1991, 1997, 1999) popularized this generative model of morphological processing by claiming that regular and irregular past tense forms are neuro-anatomically localized differently in the brain.

Our four empirical chapters strongly suggest that the difference between regular and irregular verbs is much more complex than Pinker and colleagues claim.

Effects of inflectional frequency and inflectional entropy show that inflected forms, whether regular or irregular, have their own representation in the mental lexicon and that inflection apparently is not driven by ’derivational’ rules only. Our findings are consistent with the approach of Stemberger and the Word and Paradigm morphology.

It is more difficult to answer the question on how the difference between regular and irregular verbs is influenced by semantics. We did find effects of semantic differentiation between regular and irregular verbs in our lexical statistical survey and these effects are reflected in comprehension, as the results of our visual lexical decision study have shown. In production, however, the possible effects of semantical differentiation may have been masked by formal effects (in our word naming experiments) and interpretational effects (in our picture naming experiments).

Our results lead us to assume that semantic properties differentially influence the processing of regular and irregular verbs across comprehension and speech production, but that the production tasks that we have used are not sensitive enough to allow these effects to emerge.

Thus it might be worthwhile to study the production of regular and irregular verbs in (even) more detail than we did for the present thesis. A picture naming study using ERPs (Hendrix, Tabak, Baayen and Schreuder, in preparation), investigated the electroencephalographic (EEG) correlates of a range of psycholinguistic predictors, using the same picture naming paradigm and the same materials as we did in the previous chapters. Hendrix et al. observed both form and meaning related frequency effects and, most importantly, they found that semantic properties influence both regular and irregular past tense forms, consistent with our findings.

We conclude that semantics is a crucial factor in processing both regular and irregular inflection.

References

Aronoff, M.: 1994, *Morphology by itself: stems and inflectional classes*, The MIT Press, Cambridge, Mass.

Beard, R.: 1995, *Lexeme-morpheme base morphology: a general theory of inflection and word formation*, State University of New York Press, Albany, NY.

Bierwisch, M. and Schreuder, R.: 1992, From concepts to lexical items, *Cognition* **42**, 23–60.

Blevins, J. P.: 2003, Stems and paradigms, *Language* **79**, 737–767.

Hendrix, P., Tabak, W., Schreuder, R., and Baayen, R. H.: n.d., Electrophysiological effects in language production: a picture naming study using generalized additive modeling. in preparation.

Kimura, D.: 2000, *Sex and cognition*, The MIT press, Cambridge, MA.

Levelt, W. J. M.: 1989, *Speaking. From intention to articulation*, The MIT Press, Cambridge, Mass.

Matthews, P. H.: 1974, *Morphology. An Introduction to the Theory of Word Structure*, Cambridge University Press, London.

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.

Stemberger, J. and Middleton, C.: 2003, Vowel dominance and morphological processing, *Language and Cognitive Processes* **18**(4), 369–404.

Stemberger, J. P.: 2002, Overtensing and the effect of regularity, *Cognitive Science* **26**, 737–766.

Stemberger, J. P.: 2004, Phonological priming and irregular past, *Journal of Memory and Language* **50**, 82–95.

Samenvatting en conclusie

Chapter 7

Het doel van dit proefschrift was om meer inzicht te krijgen in de manier waarop regelmatige en onregelmatige geproduceerd en verwerkt worden en met name hoe ze in productie en verwerking beïnvloed worden door de verschillende distributionele eigenschappen van alle inflectionele varianten. Onze hypothese was dat regelmatige en onregelmatige werkwoorden niet op verschillende wijze verwerkt worden en dat regelmaat niet onafhankelijk is van semantiek. We hebben deze hypothese getest door middel van een serie experimenten in het Nederlands, maar we hebben ook experimenten in het Spaans uitgevoerd.

In dit hoofdstuk geven we een samenvatting van de inhoud en implicaties van de voorafgaande hoofdstukken. Bovendien trekken we een algemene conclusie van alle onderzoeken tezamen.

