Morphological resonance in the mental lexicon

Nivja H. de Jong, Robert Schreuder and R. Harald Baayen

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

1. Introduction

Various token frequency effects are known to influence cognitive processing (Hasher and Zacks 1984). In the lexical domain, it is well-known that high-frequency words are processed faster than words with low frequencies of occurence (e.g., Rubenstein and Pollack 1963; Whaley 1978; Taft 1979). Recently, another phenomenon has been observed to play a role in lexical processing: Simplex words which occur as constituents in many complex words are responded to faster in a visual lexical decision task than words with only a few morphological family members. Words with many morphological family members also receive higher subjective frequency ratings than words with only a few morphological family members. The token frequencies of the family members are found to be irrelevant, only the number of family members plays a role. This effect of morphological family size is especially interesting from a cognitive perspective in that it is a type frequency effect without a concomitant token frequency effect, in that the type count of morphologically related family members plays a role, but their token frequencies do not. The effect has been observed both for monomorphemic words and for stems in complex words (Schreuder and Baayen 1997; Baayen, Lieber, and Schreuder 1997; Bertram, Baayen, and Schreuder 1999; De Jong, Schreuder, and Baayen 2000).

Three lines of evidence indicate that the family size effect is semantic in nature, and arises at post-identification stages of lexical processing due to activation spreading along lines of shared morpho-syntactic representations. First, the effect disappears when progressive demasking is used instead of lexical decision (Schreuder and Baayen 1997). In a progressive demasking task, participants are asked to identify target words which are masked. In successive cycles, the mask is shown for shorter latencies and the words are shown for longer latencies, such that the target words gradually seem to emerge from the mask. If progressive demasking is primarily sensitive to the early processes of form identification as argued by Grainger and Jacobs (1996; see De Jong, Schreuder, and Baayen 2000, for discussion), the absence of a family size effect in progressive demasking can be understood if it arises after form identification has been completed. Second, correlation studies show that semantically opaque family members (such as *business* as morphological relative of *busy*) do not contribute to the effect. Correlations of family size and response latencies are higher when opaque family members are removed from the counts (Schreuder and Baayen 1997; Bertram, Baayen, and Schreuder 1999). Third, De Jong, Schreuder and Baayen (2000) show that the family size effect is carried by the underlying lemma (Levelt, 1989) and not by the actual phonological and orthographic form of the word. For instance, irregular Dutch past participles such as *gevochten*, derived from the verb stem *vecht*, 'to fight', co-activate all morphologically complex words derived from *vecht*, even though almost all these morphological relatives contain the form *vecht* and not a form containing the string *vocht*. In fact, *vocht* happens to be an independent Dutch noun ('moisture'), and no correlation appears to exist between the family size of such nouns and the response latencies to semantically unrelated past participles such as *gevochten*.

The present study was prompted by two at first sight unrelated findings. First, Bertram et al. (1999) report that the family size effect for Dutch de-adjectival abstract nouns with the suffix *-heid* (*eenzaam-heid*, 'loneli-ness') seems to be restricted to a specific subset of morphological relatives. This subset includes complex nouns and verbs with adjectival stems (e.g., *ver-eenzaam-en*, 'to become lonely') without further restrictions. The set of adjectives in the family, however, appears to exclude color compounds such as *blauw-groen*, 'blue-green', and intensified adjectives such as *ijs-koud*, 'ice-cold'. Interestingly, these are exactly the adjectives to which *-heid* is hardly ever attached (? *ijskoudheid*, ? *ice-coldness*). Bertram et al. (1999) point out that a semantic restriction on suffixation of *-heid* (avoid intensified adjectives and color compounds) that must be operative in language production also seems to play a role during comprehension (intensified adjectives and color compounds are not co-activated). In what follows, we will refer to the intensified adjectives and color compounds as scale-focusing adjectives, and to the remaining adjectives as general adjectives. We refer to the intensified adjectives and the color-compounds as scale-focusing because the modifiers in these complex adjectives narrow down the general meaning of the adjectival head to a specific part of the scale covered by the head: *ice-cold* denotes an extreme location on the scale of coldness, and likewise *blue-green* denotes a particular shade of green in the range of hues covered by the general term *green*. Second, De Jong et al. (2000) report that the presence of an overt verbal inflectional suffix in Dutch verbs triggers greater co-activation of verbal family members compared to Dutch verb forms without an overt verbal marker. In Dutch, first person present tense verb forms do not carry an affix, whereas the third person present tense is the stem plus the inflectional suffix *-t*. In De Jong et al. (2000), we presented both forms of the same verb stems (e.g., *sjouw* and *sjouwt*, 'drag' and 'drags') and found that verbal family members (e.g., *wegsjouwen*, 'to drag away') only contributed to the effect of family size in the case that the verbs were presented with the overt inflectional marker.

