Abstract

Are regular morphologically complex words stored in the mental lexicon? Answers to this question have ranged from full listing to parsing for every regular complex word. We investigated the roles of storage and parsing in the visual domain for the productive Dutch plural suffix -en. Two experiments are reported that show that storage occurs for high-frequency noun plurals. A mathematical formalization of a parallel dual route race model is presented that accounts for the patterns in the observed reaction time data with essentially one free parameter, the speed of the parsing route. Parsing for noun plurals appears to be a time-costly process, which we attribute to the ambiguity of -en, a suffix that is predominantly used as a verbal ending. A third experiment contrasted nouns and verbs. This experiment revealed no effect of surface frequency for verbs, but again a solid effect for nouns. Together, our result suggest that many noun plurals are stored in order to avoid the time-costly resolution of the subcategorization conflict that arises when the -en suffix is attached to nouns.

Many mental activities require forms of computation. Some computations must be carried out so many times that it becomes advantageous to store the results, so that the outcome can immediately be retrieved from memory. This shortcut is especially useful in case the computations require a substantial amount of time. For instance, in the time course of movement planning, the computation of manual reaches has been found to be expensive compared to storing known reaches (Rosenbaum, Vaughan, Barnes & Jorgensen, 1992). Similarly, in carrying out arithmetic operations, adult performance on single digit operand problems appears to reflect direct retrieval of facts from memory, whereas more complex calculations require computational effort beyond memory retrieval (see Rickard, Healy & Bourne, 1994, and references cited there). In the domain of psycholinguistics, similar issues arise with respect to the processing of regular morphologically complex words. Are regular complex words stored as wholes in the mental lexicon, or does some form of computational decomposition take place in perception, and some form of computational composition in production? Again the question is how the relative costs and advantages of storage and computation are balanced (Frauenfelder & Schreuder, 1992).

Pinker (1991) and Pinker and Prince (1991) have argued for language production that morphologically complex words with at least one idiosyncratic property have to be stored in the mental lexicon. Conversely, words that are fully regular and transparent, both with respect to their phonological form and with respect to their semantics, are never stored as such. They are perceived and produced via their constituent morphemes. According to Pinker and Prince, the absence of storage is characteristic for the domain of inflection. For derived words, where individual complex words often assume unpredictable shades of meaning, storage might be more pervasive.

Others have argued that even fully regular and phonologically and semantically transparent complex words, especially in the higher-frequency ranges, are stored in the mental lexicon. Stemberger and MacWhinney (1986, 1988) have observed empirically that in speech production a high frequency of use seems to protect inflected verbs against mispronunciations. This, they argue, is best understood in terms of storage of the full inflected form.

High-frequency inflected words are retrieved as wholes from the mental lexicon. Hence, mispronunciations of high-frequency words that would arise as a consequence of the morphological composition process cannot occur. Conversely, low-frequency inflected forms have to be constructed on-line, and are thus more prone to speech errors.

Within the domain of language comprehension, similar claims have been made, varying from pervasive storage (e.g., Seidenberg, 1987) to minimal storage (Taft and Forster, 1975) and various intermediate positions (e.g., Frauenfelder and Schreuder, 1992). In this paper, we address the issue of the relative roles of both storage and computation for the domain of visual word recognition. Our research focus will be on the early stages of word recognition and on the possible functional role of modality-specific access representations for fully transparent regular complex words. We will first review a number of influential models of morphological processing that describe these early, form-sensitive, stages of word recognition. We then present two experiments in which the effects of whole-word frequency and stem frequency are explored for noun singulars and plurals, as well as one experiment in which nouns are contrasted with verbs. The experimental results are interpreted within the modeling framework presented in Schreuder and Baayen (1995). By means of mathematical formalizations of this verbal model, accurate estimates of mean reaction time are obtained on the basis of surface and stem frequencies in combination with essentially one free parameter, the speed of the parsing route.

PROCESS MODELS

All process models of word recognition have to account for the importance of word frequency as a factor determining the speed with which a word is recognized. This phenomenon is generally modeled in terms of the resting activation level of a word's access representation. High-frequency words have higher resting activation levels than low-frequency words. Hence, high-frequency words need less incoming information from the signal to reach threshold activation level. While there is general consensus that such frequency-sensitive access representations are crucially involved in the recognition of monomorphemic words, there is an ongoing debate on whether polymorphemic words have their own access representations.

Three different basic architectures for the processing of morphologically complex words have been proposed. The first architecture, the full listing model, is a very simple one in which all words are stored, irrespective of their morphological constituency. Apart from lexical look-up, no computations are involved. This position has been argued for by Butterworth (1983). Butterworth leaves open the possibility that substantive morphological rules are available as back-up procedures, but he slightly favors the view according to which these backup rules are "meta-rules" of an analogical nature that exploit similarities between stored forms (p. 290). Full listing models predict that the frequency of a word, be it monomorphemic or morphologically complex, should be a main determinant of the speed with which it is recognized. Some researchers have reported that the summed frequencies of a stem and all words in which that stem occurs, the so-called cumulative stem frequency, is a factor co-determining response times (e.g., Burani & Caramazza, 1987; Colé, Beauvillain & Segui, 1989; Taft, 1979). The full-listing model, however, excludes the possibility that such cumulative frequency effects arise. Since the constituents of a complex word do not play an independent role in the early stages of word recognition, it is only the frequency of the complex word itself, its so-called surface frequency, that can influence the speed of word recognition in the full-listing architecture.

The second architecture is one in which fully regular and transparent words are always recognized on the basis of their constituents. Central to this architecture is a parser that identifies the constituents of a complex word and that computes the meaning of that complex word on the basis of the meanings of the parts that have been identified. This type of model, which we will refer to as a full parsing model, predicts that the time required to access the meaning of a stem is determined by the cumulative frequency of that stem. All regular complex words containing a given stem always require identification and access to that stem in order to compute their meaning. Hence, the activation level of the stem is a function of the summed frequency of all the (transparent) words in which it occurs. Comparing processing times for a monomorphemic stem and a complex word in which that stem occurs, this kind of model predicts that processing the complex word should require more time. Parsing the complex word and computing its meaning on the basis of its constituents probably takes some extra time in addition to the time required for identification of the constituents themselves.

Note that full parsing models predict that surface frequency should not play a role in word recognition. Only the elementary building blocks of words, the morphemes, are frequency-sensitive.

The third logically possible architecture combines whole-word access with parsing in a dual route model. In dual route models, a direct route that makes use of full-form access representations is combined with a parsing route. One such model is the Augmented Addressed Morphology model (AAM). In earlier papers describing the model (Burani & Caramazza, 1987; Caramazza, Laudanna & Romani, 1988; Laudanna & Burani, 1985), "known" words are all handled by the direct route, and the parsing route is a backup option for rare or novel morphologically regular and orthographically transparent complex words. We will refer to this type of model as a cascaded dual route model: The second route comes into play only after completion of the first. A slightly less restrictive position is taken by Burani and Laudanna (1995), Chialant and Caramazza (1995), and Laudanna and Burani (1992), who suggest that words with a low surface frequency but with high-frequency constituents might be effectively processed via the parsing route.

In the AAM, surface frequency effects are attributed to different resting activation levels of access representations. With respect to cumulative stem frequency effects, various conflicting ideas appear to be entertained. Some papers suggest that representations at the central level prime all access representations with which they are linked (Burani & Caramazza, 1987; Laudanna & Burani 1985). However, Caramazza et al. (1988) suggest two possible sources for the cumulative frequency effect, namely, sensitivity to frequency for orthographic access representations in combination with sensitivity to frequency at the central level of decomposed lexical entries. In Laudanna, Badecker, and Caramazza (1992), the cumulative frequency effect is located at the central level only.

Taft (1979, see also Taft, 1994), argues for another kind of cascaded dual route model, one in which an obligatory parsing route precedes full-word retrieval in the central lexicon. In this model, the surface frequency of a simplex or complex word plays a role in the central lexicon, which lists the morphemes with which a given morpheme may combine. Stem frequency effects play a role at the access level: Words with the same cumulative stem frequency will become available to the central system in the same amount of time.

Schreuder and Baayen (1995) outline a race model with fully parallel routes. Their model is based on a spreading activation network with three representational layers: a layer of form-based modality-specific access representations (lexemes) and a layer of integration nodes (lemmas), that in turn is linked to a third layer of semantic and syntactic representations. The direct route maps a full-form access representation onto its associated lemma node, which in its turn activates its semantic and syntactic representations. In addition, the model contains a parsing route that operates in parallel with the direct route. Three stages in the parsing process are distinguished. In the first stage, access representations of affixes and stems become active along with full-form representations, leading to the activation of the lemma nodes of the stems and the affixes involved. We refer to this as the segmentation process. Following segmentation, two additional processes take place: licensing and composition. In the licensing process, the compatibility of subcategorization features of the activated constituents is checked. (The subcategorization properties of a morpheme specify the properties that another morpheme should have if the two are to be combined into a single word. For instance, the English suffix -ness is subcategorized for attaching to adjectives, and hence is excluded from attaching to, e.g., verbs: kindness, *thinkness.) In what we have called the composition process, the meaning of a complex word is computed from the meanings of its constituents. Finally, Schreuder and Baayen's (1995) model contains a mechanism of activation feedback from the syntactic and semantic layer via the lemma nodes to the access representations of constituents that are fully present in the signal. Activation feedback is hypothesized to allow cumulative frequency effects for transparent complex words only. Over time, activation feedback tunes the system such that an advantage for the parsing route results for transparent words, but a disadvantage for semantically opaque words. The most important difference between our model and the AAM and other cascaded dual route models is that the parsing route and the direct route are engaged in the recognition process from the very beginning of the recognition process.