In de psycholinguïstiek zijn de meningen verdeeld over hoe geïnflecteerde vormen gevormd worden. Volgens het *dual mechanism model* (Pinker, 1991; Pinker, 1997; Pinker, 1999) worden regelmatig geïnflecteerde vormen met behulp van regels gecreëerd vanuit hun stam, terwijl onregelmatige werkwoorden opgehaald worden uit het mentale lexicon. Andere theorieën, zoals de benaderingswijze van Stemberger and collega's (Stemberger, 2002; Stemberger, 2004; Stemberger & Middleton, 2003), beweren dat geïnflecteerde vormen naast elkaar bestaan en dat zowel onregelmatig als regelmatig geïnflecteerde vormen direct vanuit het mentale lexicon opgediept worden. De *Word and Paradigm morphology* (Matthews, 1974; Aronoff, 1994; Beard, 1995; Blevins, 2003) is in overeenstemming met deze benaderingswijze en richt zich op de paradigmatische organisatie van geïnflecteerde vormen en neemt het woord zelf (zowel eenvoudige als complexe woorden) als de basale betekenisvolle eenheid van een taal.

Werkwoorden in productie en verwerking

In hoofdstuk 2 doen we verslag van een visueel lexicale-decisie-experiment, waarin we de voorspelbaarheid onderzocht hebben van de distributionele eigenschappen in de lexicale verwerking bij het lezen van op zichzelf staande woorden. In dit onderzoek hebben we een orthogonaal contrast aangebracht met behulp van het enkel- of meervoud van het werkwoord, de tegenwoordige of verleden tijd en de regelmaat dan wel onregelmaat. Tegelijkertijd hebben we een groot aantal covariaten aan de analyse toegevoegd. Zowel de foutenanalyse als de analyses van de benoemingstijden lieten een winst in de verwerkingssnelheid zien ten gunste van de regelmatige werkwoorden. In beide analyses was deze winst het meest prominent bij de verleden tijdsvormen. Dit effect weerlegt theoriën waarin regelmatige werkwoordsvormen met behulp van regels gecreëerd worden en onregelmatige werkwoordsvormen uit het geheugen opgediept worden. Als we deze theorieën volgen, zouden we verwachten dat regelmatige verledentijdsvormen langere verwerkingstijden vereisen dan onregelmatige verleden-tijdsvormen. Wij hebben frequentie-effecten en effecten van de inflectionele entropie gevonden voor zowel regelmatige als onregelmatige werkwoorden in de tegenwoordige en verleden tijd. Hierdoor konden wij concluderen dat we bewijs gevonden hebben voor geheugensporen in het mentale lexicon voor zowel regelmatige als onregelmatige werkwoorden, wat in tegenspraak is met theorieën over het mentale lexicon waarin opslag beperkt is tot de onregelmatige vormen. In deze studie hebben we ook een effect van geslacht waargenomen. Vergeleken met mannen bleken vrouwen sterker te reageren op de inflectionele entropie van een werkwoord. Verder reageren vrouwen sneller op de meervoudsvormen dan mannen. Dit zou kunnen betekenen dat vrouwen niet alleen een beter verbaal geheugen hebben (Kimura, 1999), maar dat ze ook beter zijn in het morfologisch decoderen van woorden.

In hoofdstuk 3 gaat de focus van verwerking naar productie. In dit hoofdstuk doen we verslag van de resultaten van twee woordbenoemingsexperimenten. Het eerste experiment is een visuele woordbenoemingstaak. Het tweede experiment is een taak waarin de proefpersonen de tijd van de werkwoordsvorm (tegenwoordige of verleden tijd) om moeten zetten. Wanneer ze een werkwoord in de tegenwoordige tijd aangeboden krijgen, moesten ze dit werkwoord benoemen in de verleden tijd en vice versa. Voor beide experimenten geldt dat de belangrijkste variabele de kansverdeling van de tegenwoordige en verleden tijdsvormen in hun paradigma's is. Onze hypothese was dat deze kansen als significante voorspellers op zouden treden tussen regelmatige en onregelmatige werkwoorden. De resultaten van onze twee benoemingstaken zijn niet helemaal eenduidig. We hebben zowel bewijs gevonden voor een zuinig lexicon waarin geïnflecteerde vormen met behulp van regels aan hun stamvormen ontleend worden, als voor een rijk lexicon waarin geïnflecteerde vormen vanuit het geheugen opgehaald worden. In onze visuele woordbenoemingstaak hebben we geen onafhankelijk effect van de frequenties van de geïnflecteerde vormen gevonden. In de taak waarbij de werkwoordstijd omgezet moest worden bleken de maten die informatie gaven over de paradigmatische eigenschappen van geïnflecteerde woorden alleen de onregelmatige werkwoorden te beïnvloeden. Deze twee resultaten komen overeen met het *dual mechanism model* van Pinker. Ons meest overtuigende bewijs tegen een zuinig lexicon is het effect van inflectionele entropie, dat aanwezig was in ons visuele woordbenoemingsexperiment en, wat vooral belangrijk is, voor zowel regelmatige als onregelmatige werkwoorden. Dit effect werd verzwakt door interacties met de dichtheid van de buren, wat suggereert dat inflectionele entropie de paradigmatische aspecten van lexicale competitie weerspiegelt (als het aantal initiële concurrenten toeneemt, lijkt een paradigma met een grote informatiedichtheid de selectie van de doelvorm tussen zijn buren te bemoeilijken). Verder bleken in het tweede experiment, waarin proefpersonen van tijd moesten wisselen, de reactietijden van de meest frequente regelmatig geïnflecteerde vormen af te nemen wanneer de lemmafrequentie van de stamvorm constant gehouden werd. Dit duidt erop dat de hersenen gevoelig zijn voor de kansenverdeling van individuele geïnflecteerde vormen en dat dit onafhankelijk is van de regelmaat van deze vormen.

