LANGUAGE AND COGNITIVE PROCESSES, 2000, *15* (4/5), 329–365

The morphological family size effect and morphology

Nivja H. de Jong, Robert Schreuder, and R. Harald Baayen
Interfaculty Research Unit for Language and Speech,
University of Nijmegen, The Netherlands

It has been reported 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.

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 (Colé, Beauvillain, & Segui, 1989; Taft, 1979). 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.

Requests for reprints should be addressed to Nivja H. de Jong, Max Planck Institute for Psycholinguistics, P.O. BOX 310, 6500 AH Nijmegen, The Netherlands.
E-mail: nivjadj@mpi.nl

We wish to thank Andrea Krott for her helpful comments and criticisms of an earlier version of this paper. We are also indebted to Ken Paap, Ram Frost, and an anonymous reviewer for their careful criticisms. The first and third author were financially supported by the Dutch National Research Council NWO (PIONIER grant to the third author).

http://www.tandf.co.uk/journals/pp/01690965.html

Table 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 (Baayen, Lieber, & Schreuder, 1997; Schreuder & Baayen, 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

TABLE 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

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

[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 ($t_1(34) = -4.46$, $p = .0001$; $t_2(32) = -1.15$, $p = .2571$). 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).

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 thus far 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's (1975) 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 organised 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.

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 organised at the central, semantic level. In Bybee's (1985, 1995) 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 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

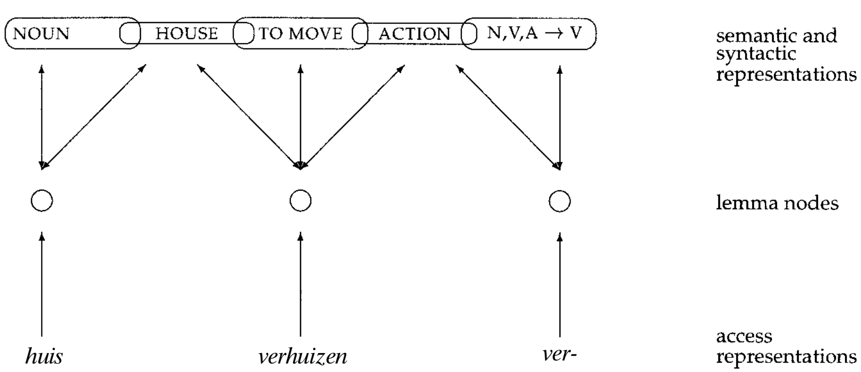

Figure 1. Representations for *huis*, "house", *verhuizen*, "move house", and the prefix *ver-* in a spreading activation model of morphological processing.

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.

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 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 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 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 4.

EXPERIMENT 1

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 3 and 4 in which we study inflected verb forms.

Method

Participants. Fourteen 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 20 verbs with a high Family Size of on average 34.7 (range 12–92, *SD* 24.5) descendants. Similarly, 20 nouns had a high Family Size with on average 36.3 (range 10–78, *SD* 19.5) descendants, and the remaining 20 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 standardised 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, six words and six 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 preceeded 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 Nec Multisync colour 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 shows the mean response latencies (calculated for the

TABLE 2
Results of Experiment 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			

correct responses) and error scores (calculated for all responses) for the four experimental conditions.

An analysis of variance revealed main effects of Word Category and Family Size, but no interaction of these two factors (F_1, $F_2 < 1$). Nouns were responded to more quickly than verbs ($F_1(1, 13) = 10.01$, $MSE = 10412.0$, $p = .0075$; $F_2(1, 76) = 9.92$, $MSE = 18457.7$, $p = .0023$), and words with a high Family Size elicited shorter response latencies than words with a low Family Size ($F_1(1, 13) = 20.73$, $MSE = 6175.5$, $p = .0005$; $F_2(1, 76) = 4.94$, $MSE = 9196.0$, $p = .0292$). 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 ($F_1(1, 13) = 13.73$, $MSE = 0.0302$, $p = .0026$; $F_2(1,76) = 8.26$, $MSE = 0.0430$, $p = .0052$). 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 respect to Family Size.

EXPERIMENT 2

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's (1997) 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, 1999).

Experiment 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 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 2A

Method

Participants. Twenty 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 comparative 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 comparatives. 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 preceeded 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 1.

Results and discussion

The participants performed the experiment with an error rate less than 15%. Table 3 lists mean reaction times (calculated for the correct responses) and mean error scores (calculated over all responses) for the

TABLE 3
Results of Experiment 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			

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 ($t_1(19) = 1.82$, $p = .0839$; $t_2(68) < 1$), nor was there any significant difference in the error scores ($t_1 = -1.34$, $p = .1956$; $t_2 < 1$). 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.