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

First, we expect that the adjectives in the morphological family will contribute more strongly to the family size effect than in the neutral condition. This expectation is based on the finding that verbal family members contribute more to the family size effect in the presence of the overt verbal inflectional suffix *-t*. Just as the *-t* boosts the contribution of the verbs in the family, the comparative suffix *-er* might boost the contribution of the adjectives in the family. Likewise, the adjectival modifier *heel* might also boost the contribution of adjectival family members. This hypothesis is based on the observation that *heel* predominantly precedes adjectives, whereas *niet* does not show such a prevalence. Indeed, in bigram probabilities derived from the corpus used for CELEX, *heel* and *niet* are found to combine with different word classes.

Table 1. Number of word-form types within different word classes which follow *heel* and *niet* according to bigram probabilities in the corpus used for CELEX, percentages given in parentheses.

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

Table 1 shows the number of times *heel* and *niet* are followed by different word-form types of adjectives, verbs, nouns, and other words. The number of word types following *heel* is lower than for *niet* (631 and 7093 respectively). A X^2-test revealed that the distributions among the word classes for these two contexts differ significantly $X^2(3)=896.55$, $p=0.000$). As can be seen in Table 1, which also shows the percentages of word types following *heel* and *niet* in these different word classes, this difference in distribution is mainly due to the fact that *heel* preceding adjectives is overrepresented, and *niet* preceding verbs is overrepresented. It should be noted that these numbers only provide a rough estimation, as many word forms are ambiguous with respect to their word class. For instance, as participles (which we counted as verbs) can functionally be adjectives, the number of word types that follow *heel* and that are unambiguously verb forms, reduces from 124 to a mere 14. From these distributional properties, we hypothesize that the VERY condition will serve as a non-neutral condition, in that the adjectival subfamilies of the targets might be boosted, wheras in the *not* condition, such a preference for adjectival family members should not occur (and, perhaps, a preference for verbal family members can be expected).

Second, we also expect that scale-focusing adjectives (color compounds and intensified adjectives) might not contribute to the family size effect in the non-neutral conditions. The comparative suffix may well be subject to the same semantic restrictions as reported for *-heid* by Bertram et al. (1999), as formations such as ? *ijs-koud-er*, '? ice-cold-er' and ? *blauw-groen-er*, '? blue-green-er' seem ungrammatical. Likewise, phrases such as *very icecold* and *very blue-green* seem odd, possibly because, e.g., *icecold* itself is already as cold as you can get. Note that for all four conditions, the focus of our interest is on the way in which subsets of morphological family members are activated as a function of morphological and phrasal context in which the adjective stems occur.

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

We carried out two experiments in order to ascertain the role of the morphological and phrasal context on the activation of the morphological family members. The first experiment contrasted simplex adjectives (the BASE condition) and the same adjectives in the context of the comparative suffix *-er* (the COMPARATIVE condition). Experiment 2 used the same set of adjectives and varied the phrasal context. In the NOT condition, the adjective was preceded by *niet*, and in the VERY condition, it was preceded by *heel*. Both experiments made use of a within-subject design.

2. Experiment 1

2.1. Method

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

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

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

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

2.2. Results

The participants performed the experiment with an overall error rate less than 15%. For each word we calculated mean response latencies (over the correct responses) and error scores (over all responses). The upper half of Table 2 shows the means and standard deviations for the experimental conditions of the experiment.

Table 2. Results of Experiment 1 and 2: Means and standard deviations of response latencies and error proportions (by participants).