PREDICTIONS OF THE PROCESS MODELS

Teasing apart predictions of the models discussed here requires considering in some detail the effects of surface and cumulative frequency for individual stems and the complex words derived from these stems. In what follows, we will concentrate on monomorphemic nouns and verbs and their corresponding plurals as found in Dutch. In Experiment 1, we varied the surface frequencies of singular nouns and their corresponding plurals, while keeping the cumulative stem frequency constant. Either the singulars were much more frequent than their plurals (singular-dominant pairs), or the plurals had a much higher surface frequency than their singulars (plural-dominant pairs). This variation in dominance leads to different predictions for all the models outlined above, as summarized in Figure 1.

PLACE FIGURE 1 APPROXIMATELY HERE

For ease of exposition, assume that the dominance relation is symmetrical, and that we have a surface frequency for singular-dominant pairs of 100 for the singulars and 10 for the plurals, and that for plural dominant pairs we have a surface frequency of 10 for the singulars and 100 for the plurals. What would the predictions of these models be? The top left hand panel shows the predictions of the full listing model. The horizontal axis indicates number (singular or plural), and the line type indicates what kind of frequency relation is involved, a solid line for singular-dominant pairs, a dotted line for plural-dominant pairs. Since the singular of a singular-dominant pair is more frequent than its plural, its reaction time should be shorter. Similarly, the plural of a plural-dominant pair should be reacted to more quickly than its lower frequency singular.

The upper right panel shows the predictions of the full parsing model. Variations in surface frequency are irrelevant here, as it is only the cumulative stem frequency that can influence reaction times. Therefore, the lines of the singular-dominant and plural-dominant pairs coincide: singular-dominant and plural-dominant pairs share the same cumulative frequency. However, since parsing may be assumed to require some additional processing time, plurals should reveal somewhat longer reaction times than their corresponding singulars.

In most of the versions of the AAM model outlined above, surface frequency effects should emerge for words with the same cumulative stem frequency. Thus, a crossover pattern similar to that predicted for the full-listing model is expected. However, since singular-dominant plurals may require parsing, and since parsing presumably is a time costly backup route, the response latencies to these plurals might be longer than those of plural-dominant singulars, even though the surface frequencies of these plurals and singulars are matched. Hence, an asymmetrical rather than a symmetrical crossover pattern is a possible prediction. A crossover interaction is also predicted by Taft's cascaded dual route model, which finds some experimental support from English (Taft 1979, but see below).

The model proposed in Schreuder and Baayen (1995) predicts that for transparent singular and plural pairs the summed frequency of the singular and plural form determines the speed with which a singular is recognized. Since regular plurals are fully compositional with respect to their singulars, each time a plural is recognized activation feedback will increase the resting activation level of the corresponding singular. (Note that a plural representation will not receive feedback from its singular, because the plural form is longer than that of the singular and hence not fully contained in the visual signal. For extensive discussion of the details of activation feedback, see Schreuder and Baayen, 1995, and Baayen, Lieber, and Schreuder, 1997). Hence, for singulars our prediction is identical to that of full parsing models, but different from the predictions for full listing models and the AAM.

Turning to plural forms, two possibilities arise. In our model, the balance of storage and computation is not fixed a priori, but depends on the properties of a given affix and the kind of base words to which it attaches. If the parsing process for a given combination of base word and affix is extremely time efficient, then there is no reason to suppose that full forms would be stored. For this situation, our model predicts the same kind of pattern as the full parsing model does. However, if the parsing process turns out to be very time costly, it may be more efficient to store frequent full forms than to process them on-line by rule. In this case, a pattern along the lines shown in the bottom right panel of Figure 1 is expected. If such a pattern is observed, we know that the other models are too restrictive. What we do not know, however, is the relative contribution of the parsing route compared to the direct route. In principle, this pattern is compatible with across-the-board storage, or with a combination of storage and computation. These possibilities can be teased apart by means of a mathematical formalization of our parallel dual route model.

THE PROCESSING OF PLURALS IN DUTCH

We will now test the predictions of these models using Dutch singular and plural nouns and verbs. In Dutch, three plural morphemes are in use for nouns: -en, -s, and -eren. The -eren plural is unproductive and restricted to some 15 nouns only. Both the -en and -s plurals are productive. In this paper, we study the (fully regular) -en plural, for two reasons. First, it is more frequently used than the -s plural. Second, the -en suffix also appears as the marker for plurality and the infinitive for verbs, so that by comparing nouns and verbs we will be able to unearth possible processing differences between these two word categories with a single suffix.

In Experiment 1, we contrasted noun singulars and plurals using the dominance design outlined above for two different stem frequency classes. For each stem frequency class, we predict the same pattern of results, although the response latencies for the high stem frequency condition will be faster than the corresponding response latencies for the low stem frequency condition.

Experiment 1

Method

Participants. One hundred participants, mostly undergraduates at Nijmegen University, were paid to participate in the experiment. All were native speakers of Dutch.

Materials. Ninety-three singular nouns and their corresponding plural forms were selected from the CELEX-database (Baayen, Piepenbrock & van Rijn, 1993) to construct four sets of singular-plural pairs, which were orthogonal with respect to stem frequency and dominance. The words within the first two sets and similarly the words within the last two sets were matched with respect to stem frequency. The stem frequency was computed as the sum of the frequencies of the singular, plural, and diminutive forms as they occur in a corpus of 42 million word tokens of written Dutch. For the first set, the 21 singular-dominant pairs had a high stem frequency (147 per million), while the frequency of the singular form exceeded that of the plural (singular: 126; plural: 13). For the second set, the 24 plural-dominant pairs had a high stem frequency (140), while the frequency of the plural form exceded that of the singular (singular: 53; plural: 86). For the third set, the 25 singular-dominant pairs had a low stem frequency (6), while the frequency of the singular form exceded that of the plural (singular: 5; plural: 1). For the fourth set, the 23 plural-dominant pairs had a low stem frequency (6), while the frequency of the plural form exceded that of the singular (singular: 2; plural: 4).

The four sets were matched with respect to length and bigram frequency of the singular and plural nouns. For all pairs of nouns the plural consisted of the orthographic form of the singular with the plural morpheme -en added (e.g., kelk – kelken, "chalice"). No other orthographic changes were involved.

In Dutch, as in English, verbs can be derived from nouns without overt affixation (e.g., the bike, to bike). All noun stems in the experiments reported here are unambiguously nouns only.

The stimulus list consisted of 186 test words, namely 93 singulars and the 93 corresponding plurals. For each of these words, a pseudoword was derived by changing one or more letters in the base word. This resulted in an additional 186 items. Furthermore, 123 filler words were added, as well as 123 pseudoword fillers derived from the filler words. Thus, the total number of stimuli was 618. The filler words consisted of adverbs, uninflected monomorphemic adjectives, and a number of singular and plural nouns all of which take the -s plural. All pseudowords consisted of orthographically and phonotactically legal letter strings.

The resulting stimulus material was divided over two experimental lists of 432 items each, in such a way that the singular form of each word pair was incorporated in one list and the plural form in the other. In this way we ensured that no participant saw the singular and plural form of the same stem. In each list, 93 different target nouns appeared (either in singular or in plural form, in approximately the same numbers), as well as 123 filler words (either in singular or plural form) together with a similar distribution of nonwords. Both lists were pseudorandomized, making sure that not more than three items of the same type (either word or pseudoword) occurred in sequence, and that no semantic associations of any kind existed between consecutive items. Finally, 40 practice items (20 words and 20 pseudowords, including singular and plural forms) were selected to precede the test material.

Procedure. Participants were tested in groups of three in individual noise-proof experimentation booths. They received a standard lexical decision instruction, specifying that they had to decide as quickly and as accurately as possible whether a presented letter string was a Dutch word or not. If it was a word, they had to push the right one of two response keys, otherwise the left one. For left-handed participants, the order of the response buttons was reversed.

Each trial consisted of the presentation of a fixation mark (asterisk) in the middle of the screen, followed after 600 ms by the stimulus centered at the same position. Stimuli were presented on Nec Multisync color monitors in white upper-case letters (font: triplex; size: 12 millimeters) on a dark background and remained on the screen until a participant pressed one of the two response buttons, or disappeared after a time period of 2 seconds if no response was given. A new trial was initiated 1200 ms after responding or time-out. When an error was made (a YES-response to a pseudoword or NO to a word), a dummy trial from an additional list of extra filler words was inserted to attenuate effects of error responses on the following test items.

Three pauses were included in the experiment: one between the practice and test set, and two during the experiment. After each break, participants continued the experiment when they were ready. The total duration of the experimental session was approximately half an hour.