In hoofdstuk 4 besteden we aandacht aan de productie van regelmatige en onregelmatige geïnflecteerde vormen waarbij we gebruik maken van een fotobenoemingstaak. Dit is een taak waarbij het volledige productieproces betrokken wordt, van conceptualisatie tot articulatie. We hebben fotobenoemingstaken uitgevoerd in zowel de tegenwoordige als de verleden tijd. Volgens het *dual mechanism model* van Pinker zouden we verwachten dat regelmatige verleden-tijdsvormen langere verwerkingstijden vereisen dan onregelmatige verleden-tijdsvormen. We hebben echter geen effect van regelmaat gevonden in een van onze fotobenoemingstaken waarbij de verleden tijd vereist werd. We vonden wel een effect van regelmaat in de andere fotobenoemingstaak in de verleden tijd, maar deze was in de tegengestelde richting: regelmatige werkwoorden werden sneller benoemd dan onregelmatige werkwoorden. Voorts hebben we frequentie-effecten gevonden voor de verleden tijdsvormen en dit was onafhankelijk van de regelmaat van deze vormen. Deze frequentie effecten beperkten zich niet alleen tot de regelmatige werkwoorden die fonologisch gezien gelijk zijn aan onregelmatige werkwoorden. In twee van de vier fotobenoemingstaken hebben we een effect van inflectionele frequentie gevonden. Dit levert direct bewijs op voor informatie in het lange termijngeheugen voor de kansverdeling van individuele geïnflecteerde vormen.

De resultaten van ons visueel lexicale-decisie-experiment en onze fotobenoemingsexperimenten komen overeen met onze hoofdhypothese dat regelmatige en onregelmatige werkwoordsvormen naast elkaar kunnen bestaan en dat ze direct vanuit het mentale lexicon opgediept worden. In beide experimenten hebben we geen voordeel in de verwerking gevonden voor de onregelmatige werkwoorden en we hebben aangetoond dat er frequentie-effecten zijn voor zowel regelmatige als onregelmatige werkwoorden. In al deze experimenten en ook in onze woordbenoemingsexperimenten hebben we effecten gevonden van inflectionele entropie. Deze effecten waren onafhankelijk van de regelmaat van het werkwoord. Gezamenlijk bekeken leveren deze bevindingen meer bewijs tegen een zuinig lexicon.

De rol van semantiek

In hoofdstuk 2 doen we verslag van een lexicaal-statistisch onderzoek naar de distributionele eigenschappen van regelmatige en onregelmatige werkwoorden. Dit onderzoek heeft de systematische semantische verschillen tussen regelmatige en onregelmatige werkwoorden laten zien. Allereerst hebben we aangetoond dat we, wanneer we een logistisch regressiemodel gebruikten, de regelmaat van een een werkwoord konden voorspellen, niet alleen met lemmafrequentie, maar ook met verschillende andere linguïstische variabelen, zoals familiegrootte, de dichtheid van de buren, argumentstructuren, hulpwerkwoorden, inflectionele entropie, de zelfstandig naamwoord-werkwoord frequentieverhouding en de geschreven-gesproken frequentieverhouding. Ten tweede heeft het onderzoek aangetoond dat de historische leeftijd van werkwoorden een verdienstelijke voorspeller is van regelmaat. Onregelmatige werkwoorden hadden een grotere kans om oude werkwoorden te zijn dan regelmatige werkwoorden en dit effect werd groter wanneer de fonologische regelmaat (gemeten met behulp van de dichtheid van de buren) toenam.

