Morphological families in the mental lexicon

ISBN: 90-6464-813-1
Cover illustration: ”Wil je appeltaart van de appelboom?” by John de Jong, 2002. After Albrecht Dürer, *Adam and Eve* 1507; Oil on panel, Museo Nacional del Prado, Madrid
Printed and bound by Ponsen & Looijen bv, Wageningen

Morphological families in the mental lexicon

een wetenschappelijke proeve
op het gebied van Letteren

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen
op gezag van de Rector Magnificus Prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op maandag 16 september 2002
des namiddags om 1.30 uur precies

door
Nivja Helena de Jong
geboren op 11 juni 1976 te Meppel

Promotor: Prof. dr. R. Schreuder
Co-promotor: Dr. R. H. Baayen

Manuscriptcommissie: Prof. dr. W. Vonk
Prof. dr. H. J. Schriefers
Prof. dr. D. Sandra (Universiteit Antwerpen – UFSIA)

The research reported in this thesis was supported by a PIONIER grant from the Dutch National Research Council NWO, the Faculty of Arts of the University of Nijmegen (The Netherlands) and the Max-Planck-Institut für Psycholinguistik, Nijmegen (The Netherlands).

Voorwoord

De afgelopen jaren bij het Max Planck Instituut heb ik veel plezier gehad. Ik denk bijvoorbeeld aan het congres in Aix-en-Provence tijdens mijn allereerste werkweek als Ph.D., de andere congressen en reizen, de woensdagavond Ph.Beer, de half-één-IWTS-lunch, mijn onvergetelijke verjaardag tijdens de workshop in Nijmegen, de (werk-)sfeer op het MPI, en die in de pioniergroep in het bijzonder. De Ph.D's, IWTS'ers, MPI'ers, student-assistenten en buitenlandse collega's wil ik bedanken voor dit plezier en de goede samenwerking.

Maar het plezier dat dit proefschrift tot stand deed komen kwam vooral door de samenwerking met Harald en Rob. Ik was verbaasd toen ze mij al op de eerste dag van mijn stage serieus namen. Mijn opmerkingen sloegen waarschijnlijk de plank vreselijk mis, maar de discussies die ik door deze openheid van Harald en Rob al vanaf het begin kon voeren waren vreselijk leerzaam en werden alleen maar leuker.

Thanks also to Laurie and Matt, who made my stays in Albany very fruitful and pleasant, despite the difficult time after Brendon's accident. I thank Anke and Mark for their hospitality in Stuttgart and Seattle, en Mandy voor de prettige samenwerking in Gent. Mijn kamergenoten Andrea, Fermín, en Rachel bedank ik voor het dagelijks luisterend oor zodat ideetjes of klachtjes meteen een klankbord kregen.

Fermín heeft er bovendien voor gezorgd dat ik behalve in Leiden, ook gezelligheid in Nijmegen kon vinden. Samen met Fenneke als de Leidse paranimf, is hij de Nijmeegse. Leiden in zijn geheel wil ik ook bedanken, waarbij Amsterdam en Den Haag als dépendances gelden. Voor er elk weekend zijn, zodat ik met niks aan of in mijn hoofd het weekend kon vieren. Niks behalve de avonden van Literis, liefdes, terrasjes, Utrecht halverwege, 't Plantsoen, mooie brieven, dansen, en semesters die op schijnbaar elk willekeurig moment konden aflopen.

Voor de fantastische illustratie op de voorkant bedank ik John. En hoezeer Leiden soms als een hechte familie fungeerde, uiteindelijk zocht en vond ik als het volgens mijn denkraam tegenzat steun bij Anne Mie en John.

Nijmegen, juni 2002.

Contents

Voorwoord **1**

1 Introduction **7**

References 14

2 Explorations on the family size effect **17**

Introduction 18

Experiment 2.1: Comparing nouns and verbs 23

Experiment 2.2: Family size or family frequency for complex words 26

Experiment 2.2a: Contrasting family frequency 26

Experiment 2.2b: Contrasting family size 28

Experiment 2.3: The role of an inflectional suffix 30

Experiment 2.3a: Verbs with an inflectional suffix 30

Experiment 2.3b: Comparing verbs with and without an inflectional suffix 32

Experiment 2.4: Is the family size effect dependent on form? 34

Experiment 2.4a: Regular past participles 34

Experiment 2.4b: Irregular past participles 36

Post-hoc analyses 39

General Discussion 43

References 50

3 Dutch and English compounds **59**

Introduction 60

Experiment 3.1: Type or token counts for constituents of Dutch compounds 62

Experiment 3.2: English concatenated and open compounds 67

Experiment 3.3: Dutch artifical open compounds 70

Experiment 3.4: The status of diffently-spelled family members 71

General Discussion 75

References 78

4 The effect of morphological and sentential context 81
Introduction 82
Experiment 4.1: Morphological context 86
Experiment 4.2: Sentential context 87
Post-hoc correlations 90
Simulation studies 94
General Discussion 100
References 103

5 Homonyms in context 105
Introduction 106
Experiment 5.1: Noun-noun homonyms 112
Experiment 5.2: Noun-verb homonyms 122
General Discussion 131
References 139

6 Effects of frequency, family size, and age of acquisition in a range of tasks 147
Introduction 148
Experiment 6.1: Perceptual Identification 151
Experiment 6.2: Lexical Decision with illegal nonwords 154
Experiment 6.3: Lexical Decision with legal nonwords 156
Experiment 6.4: subjective frequency Rating 156
Experiment 6.5: Semantic Association generation 157
Experiment 6.6: Semantic Decision 158
Experiment 6.7: Lexical Decision with novel derivations 159
Discussion of the Results 161
Regression analyses 163
General Discussion 167
References 172

7 Summary and discussion 187
The family size effect is robust 188
No effect for constituents in compounds 190
Words differ from phrases 191
Context defines the co-activation 193

Family size *and* frequency are context-sensitive 195
The effect is independent from age of acquisition 197
The meanings of family members are co-activated 199
References 203

Samenvatting **207**
References 212

Curriculum Vitae **213**

Introduction

Chapter 1

Words occur in many different contexts. The same word may show up in newspapers, novels, and menus. It can surface next to different other words. A word may even occur within different other words. For instance, the Dutch word *werk*, ’work’, is a constituent of many morphologically complex words, among which are *huiswerk*, ’homework’, *werkbaar*, ’workable’, and *verwerken*, ’to process’. In fact, in a corpus of 42 million words (CELEX; Baayen, Piepenbrock, & Gulickers, 1995), *werk* is the most productive Dutch word and has about 500 morphological descendants. The morphological productivity of most words in Dutch, however, is much more restricted. Half of the monomorphemic words occurs in at most 3 complex words.

This thesis is concerned with the role of a word’s morphological productivity during the lexical processing of the word itself. Using a database such as CELEX, we can obtain two different counts of the morphological productivity of a word. The first is a simple type count of the number of words in which a word occurs, the family size. The second is a token count of the total number of times these family members occur, the family frequency.

Schreuder & Baayen (1997) first disentangled these two counts and showed that for Dutch monomorphemic words in a visual lexical decision task, it is the family size that predicts reaction times. The larger the morphological family, the faster and more accurate participants decide whether a word is an existing word. The summed frequency of the family members, the family frequency, was found to be irrelevant. In a subjective frequency rating task, the family size also emerges as a predictor: The larger the family size of a word, the more likely participants are to estimate the word as being high in frequency. The effect is independent from other effects that are known to play a role in visual lexical decision or rating, such as frequency of occurrence or word length. The family size effect shows that monomorphemic words are not independent islands in the mental lexicon. They are connected with morphologically complex relatives. In Bybee's model of morphological structures in the mental lexicon (1988), morphologically related words are indeed linked in form as well as meaning. Surprisingly, these connections play a role when processing the monomorphemic words themselves.

The frequency of occurrence of a word has widely been recognized as an important variable in lexical processing (e.g., Forster & Chambers, 1973; Monsell, Doyle, & Haggard, 1989; Whaley, 1978). Many models of lexical processing posit its facilitatory role at the level of access (e.g., Baayen, Dijkstra, & Schreuder, 1997; Bradley & Forster, 1987; Morton, 1969): It is easier to recognize a string of letters if this string of letters is frequent. If the role of the morphological family likewise plays a role at the level of access, one would expect the family frequency, a variable that captures the string familiarity of a word within other words, to play a role. The fact that it is the type count of the morphological family members that reduces reaction times suggests that the role of the morphological family is located at a more central level of lexical processing.

Therefore, Schreuder and Baayen (1997) investigated whether the effect of family size would also be apparent in a task that presumably exclusively taps into form levels of processing. They found that in a progressive demasking task (in which participants have to identify visually presented words that gradually emerge from a mask of hash-marks on the computer screen), no effect of family size could be observed. Furthermore, in post-hoc correlations, both Schreuder and Baayen (1997) as well as Bertram, Baayen, and Schreuder (2000) reported that excluding semantically opaque words from the count of family size improves the correlations with reaction times. This suggests that it is not the number of words that overlap in form that reduces reaction times, but rather, it is the number of words that overlap in meaning.

In a rating task, Baayen, Lieber, and Schreuder (1997) replicate the effect of family size for English, as well as the lack of an effect of family frequency. For other, non-germanic languages, effects similar to the family size effect have been found.

Feldman, Pnini, and Frost (1995) reported that for Hebrew, the productivity of the root plays a role in the segment-shifting task. In this task, participants are asked to shift a (marked) morphological pattern, the root, from an existing word to another word or pseudoword. They found that if the morphological pattern is opaque (i.e., if it occurs in one single word only), participants are significantly slower than when the pattern is transparent (i.e., if it occurs in more than one word). They also reported on a lexical decision task in which they find that words containing roots that only occur in one single word are responded to more slowly than words with productive roots.

For Chinese, Taft and Zhu (1995) showed that Chinese compounds containing characters that occur in several other compounds are named faster than compounds with characters that occur in that one compound only. Similarly, Taft and Zhu (1997) and Feldman and Siok (1997) showed that words containing Chinese radicals that are productive are responded to faster in visual lexical decision than words containing radicals that are not productive. For Japanese, Yamada and Kayamoto (1998) showed that two-kanji words are responded to faster if the productivity of the kanjis is high.

For Finnish, Hyönä and Pollatsek (1998) showed that the productivity of the first constituent of Finnish compound words affects eye-movements in reading. If the first constituent of a Finnish compound has a large family, eye-fixation latencies on this first constituent are shorter. In none of these studies, however, the family frequency of the constituents (roots, characters, radicals, kanjis, or words), has been taken into account. This leaves open the question whether the reported effects in Hebrew, Chinese, Japanese, and Finnish are true type count effects of productivity, or rather effects of the corresponding token frequencies.

The aim of this thesis is to provide a better understanding of what the effect of the type count of the morphological family reflects. As already mentioned, the morphological family size effect is likely to play a role at the semantic level of lexical processing. The interpretation of this effect as offered by Schreuder and Baayen (1997) is that upon reading a word, its family members become co-activated. This would explain the facilitatory effect in the lexical decision task as well as its effect in the rating task: Words that co-activate many other words lead to a larger global activation in the mental lexicon. The more global activation, the faster participants are able to respond (see Grainger & Jacobs, 1996), and the more likely they are to rate a word as being high-frequent. This interpretation leaves us with the question why family members might be co-activated. The function of the effect can hardly be making the lexical decision task easier. This thesis will show that the effect of family size reflects spreading of semantic activation in the sense of Collins and Loftus (1975), along the lines of morphologically related words. In fact, this will lead to the prediction that the morphological structure as proposed by Bybee (1988) has implications in on-line processing. Furthermore, the restraining effects of context on the co-activation of family members reveals that the spreading of activation serves a functional purpose.

This thesis is organized as follows. After assessing that the effect is quite robust (Chapter 2) and investigating the effect for constituents in compound words (Chapter 3), we will turn to the role of context (Chapters 4 and 5). If the effect of family size reflects semantic spreading of activation, it is likely that the context in which a word occurs has an influence. Finally, we will extend the experimental paradigm from visual lexical decision to different tasks, ranging from tasks presumably sensitive to effects of form characteristics, to tasks that are sensitive to semantic levels of processing (Chapter 6).

In Chapter 2, we investigate the robustness of the family size effect, using visual lexical decision. First, we will ascertain whether the effect, which has sofar been reported for nouns (or for verbs with nominal conversion alternants), extends to Dutch verbs. Second, we will look at the role of family size of base words in suffixed words. Third, we will investigate the role of morphological context for the effect of family size. We will compare the effect of family size for verbs with and without the third personal singular marker, the inflectional suffix *-t* (e.g., *sjouw*, ’drag’ and *sjouwt*, ’drags’). Finally, if the effect of family size is truly semantic, it should not depend on exact form overlap between the target word to which the participants respond and its morphological family members. Therefore, we compare the effect of family size for regularly formed past participles with the effect for irregularly formed past participles. The regular past participles overlap in form with their family members, whereas the irregular past participles do not (e.g., the irregular past participle *gevochten*, ’fought’, has family members such as *vechter*, ’fighter’).

In Chapter 3, we turn to the morphological family of constituents in Dutch and English compound words, again using visual lexical decision. Van Jaarsveld and Rattink (1988) showed that the frequency of the constituents of Dutch compounds does not affect reaction times. They only find an effect of the frequency of the compound as a whole. At the same time, in priming studies, Zwitserlood (1994) and Sandra (1990) reported that both constituents of a compound word can be semantically primed. For novel compounds, Van Jaarsveld, Coolen, and Schreuder (1994) showed that the productivity of constituents in compounds can have an effect. They presented lexicalized and novel Dutch compounds. If the novel compounds contained highly-productive constituents, participants took longer to decide that the novel compound did not exist. This inhibitory effect of the productivity can be explained as an effect of family size: The higher the productivity of the constituents, the more family members are activated. This high activation in the mental lexicon makes it harder for participants to decide that the compound does not exist. However, it is unclear whether it is the type count that affects reaction times or whether it is the summed frequencies of the family members. The same holds for the studies by Hyönä and Pollatsek (1998) who reported on an effect of the productivity of Finnish constituents, and for the studies by Taft and Zhu (1997) and Feldman and Siok (1997) who reported on an effect of the productivity of radicals in Chinese compound characters. We will explore for Dutch and English which effects of the morphological families of the constituents play a role.

The constituents of Dutch compounds are generally concatenated. English compound words can be spelled in different ways: The constituents can be concatenated as in Dutch, they can be hyphenated, or they can be written with a space between the constituents. Inhoff, Radach, and Heller (2000) reported that initial reading times for German compounds, which are normally concatenated, speed up if the constituents are written with an unconventional space between the constituents. The semantic interpretation of the compound words, however, is hindered. As we hypothesize that the morphological family size effect is semantic in nature, we will compare the effects of the morphological family for constituents in English that are conventionally concatenated with those of constituents in compounds that are conventionally written with a space (e.g., *applesauce* and *apple pie*). Surprisingly, we do find differences, in that the constituents in open compounds are processed more like words in isolation. Finally, we will explore whether the status of family members that are written with a space or a hyphen differs from family members which are concatenated.

In Chapter 4, we will systematically investigate the role of morphological and sentential context on the effect of family size for the same set of target words. Bertram et al. (2000) reported that a specific subset of morphological family members for de-adjectival nouns ending in *-heid* (similar to the English *-ness*) did not contribute to the effect of family size. In post-hoc correlations, they found that a semantically defined subset of family members should be excluded from the family size counts. We therefore investigate adjectives in four different contexts, using visual lexical decision. We compare the effects of family size for adjectives presented in their base form, in the comparative form, and in two minimal sentential contexts. In the sentential contexts, we present the adjectives with the words *heel*, ’very’, and *niet*, ’not’. In order to gain insight in differential effects for different subsets of family members, we divide the total family size counts into four subsets: nouns, verbs, and two kinds of adjectives. We will correlate the reaction times with these subsets of family members to assess which of the family members contribute to the effects of family size. We will see that across contexts, the role of the different subsets of family members differs dramatically. In the latter part of this Chapter, we will offer simulation studies using a computationally simple model that explains the effect of family size as spreading of activation along morphological relations.

For homonyms, it is clear that the context must narrow down the meaning. In Chapter 5, we will compare the effects of the two semantically distinct families of the two meanings of a homonym in different contexts. We can also obtain differential frequency counts for these two meanings. We first investigate the role of the two frequency and family size counts for noun-noun homonyms, such as *wapen* (meaning ’weapon’ or ’coat of arms’) and use primes as disambiguating context in visual lexical decision and rating tasks. As we find in Chapter 4 that minimal context can already have a big influence on which of the family members contribute, we also consider noun-verb homonyms, such as *last* (meaning ’(the) burden’ or ’(he/she) welds’). These homonyms can be disambiguated using minimal context, unlike the noun-noun homonyms. We will use different experimental lists as disambiguation, by adding filler words that are unambiguously nouns or that are unambiguously verb forms in visual lexical decision and rating tasks. In a visual lexical decision task, we also use minimal sentential context to disambiguate, namely by preceding the homonym by either a personal pronoun or definite article (e.g., *de last* or *hij last*, respectively ’the burden’ of ’he welds’) .

The role of context for initial meaning-activation for homonyms has been the topic of hot debate in psycholinguistics for a long time (for a recent overview, see Twilley and Dixon, 2000). We do not claim that effects of family size and frequency show initial meaning-activation, but we do find that both counts are extremely sensitive to context. The fact that besides family size, frequency counts are sensitive to context as well, casts serious doubt on the claim (e.g., Baayen, Dijkstra, & Schreuder, 1997; Bradley & Forster, 1987; Morton, 1969) that frequency effects are reflecting form levels of processing exclusively.

Besides frequency of occurrence and family size, the age at which a word is first learned, also affects performance in various tasks (e.g., Brysbaert, De Lange, & Van Wijnendaele, 2000; Carrol & White, 1973). In Chapter 6, we tease apart the effects of frequency, family size, and age of acquisition. Ghyselinck, Lewis, & Brysbaert (submitted) and Ellis and Lambon Ralph (2000) claim that age of acquisition and frequency affect the same stages of lexical processing. However, in the studies that investigate these two variables, the family size of the words has not been taken into account. Similarly, in the studies that investigate family size, the age of acquisition of the words has not been taken into account. It is very likely that there is, indeed, a confound between these variables. Words that are acquired early in life not only tend to be high in frequency, but probably also tend to be the words with the highest degree of morphologically productivity. It seems difficult, however, to understand how the observed effects of context for the morphological families can be accounted for in terms of age of acquisition. We therefore systematically investigate the three variables in seven different tasks, which all use visual presentation as a starting point. The tasks range from those that presumably tap into early stages of processing to tasks that require semantic processing. We find all three variables to be independent reliable predictors. We show that the effects of the morphological family members in the semantic tasks provide very strong support for the claim that the effect of family size reflects spreading of semantic activation in the mental lexicon.

References

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., Lieber, R. and Schreuder, R.: 1997, The morphological complexity of simplex nouns, *Linguistics* **35**, 861–877.

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

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

Bradley, D. C. and Forster, K. I.: 1987, A reader's view of listening, *Cognition* **25**, 103–134.

Brysbaert, M., Lange, M. and Van Wijnendaele, I. .: 2000a, The effects of age-of-acquisition and frequency-of-occurence in visual word recognition: Further evidence from the Dutch language, *European Journal of Cognitive Psychology* **12**, 65–85.

Bybee, J. L.: 1988, Morphology as lexical organization, *in* M. Hammond and M. Noonan (eds), *Theoretical Morphology: Approaches in Modern Linguistics*, Academic Press, London, pp. 119–141.

Carroll, J. B. and White, M. N.: 1973, Word frequency and age of acquisition as determiners of picture-naming latency, *Quaterly Journal of Experimental Psychology* **25**, 85–95.

Collins, A. M. and Loftus, E. F.: 1975, A spreading-activation theory of semantic processing, *Psychological review* **82**, 407–428.

Ellis, A. W. and Lambon Ralph, M. A.: 2000, Age of acquisition effects in adult lexical processing reflects loss of plasticity in maturing systems: Insights from connectionist networks, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26**, 1103–1123.

Feldman, L. B. and Siok, W. W. T.: 1997, The role of component function in visual recognition of Chinese characters, *Journal of Experimental Psychology: Learning, Memory and Cognition* **23**, 778–781.

Feldman, L. B., Pnini, T. and Frost, R.: 1995, Decomposing words into their constituent morphemes: Evidence from English and Hebrew, *Journal of Experi-*

mental Psychology: Learning, Memory, and Cognition **21**, 947–960.

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

Ghyselinck, M., Lewis, M. B. and Brysbaert, M.: *submitted*, Age of acquisition and the cumulative-frequency hypothesis: A multi-task investigation.

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

Hyönä, J. and Pollatsek, A.: 1998, Reading finnish compound words: Eye fixations are affected by component morphemes, *Journal of Experimental Psychology: Human Perception and Performance* **24**, 1612–1627.

Inhoff, A. W., Radach, R. and Heller, D.: 2000, Complex compounds in german: Interword spaces facilitate segmentation but hinder assignment of meaning, *Journal of Memory and Language* **42**, 23–50.

Monsell, S., Doyle, M. C. and Haggard, P. N.: 1989, Effects of frequency on visual word recognition tasks, *Journal of Experimental Psychology: General* **118**, 43–71.

Morton, J.: 1969, The interaction of information in word recognition, *Psychological Review* **76**, 165–178.

Sandra, D.: 1990, On the representation and processing of compound words: automatic access to constituent morphemes does not occur, *Quarterly Journal of Experimental Psychology* **42A**, 529–567.

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

Taft, M. and Zhu, X.: 1995, The representation of bound morphemes in the mental lexicon: A Chinese study, *in* L. B. Feldman (ed.), *Morphological Aspects of Language Processing*, Lawrence Erlbaum, Hillsdale, New Jersey, pp. 293–316.

Taft, M. and Zhu, X.: 1997, Submorphemic processing in reading Chinese, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **23**, 761–775.

Twilley, L. C. and Dixon, P.: 2000, Meaning resolution processes for words: A parallel independent model, *Psychonomic Bulletin & Review* **7**, 49–82.

Van Jaarsveld, H. J. and Rattink, G. E.: 1988, Frequency effects in the processing of lexicalized and novel nominal compounds, *Journal of Psycholinguistic Research* **17**, 447–473.

Van Jaarsveld, H. J., Coolen, R. and Schreuder, R.: 1994, The role of analogy in the interpretation of novel compounds, *Journal of Psycholinguistic Research* **23**, 111–137.

Whaley, C. P.: 1978, Word–nonword classification time, *Journal of Verbal Language and Verbal Behavior* **17**, 143–154.

Yamada, J. and Kayamoto, Y.: 1998, Valency, secondary frequency, and lexical access: A Japanese study, *Applied Psycholinguistics* **19**, 87–97.

Zwitserlood, P.: 1994, The role of semantic transparency in the processing and representation of Dutch compounds, *Language and Cognitive Processes* **9**, 341–368.

Explorations on the family size effect

CHAPTER 2

This chapter has been published as Nivja H. De Jong, Robert Schreuder, and R. Harald Baayen (2000): The morphological family size effect and morphology, *Language and Cognitive Processes*, **15**, 329–365.

Abstract

Schreuder and Baayen (1997) report that in visual lexical decision response latencies to simplex nouns are shorter when these nouns have large morphological families, i.e., when they appear as constituents in large numbers of derived words and compounds. This study presents the results of four experiments that show that verbs have a Family Size effect independently of nominal conversion alternants, that this effect is a strict type frequency effect and not a token frequency effect, that the effect is co-determined by the morphological structure of the inflected verb, and that it occurs irrespective of the orthographic shape of the base word.

Introduction

It is well known that various token frequency counts affect response latencies to simplex and complex words in visual lexical decision. Taft (1979), and more recently Baayen, Dijkstra, and Schreuder (1997) showed that the Surface Frequency of a complex word, i.e., its own string frequency, as well as its Base Frequency, i.e., the summed frequency of the inflectional variants of a word, co-determine lexical processing. A third frequency measure, the Cumulative Root Frequency, the summed frequencies of all forms in which a free or bound stem occurs, has also been found to influence response latencies of complex words (Taft, 1979; Colé, Beauvillain, & Segui, 1989). The Cumulative Root Frequency and the Base Frequency are not independent counts: the Cumulative Root Frequency is equal to the sum of the Base Frequency on the one hand, and the cumulated frequencies of morphologically related family members on the other hand, to which we will refer as the Family Frequency.

Table 2.1: Inflectional variants and family members of *calculate* in the CELEX lexical database. Token counts based on a corpus of 18 million words.

Inflected Forms	Surface Frequency	Family members	Base Frequency
calculate	108	*calculate*	574
calculated	340	*calculable*	4
calculates	21	*calculation*	343
calculating	105	*calculator*	89
		calculus	50
		incalculable	26
		incalculably	1
		miscalculate	5
		miscalculation	25

Table 2.1 illustrates these counts for the English verb *calculate* using the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). This verb has four inflectional variants, listed in the first column together with their individual Surface Frequencies. Their summed Surface Frequencies constitute the Base Frequency of the lemma *calculate*. In addition, the third column lists the eight morphological family members of *calculate* together with their own Base Frequencies. The Cumulative Root Frequency of *calculate* is obtained by summation of all Base Frequencies: 1117. The Family Frequency equals Cumulative Root Frequency minus

the Base Frequency of *calculate* itself: 543.

Recently, it has been observed (Schreuder & Baayen, 1997; Baayen, Lieber, & Schreuder, 1997) that for Dutch and English simplex words the Family Frequency does not affect response latencies. Instead of this token count, a type count of the number of morphological family members, 8 for *calculate*, to which we shall refer as the Family Size, has been found to be a strong independent co-determinant of response latencies. (A morphological family member is a complex word in which a given simplex word appears as a constituent.) Thus, a word with a high Family Size such as *acid* (Surface Frequency 277, Base Frequency 384, Family Size 24, and Family Frequency 50) is responded to more quickly in visual lexical decision and rated higher in subjective frequency estimation than a matched control word with a low Family Size such as *skull* (Surface Frequency 305, Base Frequency 370, Family Size 2, and Family Frequency 21).

The observation that not a token count but a type count of the family members is the crucial variable suggests that, surprisingly, words that are not present in the visual input, but that are morphologically related to the target word in the input, are co-activated in the mental lexicon. Schreuder and Baayen (1997) discuss evidence that the effect of Family Size is a central, semantic effect. They found that the Family Size effect disappears in progressive demasking, and they argue that this suggests that it arises at more central levels of lexical processing.

Grainger and Jacobs (1996) and Grainger and Segui (1990) argue that progressive demasking is a perceptual identification task which taps into the early stages of visual processing. This interpretation of progressive demasking is not completely self-evident, however, as the long time period during which a word gradually emerges from its mask as well as the initially highly degraded nature of the input might allow central processes and metalinguistic cognitive strategies to influence the decision to initiate response execution (see Paap & Johansen, 1994).

In this alternative interpretation of progressive demasking, the absence of a Family Size effect can be explained as follows. Suppose that in progressive demasking, just as in auditory lexical decision, multiple lexical candidates are considered over time. In both tasks, the input is not completely and immediately available in the signal, in contrast to what happens in visual lexical decision. Suppose, furthermore, that semantic representations of lexical candidates are activated along with their forms, as has been claimed for the auditory modality by, for instance, Zwitserlood (1989) and Marslen-Wilson, Zhou, and Ford (1997). In these circumstances, the activation of multiple candidates would lead to the activation of multiple morphological families, masking the specific Family Size effect of the target word itself.[1]

This alternative interpretation of progressive demasking is not at odds with the hypothesis that the effect of Family Size in visual lexical decision is a central, semantic effect. Independent evidence for this hypothesis is that Schreuder and Baayen (1997) found that the removal of semantically opaque family members from the count of the Family Size leads to somewhat improved correlations with response latencies in visual lexical decision.

Recently, Bertram, Baayen and Schreuder (2000) report further independent evidence for the semantic nature of the Family Size effect. This study extends the investigation of Family Size effects from monomorphemic nouns to complex words with different affixes. For the derivational suffix *-heid* ('-ness'), the removal of opaque family members of the adjectival base words was crucial for obtaining a reliable correlation of reaction time in visual lexical decision and the count of family members in the by-item analysis. Furthermore, semantic selection restrictions on the affixation of *-heid* were observed to determine which family members contribute to the Family Size effect. In the present paper, further evidence for the central nature of the Family Size effect will be presented.

The effects of Family Size observed thusfar are not without consequences for theories of the processing of morphologically complex words and of the way in which such words are represented in the mental lexicon. First, consider Taft and Forster (1975)'s classic serial search model. In this model, affixes are removed from the visual input and the resulting stem is used to access a central bin, in which all complex forms containing that stem as a constituent are listed in order of decreasing frequency. Such a list is searched serially until a match with the input is obtained. Interestingly, these bins are organized on the principle of what we have called morphological families. Although the work of Taft and Forster is to our knowledge the first to explicitly accord a role to morphological families, the prediction that follows from a serial search through family bins is that words with large families should on average give rise to longer response latencies than words with small families, contrary to what Bertram et al. (2000) observed.

---

[1] We ran an auditory replication of Experiment 3 of Schreuder and Baayen (1997), which had revealed a solid effect of Family Size in the visual modality. This time, using exactly the same materials, no fully reliable effect of Family Size was obtained in the item analysis ($t1(34) = -4.46, p = 0.0001; t2(32) = -1.15, p = 0.2571$). This result supports the hypothesis that in auditory lexical decision the activation of multiple candidates, each activating its own family, masks (at least to some extent) the specific Family Size effect of the target word. In addition, the Family Size effect might be obscured by lexical competition caused by higher-frequency family members in the cohort, slowing down the recognition of the target along the lines suggested by Meunier and Segui (1999).

Second, full-parsing models along the lines proposed by Clahsen (1999) for inflection are likewise challenged by the Family Size effect. If one assumes that regular and predictable complex words are processed by rule, and that only irregular complex words are processed by rote, then the effect of Family Size cannot be explained. If regular complex words do not have their own representations in the central mental lexicon, then there is no information in the lexicon that could give rise to an effect of a type count of the family members of a given stem. The only type count of complex words that is available in full-parsing models is the count of semantically opaque words, which, being unpredictable and irregular, are assumed to be stored. However, semantically opaque words have been found not to contribute to the effect of Family Size.

Third, in their simplest form, full-listing models (e.g., Niemi, Laine, & Tuominen, 1994) have nothing to say about the way in which an effect of Family Size might arise without further assumptions about the way in which full forms are organized at the central, semantic level. In Bybee (1985, 1995)'s full-listing model, complex words that share aspects of form and meaning are linked by form connections and by semantic connections. Such a network architecture, when enriched with the mechanism of spreading activation, might be able to account for effects of Family Size. Because family members share many form connections and many semantic connections with the target word, activation might be taken to spread most strongly to its morphological family. Thanks to larger numbers of co-activated family members, and the concomitant more extensive patterns of activation in the mental lexicon, words with large families would then be responded to more quickly in visual lexical decision than words with small families. It remains unclear, however, how the dissociation between the emergence of a type effect of Family Size and the absence of a token effect of Family Frequency, observed by Schreuder and Baayen (1997) for monomorphemic nouns should be accounted for, given that the connections in Bybee's model encode token frequencies. Distributed connectionist models of morphological processing (e.g., Seidenberg, 1987) are similarly challenged, the more so as they do not embody distinct representations that could underlie a type count effect.

Of the hybrid models in which both rules and direct look-up play a role, the Augmented Addressed Morphology model (Caramazza, Laudanna, & Romani, 1988) is concerned primarily with the processing of inflected words in morphologically rich languages. Since the Family Size effect is derivational in nature, we turn to the parallel dual route architecture worked out by Schreuder and Baayen (1995).

In this approach, modality-specific access representations are connected to more central, modality-free lemma nodes. In turn, a lemma node is connected with many different semantic and syntactic representations. The lemma nodes of words with similar meanings have overlapping sets of semantic representations. The more similar in meaning, the larger the intersect of these sets will be. Figure 2.1 illustrates this architecture for the Dutch words *huis*, ’house’, and *verhuizen*, ’to move house’ and the verbal prefix *ver-*. Transparent family members have substantial parts of their meaning in common, which means that their lemma nodes are all connected through their shared semantic representations. For instance, *huis* and *verhuizen* share many semantic properties pertaining to the concept HOUSE. The Family Size effect can be understood in terms of spreading activation. Upon activation of the lemma node of *huis*, activation spreads to the semantic properties of *huis*, from where it spreads to other lemma nodes such as *verhuizen*. The larger the number of co-activated lemma nodes becomes, the larger the amount of activation in the mental lexicon, and the easier it becomes in visual lexical decision to decide that an existing word is presented.

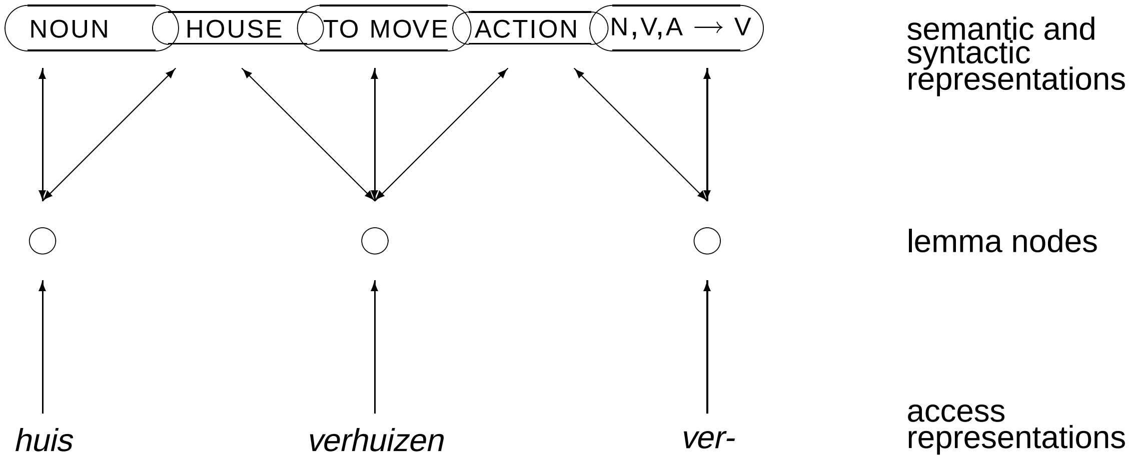

Figure 2.1: Representations for *huis*, ’house’, *verhuizen*, ’move house’, and the prefix *ver-* in a spreading activation model of morphological processing.

This explanation of the Family Size effect is still tentative and in need of further detailing. The aim of the present paper is to contribute to a better understanding of the nature of the Family Size effect by charting in more detail the effects of morphological structure on the activation of the family members. Experiment 2.1 compares the Family Size effect for simplex nouns (without verbal conversion alternants) with the effect for simplex verbs (without nominal conversion alternants), in order to ascertain whether the word category of a simplex word affects the extent to which its family members are activated in the mental lexicon. Experiment 2.2 shifts attention from monomorphemic words to complex words. The aim of this experiment is to ascertain whether response latencies to complex words are co-determined only by Family Size and not by Family Frequency, as has been observed by Schreuder and Baayen (1997) for simplex nouns. Experiment 2.3 addresses the question to what extent the presence or absence of an inflectional suffix affects the activation of the family members. Finally, if the Family Size effect is truly semantic in nature, we would expect that regular as well as irregular participles show an equally strong effect of Family Size, even though the family members of the irregular participles contain a different orthographic and phonological form of the stem than the form that appears in the participle itself. This prediction is tested in Experiment 2.4.

Experiment 2.1: Comparing nouns and verbs

Schreuder and Baayen (1997) call attention to the fact that the morphological families of simplex nouns consist mainly of nominal compounds. Because verbs by themselves do not appear as constituents of noun-noun compounds, their morphological families tend to be much smaller than for nouns. This raises the question to what extent a Family Size effect can be observed for simplex verbs. More specifically, is Family Size a relevant variable for verbs without nominal conversion alternants such as *think*, which exists only as a verb, versus *work*, which exists both as a noun and as a verb? If a Family Size effect for verbs such as *work* is observed, it is unclear to what extent this effect is due to the family of the noun *work*. Experiment 1 of Bertram et al. (2000) gives evidence for a Family Size effect for inflected Dutch verbs with the past tense suffix *-te*. However, in their experiment almost all verbs with a high Family Size happen to have nominal conversion alternants, while those with a low Family Size tend not to have such alternants. This suggests that the observed Family Size effect might in fact be carried by conversion nouns. The aim of the present experiment is to ascertain whether the Family Size effect for verbs without a nominal conversion alternant is comparable to that of nouns without a verbal conversion alternant, or whether the family size effect for verbs observed by Bertram et al. (2000) is in fact due to a nominal family size effect based on their nominal conversion alternants. The results of this experiment will also serve as a baseline for Experiments 2.3 and 2.4 in which we study inflected verb forms.

Method

Participants. 14 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We selected two sets of monomorphemic (and uninflected) words from the CELEX lexical database. The first set consisted of 40 monomorphemic nouns of the type *muur*, 'wall'. These nouns did not have a homographic conversion verb. The second set consisted of 40 monomorphemic verbs of the type *reken*, 'calculate'. These verbs did not have a homographic conversion noun. We partitioned the sets of nouns and verbs into subsets of words with a high versus a low Family Size. Twenty verbs had a low Family Size with on average 3.7 (range 1–7, *SD* 3.3) descendants, contrasting with twenty verbs with a high Family Size of on average 34.7 (range 12–92, *SD* 24.5) descendants. Similarly, twenty nouns had a high Family Size with on average 36.3 (range 10–78, *SD* 19.5) descendants, and the remaining twenty nouns had a low Family Size with on average 3.6 descendants (range 0–7, *SD* 2.2).

We matched the nouns and verbs in the four subsets for mean Base Frequency (nouns with a high Family Size 37.8, range 3.8–122.8, *SD* 33.8; nouns with a low Family Size 38.3, range 3.3–129.2, *SD* 35.9; verbs with a high Family Size 37.6, range 3.1–171.1, *SD* 39.5; verbs with a low Family Size 37.6, range 3.2–183.1, *SD* 43.5, all frequency counts standardized per million). The nouns and the verbs were also matched for word length in letters (5.0, 5.1, 4.6, and 5.0 respectively). We could not match the two subsets of nouns with the two subsets of verbs with respect to Surface Frequency, because monomorphemic (and uninflected) nouns tend to have a substantially higher Surface Frequency than monomorphemic (and uninflected) verbs when they have to be matched simultaneously with respect to Base Frequency. However, the two sets of verbs and similarly the two sets of nouns were matched in the mean for Surface Frequency across the high and low Family Size conditions (verbs, high Family Size 1.6, range 0.1–9.5, *SD* 2.3; verbs low Family Size 1.4, range 0.0–12.0, *SD* 2.6; nouns high Family Size 29.9, range 1.7–94.0, *SD* 26.4; nouns low Family Size 30.1, range 2.9–97.7, *SD* 28.7). The materials are listed in Appendix A. No additional words appeared in the experiment as fillers. Each word was paired with a pseudo word, the phonotactics of which did not violate the phonology of Dutch. Twelve practice trials, 6 words and 6 nonwords, were run before the actual experiment, which was divided in three blocks of roughly 50 items. There was a short pause between the blocks. In total, the experiment lasted approximately 15 minutes.

Procedure. Participants were tested in noise-proof experimental rooms. They were asked to decide as quickly and accurately as possible whether a letter string appearing on the computer screen was a real Dutch word. Each stimulus was preceded by a fixation mark in the middle of the screen for 50 ms. After 500 ms, the stimulus appeared at the same position. Stimuli were presented on Nec Multisync color monitors in white lowercase 36 point Helvetica letters on a dark background and they remained on the screen for 1500 ms. The maximum time span allowed for a response was 2000 ms from stimulus onset.

Results and Discussion

All participants performed the experiment with an overall error rate less than 15%. Table 2.2 shows the mean response latencies (calculated for the correct responses) and error scores (calculated for all responses) for the four experimental conditions.

Table 2.2: Results of Experiment 2.1: Means and standard deviations of response latencies and error proportions (by participants).

		RT	Error	*SD* RT	*SD* Error
Nouns	High Family Size	502	0.02	51	0.03
	Low Family Size	521	0.03	48	0.03
	Difference	-19			
Verbs	High Family Size	527	0.07	51	0.07
	Low Family Size	551	0.07	53	0.07
	Difference	-24			

An analysis of variance revealed main effects of Word Category and Family Size, but no interaction of these two factors ($F1, F2 < 1$). Nouns were responded to more quickly than verbs ($F1(1, 13) = 10.01$, MSE $= 10412.0, p = 0.0075; F2(1, 76) = 9.92$, MSE $= 18457.7, p = 0.0023$), and words with a high Family Size elicited shorter response latencies than words with a low Family Size ($F1(1, 13) = 20.73$, MSE $= 6175.5, p = 0.0005; F2(1, 76) = 4.94$, MSE $= 9196.0, p = 0.0292$). The main effect of Word Category is in line with that observed in Baayen et al. (1997) for noun and verb plurals, although in the present experiment the difference in Surface Frequency between the higher-frequency nouns and the lower-frequency verbs probably plays a more important role. An analysis of variance of the error scores revealed a reliable effect of Word Category only ($F1(1,13) = 13.73$, MSE $= 0.0302, p = 0.0026; F2(1,76) = 8.26$, MSE $= 0.0430, p = 0.0052$). We conclude that the magnitude of the Family Size effect does not differ between nouns and verbs. Apparently, the Family Size effect does not depend on a verb having a nominal conversion alternant: Verbs without a nominal conversion alternant behave similarly as nouns without a verbal conversion alternant with repect to Family Size.

Experiment 2.2: Family size or family frequency for complex words

Although Bertram et al. (2000) report an effect of Family Size for various kinds of complex words, their materials did not control for possible effects of Family Frequency. Before considering the role of the morphological family in detail for two kinds of inflected verb forms, we first address the question whether the Family Frequency of the base of a complex word does not co-determine response latencies, in contrast to the Family Size. In other words, is the role of the morphological family truly an exclusive type-frequency effect, and is the summed token-frequency of the family members really irrelevant? If we succeed in replicating Schreuder and Baayen (1997)'s results for simplex words now using complex words, then this would be problematic for models that take all token frequency effects to arise at central levels of representation (e.g., Zhou & Marslen-Wilson, 2000).

Experiment 2.2a contrasts in a factorial design high and low Family Frequency words matched for Family Size, i.e., the type count of family members. Experiment 2.2b contrasts high and low Family Size words matched for the token count of their family members. What we expect to find is an effect for Family Size only, which would be in line with the results obtained by Schreuder and Baayen (1997) for simplex words.

Experiment 2.2a: Contrasting family frequency

Method

Participants. 20 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We selected three kinds of inflected words for this experiment: inflected verbs in the second and third person singular (e.g., *sloopt*, 'you wreck, he wrecks'), inflected verbs in the past tense singular (e.g., *snapte*, 'understood'), and also adjectives in the comparitive form (e.g., *kalmer*, 'calmer') in order to obtain sufficient experimental words under the severe matching constraints of our experimental design. Of these 70 words, 35 had a high Family Frequency with an average of 30.2 (range 2.8–114.1, *SD* 28.2), and 35 words had a low Family Frequency with an average of 1.5 (range 0.0–5.5, *SD* 1.5). We matched the two subsets of complex words for mean Base Frequency (words with a high Family Frequency: 15.5, range 1.9–62.2, *SD* 13.5; words with a low Family Frequency: 15.6, range 2.2–60.2, *SD* 13.1), mean Surface Frequency (high: 2.1, range 0.1–11.1, *SD* 2.6; low: 2.1, range 0.1–11.5, *SD* 2.6), mean Family Size (high: 6.7, range 2–22, *SD* 4.5; low: 6.1, range 1–22, *SD* 4.4), and for mean length in letters (high 5.7, low: 5.7). We also matched the two subsets of words with respect to the different affixes. The subset with the high Family Frequency consisted of 24 third person singular verbs, 8 past tense verbs, and 3 comparitives. For the subset with the low Family Frequency, these numbers were 22, 9, and 4 respectively. The materials are listed in Appendix B.

As fillers, we added 56 comparatives, so that the experimental list contained the same number of inflected verbs as comparatives. Each word was paired with a pseudo word consisting of a pseudo stem followed by one of the three inflectional affixes *-t, -te*, and *-er* such that the resulting pseudo word did not violate the phonotactic rules of Dutch. The experiment was preceded by 26 practice items. There was a short pause after the practice session, and a short pause halfway through the experimental list. In total, the experiment lasted approximately 20 minutes.

Procedure. The procedure was identical to that of Experiment 2.1.

Results and Discussion

The participants performed the experiment with an error rate less than 15%. Table 2.3 lists mean reaction times (calculated for the correct responses) and mean error scores (calculated over all responses) for the two experimental conditions. The words with a high Family Frequency required slightly longer response latencies than the words with a low Family Frequency, but this difference was not reliable ($t1(19) = 1.82, p = 0.0839; t2(68) < 1$), nor was there any significant difference in the error scores ($t1 = -1.34, p = 0.1956; t2 < 1$). Here and elsewhere where we report t-tests, we use two-tailed tests with $\alpha = 0.05$. This amounts to conservative testing for frequency effects that, when present, are expected to be facilitatory.

Table 2.3: Results of Experiment 2.2a: Means and standard deviations of response latencies and error proportions (by participants).

		RT	Error	*SD* RT	*SD* Error
Complex words	High Family Frequency	615	0.12	85	0.07
	Low Family Frequency	602	0.14	80	0.08
	Difference	+13			

We conclude that Family Frequency does not have any facilitative effect on the response latencies. This result is in line with the results obtained by Schreuder and Baayen (1997) for monomorphemic words. The next experiment shows that an effect of Family Size is observed when Family Frequency is held constant.

Experiment 2.2b: Contrasting family size

Method

Participants. 20 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We selected the same three kinds of complex words: inflected verbs in the third person singular, inflected verbs in the past tense singular and adjectives in the comparative form. We constructed a contrast in Family Size, while matching for Family Frequency. Forty-five words had a high Family Size with an average of 23.4 (range 10–67, *SD* 11.4), and 45 words had a low Family Size with an average of 4.2 (range 1–9, *SD* 1.8) family members. We matched the two sets for mean Base Frequency (high: 37.0, range 1.9–262.9, *SD* 48.5; low: 38.1, range 1.0–403.0, *SD* 66.8), mean Surface Frequency (high: 4.0, range 0.0–40.5, *SD* 7.5; low: 3.9, range 0.1–46.5, *SD* 7.8), mean Family Frequency (high: 15.1, range 2.1–55.3, *SD* 13.6, low: 15.1, range 0.9–66.0, *SD* 16.3) and for mean length in letters (high: 5.7, low: 6.2). The subset of words with a high Family Size consisted of 27 third person singular verbs, 14 past tense verbs and 4 comparatives. The numbers for the subset of the words with a low Family Size were 26, 8, and 11 respectively. The materials are listed in Appendix C.

We added 60 comparatives as fillers, in order to keep the number of comparatives and inflected verbs in the experimental list the same. Each word was paired with a pseudo word with a similar morphological structure. The experiment was preceded by 26 practice items. There was a short pause after the practice session and two short pauses during the actual experiment. In total, the experiment lasted about 30 minutes.

Procedure. The procedure was identical to that of Experiment 2.1.

Results and Discussion

All participants performed the experiment with an error rate less than 8%. Table 2.4 shows the mean reaction times (calculated for the correct responses) and error scores (calculated over all responses). The words with a high Family Size elicited significant shorter response latencies than the words with a low Family Size ($t1(19) = -4.04, p = 0.0007; t2(88) = -2.20, p = 0.0303$). The error scores showed no significant difference between the two experimental conditions ($t1, t2 < 1$).

Table 2.4: Results of Experiment 2.2b: Means and standard deviations of response latencies and error proportions (by participants).

		RT	Error	*SD* RT	*SD* Error
Complex words	High Family Size	563	0.07	64	0.04
	Low Family Size	583	0.08	66	0.05
	Difference	-20			

Considered jointly, Experiments 2.2a and 2.2b show that the effect of morphological descendents on the processing of complex inflected words should be measured in terms of a type count only (Family Size), and not in terms of a token count (Family Frequency).

If it is indeed the case that the Family Size effect arises at semantic levels of representation (for further evidence for this hypothesis, see the discussion of Experiment 2.4 below), then the results of Experiments 2.2a and 2.2b show that token frequency information is not relevant at these central levels. In the model of Schreuder and Baayen (1995), this can be accounted for by restricting token frequency effects to the level of modality-specific access representations only. At the central level, activation spreads to morphologically related lemma representations. As more lemma representations are activated, subjects are able to respond more quickly.

This dissociation of type and token effects is difficult to account for in models in which all frequency effects are claimed to arise in the central lexicon. In the model proposed by Zhou and Marslen-Wilson (2000), for instance, complex words do not have independent representations at the form level. It is only at the semantic level that information about the co-occurence of constituents in complex words is available. However, with only one representational layer for encoding frequencies, it is unclear how the dissociation between type and token frequency that we have observed might be accounted for.

Finally, the results of Experiments 2.2a and 2.2b allow us to investigate effects of Family Size in the experiments following below without having to impose the severe constraint of matching for Family Frequency.

Experiment 2.3: The role of an inflectional suffix

The aim of Experiment 2.3 is to investigate the potential effect of a frequent and productive inflectional suffix, the second and third person singular present tense marker *-t* on the activation of the family members of the base word in the mental lexicon, compared to the activation of the morphological family when only the bare stem is presented. Experiment 2.3a uses a factorial design and parallels the monomorphemic verbs studied in Experiment 2.1, while Experiment 2.3b uses a modified regression design and compares inflected and simplex verbs directly within the same experiment.

Experiment 2.3a: Verbs with an inflectional suffix

Method

Participants. 30 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We selected 48 inflected verbs from the CELEX lexical database in the third person singular (e.g., *leunt*, ’leans’). Twenty-four of these verbs had a high Family Size with an average of 24.3 family members (range 10–113, *SD* 20.2) and 24 had a low Family Size with an average of 3.0 (range 1–8, *SD* 1.8) family members. We matched the two subsets with respect to mean Surface Frequency (high Family Size: 1.0, range 0.0–5.7, *SD* 1.2; low Family Size: 1.0, range 0.0–5.2, *SD* 1.2), mean Base Frequency (high: 16.3, range 2.9–48.5, *SD* 13.0; low: 14.9, range 2.0–50.0, *SD* 10.9), and mean length in letters (high: 5.5, low: 5.5). The materials are listed in Appendix D.

We added 100 fillers of different word types, mostly nouns. Each word was paired with a pseudo word with the same morphological structure which did not violate the phonotactic rules of Dutch. The experiment was preceded by 14 practice trials. There was a short pause after the practice session, and once during the experiment. In total, the experiment lasted about 25 minutes.

Procedure. The procedure was identical to that of Experiment 2.1.

Results and Discussion

All participants performed the experiment with an error rate less than 14%. Two items (one in each subset) had a mean error rate differing more than 3 standard deviations from the mean error rate in their respective conditions. One of these also differed more than 3 standard deviations from the mean reaction time. Both items were excluded from further analyses. This did not affect the matching. Table 2.5 lists the mean reaction times (calculated over the correct responses) and error scores (calculated over all responses) for both experimental conditions.

Table 2.5: Results of Experiment 2.3a: Means and standard deviations of response latencies and error proportions (by participants).

		RT	Error	*SD* RT	*SD* Error
Inflected Verbs	High Family Size	584	0.05	86	0.06
	Low Family Size	604	0.10	76	0.11
	Difference	-20			

As expected, the response latencies for the verbs with a high Family Size were shorter than those for the verbs with a low Family Size. This difference was fully reliable by participants ($t1(29) = -3.3372, p = 0.0023$) and marginally reliable by items ($t2(44) = -1.7609, p = 0.0852$). The verbs with a low Family Size elicited significantly more erroneous responses than the verbs with a high Family Size ($t1(29) = -3.2538, p = 0.0029; t2(44) = -3.1983, p = 0.0026$). Considered jointly, the shorter response latencies and lower error scores for words with a higher Family Size allow us to conclude that the Family Size affects responses in visual lexical decision.

The effect of Family Size is of the same order of magnitude as that observed for the verbs in Experiment 2.1. However, the two experiments should not be compared directly because the average Base Frequency in Experiment 2.3a is twice that of Experiment 2.1. In addition, the contrast in Family Size is slightly larger in Experiment 2.1. We therefore directly compared verbs with and without an inflectional *-t* in Experiment 2.3b using a factorial regression design.

Experiment 2.3b: Comparing verbs with and without an inflectional suffix

Method

Participants. 32 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We selected two sets of monomorphemic verbs from the CELEX lexical database. The first set contained 68 uninflected verbs of the type *kwets*, 'hurt'. None of these verbs had a familiar homographic conversion noun.[2] For the second set we used the third person singular inflection of the same 68 verbs: *kwetst*, 'hurts'.

Thus we created two sets of verbs with equal Base Frequencies (mean 6.9, range 1.7–19.5, *SD* 4.0), equal Family Sizes (mean 7.6, range 0–46, *SD* 9.3), and with somewhat different Surface Frequencies. The inflected verbs were more frequent (mean 0.5, range 0.0–1.0, *SD* 0.3) than the uninflected verbs (mean 0.3, range 0.0–2.0, *SD* 0.3). Due to the inflectional suffix *-t*, the length of the inflected verbs was one character more (mean 5.8) than the length of the uninflected verbs (mean 4.8). In general, Surface Frequency, Base Frequency, and Family Size are strongly mutually correlated. We therefore selected our materials such that the correlations between log Family Size and log Base Frequency, as well as the correlations between log Family Size and log Surface Frequency were absent in the data set. The correlations between Family Size and Base Frequency for both sets of verbs was $r = 0.110(t(66) = 0.90, p = 0.37)$; the correlation between Family Size and Surface Frequency for the subset of uninflected verbs was $r = 0.193(t(66) = 1.60, p = 0.11)$; the correlation between Family Size and Surface Frequency for the subset of inflected verbs was $r = -0.068(t(66) = -0.55, p = 0.58)$. We will call this design, in which we have taken care to remove unwanted collinearity from the data matrix, a factorial regression design, as it allows us to focus specifically on the correlation between Family Size and reaction time. If we find a reliable effect of Family Size,

---

[2] Of the 68 verbs, 18 had a conversion noun, all of which occur with a frequency of less than one per million. Of these 18 conversion nouns, 14 occur with a frequency of less than 0.25 per million.

this effect cannot be contributed to a confound with Surface Frequency or Base Frequency. The materials are listed in Appendix E.