Results

For each participant, the proportion of incorrect responses and missing data was calculated for all items in the experiment. The data from three participants, for which this proportion exceeded 10%, were excluded from further analysis. Using the remaining 97 participants, the distribution of reaction times for all items was obtained and four extreme outliers were removed from the data. The remaining observations were used to calculate item and participant mean reaction times and error scores. Table 1 shows the mean reaction times and error scores for the different test conditions.

PLACE TABLE 1 ABOUT HERE

By-participant and by-item analyses of variance showed that singular forms were reacted to faster than plurals: $F1(1, 96) = 160.77, p < .001$; $F2(1, 178) = 45.15, p < .001$. Also, higher stem frequency forms were responded to faster than lower stem frequency forms: $F1(1, 96) =$

$446.10, p < .001$; $F2(1,178) = 89.59, p < .001$; and singular-dominant forms were faster than plural-dominant forms: $F1(1,96) = 115.70, p < .001$; $F2(1,178) = 21.87, p < .001$. Significant interactions were found between Number and Stem Frequency [$F1(1,96) = 28.41, p < .001$; $F2(1,178) = 7.95, p < .01$], and between Number and Dominance [$F1(1,96) = 67.27, p < .001$; $F2(1,178) = 13.80, p < .001$].

Similar analyses were run for items with high and low-frequency stems separately. Analyses on the high-frequency stems resulted in significant main effects for Number [$F1(1,96) = 34.40, p < .001$; $F2(1,86) = 13.07, p < .001$] and Dominance [$F1(1,96) = 80.76, p < .001$; $F2(1,86) = 17.20, p < .001$], and in a significant interaction between these two factors [$F1(1,96) = 33.94, p < .001$; $F2(1,86) = 7.97, p < .01$].

For low-frequency stems only, the analyses also resulted in significant main effects for Number [$F1(1,96) = 161.76, p < .001$; $F2(1,92) = 33.42, p < .001$] and Dominance [$F1(1,96) = 39.00, p < .001$; $F2(1,92) = 8.75, p < .01$], and in a significant interaction between these factors [$F1(1,96) = 31.80, p < .001$; $F2(1,92) = 7.06, p < .01$].

We further tested whether for the singular condition the reaction times in the plural-dominant and singular-dominant conditions differed by means of t-tests on the item means. Significant differences were observed neither for the High Stem Frequency condition [$t(43) = 1.08, p = .29$], nor for the Low Stem Frequency condition [$t(46) = .29, p = .77$].

In this experiment, as in the experiments reported below, the pattern in the error data is virtually identical to that in the reaction time data. Analyses of variance by participants and by items on the error scores did not lead to any additional insights, and are therefore not reported.

Discussion

The pattern of results for each of the two stem frequency conditions is similar to the pattern predicted in the bottom right panel of Figure 1 (see also Figure 2, below). Within a stem frequency condition, singulars are processed about equally fast despite differences in their surface frequencies, while the plural forms show an effect of surface frequency. Plurals with a high surface frequency showed substantially shorter response latencies than those with a low surface frequency. (In fact, plural-dominant plurals are processed equally fast as their singulars in the High Stem Frequency condition, which shows that orthographic length as such does not underlie the effect of Number.) Finally, when we compare the differences between the singular-dominant plurals and their corresponding singulars in the High and Low Stem Frequency conditions, we find that the difference in the Low condition (96 ms) is larger than that in the High condition (54 ms).

Experiment 1 used a mixed design in which no participant saw both the singular and the plural form of a particular item. However, each participant encountered both singular and plural forms, as in normal language use. Might the balance between full-form retrieval and parsing be affected by list composition? We tested this possibility in an experiment in which the stimulus material of Experiment 1 was used in a between-participant design where participants either saw all target stems in the singular form, or all target stems in the plural form. The results of this experiment were identical to the results of Experiment 1, suggesting that the pattern observed in Experiment 1 is robust with respect to list composition.

The large effect of 96 ms obtained for the Low Stem Frequency is, at first sight, surprising, given the small difference in the mean absolute frequency of the singulars (5 per million) and the plurals (1 per million), compared to the much larger difference in the High Stem Frequency condition (126 per million for the singulars, 13 per million for the plurals). However, the differences in the mean of the logarithmically transformed frequencies (see Rubenstein & Pollack, 1963; Shapiro, 1969; and Scarborough, Cortese, & Scarborough, 1977) for evidence that the human processing is sensitive to log frequency rather than absolute frequency) are far less pronounced (1.79 for the Low Stem Frequency condition, and 2.27 for the High Stem Frequency condition, using natural logarithms). Also, our frequency counts are based on a corpus of 42 million words, so that differences in the low-frequency range that seem to be small after rescaling to the standard corpus size of one million are nevertheless statistically highly reliable. Most importantly, it is precisely for the singular-dominant plurals in the Low Stem Frequency condition that effects of morphological processing are most likely to emerge. As we shall see below, this is exactly what we find when we model our experimental results.

Modeling

Recall that the wedge-shaped pattern that we observe for noun singulars and their plurals in -en is compatible with two possible states of affairs. On the one hand, all plural forms might effectively be recognized on the basis of their full forms. On the other hand, it is also possible that both the parsing and the direct route are effectively involved. To tease apart these possibilities, we have constructed a mathematically formalized version of our model.

In this approach, two simple assumptions are crucial. First, we will assume that the resting activation level a_ω of a lexical representation ω is proportional to the logarithm of its frequency f_ω (cf. Rubenstein and Pollack, 1963; Shapiro, 1969; Scarborough, Cortese, and Scarborough, 1977). In what follows, we will assume that this proportionality holds in its simplest form,

$$a_\omega = \log f_\omega. \tag{1}$$

Second, we will assume that the time t_ω required for a lexical representation to reach threshold activation level is inversely proportional to its activation level, in the following way:

$$t_\omega = \frac{1}{1 + a_\omega}. \tag{2}$$

Note that (1) and (2) assign to each lexical representation a unique value in the interval $(0;1]$, the response latency of that representation in "model time."

To derive the predicted latencies for singular and plural nouns, consider the first task of the recognition system, the mapping of an incoming signal onto the correct access representations. We will assume that this initial mapping is achieved relatively quickly, and that the time required to complete the mapping of words of approximately the same length on their full-form access representations is a constant, ε_m. The segmentation of a plural such as handen, "hands," into its base word hand and its plural morpheme -en requires some additional processing time δ_s.

The segmentation process leads to the activation of one or more lexemes. The time for a particular lexeme λ to reach threshold activation level is given by (2). Recall that the activation feedback mechanism of the Schreuder and Baayen (1995) model entails that the cumulative stem frequency determines the resting activation level of the lexeme representation of the singular form. The surface frequency of the plural determines the resting activation level of the plural lexeme. Hence, for a singular, $t_{\lambda,\mathrm{sg}} = 1/(1+\log(f_{\mathrm{cum}}))$, for the full-form representation of a plural, $t_{\lambda,\mathrm{pl}} = 1/(1+\log(f_{\mathrm{plur}}))$.

In turn, the lexemes initiate the activation of their corresponding lemma representations. These lemma representations may or may not be frequency sensitive. In addition, plurals may or may not have their own lemma representations. Our computations have shown that an optimal fit to the obtained response latencies is obtained under the following assumptions. First, lemma representations are not frequency-sensitive (see Jescheniak, 1994; Jescheniak & Levelt, 1994, for a similar finding for lemma retrieval in speech production). Second, plurals have their own lemma representations (see Booij, 1993, who argues that pluralization involves concept formation; we will return to this issue in the general discussion). We will focus on this optimal model here. The reader is referred to Appendix A for details on the other variant model architectures.

In the optimal architecture, the response latency $\mathrm{RT}_{\mathrm{sg}}$ of a singular noun, expressed in "model time," can be defined as

$$\mathrm{RT}_{\mathrm{sg}} = \varepsilon + t_{\lambda,\mathrm{sg}} \tag{3}$$

where $\varepsilon = \varepsilon_m + \varepsilon_r$ denotes the (constant) initial mapping time ε_m and the (constant) time ε_r required for response execution. The response latency for a plural is determined by the fastest route:

$$\mathrm{RT}_{\mathrm{pl}} = \varepsilon + \min[t_{\lambda,\mathrm{pl}}, t_{\lambda,\mathrm{sg}} + \Delta_p]. \tag{4}$$

Here Δ_p is the total amount of time required by the parsing process, the sum of the time δ_s required for segmentation and the time δ_c required for licensing and composition.