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 2B

Method

Participants. Twenty 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 1.

Results and discussion

All participants performed the experiment with an error rate less than 8%. Table 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 ($t_1(19) = -4.04$, $p = .0007$; $t_2(88) = -2.20$, $p = .0303$). The error scores showed no significant difference between the two experimental conditions (t_1, $t_2 < 1$).

Considered jointly, Experiments 2a and 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 4 below), then the results of Experiments 2a and 2b show that token frequency information is not relevant at these central levels. In the model of Schreuder & 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,

TABLE 4
Results of Experiment 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			

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 (1999), 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-occurrence 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 2a and 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 3

The aim of Experiment 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 3a uses a factorial design and parallels the monomorphemic verbs studied in Experiment 1, while Experiment 3b uses a modified regression design and compares inflected and simplex verbs directly within the same experiment.

EXPERIMENT 3A

Method

Participants. Thirty 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 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 5 lists the mean reaction times (calculated over the correct responses) and error scores (calculated over all responses) for both experimental conditions.

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 ($t_1(29) = -3.3372$, $p = .0023$) and marginally reliable by items ($t_2(44) = -1.7609$, $p = .0852$). The verbs with a low Family Size elicited significantly more erroneous responses than the verbs with a high Family Size ($t_1(29) = -3.2538$, $p = .0029$; $t_2(44) = -3.1983$, $p = .0026$). 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 1. However, the two experiments should not be compared directly because the average Base Frequency in

TABLE 5
Results of Experiment 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			

Experiment 3a is twice that of Experiment 1. In addition, the contrast in Family Size is slightly larger in Experiment 1. We therefore directly compared verbs with and without an inflectional *-t* in Experiment 3b using a factorial regression design.

EXPERIMENT 3B

Method

Participants. Thirty-two 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 $r = 0.110$ ($t(66) = 0.90$, $p = .37$); the correlation between Family Size and Surface Frequency for the subset of uninflected verbs was $r = 0.193$ ($t(66) = 1.60$, $p = .11$); the correlation between Family Size and Surface Frequency for the subset of inflected verbs was $r = -0.068$ ($t(66) = -0.55$, $p = .58$). 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, this effect cannot be contributed to a confound with Surface Frequency or Base Frequency. The materials are listed in Appendix E.

[2] Of the 68 verbs, 18 had a conversion noun, all of which occur with a frequency of less than 1 per million. Of these 18 conversion nouns, 14 occur with a frequency of less than 0.25 per million.

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 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: $r = -0.440$, $t(66) = -3.98$, $p = .0002$; uninflected verbs: $r = -0.332$, $t(66) = -2.86$, $p = .0056$). For the error scores, we find a strong trend of the expected negative correlation between log Family Size and error scores (inflected verbs: $r = -0.223$, $t(66) = -1.86$, $p = .0677$; uninflected verbs: $r = -0.228$, $t(66) = -1.90$, $p = .0615$). 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 ($t_2(67) = 4.69$, $p = .000$), higher error scores ($t_2(67) = 6.72$, $p = .000$), 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 4

The aim of this experiment is to explore the Family Size effect for perfect participles. Experiment 4a investigates the influence of the regular circumfix *ge- -d* on the activation of the family members using a factorial regression design, as Experiment 3 has shown that this design is somewhat more powerful than a factorial design. Experiment 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 4A

Method

Participants. Forty-one 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 ($r = 0.154$; $t(98) = 1.54$, $p = .13$), as well as the correlation between log Family Size and log Base Frequency ($r = 0.156$; $t(98) = 1.56$, $p = .12$) 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 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 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%. 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 ($r = -0.179$; $t(98) = -1.80$, $p = .076$). There was no correlation between error scores and Family Size ($r = -0.042$; $t(98) = -0.412$, $p = .68$).

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 than we had expected. This suggests that the Family Size effect for irregular participles might also be attenuated.

EXPERIMENT 4B

Method

Participants. Forty-one 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 which 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 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 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 ($t_1(39) = -3.67, p = .0007, t_2(37) = -2.29, p = .028$) and the participles in the high condition elicited significantly less erroneous responses than those in the low condition ($t_1(39) = -2.71, p = .0099, t_2(37) = -2.78, p = .0085$).