We added 116 fillers to the experimental list, 36 monomorphemic nouns, 40 monomorphemic adjectives and the same 40 adjectives in the comparative form. A participant had to respond either to the inflected form of a verb or to the uninflected form, but never to both. The same holds for the adjectives and their comparative forms. Each word was paired with a pseudo word with the same morphological structure, the phonotactics of which did not violate the phonology of Dutch. Thirty-two practice trials, 16 words and 16 nonwords, were run before the actual experiment, which was divided into three blocks of 98 items. There was a short pause between the blocks. In total, the experiment lasted approximately 30 minutes.

Procedure. The procedure was identical to that of Experiment 2.1.

Results and Discussion

The participants performed the experiment with an overall error rate less than 15%. For each word we calculated mean response latencies (over the correct responses) and error scores (over all responses). The average RT for the uninflected verbs was 709 ms. (error rate 22%), for the inflected verbs, the average RT was 664 ms. (error rate 10%).

For both subsets, we observe a reliable negative correlation between log Family Size and RT (inflected verbs: $r = -0.440, t(66) = -3.98, p = 0.0002$; uninflected verbs: $r = -0.332, t(66) = -2.86, p = 0.0056$). For the error scores, we find a strong trend of the expected negative correlation between log Family Size and error scores (inflected verbs: $r = -0.223, t(66) = -1.86, p = 0.0677$; uninflected verbs: $r = -0.228, t(66) = -1.90, p = 0.0615$). Because Family Size is uncorrelated with both Surface Frequency and Base Frequency in our materials, we can conclude that the observed correlations for both the inflected and the uninflected verbs are due to Family Size only. This implies that we can account for 11% (uninflected verbs) up to 19% (inflected verbs) of the variance in the response latencies purely in terms of the Family Size, a count of morphologically complex word types that are not themselves present in the signal, but that apparently are all stored in the mental lexicon.

Note that the uninflected verbs show longer reaction times ($t2(67) = 4.69, p = 0.000$), higher error scores ($t2(67) = 6.72, p = 0.000$), and a lower correlation of Family Size with reaction time. Apart from the absence of the inflectional *-t*, the uninflected verbs hardly differ from their inflected counterparts. They share the same

Base Frequency and have the same Family Size, whereas it is unlikely that the small difference in Surface Frequency (0.26 per million for the uninflected verbs, and 0.48 per million for the inflected forms) might explain this difference. In fact, we expected to find longer response latencies for the inflected verbs because these very low-frequency forms are likely to require on-line parsing. We will offer a tentative explanation below. First, however, we present a final experiment investigating the effect of Family Size for both regular and irregular participles.

Experiment 2.4: Is the family size effect dependent on form?

The aim of this experiment is to explore the Family Size effect for perfect participles. Experiment 2.4a investigates the influence of the regular circumfix *ge- -d* on the activation of the family members using a factorial regression design, as Experiment 2.3 has shown that this design is somewhat more powerful than a factorial design. Experiment 2.4b investigates whether irregular participles with unpredictable vocalic alternation activate their family members, using a factorial design because there are not enough irregular participles to construct a factorial regression design. A base word such as *roei* ('to row') is phonologically and orthographically completely present in its participle *geroeid*. By contrast, a base word such as *zwem* ('to swim') is not fully retained in its participle *gezwommen*, not phonologically nor orthographically. If the Family Size effect is mediated by the exact form of the base word, then family members such as *zwembad* ('swimming pool') will not be activated by a form such as *gezwommen*. However, if the Family Size effect is a truly central morphological effect sensitive to an abstract stem representation, then we should obtain Family Size effects when counting the family members with the regular stem form for both the regular and the irregular participles.

Experiment 2.4a: Regular past participles

Method

Participants. 41 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We selected 100 regular participles from the CELEX lexical database (e.g., *geroeid*, 'rowed'). These participles had a mean Base Frequency of 6.6 (range 2.1–15.2, *SD* 3.1), a mean Surface Frequency of 0.2 (range 0.0–3.9, *SD* 0.5), a mean Family Size of 5.9 (range 0–28, *SD* 5.4), and a mean length in letters of 8.0. We selected the materials such that the correlation between log Family Size and log Surface Frequency ($r = 0.154; t(98) = 1.54, p = 0.13$), as well as the correlation between log Family Size and log Base Frequency ($r = 0.156; t(98) = 1.56, p = 0.12$) was statistically not reliable. The materials are listed in Appendix F.

As fillers we added 140 inflected verbs in the third person singular and 40 irregular participles. These verbs acted as targets in other experiments, of which the irregular participles will be discussed in Experiment 2.4b. Seventy-five words, mostly nouns, were also added as fillers to the experimental list. For reasons concerning the other experiments not reported in this article, the verbs were divided over two different experimental lists, so that 50 (randomly chosen) participles were responded to by different participants than the remaining 50. Each word was paired with a pseudo word with the same morphological structure, which did not violate the phonotactic rules of Dutch. The experiment was preceded by 22 practice items. There was a short pause after the practice session and two short pauses during the experiment. In total, the experiment lasted approximately 30 minutes.

Procedure. The procedure was identical to that of Experiment 2.1.

Results and Discussion

One participant performed the experiment with an error rate of 38% and was excluded from further analyses. The error rate of the remainig 40 participants was less than 18%. For each item, we calculated the mean response latencies (over correct responses) and the mean error scores (over all responses). The mean reaction time was 741 ms. and the mean error score was 17%. The correlation between reaction times and log Family Size shows the expected negative correlation, which was not fully reliable in a conservative two-tailed test ($r = -0.179; t(98) = -1.80, p = 0.076$). There was no correlation between error scores and Family Size ($r = -0.042; t(98) = -0.412, p = 0.68$).

Given that the effect of Family Size in our experiments is always facilitatory in nature, we conclude that regular participles also activate their family members, albeit somewhat less reliable then we had expected. This suggests that the Family Size effect for irregular participles might also be attenuated.

Experiment 2.4b: Irregular past participles

Method

Participants. 41 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We selected two sets of irregular participles from the CELEX lexical database (e.g. *gezwommen*, ’swum’). All of these participles have a stem allomorph that differs from the stem allomorph of the present tense and infinitive forms (*zwem*). Twenty participles had a high Family Size with an average of 50.1 (range 15–130, *SD* 32.1) and 20 participles had a low Family Size with an average of 7.1 (range 1–16, *SD* 4.2) family members. We matched the two sets for mean Surface Frequency (high: 10.3, range 0.2–40.2, *SD* 11.3; low: 9.5, range 0.1–47.4, *SD* 13.6), mean Base Frequency (high: 66.8, range 5.5–287.8, *SD* 74.0; low: 70.9, range 4.2–485.7, *SD* 115.8), and for mean length in letters (high: 7.95, low: 8.55). The materials are listed in Appendix G.

We added 140 verbs in the third person singular, and 100 regular participles. These verbs acted as targets for other experiments, of wich the regular participles were discussed in the previous experiment. Besides these inflected verbs, we added 75 fillers of different word sorts, mostly nouns. Each word was paired with a pseudo word with the same morphological structure, which did not violate the phonotactic rules of Dutch. For reasons concerning the other experiments not discussed here, the verbs were divided over two experimental lists, so that 20 (randomly chosen) irregular participles were responded to by different participants than the remaining 20. The experiment was preceded by 22 practice items. There was a short pause after the practice session and there were two short pauses during the experiment. In total, the experiment lasted about 30 minutes.

Procedure. The procedure was identical to that of Experiment 2.1.

Results and Discussion

One participant performed the experiment with an error rate of 38% and was excluded from further analyses. The error rate of the remaining 40 participants was less than 18%. The mean reaction time of one item in the high condition (*geworven*, ’recruited’) differed 3.0 standard deviations from the mean reaction time in this condition and was also excluded from further analyses. This did not influence the matching of the two subsets. Table 2.6 shows the mean reaction times (over the correct responses) and error scores (over all responses). The response latencies of the irregular participles with a high Family Size were significantly shorter than those with a low Family Size ($t1(39) = -3.67, p = 0.0007, t2(37) = -2.29, p = 0.028$) and the participles in the high condition elicited significantly less erroneous responses than those in the low condition ($t1(39) = -2.71, p = 0.0099, t2(37) = -2.78, p = 0.0085$).

Table 2.6: Results of Experiment 2.4b: Means and standard deviations of response latencies and error proportions (by participants).

		RT	Error	*SD* RT	*SD* Error
Irregular Participles	High Family Size	641	0.03	115	0.06
	Low Family Size	678	0.08	114	0.09
	Difference	-37			

The regular participles in Experiment 2.4a and the irregular participles in Experiment 2.4b cannot be compared directly. Experiment 2.4a uses a factorial regression design, in which no correlation between Family Size on the one hand and Surface Frequency or Base Frequency on the other hand exists. By contrast, Experiment 2.4b uses a standard orthogonal design with pairwise matching for Surface and Base Frequency between the high and low Family Size conditions. When the items of the two conditions in Experiment 2.4b are pooled, we obtain an item set in which Surface Frequency and Base Frequency correlate both with Family Size and with the response latency. Consequently, a post-hoc correlation of Family Size and response latency for Experiment 2.4b does not measure the effect of Family Size only, invalidating a direct comparison with the correlation obtained in Experiment 2.4a. What we can do, however, is use the linear regression fit to the data of Experiment 2.4a,

$$\text{RT} = 780.50 - 17.78 \log(\text{Family Size} + 1),$$

$F(1, 98) = 3.22, p = 0.076$, which shows that the model fits the data quite well, to calculate the expected difference in reaction time for words with a Family Size of 50.6 compared to words with a Family Size of 7.6, the average Family Size of the orthogonal contrast of Experiment 2.4b. This difference, 32 ms., is of the same order of magnitude as the observed difference in Experiment 2.4b, 37 ms. This suggests that the irregular participles activate their family members to the same extent as the regular participles, even though their base appears in an irregular form that is not shared by most of these family members. In line with the results obtained in Experiment 2.2, which showed that mere string familiarity of the family members does not affect response latencies in visual lexical decision, the present experiment shows that the orthographic form of the base need not be maintained for an effect of Family Size to be obtained.

Further support for the hypothesis that the full family of the abstract form of the base is activated even when the form of the base is not identical to the form in which it appears in most family members, can be obtained by some further correlational analyses. Because Surface Frequency, Base Frequency, and Family Size are all mutually correlated, we first used a stepwise multiple regression analysis as well as a tree-based analysis (Breiman, Friedman, Olshen, & Stone, 1984) to ascertain the relative importance of Base Frequency, Family Size, and Surface Frequency. Both regression techniques pointed out that Base Frequency is not a reliable independent predictor of the response latencies in our data. In order to gauge the correlation of Family Size with reaction time, we need to partial out the correlation of Surface Frequency and reaction time. The partial correlation of Family Size and reaction time, partialling out the contribution of Surface Frequency, is reliable ($r = -0.294, t(36) = -1.85, p = 0.0365$, one-tailed test). Moreover, when we count only the nominal family members, none of which contain the irregular stem form, we also observe a reliable correlation with reaction time after partialling out the correlation of this count with Surface Frequency ($r = -0.28, t(36) = -1.74, p = 0.046$), which shows that indeed family members that do not share the same irregular stem nevertheless crucially contribute to the Family Size effect. Finally, the correlation with reaction time for the counts of those family members that belong to homographs of the irregular verbal stems (e.g., *vocht*, ‘moisture’, in *ge-vocht-en*, ‘fought’, the participle of *vecht*, ‘to fight’) is small and statistically not reliable ($r = 0.02; t < 1$). Apparently, the circumfix *ge-X-en* has prevented such irrelevant false friends of the morphological family to be activated.

Post-hoc analyses

In this section we present two post-hoc analyses that allow us to investigate the Family Size effect in greater detail. Thus far, we have counted family members in a very crude way. Any family member listed in the CELEX lexical database with a frequency greater than one per 42 million was included in the family count.[3]

A first question that we have to address is whether it is realistic to include very low-frequency words in the counts of the Family Size. Including such words implies that we assume that these very low-frequency complex words are stored in the mental lexicon. We therefore calculated the correlation of Family Size with reaction times for a range of frequency thresholds. A frequency threshold of 10 means that a family member should have a frequency of at least 10 per 42 million for it to be included in the count. Figure 2.2 plots the results for a range of thresholds for Experiment 2.4b (upper left panel) and the verbs in Experiment 2.1 (upper right panel). The correlational pattern observed in the left panel is the one that we observe for all other experiments as well. The large dots represent the amount of variance explained by means of Pearson correlations (r^2). What we observe is that removing even the lowest-frequency family members results in a decrease in the amount of variance explained by Family Size.

The only exception in our data is shown in the upper right panel of Figure 2.2. For the verbs in Experiment 2.1 we observe that removing low-frequency family members leads to improved correlations. In order to ascertain that this improvement in the correlation is not an artifact, we ran a randomization test for each threshold. One such randomization test consisted of 1000 permutation runs in which the empirical frequencies of the pooled family members of all our target words were randomly re-assigned to these pooled family members. For each permutation run, a new family count was made in which only those family members were included which had an (artificial) frequency not less than the frequency threshold. These new counts were used to calculate the squared Pearson correlation of reaction times and Family Size. The upper panels of Figure 2.2 show the 95% Monte Carlo confidence intervals of r^2 by means of a vertical solid line. The dots above and below this line denote the 99% Monte Carlo intervals, and the minus signs the corresponding ranges. The upper left panel shows that removing low-frequency family members leads to consistently lower r^2 values than one would expect on the basis

---

[3] The CELEX lexical database does not list words that occur once only in the text corpus on which its frequency counts are based. At the same time, this database does list words occuring in a dictionary of Dutch that do not occur in the corpus. These words are listed as having zero frequency.

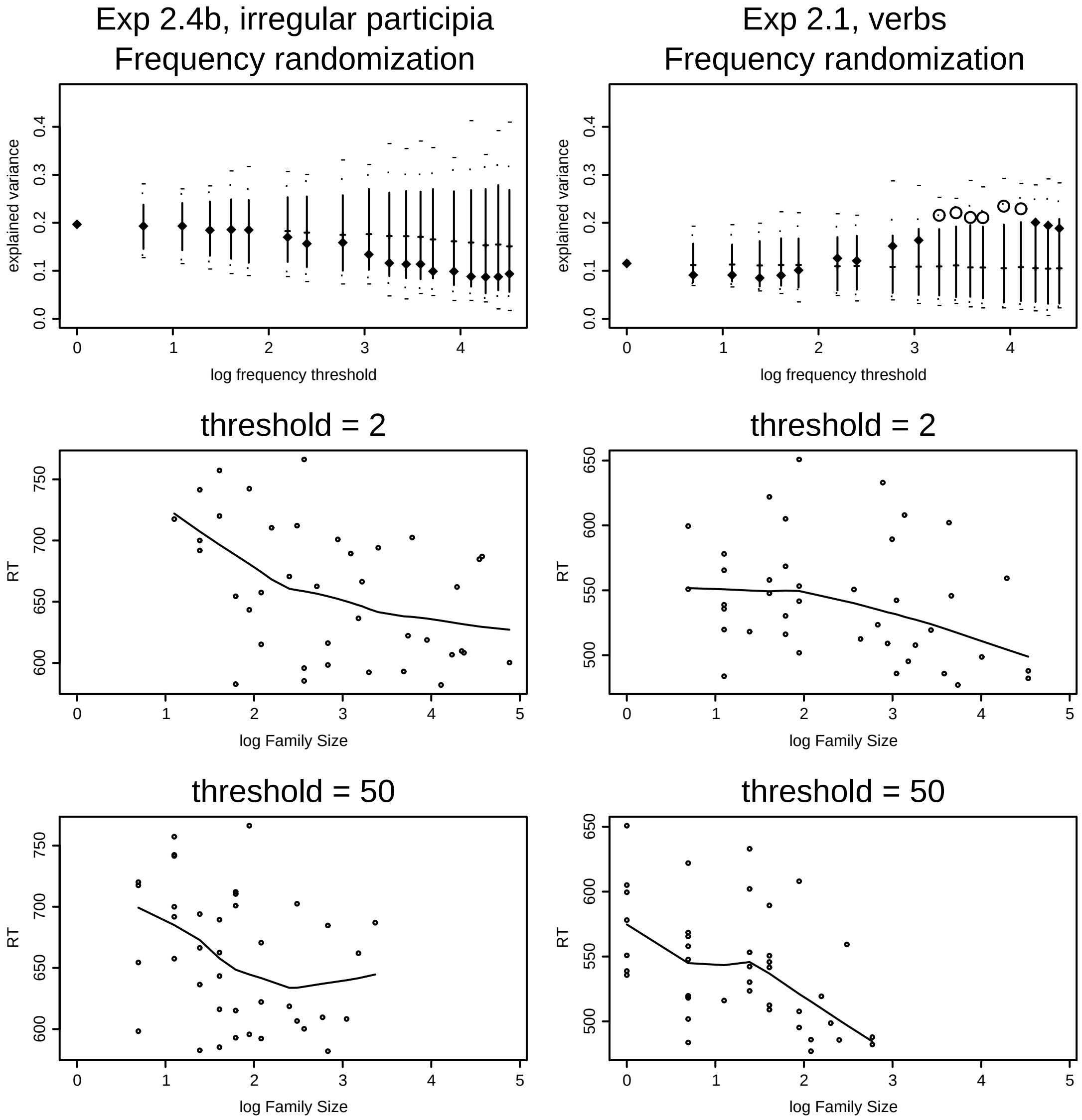

Figure 2.2: Frequency randomization results for Experiment 2.4b (upper left) and Experiment 2.1 (upper right), and scatterplots of log Family Size and RT for frequency thresholds 2 (center panels) and 50 (bottom panels) for Experiment 2.4b (left) and Experiment 2.1 (right). Frequency thresholds per 42 million.

of chance. Turning to the upper right panel, we find that the highest r^2 values are significantly higher than one would expect by chance. These data points, represented by circles, are located in the upper 2.5% of the Monte Carlo distributions.

Table 2.7: Pearson and Spearman correlations for Experiment 2.4b and the verbs in Experiment 2.1 for two different frequency thresholds.

	Experiment 2.4b
threshold 2	$r = -0.440, t(37) = -2.98, p = 0.0051$
	$r_s = -0.428, z = -2.64, p = 0.0084$
threshold 50	$r = -0.314, t(37) = -2.01, p = 0.0514$
	$r_s = -0.331, z = -2.04, p = 0.0411$
	Verbs in Experiment 2.1
threshold 2	$r = -0.302, t(38) = -1.95, p = 0.0583$
	$r_s = -0.332, z = -0.208, p = 0.0379$
threshold 50	$r = -0.484, t(38) = -3.41, p = 0.0015$
	$r_s = -0.483, z = -3.02, p = 0.0025$

We further inspected the data at thresholds 50 per 42 million (3.9 on the log scale) and 2 per 42 million (1.1 on the log scale) to make sure that the conditions for applying the Pearson correlation test are met. The scatterplots shown in Figure 2.2 do not suggest severe violations of homoscedasticity and the non-parametric regression smoothers (Cleveland, 1979) likewise suggest roughly linear trends. Table 2.7 lists the Pearson and Spearman correlation statistics corresponding to these scatterplots. Considered jointly, we may conclude that, apparently, in Experiment 2.1, only the higher-frequency verbs in the family play an effective role. At present we do not understand why this might be so, especially as the uninflected verbs in Experiment 2.3b do not show the same pattern. Possibly, the higher Base and Surface frequencies of the target verbs in Experiment 2.1 are responsible. Further research is clearly required here.

Having ascertained the appropriate frequency thresholds for the family counts of our data sets, we now turn to consider the role of the word category of the family members and the role of an explicit inflectional suffix in some more detail. Recall that in Experiment 2.3b the correlation of Family Size with response latencies turned out to be higher for the verbs with an overt inflectional suffix

($r = -0.440, t(66) = -3.98, p = 0.0002$) than for the same verbs presented in their base form ($r = -0.332, t(66) = -2.86, p = 0.0056$, two-tailed tests). When we consider the correlations of the family counts of the verbal and nominal family members with the response latencies separately, we observe the following pattern. For the target words presented in the base form, without an overt affix that singles them out as verbs, only the nominal family members appear to be activated (Nominal family members: $r = -0.367, t(66) = -3.20, p = 0.0021; r_s = -0.389, z = -3.19, p = 0.0014$. Verbal family members: $r = -0.100, t(66) = -0.81, p = 0.4182$; $r_s = -0.032, z = -0.26, p = 0.7942$, two-tailed tests). However, when the inflectional -t is present both the count of nominal family members and the count of verbal family members show reliable correlations with reaction times (Nominal family members: $r = -0.435, t(66) = -3.93, p = 2e-04; r_s = -0.441, z = -3.61, p = 3e-04$. Verbal family members: $r = -0.283, t(66) = -2.40, p = 0.0193; r_s = -0.196, z = -1.60, p = 0.1085$, two-tailed tests). The fact that the nominal family members always show a reliable correlation may well be due to the larger number of nominal family members (390 nominal versus 151 verbal family members). More interesting is the observation that apparently the presence of an overt verbal suffix is required for the verbal family members of our materials to become activated. Within the framework of our parallel dual route model (Schreuder & Baayen, 1995; Baayen, Schreuder, & Sproat, 1999; Baayen & Schreuder, 1999), we can interpret this finding as follows. Because of their substantially higher frequencies of use, the access representations of affixes will reach threshold activation level long before the base words to which they are attached. After reaching threshold, the corresponding central semantic and syntactic representations are activated. The syntactic representation of the word category VERB is connected to the representation of the suffix *-t* as well as with the representations of all verbs in the lexicon. Once the VERB node is activated, it will activate the verb representations with which it is connected in turn. In the visual lexical decision task, this additional activation of the verbal family members allows participants to respond more quickly.[4]

---

[4] One of our reviewers suggested that the uninflected and inflected verbs might be differentially affected by some other factor. One such factor might be that in visual lexical decision verbs require longer response latencies and elicit more errors than matched nouns (see, e.g., Baayen et al., 1997). Nouns, and adjectives as well, often occur in isolation in natural language, whereas verbs require syntactic context with an overt subject. Without overt verbal marking, verbs presented in isolation are somewhat strange and elicit longer response latencies and more errors, because the default expection of subjects is to encounter nouns or adjectives. Possibly, the presence of an overt inflectional marking on the verb helps to process non-default cases. We might even speculate that the default expectation of encountering nouns or adjectives is in part to be held responsible for the absence of the activation of verbal family members for the uninflected verb forms.

General Discussion

This paper addresses the role of a new factor in visual word recognition, the Family Size effect, for uninflected and inflected words in Dutch. Experiment 2.1 investigated whether the effect of Family Size for verbs depends on the verb having a nominal conversion alternant. Using an orthogonal design, we found a similar effect of Family Size for verbs without a nominal conversion alternant as for nouns without a verbal conversion alternant. This shows that the Family Size effect is not driven by the presence of a noun in the visual input.

Experiment 2.2 proceeded to ascertain whether the effect of the Family Size for complex words is truly a type count effect, and whether the summed token frequencies of the family members, the Family Frequency, do not codetermine the response latencies in visual lexical decision. We first carried out a factorial experiment that contrasted Family Frequency, while matching for Base Frequency, Surface Frequency, and Family Size. No significant difference could be observed in the response latencies. However, an experiment contrasting Family Size while matching for the other three factors revealed a fully reliable difference. This shows that the Family Size effect is not based on string familiarity.

Experiment 2.3 studied the influence of the presence of a verbal inflectional suffix on the Family Size effect. A comparison of verbs in their base form with the same verbs followed by the suffix *-t* revealed a larger effect of Family Size in the presence of the *-t*. A post-hoc analysis indicates that the verbal family members of these verbs are co-activated only when the *-t* is present. Their activation leads to a more substantial overall Family Size effect. This experiment shows that the Family Size effect for complex words has a genuine morphological component independent of a semantic component: only the presence of a by itself meaningless inflectional suffix leads to the activation of the verbs in the morphological families, a set of family members that is defined morpho-syntactically and not semantically. For a similar morphological component to the Family Size effect for a derivational suffix, see Bertram et al. (2000).

Experiment 2.4 studies regular and irregular participles. Both kinds of participles revealed an effect of Family Size. As the irregular participles do not contain the base in the orthographic and phonological form in which it appears in the present tense paradigm and in derived words and compounds, this experiment shows that the effect of Family Size is not mediated by the exact form of the base word, but by a more abstract central morphological representation. This result is comparable to the observation that in repetition priming studies irregular inflected words prime forms with an orthographically different stem (see, e.g., the review in Stolz & Feldman, 1995).

The present Family Size effect is probably related to a type count effect observed by Van Jaarsveld, Coolen, and Schreuder (1994). They observed that novel compounds with constituents that occur in many other compounds are more difficult to reject as existing words than novel compounds with constituents that occur in only a few existing compounds. We understand this result as a Family Size effect: Novel compounds with large morphological families are very word-like and hence difficult to reject as existing words.

It is important to distinguish the facilitatory effect of a large Family Size in visual lexical decision for inflected words from various inhibitory effects that have been interpreted as affecting the early stages of word recognition. For example, Carreiras, Alvarez, and De Vega (1993) report that words with high-frequency syllables are responded to more slowly than words with low-frequency syllables. We understand this effect to arise at the level of access representations. Words with high-frequency syllables activate larger competitor sets, as they occur in more words, resulting in longer response latencies. A related phenomenon is the lexical competition between orthographic neighbors, which are generally defined as words of the same length as a given target word but differing from the target word with respect to exactly one letter (Coltheart, Davelaar, Jonasson, & Besner, 1977). Various studies suggest that words with a large number of neighbors require longer processing times than words with a small number of neighbors (e.g., Goldinger, Luce & Pisoni, 1989; Grainger & Jacobs (1996); Grainger, 1990), although facilitation has also been reported (Andrews, 1989). Grainger and Jacobs (1996) show that these effects can be understood in terms of lexical competition at the access level. Another type count has been studied by Sánchez-Casas, García-Albea, and Bradley (1991) and Sánchez-Casas (1996). These authors report, for instance, that highly restrictive strings, i.e., strings that occur in relatively few word types, are more effective primes than non-restrictive strings. Again, this effect reflects competition at the early stages of visual word identification. In contrast to all these early effects, the effect of Family Size studied in the present paper is a central effect. For instance, Bertram et al. (2000) show that the Family Size effect crucially depends on the semantically transparent family members. Experiment 2.3b of the present paper shows, furthermore, that the effect is mediated by morphology: The verbs in the morphological family are activated only in the presence of the inflectional suffix *-t*. Finally, Experiment 2.4b shows that the Family Size effect is independent of the orthographic and phonological form of the base word. Also note that the Family Size effect is facilitatory in nature, to be distinguished from various inhibitory effects arising from lexical competition at the access level (see also Meunier & Segui, 1999, for lexical competition in auditory processing).

The Family Size effect in the present paper should probably also be distinguished from a family effect reported for Hebrew by Feldman, Frost, and Pnini (1995). Using the segment shifting task, they found that it is easier to shift a word pattern to a nonword consonantal root when the stimulus contains a root that occurs in many different words than when it occurs in only one word. A direct comparison with the Family Size effect discussed in the present paper is difficult to make because it is unclear whether the effect in Hebrew is a token driven effect or a type driven effect. In the absence of token frequency counts for Hebrew roots and words, it is impossible to disentangle the relative contributions of token frequencies on the one hand, and type counts on the other hand. In what folllows we will sketch a tentative interpretation of the Hebrew data in relation to the data from Dutch.

Let’s assume that the segment-shifting task as used by Feldman et al. (1995) taps into the segmentation process at the access level. In our model, the access level is the level at which token frequencies are coded. This leads us to suspect that the effect observed for Hebrew might well be a token frequency effect and not a type frequency effect. This hypothesis is supported by the observation that in Dutch the Family Size effect crucially hinges on the semantic transparancy of the family members, while in Hebrew the semantic relation between derivations sharing a given root appears to be irrelevant, as shown by Frost, Forster, and Deutsch (1997) in a priming study. The effect in Hebrew appears to be a genuine morphological form-effect, evidence for the claim advanced by Aronoff (1994) that there are morphological regularities at the form-level that operate independently of semantics.

Within the framework proposed by Schreuder and Baayen (1995), we can understand the Hebrew data along the lines shown in Figure 2.3, using as example the noun *mrgl*, ’spy’. At the access level, we have three representations, the full form *mrgl*, the root *-r-g-l-*, ’foot’, and the participial prefix *m-*. The root *-r-g-l-* is connected with many different lemma nodes, including the noun *trgl*, ’exercise’, and the noun *mrgl*. The full form of *mrgl* also points to the lemma node of *mrgl*. The access representation of the participial prefix likewise points to its own lemma node. Note that the root does not point to a unique lemma node: The root representation is a form representation only, without its own semantics. In the original model of Schreuder and Baayen (1995), access representations for morphemes are always linked up to their own lemma representations. The Hebrew data show that this coupling is too restrictive. Morphemes that have no independent meanings should not be linked up to independent lemma nodes, but to the lemma nodes of the words in which they occur.

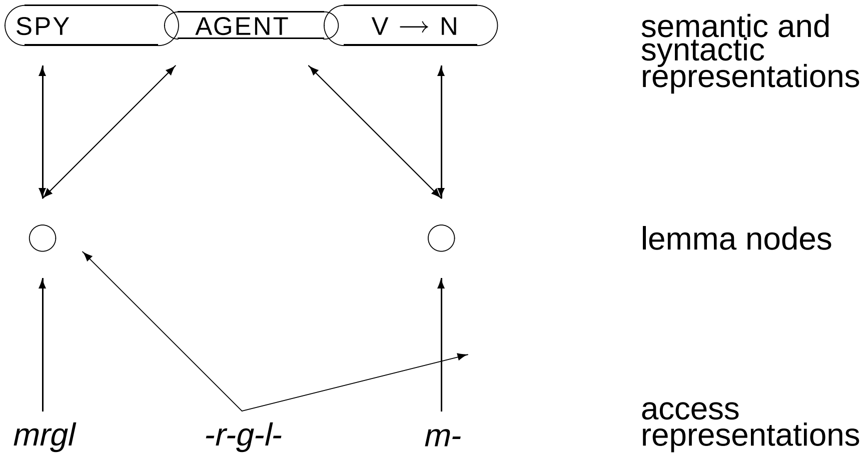

Figure 2.3: Representations for *mgrl*, ‘spy’, the root *rgl*, and the prefix *m-*, in a spreading activation model of morphological processing.

The lemma node of *mrgl* is connected with the various semantic and syntactic representations that come with the noun SPY. Similarly, the lemma node of the prefix *m-* points to the semantic and syntactic representations that come with agentive participles. Note that the semantic features representing ACTION are shared by the lemma’s of *mrgl* and *m-*. The resting activation levels of the access representations are determined by the token frequencies with which they are activated by the visual input. Roots with large families will have high resting activation levels. Hence, in the segment shifting task, such root forms are more easily detected leading to faster segment shifting. Similarly, in priming tasks, the root has been pre-activated and will therefore facilitate the activation of the target lemma node. If this interpretation is correct, the Hebrew data evidence a morphological Family Frequency effect at the form level, whereas the Dutch data discussed in the present paper evidence a morphological Family Size effect at the semantic level.

This explanation raises two questions. First, how specific to Hebrew is the architecture of Figure 2.3? Second, how specific to Dutch is the Family Size effect?

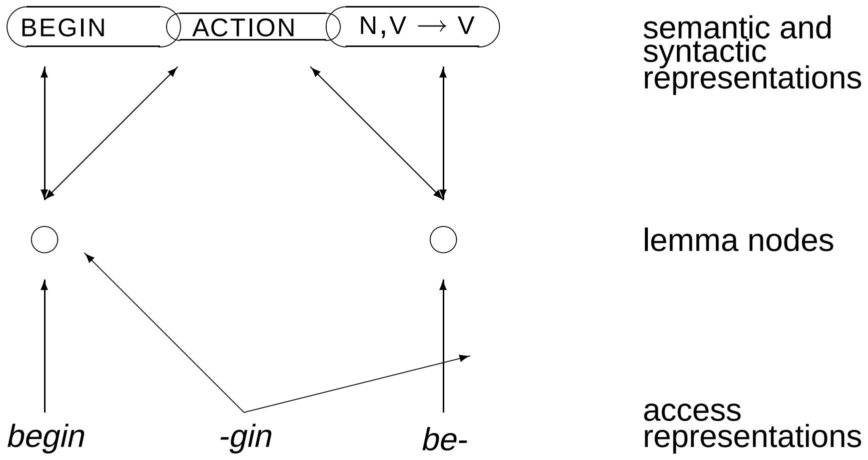

Figure 2.4: Representations for *begin*, 'begin', the meaningless stem *-gin*, and the prefix *be-*, in a spreading activation model of morphological processing.

With respect to the first question, we note that a similar architecture is independently motivated for Dutch. Figure 2.4 illustrates this for the Dutch verb *begin*, 'begin', which contains the bound stem *-gin* and the prefix *be-*. The bound stem *-gin* has no clear meaning of its own, but it also occurs in the verb *ontgin* 'develop, cultivate, exploit'. Interestingly, in spite of a lack of semantic compositionality, the verbs *begin* and *ontgin* are morphologically complex at the form level. In Dutch, past participles normally have the prefix *ge-*, except when another prefix, such as *be-* and *ont-*, is present. The participles of *begin* and *ontgin* are not *ge-begonnen* and *ge-ontgonnen*, as one would expect if these verbs were monomorphemic, but *begonnen* and *ontgonnen*. Thus, *-gin* in Dutch is the (exceptional) concatenative parallel of Hebrew roots such as *-r-g-l-*.

Next consider the question how language-specific the Family Size effect discussed in the present paper is. We suspect that the key to this question is the semantic consistency of the morphological families. In Dutch, the majority of family members of a given stem are semantically transparent. As we have seen, the family of a given root in Hebrew often contains words with unpredictable opaque meanings, words that do not stand in a transparent relation to their root. For instance, *mrgl*, 'spy', stands in no obvious semantic relation to *trgl*, 'exercise', or the noun *rgl*, 'foot'. Possibly, there are small semantically consistent subfamilies in Hebrew for which a Family Size effect might be obtained. Interestingly, the problem of semantic consistency may also arise in Finnish, a language with a very rich morphological system. Consider the stem *kirja*, ’book’, which has roughly 1100 family members. However, these family members belong to a wide range of semantic domains, as illustrated by the following examples, all of which require a translation equivalent with a different English stem: *kirjaaminen*, ’registration’; *kirjailija*, ’author’; *kirjaimellinen*, ’litterally’; *kirjaimisto*, ’alphabet’; *kirjain*, ’character’; *kirjaltaja*, ’typographer’; *kirjasto*, ’library’. With such semantic diversity within one family, we would not be surprised to find that the raw family count of *kirja* is not a reliable predictor of response latencies in visual lexical decision. As we hypothesized for Hebrew, however, such an effect might perhaps be obtained for semantically consistent subfamilies.

Experiment 2.4b invites another comparison between Hebrew and Dutch. In the process of reading the past participle *gevochten*, the morphological family of the base verb *vecht* influenced the response latencies. The Family Size of the unrelated embedded noun *vocht* was found to be irrelevant. Apparently, the presence of the circumfix *ge- -en* was sufficient to activate only the relevant meaning of *gevochten*. In Hebrew, the morphological context in which a root appears might similarly condition the activation of the correct meaning. For instance, in the morphological context *m-* the meaning SPY is activated, while in the context *t-* the meaning EXERCISE is activated.

In the introduction, we have pointed out a number of implications of the Family Size effect for current theories of morphological processing. The results of the present paper have additional theoretical consequences.

First, models that assume maximal decomposition at the identification stages of word recognition and that posit all knowledge of morphemic combinations to be stored at the central, semantic level of representation (e.g., Zhou & Marslen-Wilson, 2000), are severely challenged by the results of Experiment 2.2. This experiment showed for complex words that the token frequencies of the family members do not influence visual lexical decision latencies. This result suggests that token frequencies are not relevant at the level of semantic representations, in line with the conclusions of Schreuder and Baayen (1997) for simplex words.

Second, consider the results of Experiment 2.3b, which showed that the presence of the inflectional suffix *-t* leads to a larger Family Size effect for verbs. We interpret the effect of the *-t* as a result of this suffix being detected by the parsing route. Thanks to the parsing route, there is more evidence that the input is a verb than can be provided by the direct route in isolation, leading to a larger Family Size effect for verbs. Interestingly, Schreuder, De Jong, Krott, and Baayen (1999) report solid effects of Surface Frequency for verbs with the suffix *-t*. Although many regular inflected forms may have full form access representations, Experiment 2.3b shows that parsing can simultaneously play a role, suggesting that both full forms and morphemes are present in the mental lexicon. The balance of storage and computation is not an either-or phenomenon.

Third, Experiment 2.4b, which showed that only the genuine families of the past participles influenced response latencies, has further implications. In an affix-stripping model (Taft & Forster, 1975), the stripping of the affixes of *ge-vocht-en* would lead to a serial search in a bin containing words with the stem *vocht*, both words with the noun *vocht* and forms of the verb *vecht* with the allomorph *vocht*. The serial search mechanism predicts that both kinds of *vocht* are treated identically, whereas our experiment shows that in the presence of the circumfix *ge- -en* the family members of the noun *vocht* are not co-activated. The implementation of this morphological context sensitivity of the Family Size effect poses an interesting challenge for the future development of distributed connectionist models as well. In our model, the circumfix *ge- -en* activates the syntactic representation VERB, which we take to have an inhibitory connection with the syntactic representation NOUN. Consequently, the NOUN representation that is crucial for mediating the flow of activation from the noun lemma *vocht* to its family members is inhibited, effectively blocking activation from spreading to the family members of this noun.

Finally, in the study addressing the Family Size effect for monomorphemic words, Schreuder and Baayen (1997) propose to understand the Family Size effect as resulting from semantic activation spreading from the monomorphemic word to its family members. Their study does not allow us to rule out that this effect might be a general semantic effect, as semantically transparent morphologically related words are strongly semantically related. The present study makes clear that the immediate morphological context in which a monomorphemic verb appears mediates the Family Size effect. The context of the inflectional suffix *-t* or the circumfix *ge- -en* clearly influences the activation of family members. We therefore conclude that the Family Size effect is a semantic effect with a genuine morpho-syntactic component.

References

Andrews, S.: 1989, Frequency and neighborhood size effects on lexical access: Activation or search?, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**, 802–814.

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

Baayen, R. H. and Schreuder, R.: 1999, War and peace: morphemes and full forms in a non-interactive activation parallel dual route model, *Brain and Language* **68**, 27–32.

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., Lieber, R. and Schreuder, R.: 1997, The morphological complexity of simplex nouns, *Linguistics* **35**, 861–877.

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. and Sproat, R.: 2000, Morphology in the mental lexicon: a computational model for visual word recognition, *in* F. van Eynde and D. Gibbon (eds), *Lexicon Development for Speech and Language Processing*, Kluwer Academic Publishers, pp. 267–291.

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

Breiman, L., Friedman, J. H., Olshen, R. and Stone, C. J.: 1984, *Classification and Regression Trees*, Wadsworth International Group, Belmont, California.

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

Bybee, J. L.: 1995, Diachronic and typological properties of morphology and their implications for representation, *in* L. B. Feldman (ed.), *Morphological Aspects of Language Processing*, Lawrence Erlbaum Associates, Hillsdale, N. J., pp. 225–246.

Caramazza, A., Laudanna, A. and Romani, C.: 1988, Lexical access and inflectional morphology, *Cognition* **28**, 297–332.

Carreiras, M., Alvarez, C. J. and de Vega, M.: 1993, Syllable frequency and visual word recognition in Spanish, *Journal of Memory and Language* **32**(6), 766–780.

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

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

Colé, P., Beauvillain, C. and Segui, J.: 1989, On the representation and processing of prefixed and suffixed derived words: A differential frequency effect, *Journal of Memory and Language* **28**, 1–13.

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.

Feldman, L. B., Pnini, T. and Frost, R.: 1995, Decomposing words into their constituent morphemes: Evidence from English and Hebrew, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 947–960.

Frost, R., Forster, K. I. and Deutsch, A.: 1997, What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **23**, 829–856.

Goldinger, S. D., Luce, P. A. and Pisoni, D. B.: 1989, Priming lexical neighbors of spoken words: Effects of competition and inhibition, *Journal of Memory and Language* **28**, 501–518.

Grainger, J.: 1990, Word frequency and neighborhood frequency effects in lexical decision and naming, *Journal of Memory and Language* **29**, 228–244.

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

Grainger, J. and Segui, J.: 1990, Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies, *Perception and psychophysics* **47**, 191–198.

Marslen-Wilson, W., Zhou, X. and Ford, M.: 1997, Morphology, modality, and lexical architecture, *in* G. E. Booij and J. Van Marle (eds), *Yearbook of Morphology 1996*, Kluwer, Dordrecht, pp. 117–134.

Meunier, F. and Segui, J.: 1999, Frequency effects in auditory word recognition: The case of suffixed words, *Journal of Memory and Language* **41**, 327–344.

Niemi, J., Laine, M. and Tuominen, J.: 1994, Cognitive morphology in Finnish: foundations of a new model, *Language and Cognitive Processes* **9**, 423–446.

Paap, K. R. and Johansen, L. S.: 1994, The case of the vanishing frequency effect: A retest of the verification model, *Journal of Experimental Psychology: Human Perception and Performance* **20 (6)**, 1129–1157.

Sanchez-Casas, R. M.: 1996, Lexical access in visual word recognition: The contribution of word form, *in* M. Carreiras, J. E. García-Albea and N. Sebastián-Gallés (eds), *Language processing in Spanish*, Erlbaum, Mahwah, NJ, pp. 21–59.

Sanchéz-Casas, R. M., García-Albea, J. E. and Bradley, D. C.: 1991, On access representation in visual word recognition: The temporal separation technique, *Psychological Research* **53**(1), 53–61.

Schreuder, R. and Baayen, R. H.: 1995, Modeling morphological processing, *in* L. B. Feldman (ed.), *Morphological Aspects of Language Processing*, Lawrence Erlbaum, Hillsdale, New Jersey, pp. 131–154.

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.

Seidenberg, M.: 1987, Sublexical structures in visual word recognition: Access units or orthographic redundancy, *in* M. Coltheart (ed.), *Attention and Performance XII*, Lawrence Erlbaum Associates, Hove, pp. 245–263.

Stolz, J. A. and Feldman, L. B.: 1995, The role of orthograhic and semantic transparency of the base morpheme in morphological processing, *in* L. B. Feldman (ed.), *Morphological aspects of language processing*, Erlbaum, Hillsdale, NJ, pp. 109–129.

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

Taft, M. and Forster, K. I.: 1975, Lexical storage and retrieval of prefixed words, *Journal of Verbal Learning and Verbal Behavior* **14**, 638–647.

Van Jaarsveld, H. J., Coolen, R. and Schreuder, R.: 1994, The role of analogy in the interpretation of novel compounds, *Journal of Psycholinguistic Research* **23**, 111–137.

Zhou, X. and Marslen-Wilson, W.: 2000, Lexical representation of compound words: cross-linguistic evidence, *Psychologia* **43**, 47–66.

Zwitserlood, P.: 1989, The locus of the effects of sentential-semantic context in spoken-word processing, *Cognition* **32**, 25–64.

Appendix A, target words with reaction time.

Experiment 2.1, nouns with a high Family Size:
mode (fashion) 483, heil (welfare) 541, plicht (duty) 502, koning (king) 486, spion (spy) 489, berk (birch) 543, theorie (theory) 512, schema (scheme) 467, broek (trousers) 487, park (park) 466, muts (hat) 498, alarm (alarm) 456, plein (square) 502, klimaat (climate) 507, ketel (kettle) 554, kantoor (office) 479, bord (plate) 499, band (band) 557, vee (cattle) 512, rente (interest) 510.
Experiment 2.1, nouns with a low Family Size:
veranda (porch) 578, kerel (chap) 465, sofa (sofa) 506, maizena (corn flour) 675, broer (brother) 456, tante (aunt) 493, gazon (lawn) 535, dal (valley) 506, villa (villa) 488, term (term) 491, kolonel (kolonel) 534, humor (humour) 500, ellende (misery) 476, vacht (fur) 518, lies (groin) 621, spul (stuff) 543, neef (nephew) 522, reeks (series) 497, prooi (prey) 500, atlas (atlas) 532.
Experiment 2.1, verbs with a high Family Size:
vorder (progress) 608, stook (stoke) 589, schaam (feel ashamed) 513, meng (mix) 495, bind (tie) 559, weef (weave) 602, win (win) 482, reken (calculate) 488, zwem (swim) 546, martel (torture) 524, klaag (complain) 486, giet (pour) 508, zuig (suck) 477, woel (toss) 633, stuif (blow) 542, metsel (build with bricks) 551, lijd (suffer) 519, meld (report) 509, jaag (hunt) 499, laad (load) 486.
Experiment 2.1, verbs with a low Family Size:
weifel (waver) 651, dwing (force) 530, tuur (peer) 551, streel (carress) 558, hunker (yearn) 566, dender (rumble) 600, beef (tremble) 542, sis (hiss) 548, koester (cherish) 578, mompel (mumble) 520, raas (rage) 569, kreun (moan) 518, hijg (pant) 484, kneed (knead) 536, bied (offer) 516, tracht (endeavour) 622, daag (dawn) 553, wuif (wave) 605, pieker (cogitate) 539, knaag (knaw) 502.

Appendix B, target words with reaction time.

Experiment 2.2a, complex words with a high Family Frequency:
tooit (adorns) 691, snapte (understood) 551, raker (more home) 615, neigt (inclines) 602, kalmer (calmer) 557, schaadt (damages) 626, negeert (ignores) 589, hult (wraps) 720, huivert (shivers) 554, deinst (winces) 649, botste (bumped) 612, waait (blows) 544, siste (hissed) 628, roemt (praises) 649, duizelt (spins) 573, woester (more savage) 580, ruist (rustles) 599, zoemt (buzzes) 636, hindert (impedes) 554, drenkte (soaked) 679, wenkte (beckoned) 589, ronkte (snored) 718,

roert (stirs) 593, mijdt (avoids) 628, kaatst (bounces) 659, poogt (endeavours) 631, wendt (turns) 621, strikte (tied) 583, daagt (dawns) 713, ploft (thuds) 534, mankt (limps) 633, kwetste (hurt) 571, zoogt (breastfeeds) 650, toeft (stays) 767, gruwt (abhors) 587.

Experiment 2.2a, complex words with a low Family Frequency:

snurkte (snored) 612, slaakte (heaved) 576, faalt (fails) 620, boft (flukes) 558, schopt (kicks) 582, gaapt (yawns) 522, briest (roars) 697, zotter (sillier) 656, vlotte (proceeded) 655, tuurt (peers) 641, sabbelt (sucks) 669, kloeker (stouter) 798, streelt (carresses) 537, bulkt (teems) 676, plast (pees) 581, zwaait (swings) 522, valer (paler) 684, sloft (shuffles) 598, lijmt (glues) 590, kwakt (bumps) 729, bukte (ducked) 612, zoent (kisses) 578, smakte (smacked) 607, sloopt (wrecks) 544, plukte (plucked) 549, hapt (bites) 548, grifte (engraved) 990, dempte (filled) 659, zweeft (floats) 525, koert (coos) 809, hinkte (limped) 579, borrelt (bubbles) 573, enger (creapier) 540, stinkt (stinks) 528, kraait (crows) 561.

Appendix C, target words with reaction time.

Experiment 2.2b, complex words with a high Family Size:

zoeter (sweeter) 504, likt (licks) 587, klapte (clapped) 516, stuift (blows) 630, bokste (boxed) 573, slijpt (grinds) 554, rouwt (mourns) 578, cirkelt (cirkles) 570, smaakt (tastes) 511, wilder (wilder) 510, vetter (fatter) 515, scheert (shaves) 558, trilt (vibrates) 528, scherper (sharper) 571, kuste (kissed) 530, kalkte (plastered) 630, stopte (stopped) 588, seint (signals) 653, beukt (batters) 557, baast (bosses) 541, prikte (pricked) 539, oogstte (harvested) 641, damt (plays checkers) 594, trapt (steps) 510, sleept (drags) 564, rolt (rolls) 505, plakte (sticked) 513, hakte (chopped) 576, schopte (kicked) 563, rekt (stretches) 569, pompte (pumped) 645, danst (dances) 494, woelt (tosses) 675, spint (spins) 591, schaamt (feels ashamed) 539, poedert (powders) 609, lakte (polished) 671, glijdt (slides) 538, siert (adorns) 533, boort (drills) 576, stinkt (stinks) 501, spookte (haunted) 529, schaakte (played chess) 634, rijmt (rimes) 495, kamt (combes) 549.

Experiment 2.2b, complex words with a low Family Size:

wreder (crueller) 566, ruiste (rustled) 639, neigt (inclines) 573, kalmer (calmer) 520, dwingt (forces) 538, schaadt (damages) 659, negeert (ignores) 555, huivert (shivers) 606, deert (harms) 693, botste (bumped) 544, jankt (whines) 495, brult (roars) 560, siste (hissed) 548, juister (juster) 551, walgt (dispises) 538, deint (heaves) 665, aarzelt (hesitates) 538, woester (more savage) 526, juichte (cheered) 554, deugt (is good) 527, vromer (piouser) 688, vrolijker (happier) 482, trachtte (endeavours) 577, kaatste (bounced) 644, hapert (gets stuck) 629, zoemt (buzzes) 656, blufte (bluffed) 581, triester (sadder) 556, katholieker (more catholic) 685, zwijgt (is silent) 504, weigert (refuses) 529, rinkelt (jingles) 560, biedt (offers) 554, slapper (slacker) 548, mankt (limps) 783, hurkte (squatted) 638, gluurt (peeks) 536, druist (roars) 648, brouwt (brews) 653, toeft (stays) 727, soepeler (more supple) 584, laffer (more cowardly) 674, kreunt (moans) 549, gruwt (abhors) 631, beeft (trembles) 574.

Appendix D, target words with reaction time.

Experiment 2.3a, inflected verbs with a high Family Size:
spoedt (urges) 578, raapt (gathers) 590, kapt (does one's hair) 642, braakt (barves) 579, smeert (smears) 548, haakt (crochets) 586, ijvert (devotes) 551, stroopt (poaches) 570, spitst (pricks) 641, bokst (boxes) 559, seint (signals) 584, scheurt (tears) 582, naait (sews) 508, knoopt (ties) 529, duikt (dives) 562, waant (imagines) 594, veert (is springy) 656, pompt (pumps) 558, tuigt (harnesses) 656, boort (drills) 606, woekert (grows rank) 664, slijmt (lays it on) 566, rijmt (rimes) 548.
Experiment 2.3a, inflected verbs with a low Family Size:
wreekt (avenges) 601, krenkt (offends) 579, juicht (cheers) 536, dempt (fills) 588, rept (mentions) 646, knielt (kneels) 601, hurkt (squats) 657, smoort (suffocates) 625, leunt (leans) 556, fronst (frowns) 634, sist (hisses) 618, glooit (slopes) 621, zwiept (bounces) 656, mikt (aimes) 576, ketst (glances off) 712, schrapt (scrapes) 635, krijst (shrieks) 566, bukt (ducks) 599, tergt (provokes) 615, scheelt (is the matter) 551, loeit (moos) 575, kneedt (kneads) 622, snikt (gasps) 582.

Appendix E, target words with reaction times for the uninflected and inflected variant, as well as Family Size.

Experiment 2.3b, uninflected (and inflected) verbs:
baad(t) (bathe) 871 756 4, blus(t) (extinguish) 632 659 9, broei(t) (heat) 632 592 9, brouw(t) (brew) 932 633 7, bruis(t) (foam) 648 608 4, buitel(t) (tumble) 722 666 3, bulder(t) (roar) 667 725 3, dein(t) (heave) 865 668 2, demp(t) (fill) 759 615 6, dommel(t) (doze) 742 632 4, dool(t) (wander) 937 663 7, dweep(t) (idolize) 950 728 7, folter(t) (torture) 748 758 6, hakkel(t) (stammer) 683 680 3, huldig(t) (honour) 696 686 2, hunker(t) (yearn) 641 646 2, huw(t) (marry) 738 778 2, ijk(t) (calibrate) 872

727 5, jank(t) (whine) 653 550 3, kantel(t) (cant) 609 596 3, kneed(t) (knead) 745 676 2, knoei(t) (make a mess) 598 578 9, knok(t) (fight) 635 645 4, kwets(t) (hurt) 607 589 7, kwijn(t) (languish) 810 726 3, laad(t) (load) 583 626 35, maai(t) (mow) 631 577 8, martel(t) (torture) 612 613 16, mijmer(t) (muse) 698 709 3, mors(t) (spill) 633 681 5, neurie(t) (hum) 745 783 1, orden(t) (arrange) 741 723 26, poch(t) (boast) 735 768 3, pronk(t) (flaunt) 625 624 10, pruil(t) (pout) 891 781 3, rijg(t) (thread) 749 655 7, ritsel(t) (rustle) 733 654 3, rooi(t) (dig up) 769 643 6, schrap(t) (scrape) 678 635 4, schrob(t) (scrub) 695 678 3, sidder(t) (shiver) 705 722 3, sjouw(t) (lug) 670 623 9, slaak(t) (heave) 942 774 1, slijp(t) (grind) 657 570 15, snoei(t) (prune) 585 615 7, speur(t) (investigate) 658 658 13, spied(t) (spy) 773 656 5, spuw(t) (spew) 762 586 6, stamp(t) (stamp) 635 601 7, stoei(t) (play about) 544 631 3, sus(t) (soothe) 740 740 0, taxeer(t) (evaluate) 654 666 4, tier(t) (rage) 688 717 3, tintel(t) (tingle) 664 606 3, tob(t) (worry) 661 656 5, tors(t) (haul) 730 773 2, tover(t) (work magic) 626 561 46, train(t) (train) 540 603 43, treur(t) (grieve) 678 636 7, tuimel(t) (tumble) 714 625 4, waad(t) (wade) 867 774 3, walg(t) (dispise) 651 599 4, weef(t) (weave) 631 678 37, ween(t) (cry) 691 637 2, weifel(t) (waver) 658 671 6, woel(t) (toss) 692 656 17, wrik(t) (lever) 860 746 6, wurg(t) (strangle) 612 628 6.