In dual route race models statistical facilitation takes place when the distributions of the processing times of the two routes overlap (Raab, 1962). In order to take the effects of statistical facilitation into account, the parameters $\varepsilon, t_{\lambda,\mathrm{sg}}, t_{\lambda,\mathrm{pl}}$ and Δ_p should be interpreted as the means of random variables. In our computations we have assumed that these random variables are normally distributed. Rather than introducing additional parameters for the standard deviations of these random variables, we have made the simplifying assumption that for each random variable the standard deviation is one fourth of the mean. This parameter reflects the empirically obtained reaction time data of our experiments. Equations (3) and (4) have been applied to all items in our experiments. For each item, the mean response latency in model time was determined on the basis of $K = 500$ random samples for each variable. Let $\vec{\imath}$ be a vector containing a random sample of 500 normally distributed "model times" with mean i and standard deviation $i/4$. The expected response latency for singulars (see (3)) now becomes

$$\mathrm{E}[\mathrm{RT}_{\mathrm{sg}}] = \frac{1}{K}\sum_{k=1}^{K}(\vec{\varepsilon} + \vec{t}_{\lambda,\mathrm{sg}}). \tag{5}$$

The expected response latency for plurals (see (4)) is

$$\mathrm{E}[\mathrm{RT}_{\mathrm{pl}}] = \frac{1}{K}\sum_{k=1}^{K}(\vec{\varepsilon} + \min[\vec{t}_{\lambda,\mathrm{pl}}, \vec{t}_{\lambda,\mathrm{sg}} + \vec{\Delta}_p]). \tag{6}$$

These response latencies are expressed in "model time." To compare them with the observed reaction times, they have to be rescaled from "model time" to milliseconds. This can be accomplished by mapping the range $[\mathrm{E}[\mathrm{RT}]_{\mathrm{min}}, \mathrm{E}[\mathrm{RT}]_{\mathrm{max}}]$ of expected response latencies onto the range $[\mathrm{RT}_{\mathrm{min}}, \mathrm{RT}_{\mathrm{max}}]$ of observed reaction times.[1] (Due to the subtractions of expected reaction times while rescaling, the initial mapping parameter $\vec{\varepsilon}$ vanishes in the rescaled reaction times. The same applies to the constant time t_Λ that a lemma representation requires to become active. For ease of exposition, this constant is not entered into the equations (3)–(6).) The rescaled expected response latencies can be compared with the observed reaction times. Using maximum likelihood statistics, the optimal parse time Δ_p can be determined (see Wickens, 1982).

Should the optimal parse time be estimated on the basis of the 186 items in our experiments, or on the basis of the eight cell means? The former option is preferable, but unrealistic. It would imply that reaction times of individual words could be accurately predicted solely on the basis of word frequency counts. Although word frequency is a strong predictor of response latencies <u>in the mean</u>, it is well known that it is not the only relevant factor. Without a fully articulated theory of the way in which other factors such as neighborhood density, number of embedded words, orthographic consistency, syllable structure, word length, etc. influence the initial stages of perception that lead up to the activation of lexemes, predictions based on the frequencies of words only cannot be expected to fit the response latencies to the individual items in our experiments. However, since the analyses of variance of our experiment have shown that the variations in frequency underlying the factors Dominance and Stem Frequency result in highly significant reaction time differences for our eight cells, we will likewise ascertain the goodness-of-fit of the model on the basis of the cell means of our experiments, using a chi-square test as a measure of goodness-of-fit.[2]

PLACE FIGURE 2 APPROXIMATELY HERE

Figure 2 shows the predicted and observed mean reaction time for the eight cells of our experimental design. Visual inspection suggest that a good fit has been obtained. This is confirmed by a chi-square test of goodness-of-fit ($\chi^2_{(5)} = 5.06, p > 0.40$).[3] This fit was obtained for a surprisingly high parse time of 317 milliseconds, in other words, the processes of segmentation, licensing, and composition jointly (Δ_p) require a large proportion of the total response time. With such a high parse time, the overall majority of plural forms must have been recognized on the basis of their full-form representations. Nevertheless, the parsing route must also have been effectively involved, especially for a number of singular-dominant plurals with a low Stem Frequency. This contribution of the parsing route can be studied by comparing the predicted response latencies of the dual route model with the response latencies predicted by a model with only a direct route. In this way we can calculate, for each plural form, the advantage of having two routes in parallel. What we found is that for the plural-dominant plurals the parsing route never won the race. For the singular-dominant plurals with a high stem frequency, the (slow) parsing route won the race in 2% of the cases. For the singular-dominant plurals with a low stem frequency, the parser won the race in 14% of the cases. (These percentages were calculated by averaging over the mean number of cases for each of the plural forms in which the parsing route won the race in the 500 sample trials of the stochastic estimation of the predicted response time.) Three items, for which the surface frequency of the plural form is extremely low, and which by consequence have very long activation times, benefit most from the availability of the parsing route. Without the availability of the parsing route, it is impossible to obtain a good fit of the model to the data. Nevertheless, it remains a striking fact that the contribution of the parsing route in our computational simulations is small. We will return to this issue below.

Experiment 2

Experiment 1 showed that singular nouns of different frequencies are processed equally fast when matched for the summed frequencies of the singular and plural forms. For plural nouns, this experiment revealed an effect of surface frequency. Our mathematical model suggests that the parsing route is surprisingly time costly, and only wins the race for very low-frequency plurals. Experiment 2 addresses two issues. First, can the results of Experiment 1 be replicated with a new set of materials and a new design? Second, do the parameters of our model as determined for Experiment 1 provide a good fit to the new data of Experiment 2?

In Experiment 2, we matched sets of nouns for the surface frequency of the singular form while varying the dominance relation between the singular and the plural. We investigated three conditions. In the first condition, the plural was substantially more frequent than the singular. In the second condition, the singular and plural were of approximately equal frequency. In the third condition, the plural was less frequent than the singular. Because the summed frequencies of singular and plural differ for the three conditions, we predict that, contrary to Experiment 1, response latencies to the singulars will not be equal, despite equal surface frequencies. Furthermore, we predict an interaction between Number and Dominance. Since very low-frequency plural forms require the (slow) parsing route, we expect larger differences between the singular and the plural for the singular-dominant condition than for the plural-dominant condition. The second dominance condition was included in order to check whether our mathematical model would yield the right predictions for nouns without a strong dominance imbalance.

Method

Participants. Eighty-five participants, mostly undergraduates at Nijmegen University, were paid to participate in the experiment. All were native speakers of Dutch.

Materials. Seventy-two singular nouns and their corresponding plural forms in -en were selected from the CELEX-database to construct three sets of singular-plural pairs. In the first (plural-dominant) set, the mean frequency of the singular was 9 and that of the plural 40 per million. In the second (dominance-neutral) set, the mean frequency of the singular was 10 per million and that of the plural 11 per million. In the third (singular-dominant) set, the frequency of the singular was again 10 per million, but now the average frequency of the plural was 4 per million. The first set contained 23, the second set 24, and the third set 25 pairs of singulars and plurals. All three sets were matched for the frequency of the singular form. Seventy-two orthographically and phonotactically legal pseudoword stems were constructed to match the target words. We also added 90 filler words to the list of 144 targets, and 90 orthographically and phonotactically legal nonwords, resulting in 324 trials. The experiment was preceded by a practice session with 20 words and 20 pseudowords. No participant was exposed to both the singular and plural form of the same stem or pseudo-stem.

Procedure. The procedure was identical to that of Experiment 1.

Results

For each participant, we calculated the proportion of incorrect responses and missing data for all items in the experiment. The data from seven participants, for which this proportion exceeded 10%, were excluded from further analysis. The distribution of reaction times for all items was obtained and the six forms of three nouns for which the percentage of erroneous responses exceeded 30% for either the singular or for the plural form were removed from the data (one noun from the plural-dominant condition and two nouns from the dominance-neutral condition). Removal of these six items did not significantly affect the mean frequencies per dominance category. For the plural-dominant plurals, the mean frequency changed from 40 to 42 per million, and for the dominance-neutral condition, the mean plural frequency increased from 11 to 12 per million. All other mean frequencies remained unchanged. The remaining observations were used to calculate item and participant mean reaction times and error scores. Table 2 shows the mean reaction times and error scores for the different test conditions.

PLACE TABLE 2 ABOUT HERE

By-participant and by-item analyses of variance revealed that singular forms were reacted to faster than plurals: $F1(1,77) = 87.45, p < .001$; $F2(1,66) = 55.26, p < .001$, and that as the frequency of the plural increases, the stem frequency increases and hence response latencies decrease: $F1(2,154) = 72.00, p < .001$; $F2(2,66) = 8.3, p < .001$. The interaction of Number and Dominance shows that the extra processing time required for plurals compared to their singulars increased for decreasing frequency of the plural form: $F1(2,154) = 4.95, p < .01$;

$F2(2,66) = 3.32, p < .05$. The singular-dominant singulars required marginally significant longer response latencies than the dominant-neutral singulars ($t(45) = 1.83$, $p < .08$, two-tailed test) and significantly longer RTs than the plural-dominant singulars ($t(45) = 2.98$, $p < .01$, two-tailed test). The difference between the plural-dominant singulars and the dominance-neutral singulars failed to reach significance ($t(42) = 1.05$, $p > .20$, two-tailed test). Using Bonferroni's inequality, at the very least the difference between the plural-dominant singulars and the singular-dominant singulars is significant at the 5% level.

Discussion

Experiment 1 revealed that when singulars were matched for Stem Frequency, response times did not differ despite substantial differences in surface frequency. Experiment 2 shows the reverse pattern. When the surface frequency of the singular is kept constant, response latencies differ significantly as a function of the frequency of their plurals, as illustrated in the left panel of Figure 3. As expected, we also observed the interaction between Number and Dominance: The lowest-frequency plurals had relatively long response latencies.