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

An analysis of variance for reaction times revealed a main effect for Context (BASE and COMPARATIVE) and a main effect for Family Size, but no interactions (by participants and by items, $F1$ and $F2<1$). The monomorphemic adjectives were responded to faster than the comparatives ($F1(1,31)=61.33$, MSE=91123.9, $p=0.00$; $F2(1,76)=19.69$, MSE=68862.7, $p=0.00$), probably due to the higher Surface Frequencies of the monomorphemic adjectives. Words with a high Family Size were responded to faster than words with a low Family Size ($F1(1,31)=29.30$, MSE=32420.1, $p=0.00$;$F2(1,76)=7.97$, MSE=27865.0, $p=0.01$). An analysis of variance for the error scores shows a similar pattern, but the main effect of Family Size is not reliable in the item-analysis.

The comparatives elicited significantly higher error scores than the monomorphemic adjectives ($F1(1,31)=22.72$, MSE=0.08, $p=0.00$; $F2(1,76)=19.69$,MSE=68862.7, $p=0.00$) and the words with a low Family Size elicited higher error scores than the words with a high Family Size ($F1(1,31)=8.03$, MSE=0.03, $p=0.01$; $F2(1,76)=2.35$, MSE=0.02, $p=0.13$). Again, no interactions were found ($F1$,$F2<1$).

3. Experiment 2

3.1. Method

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

Materials. We used the same 40 monomorphemic adjectives as in Experiment 1, but presented them following the word *heel*, 'very', or following the word *niet*, 'not', creating combinations like *heel mooi*, 'very beautiful', or *niet mooi*, 'not beautiful'. The two words were presented simultaneously on the computer screen. We added the same 104 fillers of nouns and verbs. The filler nouns were presented with either a definite article *de, het*, 'the', or with the indefinite article *een*, 'a'. For mass nouns, which syntactically cannot be presented with the indefinite article, we used *wat*, 'some'. The filler verbs were presented following the personal pronoun *ik* or followed by the personal pronoun *jij?* (and a question mark). For example *ik kwets*, 'I hurt', or *kwets jij?* ', 'do you hurt? '. A participant had to respond either to the adjective presented together with *heel*, or to the adjective presented together with *niet* but never to both. The pseudowords (identical to the ones used in the previous experiment) were presented with the same contexts as the words and in the same proportions. In this way, if a phrase contained a nonword, the nonword was in the majority of the cases the second word in the phrase (following *heel, niet, de, het, een, wat*, or *ik*), but could also be presented in the first position (followed by *jij?*). The experiment was preceded by 24 practise items. There was a short pause after the practise session, and a short pause halfway through the experimental list. In total, the experiment lasted approximately 15 minutes.

Procedure. The procedure was almost identical to that of Experiment 1, except that participants were now asked to decide whether the two letter strings that appeared on the computer screen were real Dutch words. Each stimulus was preceded by a fixation mark in the middle of the screen for 500 ms. After 50 ms, the two words (the whole phrase) appeared centered at the same position. Stimuli were presented on Nec Multisync color monitors in white lowercase 21 point Helvetica letters (instead of 36 in Experiment 1) on a dark background and they remained on the screen for 1500 ms. The maximum time span allowed for a response was again 2000 ms from stimulus onset.

3.2. Results

The participants performed the experiment with an overall error rate less than 15%. For each word we calculated mean response latencies (over the correct responses) and error scores (over all responses). The bottom half of Table 2 shows the means and standard deviations for the experimental conditions of the experiment. An analysis of variance revealed a reliable main effect for Context (VERY and NOT): Adjectives presented with the word *heel* were responded to faster than adjectives presented with the word *niet* ($F1(1,31) = 18.39$, $MSE = 40893.7$, $p = 0.000$; $F2(1,76) = 7.01$, $MSE = 26949.9$, $p = 0.010$). This may be due to the increased difficulty in the NOT condition to respond with "yes" while processing a word meaning "not". Adjectives with a high Family Size were responded to faster than adjectives with a low Family Size, but this main effect of Family Size was not reliable in the by-item analysis ($F1(1,31) = 14.00$, $MSE = 16375.5$, $p = 0.001$; $F2(1,76) = 2.57$, $MSE = 9891.2$, $p = 0.113$). No interactions were found ($F1$, $F2 < 1$). An analysis of variance of the error scores revealed a main effect for Word Context only: Words presented with *niet* elicited more erroneous responses than the words presented with *heel* ($F1(1,31) = 18.39$, $MSE = 40893.7$, $p = 0.000$; $F2(1,76) = 7.01$, $MSE = 26949.9$, $p = 0.010$), suggesting that indeed the semantics of *niet* interfered with providing the correct response. All other F-values for the error analysis were less than 1. Alternatively, as one of our reviewers pointed out to us, there might be a difference in reaction times and error scores between these two phrasal contexts due to a difference in scope. The word *niet* can have scope over a single constituent or over an entire sentence, whereas the intensifier *heel* predominantly has a narrow scope. Thus, the phrases of the adjectives in the *niet* context in our experiment were ambiguous, which might have affected response latencies and error scores, especially since the wide scope reading entails treating phrases such as *niet mooi*, 'not beautiful' as truncated elliptical sentences.