The regular participles in Experiment 4a and the irregular participles in Experiment 4b cannot be compared directly. Experiment 4a uses a factorial regression design, in which no correlation between Family Size on

TABLE 6
Results of Experiment 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 one hand and Surface Frequency or Base Frequency on the other hand exists. By contrast, Experiment 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 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 4b does not measure the effect of Family Size only, invalidating a direct comparison with the correlation obtained in Experiment 4a. What we can do, however, is use the linear regression fit to the data of Experiment 4a,

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

$F(1,98) = 3.22$, $p = .076$, 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 4b. This difference, 32 ms, is of the same order of magnitude as the observed difference in Experiment 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, 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 ($r = -0.294$, $t(36) = -1.85$, $p = .0365$, 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 ($r = -0.28$, $t(36) = -1.74$, $p = .046$), 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 plots the results for a range of thresholds for Experiment 4b (upper left panel) and the verbs in Experiment 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

[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 occurring in a dictionary of Dutch that do not occur in the corpus. These words are listed as having zero frequency.

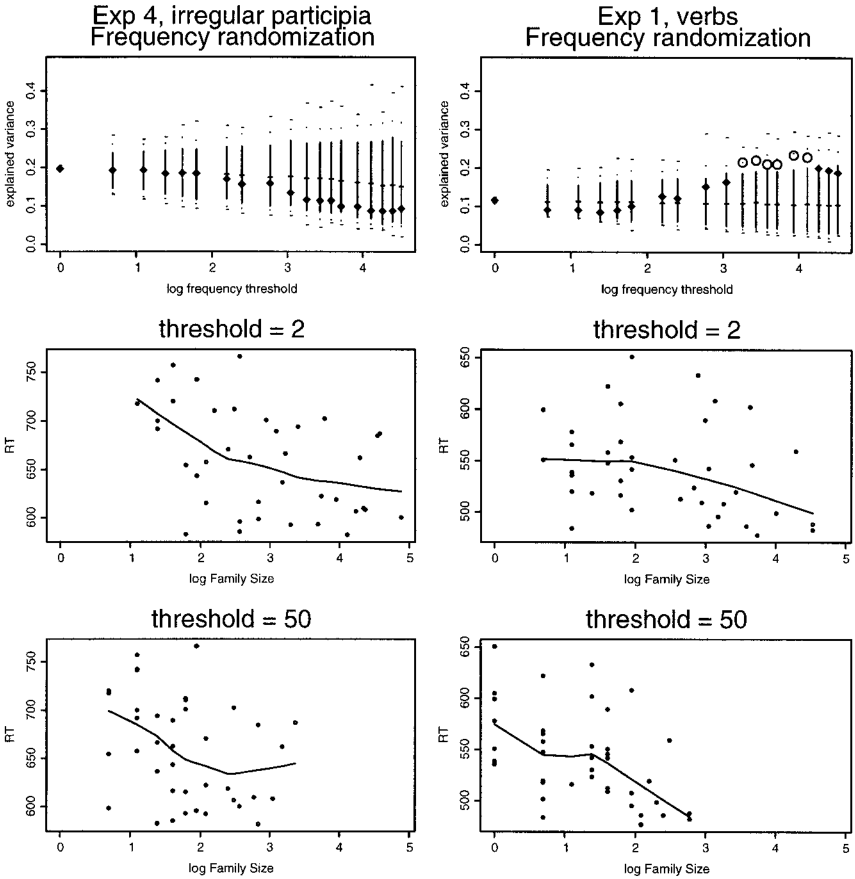

Figure 2. Frequency randomisation results for Experiment 4 (upper left) and Experiment 1 (upper right), and scatterpots of log Family Size and RT for frequency thresholds 2 (centre panels) and 50 (bottom panels) for Experiment 4 (left) and Experiment 1 (right). Frequency thresholds per 42 million.

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. For the verbs in Experiment 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 randomisation test for each threshold. One such randomisation 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 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 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.

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 do not suggest severe violations of homoscedasticity and the non-parametric regression smoothers (Cleveland, 1979) likewise suggest roughly linear trends. Table 7 lists the Pearson and Spearman correlation statistics corresponding to these scatterplots. Considered jointly, we may conclude that, apparently, in Experiment 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 3b do not show the same pattern. Possibly, the higher Base and Surface frequencies of the target verbs in Experiment 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 3b the correlation of Family Size with response latencies turned out to be higher for the verbs with an overt inflectional suffix ($r = -0.440$, $t(66) = -3.98$, $p = .0002$) than

TABLE 7

Pearson and Spearman correlations for Experiment 4b and the verbs in Experiment 1 for two different frequency thresholds