PLACE FIGURE 3 APPROXIMATELY HERE

The right panel of Figure 3 plots the response latencies in model time predicted given the parameters determined on the basis of Experiment 1. (The parameters a and Δ_p were left unchanged, but the ratio of mean and standard deviation of the response latencies was estimated anew for the data of Experiment 2.) Our model captures the observed pattern in these new data quite well, both qualitatively and quantitatively ($X^2(2) = 4.38$, $p > .10$). Note that the observed pattern of results is incompatible with full-listing models, which predict that the singulars should be processed equally fast. It is also incompatible with full-parsing models, which predict the absence of an interaction between Number and Dominance, since they assume that the summed frequency of the singular and plural form is the only relevant independent variable. The pattern of results is not incompatible with the AAM or Taft's cascaded dual route model — to tease these models and the Schreuder and Baayen (1995) model apart the design of Experiment 1 is essential. We will return to the differentiation between these models below.

Summing up, Experiments 1 and 2 reveal that the summed frequencies of singular and plural determine response latencies to singulars, and that the surface frequency of the plural form, in combination with a parsing route, determines response latencies to plural forms. By means of mathematical modeling, we have investigated the speed of the parsing route. Simulation results for Experiments 1 and 2 suggest that the time required by the parsing route is surprisingly long, roughly 300 milliseconds. Given response latencies in the order of magnitude of 600 milliseconds, the question arises why the parsing route might be so slow.

In our view, the polyfunctionality of -en might be the crucial factor. The -en suffix not only appears as a plural ending on nouns, it also marks plurality on verbs, it marks the verbal infinitive, and in combination with the prefix ge- it marks the past participle for a number of verb classes (vang-en, ge-vang-en, "to catch," "caught"). A count in the CELEX lexical database (Baayen et al., 1993) reveals that of all word tokens ending in the suffix -en, 36% are noun plurals, and 64% verbal forms. If we assume that the verbal reading of -en is the default reading of this morpheme — which seems to be a reasonable assumption given that the verbal reading is encountered nearly twice as often as the nominal one — then a conflict will arise when the stem turns out to be a noun. (Interestingly, participants report that pseudowords ending in -en "feel" like verbs.) The time required to resolve this conflict of subcategorization requirements may underlie the very long parse times observed in Experiment 1 and 2.

If the long parse time is indeed due to a subcategorization conflict, then we do not expect to find such a long parse time for verb plurals compared to their singulars. This prediction is tested in Experiment 3.

Experiment 3

Method

Participants. Forty participants, mostly undergraduates at Nijmegen University, were paid to participate in the experiment. All were native speakers of Dutch, none had participated in any of the previous experiments.

Materials. Twenty-six noun singulars and their corresponding plurals, as well as 26 verb singulars and their corresponding plurals were selected from the CELEX lexical database.

All singular forms were matched for frequency (noun singulars: 27, verb singulars: 27), and the same holds for the frequencies of the plural forms (noun plurals: 7, verb plurals: 7). The verbs were all singular and plural past tense forms of irregular verbs. For instance, for the verb lopen, "to walk," the past tense forms liep (singular) and liepen (plural) were selected. The selection of these irregular past tense forms was motivated by the consideration that it is only for irregular past tense forms that the -en plural unambiguously marks the difference between singular and plural number. Because the past tense is indicated by the vocalic alternation in the stem, -en does not supply any information on tense, but, as for nouns, only on plurality. Since all verbal past tense stems in the experiment were irregular, they must have independent access representations in the mental lexicon. Except for the potential difference in the speed of the parsing route for nouns and verbs in -en, the processing of the verbs in our experiment may be expected to proceed along similar lines as that of the nouns. Note that all target words fall into the category of singular dominance. For verbs, it turns out that there are no real plural-dominant stems in the language. Hence, dominance could not be included as a factor in the experiment.

Each participant saw half of the target materials: 13 noun singulars, 13 noun plurals, 13 verb singulars, and 13 verb plurals. No participant was exposed to the singular and plural form of the same stem.

The 52 targets were embedded in a list of 130 filler materials (26 adverbs, 26 noun singulars and plurals, all from different stems, and 26 verb singulars and verb plurals, all from different stems). An additional list of 182 nonwords was created. Of these nonwords, 78 ended in the string en and 19 in the string s, as the word fillers contained noun plurals with the plural suffix -s. Finally, a practice list was created containing 20 words and 20 nonwords.

Procedure. Participants were tested in groups of four in individual noise-proof experimental booths. They received standard instructions for visual lexical decision. Stimuli were presented on Nec Multisync color monitors in white upper case 36 points Helvetica letters against a dark background. A trial consisted of a fixation character (asterisk) of 1000 ms, followed after 50 ms by the stimulus word, which was left on the screen for 1500 ms. Time-out was set at 2000 ms after stimulus onset. The next trial was initiated 650 ms after time-out.

Three short pauses were included in the experiment, one following the practice items, and two during the experiment itself. The duration of the complete experiment was approximately 40 minutes.

Results

For each participant, the proportion of incorrect responses and missing data was calculated for all items in the experiment. Two participants, for which this proportion exceeded 10%, were excluded from further analysis. Using the remaining 38 participants, the distribution of reaction times for all items was obtained, and two extreme outliers were removed. The remaining observations were used to calculate participant and item mean reaction times and error scores. Table 3 shows the mean reaction times and error scores for the different experimental conditions.

INSERT TABLE 3 APPROXIMATELY HERE

By-participant and by-item analyses of variance with the factors Number (singular versus plural) and Word Category (noun versus verb) showed significant main effects for Number [$F1(1, 37) = 30.58, p < .001$; $F2(1, 100) = 8.45, p < .01$] and Word Category [$F1(1, 37) = 25.57, p < .001$; $F2(1, 100) = 5.86, p < .02$], as well as a significant interaction between Number and Word Category [$F1(1, 37) = 20.67, p < .001$; $F2(1, 100) = 4.38, p < .05$].

Within the set of nouns, the plurals required significantly longer response latencies than the singulars both by participants [$F1(1, 37) = 46.93, p < .001$] and by items [$F2(1, 50) = 14.22, p < .001$]. For the verbs, no significant difference emerged (for both participants and items, $F < 1$).

Discussion

As expected for singular-dominant stems, the plural nouns in Experiment 3 required longer processing times than their singulars. For verbs, however, plurals were as fast as their singulars. This result provides clear support for our hypothesis that a subcategorization conflict arises for noun plurals but not for verb plurals. In our model, we account for this conflict in the following way.

We assume that there is a single access representation for -en that is linked to two lemma representations, one for verbal -en, and one for nominal -en. The former is subcategorized for verbs, the latter for nouns. The connection strengths between the access representation and its lemmas reflect the relative frequencies with which these two readings occur in the language. Once the access representation of -en has fired, the first lemma to become available is the verbal one, as it has the higher connection strength. When the lemma representation of a verbal base becomes available, the subcategorization requirements of the verbal -en lemma are checked. Since they are met (licensed), the constituents can be passed on for further syntactic processing. When the lemma representation of a noun base becomes available, however, a subcategorization conflict arises. In this case, checking of subcategorization compatibility reveals that the "default" verbal lemma is in conflict with the categorial value of the noun stem. The verbal -en lemma has to be deactivated, in favor of the nominal -en lemma. This requires additional processing time. Once the noun lemma of -en has become available, the subcategorization conflict disappears. The licensing requirements being met, the constituents can now be passed on for further semantic and syntactic processing.

In sum, for noun plurals, the licensing time consists of the processing time required for checking the subcategorization requirements combined with the additional processing time for undoing the default verbal setting. For verbs, on the other hand, the licensing costs consist of the default checking only. Our hypothesis, then, is that for noun plurals most of the overall parse time Δ_p, which cumulates segmentation time, licensing time, and composition time, consists of the extra cost associated with undoing the default verbal subcategorization setting.

Surprisingly, verb plurals are processed nearly equally fast as their singulars, in spite of their lower surface frequency. This suggests that lexical access does not take place via a surface representation of the full form, but through the access representation of the higher-frequency singular form and that of the very high-frequency plural suffix. (In fact, the frequency of the -en suffix is of the same order of magnitude as that of the highest frequency function words in the language. For instance, the total frequency of the -en suffix equals 3.7 million in a 42 million corpus, that of the most frequent definite article (de) equals 2.3 million. This suggests that in the visual modality the segmentation of -en will proceed very rapidly and that the suffix will be identified before the stem.) Following the very rapid segmentation of the input into its constituents, an equally rapid licensing of the default combination of -en with a verbal base takes place. Apparently, no composition is involved for verb plurals. This ties in with the syntactic function of plural marking on verbs. Booij (1993) makes a distinction between inherent inflection, such as pluralization for nouns, and contextual inflection, such as person and number inflection on verbs. In the case of inherent inflection, the noun plural adds to the meaning of the base noun. For instance, shoes typically occur in pairs. For words such as shoes, the plural suffix indicates a "dual" for symmetrically paired objects rather than plurality ("more than 1") only. In the case of contextual inflection, the meaning of the verb is left unchanged. No composition of meaning is involved. The function of the plural ending for verbs is to supply information to the syntax concerning the external arguments of the verb. Within our process model, the absence of composition implies that the meaning of the base and the meaning of the plural suffix are immediately passed on to postlexical processing stages. Thus verb plurals are quickly segmented into plural affix and verbal base, without further composition. The only task of the mental lexicon is to pass on the lemma information of the stem and the plural agreement feature to postlexical syntactic processes.