Summing up the results for Experiment 1 and 2 with respect to the effect of family size, the factorial contrast between a high and a low family size was reliably reflected in the response latencies to the adjectives in the BASE and COMPARATIVE conditions. For the phrasal conditions, the family size effect was weaker and did not reach significance in the by-item analysis. The reason why the family size effect might be attenuated for words presented in a phrasal context becomes apparent when we compare the correlational structure between the four contexts of these experiments.

4. Post-hoc correlations

Table 3 shows the by-item Spearman correlations between response latencies (RT) and Family Size for the conditions in both Experiments, reliable except for the context VERY in Experiment 2. To understand why the VERY context behaves differently, we consider, for each of the four context conditions, the correlational structure of the response latencies to the 40 targets with different subsets of family members. We divided the total families into four subsets of family members. First, the nominal family members (N), second, the verbal family members (V), and two kinds of adjectival family members: Scale-focusing adjectives (color compounds and intensified adjectives), and general adjectives. For example, *mooi* has in total 8 family members. Dividing them up into the four subsets, we find that *mooi* has 3 nominal family members, 1 verbal family member, 5 family members which are general adjectives, and 3 adjectival family members which are scale-focusing.

Table 3. Spearman correlations and p-values of Family Size and RT for four morphological and phrasal contexts.

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

Table 4 shows the properties (means and standard deviations) of these four subsets of family members for all 40 targets. In the four contexts, the counts for these subsets of family members will remain the same, as accross contexts, the same 40 targets were presented. This enables us to compare the correlational structures in the four different contexts. Recall that we expect adjectival family members to contribute more to the family size effect in the non-neutral conditions (VERY and COMPARATIVE) than in the neutral conditions (BASE and NOT). Also recall that we hypothesized that scale-focusing adjectives might not contribute to the family size effect in the non-neutral conditions.

Table 4. Means and standard deviations of the different subfamilies of the 40 target words.

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

Figure 1 plots $-r_s$ as a function of the size of the four subfamilies: nouns (N), verbs (V), general adjectives (A1), and scale-focusing adjectives (A2). As expected, which subgroup of family members correlates best with reaction times of the 40 target words differs from context to context. In the BASE condition, shown in the upper left panel of Figure 1, the family size effect is driven by the general adjectives, the nouns, and also to some extent the verbs in the family. The scale-focusing adjectives do not contribute at all to the family size effect. In the COMPARATIVE condition, shown in the upper right panel, the general adjectives stand out with a particularly high correlation. The lower left panel plots the correlations for the VERY condition, and shows that here, surprisingly, the scale-focusing adjectives constitute the primary subfamily responsible for the family size effect.

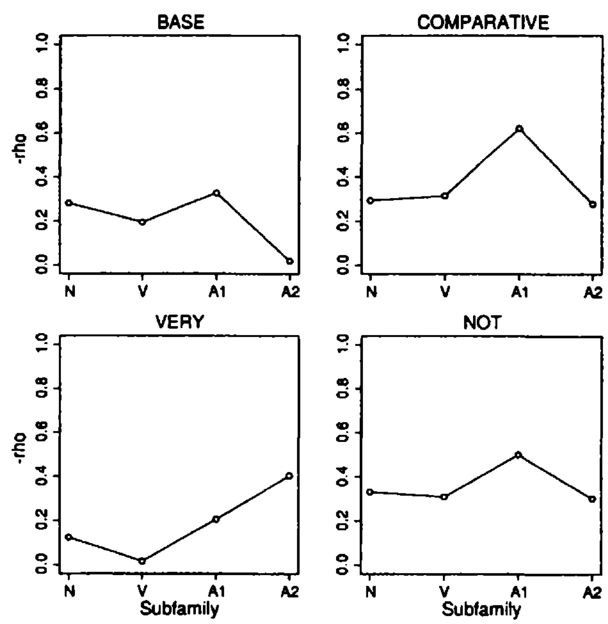

Figure 1. Pearson correlations of Reaction Times with the family counts of nouns (N), verbs (V), general adjectives (A1), and scale-focusing adjectives (A2) for four experimental conditions.