Appendix F, target words with reaction time and Family Size.

Regular participles:
geaaid (stroked) 717 2, gebaald (been fed up) 686 5, gebezemd (broomed) 829 5, gebibberd (shivered) 783 3, gebroeid (heated) 738 9, gebulderd (roared) 777 3, gebungeld (dangled) 851 0, gedamd (played checkers) 738 17, gedaverd (boomed) 963 2, gedeerd (harmed) 829 3, gedeugd (been good) 767 2, gedoold (wandered) 826 7, gedraafd (trotted) 766 12, gedraald (lingered) 646 0, gedweild (mopped) 630 4, gefonkeld (sparkled) 708 3, gegalmd (sounded) 725 8, gegeeuwd (yawned) 722 4, gegluurd (peeked) 751 3, gegonsd (buzzed) 794 1, gegraaid (grabbled) 777 3, gegraasd (grazed) 782 1, gegruweld (been horrified) 718 2, gehageld (hailed) 789 12, gehaperd (got stuck) 720 1, gehengeld (angled) 693 6, gehobbeld (bumped) 767 4, gehunkerd (yearned) 710 2, gehuppeld (skipped) 638 3, gejoeld (whooped) 851 3, gejubeld (jubilated) 750 9, gekakeld (cackled) 757 4, gekegeld (played skittles) 776 10, gekerfd (carved) 786 6, gekermd (moaned) 851 1, gekeurd (judged) 730 28, gekleefd (stuck) 685 11, gekneld (pinched) 810 8, geknoeid (made a mess) 641 9, gekrioeld (swarmed) 863 1, gekwijld (druled) 764 2, gelasterd (insulted) 812 12, gelummeld (hanged around) 916 5, gemijmerd (mused) 733 3, gemop-

perd (grumbled) 626 3, gemord (muttered) 873 1, gemurmeld (mumbled) 768 1, geneuried (hummed) 808 1, geniesd (sneezed) 783 5, gepareld (pearled) 868 15, gepeddeld (peddled) 814 1, gepiekerd (cogitated) 632 2, geplonsd (splashed) 762 3, gepluisd (given off fluff) 662 8, gepokerd (played poker) 776 4, gepraald (flaunted) 780 8, gepriemd (pierced) 694 2, gepuzzeld (puzzled) 591 5, geranseld (flogged) 706 5, gerijmd (rimed) 589 14, gerild (shivered) 785 3, geritseld (rustled) 724 3, geroeid (rowed) 746 13, geroffeld (ruffled) 777 3, gerouwd (mourned) 769 21, gesabeld (sabred) 833 5, geschuimd (foamed) 757 22, geseind (signaled) 729 19, gesidderd (shuddered) 680 3, gesijpeld (trickled) 755 2, geslijmd (layed it on) 667 14, gesloofd (drudged) 740 8, gesluisd (channeled) 757 6, gesmeuld (smouldered) 880 1, gesmoesd (whispered) 770 2, gesold (trifled) 820 5, gespeurd (investigated) 728 13, gesproeid (sprayed) 753 7, gestoeid (fought) 628 3, gestuwd (dammed) 660 12, gesuisd (rustled) 834 2, getierd (raged) 745 3, getijgerd (crawled) 782 8, getinteld (tingled) 704 3, getobd (worried) 708 5, getoerd (went for a ride) 720 24, getreurd (grieved) 700 7, getroefd (played trumps) 776 5, getuimeld (tumbled) 758 4, geturfd (tallied) 759 6, gevleid (flattered) 664 4, gewaggeld (tottered) 672 1, gewalmd (smoked) 755 1, gewapperd (flapped) 716 2, geweifeld (wavered) 741 6, gewemeld (teemed) 895 2, gewoekerd (been rank) 669 10, gewurmd (squeezed) 722 6, gezwierd (swayed) 834 6, gezwoegd (laboured) 667 2.

Appendix G, target words with reaction time.

Experiment 2.4b, irregular participles with a high Family Size:
gevroren (frozen) 694, gezwommen (swum) 593, gezogen (sucked) 622, geweken (given in) 702, gestoven (blown) 689, geschoten (shot) 608, gezonden (sent) 610, geslepen (grinded) 598, gereden (driven) 687, gevochten (fought) 607, gefloten (whistled) 666, gezworven (drifted) 701, geslopen (sneaked) 636, gevlogen (flew) 600, gewonnen (won) 685, gesneden (cut) 619, gegoten (poared) 592, gewezen (pointed) 662, gedreven (floated) 582.
Experiment 2.4b, irregular participles with a low Family Size:
getogen (set forth) 692, gehesen (hoisted) 720, gebeten (bitten) 616, gesnoten (blown) 700, geslonken (shrunken) 718, gezwollen (swollen) 712, geroken (smelled) 741, gekrompen (shrunk) 615, gestolen (stolen) 585, gelogen (lied) 654, gegleden (slid) 766, geblonken (shone) 757, gevlochten (braided) 658, gedwongen (forced) 643, gewreven (rubbed) 663, gebleken (appeared) 583, gesnoven (sniffed) 742, gezwegen (been silent) 671, gerezen (risen) 710, geholpen (helped) 596.

Dutch and English compounds

CHAPTER 3

This chapter has been published as Nivja H. De Jong, Laurie B. Feldman, Robert Schreuder, Matthew Pastizzo, and R. Harald Baayen (2002): The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects, *Brain and Language*, **81**, 555-567.

Abstract

In this study, we use the association between various measures of the morphological family and decision latencies to reveal the way in which the components of Dutch and English compounds are processed. The results show that for constituents of concatenated compounds in both languages, a position-related token count of the morphological family plays a role, whereas English open compounds show an effect of a type count, similar to the effect of family size for simplex words. When Dutch compounds are written with an artificial space, they do not reveal an effect of type count, which shows that the differential effect for the English open compounds is not superficial. The final experiment provides converging evidence for the lexical consequences of the space in English compounds. Decision latencies for English simplex words are better predicted from counts of the morphological family that include concatenated and hyphenated but not open family members.

Introduction

Several frequency effects have been reported in the domain of word recognition. The string frequency of the presented word itself (the surface frequency) as well as the summed frequency of all its inflectional variants (the base frequency) have been found to influence response latencies (e.g., Taft 1979; Baayen, Dijkstra, & Schreuder 1997).

Another effect which has recently been found to play a role in lexical processing is a type count effect: the morphological family size. The morphological family of a monomorphemic word consists of all words containing that word as a morpheme. For Dutch simplex words and for stems in derived words, participants respond faster in visual lexical decision to words with large families than to words with small families (matched for surface and base frequency). The summed frequencies of the morphological family members, the family frequency, does not influence reaction times (Schreuder & Baayen, 1997; De Jong, Schreuder, & Baayen, 2000, also Chapter 2). For English, Baayen, Lieber, and Schreuder (1997) showed that simplex nouns with a high family frequency are rated equally high in a subjective frequency rating as nouns with a low family frequency, but nouns with a high family size are rated higher than nouns with a low family size. We have now replicated these results using the visual lexical decision task: The same stimuli as used by Baayen et al. (1997) again showed an effect of family size, and an absence of an effect of family frequency.[1]

There are three independent kinds of evidence suggesting that the family size effect permeates semantic levels of lexical processing. First, the family size effect occurs in tasks requiring central levels of processing (visual lexical decision and subjective frequency rating) but not in a task tapping into form-related stages of lexical processing such as visual progressive demasking (Schreuder & Baayen, 1997). Second, the effect is not mediated by form. Dutch irregular past participles which differ in their orthographic and phonological form from their morphological family members nevertheless show an effect of family size. Furthermore, in the case that the stem-allomorph used in the irregular past participle is by itself a word with a different meaning (e.g., the noun *vocht*, 'moisture', is embedded in the past

---

[1] We conducted two experiments (visual lexical decision: see the procedure reported for Experiment 3.4 of the present paper; 17 participants each), using the materials from Experiments 2 and 3 of Baayen et al. (1997). The words with a high family frequency were responded to equally fast as the words with a low family frequency (661 and 649 milliseconds respectively: $t1, t2 < 1$), but the words with large morphological families were responded to faster than words with small families (643 and 699 milliseconds respectively: $t1(16) = -3.17, p = 0.003; t2(40) = -2.71, p = 0.005$).

participle *gevochten*, ’fought’, which is derived from *vechten*, ’to fight’), a count of morphological family members of such a form-related but not semantically related embedded word does not influence response latencies (De Jong et al., 2000, also Chapter 2). Third, removing opaque family members from the count of family size enhances correlations with reaction times (Bertram et al., 2000).

In the present study, we use variables characterizing the morphological family of the constituents to investigate how compounds are processed and stored. In the literature on the processing of compounds in Dutch, two contrasting results have been reported. On the one hand, Van Jaarsveld and Rattink (1988) report frequency effects for compounds, and the absence of a frequency effect for the constituents of these compounds. These results suggest that compounds are accessed as wholes. For a similar view with respect to compounds in English and other languages, see Marslen-Wilson (2001), but see Taft and Forster (1976). On the other hand, several semantic priming studies have shown that both constituents of semantically transparent (and to some extent opaque) compounds can be primed (Sandra 1990, Zwitserlood 1994).[2] Similar contrasting results have been reported for comparatives in Dutch, for which the frequency of the base form appears to be irrelevant (Bertram, Schreuder, & Baayen, 2000) while at the same time, the family size of the base predicts response latencies (Bertram et al., 2000). Given these results for Dutch comparatives, the present paper investigates whether the family size of the constituent affects the response latencies. If so, this would provide further evidence that the constituents of compounds play a role in the processing of compounds.

Indeed, the first experiment reported below revealed the anticipated outcome: larger left or right constituent family sizes led to shorter reaction times. However, post-hoc analyses show that a different interpretation of the results is called for. The position family frequency, the family frequency of a constituent constrained by position within the compound, is a better predictor of reaction times than is the family size of the constituent. We will argue that this position family frequency effect is a diagnostic of peripheral on-line decomposition of the compound, and that, nevertheless, the frequency of the constituent itself is irrelevant.

Experiment 3.2 replicates the position family frequency effect for English compounds, but only for those compounds which are written without an intervening

---

[2] To eliminate possible confounds with semantic transparency, we used only semantically reasonably transparent compounds in the present study. Further research is required to investigate possible differences in processing the constituents of opaque and transparent compounds with respect to variables referring to the morphological family.

space (henceforth concatenated compounds). For compounds written with a space between the constituents (henceforth open compounds), this experiment revealed a position family size effect rather than a position family frequency effect. Experiment 3.3 shows that the insertion of an artificial space into Dutch compounds does not lead to a position family size effect. Experiment 3.4, finally, shows that English open compounds do not belong to the morphological families of simplex words, providing converging evidence that the different kinds of orthographic conventions of English compounds correlate with different kinds of central representations.

Experiment 3.1: Type or token counts for constituents of Dutch compounds

In two sub-experiments, we investigate the role of the family size of constituents in Dutch compounds using standard visual lexical decision. In post-hoc analyses, we investigate the influence of several frequency counts. In Experiment 3.1a, we contrast the family size of the left constituent. In Experiment 3.1b, we contrast the right constituent family size.

Method

Participants. Twenty-four participants responded to the set of compounds of Experiment 3.1a, and another 24 participants responded to the set of compounds of Experiment 3.1b. Most participants were undergraduates at Nijmegen University and all were native speakers of Dutch.

Materials of Experiment 3.1a. We selected 112 transparent Dutch compounds from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). None of these compounds contained a linking morpheme,[3] and all were compounds composed of two nouns. We built a contrast in family size of the left constituent, while keeping the right constituent constant (e.g., the pair *ijzerwinkel*, ’hardware store’, and *antiekwinkel*, ’antique shop’). Fifty-six left constituents had a mean family size of 52 (30), and the mean family size of the 56 left constituents with a small family was 7 (4), standard deviations between parentheses. The compounds were

---

[3] In Dutch, compounds can occur with the linking morpheme *en* or *s* between the constituents, or without a linking morpheme.

matched on the frequency of the left constituent[4] (high: mean 29.0 (27.8); low: mean 27.2 (28.2)), the compound frequency (high: mean 0.5 (0.5); low: mean 0.5 (0.6)), and for mean length in letters (high: 10.1, low: 10.5). We did not match for the left constituent family frequency, as previous studies had shown that the family frequency does not affect response latencies.

Materials of Experiment 3.1b. We selected a different set of 112 transparent Dutch compounds from the CELEX lexical database. None of these compounds contained a linking morpheme and all were noun-noun compounds. Fifty-six of these compounds had a right constituent with a high family size such as *molen* in *windmolen*, 'windmill' (mean 55 (41)), and 56 compounds with identical left constituents had a right constituent with a low family size such as *vlaag* in *windvlaag*, 'gust of wind', (mean 7 (4)). We matched the compounds on the frequency of the right constituent (high: mean 33.2 (25.8); low: mean 31.9 (30.2)), compound frequency (high: mean 0.5 (0.8); low: mean 0.5 (0.6)), and mean length in letters (high: 10.6, low: 11.3). Note that the properties of these compounds are comparable to the properties of the compounds used in Experiment 3.1a, including the contrast in family size for the left (Experiment 3.1a) and right (Experiment 3.1b) constituent. As in Experiment 3.1a, we did not match for the (right) constituent family frequency, as previous studies had shown that the family frequency does not affect response latencies.

For each experiment, we constructed two lists such that the same constituent occurred in one list only, and such that the number of compounds falling into the two family size conditions were evenly distributed. Each compound was paired with a pseudo compound, which consisted of an existing constituent at either the left or right position combined with a pseudo-constituent which did not violate the phonotactical rules of Dutch.

Procedure. Participants were tested in noise-proof experimental rooms. They were asked to decide as quickly and accurately as possible whether a letter string appearing on the computer screen was a real Dutch word. Each 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 lowercase 21 point Helvetica letters on a dark background and they remained on the screen for 1500 ms. The maximum time span allowed for

---

[4] Throughout this paper, the frequency of a noun (either monomorphemic or compound) is the summed frequency of all its inflectional variants. But note that, besides matching for this frequency, we always also matched on the singular form. All frequency counts are standardized per million.

a response was 2000 ms from stimulus onset.

Results and Discussion

The participants performed both experiments with an overall error rate that was less than 20%. In Experiment 3.1a, the mean reaction time of one item in the high condition (*propagandaliteratuur*, 'propaganda literature') differed by more than 3 standard deviations from the mean reaction time and was excluded from further analyses. Its matched pair (*bekentenisliteratuur*, 'confessional literature') was also excluded from further analyses. Remaining items were entered into analyses by subjects ($t1$) and by items ($t2$). The mean reaction time for the high condition in Experiment 3.1a was significantly shorter than the mean reaction time of the low condition (660 (101) and 698 (96) ms respectively: $t1(23) = -6.09, p = 0.000; t2(108) = -2.47, p = 0.000$) and elicited fewer erroneous responses (0.04 (0.04) and 0.09 (0.08) error proportions respectively: $t1(23) = -3.85, p = 0.001; t2(108) = -2.30, p = 0.012$).[5] For Experiment 3.1b, we also find an effect of 38 milliseconds between the high and low condition (674 (111) and 712 (129) ms respectively: $t1(23) = -5.49, p = 0.000; t2(110) = -2.23, p = 0.014$) and a difference in error proportions, although not reliably so in the by item analysis (0.06 (0.06) and 0.09 (0.06) error proportions respectively: $t1(23) = -2.51, p = 0.010; t2(110) = -1.09, p = 0.140$).

A post-hoc analysis of Experiments 3.1a and 3.1b revealed that besides the family size of both the left and right constituent, the compound frequency was an important factor in determining response latencies.[6] The correlation between compound frequency and reaction times for the words in Experiment 3.1a was $r = -0.383$ ($t(108) = -4.31, p = 0.000$) and for the words in Experiment 3.1b $r = -0.347$ ($t(110) = -3.89, p = 0.000$). Further correlation analyses revealed two important results. First, the correlation of the position family size with reaction times in both experiments was higher than or comparable to the correlation of the family size. The position family of a constituent in a compound consists of a count of family members in which the constituent appears at the same position as it does in the target compound. For a constituent such as *molen*, 'mill', in *windmolen*, the position family would include family members such as *watermolen*, 'water mill', and *koffiemolen*, 'coffee grinder', but not a word such as *molensteen*, 'millstone'. Second, the correlation of the summed frequencies of the position family members (the

---

[5] Throughout this paper we report one-sided t-tests (except when mentioned otherwise), as the studied effects are expected to be facilitatory.

[6] Throughout this paper, all correlation analyses are carried out with log-transformations on all frequencies and family size counts.

position family frequency) was higher than the correlation of the position family size with reaction times.

In previous studies (Schreuder & Baayen, 1997; De Jong et al. 2000, also Chapter 2), the family frequency never played a significant role. In post-hoc correlation studies, a significant correlation might be obtained for the family frequency, but this correlation would always be lower than the correlation with family size and could always be accounted for as being a spurious correlation resulting from the high intercorrelation between family size and family frequency. In the present experiments, however, the correlation for the position family frequency is higher than the correlation for the position family size, suggesting that this position family frequency cannot be a mere spurious correlation. Instead of reporting in detail the subtle differences in these correlations, we constructed, after having pooled all the compounds of both experiments, four post-hoc factorial designs. In the first design, we contrasted the left position family frequency, in the second we contrasted right position family frequency, in the third, left position family size, and in the fourth, right position family size. In all four contrasts, we matched, as best as we could, on all other properties. We were not able to match for identical constituents as we did in the actual experiments, but we were able to match for the properties of the constituents. Table 3.1 summarizes the characteristics of the four post-hoc contrasts, as well as the observed mean reaction times (item means were calculated by subjects). Only the contrast of the left position family frequency yielded a significant difference in reaction times (Contrast 1: $t2(26) = -2.88, p = 0.008$; Contrast 2: $t2(28) = -1.94, p = 0.062$). The results of both post-hoc designs contrasting position family size did not yield in a significant difference (Contrast 3, 4: $t2 < 1$).

In Contrast 5 and 6 of Table 3.1, we compare the lemma frequency of the left and right constituents themselves. Similar to the results of Van Jaarsveld and Rattink (1988), the reaction times for both these contrasts were not significantly different (Contrast 5, 6: $t2 < 1$). In a separate Experiment,[7] we presented all left and right constituents of the compounds of Experiment 3.1a and 3.1b in isolation, as simplex nouns. In this Experiment, both frequency contrasts of Table 3.1 yielded a significant difference in reaction times (Contrast 5 526 and 585 ms respectively: $t2(34) = -3.13, p = 0.002$; Contrast 6 531 and 560 ms respectively:

---

[7] We conducted a standard VLD experiment, see the procedure for Experiment 3.1, except that the size of the letters on the screen for these monomorphemic words was 36 instead of 21 pt. Sixteen participants responded to a list consisting of the constituents of Experiment 3.1a and the pseudo constituents. Sixteen other participants responded to a list consisting of the constituents of Experiment 3.1b and the pseudo constituents.

Table 3.1: Properties and mean reaction times of the compounds in six post-hoc factorial contrasts of Experiment 3.1.

Variable	Contrast 1		Contrast 2		Contrast 3		Contrast 4		Contrast 5		Contrast 6	
	H	L	H	L	H	L	H	L	H	L	H	L
compound frequency	0.4	0.2	0.3	0.2	0.3	0.4	0.2	0.2	0.3	0.3	0.4	0.4
Left constituent												
frequency	70.1	55.9	93.3	123.6	79.1	89.4	84.2	64.8	**127.2**	**6.3**	56.8	62.2
family size	64.2	47.9	45.5	63.2	60.6	22.2	59.6	45.7	34.1	29.1	53.1	43.3
position family size	21.3	18.1	17.6	19.9	**19.9**	**5.7**	17.9	17.5	13.3	14.2	20.4	17.9
position family frequency	**25.5**	**4.1**	11.8	10.6	8.8	8.0	7.8	10.5	11.6	8.9	13.3	13.0
Right constituent												
frequency	92.3	135.3	82.7	71.9	53.0	71.1	53.4	39.5	71.8	79.9	**81.2**	**10.7**
family size	76.8	60.5	59.5	42.4	40.5	42.7	46.6	30.5	61.7	78.9	28.2	24.7
position family size	20.6	22.9	15.5	17.5	16.2	13.4	**19.5**	**5.3**	15.2	21.1	11.9	9.1
position family frequency	10.0	14.2	**32.1**	**4.1**	8.4	10.1	4.2	2.7	14.7	10.5	8.6	7.9
n	14	14	15	15	17	17	17	17	18	18	18	18
Reaction time	643	753	671	744	708	701	687	687	700	686	731	725

Note. Bold underlined entries mark the relevant contrasts.

$t2(36) = -1.74, p = 0.045$). This implies, surprisingly, that the position family frequency of the left constituent plays a facilitatory role while at the same time the frequency of occurrence of this constituent itself when embedded in a compound appears to be irrelevant. We will return to this unexpected finding in the General Discussion.

Does the morphological family affect the processing of English compounds in the same way? Note that English compounds can be written with a space between the constituents, with a hyphen, or they can be concatenated. The question addressed in Experiment 3.2 is whether concatenated and open compounds are processed differently as revealed by differing contributions of position family size and position family frequency, as the constituents of open compounds are represented in their orthography as being simplex words.

Experiment 3.2: English concatenated and open compounds

Method

Participants. 20 participants, mostly undergraduates at the State University of New York, Albany, were paid to take part in this experiment or received partial course credits. All were native speakers of English.

Materials. We selected 120 transparent English compounds from the CELEX lexical database. Sixty of these compounds were written with a space between the two constituents and 60 were concatenated. The concatenated and open compounds always shared one constituent, creating 40 pairs such as *cornflake* and *corn bread* with identical left constituents and 20 pairs such as *tinplate* and *silver plate* with identical right constituents. The concatenated and open compounds were matched on mean compound frequency (concatenated 0.66 (1.19); open 0.63 (1.48)[8]), mean

---

[8] The CELEX lexical database (Baayen et al. 1995) lists all compounds, but does not provide the frequency of occurrence of the open compounds. In order to obtain comparable frequency counts for the open compounds, we made a list of 'word-bigrams' of the corpus used for CELEX and their frequencies. Thus, we obtained the frequency with which a given word, for instance *apple*, was followed by another word (e.g., *pie*). To check whether the concatenated compounds we used were indeed written without spaces in the CELEX corpus, we also calculated the frequency of these concatenated compounds when written with a space. This mean frequency of occurrence was very low: 0.06. The open compounds of this study were never written as concatenated compounds in the CELEX database.

constituent frequency of the not-shared constituent (concatenated 88.2 (119.2); open 103.8 (119.7)), mean family size of the not-shared constituent (concatenated 14.7 (14.3); open 12.0 (17.3)), and mean length in letters of the not-shared constituent (concatenated: 4.7; open: 5.1). The properties of the shared constituents were as follows: mean constituent frequency 232.5 (274.7), mean family size 29.7 (31.8), and mean length in letters 4.6.

We constructed two experimental lists, such that the same constituents occurred in one list only and such that each list had the same number of concatenated and open compounds. Each compound was paired with a pseudo compound, which consisted of an existing constituent either in the left or right position combined with a pseudo constituent which did not violate the phonotactical rules of English. In each list, half of the pseudo compounds were open, and half of them were concatenated.

Procedure. Participants were tested in noise-proof experimental rooms. They were asked to decide as quickly and accurately as possible whether a letter string appearing on the computer screen was a real English word. Each stimulus was preceded by a fixation mark in the middle of the screen for 500 ms. After 500 ms, the stimulus appeared at the same position. Stimuli were presented on a Power Macintosh 6100/60AV personal computer in black lowercase 21 point Helvetica letters on a white background and they remained on the screen for 1500 ms. The maximum time span allowed for a response was 2000 ms from stimulus onset. Responses were registered on a Psyscope response box.

Results and Discussion

One participant performed the experiment with an overall error rate greater than 25%. All data of this participant were excluded from further analyses. The mean reaction time for two compounds (one concatenated and one open compound) exceeded three standard deviations from the mean reaction time and were also excluded from further analyses. The open compounds were responded to faster than the concatenated compounds (mean reaction times 747 (119) and 785 (131) ms respectively: $t1(34) = 4.02, p = 0.000; t2(116) = 2.35, p = 0.020$, using two-sided t-tests) and elicited fewer erroneous responses, although not reliably so (mean error proportions 0.023 (0.040) and 0.042 (0.078) respectively: $t1(34) = 1.55, p = 0.131; t2(116) = 1.95, p = 0.054$).

In the correlation analysis, the correlation for the concatenated and open compounds between reaction times and compound frequency was identical (for both the concatenated and open compounds $r = -0.332; t(57) = -2.66, p = 0.005$). Comparable to the results of the Dutch compounds, the correlation of the reaction times of the concatenated compounds and left position family frequency was slightly higher ($r = -0.328; t(57) = -2.62, p = 0.006$) than the correlation with left position family size ($r = -0.294, t(57) = -2.32, p = 0.012$). In a stepwise linear regression analysis, only the compound frequency remains as a reliable predictor of response latencies. The high intercorrelation of left position family size and left position family frequency ($r = 0.863, t(57) = 12.92, p = 0.000$) leads to substantial collinearity in the data. We therefore also carried out a non-parametric tree-based regression analysis (Breiman, Friedman, Olshen, & Stone, 1984), which singles out left position family frequency and compound frequency as reliable predictors.[9] This suggests that the English concatenated compounds are processed in a similar way as the Dutch compounds. Turning to the English open compounds, however, we see a different pattern. For the English open compounds, we find that the correlation between reaction times and left position family frequency is lower ($r = -0.290, t(57) = -2.29, p = 0.013$) than the correlation for left position family size ($r = -0.350, t(57) = -2.82, p = 0.003$). A stepwise linear regression analysis as well as a non-parametric tree-based regression analysis single out left position family size and compound frequency as reliable predictors.[10] For both the concatenated and open compounds, there were no reliable effects for the right constituents.

Although English concatenated and open compounds are processed similarly with respect to their full form (both kinds show an equal effect of compound frequency), the present results suggest that the left constituents of English open compounds are processed differently than the left constituents of concatenated compounds. Whereas the processing of constituents of concatenated compounds (of English as well as of Dutch compounds) is influenced by position family frequency, the processing of left constituents of open compounds is influenced by position family size. This type effect of the morphological family is comparable to the family size effect of simplex words, suggesting that left constituents of English open compounds are processed more similarly to simplex words than is the case for

---

[9] To gauge the independent influence of the two variables left position family size and left position family frequency for the Dutch compounds in Experiment 3.1, we were able to make post-hoc factorial contrasts. As the number of items for the English compounds was substantially reduced, such post-hoc factorial contrasts were impossible to make.

[10] Some English compounds bear stress on the second constituent rather than on the first constituent (apple ’pie versus ’side walk). In our data, eleven open compounds had primary stress on the second constituent. Excluding these open compounds from the correlational analyses did not alter the pattern of results.

constituents of concatenated compounds. We will return to this issue in the general discussion.

In Experiment 3.3, we investigate whether the difference between concatenated and open compounds in English can be interpreted as a superficial effect of the orthography by inserting an artificial space between constituents of Dutch compounds. If the different effect found for English concatenated and open compounds is due to a superficial orthographic effect, we should be able to induce a similar difference for Dutch concatenated and (artificial) open compounds.

Experiment 3.3: Dutch artifical open compounds

Method

Participants. 24 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We used the same compounds and pseudo compounds as in Experiment 3.1a. We added a space between the constituents of the compounds and pseudo compounds.

Procedure. The procedure was identical to that of Experiment 3.1a, except that we asked the participants to decide as quickly as possible whether the **two** letter strings appearing on the screen were real Dutch words.

Results and Discussion

All participants performed the experiment with an overall error rate less than 11%. The reaction times in this experiment are longer than the reaction times of Experiment 3.1a, in which these compounds were presented without the artificial space (739 and 679 ms respectively; $t1(94) = 3.35, p = 0.000; t2(218) = 4.66, p = 0.000$, using two-sided t-tests).

A correlation analysis showed that, although the inserted space changed the standard visual form of the compound, the correlation between full-form frequency (compound frequency) and reaction times was still highly significant ($r = -0.376$, $t(108) = -4.21, p = 0.000$). Comparable to the results of Experiment 3.1a, in which these compounds were written without an artificial space, the correlation between reaction times and left position family frequency ($r = -0.257, t(108) =$

$-2.76, p = 0.003$) was higher than the correlation with left position family size ($r = -0.180, t(108) = -1.91, p = 0.030$). A partial correlation of the left position family frequency, after partialling out the influence of left position family size showed a significant correlation ($r = -0.198, t(107) = -2.09, p = 0.019$), but the partial correlation of left position family size, after partialling out the influence of left position family frequency, was not significant ($r = 0.071, t < 1$). Similarly, a tree-based analysis ranked left position family frequency above left position family size. These results suggest that although the space between the constituents made the lexical decision task more difficult, as evidenced by the longer reaction times in this experiment in comparison to the reaction times of Experiment 3.1a, the artificial space had no clear effect on the way in which these compounds were processed. Similar to Dutch concatenated compounds and contrary to English open compounds, the processing of Dutch artificial open compounds is influenced by a token effect of the morphological family. We generalize from these results in Dutch that the different contributions of family size and family frequency for concatenated and open compounds in English cannot be attributed to the mere presence of a space between the constituents of a compound in the visual input.

If the difference in processing of English concatenated and open compounds is not due to some superficial effect of the orthography, but rather due to a difference at a deeper level of processing, this suggests that open compounds might be, at a central level, represented in a different way than concatenated compounds. In the three sub-experiments of Experiment 3.4, we investigate this possibility by presenting English simplex words, which are factorially contrasted with respect to the number of different kinds of family members (concatenated, hyphenated, and open family members) while keeping the total family size constant.

Experiment 3.4: The status of diffently-spelled family members

Method

Participants. 20 participants, mostly undergraduates at the State University of New York, Albany, were paid to take part in these experiments or received partial course credits. All were native speakers of English.

Materials of Experiment 3.4a. We selected 34 simplex English nouns divided into two groups of 17 words each, using the CELEX lexical database. These two groups of words were matched on total family size (mean 17.6 (8.4) and mean 15.9 (7.6)) but differed in the proportion within this family of concatenated family members versus the sum of open and hyphenated family members. Seventeen words had a high number of concatenated family members (mean 15.4 (6.3)) and 17 contained only a few concatenated family members (mean 2.6 (1.8)). Note that by contrasting the concatenated family members and at the same time keeping the total family size constant, we also contrasted the sum of open and hyphenated family members together (in the high condition 2.2 and in the low condition 13.3.) In addition to the total family size, the words were also matched on frequency (high: 104.3 (125.6); low: 113.5 (131.1)), and mean length in letters (high: 5.5, low: 4.8).

Materials of Experiment 3.4b. We selected 34 simplex English nouns, using the CELEX lexical database divided into two groups of 17 words each. These two groups of words were matched on total family size (mean 36.1 (23.8), and mean 35.4 (22.7)) but differed in the proportion within this family of family members written with a hyphen versus concatenated or open family members. Seventeen words had a high number of hyphenated family members (mean 19.4 (21.2)) and 17 contained only a few hyphenated family members (mean 1.2 (1.8)). Note that by contrasting the number of hyphenated family members while keeping the total family size constant, we at the same time contrasted the sum of concatenated and open family members (in the high condition 16.7 and in the low condition 34.2). In addition to the total family size, the words were matched on frequency (high: 133.2 (120.7); low: 134.9 (126.2)) and mean length in letters (high: 4.2, low: 4.4).

Materials of Experiment 3.4c. We selected 48 simplex English nouns from the CELEX lexical database. We divided these into two groups of 24 words each matched on total family size (mean 30.7 (21.5) and mean 31.0 (22.0)). The first group had a high number of open family members within this family (mean 11.5 (5.1)) and within the families of the second group only a few family members were written with a space (mean 1.3 (1.3)). Note that by contrasting the number of open family members within the family while keeping the total family size constant, we also contrasted the number of concatenated and hyphenated family members jointly (high condition: 19.2, low condition: 29.7). In addition to matching for the total family size, we also matched the two groups of words on frequency (high: 109.0 (108.5); low: 106.4 (98.2)) and mean length in letters (high: 4.6, low: 4.2).

We combined all simplex nouns of Experiments 3.4a to 3.4c into one list, in which the targets of one experiment served as the fillers of another experiment. Each word was paired with a pseudo word which did not violate the phonotactical rules of English.

Procedure. The procedure was identical to that of Experiment 3.2, except that the size of the letters appearing on the screen for these monomorphemic words was 36 instead of 21 pt.

Results and Discussion

All participants performed the experiment with an overall error rate less than 11%. One word in Experiment 3.4a, *centre*, was excluded from further analysis because its spelling did not conform to the American spelling conventions (resulting in a high error score and long reaction times). The reaction times of three words (one from Experiment 3.4a, and two from Experiment 3.4b) exceeded 3 standard deviations from the mean reaction time and were also excluded from further analysis. Table 3.2 shows the mean reaction times (calculated over the correct responses) and error scores (calculated over all responses) for the two experimental conditions of all three sub-experiments. We used two-tailed t-tests as we have no a-priori hypothesis concerning the direction of potential differences. The difference in reaction times and error scores of Experiment 3.4a (reaction times: $t1, t2 < 1$; error scores: $t1(19) = -1.64, p = 0.117, t2(30) = -1.09, p = 0.286$) and Experiment 3.4b (reaction times: $t1 \approx 1, t2 < 1$; error scores: $t1, t2 < 1$) were not significant. The difference in reaction times of Experiment 3.4c, however, was significant. The words with a low number of open family members were responded to faster than those with a high number of open family members ($t1(19) = 2.93, p = 0.009; t2(46) = 2.89, p = 0.006$). There were no reliable differences in the error scores ($t1, t2 < 1$). Using Bonferroni adjustments, we find that, across the three sub-experiments, the result of Experiment 3.4c remains significant at the 5% level.

Combining these three experiments, we see the same pattern in a post-hoc correlation analysis. If we count the family size of only the concatenated family members, we obtain a marginally significant negative correlation with reaction times ($r = -0.200, t(93) = -1.96, p = 0.053$). Counting only the hyphenated family members results in a similar correlation with reaction times ($r = -0.185, t(93) =$

Table 3.2: Results of Experiment 3.4. Means and standard deviations of response latencies and error proportions (by participants).

Experiment	Amount of family members	RT	(SD)	Error	(SD)
3.4a	Many concatenated	577	(104)	0.00	(0.01)
	Few concatenated	586	(98)	0.01	(0.03)
3.4b	Many hyphenated	556	(101)	0.01	(0.03)
	Few hyphenated	564	(84)	0.02	(0.04)
3.4c	Many open	586	(101)	0.00	(0.01)
	Few open	561	(92)	0.01	(0.02)

$-1.82, p = 0.073$). But restricting the count to only the open family members yields no reliable correlation ($r = -0.004, t < 1$). The best correlation was obtained if the family count was based on the sum of concatenated and hyphenated family members together ($r = -0.306, t(93) = -3.10, p = 0.003$). Just as shown by the reaction times of the factorial designs, these correlation analyses indicate that open family members do not contribute to the effect of family size of a monomorphemic word.

We inspected the family members of the words in these experiments with respect to frequency. The concatenated family members had the highest mean frequency (9.7), but the mean frequencies of the hyphenated (0.3) and open (0.4) family members were not reliably different ($t < 1$). This eliminates the possibility that merely the (low) frequency of these family members would account for the difference in the effects of family size: The open family members were just as (in)frequent as the hyphenated family members, but only the latter showed a reliable correlation with reaction times.

It is well known that that there is variability in the orthography of compounds with respect to spaces, hyphens, or concatenation. Diachronically, phrases which started out with spaces between the words, but which became very common, can now be written as one single word. Synchronically, variation in spelling of compounds is also apparent. In our materials, we checked the variability in spellings of the family members. Of all 2254 family members of the items in Experiment 3.4a to 3.4c, a total of 433 family members also occurred with an alternative spelling. Most of these were hyphenated family members for which the alternative spelling was with a space rather than a hyphen. Of these 433 family members, the frequency of occurrence of the alternative spelling was higher than the frequency of the regular spelling for 174 of these family members. Changing the status (concatenated, hyphenated, or open) of the family members according to these frequencies, such that the highest frequency of the different spellings decides the status, did not change the pattern of results.

General Discussion

This study addresses the processing and representation of compounds in Dutch and English using various measures relating to the morphological family as diagnostics. For simplex words and for stems in derived words the family size, a type count of the morphological family co-determines response latencies. Experiment 3.1 revealed that for constituents in Dutch compounds, such a type count does not predict response latencies. Instead, the position family frequency predicts decision latencies.

Constituents of Dutch compounds are always concatenated, but in English, compounds can be written in three different ways: concatenated, hyphenated, or with a space between the constituents. In Experiment 3.2, we compared English concatenated and open compounds. Similar to the Dutch results of Experiment 3.1, English compounds (concatenated and open) show an effect of compound frequency. For the English concatenated compounds, again similar to the Dutch results of Experiment 3.1, the position family frequency was a better predictor than the position family size. For the English open compounds, by contrast, the effect of position family size was more important. This outcome suggests that the presence of a space between the constituents of open compounds renders the processing of the constituents more comparable to the processing of simplex words, as simplex words also show a type effect of the morphological family, rather than a token effect.

In Experiment 3.3, we inserted an artificial space between Dutch compounds (see also Inhoff, Radach, & Heller, 1999). Comparable to the results of the Dutch concatenated compounds, and contrary to the results of the English open compounds, these Dutch artificial open compounds showed, besides an effect of compound frequency, an effect of position family frequency only. This suggests that the differential effects of the English open and concatenated compounds are not due to a superficial effect of the orthography. Converging evidence for the central level of the observed difference between concatenated and open compounds comes from Experiment 3.4. In this Experiment, English simplex words were presented.

The results of Experiment 3.4c and the post-hoc correlations of all the combined Experiments 3.4a to 3.4c showed that for English simplex words, only the number of concatenated and hyphenated family members plays a role. The more concatenated and hyphenated family members a word has, the faster participants are able to respond. The number of open family members, however, appears to be irrelevant.

Our hypothesis is that the observed effect of position family frequency of concatenated compounds in Dutch and English taps into a selection process based on conditional constituent probabilities. To see this, consider as an example the mini-lexicon of Table 3.3. Imagine that this lexicon represents all compounds. The effect of position family frequency of any compound with the left constituent *molen* can now be formulated as follows, keeping in mind that our participants knew that a target word in Experiments 3.1–3.3 would always be a compound. Conditional on the set of compounds, the probability that a compound has *molen* as its left constituent is $40/100$. Apparently, participants made use of this kind of position-dependent probabilities to speed up their responses.

Table 3.3: Artificial mini-lexicon.

Lemma	frequency
molenwiek	10
molensteen	20
molenrad	10
windmolen	20
windvaan	10
driehoek	30
total	100

Recall that the constituent frequencies of both the left and right constituent themselves are irrelevant, but that the frequency of the compound as a whole does play a role. This seems to suggest that compounds are processed as wholes and that their morphological structure does not play a role. However, although the 'unigram' frequency of the constituents is irrelevant, the existence of the effect of the positional family frequency, a conditional 'word-bigram' probability, shows that lexical processing is nevertheless sensitive to morphological structure. We interpret this conditional ’word-bigram’ effect to occur peripheral to the central lexicon, at the level of the access representations, possibly even affecting eye movements (Hyönä & Pollatsek, 1998).

The differential effect for the English open compounds, as well as the finding that open compounds do not contribute to a simplex word’s family, leads to an interesting question. Surprisingly, the orthographical space between the constituents in English leads to different processes for recognizing the compound itself, and to a differential status in the central mental lexicon of these kinds of words. Possibly, the open compounds are phrase-like entities and therefore fall outside the morphological family of a simplex word. The question which arises in this context is whether the phrase-like status of these compounds has led to writing a space between the constituents, or whether the space in the orthography has led to the phrase-like status. Whatever the direction of causality may turn out to be, the fact remains that, apparently, orthographic phrases are processed and stored differently from orthographic words.

References

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., Lieber, R. and Schreuder, R.: 1997, The morphological complexity of simplex nouns, *Linguistics* **35**, 861–877.

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

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

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**, 419–511.

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.

Hyönä, J. and Pollatsek, A.: 1998, Reading finnish compound words: Eye fixations are affected by component morphemes, *Journal of Experimental Psychology: Human Perception and Performance* **24**, 1612–1627.

Inhoff, A. W., Radach, R. and Heller, D.: 2000, Complex compounds in german: Interword spaces facilitate segmentation but hinder assignment of meaning, *Journal of Memory and Language* **42**, 23–50.

Marslen-Wilson, W.: 2001, Access to lexical representations: Cross-linguistic issues, *Language and Cognitive Processes* **16**, 699–708.

Sandra, D.: 1990, On the representation and processing of compound words: automatic access to constituent morphemes does not occur, *Quarterly Journal of Experimental Psychology* **42A**, 529–567.

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

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

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.

Van Jaarsveld, H. J. and Rattink, G. E.: 1988, Frequency effects in the processing of lexicalized and novel nominal compounds, *Journal of Psycholinguistic Research* **17**, 447–473.

Zwitserlood, P.: 1994, The role of semantic transparency in the processing and representation of Dutch compounds, *Language and Cognitive Processes* **9**, 341–368.

The effect of morphological and sentential context

CHAPTER 4

This chapter is to appear as Nivja H. De Jong, Robert Schreuder, and R. Harald Baayen: Morphological resonance in the mental lexicon, *Linguistics*.

Abstract

Words occur as morphological constituents in other words. The number of complex words (e.g., *great-ness, great-ly, . . .*) in which a base word (e.g., *great*) occurs, its morphological family size, is a strong co-determinant of response latencies in visual lexical processing. Words that occur in many other words are responded to faster than words that occur in only a few other words. Surprisingly, the morphological family size effect is independent of the frequencies of use of the base word and the frequencies of its family members. We report two experiments with adjectives such as *great* presented in different morphological and phrasal contexts. A partition of the morphological family members into nouns, verbs, and two kinds of adjectives revealed differential effects on the response latencies across these contexts. These results imply that the family size effect is context-sensitive. A simulation model shows that the observed effects can be understood as the result of activation resonance in contextually restricted networks of morphologically related words in the mental lexicon. Possibly, the contextually determined co-activation of a word's family members is part and parcel of its overall meaning percept in the brain.

Introduction

Various token frequency effects are known to influence cognitive processing (Hasher & Zacks, 1984). In the lexical domain, it is well-known that high-frequency words are processed faster than words with low frequencies of occurence (e.g., Rubenstein & Pollack, 1963; Whaley 1978; Taft 1979). Recently, another phenomenon has been observed to play a role in lexical processing: Simplex words which occur as constituents in many complex words are responded to faster in a visual lexical decision task than words with only a few morphological family members. Words with many morphological family members also receive higher subjective frequency ratings than words with only a few morphological family members. The token frequencies of the family members are found to be irrelevant, only the number of family members plays a role. This effect of morphological family size is especially interesting from a cognitive perspective in that it is a type frequency effect without a concomitant token frequency effect, in that the type count of morphologically related family members plays a role, but their token frequencies do not. The effect has been observed both for monomorphemic words and for stems in complex words (Schreuder & Baayen, 1997; Baayen, Lieber, & Schreuder, 1997; Bertram, Baayen, & Schreuder, 2000; De Jong, Schreuder, & Baayen, 2000, also Chapter 2).

Three lines of evidence indicate that the family size effect is semantic in nature, and arises at post-identification stages of lexical processing due to activation spreading along lines of shared morpho-syntactic representations. First, the effect disappears when progressive demasking is used instead of lexical decision (Schreuder & Baayen, 1997). In a progressive demasking task, participants are asked to identify target words which are masked. In successive cycles, the mask is shown for shorter latencies and the words are shown for longer latencies, such that the target words gradually seem to emerge from the mask. If progressive demasking is primarily sensitive to the early processes of form identification as argued by Grainger & Jacobs (1996; see De Jong, Schreuder, & Baayen, 2000, also Chapter 2, for discussion), the absence of a family size effect in progressive demasking can be understood if it arises after form identification has been completed. Second, correlation studies show that semantically opaque family members (such as *business* as morphological relative of *busy*) do not contribute to the effect. Correlations of family size and response latencies are higher when opaque family members are removed from the counts (Schreuder & Baayen, 1997; Bertram et al., 2000). Third, De Jong, Schreuder, & Baayen (2000, also Chapter 2) show that the family size effect is carried by the underlying lemma (Levelt, 1989) and not by the actual phonological and orthographic form of the word. For instance, irregular Dutch past participles such as *gevochten*, derived from the verb stem *vecht*, ’to fight’, co-activate all morphologically complex words derived from *vecht*, even though almost all these morphological relatives contain the form *vecht* and not a form containing the string *vocht*. In fact, *vocht* happens to be an independent Dutch noun (’moisture’), and no correlation appears to exist between the family size of such nouns and the response latencies to semantically unrelated past participles such as *gevochten*.

The present study was prompted by two at first sight unrelated findings. First, Bertram et al. (2000) report that the family size effect for Dutch de-adjectival abstract nouns with the suffix *-heid* (*eenzaam-heid*, ’loneli-ness’) seems to be restricted to a specific subset of morphological relatives. This subset includes complex nouns and verbs with adjectival stems (e.g., *ver-eenzaam-en*, ’to become lonely’) without further restrictions. The set of adjectives in the family, however, appears to exclude color compounds such as *blauw-groen*, ’blue-green’, and intensified adjectives such as *ijs-koud*, ’ice-cold’. Interestingly, these are exactly the adjectives to which *-heid* is hardly ever attached (?*ijskoudheid*, ?*ice-coldness*). Bertram et al. (2000) point out that a semantic restriction on suffixation of *-heid* (avoid intensified adjectives and color compounds) that must be operative in language production also seems to play a role during comprehension (intensified adjectives and color compounds are not co-activated). In what follows, we will refer to the intensified adjectives and color compounds as scale-focusing adjectives, and to the remaining adjectives as general adjectives. We refer to the intensified adjectives and the color-compounds as scale-focusing because the modifiers in these complex adjectives narrow down the general meaning of the adjectival head to a specific part of the scale covered by the head: *ice-cold* denotes an extreme location on the scale of coldness, and likewise *blue-green* denotes a particular shade of green in the range of hues covered by the general term *green*. Second, De Jong et al. (2000, also Chapter 2) report that the presence of an overt verbal inflectional suffix in Dutch verbs triggers greater co-activation of verbal family members compared to Dutch verb forms without an overt verbal marker. In Dutch, first person present tense verb forms do not carry an affix, whereas the third person present tense is the stem plus the inflectional suffix *-t*. In De Jong et al. (2000, also Chapter 2), we presented both forms of the same verb stems (e.g., *sjouw* and *sjouwt*, ’drag’ and ’drags’) and found that verbal family members (e.g., *wegsjouwen*, ’to drag away’) only contributed to the effect of family size in the case that the verbs were presented with the overt inflectional marker.

What these two findings have in common is that the presence of a suffix appears to condition which morphological family members may become co-activated. This suggests that the family size effect might be context-sensitive. In the present study, we systematically investigate this possible context-sensitivity not only for morphological contexts, but also for small phrasal contexts. Visual lexical decision experiments using the same 40 Dutch monomorphemic adjectives were conducted using four contexts: BASE (the simplex adjective without context), COMPARATIVE (Base followed by *-er*), VERY (Base preceded by the modifier *heel*, 'very'), and NOT (Base preceded by the negation *niet*, 'not'). In addition to the contrast between morphological versus phrasal contexts, we have a contrast between a neutral condition (BASE, NOT) and a non-neutral condition (COMPARATIVE, VERY). For the non-neutral condition, we have two expectations.

First, we expect that the adjectives in the morphological family will contribute more strongly to the family size effect than in the neutral condition. This expectation is based on the finding that verbal family members contribute more to the family size effect in the presence of the overt verbal inflectional suffix *-t*. Just as the *-t* boosts the contribution of the verbs in the family, the comparative suffix *-er* might boost the contribution of the adjectives in the family. Likewise, the adjectival modifier *heel* might also boost the contribution of adjectival family members. This hypothesis is based on the observation that *heel* predominantly precedes adjectives, whereas *niet* does not show such a prevalence. Indeed, in 'word-bigram' probabilities derived from the corpus used for CELEX, *heel* and *niet* are found to combine with different word classes. Table 4.1 shows the number of times *heel* and *niet* are followed by different word-form types of adjectives, verbs, nouns, and other words. The number of word types following *heel* is lower than for *niet* (631 and 7093 respectively). A χ^2-test revealed that the distributions among the word classes for these two contexts differ significantly ($\chi^2(3) = 896.55, p = 0.000$). As can be seen in Table 4.1, which also shows the percentages of word types following *heel* and *niet* in these different word classes, this difference in distribution is mainly due to the fact that *heel* preceding adjectives is overrepresented, and *niet* preceding verbs is overrepresented. It should be noted that these numbers only provide a rough estimation, as many word forms are ambiguous with respect to their word class. For instance, as participles (which we counted as verbs) can functionally be adjectives, the number of word types that follow *heel* and that are unambiguously verb forms, reduces from 165 to a mere 14. From these distributional properties, we hypothesize that the VERY condition will serve as a non-neutral condition, in that the adjectival subfamilies of the targets might be boosted, wheras in the *not* condition, such a preference for adjectival family members should not occur (and, perhaps, a preference for verbal family members can be expected).

Table 4.1: Number of word-form types within different word classes which follow *heel* and *niet* according to 'word-bigram' probabilities in the corpus used for CELEX, percentages given in parentheses.

Word class	*niet*	%	*heel*	%
Adjectives	1580	(22)	567	(61)
Nouns	577	(8)	165	(18)
Verbs	4459	(63)	124	(13)
Other	477	(7)	75	(8)

Second, we also expect that scale-focusing adjectives (color compounds and intensified adjectives) might not contribute to the family size effect in the non-neutral conditions. The comparative suffix may well be subject to the same semantic restrictions as reported for *-heid* by Bertram et al. (2000), as formations such as ?*ijs-koud-er*, '?ice-cold-er' and ?*blauw-groen-er*, '?blue-green-er' seem ungrammatical. Likewise, phrases such as *very icecold* and *very blue-green* seem odd, possibly because, e.g., *icecold* itself is already as cold as you can get. Note that for all four conditions, the focus of our interest is on the way in which subsets of morphological family members are activated as a function of morphological and phrasal context in which the adjective stems occur.

In what follows, we first present the experiments, which replicate the finding that words with a large morphological family are responded to faster than words with a small morphological family. We then proceed to show that, depending on the context, different morphological subfamilies indeed affect the response latencies, albeit not necessarily in the way we originally predicted. Finally, we present a new interactive activation model which provides excellent fits to the reaction time data. This computational model is a first attempt to chart the kind of lexical organization in the mind that underlies the family size effect.

We carried out two experiments in order to ascertain the role of the morphological and phrasal context on the activation of the morphological family members. The first experiment contrasted simplex adjectives (the BASE condition) and the same adjectives in the context of the comparative suffix *-er* (the COMPARATIVE condition). Experiment 4.2 used the same set of adjectives and varied the phrasal context. In the NOT condition, the adjective was preceded by *niet*, and in the VERY condition, it was preceded by *heel*. Both experiments made use of a within-subject design.

Experiment 4.1: Morphological context

Method

Participants. 32 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We selected 40 monomorphemic adjectives from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995) of the type *mooi*, ’beautiful’, and the same 40 adjectives in the comparative form (e.g., *mooier*, ’more beautiful’). Twenty adjectives had a high Family Size (mean 52, range 10–171, *SD* 42) and 20 had a low Family Size (mean 3, range 0–9, *SD* 2). The two subsets of high and low Family Size were matched with respect to Base Frequency (high: mean 86.4 (All frequency counts standardized per million.), range 3.2–405.7, *SD* 111.0; low: mean 86.5, range 3.2–403.1, *SD* 111.3) and mean length in letters. The length for the monomorphemic adjectives was 4.7 and 5.1 in the high and low condition respectively. The two subsets were also matched with respect to Surface Frequency for both the monomorphemic adjectives (high: mean 84.5, range 0.8–452.3, *SD* 115.0; low: mean 57.0, range 0.8–315.1, *SD* 80.8) and the comparatives (high: mean 2.3, range 0.0–11.7, *SD* 3.0; low: mean 2.6, range 0.0–18.1, *SD* 4.3). The materials are listed in the Appendix.

We added 104 fillers to the experimental list: 36 monomorphemic nouns and 68 inflected and uninflected verbs. A participant had to respond either to the comparative form of the adjective or to the uninflected form, but never to both. Each word was paired with a pseudo word, with the same morphological structure. The phonotactics of the pseudo words did not violate the phonology of Dutch. The experiment was preceded by 24 practice items. There was a short pause after the practice session, and a short pause halfway through the experimental list. In total, the experiment lasted approximately 15 minutes.

Procedure. Participants were tested in noise-proof experimental rooms. They were asked to decide as quickly and accurately as possible whether a letter string appearing on the computer screen was a real Dutch word. Each 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 lowercase 36 point Helvetica letters on a dark background and they remained on the screen for 1500 ms. The maximum time span allowed for a response was 2000 ms from stimulus onset.