In het tweede deel van hoofdstuk 2 bespreken we de resultaten van ons visuele lexicale-decisie-experiment. Covariantie-analyses hebben aangetoond dat de hierboven genoemde distributionele eigenschappen van regelmatige en onregelmatige werkwoorden goede voorspellers waren voor het aantal gemaakte fouten en de reactietijden in visuele lexicale decisie. Een grotere semantische dichtheid, gemeten met behulp van het aantal sets van synoniemen en karakteriserend voor onregelmatige werkwoorden, leidde tot snellere reactietijden voor onregelmatige verledentijdsvormen.

In hoofdstuk 3 rapporteren we de resultaten van onze woordbenoemingstaken. We hebben ontdekt dat de semantische maten, zoals het aantal sets van synoniemen en de argumentstructuren, niet langer significant zijn, waarschijnlijk doordat ze minder relevant zijn voor een taak zoals woordbenoeming (waarvan bekend is dat de semantische variabelen er minder bij betrokken lijken te zijn dan bijvoorbeeld in lexicale decisie). De paradigmatische effecten bleven wel aanwezig (hoewel het effect van familiegrootte verdwenen is in het experiment waarbij er van tijd gewisseld diende te worden, waarschijnlijk vanwege de hoeveelheid ruis in deze taak). Desalniettemin lijken de voorspellers die gerelateerd kunnen worden aan de vorm van de werkwoorden te domineren in onze woordbenoemingstaken.

In hoofdstuk 4 hebben we de benoeming van foto’s bestudeerd. Semantische maten kwamen weer tevoorschijn in onze analyses (in tegenstelling tot onze woordbenoemingstaken). De onzekerheid over de foto en de complexiteit van de foto bleken een vertragend effect te hebben op de reactietijden. Deze maten waren niet in interactie met regelmaat, wat wederom bewijs levert tegen een zuinig lexicon.

In hoofdstuk 5 gaan we van het Nederlands over naar het Spaans, een taal met een ander lexicalisatiepatroon dan het Nederlands. In dit hoofdstuk presenteren we de resultaten van twee verschillende experimenten die we uitgevoerd hebben: fotobenoeming en visuele lexicale decisie. Voor deze twee experimenten hebben we precies dezelfde materialen gebruikt als voor de corresponderende experimenten in het Nederlands, waarover we in de hoofdstukken 2 en 4 hebben gesproken. Onze hoofdonderzoeksvraag in dit experiment was of de moedertaalsprekers van het Spaans beïnvloed werden door de semantische eigenschappen van een werkwoord dat in het Nederlands wel, maar in het Spaans niet, in het werkwoord zelf gelexicaliseerd kan worden. Onze voorspelling was dat dit verschil tussen de twee talen zou kunnen leiden tot een langzamere verwerkingssnelheid voor de moedertaalsprekers van het Spaans. In beide experimenten bleken er variabelen te zijn die gerelateerd zijn aan het Nederlands maar een effect hadden op de reactietijden in het Spaans. Deze variabelen waren ofwel gebaseerd op de lexicale eigenschappen van de werkwoorden, zoals de Nederlandse derivationele entropie en de morfologische familiegrootte, ofwel verkregen uit de voorafgaande experimenten (zie hoofdstuk 4), zoals de Nederlandse foto-entropie, die de onzekerheid beschrijft van alle Nederlandse proefpersonen samen ten opzichte van een specifiek werkwoord dat op een foto is afgebeeld. Deze effecten waren inhiberend en doen ons, interessant genoeg, vermoeden dat de moedertaalsprekers van het Spaans wel degelijk gevoelig zijn voor de semantische eigenschappen van een werkwoord, zelfs al kunnen deze semantische details in het Spaans niet in het werkwoord zelf gelexicaliseerd worden, wat leidt tot extra onzekerheid bij de Spaanstaligen in verhouding tot de Nederlandstaligen. De *verbalisator* van Bierwisch en Schreuder (1992), die een aanpassing is aan het model van Levelt (1989), zou deze semantische details voor zijn rekening kunnen nemen. We concluderen dat de interpretatie van een foto en mogelijkerwijs ook de conceptualisatie van het daarbijbehorende item gelijk is voor de moedertaalspreker van het Spaans en het Nederlands, maar dat beide talen taalspecifieke strategieën hebben voor de lexicalisatie van dit item.