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

Interestingly, our original predictions are in part confirmed by the data, and in part refuted. What is confirmed is the prediction that adjectives should contribute more strongly to the family size effect in the non-neutral conditions. Although we initially classified the NOT condition as neutral, it appears to behave as a non-neutral condition similar to the COMPARITIVE condition. This may be due to the actual experimental context, in which *niet* ('not') was always followed by an adjective, effectively turning this supposed neutral condition into a non-neutral condition in the sense that it favors adjectival family members to become co-activated. Our predictions concerning the scale-focusing adjectives turned out to be wrong. First consider the BASE condition. In the absence of a suffix such as *-heid* ('-ness') that might exclude incompatible family members from contributing to the family size effect, we expected the scale-focusing adjectives to fully participate, contrary to fact. This behavior of the scale-focusing adjectives is probably due to the small number of scale-focusing adjectives in the pooled families of our experimental words. In the study of Bertram et al. (1999), the scale-focusing adjectives comprised half of all adjectival family members. In the present study, they comprise roughly one fifth of all the adjectives, which suggests that there might be too few scale-focusing adjectives to effectively co-determine the response latencies in the BASE condition.

However, when we turn to the lower left panel of Figure 1, we find that the scale-focusing adjectives reveal the strongest correlation of all subfamilies when the adjective is preceded by *heel* ('very'). Apparently, a small number of scale-focusing adjectival family members is still able to give rise to strong correlations with the response latencies, provided that the experimental adjective appears in the right context. The crucial property of the context supplied in the VERY condition seems to be the near synonymy of *very* with intensifiers such as *ice* in *ice-cold*, which has a meaning that comes close to that of *very cold*. For the color compounds within the subfamily of scale-focusing adjectives (e.g., *blauwgroen*, 'blue-green'), it can be argued that although the first constituent of the compound does not intensify the color of the second constituent, it does narrow down on the scale of (actually both) colors. This scale-focusing property is shared with the intensifier *heel*, 'very'. It should be mentioned, however, that only two target words in our experiments were colors themselves (*blauw*, 'blue' and *groen*, 'green') and excluding these two targets from the analysis does not change the results. Therefore, with these results we can only speculate as to whether these few color compounds within the subset of scale-focusing adjectives were actually contributing to the family size effect in the VERY condition. Nevertheless, the BASE and VERY conditions clearly show that the contribution of a subfamily may be crucially determined by phrasal context.

The observation that different subfamilies are primarily responsible for the family size effect, an observation which also figures in the simulation study reported below, sheds light on why the main effect of family size is not reliable in the *F*2-analysis in Experiment 2. The orthogonal contrast built into the experiment assumed that all family members would play a role, a contrast of 52 versus 3. However, if we count exclusively general family members, the contrast is 7.3 in the high condition and 0.3 in the low condition. Restricting the count to the scale-focusing family members results in a contrast of 3.6 versus 0.7. Counting the family size in this restricted way results in 'high' conditions with hardly any items with a truly large family size. We suspect that this effective orthogonal manipulation was too weak to show up in the by-item analysis of variance. Nevertheless, by using the actually relevant subfamily counts, reliable correlations emerge.

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

5. Simulation studies

The architecture of the Morphological Family Resonance Model, henceforth MFRM, is sketched in Figure 2. As the family size effect is a central semantic effect, the MFRM focuses on the lemma representations of words in the sense of Levelt (1989) and their associated syntactic and semantic representations in the sense of Schreuder and Baayen (1995). It is assumed that the visual presentation of a target word leads to the activation of the corresponding access representation, which in turn activates the target's lemma. The MFRM models what happens once the target lemma has been activated.

Figure 2 displays four lemma representations at the left hand side of the graph, in the array labelled *L*. The target lemma, *greatness*, is underlined, and its family members are shown in italics. One word, *think*, does not belong to the morphological family of *greatness*. At the right hand side of the graph, three arrays of central representations are shown: syntactic representations labelled *S*, affix representations labelled *D*, and meaning representations *M*. As in the model outlined in Schreuder and Baayen (1995), central representations are shared by the lemmas. For instance, the meaning representation *great* is shared by the lemmas *great*, *greater*, and *greatness*.