Results

The participants performed the experiment with an overall error rate less than 15%. For each word we calculated mean response latencies (over the correct responses) and error scores (over all responses). The upper half of Table 4.2 shows the means and standard deviations for the experimental conditions of the experiment. An analysis of variance for reaction times revealed a main effect for Context (BASE and COMPARATIVE) and a main effect for Family Size, but no interactions (by participants and by items, $F1$ and $F2 < 1$). The monomorphemic adjectives were responded to faster than the comparatives ($F1(1, 31) = 61.33,$ MSE $= 91123.9, p = 0.00; F2(1, 76) = 19.69,$ MSE $= 68862.7, p = 0.00$), probably due to the higher Surface Frequencies of the monomorphemic adjectives. Words with a high Family Size were responded to faster than words with a low Family Size ($F1(1, 31) = 29.30,$ MSE $= 32420.1, p = 0.00; F2(1, 76) = 7.97,$ MSE $= 27865.0, p = 0.01$). An analysis of variance for the error scores shows a similar pattern, but the main effect of Family Size is not reliable in the item-analysis. The comparatives elicited significantly higher error scores than the monomorphemic adjectives ($F1(1, 31) = 22.72,$ MSE $= 0.08, p = 0.00; F2(1, 76) = 19.69,$ MSE $= 68862.7, p = 0.00$) and the words with a low Family Size elicited higher error scores than the words with a high Family Size ($F1(1, 31) = 8.03,$ MSE $= 0.03, p = 0.01; F2(1, 76) = 2.35,$ MSE $= 0.02, p = 0.13$). Again, no interactions were found ($F1, F2 < 1$).

Experiment 4.2: Sentential context

Method

Participants. 32 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. We used the same 40 monomorphemic adjectives as in Experiment 4.1, but presented them following the word *heel*, ’very’, or following the word *niet*, ’not’, creating combinations like *heel mooi*, ’very beautiful’, or *niet mooi*, ’not beautiful’.

Table 4.2: Results of Experiment 4.1 and 4.2: Means and standard deviations of response latencies and error proportions (by participants).

		RT	Error	*SD* RT	*SD* Error
BASE	High Family Size	560	0.02	80	0.04
	Low Family Size	596	0.04	84	0.07
COMPARATIVE	High Family Size	617	0.05	91	0.06
	Low Family Size	646	0.10	79	0.09
VERY	High Family Size	602	0.01	80	0.03
	Low Family Size	626	0.01	84	0.03
NOT	High Family Size	639	0.05	103	0.07
	Low Family Size	660	0.04	100	0.06

The two words were presented simultaneously on the computer screen. We added the same 104 fillers of nouns and verbs. The filler nouns were presented with either a definite article *de, het*, ’the’, or with the indefinite article *een*, ’a’. For mass nouns, which syntactically cannot be presented with the indefinite article, we used *wat*, ’some’. The filler verbs were presented following the personal pronoun *ik* or followed by the personal pronoun *jij?* (and a question mark). For example *ik kwets*, ’I hurt’, or *kwets jij?*’, ’do you hurt?’. A participant had to respond either to the adjective presented together with *heel*, or to the adjective presented together with *niet* but never to both. The pseudowords (identical to the ones used in the previous experiment) were presented with the same contexts as the words and in the same proportions. In this way, if a phrase contained a nonword, the nonword was in the majority of the cases the second word in the phrase (following *heel, niet, de, het, een, wat*, or *ik*), but could also be presented in the first position (followed by *jij?*). The experiment was preceded by 24 practise items. There was a short pause after the practise session, and a short pause halfway through the experimental list. In total, the experiment lasted approximately 15 minutes.

Procedure. The procedure was almost identical to that of Experiment 4.1, except that participants were now asked to decide whether the two letter strings that appeared on the computer screen were real Dutch words. Each stimulus was preceded by a fixation mark in the middle of the screen for 500 ms. After 50 ms, the two words (the whole phrase) appeared centered at the same position. Stimuli were presented on Nec Multisync color monitors in white lowercase 21 point Helvetica letters (instead of 36 in Experiment 4.1) on a dark background and they remained on the screen for 1500 ms. The maximum time span allowed for a response was again 2000 ms from stimulus onset.

Results

The participants performed the experiment with an overall error rate less than 15%. For each word we calculated mean response latencies (over the correct responses) and error scores (over all responses). The bottom half of Table 4.2 shows the means and standard deviations for the experimental conditions of the experiment. An analysis of variance revealed a reliable main effect for Context (VERY and NOT): Adjectives presented with the word *heel* were responded to faster than adjectives presented with the word *niet* ($F1(1,31) = 18.39, MSE = 40893.7, p = 0.000; F2(1,76) = 7.01, MSE = 26949.9, p = 0.010$). This may be due to the increased difficulty in the NOT condition to respond with ”yes” while processing a word meaning ”not”. Adjectives with a high Family Size were responded to faster than adjectives with a low Family Size, but this main effect of Family Size was not reliable in the by-item analysis ($F1(1,31) = 14.00, MSE = 16375.5, p = 0.001; F2(1,76) = 2.57, MSE = 9891.2, p = 0.113$). No interactions were found ($F1, F2 < 1$). An analysis of variance of the error scores revealed a main effect for Word Context only: Words presented with *niet* elicited more erroneous responses than the words presented with *heel* ($F1(1,31) = 18.39, MSE = 40893.7, p = 0.000; F2(1,76) = 7.01, MSE = 26949.9, p = 0.010$), suggesting that indeed the semantics of *niet* interfered with providing the correct response. All other F-values for the error analysis were less than 1. Alternatively, as one of our reviewers pointed out to us, there might be a difference in reaction times and error scores between these two phrasal contexts due to a difference in scope. The word *niet* can have scope over a single constituent or over an entire sentence, whereas the intensifier *heel* predominantly has a narrow scope. Thus, the phrases of the adjectives in the *niet* context in our experiment were ambiguous, which might have affected response latencies and error scores, especially since the wide scope reading entails treating phrases such as *niet mooi*, ’not beautiful’ as truncated elliptical sentences.

Summing up the results for Experiment 4.1 and 4.2 with respect to the effect of family size, the factorial contrast between a high and a low family size was reliably reflected in the response latencies to the adjectives in the BASE and COMPARATIVE conditions. For the phrasal conditions, the family size effect was weaker and did not reach significance in the by-item analysis. The reason why the family size effect might be attenuated for words presented in a phrasal context becomes apparent when we compare the correlational structure between the four contexts of these experiments.

Post-hoc correlations

Table 4.3 shows the by-item Spearman correlations between response latencies (RT) and Family Size for the conditions in both Experiments, reliable except for the context VERY in Experiment 4.2. To understand why the VERY context behaves differently, we consider, for each of the four context conditions, the correlational structure of the response latencies to the 40 targets with different subsets of family members. We divided the total families into four subsets of family members. First, the nominal family members (N), second, the verbal family members (V), and two kinds of adjectival family members: Scale-focusing adjectives (color compounds and intensified adjectives), and general adjectives. For example, *mooi* has in total 8 family members. Dividing them up into the four subsets, we find that *mooi* has 3 nominal family members, 1 verbal family member, 5 family members which are general adjectives, and 3 adjectival family members which are scale-focusing. Table 4.4 shows the properties (means and standard deviations) of these four subsets of family members for all 40 targets. In the four contexts, the counts for these subsets of family members will remain the same, as accross contexts, the same 40 targets were presented. This enables us to compare the correlational structures in the four different contexts. Recall that we expect adjectival family members to contribute more to the family size effect in the non-neutral conditions (VERY and COMPARATIVE) than in the neutral conditions (BASE and NOT). Also recall that we hypothesized that scale-focusing adjectives might not contribute to the family size effect in the non-neutral conditions.

Figure 4.1 plots $-r_s$ as a function of the size of the four subfamilies: nouns (N), verbs (V), general adjectives (A1), and scale-focusing adjectives (A2). As expected, which subgroup of family members correlates best with reaction times of the 40 target words differs from context to context. In the BASE condition, shown in the upper left panel of Figure 4.1, the family size effect is driven by the general adjectives, the nouns, and also to some extent the verbs in the family. The scale-focusing adjectives do not contribute at all to the family size effect. In the COMPARATIVE condition, shown in the upper right panel, the general adjectives stand out with a particularly high correlation. The lower left panel plots the correlations for the VERY condition, and shows that here, surprisingly, the scale-focusing adjectives constitute

Table 4.3: Spearman correlations and p-values of Family Size and RT for four morphological and phrasal contexts.

Context	Family Size	
	r_s	p
BASE	-.32	.043
COMPARATIVE	-.39	.016
VERY	-.20	.210
NOT	-.36	.023

Table 4.4: Means and standard deviations of the different subfamilies of the 40 target words.

Subfamily	mean	SD
Nouns (N)	15.9	22.4
Verbs (V)	3.2	7.9
General adjectives (A1)	6.8	13.9
Scale-focusing adjectives (A2)	2.0	5.5

the primary subfamily responsible for the family size effect. Finally, as summarized in the lower right panel, the general adjectives seem most prominent for the NOT condition, the overall pattern being remarkably similar to that of the COMPARATIVE condition.

Interestingly, our original predictions are in part confirmed by the data, and in part refuted. What is confirmed is the prediction that adjectives should contribute more strongly to the family size effect in the non-neutral conditions. Although we initially classified the NOT condition as neutral, it appears to behave as a non-neutral condition similar to the COMPARITIVE condition. This may be due to the actual experimental context, in which *niet* ('not') was always followed by an adjective, effectively turning this supposed neutral condition into a non-neutral condition in the sense that it favors adjectival family members to become co-activated.

Our predictions concerning the scale-focusing adjectives turned out to be wrong. First consider the BASE condition. In the absence of a suffix such as *-heid* ('-ness') that might exclude incompatible family members from contributing to the

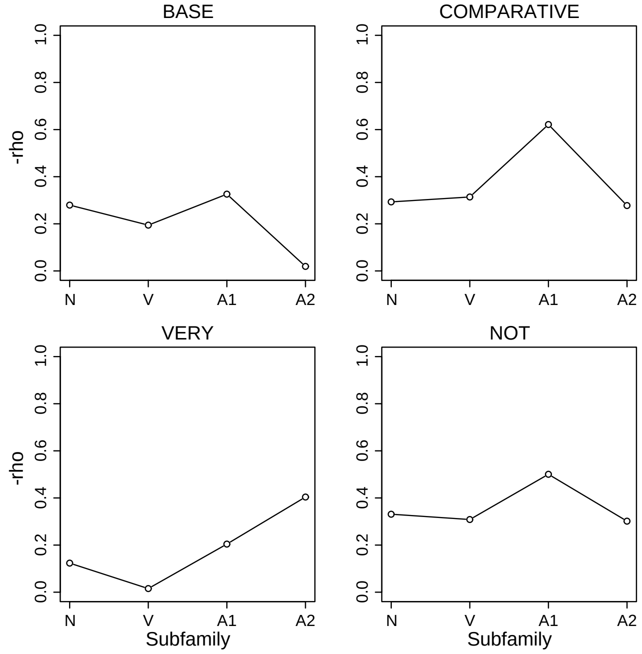

Figure 4.1: Correlations r_s of Reaction Times with the family counts of nouns (N), verbs (V), general adjectives (A1), and scale-focusing adjectives (A2) for four experimental conditions.

family size effect, we expected the scale-focusing adjectives to fully participate, contrary to fact. This behavior of the scale-focusing adjectives is probably due to the small number of scale-focusing adjectives in the pooled families of our experimental words. In the study of Bertram et al. (2000), the scale-focusing adjectives comprised half of all adjectival family members. In the present study, they comprise roughly one fifth of all the adjectives, which suggests that there might be too few scale-focusing adjectives to effectively co-determine the response latencies in the BASE condition.

However, when we turn to the lower left panel of Figure 4.1, we find that the scale-focusing adjectives reveal the strongest correlation of all subfamilies when the adjective is preceded by *heel* ('very'). Apparently, a small number of scale-focusing adjectival family members is still able to give rise to strong correlations with the response latencies, provided that the experimental adjective appears in the right context. The crucial property of the context supplied in the VERY condition seems to be the near synonymy of *very* with intensifiers such as *ice* in *ice-cold*, which has a meaning that comes close to that of *very cold*. For the color compounds within the subfamily of scale-focusing adjectives (e.g., *blauwgroen*, 'blue-green'), it can be argued that although the first constituent of the compound does not intensify the color of the second constituent, it does narrow down on the scale of (actually both) colors. This scale-focusing property is shared with the intensifier *heel*, ’very’. It should be mentioned, however, that only two target words in our experiments were colors themselves (*blauw*, ’blue’ and *groen*, ’green’) and excluding these two targets from the analysis does not change the results. Therefore, with these results we can only speculate as to whether these few color compounds within the subset of scale-focusing adjectives were actually contributing to the family size effect in the VERY condition. Nevertheless, the BASE and VERY conditions clearly show that the contribution of a subfamily may be crucially determined by phrasal context.

The observation that different subfamilies are primarily responsible for the family size effect, an observation which also figures in the simulation study reported below, sheds light on why the main effect of family size is not reliable in the $F2$-analysis in Experiment 4.2. The orthogonal contrast built into the experiment assumed that all family members would play a role, a contrast of 52 versus 3. However, if we count exclusively general family members, the contrast is 7.3 in the high condition and 0.3 in the low condition. Restricting the count to the scale-focusing family members results in a contrast of 3.6 versus 0.7. Counting the family size in this restricted way results in ’high’ conditions with hardly any items with a truly large family size. We suspect that this effective orthogonal manipulation was too weak to show up in the by-item analysis of variance. Nevertheless, by using the actually relevant subfamily counts, reliable correlations emerge.

Thus far, we have described the main patterns in the correlational structures for the four experimental conditions. A problem that arises in the analyses of the present data is the massive collinearity of the various counts of subfamilies. In general, if a word has many nominal family members, chances are high it will also have many verbal and adjectival family members. Conversely, words with hardly any verbal family members are not likely to have many adjectival or nominal family members. Due to this collinearity, it is unclear which subfamilies primarily contribute to the family size effect as a function of context. In what follows, we introduce a new interactive activation model that has proved useful for understanding the way in which context and subfamilies interact, the Morphological Family Resonance Model.

Simulation studies

The architecture of the Morphological Family Resonance Model, henceforth MFRM, is sketched in Figure 4.2. As the family size effect is a central semantic effect, the MFRM focuses on the lemma representations of words in the sense of Levelt (1989) and their associated syntactic and semantic representations in the sense of Schreuder & Baayen (1995). It is assumed that the visual presentation of a target word leads to the activation of the corresponding access representation, which in turn activates the target's lemma. The MFRM models what happens once the target lemma has been activated.

Figure 4.2 displays four lemma representations at the left hand side of the graph, in the array labelled $\mathcal{L}$. The target lemma, *greatness*, is underlined, and its family members are shown in italics. One word, *think*, does not belong to the morphological family of *greatness*. At the right hand side of the graph, three arrays of central representations are shown: syntactic representations labelled $\mathcal{S}$, affix representations labelled $\mathcal{D}$, and meaning representations $\mathcal{M}$. As in the model outlined in Schreuder & Baayen (1995), central representations are shared by the lemmas. For instance, the meaning representation *great* is shared by the lemmas *great*, *greater*, and *greatness*.

A resonance cycle in the MFRM consists of two stages. In the first stage, activation spreads from the lemmas to the central representations. In the second stage, activation spreads back from the central representations to the lemmas. The flow of activation during the very first resonance cycle is indicated by solid lines in Figure 4.2. The additional flow of activation that occurs during subsequent resonance cycles is indicated by dashed lines. Over time, the activation of the target lemma increases exponentially, with the rate of increase being determined by the extent of the morphological family.

First consider the situation in which a target word has many family members. During the first resonance cycle, it activates its corresponding central representations. Because the target has many family members, these central representations will activate a great many other lemma representations. During the next resonance cycle, these many lemma representations begin to contribute to the activation of the central representations, including those shared with the target. Hence, during the second stage of the second resonance cycle, the target and its family members will receive activation from highly activated central representations. During subsequent resonance cycles, this process is repeated, resulting in lemma activation levels that increase exponentially.

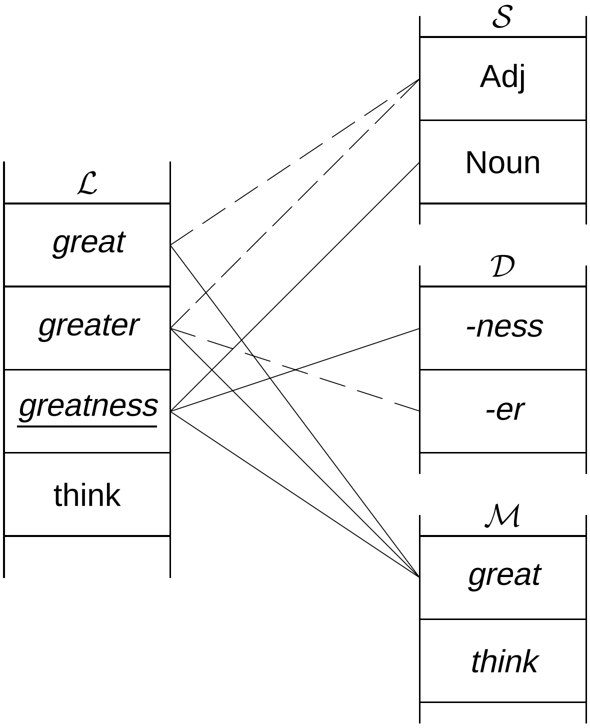

Figure 4.2: Resonance at the initial timestep (solid lines) and additional resonance at following timesteps (dotted lines) for the target word *greatness* (underlined). $\mathcal{L}$ stands for lemma representations, and $\mathcal{S}$, $\mathcal{D}$ and $\mathcal{M}$ represent central representations: syntax, affixes, and meanings respectively.

Next consider the extreme situation of a target word without a morphological family. In this case, the resonance in the system is restricted to the flow of activation between the lemma target and its central representations. Because there are no other lemma representations to contribute to the activation levels of these central representations, the rate at which the activation level of the target increases is very small.

Model times are determined by the resonance cycle in the MFRM at which a lemma reaches a preset threshold activation level. Lemmas with a large morphological family will quickly reach threshold activation level, resulting in small model times. Conversely, lemmas with small families will require many resonance cycles to reach threshold, resulting in long model times. In what follows, we present a formal, explicit definition of the MRFM.

We first need to define the set of words that participate in the resonance, the resonance set. In the simplest case, the resonance set includes all lemmas in the morphological family. The morphological family of a lemma l_i is the set of lemmas l_j

that share a semantic representation $m \in \mathcal{M}$ with l_i and that also share at least one non-affixal (phonological) form representation $e \in \mathcal{E}$ (these form representations are not shown in Figure 4.2):

$$f(l_i) = \{l_j : \mathcal{E}_j \cap \mathcal{E}_i \neq \emptyset \wedge \mathcal{M}_j \cap \mathcal{M}_i \neq \emptyset\}. \tag{4.1}$$

This definition formalizes the linguistic insight that morphological relations consist of systematic correspondences between aspects of form and aspects of meaning. We shall see below that the context effects can be explained by imposing further restrictions on the resonance set.

Each lemma representation l_i is connected with one syntactic representation $s \in \mathcal{S}$, with zero or more affix operator representations $d \in \mathcal{D}$, and with zero or more semantic representations $m \in \mathcal{M}$. The input lexicon used to install the sets of representations $\mathcal{L}, \mathcal{S}, \mathcal{D}, \mathcal{M}$ is derived from the CELEX lexical database. For each experimental adjective, all words in CELEX with a frequency of occurrence greater than 1 in a 42 million corpus of Dutch and containing the adjectival base word as a constituent according to the CELEX parse information were selected. For each word i, a lexical entry $< l_i, \mathcal{S}_i, \mathcal{D}_i, \mathcal{M}_i >$ was created in the model's input lexicon. For a word such as *greenishness*, the lexical entry would be $<$ *greenishness*, {Noun}, {*-ish, -ness*}, {*green*} $>$.

Initially, all representations have an activation level of zero. During the first stage of each resonance cycle, two events take place. First, the target lemma l_i receives a given amount of activation α from its associated access representation. Thus, the activation $a(l_i, t)$ of target l_i is increased at each timestep t by α. Second, any lemma l_j in the resonance set (including the target lemma itself) propagates part of its activation $a(l_j, t)$ forwards to the central representations $x \in \mathcal{X}_i$, $\mathcal{X}_i \in \{\mathcal{S}_i, \mathcal{D}_i, \mathcal{M}_i\}$ to which it is connected. The activation level $a(x, t)$ of a central representation x after t resonance cycles (at timestep t) equals:

$$\begin{aligned} a(x,t) &= \delta\{a(x,t-1) + \rho_{\mathcal{X}}[\alpha + a(x,t-1) \\ &+ \sum_{l_j \in f(l_i)} \mathsf{I}_{[x \in \mathcal{X}_j]}(a(l_j, t-1) + \mathsf{I}_{[j=i]}\alpha)]\}. \end{aligned} \tag{4.2}$$

In (4.2), δ $(0 \leq \delta \leq 1)$ represents a global decay rate, and $\rho_{\mathcal{X}}$ $(0 \leq \rho_{\mathcal{X}} \leq 1)$ represents the resonance sensitivity for the different kinds of central representations $\mathcal{S}, \mathcal{D}$, and $\mathcal{M}$. The idea is that different central subsystems can participate in the resonance to different degrees. Technically, the differential resonance sensitivities allow us to avoid situations in which a particular subset of central representations becomes overly dominant in the resonance.

During the second stage of each resonance cycle, activation is propagated back from the central representations to the lemma layer. The resulting activation level $a(l_k, t)$ of a lemma l_k at the end of timestep t equals

$$\begin{aligned} a(l_k,t) &= \mathsf{I}_{[l_k \in f(l_i)]} \delta \{ a(l_k, t-1) + \mathsf{I}_{[k=i]} \alpha \\ &+ \rho_{\mathcal{L}} [\alpha + a(l_k, t-1) + a(s_k, t) + \\ &+ \frac{1}{c(\mathcal{D}_k)} \sum_{d \in \mathcal{D}_k} a(d,t) \\ &+ \frac{1}{c(\mathcal{M}_k)} \sum_{m \in \mathcal{M}_k} a(m,t)] \}. \end{aligned} \tag{4.3}$$

In (4.3), $\rho_{\mathcal{L}}$ ($0 \leq \rho_{\mathcal{L}} \leq 1$) denotes the resonance sensitivity of the lemmas. The cardinality of the sets of affixes and meanings is denoted by $c(\mathcal{D})$ and $c(\mathcal{M})$ respectively. The factors $1/c(\mathcal{D})$ and $1/c(\mathcal{M})$ in (4.3) ensure that lemma l_k will always receive the same amount of activation from any of the three sets of central representations, irrespective of the number of affix and meaning representations to which it is connected. This normalization ensures that resonance among family members sharing many semantic and/or affix representations does not become so strong that family members reach threshold activation level before the target word.

Model times are defined in terms of the first timestep t' at which $a(l_i, t) >= \theta$. Once the target lemma has reached threshold activation level, it will no longer receive activation from its corresponding access representation. In the model, this is captured by setting α to zero. No resonance takes place, and the activation levels of all representations begin to decay with rate δ. Formally, for $t > t'$,

$$a(x,t) = \delta a(x, t-1). \tag{4.4}$$

All simulations reported were run with $\alpha = 1.0$, $\delta = 0.98$, $\rho_{\mathcal{L}} = \rho_{\mathcal{M}} = 0.02$, $\rho_{\mathcal{S}} = \rho_{\mathcal{D}} = 0.01$, and $\theta = 100.0$.

Having completed the formal definition of the MFRM, we now turn to consider the model's performance. In the simulation runs, the activation levels of all units in the model are reset to their default values between trials. Figure 4.3 plots the correlations of the observed RTs and subfamily sizes for the four experimental conditions by means of solid lines. The corresponding correlations of the model times and subfamily sizes are represented by dashed lines. We observe that the correlations generated by the model are similar to the empirical correlations apart from a shift along the vertical axis. The model predicts much higher correlations than we ac-

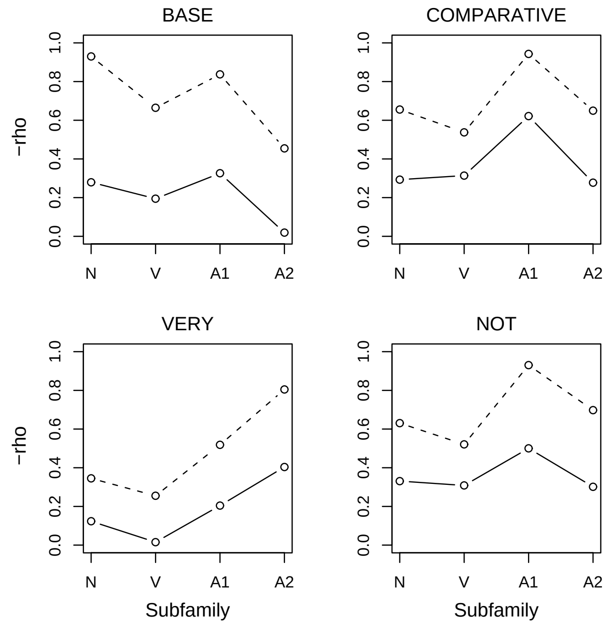

Figure 4.3: Correlations r_s of Reaction Times (solid lines) and model times (dashed lines) with the family counts of nouns (N), verbs (V), general adjectives (A1), and scale-focusing adjectives (A2) for four experimental conditions.

tually observe. This is not so surprising, as the model's predictions are based on resonance only, without taking into account the effect of word frequency and the many other factors that co-determine response latencies.

The results for the BASE context were obtained without further conditioning within the morphological family. For the other three contexts, the notion of the resonance set turned out to be crucial. No good fits can be obtained for these contexts when all family members are allowed to participate in the resonance. Much better results ensue when we specifically exclude subsets of family members from participating in the resonance by removing these family members from the model's input lexicon. Note that the model in its present form does not provide an explanation for how the restricted resonance sets arise for the various contexts. We leave this issue to further research.

For the COMPARATIVE and the NOT contexts, a good fit required restricting the resonance set to the adjectives, including both the general and the scale-focusing adjectival family members.[1] In the case of the VERY condition, by contrast, a good fit

---

[1] Similar fits are obtained when the resonance sets contain only the general adjectives. Hence, our results do not allow us to decide whether *-er* behaves similarly to *-heid*, in the sense that only

required restricting the resonance set to the scale-focusing adjectives only, i.e., to the color compounds (*blauwgroen*, ’blue-green’) and intensified adjectives (*steenkoud*, ‘stone-cold’). Intensified adjectives were coded in the model’s lexicon with ’very’ as part of their semantic representation (for *steenkoud*, the entry was $<$ *steenkoud*$, \{$Adj$\}, \emptyset, \{$*very, cold*$\} >$.

For the NOT and VERY contexts, in which *niet* and *heel* were always followed by an adjective, the model first recognizes *heel* or *niet*, both of which are specified in the lexicon as adjectival modifiers, and then proceeds to recognize the following adjective without resetting the activation levels of the units in the model to their default values.

Interestingly, subfamilies that are not included in the resonance set (and which were not available to the simulation model) nevertheless show up with high correlations with the model times. For instance, the nominal and verbal subfamilies in the case of the COMPARATIVE data set correlate quite good with model times (for both r_s is around $-.6$), although these model times were generated on the basis of the resonance with the adjectives only (general and scale-focusing). Therefore, these correlations between model times with subfamilies that were not included in the resonance set are spurious and must arise due to the intercorrelations of, in this case, nominal and verbal subfamilies on the one hand with the number of adjectives on the other hand. We suspect that the same holds for the observed correlations with the response latencies in our experiments. In fact, we propose that the model is a useful tool for ascertaining which correlations with the response latencies are driving the observed family size effects, and which correlations are mere statistical side-effects without independent explanatory value.

Table 4.5 lists the Spearman correlations and their associated p-values for the four subsets of data. The first two columns again present the correlations of the raw family counts with the response latencies. The second two columns represent the correlations of the family counts restricted to the subfamilies in the resonance sets and the response latencies. Note that the correlation for the VERY context, which is not significant given the raw family count, is significant given the appropriate subfamily (the effective family size, i.e., the resonance set as determined on the basis of the simulation model). The third two columns present the correlations of the model times with the response latencies, all of which are comparable to those

---

general adjectives are co-activated in the family. Although our results are compatible with a parsimoneous resonance set containing only general adjectives for the COMPARATIVE context, the results obtained for a larger resonance set with both general and scale-focusing adjectives are as good.

of the effective family size.

Table 4.5: Spearman correlations of Family Size and RT (left columns), of the Effective Family Size and RT (center columns), and of Model Times and RT (right columns) for four morphological and phrasal contexts.

Context	Family Size		Effective Family Size		Model	
	r_s	p	r_s	p	r_s	p
BASE	-.32	.043	-.32	.043	.38	.016
COMPARATIVE	-.39	.016	-.62	.000	.60	.000
VERY	-.20	.210	-.40	.012	.44	.004
NOT	-.36	.023	-.47	.004	.46	.003

General Discussion

This study addresses the question to what extent the morphological family size effect is modulated by its immediate morphological and phrasal context. Two visual lexical decision experiments revealed that indeed the context in which a word appears co-determines which morphological family members become co-activated. For a simplex Dutch adjective presented in isolation, all family members appear to contribute to the family size effect. When a simplex adjective is followed by the comparative suffix *-er*, the adjectival family members drive the effect. The same holds when a simplex adjective is preceded by the negation *niet*, ’not’. When preceded by *heel*, ’very’, only the color compounds and intensified adjectives (the scale-focusing adjectives) in the family are relevant.

Recall that Bertram et al. (2000) observed that the family size effect for abstract nouns with the Dutch suffix *heid* with respect to the adjectival family members was restricted to those adjectives to which *-heid* attaches, i.e., to what we have called ’general’ adjectives. The present study provides new independent support for the distinction between general and scale-focusing adjectives: In the VERY context, it is the scale-focusing adjectives that drive the family size effect, to the exclusion of the general adjectives. Note that in this case, we seem to be dealing with a form of synonymy, *very cold* and *icecold* being very similar in meaning.

To understand these data, we developed an interactive activation model, the Morphological Family Resonance Model. We regard this model as providing a reasonable functional characterization of the morpho-semantic architecture in the mental lexicon. We share with McRae, DeSa, & Seidenberg (1997; and see Halle & Marantz, 1993, for a linguistic view on distributed morphology) the assumption that word meanings are not discrete, monolithic entities (contrary to, e.g., Roelofs, 1992) and that meaning emerges as a pattern of activation across related entries within a lexical network. On the other hand, the simulations show that we do not need to postulate subsymbolic representations as in McRae et al. (1997). In fact, symbolic representations are perfectly adequate to model the functional properties of morphological resonance in a mathematically tractable and computationally simple manner. The present model can be seen as a first step towards the formalization of parts of the descriptive models proposed by Bybee (1985) and Schreuder & Baayen (1995).

The simulation studies with the MFRM revealed that the experimental results can be understood as resulting from activation spreading to restricted subsets of morphological family members, which we refer to as resonance sets. The model does not explain how resonance sets arise. However, given the resonance sets, the model provides excellent fits to the experimental results. Apparently, word-category information as well as phrasally supplied semantic information is exploited to zoom in on the appropriate subsets of family members.

In fact, the size of the resonance sets seems to be inversely proportional to the amount of information supplied by the context. If the context provides no information, as is the case when a simplex adjective is presented in isolation, all morphological family members contribute to the morphological family size effect. When the context provides information as to which word category is particularly relevant, only those family members sharing the relevant word category become co-activated. This is what we observe for the COMPARATIVE and NOT contexts. In the case of the VERY context, *heel* ('very') is a modifier which narrows down the general meaning of the adjective to a specific part of the scale. In this case, precisely those family members which express this, the scale-focusing adjectives, become co-activated. Thus, the context seems to narrow down the co-activation of morphological family members to those words whose meaning is contextually relevant.

Our results point to two important properties of the mental lexicon. First, the observed context-sensitivity of the family size effect, with resonance being restricted to sub-families, suggests a high degree of plasticity for the morpho-lexical networks in the mind. How this plasticity might be captured within the activation framework is a challenge for further research. Second, the resonance metaphor of the MFRM suggests that the percept of the meaning of a word in the mind depends not only on the activation of its own meaning, but also on the co-activated meanings of its family members.

References

Baayen, R. H., Lieber, R. and Schreuder, R.: 1997, The morphological complexity of simplex nouns, *Linguistics* **35**, 861–877.

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

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

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.

Halle, M. and Marantz, A.: 1993, Distributed morphology and the pieces of inflection, *in* K. Hale and S. J. Keyser (eds), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Vol. 24 of *Current Studies in Linguistics*, MIT Press, Cambridge, Mass, pp. 111–176.

Hasher, L. and Zacks, R. T.: 1984, Automatic processing of fundamental information. The case of frequency of occurrence, *American Psychologist* **39**, 1372–1388.

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

McRae, K., DeSa, V. and Seidenberg, M. S.: 1997, On the nature and scope of featural representations of word meaning, *Journal of Experimental Psychology:General* **126**, 99–130.

Roelofs, A.: 1992, A spreading activation theory of lemma retrieval in speaking, *Cognition* **42**, 107–142.

Rubenstein, H. and Pollack, I.: 1963, Word predictability and intelligibility, *Journal of Verbal Learning and Verbal Behavior* **2**, 147–158.

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

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

Whaley, C. P.: 1978, Word–nonword classification time, *Journal of Verbal Language and Verbal Behavior* **17**, 143–154.

Appendix

Words with a high family size with in parentheses the mean reaction times for the BASE, COMPARATIVE, HEEL, and NIET condition respectively.

blauw, ’blue’, (519 648 541 679); bol, ’round’, (587 550 622 646); fijn, ’fine’, (574 574 597 572); groen, ’green’, (534 617 614 666); kort, ’short’, (592 583 623 616); los, ’loose’, (547 570 701 558); mobiel, ’mobile’, (542 617 605 643); net, ’neat’, (590 636 599 748); plat, ’flat’, (511 626 600 722); rijk, ’rich’, (571 594 562 566); rot, ’rotten’, (563 692 593 669); schaars, ’scarce’, (565 716 626 642); spits, ’pointed’, (607 649 627 813); veilig, ’safe’, (564 597 578 593); vet, ’fat’, (570 695 605 573); vlak, ’flat’, (531 651 582 641); vol, ’full’, (564 555 575 578); vuil, ’dirty’, (519 612 558 608); zout, ’salty’, (600 588 603 586); zwak, ’weak’, (578 568 591 624).

Words with a low family size with in parentheses the mean reaction times for the BASE, COMPARATIVE, HEEL, and NIET condition respectively.

bang, ’scared’, (564 557 573 562); fel, ’fierce’, (598 626 598 658); flink, ’robust’, (611 658 592 636); gammel, ’ricket’, (655 823 712 717); gauw, ’quick’, (576 895 647 662); gering, ’petty’, (694 727 785 711); ijdel, ’vain’, (569 664 602 651); jaloers, ’jealous’, (541 641 568 618); juist, ’just’, (555 606 554 616); lauw, ’tepid’, (648 640 689 648); leuk, ’nice’, (558 591 527 620); mild, ’mild’, (543 625 609 623); modern, ’modern’, (645 686 578 726); mooi, ’beautiful’, (584 511 542 554); nors, ’grumpy’, (609 604 675 655); schril, ’shrill’, (613 829 695 859); schuin, ’slanting’, (645 633 760 683); simpel, ’simple’, (586 631 541 646); steil, ’steep’, (551 630 615 717); trots, ’proud’, (565 570 590 618).

Homonyms in context

CHAPTER 5

This chapter is submitted as Nivja H. De Jong, Robert Schreuder, and R. Harald Baayen: Local and broad activation of different meanings of an ambiguous word: Effects of word frequency and family size.

Abstract

This study makes use of frequency counts and family size counts (the number of morphologically related words) of the different meanings of Dutch ambiguous words as diagnostic tools to investigate the processing of such words in different contexts. For a particular meaning of an ambiguous word, we take an effect of its frequency as evidence for local activation of that meaning, and an effect of its family size as evidence for a more extensive, broad activation. The tasks we employed are subjective frequency rating and visual lexical decision. In Experiment 5.1, we presented noun-noun homonyms. Without context, both meanings were activated, but using related words as primes proved to be problematic. In Experiment 5.2, we presented noun-verb homonyms, enabling us to use list manipulation and minimal syntactic context for disambiguation. The effects of both frequency and family size were sensitive to these kinds of disambiguation. We interpret these effects as context mediating local and broad activation of each meaning and hypothesize that a single mechanism underlies both.

Introduction

A well-known finding in language comprehension is that high-frequency words are recognized faster and more accurately than low-frequency words (e.g., Taft, 1979). Less well-known is a similar facilitating effect of the morphological family size. In English, the word *man* has the largest family with 270 descendents, including *fireman* and *salesmanship* (We use the CELEX lexical database (Baayen, Piepenbrock, & Gulikers 1995) to count the number of morphological related words.) Recent studies have shown that words which occur as constituents in many other words are responded to faster in lexical decision and rated higher in subjective frequency rating than words that occur in only few other words (for Dutch see, e.g., Schreuder & Baayen, 1997; for English see De Jong, Feldman, Schreuder, Pastizzo, & Baayen, 2002, also Chapter 3; and for German see Lüdeling & De Jong, 2002).

Many models of visual processing locate the effects of word frequency at the access level (e.g., Baayen, Dijkstra, & Schreuder, 1997; Bradley & Forster, 1987; Morton, 1969). The effect of family size, however, clearly reflects deeper, semantic levels of processing. De Jong, Schreuder, and Baayen (2000, also Chapter 2) report that the effect of family size is not strictly mediated by form. For instance, an irregular past participle such as *gevochten*, ’fought’, is derived from the verb *vechten*, ’to fight’, and its morphological family members use the form -*vecht*-. Even though the presented words were not formally related to their morphological family, De Jong et al. (2000, also Chapter 2) report that these irregular past participles nevertheless show an effect of family size. Bertram, Baayen, and Schreuder (2000) point out that removing semantically opaque family members from the count of the family size improves correlations with reaction times. Furthermore, De Jong et al. (2000, also Chapter 2) and De Jong, Schreuder, and Baayen (to appear, also Chapter 4) show that the effect of family size is context-sensitive. Different sub-families of unambiguous words contribute to the effect depending on the morphological and immediate sentential context in which the word appears.

In this paper, we study semantically ambiguous words both in isolation, as well as in a disambiguating context, using not only frequency of occurrence but also family size as diagnostics for tracing the processing of their meanings. We count the frequencies of the different meanings and their different families with the help of corpora. The tasks that we have used are subjective frequency rating and lexical decision. The reason we use ratings in addition to the lexical decision tasks is, first, because they are similarly sensitive to a complex of factors (including a word’s family size) and not only to the frequency of the word. Second, the results for lexical decision and subjective frequency ratings may differ (Baayen et al., 1997), as we will also see in this study. Word frequency effects in both tasks can show local activation of the different meanings of a homonym, whereas an effect of family size may show how semantic activation of the different meanings of a homonym spreads in the mental lexicon to those morphologically related words that share the relevant meaning of the ambiguous word. We will refer to this more extensive activation of meaning as broad activation.

This study is carried out against the background of two debates in research on word processing. The first is on ambiguity resolution, and the second on the locus of the effect of word-frequency. We will first discuss some of the literature relevant for both issues (see Twilley & Dixon, 2000, for a recent overview on the ambiguity literature), and then proceed by introducing the present study.

Many studies on ambiguous words have addressed the question whether primary lexical access is autonomous or selective. Theories propagating the autonomous view stress the fact that many cross-modal priming studies in which ambiguous words are followed by a target related to either of their meanings show an effect of priming compared to an unrelated control word if the ISI is very short (Lucas 1987; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney 1979). The preceding context of the homonym does not affect these priming results: Even if the preceding context disambiguates the homonym, targets related to both meanings are primed. With a longer ISI (approximately 200 ms.), only the appropriate meaning is activated. The word related to the inappropriate meaning of the homonym no longer shows an effect of priming compared to unrelated control words. These studies suggest that primary access to ambiguous words is unaffected by preceding context. They conclude that only after multiple meanings have been accessed does the context begin to influence which meaning is ultimately selected. Others, however, claim that sentence context is able to direct immediate lexical access to one particular meaning. Various experimental paradigms have been used to support this claim: cross-modal priming (Tabossi & Zardon, 1993), visual-visual priming (Simpson & Krueger, 1991), a modified Stroop-task (Paul, Kellas, Martin, & Clark, 1992), and eye-movement studies (e.g., Duffy, Morris, & Rayner, 1988; Rayner & Frazier, 1989; Rayner & Morris, 1991).

Glucksberg, Kreuz, and Rho (1986), and Lucas (1987) point out that, possibly, the priming studies showing multiple meaning activation for ambiguous words do not in fact reveal an effect of autonomous lexical access. Instead, especially when very short ISI’s are used, backward (target to prime) associates may play a role.

When priming is used as a measure of meaning activation, the inappropriate meaning may well not be activated by the ambiguous word itself but by backwards priming from the target to the ambiguous prime. In this reasoning, the inappropriate sense of the ambiguous word is activated through semantic association with the target (after which this activation of the inappropriate meaning in its turn facilitates the processing of the target). To test this, Glucksberg et al. (1986) use pseudowords as targets, which should only serve as forward, but not as backward primes. They use interference to pseudowords resembling words related to ambiguous words as a measure of meaning activation (e.g., interference to *piamoe* or *kidnea* when the sentence contains *organ*). They find that interference only occurs for the pseudoword targets which resemble words related to the context-appropriate meaning of the ambiguous words. This would be evidence that in studies using related words as targets, facilitation for target words related to the inappropriate meaning of the ambiguous words indeed occurs through backwards priming. But multiple access is still a reasonable explanation in a task measuring interference for pseudowords, as Burgess, Tanenhaus, and Seidenberg (1989) show that this task is not susceptible to lexical priming, but rather to sentential context. Support for the hypothesis that with short ISI's backwards priming may play a role, on the other hand, comes from Van Petten and Kutas (1987). They used a cross-modal priming paradigm with preceding disambiguating context while measuring event-related potentials (ERP) and investigated the N400. The N400 is the electrical potential that normally occurs around 400 ms after the onset of a stimulus, and presumably mainly reflects the ease of its semantic integration (see, e.g., Chwilla, Brown, & Hagoort, 1995; and Holcomb, 1993). Similar to the other studies using cross-modal priming, Van Petten and Kutas (1987) also observed that both meanings of a homonym are activated with a short ISI: Words related to both the appropriate and inappropriate meaning are responded to faster than unrelated controls. However, the ERP data show that the onset of the N400 response was 200 ms earlier for targets related to the appropriate meaning than for the targets related to the inappropriate meaning of the ambiguous word. This suggests that, indeed, backwards priming might have played a role, in that the inappropriate meaning of the ambiguous word was activated only after at least some of the processing of the target took place. Besides the problem of backwards priming, another difficulty in the methodology of priming in these studies arises. Instead of tracking the lexical processes instantiated by the homonym itself, an experiment with such a methodology studies the lexical access of the different meanings of a homonym indirectly via the response times to related targets.

Paul et al. (1992) show that the salience of the disambiguating context is crucial in determining the amount of priming for the appropriate meaning. Rayner and Frazier (1989), Simpson and Krueger (1991), and Tabossi and Zardon (1993) show that the relative frequencies of the two meanings of the ambiguous word also determine whether lexical access is exhaustive or selective. For ambiguous words with the context directing to the dominant meaning, selective access seems to take place. When the context is directed to the subordinate meaning, however, both meanings can be activated. This is known as the subordinate bias effect, as in eye-movement paradigms gaze durations on homonyms are longer than on control words only when preceding context calls for the subordinate meaning of the homonyms. This suggests that despite the biasing context for the subordinate meaning, the dominant meaning was activated as well, leading to a time-costly competition between the dominant and subordinate meaning (e.g., Binder & Rayner, 1998; Duffy, Morris, & Rayner, 1988; Kambe, Rayner, & Duffy 2001; Rayner, Binder, & Duffy, 1999; and Rayner, Pacht, & Duffy, 1994). Others have questioned this view and reported that the subordinate bias effect can be modulated by strength of context, even to the extent that it can disappear (e.g., Martin, Vu, Kellas, & Metcalf, 1999; Vu, Kellas, Metcalf, & Herman 2000; and Vu, Kellas, & Paul, 1998).

Summing up, depending on the salience of the context and the relative frequency of the two meanings of an ambiguous word, the context can selectively direct primary lexical access, at least to the dominant meaning of an ambiguous word. Interestingly, the role of the relative frequency of the two meanings of a homonym suggests that some measure of word frequency might reflect more than mere string familiarity, supposedly effective only at the level of access in word recognition. In fact, it suggests that some aspect of word frequency is semantic in nature. Balota and Chumbley (1984) initiated the discussion about the locus of the word-frequency effect by claiming that the effect of word frequency in visual lexical decision might well be exaggerated due to post-access effects in the decision stage. But Monsell, Doyle, and Haggard (1989) showed that the effect of word frequency across tasks requiring lexical access was remarkably similar and that therefore its (major) locus should be at the access level. Balota (1990) argues that a ”magic moment” (i.e., lexical access without meaning) does not exist and Balota and Chumbley (1990) reply to Monsell et al. (1989) by pointing out that the parallel distributed processing framework presented by Monsell et al. (1989) also seems to be sensitive to effects of word frequency at several levels. Hino and Lupker (1996) argue that the effects of word frequency (as well as polysemy) do not play a role in lexical access and likewise propose an explanation in the terms of a parallel distributed network account.

In addition, several studies report a context by word frequency interaction, in that semantic context aids low-frequency words more than high-frequency words both when using sentences as primes (e.g., Stanovich & West 1981) as well as single word priming (Becker, 1979; Borowsky & Besner, 1993; Plaut & Booth, 2000). Such an interaction shows that the effect of word-frequency is modulated by semantic context, suggesting that it cannot be an effect operating on the word-form level solely (see also Plourde & Besner, 1997). Borowsky and Besner (1993), in accounting for the interacting effects of context and word frequency, propose a multi-stage activation model in which word frequency affects the mapping between orthographic input lexicon and the semantic system. Plaut and Booth (2000) present a single-mechanism distributed network account which also can accomodate the interactive effect of word frequency and context. In their model, the effect of word frequency is not located at the level of lexical access (without meaning), as the predictions of reaction times of the model are based on the amount of activation (”stress”) in the semantic system. On the other hand, Andrews and Heathcote (2001) provide distributional analyses of reaction time data, showing that differences in the effect of word frequency across several tasks could still be accounted for by models that postulate a common frequency-sensitive identification process.

Finally, Rudell (1999) and Rudell and Hu (2001) use the recognition potential (RP), an electrical response of the brain that occurs when a subject is consciously aware of a recognizable image, to study high- and low-frequency words. Words were presented at 800 ms intervals, without a task other than reading the words, and the mean RP latency was 266 for the high-, and 292 for the low-frequency words (Rudell, 1999). A very similar difference in latencies of the RP was found when a lexical decision was required on these high- and low-frequency words (Rudell & Hu, 2001). This effect of word-frequency on the RP shows that, whatever the locus may be, it must at least partially be relatively early and task-independent.

As mentioned before, another reason to assume that at least part of the effect of word-frequency is located on either the semantic level or the interaction between access and semantic levels, comes from the literature on ambiguous words in which the relative frequencies of a homonym have been found to play a role. So far, the role of frequency of these ambiguous words has only been used as a binary value (dominant versus subordinate). But the effect of frequency is a graded effect and therefore taking into account the different *absolute* frequencies of ambiguous words can provide additional insight into the processing of the different meanings of these words. Therefore, in this paper, we investigate the effect of the absolute frequency counts and family sizes of the two meanings of an ambiguous word on the comprehension of these ambiguous words themselves. To measure the effects of word frequency and family size, we will carry out correlational analyses on rating and reaction time data. We use these effects of frequency and family size as diagnostics for determining how the context (and the absence of a context) in which an ambiguous word is presented might affect lexical processing.

We take an effect of (absolute) word-frequency of one of the meanings of a homonym as evidence for the local activation of that particular meaning of the homonym. If the correlations for the word frequencies of both meanings are significant, we take this as evidence for the local activation of both word meanings. An effect of family size for one particular meaning may provide additional insight into whether semantic activation of that meaning has spread to its associated morphological family members. In other words, we can use the family size effect to ascertain whether local semantic activation has lead to broad activation of a specific meaning.

In the first experiment, we use noun-noun homonyms. We present these homonyms in a subjective frequency rating and in a lexical decision task without disambiguating context. As in Deloche, Seron, Scius, and Segui (1987), and as in Gottlob, Goldinger, Stone, and Van Orden (1999; their Experiment 3, using an association judgement task), we also presented homonyms as targets and related words as primes. We used the homonyms as targets in order to measure the effects of frequency and family size of both meanings of these noun-noun homonyms themselves, when disambiguating context (the primes) was provided. In Experiment 5.2, we turn to noun-verb homonyms, which we show can be disambiguated using list manipulation in both on-line lexical decision as well as in subjective frequency rating. In addition, we use minimal syntactic information for disambiguation. Instead of the somewhat meta-linguistic syntactic function decision task as used in Forster and Bednall (1976) in the visual modality (see also Cutler & Clifton, 1984, and Deutsch & Wijnen, 1985, for similar tasks using unambiguous words in the auditory modality), we used a lexical decision task. We shall see that besides the differential effects of word-frequency of the two meanings of an ambiguous word, the differential effects of family size increase our understanding of the lexical processing of ambiguous words.

Experiment 5.1: Noun-noun homonyms

In the first experiment, we investigate the role of frequency and family size of the two meanings of noun-noun homonyms. For the same set of words, we make use of a subjective frequency rating without disambiguating context, a subjective frequency rating in which the homonym was presented together with a word related to either of the meanings, and two lexical decision conditions. In the first lexical decision task, no disambiguating context is provided, while the second makes use of unrelated primes or primes related to both of the meanings.

Method

Participants. A total of 117 undergraduate students were paid to take part in this experiment. Eighteen students performed the subjective frequency rating condition without disambiguating context, and 18 performed the lexical decision condition without disambiguating context. In the subjective frequency rating condition with disambiguating context, 36 undergraduate students participated. Forty-five undergraduate students participated in the lexical decision condition with primes. All participants were students at the University of Nijmegen and were native speakers of Dutch.

Materials. We selected 37 Dutch noun-noun homonyms from the CELEX lexical database which had two clear distinct meanings such as *vorst* which can mean either ’monarch’ or ’frost’. For 24 of these pairs, which differ syntactically in some way, CELEX provides surface and base frequency estimates for the different meanings based on the co-occurrence frequencies with a gender-marked definite article.[1] The remaining pairs of homonyms do not differ syntactically and CELEX provides only the total surface and base frequencies for both meanings jointly. For these pairs of homonyms, we used a corpus of roughly 4.5 million words of Dutch newspaper texts. For every instance of these homonyms in this corpus (but with a maximum of 500 per homonym), we determined which meaning was intended in

---

[1] The surface frequency of a meaning of a homonym is the frequency of the singular form of the homonym in that particular meaning. The base frequency of a meaning of a homonym is the lemma frequency of the homonym in that particular meaning, i.e. the summed frequencies of all inflectional variants of the homonym in that particular meaning.

order to estimate the individual probabilities of both meanings. We then partitioned the total surface and base frequency listed in CELEX for the two meanings of each homonym proportionally to these disambiguated counts. For all homonyms, we calculated the family size of the two meanings, attributing each morphological family member to one of the meanings of a given homonym. In this way, we obtained 37 noun-noun homonyms with separate surface frequencies, base frequencies, and family size counts for the two individual meanings.[2]

For both meanings of each homonym, we selected as many strongly related words as possible (up to 5 for each meaning). We constructed two lists with all the homonyms together with the strongly related words for one of the meanings per list. We asked 28 undergraduates from Nijmegen University to rate for relatedness on a seven-point scale (14 participants in each list, none of whom participated in the actual experiment). One group of participants rated, for instance, the relatedness of *vorst* (meaning ’frost’ or ’monarch’) with *winter*, ’winter’, *vriezen*, ’freeze’ and *kou*, ’cold’. The other group of participants rated the relatedness of *vorst* with *regering*, ’parliament’, *koning*, ’king’, and *tiran*, ’tyrant’. We then selected the best pairs of related words for each homonym such that each homonym was paired with two equally related words of both meanings. For instance, the words *vriezen* and *koning* were selected as related words for the homonym *vorst*, as the average rating for relatedness with *vorst* was 6.4 and 6.6 respectively. The mean relatedness for the first list was 5.0, and 5.2 for the second list.

We then assigned the meanings of all homonyms to two new lists, such that the mean surface frequency, the mean base frequency and the mean family size for one meaning was larger (surface frequency 36.0, base frequency 41.1 and family size 29, token frequency counts standardized per million) than for the other (surface frequency 4.0, base frequency 10.7 and family size 8). We will refer to these two meanings as the dominant meaning and the subordinate meaning. The correlation between the base frequencies of the two meanings was not significant ($r = .266; t(35) = 1.63, p = .112$) and the same holds for the correlation between the two family size counts of each homonym ($r = .131; t < 1$). This will allow us to look at the contribution of the base frequency and family size of each meaning of a homonym separately, without having to worry about possible statistical side-effects due to collinearity. The correlation between the two surface frequencies, on the other hand, was significant ($r = .488, t(35) = 3.31, p = .002$).

---

[2] Nine of these noun-noun homonyms could also occur as verbal form with a mean frequency of 7.0. In most cases (all but one, which had a frequency of 0.1), however, the meaning of this verbal form was very close to the meaning as a noun.