We kunnen concluderen dat ons lexicaal-statistische onderzoek duidelijkheid heeft gegeven over de kwestie of er sprake is van effecten van semantische differentiatie tussen regelmatige en onregelmatige werkwoorden. Het onderzoek liet zien dat er inderdaad systematische semantische verschillen zijn tussen regelmatige en onregelmatige werkwoorden. Deze effecten zagen we gereflecteerd in ons visuele lexicale decisie-experiment en tot op zekere hoogte in onze fotobenoemingsexperimenten in het Nederlands en het Spaans, maar niet in onze woordbenoemingsexperimenten.

Slotopmerkingen

Het doel van dit proefschrift was om beter te begrijpen hoe regelmatige en onregelmatige werkwoorden geproduceerd en verwerkt worden, waarbij we ons specifiek gericht hebben op de interactie tussen semantiek en regelmaat. De overheersende opvatting binnen het psycholinguïstisch onderzoek naar de morfologische verwerking is dat vorm en betekenis in strikt gescheiden modules van het mentale lexicon zouden kunnen worden ondergebracht. Binnen deze opvatting worden verschillen tussen regelmatige en onregelmatige werkwoorden niet systematisch gerelateerd aan semantische verschillen. Regelmatige vormen zouden gecreëerd worden door regels en onregelmatige vormen zouden direct uit het mentale lexicon opgediept worden. Pinker (1991, 1997, 1999) heeft dit generatieve model voor morfologische verwerking gepopulariseerd door te beweren dat regelmatige en onregelmatige verleden-tijdsvormen in de hersenen neuranatomisch verschillend gelocaliseerd zijn. In de vier voorafgaande hoofdstukken is duidelijk geworden dat het verschil tussen regelmatige en onregelmatige werkwoorden veel complexer is dan Pinker en zijn collega's suggereren. De effecten van inflectionele frequentie en inflectionele entropie hebben in de voorafgaande hoofdstukken aangetoond dat elke geïnflecteerde vorm, regelmatig of onregelmatig, een eigen representatie heeft in het mentale lexicon en dat inflectie blijkbaar niet gedreven kan worden door 'derivationele' regels. Onze bevindingen zijn in lijn met de benaderingswijze van Stemberger en met de *Word and Paradigm morphology*.

Lastiger is het om de vraag te beantwoorden hoe het verschil tussen regelmatige en onregelmatige werkwoorden beïnvloed wordt door de semantiek. We hebben effecten gevonden van semantische differentiatie tussen regelmatige en onregelmatige werkwoorden in ons lexicaal-statistische onderzoek en deze effecten waren ook waarneembaar in de verwerking, zoals de resultaten van ons visuele lexicale-decisie-experiment hebben laten zien. In de productie zijn de mogelijke effecten van semantische differentiatie wellicht gemaskeerd door de effecten van vormelijke variabelen (in onze woordbenoemingstaken) en de effecten van paradigmatische variabelen (in onze fotobenoemingstaken).

Onze resultaten hebben geleid tot de aanname dat semantische eigenschappen wel degelijk de verschillen tussen regelmatige en onregelmatige werkwoorden beïnvloeden, zowel in de verwerking als in de productie, maar dat de productietaken die wij gebruikt hebben niet gevoelig genoeg zijn om deze effecten uit de ruis tevoorschijn te laten komen. Daarom kan het de moeite waard zijn om de productie van regelmatige en onregelmatige werkwoorden (nog) gedetailleerder te bestuderen dan wij voor dit proefschrift gedaan hebben. De resultaten van een fotobenoemingsexperiment waarbij ERPs gebruikt zijn (Hendrix, Tabak, Baayen en Schreuder, in voorbereiding), hebben de electroencephalografische (EEG) correlaten van een verscheidenheid aan psycholinguïstische predictoren onderzocht, waarbij ze van hetzelfde paradigma en dezelfde materialen gebruik gemaakt hebben als wij in de voorafgaande hoofdstukken. Hendrix et al. hebben zowel aan vorm als aan betekenis gerelateerde frequentie-effecten gevonden en, het belangrijkste, ze hebben gevonden dat semantische eigenschappen zowel de regelmatige als de onregelmatige werkwoorden beïnvloeden, wat overeenkomt met onze bevindingen. We concluderen dat semantiek een cruciale factor is in de verwerking van zowel regelmatige als onregelmatige inflectie.