The speed with which verb plurals are apparently parsed raises another question, namely, why these verb plurals show reaction times that are of the same order of magnitude as the reaction times to the noun plurals in the experiment. If verb plurals are so easy to parse, one would expect that they elicit shorter response latencies than noun plurals. Note, however, that the singulars of verbs show up with longer response latencies than the noun singulars, even though the sets of singulars are matched for frequency of occurrence. This suggests that verbs are more difficult to process than nouns in the list-like condition of our experiment. Nouns, in both singular and plural form, appear in isolation in natural language. Only non-finite verb forms such as infinitives and participles can do so. In addition, all verbs in our experiment appeared in the past tense rather than in the unmarked present tense. Taken jointly, these considerations suggest that the verb singulars and plurals in our experiment are more difficult to process than their frequency of use would suggest. If so, the fact that the verb plurals show up with reaction times similar to those of the noun plurals is a coincidence only. We have carried out control experiments in which participants were exposed to only nouns or to only verbs (using the same materials listed in Appendix B). These experiments showed that list composition is responsible for the nearly identical reaction times for plural nouns and verbs in Experiment 3. In these control experiments, the reaction times of the noun plurals were now significantly faster than those of the verb plurals. Both were faster than in Experiment 3. Crucially, the same pattern of results was replicated: a significant effect of Number for nouns, and no significant effect of Number for verbs. We conclude that the similar response latencies in Experiment 3 for noun and verb plurals are a coincidence, and that in general verbs without context are more difficult to process than nouns of comparable frequency.

These findings may shed light on why Taft (1979) obtained different results for uninflected words matched for the summed frequency of their inflectional variants, and contrasting with respect to surface frequency. Whereas our Experiments suggest that it is the summed frequency that crucially determines response latencies to uninflected nouns, Taft (Experiment 3) observed that the surface frequency of the uninflected form also played a role. Inspection of his critical items reveals, however, that the high surface frequency condition for uninflected words contained two verbs or nouns with a conversion alternant (milk, rear), while in the corresponding low-frequency condition, 13 items were verbs or nouns with a conversion verb (hunt, marry, clothe, nod, chew, mutter, excite, nail, shout, clap, drip, oblige, and poke). With 20 items in each condition, our findings suggest that the effect of surface frequency observed by Taft is confounded with an effect of word category. At least in Dutch, verbs are more difficult to process than nouns. Although other explanations in terms of general differences between Dutch and English might be invoked to explain why Taft obtained his pattern of results, it seems to us that the overrepresentation of verbs in his low-frequency condition is the most likely source for the divergence between his and our experiments.

Finally, note that we have assumed that nominal and verbal -en share the same access representation. This assumption implies that we also have to assume that the segmentation process for nominal -en is as fast as that for verbal -en. This implies in its turn that most of the parse time for nouns, which includes segmentation time, licensing time, and composition time, is taken up by the latter two, of which the licensing time may be the most important one.

GENERAL DISCUSSION

We have addressed the issue of storage and computation in morphological processing. A number of researchers have approached this issue by comparing regular with irregular complex words (e.g., Rumelhart & McClelland, 1986; Pinker & Prince, 1991; MacWhinney & Leinbach, 1991). There is little disagreement with respect to irregular complex words — all the experimental evidence suggests that these words are stored as wholes at some level of representation. Opinions differ, however, with respect to regular complex words. Here storage has been argued to be pervasive (Butterworth, 1983), to be likely, especially for high-frequency words (Stemberger & MacWhinney, 1986; Bybee, 1985, 1995; Baayen, 1992; Frauenfelder & Schreuder, 1992), or to be completely absent (Pinker, 1991; Pinker & Prince, 1991).

We have approached this issue by examining how regular Dutch plurals in -en and their singulars are processed. By manipulating the frequency relation between the singular and the plural form, the following pattern of results emerged. For singular nouns, response latencies are determined by the summed frequency of the singular and plural form. For plural nouns, response latencies depend predominantly on their surface frequency, but in the lowest-frequency ranges some forms are probably processed by means of the parsing route. The pattern of results obtained is incompatible with full-parsing models, full-listing models, and cascaded dual route models (AAM, Taft 1979, 1994). A mathematical model formalizing the assumptions of the verbal model outlined in Schreuder and Baayen (1995) yielded a reliable fit to the experimental data of Experiment 1 for a surprisingly high parse time of some 300 milliseconds. The same parse time also yielded a good fit to the data of Experiment 2. Our hypothesis is that the polyfunctionality of the -en suffix underlies this long parse time. In Dutch, -en occurs almost twice as often as a verbal ending than as a nominal ending. Therefore, a noun stem and the default verbal reading of -en create a subcategorization conflict. The resolution of this conflict is, apparently, time costly. This hypothesis is supported by Experiment 3, in which noun plurals and singulars were compared with verb singulars and plurals. In contrast to the nouns, the verbs did not reveal a significant difference in processing time for singulars and their corresponding plurals, even though these plurals were all lower in frequency than their singulars. This suggests that there is no real subcategorization conflict for verbs, and that verb plurals are recognized on the basis of their stems and the suffix -en. In contrast to nouns, no access representations for verb plurals seem to be involved. It is only for noun plurals that full-form access representations are available, the functionality of which is to speed up lexical processing which would otherwise be slowed down substantially by the subcategorization conflict.

The long parse time of about 300 milliseconds observed for Dutch noun plurals contrasts with a much shorter parse time of 90 milliseconds calculated for recent experimental results for Italian noun plurals and singulars. Using a lexical decision task and the same experimental design as in Experiment 1, Baayen, Burani, and Schreuder (1996) found that the response latencies to the singular-dominant and plural-dominant singulars were indistinguishable, as in Dutch. This supports our claim that response latencies to singulars are crucially determined by the summed frequency of the singular and the plural form. As expected, singular-dominant plurals in Italian required longer response latencies than their corresponding singulars, but, surprisingly, the plural-dominant plurals were now responded to significantly faster than their corresponding singulars. If the stem frequency is indeed the crucial determinant for the singular form, then the surface frequency of the plural cannot by itself explain why plural-dominant plurals are processed faster than their singulars. After all, the stem frequency is by definition higher than the surface frequency of the plural. These results, which are incompatible with full-listing models, full-parsing models, and cascaded dual route models, receive a natural interpretation within our framework. Stochastic mathematical modeling yielded a good fit to the empirical data for a parse time of some 90 milliseconds. The difference between Dutch and Italian parse times supports our hypothesis that the long parse time for Dutch is to be attributed to the subcategorization ambiguity of the -en suffix. With the much smaller parse time for Italian, we now observe a clear effect of statistical facilitation. Due to the availability of two routes with overlapping distributions of processing times, high-frequency plurals are on average processed significantly faster than their corresponding singulars (see Baayen et al., 1996 for detailed discussion).

This cross-linguistic comparison of the processing of noun plurals in Dutch and Italian illustrates a more general point. In the past, the driving research question in psycholinguistic research on morphological processing has been whether or not morphology has some role to play. Models have been developed that, at a global level and generalizing over different kinds of affixes and word formation processes, sketch how morphological processing takes place for any complex word. In our view, approaches in which the specific properties of affixes and the word formation patterns in a given language are not taken into account are severely limited. Depending on such parameters as modality, the distributional properties of affixes (Laudanna and Burani, 1995), productivity, frequency of use and frequency dominance, semantic and phonological transparency, and subcategorizational ambiguity, evidence for morphological processing may or may not appear in the data. Not surprisingly, the existing literature on morphological processing is plagued by contradictory results (see Henderson, 1985; McQueen & Cutler, in press; Sandra, 1994). Marslen-Wilson et al. (1994) argue that this is due to a general neglect to take into account the role of semantic transparency. While we fully agree that semantic transparency is a crucial determinant of morphological processing, our results show that there are many other important parameters that should not be ignored either.

We have interpreted our results in the framework of the qualitative model of morphological processing outlined in Schreuder and Baayen (1995). Quantitative simulation studies have allowed us to study its predictions in some detail. One assumption of this model, frequency sensitivity at the layer of the integration nodes (lemmas), was not supported by our simulation studies. Apparently, frequency effects are limited to the peripheral lexeme layer, the layer of visual access representations. More central representational layers, typically the layers where operations of a more symbolic nature are carried out on syntactic and semantic representations, appear to be frequency-free. This finding is in line with the claim advanced by Jescheniak and Levelt (1994) that in speech production the lemmas are not frequency sensitive, but only the word form representations. From this perspective, models which make use of two frequency-sensitive representational layers appear less parsimonious. This is an interesting domain of inquiry that clearly requires further research.