A resonance cycle in the MFRM consists of two stages. In the first stage, activation spreads from the lemmas to the central representations. In the second stage, activation spreads back from the central representations to the lemmas. The flow of activation during the very first resonance cycle is indicated by solid lines in Figure 2.

The additional flow of activation that occurs during subsequent resonance cycles is indicated by dashed lines. Over time, the activation of the target lemma increases exponentially, with the rate of increase being determined by the extent of the morphological family.

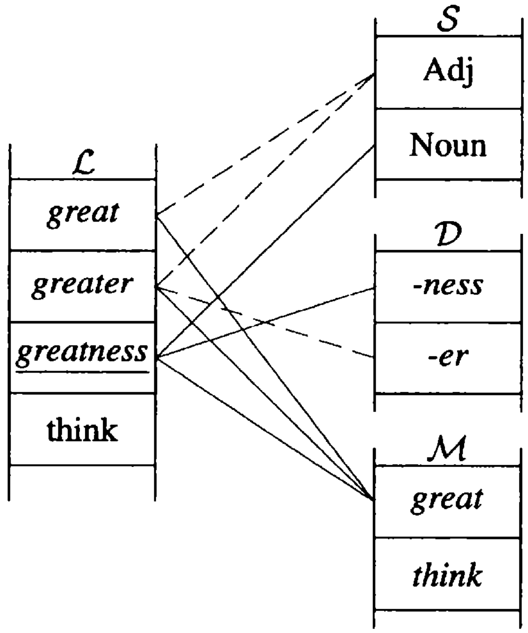

Figure 2. Resonance at the initial timestep (solid lines) and additional resonance at following timesteps (dotted lines) for the target word *greatness* (underlined). *L* stands for lemma representations, and *S*, *D* and *M* represent central representations: syntax, affixes, and meanings respectively.

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

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

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

$$f(l_i) = \{l_j : \mathcal{E}_j \cap \mathcal{E}_i \neq \emptyset \wedge \mathcal{M}_j \cap \mathcal{M}_i \neq \emptyset\}. \qquad (1)$$

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

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

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

$$\begin{aligned} a(x,t) &= \delta\{a(x,t-1) + \rho_{\mathcal{X}}[\alpha + a(x,t-1) \\ &+ \sum_{l_j \in f(l_i)} \mathrm{I}_{[x \in \mathcal{X}_j]}(a(l_j,t-1) + \mathrm{I}_{[j=i]}\alpha)]\}. \end{aligned} \tag{2}$$

In (2), δ ($0 \leq \delta \leq 1$) represents a global decay rate, and ρ_X ($0 \leq \rho_X \leq 1$) represents the resonance sensitivity for the different kinds of central representations *S,D*, and *M*. The idea is that different central subsystems can participate in the resonance to different degrees. Technically, the differential resonance sensitivities allow us to avoid situations in which a particular subset of central representations becomes overly dominant in the resonance.

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

$$\begin{aligned} a(l_k,t) &= \mathrm{I}_{[l_k \in f(l_i)]}\delta\{a(l_k,t-1) + \mathrm{I}_{[k=i]}\alpha \\ &+ \rho_{\mathcal{L}}[\alpha + a(l_k,t-1) + a(s_k,t) + \\ &+ \frac{1}{c(\mathcal{D}_k)} \sum_{d \in \mathcal{D}_k} a(d,t) \\ &+ \frac{1}{c(\mathcal{M}_k)} \sum_{m \in \mathcal{M}_k} a(m,t)]\}. \end{aligned} \tag{3}$$

In (3), ρ_L ($0 \leq \rho_L \leq 1$) denotes the resonance sensitivity of the lemmas. The cardinality of the sets of affixes and meanings is denoted by $c(D)$ and $c(M)$ respectively. The factors $1/c(D)$ and $1/c(M)$ in (3) ensure that lemma l_k will always receive the same amount of activation from any of the three sets of central representations, irrespective of the number of affix and meaning representations to which it is connected. This normalization ensures that resonance among family members sharing many semantic and/or affix representations does not become so strong that family members reach threshold activation level before the target word.