Referenties

Aronoff, M.: 1994, *Morphology by itself: stems and inflectional classes*, The MIT Press, Cambridge, Mass.

Beard, R.: 1995, *Lexeme-morpheme base morphology: a general theory of inflection and word formation*, State University of New York Press, Albany, NY.

Bierwisch, M. and Schreuder, R.: 1992, From concepts to lexical items, *Cognition* **42**, 23–60.

Blevins, J. P.: 2003, Stems and paradigms, *Language* **79**, 737–767.

Hendrix, P., Tabak, W., Schreuder, R., and Baayen, R. H.: n.d., Electrophysiological effects in language production: a picture naming study using generalized additive modeling. in preparation.

Kimura, D.: 2000, *Sex and cognition*, The MIT press, Cambridge, MA.

Levelt, W. J. M.: 1989, *Speaking. From intention to articulation*, The MIT Press, Cambridge, Mass.

Matthews, P. H.: 1974, *Morphology. An Introduction to the Theory of Word Structure*, Cambridge University Press, London.

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.

Stemberger, J. and Middleton, C.: 2003, Vowel dominance and morphological processing, *Language and Cognitive Processes* **18**(4), 369–404.

Stemberger, J. P.: 2002, Overtensing and the effect of regularity, *Cognitive Science* **26**, 737–766.

Stemberger, J. P.: 2004, Phonological priming and irregular past, *Journal of Memory and Language* **50**, 82–95.

Acknowledgments

The work is done! Finally, I can express my gratitude to all the people without whom I would never have reached this point. First of all, I would like to thank my two *promotores*, Rob Schreuder and Harald Baayen, who gave me the opportunity to learn about psycholinguistic research in an extraordinary scientific environment. I certainly learned a lot from their rich experiences. I want to thank Harald for his continuous enthusiasm and energy and for our inspiring sessions of analysing data, writing and brainstorming. I am also deeply indebted to Rob for his unconditional support and confidence. He always provided a clear view on the data and reminded me not to get lost in details.

Furthermore, I owe many thanks to all my colleagues at the CLSM research group (formerly known as the IWTS). Thanks to Anneke, Arina, Debby, Francisco, Iris, Kors, Lanneke, Loes, Lucie, Marco, Marieke, Marjolein, Mirjam, Nienke, Nicole, Patricia, Roel, Victor and all the others who have been there over the years, the CLSM always was a very pleasant working environment for me. I am indebted to Mirjam for always providing a working space for me. I want to thank Anton, Dan, Helen, Laura and Mark for being such good roommates. Special thanks go to Karen, who was my roommate during all these years at the MPI, who became a good friend and who agreed to be my paranymph. I would like to thank Else, Femke, Iris, Josje, Karen and Suzanne for our nice dinner meetings during my first few years as a PhD student. I want to thank Dieuwke for being such a good travel buddy on conferences and for joining me at the MPI so we could have lunch, coffee, talks and walks together.

For their practical support, I am indebted to a lot of student assistants that helped me run the experiments. Thank you so much, Anneli, Berit, Femke, Hanke, Hilde, Inge, Jeanne, Luuk, Marieke and Mybeth! Many thanks go also to numerous people for participating in my experiments.

I wish to thank Hanke van Buren and Peter Hendrix for allowing me to be supervisor for their master theses. Special thanks go to Hanke for her contribution

to this dissertation (Chapter 5).

Let me also acknowledge the members of the manuscript committee, Mirjam Ernestus, Ulli Frauenfelder and Roeland van Hout, for their valuable feedback on the dissertation.

I am grateful for the professional support of the technical and administrative staff of the Max Planck Institute. It is a privilege for me to publish this dissertation in the MPI Series for Psycholinguistics.

I would like to thank my friends for all those pleasant hours shared together during the past few years. Special thanks go to Aukje, Berit and Wytske who made life in Nijmegen so much fun. I am honored to have Aukje as my second paranymph.

I want to thank the nannies at the *Kinderopvang Heyendael* for their tremendous job in taking care of our children. Although it was not always easy to leave the children for something like regular and irregular verbs, I always knew they were in safe and caring hands, which made it so much easier. I would like to thank my parents and parents in law as well for taking such good care of our children during my working days. I always enjoyed the happy voices of my boys when I called home to hear if everything was alright. Thank you!