The simulation studies furthermore suggested as optimal a model in which plural forms have their own lemma representations. This finding ties in nicely with the linguistic observation that nominal pluralization involves a kind of concept formation (Booij, 1993). For some nouns, the changes in semantic and syntactic information may be minimal, amounting to a change in number only. But for many nouns, the semantics of the plural are subtly different from those for the singular. Many plural-dominant nouns are unmarked with respect to their singulars, a phenomenon that is known as markedness shift (Tiersma, 1982). Plural dominant nouns such as eyes and tulips occur in natural pairs or groups. For these nouns, the singular is marked, in that it singles out one instance from a natural pair or group. In contrast, singular-dominant nouns such as nose and mouth can be pluralized, but are semantically marked in the plural: Noses and mouths do not occur in natural groups. Interestingly, there are languages in which plural-dominant plurals are not only semantically, but also formally unmarked. In these languages, plural-dominant plurals are monomorphemic words, the singulars of which are obtained by suffixation of a so-called singulative marker. For instance, in Bari, an Eastern Nilotic language, the plural of the singular-dominant noun kupö, "large basket," is kupö-jin, whereas the singular of the plural-dominant noun kuru, "worms," is kuru-töt (Dimmendaal, 1993).

Finally, our experimental evidence for storage of fully regular complex words sheds new light on the controversy between connectionist and symbolic approaches to language processing. While in most connectionist approaches morphological rules emerge as an epiphenomenon of distributed joint storage of monomorphemic and morphologically complex words, symbolic models make a fundamental distinction between representations, which may be frequency sensitive, and rules which act on these representations. Pinker (1991) has taken the symbolic position to its extreme by claiming that no regular complex words are stored as wholes in the mental lexicon. Applying what is essentially Bloomfield's (1933) view of the linguistic lexicon to psycholinguistics, Pinker argues for language production that storage of complex words in the mental lexicon is restricted to forms that have at least one unpredictable, idiosyncratic property. Our experimental results show that, at least with respect to language comprehension, the Bloomfieldian view is too restricted. To account for the observed effects of storage and computation, we have developed a theoretical framework that exploits the flexibility of parallel dual route models to explain the intricate pattern of reaction times obtained for Dutch nominal and verbal pluralization. It is difficult to see how these patterns could be understood using monolithic neural nets (see also Gasser, 1993, 1994), modeling in one pass what in our view is a complex multilayered system. At the same time, our data clearly show that models that categorically deny storage of regular complex words cannot be maintained either. As in other domains of cognitive processing, storing complex representations may significantly speed up real-time performance as a function of the complexity of the computations involved.

,

Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In G. E. Booij & J. van Marle (Eds.), Yearbook of morphology 1991 (pp. 109–149). Dordrecht: Kluwer Academic Publishers.

Baayen, R. H., Burani, C. & Schreuder, R. (1996). Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. To appear in G. E. Booij & J. v. Marle (Eds.), Yearbook of morphology 1996. Dordrecht: Kluwer Academic Publishers.

Baayen, R. H., Lieber, R., and Schreuder, R. (1997). The morphological complexity of simplex nouns (submitted).

Baayen, R. H., Piepenbrock, R. & van Rijn, H. (1993). The CELEX lexical database (CD-ROM), Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bloomfield, L. (1933). Language. London: Allen and Unwin.

Booij, G. E. (1993). Against split morphology. In G. E. Booij & J. v. Marle (Eds.), Yearbook of morphology 1993 (pp. 27–49). Dordrecht: Kluwer Academic Publishers.

Burani, C. & Caramazza, A. (1987). Representation and processing of derived words. Language and Cognitive Processes, 2, 217–227.

Burani, C. & Laudanna, A. (1992). Units of representation of derived words in the lexicon. In R. Frost & L. Katz (Eds.), Orthography, phonology, morphology, and meaning (pp. 361–376). Amsterdam: Elsevier.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), Language production (Vol. II): Development, writing and other language processes (pp. 257–294). London: Academic Press.

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, N.J.: Lawrence Erlbaum Associates.

Caramazza, A., Laudanna, A. & Romani, C. (1988). Lexical access and inflectional morphology. Cognition, 28, 297–332.

Chialant, D. & Caramazza, A. (1995). Where is morphology and how is it processed? The case of written word recognition. In L. B. Feldman (Ed.), Morphological aspects of language processing (pp. 55–78). Hillsdale, N.J.: Lawrence Erlbaum Associates.

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.

Dimmendaal, G. J. (1993). On language. [Review of the book: On language. Selected writings of Joseph H. Greenberg.] Language, 4, 796–807.

Frauenfelder, U. H. & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. E. Booij & J. v. Marle (Eds.), Yearbook of morphology 1991 (pp. 165–183). Dordrecht: Kluwer Academic Publishers.

Gasser, M. (1993). Learning words in time: towards a modular connectionist account of the acquisition of receptive morphology. Indiana, Bloomington, Indiana University: Computer Science and Linguistics Departments.

Gasser, M. (1994). Modularity in a connectionist model of morphology acquisition. Proceedings of the International Conference on Computational Linguistics, 15, 214–220.

Goldinger, S. D., Luce, P. A., Pisoni, D. B. & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 1211–1238.

Henderson, L. (1985). Towards a psychology of morphemes. In A. W. Ellis (Ed.), Progress in the psychology of language, Vol. 1 (pp. 15–72). London: Lawrence Erlbaum.

Jescheniak, J. D. (1994). Word frequency effects in speech production. Unpublished Doctoral Dissertation, Catholic University of Nijmegen, Nijmegen.

Jescheniak, J. D. & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. Journal of Experimental Psychology: Learning, Memory and Cognition, 20(4), 824–843.

Laudanna, A., Badecker, W., & Caramazza, A. (1992). Processing inflectional and derivational morphology. Journal of Memory and Language, 31, 333–348.

Laudanna, A. & Burani, C. (1985). Address mechanisms to decomposed lexical entries. Linguistics, 23, 775–792.

Laudanna, A. & Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In L. B. Feldman (Ed.), Morphological aspects of language processing (pp. 345–364). Hillsdale, N.J.: Lawrence Erlbaum.

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

Marslen-Wilson, W., Tyler, L. K., Waksler, R. & Older, L. (1994). Morphology and meaning in the English mental lexicon. Psychological Review, 101, 3–33.

McClelland, J. L. & Rumelhart, D. E. (Eds.) (1986). Parallel distributed processing. Explorations in the microstructure of cognition. Vol. 2: psychological and biological models. Cambridge, Mass.: The MIT Press, Cambridge, Mass.

McQueen, J. M. & Cutler, A. (in press). Morphology in word recognition. In A. M. Zwicky & A. Spencer (Eds.), The Handbook of morphology. Oxford: Basil Blackwell.

Pinker, S. (1991). Rules of language. Science, 153, 530–535.

Pinker, S. & Prince, A. (1991). Regular and irregular morphology and the psychological status of rules of grammar. In Proceedings of the 1991 meeting of the Berkeley Linguistics Society.

Raab, D. H. (1962). Statistical facilitation of simple reaction time. Transactions of the New York Academy of Sciences, 24, 574–590.

Rickard, T. C., Healy, A. F. & Bourne Jr., E. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 1139–1153.

Rosenbaum, D. A., Vaughan, J., Barnes, H. J. & Jorgensen, M. J. (1992). Time course of movement planning: Selection of handgrips for object manipulation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 1058–1073.

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

Rumelhart, D. E. & McClelland, J. L. (1986). On learning the past tenses of English verbs. In McClelland & Rumelhart (1986), pp. 216–271.

Sandra, D. (1994). The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. Language and Cognitive Processes, 9, 227–269.

Scarborough, D. L., Cortese, C. & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 3, 1–17.

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

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: Lawrence Erlbaum Associates.

Shapiro, B. J. (1969). The subjective estimation of word frequency. Journal of Verbal Learning and Verbal Behavior, 8, 248–251.

Stemberger, J. P. & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. Memory & Cognition, 14, 17–26.

Stemberger, J. P. & MacWhinney, B. (1988). Are lexical forms stored in the lexicon? In M. Hammond & M. Noonan (Eds.), Theoretical Morphology: Approaches in Modern Linguistics (pp. 101–116). London: Academic Press.

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

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. Language and Cognitive Processes, 9, 271–294.

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

Tiersma, P. M. (1982). Local and general markedness. Language, 58, 832–849.

Wickens, T. D. (1982). Models for behavior: stochastic processes in psychology. San Fransisco: W.H. Freeman and Company.

Author Note

We are indebted to Cristina Burani, Anne Cutler, Stef van Halen, Pim Levelt, James McQueen, Richard Sproat, and three anonymous reviewers for invaluable discussion and criticism. We are especially in debt to Richard Sproat, who wrote the operational parser for Dutch plurals that inspired our simulation studies. Furthermore, we wish to thank Leonore Biegstraaten and Maarten van Casteren for their aid in preparing and running the experiments.

Correspondence concerning this article should be addressed to R. Harald Baayen, who is at the Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands. Electronic mail may be sent to baayen@mpi.nl.