We added 58 noun fillers to the experimental list of the subjective frequency rating. Half of the fillers were very low in base frequency (under 2), and half of them were very high (over 250) in order to have the ratings for the homonyms to be estimated somewhere around the middle of the scale. For the subjective frequency rating with disambiguating context, we made two experimental lists and presented the homonyms with one of the two strongly related words without adding filler words. For the lexical decision task without primes, we added 78 noun fillers to the experimental list and paired each word with a pseudoword with legal phonotactics. The mean base frequency of these filler words was 18.4. For the lexical decision task with primes, we made three experimental lists. In each list, a homonym was primed by either a word related to the dominant meaning, a word related to the subordinate meaning, or by an unrelated prime. As 37 cannot be divided by three, in each list twelve target words of two conditions, and thirteen of a third condition appeared. We added 35 noun fillers to all three experimental lists (mean base frequency 48.0). These noun fillers were preceded by unrelated primes. The total of 72 words was paired with an equal number of pseudowords with legal phonotactics. These pseudowords were preceded by word primes. In this way, the ratio of related primes in each experimental list was about one third of the word targets (and one to six taking into account the prime-pseudoword pairs).

Procedure. For the subjective frequency rating, participants were asked to indicate on a seven-point scale how frequently they thought a word was used in current Dutch. For the subjective frequency rating with disambiguating context, participants were asked to indicate how frequently they thought a particular meaning of a word was used in current Dutch by presenting the homonym with a strongly related word of either the dominant or the subordinate meaning and asking the participant to rate the frequency of the word with respect to this particular meaning.

For the lexical decision task, participants were tested in noise-proof experimental rooms. They were asked to decide as quickly and accurately as possible whether a letter string appearing on the computer screen was a real Dutch word. Each stimulus was preceded by a fixation mark in the middle of the screen for 50 ms. After 500 ms, the stimulus appeared at the same position. Stimuli were presented on Nec Multisync color monitors in white lowercase 36 point Helvetica letters on a dark background and they remained on the screen for 1500 ms. The maximum time span allowed for a response was 2000 ms from stimulus onset. For the lexical decision task with primes, each stimulus was similarly preceded by a fixation mark in the middle of the screen for 50 ms. After 500 ms, the prime appeared at the same position for 100 ms. After another 50 ms, the target word appeared at the same position. We used this SOA of 150 ms., as facilitation from semantically related primes should occur, whereas inhibition from unrelated primes presumably would be minimal (see Neely, 1991). The targets remained on the screen for 1500 ms and the maximum time span allowed for a response was 2000 ms from target onset. Participants were asked to decide as quickly and accurately as possible whether the second letter string appearing on the computer screen was a real Dutch word.

Results and Discussion

First we will report the overall results of the different experimental conditions, then we will continue with correlational analyses, comparing the correlational results for the three variables surface frequency, base frequency and family size of the two different meanings across conditions.

All participants performed the lexical decision and the priming lexical decision with an error rate below 8%. Two items in the lexical decision priming task (one from the unrelated priming condition and one from the condition in which the dominant meaning was primed) had reaction times exceeding 3 standard deviations from the mean reaction time. We excluded these data points from further analyses. Table 5.1 shows the means and standard deviations of the ratings and response times for all experimental conditions.

The targets presented with words related to the dominant meaning were rated significantly higher than when they were presented with words related to the subordinate meaning (mean 4.5 and 4.0 respectively: $t1(34) = -3.15, p = .003; t2(72) = -2.28, p = .026$). The reaction times for the targets following both kinds of primes were faster than the targets following unrelated primes. For primes related to the dominant meaning: $t1(44) = 6.08, p = .000; t2(70) = 3.88, p = .000$. For primes related to the subordinate meaning: $t1(44) = 5.17, p = .000; t2(71) = 2.31, p = .024$. Contrary to the rating study with disambiguating context, there was no significant difference in reaction times between the two related prime conditions (mean 521 and 529 respectively: $t1(44) = 1.47, p = .149; t2(71) = 1.11, p = .271$).

These results do not confirm the subordinate bias effect (e.g., Binder & Rayner, 1998; Kambe, Rayner, & Duffy 2001; Rayner, Binder, & Duffy, 1999; and Rayner, Pacht, & Duffy, 1994; but see, e.g., Martin, Vu, Kellas, & Metcalf, 1999; Vu, Kellas, Metcalf, & Herman 2000; and Vu, Kellas, & Paul, 1998). Evidently, our paradigm was very different from the ones used in the above-mentioned papers studying the subordinate bias effect. But the paradigm in the study by Gottlob et al. (1999, their

Table 5.1: Results of Experiment 5.1: Means and standard deviations of ratings, response latencies and error proportions (by participants).

Experimental condition	RT/rating (*SD*)	error (*SD*)
Simple rating	4.3 (1.0)	
Simple lexical decision	581 (36)	.02 (.03)
Lexical decision (unrelated prime)	560 (71)	.04 (.06)
Rating with context favoring dominant meaning	4.5 (.9)	
Lexical decision (dominant meaning prime)	521 (57)	.03 (.04)
Rating with context favoring subordinate meaning	4.0 (1.2)	
Lexical decision (subordinate meaning prime)	529 (63)	.02 (.04)

Experiment 3) and Deloche et al. (1987) is more similar. However, Deloche et al. (1987) examined the interaction between visual field presentation and imageability of the high or low imageable meaning of a homonym, while matching on meaning frequency. The study by Gottlob et al. (1999) is more comparable to the present experiment, as they studied the amount of facilitation for dominant and subordinate primed meanings of homonyms. Using an association judgement task, they also found equal facilitation for targets following a prime related to the dominant or subordinate meaning of the homonym.

In order to ascertain what factors might have played a role during the processing of these homonyms across the different conditions, we now turn to the correlational analyses. We report one-tailed t-tests, as we expect all factors to be facilitatory (a positive correlation coefficient for the rating studies for all variables and a negative correlation coefficient for the lexical decision studies). As we will perform two correlations on each variable per condition, we apply Bonferroni adjustments and set our alpha level to $.025$. We only report correlations which are significant with respect to this alpha-level. In addition, we report stepwise linear regression analyses. It does not make sense to carry out linear regression models entering the six variables jointly, as there was pervasive collinearity in the data matrix, with a very high condition number, 43.57 (Belsley, Kuh, and Welsch, 1980). But for for the dominant and subordinate base frequency, as well as for the two family size variables, there is no collinearity. This allows us to compare the contribution of these variables with linear regression models.

Figure 5.1 provides a graphical summary of the correlational results. The y-axes represent the values of the Pearson correlation coefficient. The x-axes represent the variables surface frequency (Surf), base frequency (Base) and family size (Fam) for both the dominant (D, dark bars) and subordinate (S, light bars) meanings of the homonyms. The first column of panels shows the correlational results for the three different subjective frequency rating conditions (no context, presented with disambiguating primes favoring the dominant meaning, and with disambiguating primes favoring the subordinate meaning). The panel in the second column represents the correlational results for the lexical decision condition without primes, and the third column of panels shows the correlational results for the lexical decision conditions with primes (unrelated primes, primes related to the dominant meaning, and primes related to the subordinate meaning). The correlations which are marked with an asterisk are significant at the $\alpha = .025$ level AND are retained as explanatory variables in a stepwise linear regression analysis in which both the dominant and subordinate variants of a variable entered.

Consider the first row of Figure 5.1, which summarizes the results for the conditions in which no related primes were presented and focus on the subjective frequency rating (The first panel of the first row). Both the surface frequency of the dominant and the surface frequency of the subordinate meaning correlate significantly with the ratings ($r = .586, t(35) = 4.27, p = .000$ and $r = .437, t(35) = 2.88, p = .003$). Similarly, both base frequencies correlate significantly with the ratings ($r = .641; t(35) = 4.94, p = .000$ and $r = .403; t(35) = 2.61, p = .007$), and both variables are retained in a stepwise linear regression model ($t(34) = 4.43, p = .000$ and $t(34) = 1.93, p = .031$). Both the dominant and subordinate family size counts also correlate significantly with the ratings ($r = .485; t(35) = 3.28, p = .001$ and $r = .347, t(35) = 2.19, p = .018$), and both were retained in a stepwise linear regression analysis ($t(34) = 3.13, p = .002$ and $t(34) = 2.01, p = .026$).

Turning to the lexical decision data without primes (the second panel of the first row), we see that all correlations are negative and reduced somewhat in absolute value, but the pattern remains very similar. Both the dominant and subordinate surface frequency correlate significantly with the reaction times ($r = -.462, t(35) = -3.08, p = .002$ and $r = -.357, t(35) = -2.26, p = .002$) and both base frequencies correlate significantly with the reaction times ($r = -.452, t(35) = -3.00, p = .003$ and $r = -.332, t(35) = -2.08, p = .022$), which both are retained in a stepwise linear regression model, although the coefficient for the subordinate base frequency is not significant ($t(34) = -2.55, p = .008$ and $t(34) = -1.48, p = .074$). Both family size counts correlate significantly with the reaction times ($r = -.330, t(35) = -2.07, p = .023$ and $r = -.422, t(35) = -2.75, p = .005$), and both are retained in a stepwise

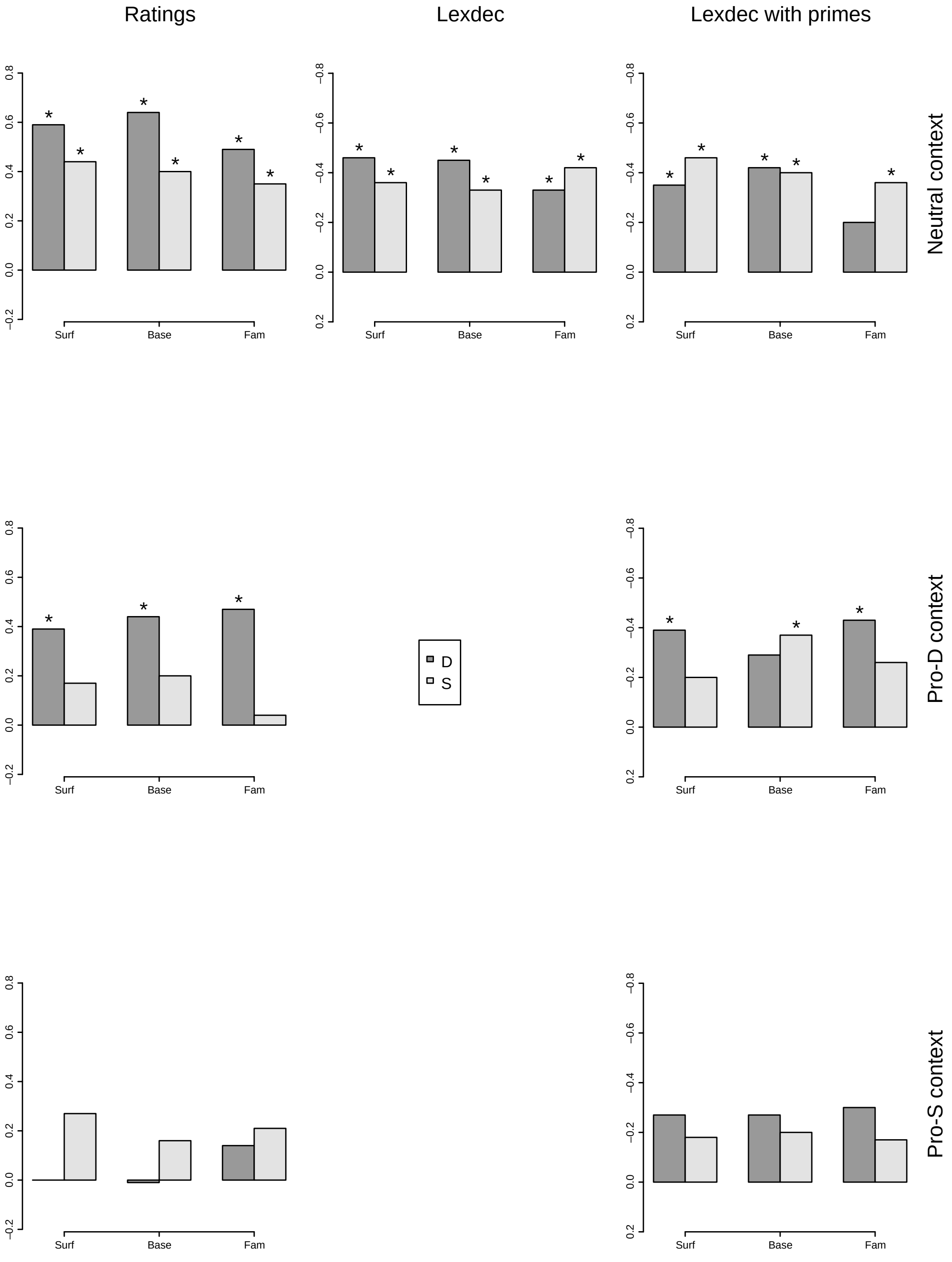

Figure 5.1: Values (r) of Pearson correlations of ratings (first column of panels) and reaction times (the remaining panels) with the dominant (D, dark bars) and subordinate (S, light bars) surface frequencies (Surf), base frequencies (Base), and family size counts (Fam) of noun-noun homonyms in Experiment 5.1 in different contexts. Asterisks mark the significant correlations ($p < 0.025$).

linear regression model ($t(34) = -1.87, p = .035$ and $t(34) = -2.58, p = .005$).

The pattern of correlations of the lexical decision data with unrelated primes (the third panel of the first row) is somewhat different. As in the condition without primes, the surface frequency counts of both the dominant and the subordinate meaning are significant ($r = -.348, t(34) = -2.17, p = .019$ and $r = -.460, t(34) = -3.02, p = .002$) and both base frequency counts correlate significantly with the reaction times ($r = -.423, t(34) = -2.72, p = .005$ and $r = -.396, t(34) = -2.51, p = .008$), which are both retained in a stepwise linear regression model ($t(33) = -2.19, p = .018$ and $t(33) = -1.94, p = .030$). But for the family size correlations, we observe that only the count for the subordinate meaning correlates significantly with reaction times ($r = -.357, t(34) = -2.23, p = .016$), and only the subordinate family size count is retained in the stepwise linear regression model. We can nevertheless conclude that in the subjective frequency rating as well as in both lexical decision tasks, when no related prime is presented, neither of the two meanings seems to be exclusively favored over the other.

The second row represents the results for the conditions in which the dominant meaning was primed. For the subjective frequency rating with primes related to the dominant meaning (the first panel in the second row), we find that only the dominant surface frequency, base frequency, and family size emerge as significant explanatory variables. Comparing the two surface frequency counts, only the dominant surface frequency correlates significantly with reaction times ($r = .386, t(35) = 2.48, p = .009$). Similarly, only the dominant base frequency correlates significantly with reaction times ($r = .44; t(35) = 2.91, p = .003$) and is retained in a stepwise linear regression model. Finally, comparing the two family size counts, only the dominant family size correlates significantly with reaction times ($r = .467, t(35) = 3.13, p = .002$) and is retained in the stepwise linear regression model.

In the lexical decision data with primes related to the dominant meaning (the last panel of the second row), we similarly find that the surface frequency and family size count of only the dominant meaning are significant explanatory variables. Comparing the dominant and subordinate surface frequency, only the dominant surface frequency correlated significantly with reaction times ($r = -.392, t(34) = -2.48, p = .009$). For the family size counts, only the dominant family size correlated significantly with reaction times ($r = -.433, t(34) = -2.80, p = .004$) and only the dominant family size was retained in the stepwise linear regression model. For the base frequencies, however, unlike the correlational results of the rating study with primes related to the dominant meaning, it is the count of base frequency of the subordinate meaning rather than that of the dominant meaning that correlates significantly with reaction times ($r = -.372, t(34) = -2.34, p = .013$), and that is retained in a stepwise linear regression model.

Turning to the data in which the primes were related to the subordinate meaning, plotted in the third column of Figure 5.1, we see no reliable correlations whatsoever, neither in the subjective frequency rating (the first panel in the third row), nor in the lexical decision with primes related to the subordinate meaning (the last panel in the third row) and the stepwise linear regression models confirmed the absence of significant predictors.

Summing up, we find that if no context is provided, both meanings are activated, as frequency counts and family size counts of both the dominant and subordinate meaning emerge as reliable explanatory variables. But in the rating study with a prime related to the dominant meaning, we observe that only the appropriate surface frequency, base frequency, and family size, those pertaining to the dominant meaning, are reliable predictors. In the lexical decision study with primes related to the dominant meaning, we do not see such a clear pattern of results. The dominant surface frequency and the dominant family size play a role, but at the same time the subordinate base frequency is a reliable explanatory variable, whereas the dominant base frequency appears to play no role. When the prime is related to the subordinate meaning, no reliable correlations emerge, not in the rating task nor in the lexical decision task.

The absence of frequency and family size effects for the meaning of a homonym apparently do not mean that this meaning is not activated, as facilitation in the lexical decision task with primes was obtained for both dominant and subordinate meanings. For unambiguous words, several studies have reported that semantic context and word frequency may interact (for sentences as primes see, e.g., Stanovich & West, 1981; for single word primes see Becker, 1979; Borowsky & Besner, 1993; Plaut & Booth, 2000). In these studies it is found that context aids low-frequency words more than high-frequency words. In other words, the effect of frequency may disappear if sufficient semantic context is available. This might be an explanation for the absence of a frequency effect for the subordinate meaning of the homonyms in the present study.

However, if this interaction played a role in our study, we would expect to find no effects of frequency for the homonyms when primed by a word related to the dominant meaning as well. However, as mentioned above, we do find effects of frequency for both the dominant and subordinate meaning in lexical decision when the dominant meaning was primed. This might suggest that the single related word did not suffice to fully disambiguate the target words. Which leaves us with an improbable conclusion: although the primes were closely matched in relatedness, the prime related to the subordinate meaning would fully disambiguate the homonym (as evidenced by the absence of any effect of frequency), whereas the prime related to the dominant meaning of the homonym would not do so (as evidenced by the effects of frequencies for both the dominant and subordinate meanings).

A potential confound in our experiment are the mean word frequencies of the primes. Giraudo and Grainger (2000) show that the amount of morphological priming in a masked priming paradigm is affected by the word frequencies of the primes: high frequency primes produce more priming than low-frequency primes. We could not control for the mean word frequencies of the primes related to the dominant meaning and primes related to the subordinate meaning (mean base frequency 132.9 and 63.1 respectively), while matching closely for relatedness. Although we did neither use morphological primes, nor the masked priming paradigm (as Giraudo & Grainger, 2000, did), it is possible that the frequencies of the primes have affected the amount of priming in our study in some way. At the moment, we conclude that in a priming paradigm, it is unclear what factors may have influenced the processing of a target word right after processing a related prime.

In the next experiment, we again look at the role of context on the processing of homonyms. This time, however, we study homonyms which change word category (noun-verb homonyms), such as *last*, which can either mean ’burden’ or ’(you/he/she) welds’. This will allow us to avoid the priming paradigm and use minimal contexts. These minimal contexts fully disambiguate the homonyms without providing too precise semantic information that might give rise to a context by frequency interaction, and without using semantic primes which we would have to control for relatedness as well as frequency. Through all experimental conditions we use the same set of noun-verb homonyms and investigate the role of the surface frequencies, base frequencies, and family size counts of the nominal and verbal meanings. We use an experimental ’mixed’ list by adding the same number of unambiguous noun and verb filler words and use a subjective frequency rating as well as a lexical decision task. For two biased experimental lists (a ’nominal’ list with only noun filler words and a ’verbal’ list with only verb filler words) we will also report a subjective frequency rating and a lexical decision task. Finally, we use an immediate syntactic context for disambiguation (as in Forster & Bednall, 1976) in a lexical decision task, presenting the homonym with either a definite article or a personal pronoun.

Experiment 5.2: Noun-verb homonyms

Method

Participants. One-hundred forty-seven undergraduate students were paid to take part in this experiment. In the subjective frequency rating condition with the mixed list, 18 students participated, and 23 in the lexical decision condition with the mixed list. In each of the two subjective frequency ratings in which all the fillers were verbs or all the fillers were nouns, 18 participants took part, and in each lexical decision condition in which all the fillers were verbs or all the fillers were nouns, 20 participants took part. Thirty undergraduate students, finally, participated in the condition in which syntactic disambiguation was presented. All participants were students at the University of Nijmegen and were native speakers of Dutch.

Materials. We selected 35 verb forms (12 forms first person present tense, 11 forms second/third person present tense and 12 forms first/second/third person irregular past tense) which also had a clear noun meaning (e.g., *last*, ’burden’ or ’welds’). Their meanings were not related. As verbal forms, these words had a mean surface frequency of 84.5, a mean base frequency of 274.9 and a mean family size of 35. The mean surface frequency of the nominal meaning was 20.5, base frequency 24.9, and the mean family size was 21. The correlation between the surface frequencies, base frequencies, and family size counts of the two meanings were not significant ($r = .062, t < 1; r = .151, t < 1;$ and $r = .035, t < 1$).

For the subjective frequency rating and the lexical decision task with the mixed list, we added 50 noun and 50 verb filler words. In all subjective frequency ratings, we chose the filler words such, that half of them (of both nouns and verbs) had a very low base frequency (less than 2) and half of them a very high one (over 250), to ensure that the homonymic targets would mostly be judged around the middle of the scale. The mean base frequencies of the noun and verb filler words in the mixed list lexical decision condition were 4.5 and 7.4 respectively. For the subjective frequency ratings and lexical decision tasks with list manipulation, we added either 86 noun filler words, or 86 verb filler words. The mean base frequencies of the noun and verb fillers in these lexical decision conditions were 55.5 and 70.0 respectively. In the conditions with syntactic disambiguation, finally, we used the same 100 fillers as in the condition without biasing information. Half of the fillers were nouns, and half of the fillers were verb forms. We presented the noun fillers with an article (either *de* or *het*, ’the’) and the verb fillers with a personal pronoun (either *ik*, ’I’, or *hij*, ’he’). We constructed two lists (15 participants per list), presenting a verb-noun homonym in the one list with an article (e.g., *de last*, ’the burden’), and in the other with a personal pronoun (e.g., *hij last*, ’he welds’) such that the same number of homonyms and filler words with a definite article or with a personal pronoun appeared in each list.

For the lexical decision tasks, each word was paired with a pseudoword, the phonotactics of which did not violate the phonology of Dutch. In the lexical decision task with syntactic context, the same pseudowords were used but now presented following either a definite article or a personal pronoun, counterbalanced over the two lists.

Procedure. For the three subjective frequency ratings, participants were asked to indicate on a seven-point scale how often they thought a word is used in Dutch. For the lexical decision tasks, the procedure was identical to that of Experiment 5.1. For the lexical decision task in which syntactic context was provided, the procedure was the same except that either a definite article or a personal pronoun was presented together with the target words. The participants were now asked to decide as quickly and accurately as possible whether the letter string following the definite article or personal pronoun was a real Dutch word. This procedure is somewhat different from the one in Forster and Bednall (1976, Experiment 2), who used a syntactic function decision in which participants had to respond whether the article or pronoun presented with a noun, verb, or noun-verb homonym, constituted a syntactically legal combination.

Results and Discussion

We first report the overall results of the different experimental conditions. We then continue with a series of correlational analyses for the three variables surface frequency, base frequency and family size of the two different meanings across conditions.

For the lexical decision tasks, all participants performed with an error rate below 14%. Table 5.2 shows the means and standard deviations of the ratings and response latencies for all experimental conditions.

The homonyms in the verbal list were rated higher than those in the nominal list, although only marginally significantly so in the by subject analysis ($t1(34) =$

$1.95, p = .059; t2(68) = 2.36, p = .021$). At the same time, in the lexical decision data, the homonyms in the verbal list were processed more slowly than in the nominal list, although again only marginally so in the by subject analysis ($t1(38) = -1.81, p = .079; t2(68) = -3.11, p = .003$). There was no significant difference between the reaction times of the two syntactic conditions ($t1, t2 < 1$). We offer an explanation of the mean differences of the ratings and response latencies between the different conditions after discussing the correlational analyses, as we need these results to understand these differences.

Table 5.2: Results of Experiment 5.2: Means and standard deviations of ratings, response latencies and error proportions (by participants).

Experimental condition	RT/rating (*SD*)	error (*SD*)
Rating with nominal and verbal fillers	4.4 (1.0)	
Lexical decision with nominal and verbal fillers	572 (36)	.03 (.05)
Rating with nominal fillers	3.8 (.9)	
Lexical decision with nominal fillers	598 (71)	.08 (.12)
Lexical decision with definite article	641 (88)	.08 (.11)
Rating with verbal fillers	4.4 (1.1)	
Lexical decision with verbal fillers	645 (52)	.09 (.12)
Lexical decision with personal pronoun	644 (80)	.10 (.11)

Turning to the correlational results, we again report one-tailed t-tests mentioning only those correlations which reach significance for an alpha-level of $.025$. In addition, we will perform stepwise linear regression analysis, entering the surface frequency, base frequency, or family size counts of both meanings as explanatory variables. As in Experiment 5.1, there was pervasive collinearity in the data matrix (condition number 37.02), which shows that linear regression models entering the six variables jointly, are not informative. Figure 5.2 summarizes the results. As in Figure 5.1, the y-axes represent the values of the Pearson correlation coefficient for the different variables in the different conditions. The x-axes represent the three different variables surface frequency (Surf), base frequency (Base), and family size (Fam) for both the nominal (N, dark bars) and verbal (V, light bars) meaning of the homonyms. The first column shows the rating conditions in which the list bias was neutral, nominal, or verbal. The second column shows the lexical decision conditions with the same list biases. The third column, finally, represents the correlational results for the experimental conditions of the syntactic disambiguation conditions.

The correlations which are significant at $\alpha = .025$ AND are retained in a stepwise linear regression model are marked with an asterisk. The bars marked with a dot represent the variables which are not significant predictors by themselves, but are nevertheless retained in a stepwise linear regression model.

In the first row of panels of Figure 5.2, we see the results for the mixed lists. For the subjective frequency correlations (the first panel of the first row), we find that the nominal and verbal surface frequency reliably correlate with the ratings ($r = .512, t(33) = 3.42, p = .001$ and $r = .444; t(33) = 2.85, p = .004$), and both are retained in a stepwise linear regression model ($t(32) = 3.64, p = .000$ and $t(32) = 3.11, p = .002$). Similarly, both base frequency counts correlate significantly with the ratings ($r = .605; t(33) = 4.36, p = .000$ and $r = .548; t(33) = 3.76, p = .000$) and are retained in a stepwise linear regression model ($t(32) = 4.61, p = .000$ and $t(32) = 4.03, p = .000$). The correlation for the family size of the nominal meaning with the ratings also is significant ($r = .602; t(33) = 4.33, p = .000$), whereas the correlation of the verbal family size with the ratings does not reach significance ($r = .276; t(33) = 1.65, p = .054$), but both counts are retained as explanatory variables in a stepwise linear regression model ($t(32) = 4.43, p = .000$ and $t(32) = 1.91, p = .033$). The absence of a reliable effect of family size of the verbal meaning suggests that a local activation of the verbal meaning of a homonym, as evidenced by the surface and base frequency effects of the verbal meaning, does not always automatically lead to broad activation of that meaning.

The results for the lexical decision condition with the mixed list (the second panel of the first row) are very different. Only the correlations of surface frequency, base frequency, and family size of the nominal meaning with reaction times are significant ($r = -.481, t(33) = -3.15, p = .002; r = -.548, t(33) = -3.76, p = .000;$ and $r = -.438, t(33) = -2.80, p = .004$). Comparing the contribution of the nominal and verbal surface frequency, base frequency, or family size in stepwise linear regression models, only the nominal variables turn up as reliable predictors of reaction times.

We now turn to the second row of panels of Figure 5.2, plotting the correlational results in which the nominal meaning was favored. For the subjective frequency rating with the nominal list (the first panel of the second row), only the surface frequency and family size of the nominal meaning emerges with a significant correlation with the ratings ($r = .727, t(33) = 6.06, p = .000$ and $r = .742, t(33) = 6.35, p = .000$) and are retained in stepwise linear regression models as explanatory variables, contrary to the results of the subjective frequency

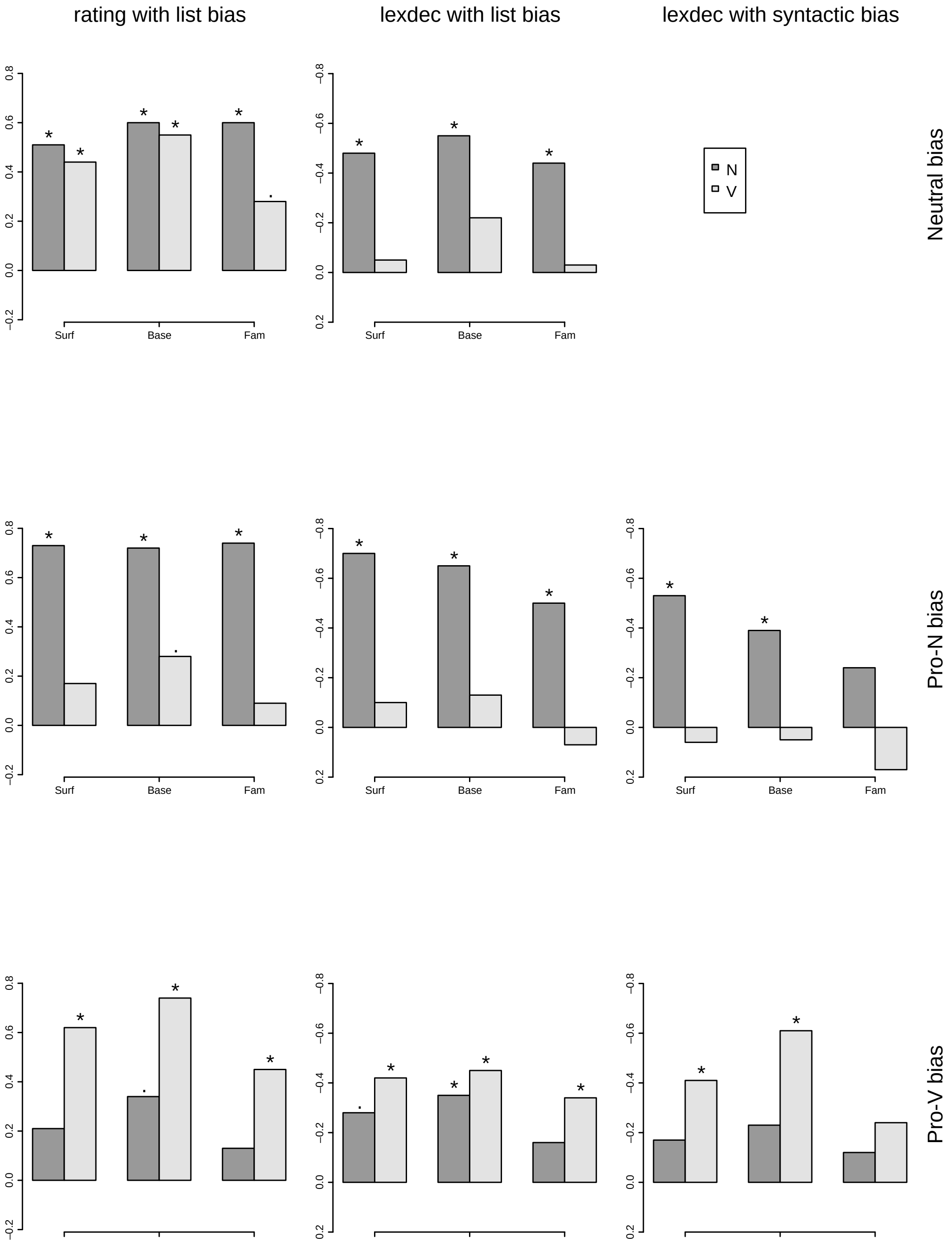

Figure 5.2: Values (r) of Pearson correlations of ratings (first column of panels) and reaction times (the remaining panels) with the nominal (N, dark bars) and verbal (V, light bars) surface frequencies (Surf), base frequencies (Base), and family size counts (Fam) of noun-verb homonyms in Experiment 5.2 in different contexts. Asterisks mark the significant correlations ($p < 0.025$) and dots represent variables that are not significant by themselves, but are nevertheless retained in a stepwise linear regression model.

rating study with the mixed list. Only the nominal base frequency reveals a significant correlation ($r = .722, t(33) = 6.67, p = .000$), but both the nominal and the verbal base frequency counts are retained in a stepwise linear regression model, although the coefficient of the verbal base frequency does not reach significance by itself ($t(32) = 5.81, p = .000$ and $t(32) = 1.49, p = .074$).

The correlational results for the lexical decision data with the nominal list (the second panel of the second row) are very similar: Only the correlations of the surface frequency, base frequency and family size of the nominal meaning with reaction times are significant ($r = -.697, t(33) = -5.59, p = .000; r = -.647, t(33) = -4.88, p = .000;$ and $r = -.503, t(33) = -3.35, p = .001$), and entering both surface frequency, base frequency, and family size counts in stepwise linear regression models, only the nominal counts are retained as explanatory variables.

The homonyms in the condition with syntactic context when preceded by a definite article (the last panel in the second row), behave somewhat differently. In this case, only the surface frequency and base frequency of the nominal meaning play a role ($r = -.532, t(33) = -3.61, p = .000$ and $r = -.386, t(33) = -2.40, p = .011$), and in stepwise linear regression models, only the nominal surface frequency and base frequency are retained as explanatory variables. The correlation with the family size of the nominal meaning has disappeared and neither family size counts in the model resulting from a stepwise linear regression analysis reaches significance. Apparently, only local activation of the nominal meaning occurred.

The third row of panels of Figure 5.2 plots the correlational results for the conditions in which the verbal meaning was favored. In this row of panels we now see a mirrored pattern of results. First consider the subjective frequency rating in which all filler words were verbs (the first panel of the third row). Interestingly, now only the correlation of surface frequency and family size of the verbal meaning with the ratings is significant ($r = .624, t(33) = 4.59, p = .000$ and $r = .446, t(33) = 2.86, p = .004$), and in the two stepwise linear regression models, only the verbal surface frequency and family size counts are retained. For the base frequency counts, only the verbal base frequency correlates significantly with the ratings ($r = .735, t(33) = 6.23, p = .000$), but in a stepwise linear regression model, both the nominal and verbal counts are retained as explanatory variables ($t(32) = 2.03, p = .025$ and $t(32) = 6.13, p = .000$).

For the lexical decision task in which all filler words were verbs (the second panel of the last row), the nominal meaning seems to emerge somewhat more strongly. Only the surface frequency count of the verbal meaning correlates significantly with reaction times ($r = -.420, t(33) = -2.19, p = .006$), but both the nominal and verbal counts are retained in a stepwise linear regression model, although the coefficient of the nominal surface frequency is not significant by itself ($t(32) = -1.62, p = .057$ and $t(32) = -2.62, p = .007$). For the base frequencies, both the nominal and verbal base frequency counts significantly correlate with reaction times ($r = -.347, t(33) = -2.12, p = .021$ and $r = -.454, t(33) = -2.92, p = .003$), and both are retained in a stepwise linear regression model ($t(32) = -1.89, p = .034$ and $t(32) = -2.72, p = .005$). For the nominal and verbal family size counts, only the verbal family size correlates significantly with reaction times ($r = -.341, t(33) = -2.08, p = .023$), and only the verbal family size count is retained in a stepwise linear regression model. This suggests that the nominal meaning was activated only locally, alongside with the local as well as broad activation of the verbal meaning.

However, in the lexical decision task in which the syntactic context (a personal pronoun) requires the verbal meaning (the third panel in the last row), we clearly see that only the verbal surface and base frequency play a role ($r = -.411, t(33) = -2.59, p = .007$ and $r = -.608, t(33) = -4.41, p = .000$). Likewise, in stepwise linear regression models, only the verbal surface and base frequency turn up as an explanatory variables. In the stepwise linear regression model with the two family size counts as predictors, neither reaches predictor significance. Note that, as for the homonyms presented with a definite article, no significant correlation emerges between the reaction times and the family counts of the nominal or the verbal meaning. Only local activation of the verbal meaning takes place. Before discussing this absence of broad activation for both syntactic conditions, we return to the mean differences in ratings and reaction times across the different conditions.

Recall that the homonyms presented in a verbal list take longer to process than when presented in a nominal list (mean 645 versus 598, see Table 5.2), despite the fact that the mean word-frequency of the verbal meaning was higher than the mean word-frequency of the nominal meaning and despite the fact that, in the rating study, the homonyms were rated higher when presented in a verbal list than when presented in a nominal list (mean 4.4 and 3.8 respectively, as expected given the difference in mean word (base) frequencies, verbal: 274.9; nominal: 24.9). Apparently, in a lexical decision task, the nominal meaning of a noun-verb homonym has an intrinsic advantage over the verbal meaning.

Another advantage of the nominal meaning over the verbal meaning emerges from the results of the lexical decision task without disambiguating context. In this case, the homonyms were processed as nouns exclusively, as evidenced by the significant correlations between response latencies and the surface frequency, base frequency and family size of the nominal meaning only. These findings are in line with the results of Baayen et al. (1997) who found that participants are faster to respond to nouns than to verbs which are matched in word frequency. What the present results also show is that the absence or presence of a context is crucial for the differences in reaction times. The differences between the response latencies of the homonyms processed as verbs or as nouns disappear when minimal syntactic context is provided (mean 644 and 641 respectively). Apparently, the absence of a context renders verbs more difficult to process than nouns, which is not surprising considering that nouns can occur in natural language in isolation, but finite verb forms in a language such as Dutch cannot. Therefore, presenting these noun-verb homonyms in isolation may also be seen as providing a biasing context, because only nouns can occur in isolation.

We now turn to the question why the morphological family of both the nominal and the verbal meaning appear to play no role when immediate sentential context is provided. De Jong et al. (2000, also Chapter 2) show that the morphological context in which a verb appears can change which subset of family members contributes effectively to the effect of family size. For instance, it is only in the presence of an overt verbal suffix that the verbal family members of the target word contribute to the effect of family size. De Jong et al. (to appear, also Chapter 4) investigate the effect of morphological and minimal sentential context on the activation of family members of adjectives. They found that it is not only morphological context that is able to narrow down the activation of family members to appropriate subfamilies, but that a minimal sentential context may do so as well. In the condition of the present experiment in which the noun-verb homonyms were likewise presented in minimal sentential context, the family size effect might also be restricted to particular subfamilies, leading to the absence of an effect of the total family size.

In fact, when the homonym is presented with a personal pronoun, we may expect a verbal subfamily effect of the verbal meaning, as the presence of a personal pronoun leads one to expect a verb to follow. We therefore restricted our count of family members of the verbal meaning to the subset of its verbal family members only. For instance, we do not include nouns such as *lasapparaat*, ’welding machine’, for the stimulus *hij last*, ’he welds’, but only verbs such as *inlassen*, ’to insert, to weld in’. Interestingly, the correlation of the verbal subfamily of the verbal meaning improves considerably compared to the correlation of the total family of the verbal meaning, and now reaches significance ($r = -.381, t(33) = -2.37, p = .024$). Apparently, the presence of the personal pronoun further narrows down the activation of the family members to precisely the contextually appropriate, verbal family members of the verbal meaning of the homonym.

We have seen that the correlation between reaction times and the complete family of the nominal meaning when the homonyms were presented with a definite article did not reach significance. In this case, we might expect a subfamily effect of the nominal family members of the nominal meaning, since the presence of a definite article creates the expectation of a noun to follow. We therefore constructed new counts in which, for instance, we did not include verbs like *belasten*, ’to burden’, for the stimulus *de last*, ’the burden’ but only nominal family members like *belasting*, ’burden’ or ’taxation’. However, hardly any family members of these nominal meanings were verbal or adjectival. Therefore, the counts of the nominal subfamily of the nominal meaning are nearly identical to the counts of its total family size. Not surprising, the correlation with reaction times for these new counts of the nominal family were not significant either.

We suspect that the absence of an effect of family size when the homonyms are presented with a definite article is due to the semantic difference between nouns presented without any context and nouns with a definite article. According to Langacker (1991), the use of an article ’grounds’ the nominal clause, i.e. it establishes the location of the thing serving as the nominal profile. Without context, the noun reflects a category, which means that all features of this category are still relevant. Upon reading *man*, words like *chairman* and *fisherman*, the morphological family members of the word, are therefore also activated. But if a specific instantiation of the category is called for, because of the definite article, only the meaning of the word itself is still relevant. In this case, only local activation of the meaning of *man* takes place, whereas the absence of a context leads to broad activation in the mental lexicon. More research is clearly required to investigate the role of specification by means of a definite article on the activation of family members.

In conclusion, Experiment 5.2 clearly shows that besides family size, the different absolute frequency counts of noun-verb homonyms influence reaction times and ratings depending on context. Forster and Bednall (1976, Experiment 2), who used a syntactic function decision task with very similar minimal syntactic disambiguation for noun-verb homonyms, report that a frequency effect is only obtained in the case that the dominant meaning is called for. However, the frequency counts used in their experiment were the summed frequencies of both meanings. In the present experiment with syntactic disambiguation, we see that the two frequency counts have different effects as a function of context. Therefore, we think that their effect of summed frequency in the case that the syntactic context directed to the dominant meaning actually was an effect of the absolute frequency of the dominant meaning, as these two counts are fairly similar. Their absence of an effect of the summed frequency when the context directed to the subordinate meaning can now be explained, as the summed frequency does not resemble the absolute frequency count of the subordinate meaning.

General Discussion

This paper addresses the question how ambiguous words are processed in different contexts. For an unambiguous word, the token count of a word's frequency, and also the type count of its morphological family members (its morphological family size), are predictors of processing times. Several studies have shown that this effect of family size is context-sensitive. In this study, we address the question how context may influence the effects of the two morphological families of an ambiguous word. Besides the effect of the morphological families, we investigate the role of absolute frequency counts of the two meanings of a homonym (rather than just the binary notion of dominant versus subordinate extrapolated from a relative frequency), in different contexts.

The tasks that we have employed are subjective frequency rating and lexical decision. An effect of the token frequency of one of the meanings of a homonym can be taken as evidence for the activation of that particular meaning. We have called this local activation. An effect of the morphological family size of one of the meanings can be seen as evidence for the spreading of activation to the morphological family members of that particular meaning. We have referred to this more extensive semantic activation as broad activation.

In the first experiment, we presented noun-noun homonyms without context, or with a prime related to either of the meanings of the ambiguous target word as disambiguating context. For the second experiment, we used noun-verb homonyms that allowed us to use list manipulation and syntactic context for disambiguation. Our main interest in these studies are the correlational analyses of the effects of the surface frequencies, base frequencies, and family size counts of the two different meanings of the ambiguous words.

Table 5.3 summarizes the correlational results of the first experiments. The rows represent the different contexts in which we placed these homonyms, and the

Table 5.3: Summary of results of Experiment 5.1, with significant effects of the dominant (D) and subordinate meanings (S) of noun-noun homonyms or with no significant effects (ns).

Context	rating		lexical decision	
	frequency	family size	frequency	family size
Neutral context	D/S	D/S	D/S	D/S
Prime-biasing for Dominant	D	D	D/S	D
Prime-biasing for Subordinate	ns	ns	ns	ns

columns represent the results for the word frequency measures and the family size measure in the rating studies and lexical decision conditions. The D's stand for a significant effect of the dominant meaning of the homonym, and the S's stand for a significant effect of the subordinate meaning. The first row shows that in the rating and the lexical decision without a context, the word frequency and the family size of both meanings play a role. Evidently, local as well as broad activation of both meanings took place. The presence of a prime related to the dominant meaning in the rating study, as can be seen in the second row of Table 5.3, was sufficient to limit the activation locally to the dominant meaning of the word, along with the broad activation of the family members related to the dominant meaning. The lexical decision data with the prime related to the dominant meaning showed somewhat different results: the base frequency of the subordinate meaning, rather than of the dominant meaning, correlated with reaction times. Turning to the third row, we see that there was no evidence whatsoever of activation when the prime was related to the subordinate meaning of the homonym.

Especially this last result leads us to conclude that this particular use of the priming paradigm might not be optimally suited for tracking the processing of these ambiguous words, for several reasons. First, for unambiguous words, a context by frequency interaction has been reported, using sentences as primes (e.g., Stanovich & West, 1981), as well as single word primes (Becker, 1979; Borowsky & Besner, 1993; and Plaut & Booth, 2000). The results of these studies suggest that the frequency effect may disappear as a function of strength of context, which also might have played a role in this study with ambiguous words. Second, several studies (see Kahan, Neely, & Forsythe, 1999) have shown that backwards priming (the influence of associates from target to prime on the processing of the target) may play a role. This is especially a problem for priming studies in which the prime is the ambiguous word, as Van Petten and Kutas (1987) show, and perhaps less of a problem when the target is the ambiguous word as in Experiment 5.1 of the present paper. Finally, the word frequency of the prime, which was not controlled for in our Experiment 5.1, may also have affected the amount of priming (Giraudo & Grainger, 2000). It is as yet unclear what the exact effect of the processing of one isolated word might be on the processing of the next semantically related or unrelated isolated word.

In the second experiment, we therefore abandoned the priming paradigm and turned to noun-verb homonyms, which lend themselves for disambiguation by list manipulation as well as disambiguation by means of a minimal syntactic context. By using such a semantically void context, which yet fully disambiguates these noun-verb homonyms, we avoid the possibility of a context by frequency interaction. Table 5.4 summarizes the correlational results of this experiment. The N's and V's in this table represent the significant effects of frequency and family size of the nominal and verbal meaning of the homonym across conditions.

Table 5.4: Summary of results of Experiment 5.2, with significant effects for the nominal and verbal meaning of noun-verb homonyms (N and V for significant effects, n and v for effects which are not significant by themselves but are nevertheless retained in a stepwise linear regression model), or no significant effects (ns).

Context	rating		lexical decision	
	frequency	family size	frequency	family size
Mixed-list	N/V	N/v	N	N
List-biasing for Noun	N/v	N	N	N
Syntactic biasing for Noun	–	–	N	ns
List-biasing for Verb	V/n	V	V/n	V
Syntactic biasing for Verb	–	–	V	verbal subset*

* explained in text

The first row shows that when no disambiguating context was provided, both in the rating and in the lexical decision, local and broad activation of the nominal meaning always took place, as evidenced by the significant effects of nominal word frequency and nominal family size. Activation of the verbal meaning was apparent in the rating study only, but the activation might be limited to local activation mainly. Apparently, when a verb-noun homonym is presented, even when the verbal meaning is the more frequent, the nominal meaning is the more susceptible one. We think that the absence of a context is more problematic for processing inflected verb forms than for processing nouns, leading to the activation of only the nominal meaning in the lexical decision task without disambiguating context.

As can be seen in the second row of Table 5.4, presenting the homonyms in a list-biasing context favoring the nominal meaning in a rating leads to local activation of mainly the nominal meaning and to broad activation of the nominal meaning exclusively. In the lexical decision task with list biasing context, locally as well as broadly, only the nominal meaning is activated.

In the third row we see the results for the lexical decision study when placing the homonyms in a syntactic context biasing the nominal meaning. Contrary to the list manipulation results, now the nominal meaning is activated only locally, while broad activation does not take place. We have offered the explanation that this absence of an effect of the family members of the nominal meaning, or of a specific subfamily of the nominal meaning, is due to the semantics conveyed by the definite article. A noun without an article refers to a category, whereas the definite article ’grounds’ the noun, making it refer to an instantiation of that category (Langacker, 1991). We think that this specification of the semantics by the definite article narrows down the range of meanings to precisely that of the noun itself. Therefore, little or no activation can spread to contextually inappropriate morphological family members. This results in local activation of the nominal meaning of the homonym exclusively.

The fourth row summarizes the results when we bias the verbal meaning by means of list manipulation. The results are reversed compared to the results of the same homonyms in a list manipulation biasing the nominal meaning. It is now mainly the verbal meaning, along with the family members of the verbal meaning that are activated. In the rating and especially the lexical decision task, the meaning of the noun appears to be activated as well. But the absence of an effect of the family members of the nominal meaning show that the activation of the nominal meaning was only local. Apparently, the advantage of the nominal meaning over the verbal meaning as evidenced in the mixed list studies is so great that even in this verbal list local activation of the nominal meaning can occasionally take place.

The fifth row shows the results when minimal syntactic context biases the verbal meaning in lexical decision by presenting the homonym following a personal pronoun. In this case, the verbal meaning seemed to be activated only locally. The total count of family members of the verbal meaning did not play a role. However, upon closer examination of the family members of the verbal meaning, we found that a specific subset, the verbal family members of the verbal meaning, did play a significant role. This finding is in line with De Jong et al. (to appear, also Chapter 4), who found that minimal sentential context can limit the effect of family size to a contextually appropriate subset of family members. In this case, the presence of a personal pronoun further narrowed down the activation within the morphological family of the verbal meaning of the homonym to precisely its verbal family members.

We conclude from the present results that the context in which a homonym appears plays a role in the processing of such a homonym. Besides the context-sensitivity of the effect of family size for such words, we see that the effect of absolute word-frequency is context-sensitive as well.[3] This has implications for studies comparing reaction times or reading times on homonyms to frequency-matched control words (e.g., Folk & Morris, 1995; Rayner et al., 1994). If the frequency effect is sensitive to context, it might well be that latencies for subordinate meanings are longer than those for control words, precisely because the control words have a higher frequency than the subordinate meaning of the homonym. Therefore, the subordinate bias effect could partly be due to an effect of concept-frequency. In this explanation of the subordinate bias effect, it is precisely the selection due to context that ensures the difference in latencies, rather than a time-costly competition between the dominant and subordinate meanings.

This explanation of absolute frequency is first suggested by Rayner and Duffy (1986). To test this hypothesis, in a post-hoc analysis, they compare gaze durations on equibiased homonyms (for which the two meanings have a similar frequency) to control words which are matched in absolute frequency and still find a 10 ms. difference. They conclude that future research is necessary to ascertain whether the absolute frequencies play a role, but that their results do not provide strong support for such a hypothesis. Kawamoto (1993), who develops a connectionist model to explain several results found for ambiguous words, accounts for such an effect of relative frequency (but see Twilley & Dixon, 2000, for a quantitative model of ambiguity resolution in which absolute frequencies might play a role). The present results, however, show that absolute frequency effects of the two different meanings of homonyms do exist.

On the other hand, we have found that when strongly related primes precede homonyms, the effects of frequency of especially the subordinate meaning can be reduced. We suggested that the interactive effect of context by frequency observed for unambiguous words might be responsible. The presence or absence of a subor-

---

[3] In fact, we also analysed the possible role of relative frequencies in all conditions of the two experiments. We used the absolute frequency counts to calculate these relative frequencies by dividing for both meanings the absolute frequency of one of the meanings by the summed frequencies of both meanings. These relative frequency counts never were significant predictors, not for the ratings, nor for the reaction times.

dinate bias effect could in this reasoning at least in part be sensitive to the strength of semantic context, as claimed by Paul et al. (1992), Martin et al. (1999), and Vu et al. (1998). Either way, the thrust of the argument based on our results is that word-frequency effects are not restricted to the access level (reflecting string-familiarity).[4] At least a part of the word-frequency effect must be located at the semantic level (see also, Balota, 1990; Borowsky & Besner, 1993; Hino & Lupker, 1996; Plaut & Booth, 2000; Plourde & Besner, 1997). Converging evidence for the effects of absolute frequency of the different meanings of homonyms in another modality comes from Caramazza, Costa, Miozzo, and Bi (2001). In a recent paper studying production of homonyms, they also report on an effect of absolute (specific) frequency in a picture naming task.

For ambiguous words such as the homonyms used in the present study, it is clear that depending on context, a different meaning is called for. But non-homonymic words may also have different shades of meaning. A recent study addressing this issue is Azuma and Van Orden (1997) who investigated different sorts of polysemous words presented without context. Following Jastrzembski (1981), they investigated not only the Number of Meanings of a word, but also the Relatedness within these meanings.[5] In Jastrzembski's study (1981), Number of Meanings and Relatedness was estimated by using dictionary-entries, a method which has been criticised as psychologically invalid (Gernsbacher, 1984). In the Azuma and Van Orden study, the Number of Meanings of a word was estimated by having participants fill out all the meanings they knew for a specific word. Relatedness was estimated by using a rating task. All subordinate meanings were paired with the dominant meaning of a word and participants were asked to rate (on a seven-point-scale) the relatedness for every pair of meanings. The dominant meaning was defined as the most frequently reported meaning in the study estimating the number of meanings. In two experiments, they orthogonally contrasted these two variables. In Experiment 1 using legal nonwords, no significant effects were found. But when pseudohomophones were used as nonwords, there was a main effect of Number

---

[4] Whether the effects of frequency we have observed in this study are in fact effects of frequency only, is not clear. A variable which is confounded with frequency is Age of Acquisition (e.g., Carroll & White, 1973). For our materials, Age of Acquisition norms do not exist. As yet, our results could therefore be explained as effects of frequency, effects of Age of Acquisition, or as a combination of the two (cumulative frequency; see Lewis, Gerhand, & Ellis, 2001).

[5] Jastrzembski (1981) also investigated the effect of number of derivations, a variable similar to the family size variable we use, and found no significant effect. But it seems that the derivations in Jastrzemski's (1981) study were mostly semantically unrelated to the target base word. Furthermore, the orthogonal contrast he uses, is very small (e.g., a mean of 4.0 to 2.1 number of derivations in his Experiment 5).

of Meanings and a main effect of Relatedness. Crucially, they also observed an interaction of the two variables. Participants responded faster to words with many meanings than to words with few meanings, but only when the relatedness for these few-meaning words was low. In their Experiment 2, with mostly different materials, they used pseudohomophones as nonwords and found a main effect of Relatedness only. There was no main effect of Number of Meanings. As in Experiment 1, they also found an interaction with Number of Meanings and again, Relatedness only mattered for words with few meanings.