	Experiment 4b	*Verbs in Experiment 1*
Threshold 2	$r = -0.440$, $t(37) = -2.98$, $p = .0051$	$r = -0.302$, $t(38) = -1.95$, $p = .0583$
	r_s -0.428, $z = -2.64$, $p = .0084$	$r_s = -0.332$, $z = -0.208$, $p = .0379$
Threshold 50	$r = -0.314$, $t(37) = -2.01$, $p = .0514$	$r = -0.484$, $t(38) = -3.41$, $p = .0015$
	$r_s = -0.331$, $z = -2.04$, $p = .0411$	$r_s = -0.483$, $z = -3.02$, $p = .0025$

for the same verbs presented in their base form ($r = -0.332$, $t(66) = -2.86$, $p = .0056$, 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: $r = -0.367$, $t(66) = -3.20$, $p = .0021$; $r_s = -0.389$, $z = -3.19$, $p = .0014$. Verbal family members: $r = -0.100$, $t(66) = -0.81$, $p = .4182$; $r_s = -0.032$, $z = -0.26$, $p = .7942$, 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: $r = -0.435$, $t(66) = -3.93$, $p = .0002$; $r_s = -0.441$, $z = -3.61$, $p = .0003$. 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 (Baayen & Schreuder, 1999; Baayen, Schreuder, & Sproat, 1999; Schreuder & Baayen, 1995), 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 expectation 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 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 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 co-determine 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 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 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 neighbours, 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 neighbours require longer processing times than words with a small number of neighbours (e.g., Goldinger, Luce, & Pisoni, 1989; Grainger, 1990; Grainger & Jacobs, 1996), 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 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 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 follows we will sketch a tentative interpretation of the Hebrew data in relation to the data from Dutch.

Let us 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 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.

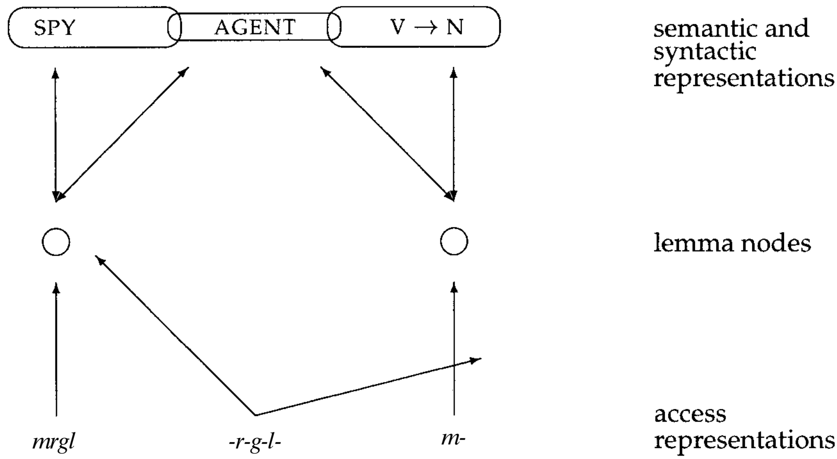

Figure 3. Representations for *mgrl*, "spy", the root *rgl*, and the prefix *m-*, in a spreading activation model of morphological processing.

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 AGENT 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 3? Second, how specific to Dutch is the Family Size effect? With respect to the first question, we note that a similar architecture is independently motivated for Dutch. Figure 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

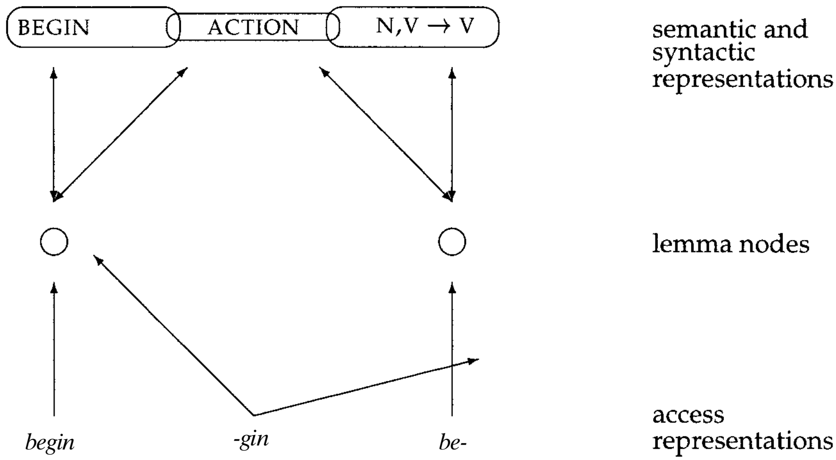

Figure 4. Representations for *begin*, "begin", the meaningless stem *-gin*, and the prefix *be-*, in a spreading activation model of morphological processing.