I want to thank my parents for always believing in me. Without their support, throughout the years, this book would not have been written. I would like to thank Jille, Nanda, Jasper and Stijn for being the best brother, sister in law and nephews one can wish.

Thijmen and Joris, thank you for being my wonderful boys. I could not love you more. Thijmen, thank you for your beautiful drawing on the cover of this book. Joris, I promise, next book will be yours!

Wouter, my dearest, how can I thank you enough? *Omnia vincit amor, et nos cedamus amori.*

Curriculum Vitae

Wieke Tabak was born in Emmen, The Netherlands, on 29 June 1979. After her graduation from the Gymnasium in Emmen in 1997, she initially wanted to become a primary school teacher and went to the Stenden Hogeschool in Groningen. In 1998 she received her propaedeutics. In 1999, she started studying Scandinavic Languages and Cultures at the University of Groningen. In 2000, she also started studying General Linguistics at the same university. Within her studies of Scandinavic Languages and Cultures she specialized in Swedish Linguistics and within her studies of General Linguistics her specialization was Neurolinguistics. Wieke went to Lund, Sweden, in 2003 to do her internship at the Department of Logopedics, Phoniatrics and Audiology at the University of Lund. She graduated in both Scandinavic Languages and Cultures and General Linguistics in 2003. Her MA thesis was on specific language impairment in Swedish and Dutch, focusing on the use of subjects and articles. In January 2004, she started a PhD project entitled "Semantics and (ir)regularity in morphological processing" at the Interfaculty Unit for Language and Speech (IWTS, now CLSM) at the Radboud University of Nijmegen. After having two sons, in 2006 and in 2008, she finished her PhD in 2010. Currently, Wieke works as a lecturer at Fontys Hogeschool in Den Bosch.

Publications

Tabak, W., Schreuder, R. and Baayen, R. H. (2005) Lexical statistics and lexical processing: Semantic density, information complexity, sex, and irregularity in Dutch. *Linguistic evidence — Empirical, Theoretical, and Computational Perspectives*, p. 529–555. Berlin: Mouton de Gruyter.

Tabak, W., Schreuder, R. and Baayen, R. H. (2005) The processing of regular and irregular verbs. Kepser, S. and Reis, M. (eds.) *Proceedings of Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, p. 121-126. Saarbrücken.

Tabak, W., Schreuder, R. and Baayen, R. H. (2010) Producing inflected verbs: a picture naming study. *The Mental Lexicon* 5(1), pp. 22-46.

Tabak, W., Schreuder, R. and Baayen, R. H. (submitted) Is inflection derivational? Evidence for a redundant lexicon from word naming. *Quarterly Journal of Experimental Psychology. Chapter 3 of this dissertation.*

Tabak, W., Van Buren, H., Schreuder, R. and Baayen, R. H. (submitted) A cross-language perspective on processing in Spanish and Dutch. *Morphology. Chapter 5 of this dissertation*

Hendrix, P., Tabak, W. and Schreuder, R. and Baayen, R. H. (in preparation) Electrophysiological effects in language production: a picture naming study using generalized additive modeling.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*

2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsaggital articulography. *Niels O. Schiller*

3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*

4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*

5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*

6. Gesture and speech production. *Jan-Peter de Ruiter*

7. Comparative intonational phonology: English and German. *Esther Grabe*

8. Finiteness in adult and child German. *Ingeborg Lasser*

9. Language input for word discovery. *Joost van de Weijer*

10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*

11. Producing past and plural inflections. *Dirk Janssen*

12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*

13. From speech to words. *Arie van der Lugt*

14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*

15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*

16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*

17. Moving eyes and naming objects. *Femke van der Meulen*

18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*

19. Morphology in speech comprehension. *Kerstin Mauth*

20. Morphological families in the mental lexicon. *Nivja H. de Jong*

21. Fixed expressions and the production of idioms. *Simone A. Sprenger*

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*

29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Broersma*

35. Retrieving semantic and syntactic word properties. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*

43. Affix reduction in spoken Dutch: Probabilistic effects in production and perception. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*

45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita E. Wagner*

50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*

51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*

52. Prososdic structure in speech production and perception. *Claudia Kuzla*

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*

54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*

55. Processing the fine temporal structure of spoken words. *Eva Reinisch*

56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*