It is revealing, however, to consider the results of Experiment 1 and 2 of Azuma and Van Orden (1997) with respect to family size. To estimate the contributions of the three variables Relatedness, Number of Meanings, and family size, we performed stepwise linear models, excluding in every step main effects and interactions which were not significant. For Experiment 1 there were no significant effects for Relatedness and Number of Meanings for the sub-experiment with legal nonwords. Family size, by contrast, can clearly account for some of the variance in this dataset. In a stepwise linear regression analysis, family size emerges as the only reliable factor ($\beta = -23.01, t(38) = -3.69, p = .001$). For the same items in the sub-experiment that used pseudohomophones as nonwords, it is again only family size that is retained as a fully reliable factor ($\beta = -36.52, t(37) = -4.21, p = .000$), but now Relatedness may perhaps also have some influence ($\beta = -10.50, t(37) = -2.01, p = .052$). Turning to Experiment 2, we see a very similar pattern: Only family size ($\beta = -29.65, t(37) = -2.40, p = .021$) and Relatedness ($\beta = -16.48, t(37) = -2.39, p = .022$) are reliable factors.

Summing up, in all these analyses Number of Meanings is discarded as a reliable factor. It is replaced by family size as a superior measure for accounting for the data. In addition to family size, Relatedness also has some effect: The more clearly related the different meanings of a word are, the faster participants are able to respond. We hypothesize that the factor Relatedness is more or less capturing the same as the effect which Bertram et al. (2000) found: that removing opaque family members from the count of morphological family members improves reaction times.

We have seen for homonyms without context that both meanings are activated. As the polysemous words used by Azuma and Van Orden (1997) occurred in isolation, all shades of meaning of these words and therefore all family members were still relevant. This resulted in the high correlations between reaction times and family size. But when the homonyms of the present study were placed in a (minimal) context, local as well as broad activation of the meanings was affected. Whether the context in which a polysemous word occurs likewise affects the local and broad activation of its (shades of) meaning, is subject for future research. Azuma and Van Orden (1997) stress the fact that the measure Number of Meanings is not obtained objectively. Importantly, the family size measure is obtained in an objective manner, namely by simply counting the number of morphological descendants of a word in a large corpus. Although a list of morphological family members may not be the adequate way to describe a word's meaning, we think that such a list does reflect the broad semantic structure of a word, and that the family size can be used as a tool to better understand the activation of the different shades of meaning of a word.

We conclude that resolving homonymy does not only occur by locally activating the appropriate meaning, but also by broader activation of precisely that morphological family that is related to the appropriate meaning. Furthermore, the immediate context affects this broad activation and narrows it down to a specific subset of contextually appropriate family members. We hypothesize that for unambiguous words with different shades of meaning, broad activation is also sensitive to context. Ambiguity resolution probably is not a process only occurring upon reading or hearing ambiguous words, but also (albeit less obviously) for unambiguous words. Previous studies investigating the role of context on the effect of family size have shown that the activation of different subfamilies depending on context also occurs for unambiguous words (De Jong et al., 2000, also Chapter 2; De Jong et al., to appear, also Chapter 4). We hypothesize that the mechanism which leads to the context-sensitive, appropriate local activation of a word's meaning which we observed for homonyms, may well be the same as the mechanism underlying the context-sensitive, appropriate broad activation in the mental lexicon.

References

Andrews, S. A. and Heathcote, A.: 2001, Distinguishing common and task-specific processes in word identification: A matter of some moment?, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **27**, 514–544.

Azuma, T. and Van Orden, G. C.: 1997, Why safe is better than fast: The relatedness of a word's meaning affects lexical decision times, *Journal of Memory and Language* **36**, 484–504.

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

Balota, D. A.: 1990, The role of meaning in word recognition, *in* D. A. Balota, G. B. Flores d'Arcais and K. Rayner (eds), *Comprehension Processes in Reading*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 9–32.

Balota, D. A. and Chumbley, J. I.: 1984, Are lexical decisions a good measure of lecical access? The role of word frequency in the neglected decision stage, *Journal of Experimental Psychology: Human Perception and Performance* **10**, 340–357.

Balota, D. A. and Chumbley, J. I.: 1990, Where are the effects of frequency in word recognition tasks? Right where we said they were! Comment on Monsell, Doyle, and Haggard (1989), *Journal of Experimental Psychology: General* **119**, 231–237.

Becker, C. A.: 1979, Semantic context and word frequency effects in visual word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **5**, 252–259.

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.

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

Binder, K. S. and Rayner, K.: 1998, Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading, *Psychological Bulletin and Review* **5**, 271–276.

Borowsky, R. and Besner, D.: 1993, Visual word recognition: A multistage activation model, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **19**, 813–840.

Bradley, D. C. and Forster, K. I.: 1987, A reader's view of listening, *Cognition* **25**, 103–134.

Burgess, C., Tanenhaus, M. K. and Seidenberg, M. S.: 1989, Context and lexical access: Implications of nonword interference for lexical ambiguity resolution, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**, 620–632.

Caramazza, A., Costa, A., Miozzo, M. and Bi, Y.: 2001, The specific word-frequency effect: Implications for the representation of homophones in speech production, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **27**, 1430–1450.

Carroll, J. B. and White, M. N.: 1973, Word frequency and age of acquisition as determiners of picture-naming latency, *Quaterly Journal of Experimental Psychology* **25**, 85–95.

Chwilla, D. J., Brown, C. M. and Hagoort, P.: 1995, The N400 as a function of the level of processing, *Psychophysiology* **32**, 274–285.

Cutler, A. and Clifton, C.: 1984, The use of prosodic information in word recognition, *in* H. Bouma and D. G. Bouwhuis (eds), *Attention and Performance X*, Hillsdale, NJ: Erlbaum.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M. and Baayen, R. H.: 2002, The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects, *Brain and Language*, **81**, 555–567.

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 Jong, N. H., Schreuder, R. and Baayen, R. H.: *to appear*, Morphological resonance in the mental lexicon, *Linguistics*.

Deloche, G., Seron, X., Scius, G. and Segui, J.: 1987, Right hemisphere language processing: Lateral difference with imageable and nonimageable ambiguous words, *Brain and Language* **30**, 197–205.

Deutsch, W. and Wijnen, F.: 1985, The article's noun and the noun's article: Explorations into the representation and access of linguistic gender in dutch, *Linguistics* **23**, 793–810.

Duffy, S. A., Morris, R. K. and Rayner, K.: 1988, Lexical ambiguity and fixation times in reading, *Journal of Memory and Language* **27**, 429–446.

Folk, J. R. and Morris, R. K.: 1995, The role of multiple lexical codes in reading: Evidence from naming time, oral reading, and eye movements, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 1412–1429.

Forster, K. I. and Bednall, E.: 1976, Terminating and exhaustive search in lexical access, *Memory and Cognition* **4**, 53–61.

Gernsbacher, M. A.: 1984, Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy, *Journal of Experimental Psychology: General* **113**, 256–281.

Giraudo, H. and Grainger, J.: 2000, Effects of prime word frequency and cumulative root frequency in masked morphological priming, *Language and Cognitive Processes* **15**, 421–444.

Glucksberg, S., Kreuz, R. J. and Rho, S. H.: 1986, Context can constrain lexical access: Implications for models of language comprehension, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **12**, 323–335.

Gottlob, R. G., Goldinger, S. D., Stone, G. O. and Van Orden, G. C.: 1999, Reading homographs: Orthographic, phonologic, and semantic dynamics, *Journal of Experimental Psychology: Human Perception and Performance* **25**, 561–574.

Hino, Y. and Lupker, S. J.: 1996, Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22**, 1331–1356.

Holcomb, P.: 1993, Semantic priming and stimulus degradation: Implications for the role of the n400 in language processing, *Psychophysiology* **30**, 47–61.

Jastrzembski, J.: 1981, Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon, *Cognitive Psychology* **13**, 278–305.

Kahan, T. A., Neely, J. H. and Forsythe, W.: 1999, Dissociated backward priming effects in lexical decision and pronunciation tasks, *Psychonomic Bulletin & Review* **6**, 105–110.

Kambe, G., Rayner, K. and Duffy, S. A.: 2001, Global context effects on processing lexically ambiguous words: Evidence from eye fixations, *Memory & Cognition* **29**, 363–372.

Kawamoto, A. H.: 1993, Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account, *Journal of Memory and Language* **32**, 474–516.

Langacker, R. W.: 1991, *Foundations of Cognitive Grammar. Vol. 2, Descriptive Application*, Stanford University Press, Stanford.

Lewis, M. B., Gerhand, S. and Ellis, H. D.: 2001, Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects?, *Cognition* **78**, 189–205.

Lucas, M. M.: 1987, Frequency effects on the processing of ambiguous words in sentence contexts, *Language and Speech* **30**, 25–46.

Lüdeling, A. and De Jong, N. H.: 2002, German particle verbs and word-formation, *in* N. Dehé, R. Jackendoff, A. McIntyre and S. Urban (eds), *Verb-particle explorations*, Mouton de Gruyter, Berlin, pp. 315–333.

Martin, C., Vu, H., Kellas, G. and Metcalf, K.: 1999, Strength of discourse context as a determinant of the subordinate bias effect, *Quaterly Journal of Experimental Psychology* **52A**, 813–839.

Monsell, S., Doyle, M. C. and Haggard, P. N.: 1989, Effects of frequency on visual word recognition tasks, *Journal of Experimental Psychology: General* **118**, 43–71.

Morton, J.: 1969, The interaction of information in word recognition, *Psychological Review* **76**, 165–178.

Neely, J. H.: 1991, Semantic priming effects in visual word recognition: A selective review of current findings and theories, *in* D. Besner and G. W. Humphreys (eds), *Basic processes in reading*, Hillsdale, NJ: Erlbaum, pp. 264–336.

Paul, S. T., Kellas, G., Martin, M. and Clark, M. B.: 1992, Influence of contextual features on the activation of ambiguous word meanings, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **18**, 703–717.

Plaut, D. C. and Booth, J. R.: 2000, Individual and developmental differences in semantic priming: Empirical and computational support for a single mechanism account of lexical processing, *Psychological Review* **107**, 786–823.

Plourde, C. E. and Besner, D.: 1997, On the locus of the word frequency effect in visual word recognition, *Canadian Journal of Experimental Psychology* **51**, 181–194.

Rayner, K. and Duffy, S. A.: 1986, Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity, *Memory and*

Coginition **14**, 191–201.

Rayner, K. and Frazier, L.: 1989, Selection mechanisms in reading lexically ambiguous words, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**, 779–790.

Rayner, K. and Morris, R. K.: 1991, Comprehension processes in reading ambiguous sentences: Reflections from eye movements, *in* G. B. Simpson (ed.), *Understanding Word and Sentence*, Elsevier, Amsterdam, pp. 175–198.

Rayner, K., Binder, K. S. and Duffy, S. A.: 1999, Contextual strength and the subordinate bias effect: Comment on Martin, Vu, Kellas, and Metcalf, *Quarterly Journal of Experimental Psychology* **52A**, 841–852.

Rayner, K., Pacht, J. M. and Duffy, S. A.: 1994, Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: evidence from eye fixations, *Journal of Memory and Language* **33**, 527–544.

Rudell, A.: 1999, The recognition potential and the word frequency effect at a high rate of word presentation, *Cognitive Brain Research* **8**, 173–175.

Rudell, A. P. and Hu, B.: 2001, Does a warning signal accelerate the processing of sensory information? evidence from recognition potential responses to high and low frequency words, *International Journal of Psychophysiology* **41**, 31–42.

Schreuder, R. and Baayen, R. H.: 1995, Modeling morphological processing, *in* L. B. Feldman (ed.), *Morphological Aspects of Language Processing*, Lawrence Erlbaum, Hillsdale, New Jersey, pp. 131–154.

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

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M. and Bienkowski, M.: 1982, Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing, *Cognitive Psychology* **14**, 489–537.

Simpson, G. B. and Krueger, M. A.: 1991, Selective access of homograph meanings in sentence context, *Journal of Memory and Language* **30**, 627–643.

Stanovich, K. E. and West, R. F.: 1981, The effect of sentence context on ongoing word recognition: Test of a two-process theory, *Journal of Experimental Psychology: Human Perception and Performance* **7**, 658–672.

Swinney, D. A.: 1979, Lexical access during sentence comprehension: (re)consideration of context effects, *Journal of Verbal Learning and Verbal Behavior* **18**, 645–659.

Tabossi, P. and Zardon, F.: 1993, Processing ambiguous words in context, *Journal of Memory and Language* **32**, 359–372.

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

Twilley, L. C. and Dixon, P.: 2000, Meaning resolution processes for words: A parallel independent model, *Psychonomic Bulletin & Review* **7**, 49–82.

Van Petten, C. and Kutas, M.: 1987, Ambiguous words in context: An event-related potential analysis of the time course of meaning activation, *Journal of Memory and Language* **26**, 188–208.

Vu, H., Kellas, G. and Paul, S.: 1998, Sources of sentence constraint on lexical ambiguity resolution, *Memory and Cognition* **26**, 979–1001.

Vu, H., Kellas, G., Metcalf, K. and Herman, R.: 2000, The influence of global discourse on lexical ambiguity resolution, *Memory and Cognition* **28**, 236–252.

Appendix A

Materials, reaction times and ratings of the different conditions of Experiment 5.1. *rating* stands for the mean ratings, *RT* for the mean reaction times (*RTunrel.* for those following unrelated primes), and *error* for the mean error proportions.

Target	Neutral context			Dominant context			Subordinate context		
	RT, error	rating	RTunrel., error	prime	RT, error	rating	prime	RT, error	rating
aas (bait or ace)	611, .11	2.56	515, .19	lokken (to entice)	565, .13	4.00	kaarten (cards)	640, .07	5.00
baan (job or path)	604, .00	5.94	517, .00	dienst (service)	497, .00	5.33	strook (strip)	521, .00	3.67
bal (ball or ball)	552, .00	5.50	517, .00	gooien (to throw)	504, .13	6.00	dansfeest (dance)	547, .00	4.00
blik (look or tin)	562, .00	4.61	527, .00	kijken (to look)	509, .00	4.22	metaal (metal)	487, .07	4.67
bloem (flower or flour)	538, .00	5.33	451, .00	roos (rose)	468, .00	5.33	meel (flour)	528, .00	4.44
bocht (bend or rubbish)	585, .00	4.50	563, .00	draai (turn)	516, .00	5.33	drank (drink)	577, .07	2.44
breuk (break or fraction)	618, .00	3.61	545, .00	vallen (to fall)	543, .07	3.83	rekenen (to calculate)	591, .00	4.33
bus (bus or tin)	532, .00	5.78	488, .00	trein (train)	488, .00	5.39	opbergen (to put away)	547, .00	3.39
das (tie or badger)	573, .06	4.44	542, .00	kou (cold)	506, .00	4.61	dier (animal)	598, .00	4.11
gang (passage or walk)	522, .06	5.22	465, .00	huis (house)	532, .07	5.11	pas (step)	552, .07	2.33
gier (vulture or liquid manure)	640, .06	2.61	594, .00	vogel (bird)	601, .07	3.44	mest (manure)	543, .00	3.89
hoop (hope or heap)	563, .06	4.89	495, .00	bidden (to pray)	539, .00	3.89	berg (mountain)	617, .07	4.22
jacht (hunting or yacht)	540, .00	3.83	558, .00	jager (hunter)	502, .00	5.22	zeilen (to sail)	506, .00	4.83
kas (cash-desk or greenhouse)	605, .11	4.28	577, .00	winkel (shop)	537, .00	4.00	bloem (flower)	550, .00	3.56
knol (tuber or nag)	578, .00	2.56	655, .07	koolraap (kohlrabi)	515, .06	2.94	merrie (mare)	571, .07	4.83
kop (head or cup)	591, .00	5.44	592, .00	hoofd (head)	460, .06	5.17	mok (mug)	534, .00	6.17
lof (praise or chicory)	599, .00	3.44	591, .13	prijzen (to praise)	522, .06	4.11	groente (vegetable)	536, .07	3.39
loods (shed or pilot)	630, .00	3.39	705, .20	garage (garage)	548, .06	3.56	boot (boat)	623, .00	3.22
maat (measure or pal)	604, .00	4.56	550, .00	meten (to measure)	505, .00	5.11	vriend (friend)	558, .00	5.00
muil (mouth or mule)	689, .00	2.50	620, .07	mond (mouth)	556, .06	3.28	schoen (shoe)	604, .13	2.22
pad (path or toad)	557, .00	4.67	585, .00	wandelen (to walk)	498, .00	4.78	dier (animal)	533, .00	4.11
rol (roll or role)	557, .06	4.22	523, .00	papier (paper)	462, .00	4.67	theater (theatre)	497, .00	4.44
ruit (window-pane or diamond)	574, .00	4.61	617, .20	venster (window)	496, .00	5.11	vierkant (square)	491, .00	3.28
schaal (scale or dish)	541, .00	4.50	508, .00	maquette (scale-model)	477, .00	3.39	lepel (spoon)	549, .07	2.78
schat (treasure or darling)	555, .00	4.56	541, .00	rijkdom (wealth)	478, .00	4.06	lieveling (darling)	452, .00	6.33
schoft (scoundrel or shoulder)	609, .06	3.72	595, .00	ellendeling (wretch)	572, .07	4.61	paard (horse)	575, .00	1.56
schop (shovel or kick)	561, .06	4.44	563, .13	tuinieren (to garden)	539, .00	4.11	pijn (pain)	488, .00	5.11
schot (shot or partition)	549, .00	3.39	580, .00	kogel (bullet)	715, .00	5.11	afscheiding (separation)	490, .06	3.17
slaap (sleep or temple)	588, .00	6.39	541, .07	bed (bed)	559, .00	6.17	gezicht (face)	465, .00	2.94
sport (sports or rung)	601, .00	5.72	542, .00	atletiek (athletics)	510, .00	5.50	ladder (ladder)	452, .00	2.39
stroom (stream or electric power)	558, .00	5.22	636, .07	rivier (river)	517, .00	3.33	spanning (tension)	508, .00	5.67
toets (test or key)	553, .00	4.11	579, .07	tentamen (examination)	497, .00	5.33	knop (button)	474, .00	5.39
vak (trade or section)	588, .06	5.11	581, .07	arbeid (labour)	534, .00	4.89	veld (field)	544, .13	3.50
vorst (monarch or frost)	660, .00	3.11	631, .00	koning (king)	516, .00	3.39	vriezen (to freeze)	476, .00	5.89
wals (waltz or roller)	593, .00	3.06	592, .07	balzaal (ballroom)	556, .13	3.50	pletten (to crush)	488, .00	3.78
wapen (weapon or blazon)	545, .00	3.89	496, .00	kogel (bullet)	526, .00	5.11	embleem (emblem)	449, .00	3.00
zijde (side or silk)	556, .00	3.00	616, .07	flank (flank)	563, .00	4.00	fluweel (velvet)	482, .00	4.61

Appendix B

Materials, reaction times and ratings of the different conditions of Experiment 5.2. *rating* stands for the mean ratings, *RT* for the mean reaction times, *error* for the mean error proportions, and *syntax* for the syntactic context used as disambiguation.

Target	neutral context		nominal context				verbal context			
	rating	RT, error	rating	RT, error	syntax	RT, error	rating	RT, error	syntax	RT, error
baal (bag or am fed up)	3.44	591, .00	3.28	625, .15	de (the)	775, .40	4.94	606, .00	ik (I)	609, .00
bad (bath or prayed)	5.94	558, .00	5.33	550, .00	het (the)	617, .00	4.17	581, .10	hij (he)	605, .07
borg (surity or stored)	4.00	592, .00	3.94	673, .20	de (the)	683, .07	3.50	655, .15	hij (he)	734, .13
dreef (avenue or floated)	3.50	646, .14	3.00	719, .10	de (the)	682, .33	4.94	625, .05	hij (he)	603, .00
goot (gutter or poured)	3.78	605, .00	3.89	585, .05	de (the)	665, .07	4.83	721, .05	hij (he)	634, .13
hak (heel or hack)	4.28	508, .04	3.39	555, .05	de (the)	576, .07	4.17	636, .05	ik (I)	575, .07
heft (handle or lifts)	3.11	610, .04	2.72	641, .20	het (the)	667, .07	3.67	604, .00	hij (he)	703, .20
hei (heath or ram)	3.72	608, .04	3.11	595, .05	de (the)	583, .07	2.50	654, .55	ik (I)	922, .53
kan (jug or can)	6.33	522, .04	4.83	559, .00	de (the)	529, .00	6.67	616, .05	ik (I)	513, .00
klink (doorhandle or sound)	4.28	570, .00	3.94	623, .05	de (the)	608, .07	4.72	653, .05	ik (I)	643, .13
krab (crab or scratch)	3.61	558, .00	3.00	660, .15	de (the)	681, .14	3.61	672, .10	ik (I)	678, .07
kust (coast or kisses)	5.00	543, .00	4.44	517, .05	de (the)	588, .07	5.67	581, .00	hij (he)	606, .00
last (burden or welds)	4.50	540, .00	4.56	569, .05	de (the)	546, .00	3.78	653, .05	hij (he)	678, .07
mat (mat or measured)	3.83	557, .00	4.11	578, .00	de (the)	570, .07	4.06	640, .15	hij (he)	677, .07
mist (fog or misses)	5.17	531, .04	3.94	534, .00	de (the)	550, .00	5.72	579, .00	hij (he)	593, .07
mol (mole or wreck)	3.72	552, .00	3.72	557, .00	de (the)	509, .00	3.61	641, .25	ik (I)	642, .07
poot (paw or plant)	5.11	576, .00	4.72	599, .00	de (the)	621, .00	3.33	725, .10	ik (I)	759, .13
port (port or prods)	3.39	626, .18	3.11	643, .30	de (the)	744, .27	2.44	781, .35	hij (he)	775, .20
rijst (rice or arises)	5.56	553, .00	4.78	536 , .00	de (the)	585, .00	3.28	653, .05	hij (he)	641, .07
schoot (lap or shot)	4.44	570, .00	3.72	575, .00	de (the)	654, .00	5.22	613, .05	hij (he)	638, .00
sla (lettuce or hit)	5.83	556, .00	5.06	530, .00	de (the)	593, .00	5.89	622, .05	ik (I)	554, .00
sloot (ditch or shut)	4.83	538, .00	4.11	549, .00	de (the)	579, .00	5.11	583, .00	hij (he)	608, .20
spant (rafter or stretches)	3.28	612, .00	1.94	854, .60	het (the)	762, .27	3.83	607, .05	hij (he)	669, .07
spel (game or spell)	5.78	529, .00	5.33	535, .05	het (the)	540, .00	4.94	624, .10	ik (I)	580, .00
staart (tail or stares)	4.72	542, .00	4.50	523, .10	de (the)	643, .07	4.06	644, .00	hij (he)	617, .07
staat (state or stands)	5.94	592, .00	4.94	566, .00	de (the)	771, .13	6.39	616, .00	hij (he)	595, .07
stal (stable or stole)	4.44	534, .00	3.78	563, .00	de (the)	595, .00	4.83	556, .00	hij (he)	596, .07
steeg (alley or rised)	4.00	555, .00	3.56	540, .00	de (the)	612, .00	5.28	643, .05	hij (he)	599, .07
steel (stem or steal)	4.83	567, .04	3.83	589, .00	de (the)	595, .13	4.89	616, .00	ik (I)	532, .00
stelt (stilt or puts)	5.11	596, .18	2.67	665, .10	de (the)	912, .33	6.00	642, .00	hij (he)	617, .27
vilt (felt or skins)	2.67	614, .14	2.28	665, .20	het (the)	805, .07	2.17	697, .00	hij (he)	728, .33
vocht (moisture or fought)	5.33	543, .00	4.28	529, .05	het (the)	581, .00	5.00	674, .10	hij (he)	572, .07
wond (wound or wound)	4.72	534, .00	4.94	550, .05	de (the)	644, .00	3.89	660, .10	hij (he)	681, .07
wreef (instep or rubbed)	3.78	646, .04	3.00	700, .10	de (the)	669, .27	4.61	676, .10	hij (he)	601, .00
zwam (fungus or drivel)	2.61	606, .09	2.67	669, .10	de (the)	695, .00	2.67	793, .35	ik (I)	756, .13

Effects of frequency, family size, and age of acquisition in a range of tasks

CHAPTER 6

This chapter is submitted as Nivja H. De Jong, Mandy Ghyselinck, Robert Schreuder, Marc Brysbaert, and R. Harald Baayen: Effects of Frequency, Family Size, and Age of Acquisition in lexical processing.

Abstract

We report a series of seven experiments in which we study three variables that are known to influence reaction times in visual word recognition: Age of Acquisition, word Frequency, and a word's morphological Family Size. The variables Frequency and Age of Acquisition have been studied and compared in numerous studies, but never taking into account a possible confound with Family Size. We compare the results of factorial analyses and those of regression analyses in seven different tasks, ranging from tasks tapping into early stages of processing to highly semantic tasks. We conclude that Age of Acquisition is the most robust variable in this range of seven tasks, but that Frequency as well as Family Size have independent predictive power. In the semantic tasks, although the total counts of Family Size cannot predict reaction times, the differential role for the morphological family members provides insight into the semantic processes at play.

Introduction

Throughout the history of psycholinguistic research, a number of variables have been proposed that would influence the speed of lexical processing in the visual modality. Among these are word length, orthographic neighborhood size, frequency of occurrence, age of acquisition, and morphological family size. In this paper, we study the effects of Frequency, Age of Acquisition, and Family Size, using seven different experimental tasks in the visual domain which range from tasks that presumably tap into early stages of lexical processing, to tasks that specially require semantic processing.

The Frequency with which a word occurs is probably the most studied and widely recognized variable: Words that occur often in corpora are easier to process than words with a low frequency of occurence. This effect has been found in a range of tasks, among which are word naming (Forster & Chambers, 1973; Monsell, Doyle, & Haggard, 1989), lexical decision (Forster & Chambers, 1973; Whaley, 1978), semantic classification (Landauer, Ross, & Didner, 1979), picture naming (Oldfield & Wingfield, 1965), and normal reading (Schilling, Rayner, & Chumbley, 1998).

Carrol and White (1973) were the first to introduce Age of Acquisition as a variable, in a study on picture naming. Pictures of objects whose names are acquired early in life were easier to name than pictures of objects whose names are late-acquired. In other tasks, Age of Acquisition has also been found to influence lexical processing, such as word recognition experiments (Gilhooly & Logie, 1981; Brown & Watson, 1987; Coltheart, Laxon, & Keating, 1988), speeded word naming (Gerhand & Barry, 1999a), lexical decision (Brysbaert, Lange, & Van Wijnendaele, 2000a; Gerhand & Barry, 1999b; Morrison & Ellis, 1995), word-associate generation (Van Loon-Vervoorn, 1989), and semantic categorization (Brysbaert, Van Wijnendaele, & De Deyne, 2000b). Morrison and Ellis (1995) even claimed that many previously observed effects of Frequency actually were effects of Age of Acquisition in disguise, as these two variables are highly correlated. In re-analyzing existing experiments and conducting new experiments in which they controlled for each of the two variables while contrasting the other, they found no effect for Frequency in naming latencies, but a strong effect of Age of Acquisition in both naming and lexical decision.

The recent view on effects of Age of Acquisition and Frequency is that they are both important predictors on lexical processing time. Both effects have been independently found in different tasks (Brysbaert, 1996; Brysbaert et al., 2000a; Brysbaert et al., 2000b; Gerhand & Barry, 1998, 1999a; Morrison & Ellis, 2000;

but see Barry, Hirsh, Johnston, & Williams, 2001). In a recent study, Ghyselinck, Lewis, and Brysbaert (submitted) systematically study the independent effects of Frequency and Age of Acquisition. They report experiments using a range of tasks involving perceptual identification, immediate and delayed naming, lexical decision with different kinds of pseudo-words, and a decision task in which participants had to decide on words versus first names. They conclude that although Age of Acquisition is the strongest predictor, both effects are apparent in the whole range of tasks. Furthermore, the magnitude of the effects are strongly correlated accross the different tasks. This suggests that Frequency and Age of Acquisition affect similar stages of processing.

The youngest and least known variable that is included in this study, is a word's morphological Family Size. The morphological family of a word consists of all words in which the target word occurs as a constituent. For instance, the word *man*, the word in English with the largest morphological family, occurs as a constituent in words such as *manly*, *snowman*, and *workmanlike*. There are two measures that can be obtained from this morphological family. First, the Family Size, which is a simple type count of the number of related words in which a word occurs. The second is a token count, the Family Frequency, which is the summed frequency of all these morphological related words. The Family Size of *man* is 270, and the Family Frequency is 506 per million (counts taken from the CELEX lexical database, Baayen, Piepenbrock, & Gulikers, 1995). Schreuder and Baayen (1997) first showed an effect of a word's Family Size in a visual lexical decision task, independently of the Frequency of the word. The larger a morphological family, the faster participants were able to respond.

This effect has been replicated for Dutch (De Jong, Schreuder, & Baayen, 2000, also Chapter 2), English (De Jong, Feldman, Schreuder, Pastizzo, & Baayen, 2002, also Chapter 3), German (Lüdeling & De Jong, 2002), Hebrew (Moscoso del Prado Martín, Deutsch, Frost, Schreuder, De Jong, & Baayen, in preparation), and a similar effect of semantic radical productivity has been found in Chinese (Feldman & Siok, 1997). If analyzed, in these studies Family Frequency did not affect reaction times, although in a recent study, Baayen, Tweedie, and Schreuder (2002) re-analyzed data from Schreuder and Baayen (1997), using more sensitive statistical techniques, and found a small but statistically reliable inhibitory effect of Family Frequency. This small inhibitory effect is probably related to effects of neighborhood inhibition, arising during early stages of processing (e.g., Grainger & Jacobs, 1996). The effect of Family Size, however, has been interpreted as a semantic effect on the basis of several indications. First, in a post-hoc correlational study, Bertram, Baayen, and Schreuder (2000) found that excluding semantically opaque family members from the count of Family Size improved correlations with reaction times. Second, De Jong et al. (2000, also Chapter 2) showed that the effect occurs both for regular formed Dutch past participles which share form with their family members (e.g., *geroeid*, ’rowed’, has family members such as *roeispaan*, ’oar’, and *roeier*, ’rower’) as well as for irregularly formed past participles that do not share the same stem with their family members (e.g., *gevochten*, ’fought’, has family members such as *gevecht*, ’(the) fight’ and *vechter*, ’fighter’). Third, minimal morphological and sentential context can narrow down the subset of family members which are contributing to the effect (De Jong, Schreuder, & Baayen, to appear, also Chapter 4). For instance, if an adjective such as *mooi*, ’beautiful’ occurs in isolation, all morphological family members contribute to the effect. But if morphological information such as the comparative suffix *-er* is added (constituting targets such as *mooier*, ’more beautiful’), it is only the adjectival family members which are contributing. In an experiment in which participants were asked to perform a double lexical decision, adding the word *heel*, ’very’, to the adjectives (e.g., *heel mooi*, ’very beautiful’) could even more drastically narrow down the subset that contributed to the effect of Family Size, namely to precisely those family members that share a similar meaning with the full (mini-)phrase (e.g., *supermooi* and *bloedmooi*, both meaning ’extremely beautiful’).

Ghyselinck et al. (submitted) claim that Age of Acquisition and Frequency affect the same levels of lexical processing, as the magnitude of the two effects throughout a series of experiments with tasks tapping into different levels of processing were highly correlated. However, in their analyses, they did not take into account the possible role of Family Size. Frequency and Family Size are highly correlated variables and it could be that the semantic effects they found for Frequency, actually were effects of Family Size. Similarly, Age of Acquisition and Family Size are likely to be correlated. Sofar, no study has taken into account the possible confound between these two variables. Studies reporting on effects of Age of Acquisition have not taken into account the role of Family Size, and vice versa. To illustrate the collinearity, Figure 6.1 shows the correlational structure of 2654 Dutch monomorphemic nouns for which both Age of Acquisition as well as Frequency and Family Size measures are available (Ghyselinck, De Moor, & Brysbaert, 2000; CELEX lexical database, Baayen et al., 1995). The correlation between log Base Frequency (the summed frequency of all inflectional variants) and log Family Size (first panel) is $r = 0.800$ ($t(2652) = 68.62, p = 0.000$), between Age of Acquisition and log Base Frequency (second panel) $r = -0.653$ ($t(2652) = -44.36, p = 0.000$), and between log Family Size and Age of Acquisition (third panel) $r = -0.598$ ($t(2652) = -38.42, p = 0.000$).

In this study, we will report on seven experimental tasks in which we investigated the role of these three variables using factorial designs. The first design contrasted Age of Acquisition (while matching on Frequency and Family Size), the second contrasted Frequency (while matching on Family Size and Age of Acquisition), and the third contrasted Family Size (while matching on Frequency and Age of Acquisition). The seven tasks that we use presumably tap into different stages of processing. We used a Perceptual Identification task and a Lexical Decision task with illegal nonwords (length-matched strings of consonants) to tap into early stages of processing, a Lexical Decision task with phonotactically legal nonwords and a Rating task to presumably tap into intermediate stages of processing, and finally an Association task and a Semantic Decision task to tap into highly semantic stages of processing, and finally we used a Lexical Decision task with newly coined derivations to see whether the effects would be sensitive to morphological processes.

In what follows, we first report the seven Experiments, and report the results of the factorial analyses per experiment. We then proceed by discussing these results of the experiments. Finally, we offer by-item regression analyses on all experimental results, after assessing in a simulation that the collinearity in our data set is not problematic for such multiple regression analyses.

Experiment 6.1: Perceptual Identification

Method

Participants. Twenty-two participants, mostly undergraduates at Ghent University, were paid to take part in this experiment. All were native speakers of Dutch.

Age of Acquisition materials. We used the CELEX lexical database for Frequency and Family Size information (Baayen et al, 1995) and the Age of Acquisition norms of 2816 monomorphemic nouns collected by Ghyselinck et al. (2000). These norms were collected from 559 undergraduate students who rated at which actual year they first learned the words, and were validated with children's knowledge by De Moor, Ghyselinck, and Brysbaert (2000). We selected 54 monomorphemic nouns. Twenty-seven of these nouns had an early Age of Acquisition according to these

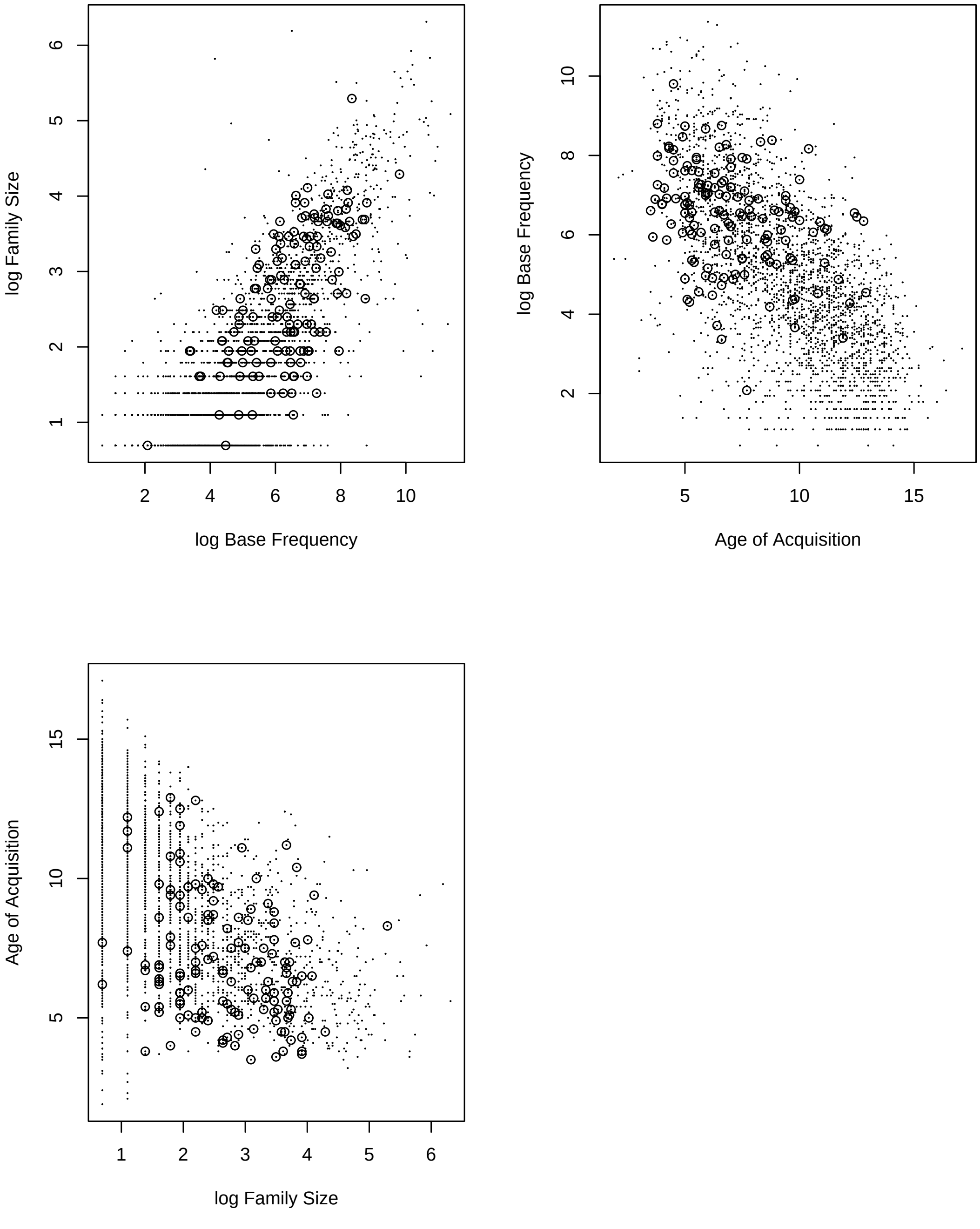

Figure 6.1: Scatterplots of Age of Acquisition, log Base Frequency, and log Family Size for 2654 Dutch monomorphemic nouns. Encircled dots represent the 141 items used in this study.

ratings (mean 4.6, SD 0.6), and 27 were words with a late Age of Acquisition (mean 9.7, SD 1.5). The early and late groups of words were matched on Surface Frequency (the frequency of the singular, mean 16.8, SD 21.0, and mean 16.4, SD 23.8; all frequency counts standardized per million), Base Frequency (mean 21.8, SD 25.2, and mean 28.5, SD 34.6), Family Size (mean 23, SD 38, and mean 21, SD 15), and on mean length in letters (4.3 and 4.6). In both conditions, 8 words represented objects made by nature, such as *appel* ('apple'), and 19 represented objects which are man-made (e.g., *tafel*, 'table').

Family Size materials. We selected 46 monomorphemic nouns from the CELEX lexical database which were also rated on Age of Acquisition in Ghyselinck et al. (2000). Twenty-three of these nouns had a high Family Size (mean 35, SD 11), and 23 had a low Family Size (mean 7, SD 3). The two groups were matched for Surface Frequency (mean 16.3, SD 9.1, and mean 15.8, SD 8.8), Base Frequency (mean 25.3, SD 13.8, and mean 23.6, SD 13.6), Age of Acquisition (mean 6.6, SD 1.6, and mean 6.4, SD 1.8), and on mean length in letters (4.5 and 4.5). In both conditions, 10 words represented objects made by nature, and 13 represented man-made objects.

Frequency materials. We selected 72 monomorphemic nouns from the CELEX lexical database which were also attested in the ratings on Age of Acquisition reported in Ghyselinck et al. (2000). Thirty-six of these nouns had a high Surface and Base Frequency (mean 49.3, SD 51.7, and mean 65.4, SD 74.1) and 36 had a low Surface and Base Frequency (mean 4.1, SD 4.9, and mean 7.1, SD 6.9). The two groups of words were matched on Age of Acquisition (mean 7.1, SD 2.3, for the high condition and mean 6.9, SD 2.2 for the low condition), Family Size (mean 20, SD 17, and mean 16, SD 14), and on mean length of letters (4.4 and 4.4). In the high condition, 25 words represented objects made by man, and 11 words represented nature-made objects. In the low condition, these numbers were 24 and 12 respectively.

For all experiments, we combined the materials of all three contrasts, such that for instance the targets for the Frequency contrast would serve as the fillers for the Age of Acquisition contrast. Some words served as targets in more than one contrast, hence, the total number of words in the experimental list was 141 (instead of $54 + 46 + 72 = 172$). The materials are listed in Appendices A to F.

Procedure. Participants were tested in noise-proof experimental rooms. They were asked to decide as accurately as possible to identify which word appeared on the computer screen. Each stimulus was preceded by a fixation mark in the middle of the screen for 500 ms. After 500 ms, the stimulus appeared at the same position. Stimuli were presented on a 14-inch color monitor in white lowercase courier letters on a dark background and they remained on the screen for 32 ms. The stimulus was followed by a mask (######) which covered the targets completely. Participants responded orally. The ITI was 500 ms. The experiment was preceded by 20 practice items, and the stimuli were randomised for each subject.

Results

The first row of Table 6.1 shows the mean error percentages of the conditions in all three experimental contrasts. Words with an early Age of Acquisition were recognized correctly more often than the words with a late Age of Acquisition, although not reliably so in the by-item analysis ($t1(21) = 2.82, p = 0.0102; t2(52) = 1.65, p = 0.1054$). Neither the contrast in Family Size ($t1, t2 < 1$), nor the contrast in Frequency, ($t1(21) = -1.37, p = 0.1854; t2 < 1$) elicited different error percentages.

Experiment 6.2: Lexical Decision with illegal nonwords

Method

Participants. 20 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. The materials were identical to those of Experiment 6.1.

Procedure. Every word was paired with a nonword, which consisted of a string of consonants. Participants were tested in noise-attenuated experimental rooms. They were asked to decide as quickly and accurately as possible whether a letter string appearing on the computer screen was a real Dutch word. Each stimulus was preceded by a fixation mark in the middle of the screen for 50 ms. After 500 ms, the stimulus appeared at the same position. Stimuli were presented on Nec Multisync color monitors in white lowercase 36 point Helvetica letters on a dark background

Table 6.1: Results of Experiments 6.1 to 6.7: means of Error percentages (Experiment 6.1), harmonic means of response latencies and means of error percentages (Experiments 6.2, 6.3, 6.5, 6.6, and 6.7), and means of ratings (Experiment 6.4). The asterisks in the columns labeled S show the significant effects ($\alpha = 0.05$, by subjects and by items).

Experiment		Age of Acquisition			Frequency			Family Size		
		Early	Late	S	High	Low	S	High	Low	S
6.1 (PerId)	Error	11.4	17.2		14.3	17.2		14.2	15.5	
6.2 (Ill)	RT	445	459	*	447	456		446	452	
	Error	1.1	2.0		1.5	2.2		1.9	1.1	
6.3 (Leg)	RT	521	573	*	533	563	*	534	534	
	Error	2.2	5.2	*	2.4	4.6		1.9	1.5	
6.4 (Rating)	rating	5.1	3.2	*	5.0	3.7	*	4.3	4.4	
6.5 (Assoc)	RT	1338	1508	*	1444	1443		1439	1322	*
	Error	1.0	4.6	*	1.4	2.2		2.8	1.0	
6.6 (Semdec)	RT	824	846		815	825		809	805	
	Error	7.3	5.0		4.9	4.7		4.3	6.9	
6.7 (Suffixed)	RT	750	816	*	772	785		737	771	*
	Error	17.6	14.6		13.7	15.4		9.2	11.2	

and they remained on the screen for 1500 ms. The maximum time span allowed for a response was 2000 ms from stimulus onset. We randomized the items three times, and presented each random order both ways, creating six experimental lists. The experiment was preceded by 16 practice items.

Results

All participants performed the experiment with an error rate lower than 4%. The second row of Table 6.1 shows the harmonic means of the reaction times (calculated over the correct responses; we use harmonic means to deal with reaction time outliers, see Ratcliff, 1993) and the mean error scores (calculated over all responses). The words with an early Age of Acquisition were responded to faster than words with a late Age of Acquisition ($t1(19) = 3.36, p = 0.003; t2(52) = 2.14, p = 0.037$). However, the Family Size contrast ($t1(19) = -1.50, p = 0.150, t2 < 1$) did not elicit a difference in reaction times. Words with a high Frequency were responded to faster than words with a low Frequency, albeit only marginally so ($t1(19) = -2.00, p = 0.0596; t2(70) = -1.83, p = 0.0721$). There were no significant differences in the error scores in any of the three contrasts (all $t1$'s < 1.5; all $t2$'s ≈ 1).

Experiment 6.3: Lexical Decision with legal nonwords

Method

Participants. 20 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. The materials were identical to those of Experiment 6.1.

Procedure. The procedure was identical to that of Experiment 6.3, except that every word was paired with a pseudoword with legal Dutch phonotactics.

Results

All participants performed the experiment with an error rate lower than 14%. The third row of Table 6.1 shows the harmonic means of the reaction times (calculated over the correct responses) and mean error scores (calculated over all reponses) of the experimental conditions of all three experimental contrasts. Only the contrast in Age of Acquisition and Frequency elicited a significant difference in reaction times (Age of Acquisition: $t1(19) = 7.86, p = 0.000; t2(52) = 4.03, p = 0.000$; Frequency: $t1(19) = -6.65, p = 0.000; t2(70) = -2.77, p = 0.007$) The reaction times of the Family Size contrast were not different ($t1, t2 < 1$). Similarly, there was an effect in the error scores for the Age of Acquisition ($t1(19) = 2.56, p = 0.019; t2(52) = 2.11, p = 0.040$) and Frequency contrast ($t1(19) = -2.44, p = 0.025; t2(70) = -1.72, p = 0.089$), albeit only marginially significant in the by-item analysis for the Frequency contrast. For the Family Size contrast, no significant differences were found ($t1, t2 < 1$).

Experiment 6.4: subjective frequency Rating

Method

Participants. 18 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. The materials were identical to those of Experiment 6.1.

Procedure. Participants were asked to indicate on a seven-point scale how frequently they thought a word was used in current Dutch. We randomized the items three times, and presented each random order both ways, creating six experimental lists.

Results

The fourth row of Table 6.1 shows the mean ratings of both conditions in all three experimental contrasts. Again, words with an early Age of Acquisition were rated higher than words with a late Age of Acquisition ($t1(17) = -15.27, p = 0.000; t2(52) = -6.38, p = 0.000$). Similarly, words with a high Frequency were rated higher than words with a low Frequency ($t1(17) = 13.30, p = 0.000; t2(70) = 4.01, p = 0.000$). The contrast in Family Size did not results in different ratings ($t1(17) = -1.68, p = 0.111; t2 < 1$).

Experiment 6.5: Semantic Association generation

Method

Participants. T participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch. wenty undergraduate students from Ghent University participated for course credits. All were native speakers of Dutch.

Materials. The materials were identical to those of Experiment 6.1.

Procedure. Participants were tested in noise-proof experimental rooms. They were asked to name as quickly as possible a semantic associate of the word appearing on the computer screen. Each stimulus was preceded by a fixation mark in the middle of the screen for 500 ms. After 500 ms, the stimulus appeared at the same position. Stimuli were presented on a 14-inch color monitor in white lowercase courier letters on a dark background and they remained on the screen for 300 ms. A voicekey response box registrated the response times and the experimentator registrated the semantic associates. The maximum time span allowed for a response was 5000 ms from stimulus onset. We randomized the items for each subject. The ITI was 1000 ms. The experiment was preceded by 20 practice items.

Results

Problems with the voicekey, no responses, and responses beyond 5 seconds were counted as errors. One participant performed the experiment with an error rate of more than 20%. Five items (three from the Frequency contrast, and two from the Age of Acquisition contrast) had an error rate of 30% or more. These data points were excluded from further analyses. The fifth row of Table 6.1 shows the harmonic mean reaction times (calculated over the correct responses) and mean error scores (calculated over all responses) for the conditions in all three contrasts. Words with an early Age of Acquisition elicited faster reaction times than words with a late Age of Acquisition ($t1(18) = 4.8, p = 0.000; t2(50) = 3.19, p = 0.003$). The contrast in Frequency did not result in a significant difference in reaction times ($t1, t2 < 1$). The words with a high Family Size, however, were responded to slower than the words with a low Family Size ($t1(18) = 6.03, p = 0.000; t2(44) = 2.34, p = 0.024$). For the analyses on the error scores, we only considered no reponses and responses beyond five seconds. For Age of Acquisition, there was a significant effect in the error scores ($t1(18) = 2.82, p = 0.011; t2(52) = 2.24, p = 0.032$). The contrast in Frequency did not show a difference in error scores ($t1(18) = -1.28, p = 0.214; t2 < 1$). For the Family Size contrast, somewhat more errors were made if the morphological family was large, although only significant in the by-subject analysis. ($t1(18) = 2.16, p = 0.044; t2(44) = 1.32, p = 0.197$).

Experiment 6.6: Semantic Decision

Method

Participants. T participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch. wenty-five undergraduate students from Ghent University were paid to take part in this experiment. All were native speakers of Dutch.

Materials. The materials were identical to those of Experiment 6.1.

Procedure. Participants were tested in noise-proof experimental rooms. They were asked to decide as quickly and accurately as possible whether a word appearing on the computer screen was a man-made object or a nature-made object. Each stimulus was preceded by a fixation mark in the middle of the screen for 500 ms. After 500 ms, the stimulus appeared at the same position. Stimuli were presented on a 14-inch color monitor in white lowercase courier letters on a dark background and they remained on the screen until a response was given. The maximum time span allowed for a response was 10000 ms from stimulus onset. The ITI was 1500 ms. For half of the subjects, the left push button was the ’man-made object’ button, and for the other half the ’nature-made object’ button. We randomized the items for each subject. The experiment was preceded by 20 practice items.

Results

All participants performed the Experiment with an error rate lower than 16 %. Three items (one from the Age of Acquisition contrast, one from the Family Size contrast, and one item both in the Family Size and Frequency contrast) elicited error percentages of over 30%. These items were excluded from further analyses. The sixth row of Table 6.1 shows the harmonic means of the reaction times (calculated over the correct responses) and the mean error percentages (calculated over all responses) for the conditions in all three experimental contrasts. None of the contrasts elicited a reliable difference in reaction times (Age of Acquisition contrast: $t1(24) = 1.37, p = 0.183; t2(51) = 1.19, p = 0.239$, Family Size contrast: $t1, t2 < 1$, Frequency contrast: $t1(24) = -1.02, p = 0.317; t2 < 1$). Likewise, no reliable differences were found in the error scores, although trends were found in the by-subjects analyses (Age of Acquisition contrast: $t1(24) = -1.89, p = 0.070; t2 < 1$, Family Size contrast: $t1(24) = -1.93, p = 0.065; t2 < 1$, Frequency contrast: $t1(24) = -2.10, p = 0.047; t2 < 1$).

Experiment 6.7: Lexical Decision with novel derivations

Method

Participants. 20 participants, mostly undergraduates at Nijmegen University, were paid to take part in this experiment. All were native speakers of Dutch.

Materials. The materials were identical to those of Experiment 6.1, except that we added a suffix to all the nouns, creating new, possible words. We used 12 different suffixes and chose for each word a suffix which would create a possible and eligible new word. For instance, for the word *wolk* ('cloud'), we used the suffix *-vormig* ('-shaped') to create the word *wolkvormig* ('cloud-shaped'). In most of the cases, we were indeed able to create a new word which was not attested in the CELEX lexical database. But for six words we had to use a word with a (low) frequency in the CELEX database. The mean surface frequency of these six words was 0.1 per million. The most frequently used suffixes were *-achtig* ('-like'), *-vormig* ('-shaped'), and *-loos* ('-less'). The materials are listed in Appendices A to F.

Procedure. The procedure was identical to that of Experiment 6.3, except that we now asked our participants to decide as quickly and accurately as possible whether a letter string appearing on the computer screen was a possible Dutch word. The pseudo words in this experiment consisted either of an existing base word with a pseudo-suffix, or of a pseudo-base with an existing suffix.

Results

Three participants performed the experiment with an error rate of 20% or more. Fourteen items of the Age of Acquisition contrast, 8 of the Family Size contrast, and 17 items of the Frequency contrast had an error rate of 30% or more. These data points were excluded from the analyses. The last row of Table 6.1 shows the harmonic means of reaction times (calculated over the correct responses) and error scores (calculated over all responses) of the experimental conditions of the three experimental contrasts. Despite the large number of items that was now excluded from the reaction time analyses, the matching in the three contrasts was hardly affected.

Words with an early Age of Acquisition were responded to faster than words with a late Age of Acquisition ($t1(16) = 5.80, p = 0.000; t2(38) = 3.58, p = 0.001$). The contrast in Family Size also resulted in a significant difference in reaction times: Words with a high Family Size were responded to faster than words with a low Family Size ($t1(16) = -34.06; p = 0.010; t2(36) = -2.16, p = 0.038$). However, no significant differences were found for the words with a high and low Frequency ($t1, t2 < 1$). No significant effects were found in the error analyses (Age of Acquisition contrast: $t1(16) = 1.48, p = 0.159; t2 < 1$; Frequency contrast: $t1, t2 < 1$; Family Size contrast: $t1(16) = -1.12, p = 0.280, t2 < 1$).

Discussion of the Results

All experiments taken together point to Age of Acquisition as the most robust variable; its effect is only absent in the Perceptual Identification task (in which it was significant by subjects) and in the Semantic Decision task. Note that in these two tasks, the Frequency and Family Size contrasts did not result in significantly different reaction times either.

Compared to Age of Acquisition, Frequency proved to be not a very strong variable, only significant in the Lexical Decision task with phonotactically legal nonwords, and in the Rating task. Likewise, the contrast for Family Size proved to be significant in only two tasks: inhibitory in the Association task, and facilitatory in the Lexical Decision task with novel derived forms.

From previous results, we would have expected to find effects for Age of Acquisition in the Perceptual Identification task (Ghyselinck et al., submitted) and the Semantic Decision task (Brysbaert et al., 2000b). We also expected to find Frequency effects in the Perceptual Identification task (Ghyselinck et al., submitted), in the Lexical Decision task with illegal nonwords (Ghyselinck et al., submitted; Stone & Van Orden, 1993) and perhaps in the Lexical Decision task with novel derivations. Our predictions for the effect of Family Size are also not born out. As previous studies report significant effects of Family Size in Lexical Decision tasks and Rating tasks, and as we argued Family Size to be a semantic variable, we expected to find significant effects for Family Size in the Lexical Decision task, the Rating task, and in the Semantic Decision task.