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

$$a(x,t) = \delta a(x,t-1). \tag{4}$$

All simulations reported were run with α=1.0, δ=0.98, ρ_L=ρ_M=0.02, ρ_S=ρ_D=0.01, and θ=100.0.

Having completed the formal definition of the MFRM, we now turn to consider the model's performance. In the simulation runs, the activation levels of all units in the model are reset to their default values between trials.

Figure 3 plots the correlations of the observed RTs and subfamily sizes for the four experimental conditions by means of solid lines. The corresponding correlations of the model times and subfamily sizes are represented by dashed lines. We observe that the correlations generated by the model are similar to the empirical correlations apart from a shift along the vertical axis. The model predicts much higher correlations than we actually observe. This is not so surprising, as the model's predictions are based on resonance only, without taking into account the effect of word frequency and the many other factors that co-determine response latencies.

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

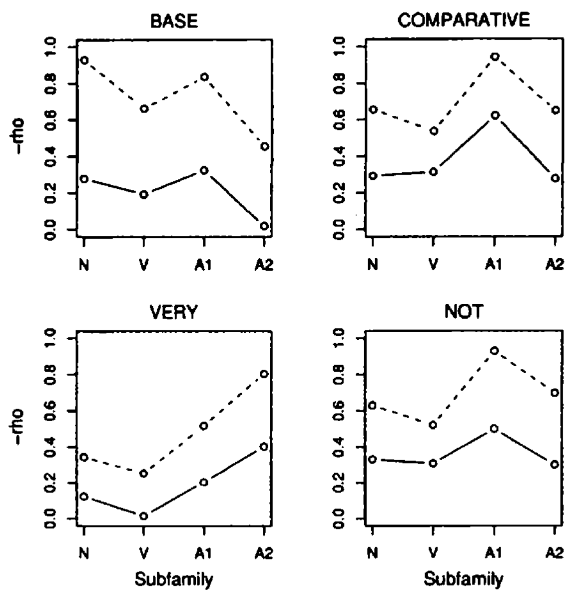

Figure 3. Pearson correlations of Reaction Times (solid lines) and model times (dashed lines) with the family counts of nouns (N), verbs (V), general adjectives (A1), and scale-focusing adjectives (A2) for four experimental conditions.

For the COMPARATIVE and the NOT contexts, a good fit required restricting the resonance set to the adjectives, including both the general and the scale-focusing adjectival family members.[1] In the case of the VERY condition, by contrast, a good fit required restricting the resonance set to the scale-focusing adjectives only, i.e., to the color compounds (*blauwgroen*, 'blue-green') and intensified adjectives (*steenkoud*, 'stone-cold'). Intensified adjectives were coded in the model's lexicon with 'very' as part of their semantic representation (for *steenkoud*, the entry was <*steenkoud*, {Adj}, Ø, {*very, cold*}>.

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

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

Table 5 lists the Spearman correlations and their associated p-values for the four subsets of data. The first two columns again present the correlations of the raw family counts with the response latencies. The second two columns represent the correlations of the family counts restricted to the subfamilies in the resonance sets and the response latencies. Note that the correlation for the VERY context, which is not significant given the raw family count, is significant given the appropriate subfamily (the effective family size, i.e., the resonance set as determined on the basis of the simulation model). The third two columns present the correlations of the model times with the response latencies, all of which are comparable to those of the effective family size.

Table 5. Spearman correlations of Family Size and RT (left columns), of the Effective Family Size and RT (center columns), and of Model Times and RT (right columns) for four morphological and phrasal contexts.

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

6. General discussion

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

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

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

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

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

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

Author note

We wish to thank the reviewers Harald Clahsen and Gary Libben for their helpful comments on an earlier version of this paper. This study was financially supported by the Dutch National Research Council NWO (PIONIER grant to the third author), the University of Nijmegen (The Netherlands), and the Max Planck Institute for Psycholinguistics (Nijmegen, The Netherlands).

Appendix

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

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

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

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

Notes

1. Similar fits are obtained when the resonance sets contain only the general adjectives. Hence, our results do not allow us to decide whether *-er* behaves similarly to *-heid*, in the sense that only general adjectives are co-activated in the family. Although our results are compatible with a parsimoneous resonance set containing only general adjectives for the COMPARATIVE context, the results obtained for a larger resonance set with both general and scale-focusing adjectives are as good.

References

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

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

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

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

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

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

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

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

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

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

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

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

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