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*, "literally"; *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 hypothesised for

Hebrew, however, such an effect might perhaps be obtained for semantically consistent subfamilies.

Experiment 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, 1999), are severely challenged by the results of Experiment 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 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 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 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*. Cambridge, MA: MIT Press.

Baayen, R.H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, *36*, 94–117.

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

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

Baayen, R.H., & 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., Schreuder, R., & Sproat, R. (1999). Modeling morphological segmentation in a parallel dual route framework for visual word recognition. In F. van Eynde & D. Gibbon (Eds.), *Lexicon development for speech and language processing*. Dordrecht: Kluwer Academic Publishers.

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

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

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

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*, pp. 225–246. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*, 297–332.
Carreiras, M., Alvarez, C.J., & 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., & 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., & Besner, D. (1977). Access to the internal lexicon. In S. Dornick (Ed.), *Attention and performance*, Vol. VI. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
Feldman, L., Frost, R., & Pnini, T. (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., & 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.
Goldringer, S.D., Luce, P.A., & 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., & Jacobs, A.M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.
Grainger, J., & 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., & Ford, M. (1997). Morphology, modality, and lexical architecture. In G.E. Booij & J. Van Marle (Eds.), *Yearbook of morphology 1996*, pp. 117–134. Dordrecht: Kluwer Academic Publishers.
Meunier, F., & 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., & Tuominen, J. (1994). Cognitive morphology in Finnish: Foundations of a new model. *Language and Cognitive Processes*, *9*, 423–446.
Paap, K.R., & 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.
Sanchéz-Casas, R.M. (1996). Lexical access in visual word recognition: The contribution of word form. In M. Carreiras, J.E. García-Albea, & N. Sebastián-Gallés (Eds.), *Langauge processing in Spanish*, pp. 21–59. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
Sanchéz-Casas, R.M., García-Albea, J.E., & Bradley, D.C. (1991). On access representation in visual word recognition: The temporal separation technique. *Psychological Research*, *53*(1), 53–61.
Schreuder, R., & Baayen, R.H. (1997). How complex simplex words can be. *Journal of Memory and Language*, *37*, 118–139.

Schreuder, R., & Baayen, R.H. (1995). Modeling morphological processing. In L.B. Feldman (Ed.), *Morphological aspects of language processing*, pp. 131–154. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Schreuder, R., De Jong, N.H., Krott, A., & 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*, pp. 245–263. Hove, UK: Lawrence Erlbaum Associates Ltd.

Stolz, J.A., & Feldman, L.B. (1995). The role of orthographic and semantic transparency of the base morpheme in morphological processing. In L.B. Feldman (Ed.), *Morphological aspects of language processing*, pp. 109–129. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Taft, M., & 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., & Schreuder, R. (1994). The role of analogy in the interpretation of novel compounds. *Journal of Psycholinguistic Research*, *23*, 111–137.

Zhou, X., & Marslen-Wilson, W. (1999). Lexical representation of compound words: cross-linguistic evidence. *Unpublished manuscript.*

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 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 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 (colonel) 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 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 1, verbs with a low Family Size

weifel (waver) 651, dwing (force) 530, tuur (peer) 551, streel (caress) 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 (gnaw) 502.

APPENDIX B

Target words with reaction time

Experiment 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 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 (caresses) 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 (creepier) 540, stinkt (stinks) 528, kraait (crows) 561.

APPENDIX C

Target words with reaction time

Experiment 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 (stuck) 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 (rhymes) 495, kamt (combs) 549.

Experiment 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 (despises) 538, deint (heaves) 665, aarzelt (hesitates) 538, woester (more savage) 526, juichte (cheered) 554, deugt (is good) 527, vromer (more pious) 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 3a, inflected verbs with a high Family Size

spoedt (urges) 578, raapt (gathers) 590, kapt (does one's hair) 642, braakt (vomits) 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 (rhymes) 548.

Experiment 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 (aims) 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 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) (idolise) 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) (despise) 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 (grabbed) 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 (drooled) 764 2, gelasterd (insulted) 812 12, gelummeld (hung around) 916 5, gemijmerd (mused) 733 3, gemopperd (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 (rhymed) 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 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 (ground) 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 (poured) 592, gewezen (pointed) 662, gedreven (floated) 582.

Experiment 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.