A possible explanation for the absence of effects for Family Size in for instance the Rating and Lexical Decision tasks is that previous reported effects of Family Size in these tasks actually were Age of Acquisition effects in disguise, as previous studies investigating Family Size did not control for Age of Acquisition. It seems premature, however, to conclude that effects of Family Size in previous studies were due to a confound with Age of Acquisition, for two reasons. First, it is unclear how the sensitivity to context of the Family Size effect as reported by De Jong et al. (to appear, also Chapter 4) can be explained if the effect actually was an effect of Age of Acquisition. Second, the power of the three designs in the present study is not comparable. Besides the difference in number of items of the three factorial designs (54 for Age of Acquisition, 72 for Frequency, and 46 for Family Size), there is a big diffence in the size of the contrasts. As it is log Frequency and log Family Size, rather than the absolute counts, that correlate with reaction times, their contrasts effectively consist of 2.3 and 1.4 log-units respectively. The effective contrast for

Age of Acquisition, however, is much larger: The mean Age of Acquisition differs 5 years between the two subsets.[1] The larger the contrast for a given variable, and the greater the number of items in a specific contrast, the easier it will be to find a significant effect for that variable. At the same time, it should be kept in mind that the factorial contrasts we designed are for all three variables the most powerful contrasts we could find. This suggests that given the current data for the Age of Acquisition of Dutch words (2654 four- and five-letter nouns from Ghyselinck et al. (2000), for which also CELEX-information is available), it is impossible to find an effect for Family Size using factorial contrasts in tasks such as Lexical Decision and Rating.

A possibility to increase the power (for all three variables), is to include all items of the three contrasts together in a regression analysis. A well-known problem with such variables in regression analyses is collinearity. The reason we were not able to increase the power of our factorial designs, is that items which contrast substantially in one dimension, will also contrast substantially the other dimensions. Consider again Figure 6.1, which shows the correlational structure of the 2654 nouns for Age of Acquisition, log Base Frequency, and log Family Size. The circles around the dots in Figure 6.1 represent the items of the three factorial contrasts in the present study. The correlation for these 141 items between log Base Frequency and log Family Size (first panel) is $r = 0.617$ ($t(139) = 9.23, p = 0.000$), between Age of Acquisition and log Base Frequency (second panel) $r = -0.344$ ($t(139) = -4.32, p = 0.000$) and between log Family Size and Age of Acquisition (third panel) $r = -0.290$ ($t(139) = -3.57, p = 0.000$). The collinearity index number (Belsley, Kuh, & Welsch, 1980) is rather high as well, 18.11. Given these correlational structures in our data matrix, conclusions based on linear regression models seem unwarranted, as these methods might be unable to tease out the effects of such highly correlated variables.

In order to gauge the possibility of spurious effects due to the collinearity in our data matrix, we carried out simulation studies. In 1000 simulation runs, we created data sets with the same characteristics as in our empirical data set. For each of these runs, we obtained simulated reaction times that were modeled from the empirical reaction times in the Lexical Desicion task with legal nonwords, in which the factorial analyses found an effect for Age of Acquisition only. The simulated reaction times were dependent on the three variables that mimicked Age of Ac-

---

[1] Note that these comparisons in contrasts are based on the assumption that the relevant unit for Age of Acquisition should be measured as the mean age in years (and not, for instance, months) at which a word is first learned.

quisition, Base Frequency, and Family Size closely, or on only two or one of these variables. In addition, to compare the power and number of Type I errors for such regression analyses with that of factorial analyses, we constructed for each randomly generated data set three factorial designs, and carried out factorial analyses. We found that with variables comparable in characteristics to our original data set (equal means and variances, similar collinearity and pair-wise correlations), regression analyses are much more powerful than factorial analyses, especially in picking up effects of a variable with small variance. The number of spurious effects was acceptable, and around the range of what can be expected using an alpha level of 0.05. In Appendix G, we report on these simulation studies. In what follows, we turn to the regression analyses on our empirical data of the seven experiments.

Regression analyses

We carried out regression analyses, using all items.[2] Table 6.2 shows the β-weights (co-efficients in the linear models) and t-values for the three variables Age of Acquisition, log Base Frequency, and log Family Size in the six reaction time Experiments. For the error analyses, we used logistic regression models, with the logit of the number of participants that made an error and the number of participants that responded correctly as a linear function of the same three independent variables; We also report the β-weights and corresponding z-values. A quick comparison of the columns labeled *S* in Tables 6.1 (asterisks marking the significant effects for the factorial contrasts) and 6.2 (asterisks marking the significant effects in the regression models) reveals that the linear models come up with significant effects more often than the factorial analyses.

First consider the variable Age of Acquisition. We now also find effects in the Perceptual Identification task, in the error analysis of the Semantic Decision and in Lexical Decision task with novel derivations. The effect is only absent in the error analysis of the Lexical Decision task with illegal nonwords (in which very few errors were made), and in the reaction time analysis of the Semantic Decision task. In summary, in seven different tasks, ranging from tasks tapping into early perceptual processes to semantic tasks, we now find effects of Age of Acquisition.

The linear regression models do not reveal many significant effects of Frequency, just as in the factorial analyses: Only in the reaction time analysis of the Lexical

---

[2] Additional simulations not reported in Appendix G showed that selecting 141 items in the regression analyses rather than the 40 as reported on did not change the results.

Table 6.2: Results of Experiments 6.1 to 6.7: β-weights, t-values for the linear models using items with error rates under 30% (on harmonic by-item means), and z-values for the logistic regression models using all items (error-data), with Age of Acquisition, Base Frequency, and Family Size as independent variables. The asterisks in the columns labeled S show the significant effects ($\alpha = 0.05$).

Experiment		Age of Acquisition			Frequency			Family Size		
		β	t/z	S	β	t/z	S	β	t/z	S
6.1 (PerId)	Error	0.08	3.66	*	-0.02	-0.33		-0.20	-2.82	*
6.2 (Ill)	RT	4.40	4.62	*	-1.33	-0.66		-3.02	-1.03	
	Error	0.08	1.23		-0.13	-0.90		-0.00	-0.01	
6.3 (Leg)	RT	10.70	7.13	*	-7.61	-2.40	*	-12.12	-2.62	*
	Error	0.10	2.11	*	-0.12	-1.21		-0.34	-2.07	*
6.4 (Rating)	rating	-0.25	-7.56	*	0.34	4.78	*	0.23	2.29	*
6.5 (Assoc)	RT	26.41	3.47	*	-16.09	-1.02		31.48	1.35	
	Error	0.30	4.91	*	-0.12	-0.82		0.25	1.20	
6.6 (Semdec)	RT	2.09	0.57		-7.76	-1.01		-4.33	-0.38	
	Error	-0.11	-2.80	*	-0.04	-0.53		-0.03	-0.26	
6.7 (Suffixed)	RT	18.85	7.12	*	-10.32	-1.95		-16.09	-2.13	*
	Error	0.06	2.76	*	-0.14	-2.87	*	0.04	0.63	

Decision task, in the Rating task, and in the error analysis of the Lexical Decision task with novel derivations do we find an independent effect of Frequency.

For Family Size, surprisingly, in one task in which the factorial analysis showed a significant effect, the effect is not significant in the regression analysis. The inhibitory effect in the Semantic Association task has disappeared. At the same time, we now do see a significant effect for Family Size in the Perceptual Identification task, in the Lexical Decision task with legal nonwords (both in the reaction time and error data), and in the Rating task. Unlike the effect of Age of Acquisition, and contrary to our expectations, the effect of Family Size is significant in a task that involves early processes of lexical identification, rather than in tasks predominantly requiring semantic processing. However, upon closer examination we find that the morphological family plays a role on specific aspects of the Semantic Association and Semantic Decision tasks.

Consider the possible role of the morphological family for the Semantic Association task. In this task, participants are asked to name a semantically related word as soon as possible. Following the examples they have been given in the task description, participants are discouraged to name a morphological relative. Despite the task description and examples, participants do occasionally come up with a word like *taart* ('pie') upon reading *appel* ('apple'), or even with *appeltaart* ('apple pie'). In fact, of all 2413 correct responses, 595 were responses in which either a family member, or part of a family member of the target word were named. Interestingly, in a logistic regression model predicting for each item the logit of the number of participants that responded with such a family member and the number of participants that did not, both Family Size and Frequency were highly significant variables (Age of Acquisition: $\beta = -0.01, z = -0.60, p = 0.547$; Frequency: $\beta = -0.45, z = -8.92, p = 0.000$; Family Size: $\beta = 0.79, z = 10.25, p = 0.000$). These β-values can be interpreted as follows: the larger the morphological family, the more likely participants are to respond with a morphological family member. At the same time, the higher the frequency of the target word, the less likely participants are to respond with a morphological family member. Possibly, high frequency targets tend to have higher frequency semantically related words and morphologically complex words are lower in frequency than monomorphemic words. Indeed, for the responses that were not morphologically related, there was a significant positive correlation between the frequency of the target, and the frequency of the response ($r = 0.364, t(139) = 4.5, p = 0.000$). In the responses of this experiment, the mean frequency of the mophologically related responses was 1.1 per million, and the responses of the morphologically unrelated responses was 195 per million. As morphologically complex words are lower in frequency than monomorphemic words are, and we found that semantically related words tend to be higher in frequency as the frequency of the target word increases, it now follows that morphological family members are more likely to become competitors with monomorphemic words if the frequency of the target is relatively low. Interestingly, the Age of Acquisition of the target words does not play a role in this competitive selection process.

Turning to the Semantic Decision task, it turns out that despite the null-effect in both the factorial as well as the regression analyses, the morphological family actually does play an interesting role. In this task, participants had to decide whether the word they read is an object made by man or made by nature. If the morphological family plays a role in this decision process, the meanings of these morphological family members are important, rather than a simple count of all family members irrespective of their meaning. Take as an example the morphological family of *appel*. Among the family members of *appel* are *appelrond* ('round as an apple'), *appelboom* ('apple tree'), *appeltaart* ('apple pie'), and *appelmoes* ('applesauce'). Upon deciding whether *appel* is an object made by man or made by nature, these different family members may play different roles. First, only objects within the family can be considered as relevant family members. For example, deciding on the noun *appel*, the semantics of a morphological relative such as the adjective *appelrond* are not relevant. Therefore, we counted for all family members the number of relevant family members (objects, hence nouns) and the number of irrelevant family members (mostly qualities and activities, hence adjectives and verbs). Second, consider the relevant family members (the nouns) of *appel*. Among those relevant family members, some are congruent with the decision that has to be made, in other words, they are nature-made objects themselves (e.g., *appelboom*). Other relevant family members of *appel*, on the other hand, are incongruent with the correct decision, they are man-made objects (e.g., *appeltaart* and *appelmoes*). For all words in our experiment, we obtained two counts that expressed the relevance and congruency within the family. For Relevance, we counted the number of relevant family members and the number of irrelevant family members. Subsequently, we obtained our measure Relevance for each item by taking the log of this proportion. In our (incomplete) example of *appel*, the Relevance would be $log(3/1)$. We calculated Congruency in the same manner, by taking the log of the proportion of congruent over incongruent family members ($log(1/2)$ in the example of *appel*). Taking as an example the count of Relevance, the log of such a proportion actually captures both the count of relevant family members, as well as the count of irrelevant family members, but with opposite signs, as $log(3/1)$ equals $log(3) - log(1)$.[3]

A linear regression model showed that Relevance, but not Congruency, co-determined the response latencies (Relevance: $\beta = -21.48, t = 2.59, p = 0.011$; Congruency: $\beta = 1.21, t = 0.20, p = 0.839$). For the error data, a logistic regression model showed a significant effect for Congruency in addition to Age of Acquisition, whereas the effect of Relevance was only marginally significant (Age of Acquisition: $\beta = -0.08, z = -2.05, p = 0.040$; Relevance: $\beta = -0.15, z = -1.77, p = 0.076$; Congruency: $\beta = -0.377, t = -6.47, p = 0.000$). We interpret these effects of the morphological family as support for the hypothesis that upon reading a word, semantic activation in the mental lexicon spreads along morpho-semantic lines. In fact, these results suggest that this spreading of activation has consequences for the meaning of the word itself. For a word with low Relevance within the family (with respect to deciding on objects), it will be more time-consuming to decide whether the word is an object made by man or made by nature. Furthermore, if the Congruency

---

[3] In fact, we added one to all counts, in order to avoid taking the log of or dividing by zero.

within the family is low (with respect to deciding on man-made or nature-made), participants are more likely to make an error. In other words, a word representing a nature-made object may be less like a prototypical nature-made thing if many of its family members are incongruent with this decision, i.e., if they represent man-made objects.

To summarize the effects of the morphological family, we find that it plays a role in six different experimental tasks, absent only in the Lexical Decision task with illegal nonwords. In most of these experiments, it is the total number of family members, the Family Size, that plays a facilitatory role in some aspect of lexical processing. In the Semantic Association task, however, the Family Size co-determines the actual response. For words with large families, participants are more likely to generate (part of) a morphological family member. In the Semantic Decision task, the word category of the morphological family members and their meaning with respect to the decision turned out to be crucial in both reaction time and error data.

General Discussion

There are many studies investigating the independent effects of Frequency and Age of Acquisition. Similarly, there are studies investigating the independent effects of Frequency and Family Size. However, there is no evidence sofar that all three variables have independent effects, as the studies investigating Age of Acquisition and Frequency did not control for Family Size, and the studies concerned with Family Size and Frequency did not control for Age of Acquisition. In this paper, we attempted to fill in this gap. In seven different experimental tasks, we systematically investigated the effects of these variables in lexical processing, using the visual modality.

We first constructed three factorial designs, in which either Age of Acquisition, Frequency, or Family Size were contrasted, while matching on the other variables. The tasks we employed were Perceptual Identification, Lexical Decision with phonotactically illegal nonwords (strings of consonants), Lexical Decision with phonotactically legal nonwords, subjective frequency Rating, Semantic Association generation, Semantic Decision (man-made or nature-made object), and Lexical Decision with newly coined derivations. The factorial analyses pointed to Age of Acquisition as the most robust variable, with occasionally significant effects for Frequency or Family Size. The factorial contrasts, however, were hardly comparable, as the contrast for Age of Acquisition was larger than the other two. We therefore carried out regression analyses, combining all items from the three designs. Prior to these analyses, we assessed in a simulation study that regression analyses on randomly generated datasets with very similar characteristics as our original dataset are not more prone to give rise to spurious effects due to collinearity than factorial analyses. Furthermore, this simulation study showed that regression analyses on three graded, correlated variables are very much more powerful. Finally, performing regression analyses on variables that are graded, we claim, makes more sense than carrying out factorial analyses on designs that represent artificially constructed dichotomies. We therefore decided to base our conclusions on the results of the regression analyses of our experimental data sets.

Both Frequency and Age of Acquisition have been reported to influence performance and reaction times in very different tasks (Brysbaert, 1996; Brysbaert et al., 2000a; Brysbaert et al., 2000b; Gerhand & Barry, 1998, 1999a; Ghyselinck et al., submitted; Morrison & Ellis, 2000; but see Barry, Hirsh, Johnston, & Williams, 2001). In line with these previous studies, we observed effects of Age of Acquisition in the full range of tasks (Perceptual Identification, Lexical Decision with illegal nonwords, Lexical Decision with legal nonwords, Rating, Semantic Association, Semantic Decision, and the Lexical Decision task with new derivations).

The effect of Frequency proved to be less strong, effective only in the Lexical Decision with legal nonwords, the Rating, and the Lexical Decision with new derivations. In six of the seven tasks, on the other hand, some measure of the morphological family was a significant predictor, including the Perceptual Identification task. In line with previous studies (e.g., Bertram et al., 2000; Schreuder & Baayen, 1997), the effect was significant in the Lexical Decision with legal nonwords, and in the Rating. In the Semantic Association task the Family Size co-determined the actual response (large families resulting in morphologically related words as a response), which supports the hypothesis that the effect of Family Size is semantic in nature (Bertram et al., 2000; De Jong et al., 2000, also Chapter 2; De Jong et al., to appear, also Chapter 4). Similarly, the very specific role of the morphological family members in the Semantic Decision task, in which their word category and meaning with respect to the decision co-determined response latencies and errors respectively, provides strong support for the hypothesis that the morphological family affects lexical processing at the semantic level. The effect of Family Size in the Lexical Decision with novel derivations also does not come as a surprise, as this task probably involves parsing the targets into stem and suffix, which we think might be easier when the stem is very productive (cf., has many family members;

see also Feldman, Pnini, & Frost, 1995).

The effect of Family Size in the Perceptual Identification task, however, was not expected. Using another task that presumably also taps into early stages of processing, Progressive Demasking, Schreuder and Baayen (1997) did not find an effect for Family Size. Possibly, the task used in the present study, Perceptual Identification, is more sensitive and therefore more likely to pick up on any effect.

The present results suggest that Family Size may be effective in early, more formal stages of processes in addition to semantic levels of processing, contrary to our expectations. Another explanation of the present results would be that semantic processes do play a role in the Perceptual Identification task. In order for participants to have recognized the word completely, they also need to have recognized at least some of the semantics. In this reasoning, the semantic processes that follow the very short presentation may help participants to identify the word and report correctly. If the effect of Family Size in the Perceptual Identification task is indeed reflecting semantic processing upon presentation, the absence of an effect of Family Size in the Lexical Decision with illegal nonwords is more easily explained. In this task, semantic processing is not likely to occur, as participants can base their judgements on low-level, pure formal information.

With respect to Age of Acquisition and Frequency, Lewis (1999) and Lewis, Gerhand, and Ellis (2001) claim that these two variables actually reflect one single variable, cumulative frequency. This cumulative frequency is defined as the total number of times a word has been encountered. However, Ghyselinck et al. (submitted) provide evidence that Age of Acquisition and Frequency cannot be captured in one cumulative frequency variable, as in their tasks, the effect of Age of Acquisition was much larger than the effect of Frequency. Simulations with connectionist models also predict a larger effect of Age of Acquisition. For example, Ellis and Lambon Ralph (2000) found that activation patterns that are introduced early in training the network, have a long-lasting advantage over later introduced patterns, regardless of the cumulative frequency of activation. Our present results provide additional evidence against the cumulative-frequency hypothesis. If Frequency and Age of Acquisition should really be combined into one single variable, these two variables should emerge jointly as significant predictors, contrary to fact. Whereas Age of Acquisition was a significant predictor in all seven tasks, the role of Frequency seems to be less robust: It is a significant predictor in only three tasks. However, keep in mind that the variance of Age of Acquisition in our dataset was larger than the variance of Frequency (and Family Size). With these differences in characteristics, finding an effect for Age of Acquisition (if apparent) will be easier than finding an effect for Frequency, as our simulation studies also examplify (see Table 6.3 in Appendix G). Therefore, a much stronger claim against the cumulative-frequency hypothesis would be if a dependent variable can be reliably predicted by Frequency but not by Age of Acquisition. The actual response in the Semantic Association task was such a variable. In predicting the responses of the participants, we found that the higher the Frequency of the target word, the less likely participants are to name (part of) a morphological relative as a semantically related word. We interpreted this as the result of a competition process in which high-frequent words are more likely to be selected than low-frequent words. As target words which are high in Frequency tend to have semantically related words that are high in Frequency, morphological relatives (which are low-frequent) tend to come into play with more likelihood if the target words are relatively low-frequent. The Age of Acquisition of the target words is irrelevant in this respect, suggesting Frequency and Age of Acquisition cannot be combined into one single variable, and probably tap into somewhat different cognitive processes. With respect to Age of Acquisition and Frequency, we conclude, in line with Ghyselinck et al. (submitted) and Ellis and Lambon Ralph (2000) that Age of Acquisition is the more robust variable of the two, and that the two variables have independent effects.

Unlike Frequency, measures of the morphological family were significant in almost the full range of tasks. In the tasks predominantly tapping into semantic processing, the morphological family plays a tractable semantic role, in line with the hypothesis that upon reading a word, the meanings of morphologically related words are co-activated. In the Semantic Association task, this co-activation gives rise to responses of morphologically related words. In the Semantic Decision task, this co-activation leads to facilitation in response times if the morphological family members are relevant with respect to the decision. Furthermore, the decisions themselves are co-determined by the proportion of family members that are congruent with respect to the decision. If only a low proportion of family members is congruent with the correct decision, participants are more likely to make an error. This suggests that the co-activation of the meanings of family members upon reading a word actually co-determines the percept of the meaning of that specific word.

In a framework like Schreuder and Baayen (1995), as outlined in De Jong et al. (2000, also Chapter 2) with respect to morphological families, and more specifically implemented computationally in De Jong et al. (to appear, also Chapter 4), this co-activation can be explained in terms of spreading of activation along morpho-semantic relations. In this framework, semantic information is shared by concepts with overlapping meanings. If we assume that in performing a Semantic Decision task in which participants have to decide on whether an object is man-made or nature-made, they will make the decision after the amount of activation for nodes such as MAN-MADE or NATURE-MADE reach a certain threshold.[4] If semantic activation spreads along morpho-semantic lines, the target word will send activation not only to the correct decision-node, but also to its morphological family members. Subsequently, these morphological family members will send activation to these nodes as well.

This spreading activation framework can explain the reaction time data as well as the error data. If many family members are relevant with respect to the decision, either of the nodes will reach threshold quickly. At the same time, if these relevant family members are predominantly congruent with the correct decision, chances are high that the ’correct’ node will reach threshold first. But if many family members are incongruent with the ’correct’ decision, it is possible that the incorrect node will reach threshold before the correct one does, leading to an ’erroneous’ response.

In this spreading activation framework, units are symbolic. At the same time, the achitecture reflects the gradedness and dynamics of word meanings. In a subsymbolic approach, these results can also be captured, but the gradedness of meaning that emerge in the symbolic account upon reading a word by spreading of activation, would rather be emerging over time in the training of the model.

In conclusion, Frequency, Age of Acquisition, and Family Size independently influence lexical processing. Both Age of Acquisition and Family Size play a significant role in tasks tapping into semantic stages of processing. In the Semantic Association task, the Age of Acquisition of a word can predict the speed of the semantic processes. The Family Size, however, can predict the actual response in the Semantic Association task, following the prediction that upon reading a word, semantic activation spreads to morphologically related words. In the Semantic Decision task, the differential roles for subsets within the morphological family of a word additionaly provides insight in its dynamic and graded meaning, in an objective and tractable manner.

---

[4] Using the same framework, Krott, Schreuder, & Baayen (in press) modeled decisions on similarly abstract nodes, namely linking elements for novel Dutch compounds.

References

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., Tweedie, F. J. and Schreuder, R.: in press, The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon, *Brain and Language*.

Barry, C., Hirsh, K. W., Johnston, R. A. and Williams, C. L.: 2001, Age of acquisition, word frequency, and the locus of repetition priming of picture naming, *Journal of Memory and Language* **44**, 350–375.

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.

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

Brown, G. D. A. and Watson, F. L.: 1987, First in, first out. Word learning age and spoken word frequency as predictors of word familiarity and word naming latency, *Memory and Cognition* **15**, 208–216.

Brysbaert, M.: 1996, Word frequency affects naming latency in Dutch with age-of-acquisition controlled, *European Journal of Cognitive Psychology* **8**, 185–193.

Brysbaert, M., Lange, M. and Van Wijnendaele, I.: 2000a, The effects of age-of-acquisition and frequency-of-occurence in visual word recognition: Further evidence from the Dutch language, *European Journal of Cognitive Psychology* **12**, 65–85.

Brysbaert, M., Van Wijnendaele, I. and De Deyne, S.: 2000b, Age-of-acquisition of words is a significant variable in semantic tasks, *Acta Psychologica* **104**, 215–226.

Carroll, J. B. and White, M. N.: 1973, Word frequency and age of acquisition as determiners of picture-naming latency, *Quaterly Journal of Experimental Psychology* **25**, 85–95.

Coltheart, V., Laxon, V. J. and Keating, C.: 1988, Effects of word imageability and age of acquisition on children’s reading, *British Journal of Psychology* **79**, 1–12.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M. and Baayen, R. H.: 2002, The processing and representation of Dutch and English compounds:

Peripheral morphological, and central orthographic effects, *Brain and Language* **81**, 555-567.

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 Jong, N. H., Schreuder, R. and Baayen, R. H.: *to appear*, Morphological resonance in the mental lexicon, *Linguistics*.

De Moor, W., Ghyselinck, M. and Brysbaert, M.: 2000, A validation study of the age-of-acquisition norms collected by ghyselinck, de moor and brysbaert, *Psychologica Belgica* **40**, 99–114.

Ellis, A. W. and Lambon Ralph, M. A.: 2000, Age of acquisition effects in adult lexical processing reflects loss of plasticity in maturing systems: Insights from connectionist networks, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26**, 1103–1123.

Feldman, L. B. and Siok, W. W. T.: 1997, The role of component function in visual recognition of Chinese characters, *Journal of Experimental Psychology: Learning, Memory and Cognition* **23**, 778–781.

Feldman, L. B., Pnini, T. and Frost, R.: 1995, Decomposing words into their constituent morphemes: Evidence from English and Hebrew, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 947–960.

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

Gerhand, S. and Barry, C.: 1998, Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **24**, 267–283.

Gerhand, S. and Barry, C.: 1999a, Age-of-acquisition and frequency effects in speeded word naming, *Cognition* **73**, 27–36.

Gerhand, S. and Barry, C.: 1999b, Age of acquisition, word frequency, and the role of phonology in the lexical decision task, *Memory & Cognition* **27**, 592–602.

Ghyselinck, M., De Moor, W. and Brysbaert, M.: 2000, Age-of-acquisition ratings for 2816 Dutch four- and five-letter nouns, *Psychologica Belgica* **40**, 77–98.

Ghyselinck, M., Lewis, M. B. and Brysbaert, M.: *submitted*, Age of acquisition and the cumulative-frequency hypothesis: A multi-task investigation.

Gilhooly, K. J. and Logie, R. H.: 1981, Word age-of-acquisition and visual recognition thresholds, *Current Psychological Research* **1**, 215–225.

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

Krott, A., Schreuder, R. and Baayen, H.: in press, Linking elements in Dutch noun-noun compounds: Constituent families as predictors for response latencies, *Brain and Language*.

Lewis, M. B.: 1999, Age of acqusition in face categorisation: is there an instance-based account?, *Cognition* **71**, B23–B39.

Lewis, M. B., Gerhand, S. and Ellis, H. D.: 2001, Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects?, *Cognition* **78**, 189–205.

Lüdeling, A. and De Jong, N. H.: 2002, German particle verbs and word-formation, *in* N. Dehé, R. Jackendoff, A. McIntyre and S. Urban (eds), *Verb-particle explorations*, Mouton de Gruyter, Berlin, pp. 315–333.

Monsell, S., Doyle, M. C. and Haggard, P. N.: 1989, Effects of frequency on visual word recognition tasks, *Journal of Experimental Psychology: General* **118**, 43–71.

Morrison, C. M. and Ellis, A. W.: 1995, Roles of word frequency and age of acquisition in word naming and lexical decision, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 116–133.

Morrison, C. M. and Ellis, A. W.: 2000, Real age of acquisition effects in word naming and lexical decision, *British Journal of Psychology* **91**, 167–180.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H. and Baayen, R. H.: *in preparation*, Changing places: A cross-language perspective on frequency and family size in Hebrew and Dutch.

Oldfield, R. C. and Wingfield, A.: 1965, Response latencies in naming objects, *Quarterly Journal of Experimental Psychology* **17**, 273–281.

Ratcliff, R.: 1993, Methods for dealing with reaction time outliers, *Psychological Bulletin* **114**, 510–532.

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.: 1995, Modeling morphological processing, *in* L. B. Feldman (ed.), *Morphological Aspects of Language Processing*, Lawrence Erlbaum, Hillsdale, New Jersey, pp. 131–154.

Schreuder, R. and Baayen, R. H.: 1997, How complex simplex words can be, *Jour-*

nal of Memory and Language **37**, 118–139.

Stone, G. O. and Van Orden, G. C.: 1993, Strategic control of processing in word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **19**, 744–774.

Van Loon-Vervoorn, W. A.: 1989, *Eigenschappen van basiswoorden*, Swets and Zeitlinger, Lisse.

Whaley, C. P.: 1978, Word–nonword classification time, *Journal of Verbal Language and Verbal Behavior* **17**, 143–154.

Appendices A to F:

All words in the three contrasts with mean error proportions, harmonic mean reaction times, and mean ratings from the seven experiments: Perceptual Identification, Lexical Decision with illegal nonwords, Lexical Decision with legal nonwords, subjective frequency Rating, Semantic Association generation, Semantic Decision, and Lexical Decision with novel derivations (suffixes used for these novel derivations are also shown).

Appendix A: Words with an early Age of Acquisition

Item	PerId	LexdecIll		LexdecLeg		Rating	Assoc		Semdec		LexdecSufffixed		
		RT	Error	RT	Error		RT	Error	RT	Error	Suffix	RT	Error
beker (beker)	0.00	455	0.00	508	0.00	5.94	1196	0.00	750	0.00	loos	790	0.20
erwt (pea)	0.21	483	0.05	562	0.05	4.61	1386	0.00	806	0.00	achtig	786	0.05
koek (cake)	0.16	424	0.05	561	0.05	5.56	1404	0.00	1025	0.16	erij	838	0.70
lepel (spoon)	0.05	450	0.00	523	0.00	5.44	1336	0.00	848	0.00	vormig	764	0.00
pijl (arrow)	0.21	491	0.00	506	0.00	4.33	1109	0.00	837	0.04	vormig	708	0.05
wiel (wheel)	0.05	444	0.00	501	0.00	5.39	1188	0.00	869	0.00	loos	695	0.05
molen (mill)	0.14	430	0.05	530	0.00	3.94	1494	0.13	771	0.00	waarts	686	0.25
slee (sledge)	0.37	414	0.00	542	0.05	2.78	1188	0.00	831	0.04	gewijs	1432	0.95
vest (vest)	0.22	448	0.00	522	0.05	4.00	1384	0.00	745	0.00	achtig	805	0.05
soep (soup)	0.19	434	0.00	550	0.11	5.83	1380	0.00	935	0.24	erig	732	0.15
kaak (jaw)	0.53	434	0.00	604	0.05	4.28	1284	0.00	807	0.12	vormig	781	0.05
duim (thumb)	0.10	424	0.00	471	0.00	5.44	1367	0.00	757	0.04	erig	798	0.30
kast (closit)	0.18	427	0.05	472	0.00	6.06	1468	0.06	704	0.00	achtig	812	0.05
teen (toe)	0.11	432	0.00	552	0.00	5.28	1375	0.00	783	0.04	vormig	696	0.10
stoel (chair)	0.05	417	0.00	461	0.00	6.50	1083	0.00	713	0.04	achtig	711	0.05
snoep (candy)	0.05	420	0.00	476	0.00	6.00	1378	0.00	931	0.04	erij	724	0.00
knie (knee)	0.00	442	0.00	540	0.05	5.78	1590	0.06	708	0.28	vormig	720	0.05
boter (butter)	0.00	457	0.00	507	0.00	5.94	1591	0.00	1122	0.44	achtig	752	0.15
broek (trousers)	0.05	451	0.00	490	0.00	6.28	1555	0.00	744	0.00	vormig	772	0.15
zeep (soap)	0.00	412	0.05	477	0.00	4.89	1101	0.00	951	0.12	sel	729	0.25
stift (stylus)	0.14	483	0.00	548	0.00	4.50	1231	0.00	709	0.00	er	998	0.75
mouw (sleeve)	0.22	446	0.00	540	0.11	4.94	1422	0.00	834	0.00	vormig	781	0.10
peer (pear)	0.05	420	0.00	508	0.00	4.94	1163	0.00	727	0.00	achtig	670	0.10
tand (tooth)	0.10	442	0.00	513	0.00	5.56	1303	0.00	777	0.04	ig	679	0.55
puree (mash)	0.05	462	0.00	550	0.00	4.67	1537	0.00	964	0.28	achtig	850	0.05
kers (cherry)	0.05	403	0.00	501	0.00	4.33	1265	0.00	739	0.00	achtig	769	0.00
kous (stocking)	0.00	438	0.05	533	0.05	4.50	1327	0.00	739	0.04	loos	730	0.10

Appendix B: Words with a late Age of Acquisition

Item	PerId	LexdecIll		LexdecLeg		Rating	Assoc		Semdec		LexdecSufffixed		
		RT	Error	RT	Error		RT	Error	RT	Error	Suffix	RT	Error
koran (Koran)	0.43	431	0.05	597	0.05	1.72	1683	0.00	844	0.08	ist	1210	0.65
diner (dinner)	0.09	471	0.00	541	0.00	4.50	1266	0.00	808	0.00	achtig	913	0.20
mythe (myth)	0.15	567	0.10	604	0.05	2.28	1418	0.06	930	0.00	achtig	818	0.25
boeg (bow)	0.44	427	0.05	613	0.16	2.50	1747	0.25	1022	0.04	vormig	899	0.20
kust (coast)	0.10	406	0.00	534	0.05	3.83	1341	0.00	1127	0.16	achtig	809	0.10
mode (fashion)	0.16	468	0.00	523	0.00	5.56	1406	0.00	697	0.00	achtig	756	0.10
lans (spear)	0.20	478	0.05	623	0.16	1.94	1943	0.05	762	0.00	vormig	824	0.15
proza (prose)	0.15	471	0.05	631	0.11	2.94	1744	0.00	838	0.04	loos	921	0.40
nota (account)	0.11	477	0.00	587	0.00	3.56	1453	0.05	728	0.00	loos	898	0.40
loods (shed)	0.14	463	0.00	628	0.00	2.61	1612	0.06	797	0.00	achtig	817	0.30
vocht (moisture)	0.00	457	0.00	517	0.00	4.83	1280	0.00	953	0.12	loos	725	0.05
takel (tackle)	0.44	425	0.00	658	0.11	2.39	1538	0.06	834	0.00	aar	785	0.60
bron (source)	0.10	452	0.00	500	0.00	3.89	1130	0.00	909	0.16	waarts	708	0.10
valk (falcon)	0.16	447	0.00	528	0.00	2.67	1269	0.00	667	0.00	erij	941	0.30
kunst (art)	0.09	426	0.00	502	0.00	5.39	1472	0.00	804	0.04	erig	809	0.25
icoon (icon)	0.14	507	0.05	668	0.11	2.00	1904	0.17	852	0.00	vormig	920	0.00
zenuw (nerve)	0.10	468	0.10	537	0.00	4.28	1730	0.00	858	0.08	erij	888	0.35
dijk (dike)	0.17	463	0.10	523	0.00	4.11	1536	0.00	766	0.16	vormig	723	0.05
hars (resin)	0.16	405	0.00	624	0.05	2.22	1285	0.11	879	0.08	ig	888	0.20
club (club)	0.06	457	0.00	511	0.05	5.11	1577	0.05	743	0.00	gewijs	810	0.05
porie (pore)	0.55	460	0.00	729	0.21	2.28	1469	0.06	1062	0.16	vormig	978	0.30
pand (premises)	0.42	431	0.00	517	0.00	3.50	1720	0.06	749	0.00	vormig	736	0.20
barak (barracks)	0.14	447	0.00	658	0.11	1.78	1830	0.26	776	0.00	achtig	952	0.15
vloot (fleet)	0.00	431	0.00	542	0.05	2.56	1510	0.00	865	0.04	gewijs	940	0.25
vloed (tide)	0.15	448	0.00	515	0.00	3.11	1364	0.00	984	0.20	loos	767	0.10
roman (novel)	0.19	464	0.00	551	0.05	4.78	1119	0.00	731	0.00	achtig	830	0.05
opera (opera)	0.05	468	0.00	542	0.00	3.17	1281	0.00	737	0.00	waarts	760	0.30

Appendix C: Words with a high Frequency

Item	PerId	LexdecIll		LexdecLeg		Rating	Assoc		Semdec		LexdecSufffixed		
		RT	Error	RT	Error		RT	Error	RT	Error	Suffix	RT	Error
tabel (table)	0.41	424	0.00	642	0.11	4.56	1684	0.00	862	0.00	erig	850	0.45
model (model)	0.25	457	0.00	524	0.00	4.72	1795	0.00	818	0.00	erig	912	0.05
diner (dinner)	0.09	471	0.00	541	0.00	4.50	1266	0.00	808	0.00	achtig	913	0.20
cent (penny)	0.19	456	0.00	530	0.00	4.33	1160	0.00	738	0.00	erig	798	0.25
dorp (village)	0.05	448	0.00	492	0.00	5.78	1454	0.00	761	0.00	waarts	757	0.20
kust (coast)	0.10	406	0.00	534	0.05	3.83	1341	0.00	1127	0.16	achtig	809	0.10
trein (train)	0.09	458	0.00	485	0.00	6.56	1584	0.00	718	0.00	er	849	0.85
proza (prose)	0.15	471	0.05	631	0.11	2.94	1744	0.00	838	0.04	loos	921	0.40
lunch (lunch)	0.05	457	0.05	504	0.00	5.61	1142	0.00	788	0.00	loos	835	0.15
nota (account)	0.11	477	0.00	587	0.00	3.56	1453	0.05	728	0.00	loos	898	0.40
hotel (hotel)	0.09	426	0.05	475	0.00	4.78	1268	0.00	718	0.00	waarts	766	0.15
naam (name)	0.14	467	0.00	615	0.21	6.78	1707	0.00	843	0.00	achtig	776	0.20
keel (throat)	0.26	422	0.05	495	0.00	5.17	1520	0.00	798	0.12	vormig	659	0.10
vocht (moisture)	0.00	457	0.00	517	0.00	4.83	1280	0.00	953	0.12	loos	725	0.05
villa (villa)	0.05	436	0.05	535	0.00	3.44	1222	0.00	662	0.04	waarts	757	0.20
grap (joke)	0.32	410	0.05	563	0.00	5.83	1168	0.00	818	0.00	gewijs	726	0.10
taak (task)	0.15	441	0.05	544	0.00	4.94	1368	0.00	775	0.00	loos	748	0.05
fles (bottle)	0.19	440	0.00	530	0.00	6.06	1365	0.00	763	0.00	vormig	667	0.05
duim (thumb)	0.10	424	0.00	471	0.00	5.44	1367	0.00	757	0.04	erig	798	0.30
kroeg (pub)	0.24	417	0.00	532	0.00	6.06	1178	0.00	720	0.00	waarts	796	0.15
tong (tongue)	0.20	433	0.05	494	0.00	5.39	1699	0.00	707	0.12	vormig	709	0.00
kader (frame)	0.00	441	0.00	561	0.00	2.94	1435	0.00	765	0.00	loos	772	0.05
knie (knee)	0.00	442	0.00	540	0.05	5.78	1590	0.06	708	0.28	vormig	720	0.05
wang (cheek)	0.29	427	0.00	529	0.11	4.78	1130	0.00	778	0.04	achtig	811	0.10
bier (beer)	0.10	421	0.05	499	0.00	6.11	1251	0.00	984	0.08	achtig	757	0.05
unie (union)	0.35	473	0.00	592	0.00	2.50	1688	0.06	858	0.00	loos	947	0.30
fort (fort)	0.10	439	0.00	592	0.16	2.11	1390	0.00	810	0.00	achtig	810	0.30
huid (skin)	0.09	415	0.00	493	0.00	5.33	1426	0.17	701	0.04	achtig	693	0.10
vuist (fist)	0.09	443	0.05	510	0.00	4.67	1263	0.06	823	0.36	vormig	781	0.05
debat (debate)	0.09	469	0.00	601	0.00	3.11	1613	0.00	775	0.00	gewijs	916	0.25
broek (trousers)	0.05	451	0.00	490	0.00	6.28	1555	0.00	744	0.00	vormig	772	0.15
grot (cave)	0.50	461	0.00	537	0.05	3.11	1604	0.06	915	0.16	achtig	725	0.00
muur (wall)	0.25	448	0.00	465	0.00	5.83	1443	0.00	748	0.00	loos	710	0.10
plein (square)	0.05	435	0.00	490	0.00	5.72	1590	0.06	788	0.00	ig	744	0.70
taxi (taxi)	0.14	440	0.05	523	0.00	4.50	1220	0.00	711	0.04	waarts	802	0.10
buik (stomach)	0.10	417	0.00	535	0.00	5.72	1592	0.00	763	0.12	ig	762	0.40

Appendix D: Words with a low frequency

Item	PerId	LexdecIll		LexdecLeg		Rating	Assoc		Semdec		LexdecSuffixed		
		RT	Error	RT	Error		RT	Error	RT	Error	Suffix	RT	Error
kever (beatle)	0.05	436	0.00	540	0.00	3.33	1624	0.00	735	0.04	achtig	771	0.10
koek (cake)	0.16	424	0.05	561	0.05	5.56	1404	0.00	1025	0.16	erij	838	0.70
pomp (pump)	0.16	466	0.05	567	0.00	3.67	1222	0.00	742	0.04	erij	821	0.45
wesp (wasp)	0.10	428	0.00	563	0.05	3.83	1257	0.00	657	0.00	achtig	753	0.05
stro (straw)	0.19	527	0.05	587	0.11	2.67	1509	0.00	814	0.04	achtig	869	0.30
kwast (brush)	0.18	450	0.00	577	0.05	3.94	1404	0.00	843	0.00	erig	841	0.20
hark (rake)	0.17	450	0.05	588	0.00	3.72	1660	0.06	949	0.08	sel	878	0.75
strip (comic)	0.15	436	0.00	537	0.00	4.33	1169	0.00	768	0.00	erig	925	0.30
kraam (stall)	0.36	448	0.00	533	0.00	3.50	1856	0.11	973	0.08	achtig	806	0.20
zebra (zebra)	0.00	441	0.00	534	0.05	3.00	1249	0.00	687	0.00	loos	791	0.15
cello (cello)	0.05	464	0.00	616	0.00	2.39	1402	0.00	756	0.00	vormig	930	0.25
wiel (wheel)	0.05	444	0.00	501	0.00	5.39	1188	0.00	869	0.00	loos	695	0.05
molen (mill)	0.14	430	0.05	530	0.00	3.94	1494	0.13	771	0.00	waarts	686	0.25
wafel (waffle)	0.00	433	0.00	553	0.00	3.83	1440	0.00	937	0.16	vormig	756	0.00
soep (soup)	0.19	434	0.00	550	0.11	5.83	1380	0.00	935	0.24	erig	732	0.15
appel (apple)	0.05	469	0.05	501	0.00	5.94	1172	0.00	701	0.00	vormig	697	0.00
riool (sewer)	0.10	417	0.00	569	0.00	3.50	1379	0.00	851	0.04	loos	803	0.20
iglo (igloo)	0.29	476	0.10	588	0.05	2.00	1398	0.00	888	0.08	vormig	851	0.15
netel (nettle)	0.14	434	0.00	609	0.16	2.06	1801	0.00	903	0.04	achtig	810	0.25
paal (pole)	0.24	424	0.00	595	0.05	4.28	1696	0.06	832	0.08	vormig	665	0.05
vijl (file)	0.53	491	0.00	556	0.11	3.06	1501	0.00	879	0.04	er	868	0.45
worm (worm)	0.26	435	0.00	526	0.05	3.50	1419	0.06	715	0.04	ig	754	0.10
polis (policy)	0.18	513	0.00	628	0.00	2.89	1641	0.11	765	0.00	loos	944	0.45
biet (beet)	0.17	435	0.05	649	0.16	2.89	1500	0.00	728	0.00	achtig	848	0.10
mast (mast)	0.65	443	0.00	564	0.05	3.17	1448	0.00	779	0.00	vormig	753	0.05
gesp (buckle)	0.53	505	0.05	699	0.16	2.50	1584	0.06	883	0.12	loos	938	0.25
zeep (soap)	0.00	412	0.05	477	0.00	4.89	1101	0.00	951	0.12	sel	729	0.25
spin (spider)	0.05	444	0.00	520	0.00	5.00	1544	0.05	676	0.04	ig	791	0.05
egel (hedgehog)	0.62	433	0.05	560	0.05	3.06	1462	0.00	726	0.00	achtig	838	0.00
video (video)	0.05	448	0.00	488	0.00	5.83	1354	0.00	755	0.00	loos	731	0.15
kiosk (kiosk)	0.05	474	0.00	603	0.00	3.78	1434	0.00	707	0.00	waarts	976	0.15
kegel (cone)	0.14	437	0.00	566	0.00	2.94	1729	0.00	973	0.00	aar	815	0.05
speld (pin)	0.05	454	0.00	536	0.00	3.89	1325	0.00	769	0.00	vormig	746	0.00
kous (stocking)	0.00	438	0.05	533	0.05	4.50	1327	0.00	739	0.04	loos	730	0.10
neon (neon)	0.28	497	0.05	713	0.21	2.00	1372	0.20	778	0.16	loos	969	0.40
kiwi (kiwi)	0.25	518	0.10	549	0.11	3.72	1334	0.00	797	0.04	achtig	875	0.10

Appendix E: Words with a high Family Size

Item	PerId	LexdecIll		LexdecLeg		Rating	Assoc		Semdec		LexdecSuffixed		
		RT	Error	RT	Error		RT	Error	RT	Error	Suffix	RT	Error
wolk (cloud)	0.20	443	0.00	499	0.00	5.33	1285	0.00	766	0.00	vormig	725	0.00
zaad (seed)	0.14	416	0.00	526	0.00	3.94	1643	0.00	777	0.00	loos	697	0.00
hoed (hat)	0.27	396	0.00	485	0.05	4.39	1411	0.00	717	0.00	vormig	733	0.00
pomp (pump)	0.16	466	0.05	567	0.00	3.67	1222	0.00	742	0.04	erij	821	0.45
rots (rock)	0.09	434	0.00	596	0.05	4.00	1366	0.05	776	0.12	erig	702	0.05
molen (mill)	0.14	430	0.05	530	0.00	3.94	1494	0.13	771	0.00	waarts	686	0.25
vest (vest)	0.22	448	0.00	522	0.05	4.00	1384	0.00	745	0.00	achtig	805	0.05
appel (apple)	0.05	469	0.05	501	0.00	5.94	1172	0.00	701	0.00	vormig	697	0.00
koren (corn)	0.09	461	0.00	555	0.00	2.83	1470	0.00	766	0.04	achtig	847	0.05
gevel (front)	0.42	468	0.00	586	0.00	3.22	1292	0.06	777	0.00	loos	775	0.10
darm (intestine)	0.05	431	0.00	515	0.00	4.22	1475	0.06	782	0.08	achtig	737	0.10
kast (closit)	0.18	427	0.05	472	0.00	6.06	1468	0.06	704	0.00	achtig	812	0.05
plank (plank)	0.10	442	0.00	514	0.00	4.83	1441	0.00	862	0.04	erig	766	0.35
rijst (rice)	0.00	481	0.05	522	0.00	5.83	1540	0.00	783	0.00	erig	876	0.30
fonds (fund)	0.19	470	0.10	569	0.00	3.28	1330	0.00	710	0.04	loos	859	0.10
paal (pole)	0.24	424	0.00	595	0.05	4.28	1696	0.06	832	0.08	vormig	665	0.05
storm (storm)	0.09	412	0.00	485	0.00	4.89	1419	0.00	854	0.04	ig	652	0.05
graf (grave)	0.30	446	0.00	503	0.00	4.22	1398	0.00	912	0.08	waarts	752	0.10
doek (cloth)	0.11	455	0.05	536	0.11	4.56	1466	0.17	810	0.00	erig	784	0.50
ijzer (iron)	0.00	476	0.05	505	0.00	4.28	1366	0.06	1103	0.44	achtig	669	0.05
kroon (crown)	0.09	424	0.00	562	0.00	3.61	1317	0.00	870	0.00	vormig	727	0.05
slang (snake)	0.05	411	0.00	506	0.00	4.28	1831	0.00	724	0.00	vormig	736	0.00
zeil (sail)	0.10	440	0.00	533	0.05	3.50	1314	0.00	917	0.00	erij	847	0.35

Appendix F: Words with a low Family Size

Item	PerId	LexdecIll		LexdecLeg		Rating	Assoc		Semdec		LexdecSuffixed		
		RT	Error	RT	Error		RT	Error	RT	Error	Suffix	RT	Error
halte (stop)	0.00	449	0.00	531	0.00	5.06	1080	0.00	767	0.04	waarts	926	0.25
beker (beker)	0.00	455	0.00	508	0.00	5.94	1196	0.00	750	0.00	loos	790	0.20
cent (penny)	0.19	456	0.00	530	0.00	4.33	1160	0.00	738	0.00	erig	798	0.25
ruzie (row)	0.05	436	0.00	514	0.00	5.11	1503	0.06	728	0.00	loos	707	0.05
pijl (arrow)	0.21	491	0.00	506	0.00	4.33	1109	0.00	837	0.04	vormig	708	0.05
hoeve (farm)	0.15	485	0.00	622	0.11	1.89	1315	0.00	823	0.00	waarts	944	0.60
vaas (vase)	0.29	438	0.00	508	0.00	4.33	1156	0.00	799	0.00	vormig	757	0.10
trui (sweater)	0.05	438	0.00	509	0.00	5.72	1272	0.00	736	0.16	vormig	814	0.15
villa (villa)	0.05	436	0.05	535	0.00	3.44	1222	0.00	662	0.04	waarts	757	0.20
hiel (heel)	0.11	429	0.00	582	0.05	3.33	1253	0.00	829	0.28	vormig	837	0.15
grap (joke)	0.32	410	0.05	563	0.00	5.83	1168	0.00	818	0.00	gewijs	726	0.10
duim (thumb)	0.10	424	0.00	471	0.00	5.44	1367	0.00	757	0.04	erig	798	0.30
vacht (fur)	0.10	449	0.00	543	0.00	3.50	1346	0.00	1078	0.12	loos	734	0.10
oever (bank)	0.33	485	0.00	543	0.05	3.28	1380	0.00	809	0.24	achtig	766	0.10
wang (cheek)	0.29	427	0.00	529	0.11	4.78	1130	0.00	778	0.04	achtig	811	0.10
lift (elevator)	0.10	456	0.05	525	0.00	4.94	1767	0.05	737	0.00	erig	748	0.45
vuist (fist)	0.09	443	0.05	510	0.00	4.67	1263	0.06	823	0.36	vormig	781	0.05
debat (debate)	0.09	469	0.00	601	0.00	3.11	1613	0.00	775	0.00	gewijs	916	0.25
poes (puss)	0.10	431	0.00	511	0.00	5.89	1273	0.00	797	0.04	ig	673	0.10
grot (cave)	0.50	461	0.00	537	0.05	3.11	1604	0.06	915	0.16	achtig	725	0.00
stoep (pavement)	0.29	432	0.00	507	0.00	5.06	1564	0.00	766	0.00	loos	698	0.10
kraai (crow)	0.18	424	0.00	534	0.00	3.61	1167	0.00	734	0.00	achtig	723	0.05
taxi (taxi)	0.14	440	0.05	523	0.00	4.50	1220	0.00	711	0.04	waarts	802	0.10

Appendix G: Simulations

The response latencies in a simulated data set depended linearly on three independent random variables, A (Age of Acquisition), B (Base Frequency), and F (Family Size):

$$\text{RT}_i = \beta_0 + \beta_1 A_i + \beta_2 B_i + \beta_3 F_i + \varepsilon_i, \;\; i = 1, 2, \ldots, 40. \tag{6.1}$$

with ε and the β-weights estimated from the empirical data set of the Lexical Decision experiment with legal nonwords ($\varepsilon \sim \mathcal{N}(0, 0.0044), \beta_0 = 6.32, \beta_1 = 0.0135, \beta_2 = 0.0217$, and $\beta_3 = 0.0198$).[5] Each simulated data set contained 40 'word' items. Let μ_A and σ_A^2 denote mean and variance of Age of Acquisition, μ_B and σ_B^2 denote mean and variance of (log) Base Frequency, and let μ_F and σ_F^2 denote mean and variance of (log) Family Size. We assigned these means and variances the values of the corresponding predictor variables in our empirical data set of 141 words: $\mu_A = 7.07, \sigma_A^2 = 5.08, \mu_B = 6.41, \sigma_B^2 = 1.69, \mu_F = 2.67$, and $\sigma_F^2 = 0.76$. In order to make the simulated data sets similar to the actual data set in terms of the pairwise correlations between the independent variables and overall collinearity, we made use of an implicit common random variable C:

$$C \sim \mathcal{N}(0, \sigma_C^2). \tag{6.2}$$

The random variables A, B, and F were all constructed from this common basis, as follows:

$$\begin{aligned} B &\sim \mathcal{N}(\mu_B, \sigma_B^2 - \sigma_C^2) + C & (6.3)\\ F &\sim \mathcal{N}(\mu_F, \sigma_F^2 - \sigma_C^2) + C & (6.4)\\ A &\sim -w_2 \mathcal{N}(\frac{1}{w_2}\mu_A, x(\sigma_A^2 - \sigma_C^2)) + w_1 C, & (6.5) \end{aligned}$$

with $x = (\sigma_A^2 - w_1^2\sigma_C^2)/(w_2^2(\sigma_A^2 - \sigma_C^2))$ and $w_1^2 + w_2^2 = 1$. In this way, the simulated means and variances are identical to the empirically observed means and variances in the data set, while at the same time ensuring that the independent variables are correlated. By choosing $w_1 = 0.9$ and $\sigma_C^2 = 0.76$, the collinearity in the simulated data sets and the pairwise correlations approximate those in our original data set.

The results reported in the column labeled YYY in Table 6.3 are based on simulated data sets in which the reaction times are obtained as just described. In order

---

[5] Reaction times were log-transformed, hence the small numbers

to evaluate the likelihood of Type I errors, we also constructed simulated reaction times in which any of the random variables A, B, or F was replaced by dummy random variables A', B', and F', introducing noise (uncorrelated with the other variables) with the same mean and variance as A, B, or F respectively. In this way, the mean and variance in the simulated reaction times remain the same. In Table 6.3, the column labeled NYY, for instance, reports the results for simulated data sets in which A, the random variable mimicking Age of Acquisition, was replaced by A'. Likewise, the last column labeled NNN describes the results for the data sets in which all three random variables were replaced by dummy random variables.

Table 6.3: Number of significant effects ($p < 0.05$) for 1000 simulated data sets with 40 items, using multiple regression and three factorial analysese for each data set. The columns specify which independent variables contributed to the simulated reaction times, with Y specifying inclusion for A, B, and F respectively. A, B, and F are the randomly simulated variables that correspond in characteristics to Age of Acquisition, Base Frequency, and Family Size.

	regression (40 items)							
	YYY	YNY	YYN	NYY	YNN	NNY	NYN	NNN
A	1000	1000	1000	41	1000	49	43	54
B	1000	50	1000	744	52	59	696	58
F	1000	854	54	445	53	354	59	53
	factorial designs (40 items each)							
	YYY	YNY	YYN	NYY	YNN	NNY	NYN	NNN
A	1000	1000	1000	57	1000	51	44	48
B	729	21	679	617	17	51	616	52
F	166	106	0	474	39	269	49	46

In all statistical analyses of the simulated data sets, the simulated reaction times were always predicted from A, B, and F, even in cases where the simulated reaction times in the data set were actually constructed from the dummy variables A', B', and F'. The first half of Table 6.3 shows that the regression analyses combine power, depending on which variable is under scope, with an acceptable chance of Type I error. The numbers in the table represent the number of times a variable is picked up as a significant variable in a multiple regression analysis in 1000 simulation runs ($p's < 0.05$). The regression analysis always detects the relevance of A if present, which, given the relatively small variance of the error (σ_ε^2) and the large variance of A itself is not very surprising. B is identified as a relevant predictor if present in 696 to 1000 of the 1000 simulation runs, depending on which other variables are contributing to the reaction times. The variable with the smallest variance, F, emerges as the most difficult to identify if contributing to the reaction times, with identification rates ranging from as low as 354 up to 1000 out of 1000 simulation runs. The Type I error rates range between 41/1000 up to 59/1000, which shows that the collinearity in our kind of data does not lead to a real Type I error problem.

To compare these regression results with results using factorial analyses, we also constructed for each of the thousand simulation runs three factorial designs, with 40 items in each: One for A (in which we contrasted A while matching on B and F), and similarly one for B, and one for F. In order to construct these factorial designs from the same 1000 datasets as the datasets on which the regression analyses were carried out, we actually made for each simulation run a sample of 500 items ('words') in the same manner as described above, with $i = 1, 2, \ldots, 500$, from which we randomly selected 40 items for the regression analyses. To construct the three factorial designs on each of the 1000 datasets, we selected 40 items for each contrast that were maximally contrasting in one dimension, while not significantly differing in the other two dimensions ($p's > .1$, two-tailed t-tests). In this way, for each simulation run, each of the three factorial designs is comparable in power in terms of number of items as the one single regression analysis. Whereas the regression analysis tests whether each of the three variables A, B, and F are significantly predicting reaction times using the same 40 items, the three factorial designs may include a total of 120 items (but note that, as we sample from the same 500 items for each of the three factorial designs, some overlap is possible).

The bottom half of Table 6.3 reports the significant results of the factorial analyses of the three variables, using t-tests ($p's < 0.05$). Equal to the regression analyses, the variable A is always a significant variable if indeed contributing to the simulated reaction times. For the variables B and F, however, the factorial analyses are less powerful than the regression analyses: B is a significant variable in 616 to 729 out of 1000 simulation runs, and F, the variable with the smallest variance, is picked up as a significant variable in a mere 106 to 474 of the 1000 simulation runs. The number of times a Type I error is made, is also somewhat lower (0/1000 up to 57/1000).

From these simulation results we conclude that, especially for a variable with a low variance, such as F (with the same characteristics as Family Size), regression analyses are much more powerful to pick up on significant effects, if present, than factorial analyses. At the same time, the chances of Type 1 errors in the regression analyses for the simulated data sets with the same characteristics as our empirical data set are around the range of what can be expected, given the alpha level of 0.05. Conclusions based on factorial analyses are too restricted, and especially for a variable with a small variance such as Family Size, the likelihood of accepting H_0 is too high. For instance, in only 106 out of 1000 simulation runs was F picked up as a significant variable if the simulated reaction times actually were dependent on A and F jointly.

Summary and discussion

CHAPTER 7

Words can serve as morphological constituents in other words. Some words have a high morphological productivity, in that they occur in many complex words, whereas others are morphological islands. In English, the word *man* has the largest family with 270 descendents, including *fireman* and *salesmanship* (CELEX; Baayen, Piepenbrock, and Gulikers, 1995). Most words in English as well as in Dutch, however, are not that productive: Half of the monomorphemic words occur in at most 3 other words.

Schreuder and Baayen (1997) showed that in Dutch, this morphological productivity is a facilitatory factor in lexical processing. The larger the morphological family, the faster participants are able to respond in a visual lexical decision task. For other languages, similar effects have been found, such as for first constituents in Finnish compound words (Hyönä and Pollatsek, 1998), for morphological roots in Hebrew words (Feldman, Pnini, and Frost, 1995), for characters and radicals in Chinese words (Taft and Zhu, 1995; Taft and Zhu, 1997; and Feldman and Siok, 1997), and also for kanji in Japanese words (Yamada and Kayamoto, 1998). However, the number of words in which a given constituent occurs, the family size, is highly correlated with the summed frequency of these morphological family members, the family frequency. For Dutch, Schreuder and Baayen (1997) showed that it is the type count that matters, whereas the family frequency is irrelevant. Baayen, Lieber, and Schreuder (1997) replicated this finding in a rating task for English words.

In this thesis, we investigate what the functionality of the observed effect of family size is. Schreuder and Baayen (1997) interpret the effect as spreading of activation along morphological lines. Upon reading a word, its morphological family members become co-activated. The more global activation in the mental lexicon (in the sense of Grainger & Jacobs, 1996), the easier it will be to decide in a lexical decision task that the target word is an existing word. The experiments in this thesis, however, show that this co-activation of family members actually reflects meaning activation, as the context in which a word occurs can define, in a semantically plausible way, which subsets of family members are activated. In Chapter 6, we find that the co-activation of family members even co-determines the meaning of the monomorphemic target word itself. Before we turn in detail to the role of context, however, Chapter 2 investigates how robust the effect of family size is, and Chapter 3 turns to effects of the morphological family of constituents in compound words.

The family size effect is robust

The studies by Schreuder and Baayen (1997) and Bertram, Baayen, and Schreuder (2000) leave open several questions that are concerned with the robustness of the effect. First, whether verbs as well as nouns show an effect of family size. Second, whether stems in suffixed words show an effect of family size. Third, how the presence of an inflectional suffix may have an influence, and finally, whether there is an effect of family size for words which do not share the same orthographic and phonological form with their family members. Chapter 2, *Explorations on the family size effect*, provides an answer to these questions, all using Dutch words in visual lexical decision.

Bertram, Baayen, and Schreuder (2000) investigated the effect of family size for inflected verb forms and found that verbs with a high family size were responded to faster than verbs with a low family size. However, most of the verbs in their high family size condition were verbs with nominal conversion alternants, whereas the verbs in the low family size condition tended not to have such nominal alternants. This confound leaves open the question whether the effect of family size truly exists for verb forms. Therefore, in Experiment 2.1, we directly compare nouns and verbs, none of which have a conversion alternant. We find that verbs as well as nouns show an effect of family size.

Bertram et al. (2000) also investigated the effect of family size for complex words, and found that the family size of the base word affected reaction times. For instance, when *trotser*, ’prouder’ was presented, the number of family members of *trots*, ’proud’ or ’pride’, reduced reaction times. However, in their materials, they did not control for the summed token frequencies of these family members, the family frequency of the base word. For monomorphemic words, Schreuder and Baayen only found an effect of the type count of the morphological family (family size), and no effect of the family frequency. Experiment 2.2 shows that for suffixed words, in line with the results for monomorphemic words, there is a facilitatory effect of the family size of the base word, and no such effect for family frequency. If anything, there is a trend that words with a high family frequency are responded to slower than words with low family frequencies. In a study re-analyzing the materials from Schreuder and Baayen (1997), Baayen, Tweedie, and Schreuder (2002) were able to show that, indeed, for monomorphemic words, the family frequency is an inhibitory factor. They explain this inhibition as an effect of competition between the monomorphemic target words and their high-frequency family members.

In Experiment 2.3, we investigate the role of the inflectional suffix *-t*. We directly compare verbs in the first person present tense, without an overt suffix, to the same verbs in the third person present tense (e.g., *sjouw* and *sjouwt*, ’drag’ and ’drags’). Although both verb forms show an effect of family size, it is attenuated for the verb forms without the inflectional marker. We can explain this difference in effect size when we divide the total family counts into nouns and verbs. For the example *sjouw(t)*, we count *wegsjouwen*, ’to drag away’, as a verbal family member, and a word like *sjouwer*, ’dragger’ is counted as a nominal family member. Correlations between these subsets of family members with reaction times reveal that verbal family members (e.g., *wegsjouwen*) only contribute to the effect of family size when the verbs are presented with the overt inflectional marker.

The fact that it is not the string familiarity of a morpheme (as captured by the family frequency), but rather the type count of its family members that reduces reaction times, leads to the hypothesis that the effect does not operate at the form level. Schreuder and Baayen (1997) as well as Bertram, Schreuder, and Baayen (2000) supported this view by showing that semantically opaque family members do not contribute to the effect of family size (e.g., not counting *casualty* as a family member of *casual* improves correlations with reaction times). In Experiment 2.4, we provide direct evidence that the effect of family size does not depend on exact overlap in orthographic and phonological form. We compare regular and irregular formed past participles (e.g., *geroeid*, which is derived from *roeien*, ’to row’, and *gevochten*, which is derived from *vechten*, ’to fight’). Although the morphological family members of the irregular past participles do not share the exact form with these past participles (e.g., *vechter*, ’fighter’, is a morphological family member of *gevochten*), the effect of family size for the regular and irregular participles is the same.

No effect for constituents in compounds

After establishing that for suffixed words, as well as for monomorphemic words, family size reduces reaction times, we turn to compound words in Chapter 3, *Dutch and English compounds*. In an earlier lexical decision study, constituents of Dutch compounds did not show an effect of the constituent frequency (Van Jaarsveld & Rattink, 1988). In priming studies, on the other hand, both constituents of a compound showed morphological priming (Sandra, 1990; Zwitserlood, 1994). If the effect of family size reflects spreading of activation, it might mirror morphological or semantic priming, in that both constituents of (at least relatively transparent) compound words would show an effect of family size.

Experiment 3.1, however, reveals that for constituents in Dutch compounds, no such effect of family size is obtained. Similarly, no effect of the frequencies of the constituents is present, in line with Van Jaarsveld and Rattink (1988). What we do find, however, is an effect for the frequency of the compound as a whole, the co-occurrence frequency of its two constituents. Furthermore, we find an effect of the positional frequency of the constituents. This positional frequency is the token count of a constituent in the same position. For instance, for the constituent *wind* in *windmolen* ('wind mill'), the positional frequency is the number of tokens in which *wind* occurs as a left constituent. This token count is remarkably similar to the family frequency token count. Its effect, however, is qualitatively different: Where the effect of family frequency for monomorphemic words (Baayen et al., 2002) and possibly for suffixed words (Chapter 2) is inhibitory and probably reflects competition at the form level, the effect of positional frequency for constituents in compounds is facilitatory instead of inhibitory. The effect of the positional frequency reflects a positional occurrence probability. If the frequency of *wind*XXXX (with XXXX standing for any constituent) is high, the probability of this string of letters being an actual compound increases. We hypothesize that the effect of positional frequency is located at the level of access and possibly even affects eye-movements. Hyönä and Pollatsek (1998) reported that a type frequency count of the first constituent in Finnish compounds has an effect on eye-movements. If the first constituent occurs in many other compounds as a first constituent (i.e., if the first constituent has a large family), initial fixations and gaze durations were shorter than when it was a compound with a unique beginning. Furthermore, the initial fixation position was slightly closer to the beginning of the word for compound words with unique beginnings. Possibly, the type-count effect they reported on actually reflects the positional token frequency effect. Likewise, the type count effects for novel compounds that Van Jaarsveld, Coolen, and Schreuder (1994) reported on (newly coined compounds are harder to reject as existing compounds if they contain highly productive constituents), might actually have been positional frequency effects as well. Interestingly, Taft and Zhu (1997) showed a very similar effect for Chinese radicals in compound characters. They report that the positional productivity of the radicals should be taken into account, rather than the total productivity. However, they were unable to tease apart the type count of the positional productivity with its token count.

Words differ from phrases

Constituents of Dutch compounds are always concatenated, whereas the spelling of English compounds varies: They can be concatenated, hyphenated or written with a space between the constituents. The latter we call open compounds. In Experiment 3.2, we compare English concatenated compounds with English open compounds.

For the concatenated compounds, we find the same effects as we find in Dutch: an effect of the frequency of the compound as a whole, plus an effect of positional frequency (for the left constituent). Compounds that are written with a space also show an effect of compound frequency, but instead of an effect of positional frequency, we now find an effect of positional family size (for the left constituent). This effect of a type count for the open compounds suggests that its constituents are processed more similar to monomorphemic words. In Experiment 3.3, we return to Dutch compounds, but we now add a space between the constituents. This simple change in the stimuli changes the lexical decision task effectively into a double lexical decision task. Nevertheless, the results are largely the same as for the concatenated compounds: an effect of the frequency of the compound as a whole, and an effect of positional frequency. The dissociation we find between the English open compounds and the Dutch artificial open compounds rules out the possibility that the effect of the space in the English open compounds is a superficial effect of the orthography.

Experiment 3.4 provides further evidence for the special status of English open compounds. In three sub-experiments, we investigate the effect of family size for English monomorphemic words. In each of the three sub-experiments, we control for total family size, while contrasting either the number of concatenated, hyphenated, or open compounds within the family. We find that while keeping the total family size constant, only the design in which the number of open compounds within the family is contrasted yields a significant difference in reaction times. If many of the family members are written with a space, reaction times are longer. A post-hoc correlation study confirms this finding in that the number of concatenated and hyphenated family members reduces reaction times, whereas the number of open family members is irrelevant.

To summarize these experiments, the constituents (at least the left constituent) of English open compounds are processed more similar to monomorphemic words. At the same time, complex words containing a space do not belong to the morphological families of its constituents. We conclude that combinations of words which are written with a space are processed and stored as phrases, which is qualitatively different from the processing and storage of complex words without a space. The question remains whether these compounds with spaces are processed and stored as phrases because of the orthography, or whether the space in the orthography is felt to be appropriate, precisely because they are processed and stored as phrases.

Converging evidence for a different status for phrases as opposed to words, but without the confound of orthography, can be found in Lüdeling and De Jong (2002). They compared family size effects for German base verbs (e.g., *stehen*, 'to stand'), with effects for semantically transparent particle verbs (e.g., *dastehen*, 'to stand there'), and semantically opaque particle verbs (e.g., *ausstehen*, 'to be missing'). Particle verbs are entities more similar to phrases than normal complex words. The most obvious reason is that the particle and the base verb can be separated in a sentence (see e.g., Booij, 1990; Van Riemsdijk, 1978; Schreuder, 1990). If particle verbs are unanalyzable units at the semantic level, we expect a severely attenuated effect of family size for the opaque particle verbs. This would be in line with Bertram et al. (2000) and Schreuder and Baayen (1997), who showed that only semantically related family members contribute to the family size effect. Lüdeling and De Jong (2002), however, showed that the effect of family size for the opaque particle verbs includes all family members of the base verbs. They conclude that particle verbs, even though they are spelled as words in the experiment, are processed as phrases.

Context defines the co-activation

If the effect of family size reflects spreading of semantic activation, we can expect the context in which a word appears to have an effect. In Chapter 4, *The effect of morphological and sentential context*, we specifically address this issue. In Chapter 2, we found that minimal morphological context for verb stems can already play a role. Bertram et al. (2000) found that for de-adjectival nouns ending in *-heid* (e.g., *droogheid*, ’dryness’), a semantically defined subset is excluded from the family size effect. This subset contained color-compounds and intensified adjectives, which we will call scale-focusing adjectives as both kinds of adjectives share the feature of focusing on a very specific part of the scale denoted by the base adjective. For *droogheid*, family members like *kurkdroog* and *gortdroog* (literally ’cork-dy’ and ’barley-dry’, but both meaning ’very dry’) are such scale-focusing adjectives. Bertram et al. (2000) found that excluding these family members from the count of family size improved the correlations with reaction times.

To further investigate the effect of context on adjectives, we present 40 adjectives in four different contexts. In Experiment 4.1, we present the adjectives in their base form and in the comparative form (e.g., *mooi*, ’beautiful’, and *mooier*, ’more beautiful’). In Experiment 4.2, we present the same adjectives in minimal sentential context: preceded by *heel*, ’very’, or *niet*, ’not’. The stimuli consist of the full phrases, such as *heel mooi*, ’very beautiful’.

To investigate the role of context, we divide the total family size counts into different subsets: family members that are nouns, verbs, general adjectives, or scale-focusing adjectives, with the term general adjectives denoting all adjectival family members that are not scale-focusing. The correlations of these different subsets with reaction times across contexts suggest that without any context (presenting the base word), all family members contribute to the family size effect. In the case that the adjectives are presented in their comparative form or with *niet*, it is predominantly the count of the general adjectives that correlates with reaction times. For the adjectives presented with *heel*, however, it is predominantly the count of scale-focusing adjectives that correlates with reaction times. It is clear from these experiments that across contexts, the pattern of correlations differs a lot. However, due to the collinearity in the dataset (a word with many nominal family members is likely to have many adjectival family members, and so on), conclusions are tentative as some correlations that turn up as significant might in fact be spurious.

In the second part of Chapter 4, we propose a computational model that simulates the results of morphological family size. This model, the Morphological Family

Resonance Model (MFRM), also proves to be a useful statistical tool to deal with the collinearity problem and enables us to estimate the contributions of the different subsets of family members.

In the MFRM, a computational implementation of the general model proposed by Schreuder & Baayen (1995), semantic and syntactic representations are shared. In this way, different lemma's (in the sense of Levelt, 1989) pointing to the same semantics are connected. The model explains the effect of family size by means of resonance between the lemma level and the semantic level. As activation from a specific lemma spreads to its semantic and syntactic representations, these representations will subsequently send activation back to the lemma's to which they are connected. Due to the resonance, the activation of the target lemma accumulates exponentially, with the rate of accumulation determined by the number of family members, until a certain threshold is reached.

For all four contexts, we obtain model times which we correlate with the different subsets of family members. In order for the MFRM to simulate the results across contexts, we find that we need different resonance sets, the size of which is inversely proportional to the amount of information provided by the context. The resonance set is defined as the set of family members that are allowed to participate in the resonance. To obtain the same pattern of correlations as the empirical pattern for the condition in which the adjectives are presented without context, the resonance set contains all morphological family members. For the other three contexts, we find that we need restrictions on the resonance sets. The context of the comparative suffix and the context of *niet*, in these experiments, provide information as to which word-category is specifically relevant (namely the adjectival family members). For the MFRM, we find that it is precisely those family members that we need to include in the resonance set. Even further restrictions are needed to obtain a reasonable fit for the adjectives presented with *heel*. In this case, only the scale-focusing adjectives are included in the resonance set, which are the family members that share meaning with the full phrase that is presented. For instance, *bloedmooi* (literally, 'blood-beautiful') is such a scale-focusing adjective in the family of *mooi* and its meaning, indeed, is *very beautiful*.

The MFRM proves to be fruitful in two different ways. First, we can simulate the effect of family size using spreading of activation along morpho-semantic relations. Second, the restrictions we are forced to make in order to get reasonable fits in the correlational patterns allow us to conclude that in the empirical data, the same restrictions are likely to play a role. We can now understand the effect of family size in a more functional way. If such severe, but semantically plausible restrictions play a role when only minimal context is provided, the spreading of activation can be understood as semantic activation co-determining the meaning of the word itself. Adjectives in different contexts activate different subsets of family members, leading to different percepts of meaning.

Family size *and* frequency are context-sensitive

If context plays a role for adjectives that are, in the linguistic sense, not ambiguous, it is very likely that context will also play a role for words that are truly ambiguous. In Chapter 5, *Homonyms in context*, we pursue this issue. We expect that if context is provided, only the context-specific family size of a homonym will correlate with reaction times. Additionally, since for the two meanings of a homonym different frequency counts can be obtained, we investigate whether besides family size, frequency can be context-sensitive as well.

In Experiment 5.1, we present Dutch noun-noun homonyms such as *wapen*, which can mean either ’weapon’ or ’coat of arms’. We divide the meanings in dominant (the most frequently encountered) and subordinate meanings. For *wapen*, ’weapon’ is the dominant meaning and ’coat of arms’ the subordinate. We present the homonyms in a visual lexical decision task without context, and in a lexical decision task using semantically related primes to both the dominant and subordinate meanings as disambiguating context, as well as unrelated primes. We also use a subjective frequency rating task without any context, and a subjective frequency rating task with semantically related words as disambiguating context.

Without disambiguating context, as evidenced by frequency and family size effects for both of the meanings, both meanings are activated. An exception is the lexical decision with unrelated primes, for which both frequency counts turn up with reliable correlations, but no reliable family size effects are found. Presenting the two meanings with a disambiguating prime in the rating task reveals that only the frequency and family size of the dominant meaning correlates significantly with reaction times if the prime indeed was related to the dominant meaning. In the lexical decision task with related primes, however, the results are less clear. If the context directs to the dominant meaning, both frequency counts correlate with reaction times, but only the family size of the dominant meaning turns up with a reliable correlation. If the prime is related to the subordinate meaning, just as in the rating task, none of the correlations are significant.

In Chapter 4 we found that the number of family members that participate in the family size effect is inversely proportional to the amount of context provided. It is likely that the highly related semantic primes in this study narrow down activation of family members to only a small subset. In the case that the subordinate meaning (with the least number of family members) is primed, this subset is probably too small to measure. Interestingly, the effect of frequency also tends to disappear (in line with results by Becker, 1979; Borowsky & Besner, 1993; Plaut & Booth, 2001). As evidenced by the specific frequency-effect of the dominant meanings in the rating study, the frequency effect, similar to the family size effect, is context-sensitive. Models of lexical processing that posit the frequency effect solely at the level of access (e.g., Baayen, Dijkstra, & Schreuder, 1997; Bradley & Forster, 1987; Morton, 1969), therefore, cannot explain this context-sensitivity. Apparently, at least part of the effect of frequency must be operating at a more central level in the mental lexicon.

In Experiment 5.2 we turn to Dutch noun-verb homonyms, such as *last* (meaning ’(the) burden’ or ’(he/she) welds’), because these homonyms can be disambiguated with minimal context. We present the noun-verb homonyms in a rating and lexical decision task without context, in a rating and lexical decision task using the experimental list as context (by providing filler words that are either all unambiguously nouns or that are all unambiguously verb forms), and finally in a lexical decision task in which the homonyms are preceded by minimal syntactic context (e.g., *de last*, ’the burden’ or *hij last*, ’he welds’).

When no context is provided, it is only in the rating task that both meanings are activated, as evidenced by frequency and family size effects for both meanings (although the effect for the family size of the verb meaning is only marginal). In the lexical decision task without context, only the counts for the nominal meaning correlate significantly with reaction times. We observe the same pattern for the rating and the two lexical decision tasks when the context directs to the nominal meaning. There is one exception, the family size effect of the nominal meaning when the homonyms are preceded by a definite article is not significant. If the context directs to the verbal meaning of the homonym, the results change dramatically. In these cases, it is the counts of frequency and family size of predominantly the verbal meaning that correlate with ratings and reaction times.

The nominal meaning appears to be dominant. In the lexical decision task without context, only the counts of this meaning influence reaction times. And even in the case that all filler words are verb inflections, besides the counts of the verb meaning, the frequency of the noun meaning still affects reaction times. We hypothesize that nouns are more easily interpretable without direct context than inflected verb forms, which is in line with results by Baayen et al. (1997). Indeed, if the homonyms in the lexical decision task are inbedded in direct context, that is if they are preceded by a personal pronoun, only the verb meaning is activated. Furthermore, in this case we find an effect of context comparable to the restricting effects of context found in Chapter 4: Only the verbal family members of the verb meaning play a role (e.g., counting *inlassen*, 'to insert, to weld in', but not *lasapparaat*, 'welding machine', for *hij last*). We hypothesize that the lack of an effect of the noun family size when the homonyms are preceded by a definite article, are likewise due to a restrictional effect of context: The definite article specifies the meaning of the noun (see Langacker, 1991), so that no effect of family size is obtained (e.g., *de last*, 'the burden', only refers to one very specific *burden*, and the meanings of family members such as *schuldenlast*, 'burden of debt' or *overbelasten*, 'overburden', may now be irrelevant).

To summarize, we find in both experiments that besides family size, also frequency is context-sensitive. In line with these results, McDonald and Shillcock (2002) showed that contextual distinctiveness, a corpus-derived summary measure of the frequency distribution of the contexts in which a word occurs, is a better predictor than the traditional count of frequency of occurrence of a word. With respect to family size we find in Chapter 6, as well as in Chapter 4, that context narrows down the activation of family members to a subset of family members that is semantically and syntactically defined.

The effect is independent from age of acquisition

When we investigate the role of family size, we attempt to control our materials with respect to other variables that are known to influence lexical processing, such as different counts of frequency of occurrence and length. Thusfar, we have not mentioned another potentially confounding variable that has recently attracted a lot of attention: age of acquisition. The age of acquisition of a word is defined as the mean age at which this word is learned. In various tasks, age of acquisition has been found to affect performance (e.g., Brysbaert, De Lange, & Van Wijnendaele, 2000; Carrol & White, 1973). The effect of age of acquisition is independent from the effect of frequency, yet has been argued to affect the same levels of processing (Ghyselinck, Lewis, & Brysbaert, submitted; Ellis & Lambon Ralph, 2000). However, in the studies investigating frequency and age of acquisition, the family size of the words has not been taken into account. Similarly, in the studies investigating family size, the age of acquisition of the words has not been taken into account.

In Chapter 6, *Effects of frequency, age of acquisition, and family size in a range of tasks*, we attempt to ascertain what the independent relative contributions of frequency, age of acquisition, and family size are. As the variables are correlated, it is possible that a part of the family size effect actually is an age of acquisition effect. On the other hand, it is difficult to see how the observed restrictional effects of context on the morphological family can be explained by age of acquisition alone. In this Chapter, we use a range of seven tasks that presumably tap into different stages of lexical processing, all using visual presentation of Dutch words as a starting point.

We first construct three factorial designs, each of which maximally contrasts one of the three variables, while controlling for the other two. The tasks we employ are perceptual identification, lexical decision with strings of consonants as nonwords, lexical decision with phonotactically legal nonwords, a subjective frequency rating, a semantic association generation task, a semantic decision task, and, finally, a lexical decision task with newly coined derivations as target words.

The results of these factorial designs reveal that age of acquisition clearly is the strongest predictor. Frequency and family size play a role in only a few of these tasks. However, the power of the factorial designs, in terms of number of items but mostly in size of the contrasts, is hardly comparable. Regression analyses, using the pooled item sets of all three contrasts, are a better means to ascertain whether these variables influence responses. However, a problem with regression analyses on variables that are correlated, as our variables are, is that it is unclear whether the effects that the analyses report on are truly independent. We therefore carry out a simulation study that compares the results for factorial designs and regression designs on simulated data sets with very similar characteristics as the empirical data set. The outcome of this simulation study unambiguously shows that for variables with the characteristics of age of acquisition, frequency, and family size as in the empirical data set, regression analyses are much more powerful than factorial analyses, especially in picking up significant effects of variables with only small variances, such as family size. At the same time, the regression analyses are not more prone to show spurious effects than the factorial analyses.

We therefore base our conclusions for the empirical data set on regression analyses. In all seven tasks, age of acquisition is a significant variable. Frequency, however, is less robust, significant only in the lexical decision task with consonant strings as nonwords, the lexical decision with legal nonwords, and the rating task. The family size of a word, as we may expect from previously observed effects, is a reliable predictor in the lexical decision task with legal nonwords, the rating task, and the lexical decision with newly coined derivations.

Surprising, however, is the significant effect for family size in the perceptual identification task, a task that is generally viewed as exclusively tapping into form stages of processing. In this task, participants see the target word for a mere 32 milliseconds and are asked what word they have seen. Schreuder and Baayen (1997) reported that family size does not influence reaction times in a progressive demasking task, a task which likewise has been argued to be sensitive to effects of characteristics of the form of the stimulus. Possibly, perceptual identification is a more sensitive task to pick up on any effect. We offer two explanations for the significant effect of family size in this task. First, the family size effect might, after all, have a form component. Second, family size might not have a form component, but in the perceptual identification task, participants are helped by a large family during the semantic processing following the short presentation of the word. In line with the latter explanation is the fact that in the lexical decision task with consonant strings as nonwords, in which participants do not need to identify the exact word and can base their decisions on pure form information, the family size is irrelevant. Furthermore, the results in Chapter 2 for the irregular past participles showed that the effect of family size cannot depend on a mere overlap in form.

The meanings of family members are co-activated

A direct test on whether the effect of family size is truly semantic and reflects spreading of semantic activation, are the semantic association generation and semantic decision tasks in Chapter 6. In the semantic association task, we find an effect of family size on the actual response: If the family size of the target word is large, participants are more likely to name (part of) a morphological family member as a semantically related word. This finding is evidence for the claim that the effect of family size in a task such as lexical decision reflects spreading of activation along morpho-semantic lines. The semantic decision task reveals further support for the claim made in Chapters 4 and 5 that this spreading of activation co-determines the meaning of the target words. In the semantic decision task, participants are asked to decide whether the words represent objects made by man or made by nature.

We find that the relevance within the family with respect to this decision (objects versus actions and qualities) co-determines the response times. The actual decisions are co-determined by the congruency within the family. For instance, the correct response to *appel*, ’apple’, would be NATURE-MADE. But family members such as *appeltaart*, ’apple pie’, and *appelmoes*, ’applesauce’, are incongruent with this decision. If a word has many of such incongruent family members, participants are more likely to make an error. Apparently, words representing objects that are nature-made but that have many family members that are man-made are intuitively felt to be less prototypically nature-made.

The role of the morphological family in the semantic association generation task and in the semantic decision task are direct evidence for the hypothesis that semantic activation spreads along morpho-semantic lines. Upon reading a word, the meanings of morphologically related words become co-activated, even to the extent that the meaning of the target word is in part defined by this co-activation.

These results are well interpretable with the linguistically inspired model of the mental lexicon by Bybee (1988), who explicitly links morphologically related words at both the level of form and the level of meaning. Our results are also easy to understand within the psycholinguistic spreading-activation theory by Collins and Loftus (1975). In this theory, semantically similar concepts are connected with links, along which spreading of activation occurs. The full meaning of a word is defined as the activation of the nodes in the whole semantic network. In this spreading activation network, and as in Roelofs’ (1992) computational implementation, the links between concepts express relations such as hyponymy and hyperonymy. The effect of the morphological family documented in this thesis shows that morphological relations between words, established objectively on the basis of a lexical database, reveal a considerable part of the semantic links in the mental lexicon.

That morphological relations establish conceptual links along which activation spreads actually should not come as a surprise. Morphology is the organizational principle in language to systematically link form with meaning. In producing a word for a new concept, the major resource is morphology. For instance, if I want to express the concept of ’with the shape of a cloud’, I’m likely to say *cloud-shaped*, using a newly coined word that is made up of two existing words, in form as well as meaning. Methodologically, the family size is a new tool that exploits the structure of the language itself to trace what kinds of semantic fields participate in the spreading of activation. Whereas in priming studies, only binary links between words can be shown, the effect of the morphological families makes it possible to chart the co-activation of larger sets of words.

A question that still remains open, however, is whether semantic relations between words that are supported by a morphological form relation are qualitatively different from other semantic relations. Is it the case that the observed effects of the morphological family indeed reflect special status for semantic relations between morphological family members, or can relations between words that are semantically, but not morphologically related show the same effect? In Chapter 2, we show that it is not mere overlap in form that ensures the family size effect. Thus, the present results do not rule out the possibility that the observed effects of the morphological family actually reflect purely semantic relations in the mental lexicon, independent from overlap in morphological form.

All chapters considered jointly, the most revealing finding is that different subsets of the morphological family show differential effects, depending on context. Context is very broadly defined: It can be morphological context (Chapters 2 and 4), minimal sentential context (Chapters 4 and 5), the context of the experimental list (Chapter 5), and the context of the experimental task itself (Chapter 6). As minimal context can drastically constrain the subset of family members that is activated, we can now also understand why in Chapter 3, the constituents of concatenated compounds do not show an effect of family size. In this case, the other constituent of the compound narrows down the activation to only those family members that are relevant given this other constituent, that is, to family members of the compound word as a whole. Moscoso del Prado Martín, Deutsch, Frost, Schreuder, De Jong, and Baayen (in preparation) showed that indeed, for derived words (including compounds), there is an effect of family size for the family members that are derived from these complex words themselves, but only a weak effect for the remaining family members of the base word.

We have also seen that the status of a fixed combination of words affects which family members are activated. If the base word is part of a phrase rather than part of a complex word, this base word is processed similarly to a word in isolation (Chapter 3 and Lüdeling & De Jong, 2002), irrespective of whether the phrase contains an actual space or not.

The overall conclusion we can draw on the basis of these findings is that the family size effect does not simply reflect some automatic global activation in the mental lexicon (Schreuder & Baayen, 1997), which merely happens to be useful when carrying out a visual lexical decision task. To the contrary, since context determines which words in the family are activated, we take the family size effect to reflect functional co-activation of meanings. Without context, a word carries the meaning of all its senses, and all these senses become co-activated. But words embedded in context only co-activate the meanings of precisely those family members that are relevant given the context.

References

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

Baayen, R. H., Tweedie, F. J. and Schreuder, R.: in press, The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon, *Brain and Language*.

Becker, C. A.: 1979, Semantic context and word frequency effects in visual word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **5**, 252–259.

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

Booij, G.: 1990, The boundary between morphology and syntax: Separable complex verbs in dutch, *in* G. E. Booij and J. van Marle (eds), *Yearbook of Morphology 1990*, Kluwer Academic Publishers, Dordrecht, pp. 45–63.

Borowsky, R. and Besner, D.: 1993, Visual word recognition: A multistage activation model, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **19**, 813–840.

Bradley, D. C. and Forster, K. I.: 1987, A reader's view of listening, *Cognition* **25**, 103–134.

Brysbaert, M., Lange, M. and Van Wijnendaele, I. .: 2000a, The effects of age-of-acquisition and frequency-of-occurence in visual word recognition: Further evidence from the Dutch language, *European Journal of Cognitive Psychology* **12**, 65–85.

Bybee, J. L.: 1988, Morphology as lexical organization, *in* M. Hammond and M. Noonan (eds), *Theoretical Morphology: Approaches in Modern Linguistics*, Academic Press, London, pp. 119–141.

Carroll, J. B. and White, M. N.: 1973, Word frequency and age of acquisition as determiners of picture-naming latency, *Quaterly Journal of Experimental Psychology* **25**, 85–95.

Collins, A. M. and Loftus, E. F.: 1975, A spreading-activation theory of semantic processing, *Psychological review* **82**, 407–428.

Ellis, A. W. and Lambon Ralph, M. A.: 2000, Age of acquisition effects in adult lexical processing reflects loss of plasticity in maturing systems: Insights from connectionist networks, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26**, 1103–1123.

Feldman, L. B. and Siok, W. W. T.: 1997, The role of component function in visual recognition of Chinese characters, *Journal of Experimental Psychology: Learning, Memory and Cognition* **23**, 778–781.

Feldman, L. B., Pnini, T. and Frost, R.: 1995, Decomposing words into their constituent morphemes: Evidence from English and Hebrew, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 947–960.

Ghyselinck, M., Lewis, M. B. and Brysbaert, M.: *submitted*, Age of acquisition and the cumulative-frequency hypothesis: A multi-task investigation.

Hyönä, J. and Pollatsek, A.: 1998, Reading finnish compound words: Eye fixations are affected by component morphemes, *Journal of Experimental Psychology: Human Perception and Performance* **24**, 1612–1627.

Langacker, R. W.: 1991, *Foundations of Cognitive Grammar. Vol. 2, Descriptive Application*, Stanford University Press, Stanford.

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

Lüdeling, A. and De Jong, N. H.: 2002, German particle verbs and word-formation, *in* N. Dehé, R. Jackendoff, A. McIntyre and S. Urban (eds), *Verb-particle explorations*, Mouton de Gruyter, Berlin, pp. 315–333.

MacDonald, S. A. and Shillcock, R. C.: 2002, Rethinking the word frequency effect: The neglected role of distributional information in lexical processing, *Language and Speech* **44**, 295–323.

Morton, J.: 1969, The interaction of information in word recognition, *Psychological Review* **76**, 165–178.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H. and Baayen, R. H.: *in preparation*, Changing places: A cross-language perspective on frequency and family size in Hebrew and Dutch.

Plaut, D. C. and Booth, J. R.: 2000, Individual and developmental differences in semantic priming: Empirical and computational support for a single mechanism account of lexical processing, *Psychological Review* **107**, 786–823.

Riemsdijk, H. C. v.: 1978, *A Case Study in Syntactic Markedness*, Peter de Ridder Press, Lisse.

Roelofs, A.: 1992, A spreading activation theory of lemma retrieval in speaking, *Cognition* **42**, 107–142.

Sandra, D.: 1990, On the representation and processing of compound words: automatic access to constituent morphemes does not occur, *Quarterly Journal of Experimental Psychology* **42A**, 529–567.

Schreuder, R.: 1990, Lexical processing of verbs with separable particles, *Yearbook of Morphology* **3**, 65–79.

Schreuder, R. and Baayen, R. H.: 1995, Modeling morphological processing, *in* L. B. Feldman (ed.), *Morphological Aspects of Language Processing*, Lawrence Erlbaum, Hillsdale, New Jersey, pp. 131–154.

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

Taft, M. and Zhu, X.: 1995, The representation of bound morphemes in the mental lexicon: A Chinese study, *in* L. B. Feldman (ed.), *Morphological Aspects of Language Processing*, Lawrence Erlbaum, Hillsdale, New Jersey, pp. 293–316.

Taft, M. and Zhu, X.: 1997, Submorphemic processing in reading Chinese, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **23**, 761–775.

Yamada, J. and Kayamoto, Y.: 1998, Valency, secondary frequency, and lexical access: A Japanese study, *Applied Psycholinguistics* **19**, 87–97.

Zwitserlood, P.: 1994, The role of semantic transparency in the processing and representation of Dutch compounds, *Language and Cognitive Processes* **9**, 341–368.

Samenvatting

De relatie tussen de vorm van een woord en zijn betekenis is arbitrair. Er is geen reden waarom de ronde vrucht die Adam en Eva het paradijs uitjoeg, ruzie tussen drie Griekse godinnen ontketende, en Sneeuwwitje in slaap sukkelde een ’appel’ heet. En zoals een appel per toeval een appel heet te zijn, is een taart per toeval een ’taart’. Een *appeltaart*, echter, is niet toevallig een appeltaart. De betekenis van ’appeltaart’ heeft te maken met de betekenissen van beide constituenten. Het organisatorisch principe van talen dat zorgt dat dit soort combinaties van woordvormen ook combinaties van woordbetekenissen kunnen zijn, is de morfologie.

Sommige woorden zijn vaak een onderdeel van een ander woord, terwijl andere woorden morfologische eilanden zijn. In het Nederlands heeft het woord *werk* de grootste morfologische familie. Voorbeelden van familieleden zijn *monnikenwerk, overwerken*, en *werkloosheidsuitkering*. In dit proefschrift worden de consequenties van morfologische relaties tussen woorden op het verwerken van die woorden onderzocht.

Een bekend effect in lexicale verwerking is het effect van de frequentie van een woord. In een lexicale decisietaak, waarbij proefpersonen zo snel mogelijk moeten beslissen of een string van letters bestaande Nederlandse woorden vormen, reageren proefpersonen sneller op woorden die vaak voorkomen dan op woorden die weinig voorkomen (Whaley, 1978). Eerder onderzoek heeft uitgewezen dat net zoals de frequentie van een woord, de grootte van de morfologische familie een faciliterend effect heeft op de verwerking van een ongeleed woord. Proefpersonen kunnen sneller reageren naarmate de morfologische familie groter is. De frequentie van de familieleden bleek irrelevant (Schreuder & Baayen, 1997). De centrale vraag in dit proefschrift is wat de functionaliteit van dit familiegrootte-effect zou kunnen zijn.

In Hoofdstuk 2 werd aangetoond dat het effect van de morfologische familie niet alleen optreedt bij de verwerking van ongelede zelfstandig naamwoorden. Ook bij werkwoorden heeft de grootte van de morfologische familie een faciliterend effect in lexicale decisie (Experiment 2.1). In Experiment 2.2 werd aangetoond dat voor gelede woorden het effect van het aantal familieleden, net zoals bij ongelede woorden, onafhankelijk is van de frequentie van de familieleden. In Experiment 2.3 werd de invloed van een inflectioneel suffix onderzocht. Het effect van de morfologische familie werd vergeleken tussen werkwoorden zonder inflectioneel suffix en diezelfde werkwoorden met het suffix *-t* (bijvoorbeeld *sjouw* en *sjouwt*). Alleen bij de werkwoorden met het inflectionele suffix bleken de werkwoordelijke familieleden (zoals *wegsjouwen*) aan het familiegrootte-effect bij te dragen. Als de proefpersonen reageerden op de variant zonder het inflectionele suffix bleken alleen familieleden zoals *sjouwer*, de zelfstandig naamwoorden, bij te dragen aan het effect. Het vierde experiment van Hoofdstuk 2 liet duidelijk zien dat het effect niet afhankelijk is van pure overeenkomsten in vorm. In dit experiment reageerden de proefpersonen op onregelmatige voltooid deelwoorden, waarbij de vorm verschilde van die van de familieleden. Bijvoorbeeld het onvoltooid deelwoord *gevochten* heeft familieleden als *vechter* of *gevecht*, woorden die niet de exacte vorm met het voltooid deelwoord delen. Ondanks dit verschil in vorm, was het effect van familiegrootte net zo sterk als bij regelmatige gevormde deelwoorden zoals *geroeid*, die wel de exacte vorm van de stam met hun familieleden delen (waaronder *roeier*, en *roeispaan*). Eerder onderzoek had al aangetoond dat alleen morfologisch gerelateerde woorden die ook semantisch gerelateerd zijn bijdragen aan het familiegrootte-effect. Bijvoorbeeld, *gemeente* is niet semantisch gereleateerd aan *gemeen* en de correlatie tussen reactietijden en familiegroottes verbetert als dit soort familieleden niet meegeteld worden (Bertram, Baayen, & Schreuder, 2000). Samen met Experiment 2.4 bewijzen deze resultaten dat het familiegrootte-effect niet een effect is op het niveau van vormherkenning.

In Hoofdstuk 3 werd onderzocht wat het effect van familiegrootte bij samenstellingen is. Neem een samenstelling als *windmolen*. Zijn het zowel de familieleden van *wind* als die van *molen* die bijdragen aan het familiegrootte-effect? Voor Nederlandse samenstellingen (Experiment 3.1) werd duidelijk dat geen van beide families invloed hebben. De resultaten wezen uit dat het niet zozeer het aantal familieleden was, zoals bij ongelede woorden, maar de positiefrequenties van de twee delen van een samenstelling die invloed hadden. Bijvoorbeeld, als *molen* vaak als linkerlid voorkomt, of *wind* vaak als rechterlid, dan is het makkelijker om het geheel *windmolen* te herkennen. In tegenstelling tot Nederlandse samenstellingen, kunnen Engelse samenstellingen op verschillende manieren geschreven worden (aan elkaar zoals bij Nederlandse samenstellingen, met een streepje tussen de twee constituenten, of met een spatie). Experiment 3.2 onderzocht of deze verschillen in schrijfwijzen ook invloed hebben op de verwerking. De samenstellingen die geschreven werden zoals de Nederlandse samenstellingen (aan elkaar) werden inderdaad op dezelfde wijze verwerkt: niet de familiegrootte, maar de positiefrequentie was van invloed. De (linker)constituenten van Engelse samenstellingen met een spatie werden echter verwerkt meer in overeenstemming met ongelede woorden. Experiment 3.3, waarin de Nederlandse samenstellingen werden aangeboden met een artificiële spatie tussen de constituenten, wees uit dat de afwijkende verwerking voor de Engelse samenstellingen met een spatie niet een puur vormeffect van de spatie was. De Nederlandse samenstellingen met spaties kostten wat meer tijd dan de samenstellingen die zonder spaties geschreven waren, maar het effect van positiefrequentie was hetzelfde. De resultaten van Experiment 3.4, waarin ongelede Engelse woorden aangeboden werden, leverden extra bewijs voor de speciale status van Engelse samenstellingen met spaties. Uit dit experiment bleek namelijk dat de familieleden die met een spatie geschreven worden niet bijdragen aan het effect van familiegrootte. Samengevat, de constituenten van samenstellingen met spaties worden meer als ongelede woorden verwerkt, en bovendien behoren deze samenstellingen niet tot de morfologische familie van ongelede woorden. Wellicht dat de Engelse samenstellingen met spaties in het mentale lexicon niet opgeslagen zijn als normale gelede woorden, maar meer als zinsdelen.

Hoofdstuk 4 onderzocht wat de invloed van context op het effect van familiegrootte kan zijn. Experiment 2.3 had al bewezen dat minieme morfologische context (de aanwezigheid van een inflectioneel suffix bij werkwoorden) bepalend kan zijn voor welke familieleden aan het effect bijdragen. In Experiment 4.1 onderzochten we de invloed van het suffix *-er* bij bijvoegelijk naamwoorden (bijvoorbeeld *mooi* en *mooier*), en in Experiment 4.2 werd minimale zinscontext aangeboden: de bijvoegelijk naamwoorden werden gepresenteerd met *niet* of *heel* (bijvoorbeeld *niet mooi* en *heel mooi*). Beide soorten context bleken een inperkende invloed te hebben op het aantal familieleden dat bijdroeg aan het effect. Zonder context waren het alle familieleden van de bijvoegelijk naamwoorden die tot het effect bijdroegen. Maar als ze met het suffix *-er* of in een zinscontext met *niet* gepresenteerd werden, waren het alleen de familieleden die zelf ook bijvoegelijk naamwoorden waren. Als de bijvoegelijk naamwoorden met *heel* gepresenteerd waren, werd de subset van familieleden die bijdroeg aan het familiegrootte-effect nog veel kleiner. In dit geval waren het vooral die familieleden die qua betekenis met het hele zinsdeel overeenkwamen, zoals *bloedmooi* bij *heel mooi*. De tweede helft van Hoofdstuk 4 beschrijft een mathematisch model dat het effect van familiegrootte verklaart als het spreiden van activatie tussen woorden die zowel morfologisch als semantisch verwant zijn.

In Hoofdstuk 5 werd de invloed van context nogmaals onder de loep genomen. Ditmaal gebruikten we homoniemen als items in lexicale decisie experimenten en ratings. Voor homoniemen is het overduidelijk dat de context invloed heeft op de betekenis. In Experiment 5.1 werden homoniemen zoals *wapen* aangeboden zowel zonder context als met een semantisch gerelateerde prime als disambiguerende context. In Experiment 5.2 gebruikten we homoniemen zoals *last* die een zelfstandig naamwoord of een vervoegd werkwoord kunnen zijn. Voor deze homoniemen gebruikten we andere vormen van context: manipulatie van de experimentele lijst (een combinatie van zelfstandig naamwoorden en werkwoorden, alleen zelfstandig naamwoorden, of alleen werkwoorden als andere woorden in het experiment) en syntactische context (*hij last* en *de last*). Het effect van familiegrootte voor de homoniemen in beide experimenten was afhankelijk van de context. Zonder context bleken familieleden van beide betekenissen invloed te kunnen hebben, terwijl de contexten de activatie van familieleden in kon perken naar precies die familie die qua betekenis conform de context was.

Voor de twee betekenissen van de homoniemen konden wij ook verschillende frequenties verkrijgen. En net zoals het effect van familiegrootte context-afhankelijk is, bleek het effect van frequentie dat ook te zijn. Bijvoorbeeld, in het geval dat *last* aangeboden werd in een lijst met alleen maar zelfstandig naamwoorden als fillers, werden alleen familieleden als *belasting* geactiveerd, en was alleen de frequentie van *last* in de betekenis van het zelfstandig naamwoord van invloed op de ratings en reactietijden. Dit effect van context op frequentie wijst uit dat niet alleen het effect van familiegrootte, maar ook het effect van frequentie op zijn minst een betekeniscomponent in zich heeft. Dit in tegenstelling tot sommige modellen van lexicale verwerking, die veronderstellen dat het effect van frequentie alleen aspecten van het herkennen van de vorm van woorden reflecteert (bijvoorbeeld Morton, 1969).

Bij het onderzoek naar het effect van bijvoorbeeld familiegrootte wordt geprobeerd te controleren voor effecten van andere variabelen. Tot dusverre in dit proefschrift is de verwervingsleeftijd van woorden buiten beschouwing gelaten. Voor verschillende taken is bewezen dat de verwervingsleefstijd van een woord invloed heeft (bijvoorbeeld Brysbaert, De Lange, & Van Wijnendaele, 2000; Carrol & Whyte, 1973). Hoe eerder een woord geleerd is, hoe makkelijker dit woord te herkennen is. Dit effect is, net zoals het effect van de familiegrootte, onafhankelijk van de frequentie van het woord. In een reeks experimenten met verschillende taken onderzochten wij wat de onafhankelijke effecten van frequentie, verwervingsleeftijd, en familiegrootte zijn. De taken varieerden van taken die vooral zicht geven op de vroege stadia van lexicale verwerking, tot taken die inzicht geven in de semantische verwerking.

De verwervingsleeftijd bleek een belangrijke voorspellende variabele voor alle experimentele taken. De frequentie en de familiegrootte van een woord bleken ieder een onafhankelijke bijdrage te leveren in een aantal taken. Bovendien gaf het specifieke effect van de familiegrootte in twee taken die vooral semantische verwerking belichten, extra inzicht in de functionaliteit van dit effect. In een semantische associatie taak, waarbij proefpersonen bij het lezen van een woord zo snel mogelijk een semantisch gerelateerd woord moesten zeggen, bleken proefpersonen vaker een (deel van een) morfologisch gerelateerd woord te noemen als het woord een grote familie had. In een semantische categorisatietaak, waarbij de proefpersonen moesten beslissen of een woord een object door de mens of door de natuur gemaakt representeerde, bleek de betekenis van de familieleden cruciaal. De snelheid waarmee proefpersonen in staat waren antwoord te geven was deels afhankelijk van de proportie familieleden die gezien deze beslissing relevant was (de proportie objecten tegenover niet-objecten). Bovendien bleek de beslissing zelf deels afhankelijk van de status van de familieleden met betrekking tot de beslissing (mens gemaakte objecten versus natuur gemaakte objecten). Bijvoorbeeld, *appel* reflecteert een door de natuur gemaakt object. Er zijn familieleden van appel die congruent zijn met deze categorie (zoals *appelboom*). Maar familieleden zoals *appeltaart* en *appelmoes* zijn objecten door de mens gemaakt en daarmee niet congruent met de beslissing die de proefpersonen moesten maken. Als de proportie van incongruente familieleden van een woord erg hoog is, zo bleek, waren proefpersonen eerder geneigd een fout te maken.

De resultaten van deze twee experimenten zijn een direct bewijs voor de claim dat het familiegrootte-effect spreiding van activatie reflecteert. Zelfs zodanig dat de betekenis van het ongelede woord beïnvloed wordt door de activatie van de semantiek van de familieleden. Een appel is ook een beetje een appeltaart, en daarom niet meer puur natuur. Bovendien is het familiegrootte-effect niet een afspiegeling van een soort automatische globale activatie. Want, zoals gebleken uit de Hoofdstukken 2, 4, en 5, kan de de context waarin een woord verschijnt de activatie beperken tot slechts die familieleden die gezien de context nog relevant zijn.

References

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

Brysbaert, M., Lange, M. and Van Wijnendaele, I. .: 2000, The effects of age-of-acquisition and frequency-of-occurence in visual word recognition: Further evidence from the Dutch language, *European Journal of Cognitive Psychology* **12**, 65–85.

Carroll, J. B. and White, M. N.: 1973, Word frequency and age of acquisition as determiners of picture-naming latency, *Quaterly Journal of Experimental Psychology* **25**, 85–95.

Morton, J.: 1969, The interaction of information in word recognition, *Psychological Review* **76**, 165–178.

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

Whaley, C. P.: 1978, Word–nonword classification time, *Journal of Verbal Language and Verbal Behavior* **17**, 143–154.

Curriculum Vitae

Nivja de Jong werd geboren in Meppel op 11 juni 1976. Na het behalen van haar VWO-diploma in 1995 aan het Stedelijk Gymnasium Arnhem studeerde zij Nederlandse Taal en Letterkunde aan de Universiteit Leiden. Tijdens de tweede helft van haar studie werkte zij als stagiaire aan het Max Planck Instituut voor Psycholinguistiek in Nijmegen. In augustus 1999 studeerde zij in Leiden af en in juni 1999 begon zij als Ph.D-student in het PIONIER project ’The balance of storage and computation in the mental lexicon’. Dit project is ondersteund door NWO, de Katholieke Universiteit Nijmegen, en het Max Planck Instituut voor Psycholinguistiek. Vanaf juli 2002 werkt zij als research associate binnen het project ’The co-ordination of self-interruption and self-repair in speech’ van de universiteit van Edinburgh in Schotland.

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 midsagittal articulography
   Niels Schiller
3. Lexical access in the production of ellipsis and pronouns
   Bernadette 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 Austronesian 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 categorization 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 de Jong