[1]The rescaled expected reaction time E[RT′] is given by

$$\mathrm{E[RT]}' = \mathrm{RT}_{\mathrm{min}} + \frac{\mathrm{E[RT]} - \mathrm{E[RT]}_{\mathrm{min}}}{\mathrm{E[RT]}_{\mathrm{max}} - \mathrm{E[RT]}_{\mathrm{min}}} \cdot (\mathrm{RT}_{\mathrm{max}} - \mathrm{RT}_{\mathrm{min}})$$

[2]We have used the test statistic

$$X^2 = \sum_{i=1}^{N} \frac{(\overline{\mathrm{RT}_i} - \overline{\mathrm{E[RT]}_i'})^2}{\frac{\mathrm{VAR[RT}_i]}{N_i} + \frac{\mathrm{VAR[E[RT]}_i']}{N_i}},$$

where N_i denotes the number of observations in cell i. With $N = 8$ and two free parameters (the parse times Δ_p and the estimated standard deviation ratio at one quarter of the mean), X^2 is $\chi^2_{(6)}$ distributed.

[3]This result is based on a model that makes use of one extra parameter. The reason for introducing this additional parameter is that, although a reasonable fit ($\chi^2_{(6)} = 7.75, p > 0.25$) can be obtained for the model defined by (2), (5) and (6), it appeared that this model overestimates reaction times for the lowest-frequency plurals. This overestimation is due to the shape of the time function (2). For decreasing word frequency in the lower frequency ranges, (2) predicts rapidly increasing activation times. By adding a new parameter a to the time function (2), the rate at which the predicted activation times increase with decreasing word frequency can be influenced so that very low frequency words will not require extremely long activation times. The revised activation function underlying Figure 2 is

$$t_\omega = \frac{1}{1 + (\log f_\omega)^a} - b,$$

where $b = 1/(1+(\log f_{max})^a) - 1/(1+\log f_{max})$. Note that b is not an independent parameter of the model, as it is completely determined by the parameter a and the maximal frequency f_{max}. By means of the constant b we ensure that both the new activation function and (2) map the frequency domain $[1, f_{max}]$ onto the same interval of activation times, $[1, 1/(1 + \log f_{max})]$. The fit of Figure 2 was obtained for $a = 0.4$.

Appendix A

Alternative Model Architectures

In the main text we have presented a model in which lemma representations are not frequency-sensitive, and in which plurals have their own lemma representations (see equations (5) and (6)). Here we briefly present the three other models that result from varying the assumptions concerning the frequency-sensitivity of lemma representations, and the assumptions concerning the presence or absence of independent lemma representations for plurals.

1. Frequency-sensitive lemmas, no independent plural lemmas

The expected RT for singulars is jointly determined by the activation times of the access representation and the lemma representation.

$$\mathrm{E}[\mathrm{RT}_{\mathrm{sg}}] = \frac{1}{K}\sum_{k=1}^{K}((\vec{t}_{\lambda,\mathrm{sg}} + \vec{t}_{\Lambda}) + \vec{\varepsilon}), \tag{7}$$

where t_{Λ} denotes the activation time of the lemma representation

$$t_{\Lambda} = \frac{1}{1 + \log(f_{\mathrm{cum}})}. \tag{8}$$

For plurals, the direct route bypasses the segmentation process. Both the direct route and the parsing route must rely on semantic composition at the level of lemma representations to obtain the meaning of the plural. Hence, the expected RT for plurals is

$$\mathrm{E}[\mathrm{RT}_{\mathrm{pl}}] = \frac{1}{K}\sum_{k=1}^{K}(\min[(\vec{t}_{\lambda,\mathrm{pl}} + \vec{t}_{\Lambda} + \vec{\delta}_c), (\vec{t}_{\lambda,\mathrm{sg}} + \vec{t}_{\Lambda} + \vec{\delta}_c + \vec{\delta}_s)] + \vec{\varepsilon}). \tag{9}$$

2. Frequency-sensitive lemmas, independent plural lemmas

For singulars, the expected RT remains unchanged:

$$\mathrm{E}[\mathrm{RT}_{\mathrm{sg}}] = \frac{1}{K}\sum_{k=1}^{K}((\vec{t}_{\lambda,\mathrm{sg}} + \vec{t}_{\Lambda}) + \vec{\varepsilon}). \tag{10}$$

For plurals, we assume that the activation level of the plural lemma is identical to that of its corresponding singular lemma. This assumption is a consequence of the co-activation of semantically related lemmas in the Schreuder and Baayen (1995) model. As the semantics of singulars and plurals are very similar, they will activate each other's lemmas to almost the same extent. As a consequence, the activation levels of their lemmas will be roughly the same. This assumption leads to the following expected RT for plurals:

$$\mathrm{E}[\mathrm{RT}_{\mathrm{pl}}] = \frac{1}{K}\sum_{k=1}^{K}(\min[(\vec{t}_{\lambda,\mathrm{pl}} + \vec{t}_{\Lambda}), (\vec{t}_{\lambda,\mathrm{sg}} + \vec{t}_{\Lambda} + \vec{\Delta}_p)] + \vec{\varepsilon}), \qquad (11)$$

Note that the direct route no longer involves composition.

3. Frequency-free lemmas, no independent plural lemmas

If lemma representations have a constant activation time irrespective of the frequency with which they are accessed, they add a constant amount of "model time." Together with $\vec{\varepsilon}$, this constant disappears after rescaling. For ease of exposition, we do not mention this constant activation time in the model definition. The expected RT for singulars is simply

$$\mathrm{E}[\mathrm{RT}_{\mathrm{sg}}] = \frac{1}{K}\sum_{k=1}^{K}(\vec{t}_{\lambda,\mathrm{sg}} + \vec{\varepsilon}), \qquad (12)$$

and that for plurals is

$$\mathrm{E}[\mathrm{RT}_{\mathrm{pl}}] = \frac{1}{K}\sum_{k=1}^{K}(\min[(\vec{t}_{\lambda,\mathrm{pl}} + \vec{\delta}_c), (\vec{t}_{\lambda,\mathrm{sg}} + \vec{\delta}_s + \vec{\delta}_c)] + \vec{\varepsilon}). \qquad (13)$$

We have tested these model variants against the experimental data. The model variants 1 and 3, in which plurals do not have their own lemma representations, performed less well than their counterparts. Reasonable fits for these models could only be obtained by making the composition time δ_c close to zero. For $\delta_c \to 0$, the χ^2 measure for these models can be made arbitrarily similar to the χ^2 values of the models with lemma representations for plurals. Note, however, that if the composition time is virtually zero, these model variants become indistinguishable from models in which no composition is required because a stored lemma representation of the plural is available.

The remaining model, variant 2, assumes that lemma representations are frequency sensitive. The optimal value of the parse time Δ_p for this model equals 306 ms for $\chi^2_{(6)} = 10.62, p = 0.101$. Recall that the model presented in the text assumes that no frequency information accumulates at the lemma level. In its simplest form (see footnote 6), a parse time of 315 ms is obtained ($\chi^2_{(6)} = 7.75, p = 0.257$).

These computational analysis of Experiment 1, together with the analysis presented in the main text, suggest that the original model proposed by Schreuder and Baayen (1995) has to be revised. First, our results suggest that a variant of this model in which lemma representations are not frequency-sensitive is to be preferred. Second, Schreuder and Baayen argued on linguistic grounds that the extremely high degree of semantic transparency characteristic of plural nouns should obviate the need of assigning separate lemma representations to plurals. However, we have seen that these two assumptions lead to non-optimal fits. Apparently, frequency information does not cumulate at the lemma level. This finding is in line with the results reported by Jescheniak and Levelt (1994) for production. In addition, plural forms appear to require their own lemma representations.

Appendix B

Stimuli used in Experiments 1–3. Each stem is followed by the response latencies to the singular and plural form.

Stems used in Experiments 1.

<u>Low Stem Frequency, Singular-Dominant Stems:</u> baai (644, 803), boeg (613, 748), bruid (538, 626), fuik (656, 795), galg (564, 672), havik (601, 662), kelk (677, 734), klerk (654, 830), korps (655, 682), loep (578, 764), lont (668, 727), muil (583, 677), muts (563, 598), part (637, 662), pont (628, 775), prei (590, 744), pruik (575, 613), romp (589, 733), sprei (640, 924), stoet (657, 936), telg (623, 810), valk (581, 693), vork (585, 614), zalm (528, 608), zeug (686, 698).

<u>Low Stem Frequency, Plural-Dominant Stems:</u> berk (692, 707), biet (581, 679), dwerg (531, 637), erwt (561, 615), flank (659, 742), friet (606, 768), geit (564, 603), gift (605, 676), halm (687, 718), kers (609, 544), klomp (588, 576), kluit (656, 700), kous (604, 559), kuit (645, 630), lakei (646, 712), meeuw (537, 626), nier (587, 607), rups (613, 708), twijg (643, 681), welp (682, 681), wesp (564, 589), wilg (620, 655), worm (514, 540).

<u>High Stem Frequency, Singular-Dominant Stems:</u> ambt (663, 715), buik(544, 621), drank (519, 562), eeuw (600, 616), feit (614, 583), gang (556, 630), helft (583, 725), hemd (552, 594), hoofd (496, 536), huid (498, 582), kast (567, 568), kern (635, 714), nest (545, 528), park (540, 623), plein (584, 709), pond (559, 637), soep (541, 597), stijl (545, 621), tijd (536, 587), tong (534, 563), voogd (603, 682).

<u>High Stem Frequency, Plural-Dominant Stems:</u> darm (563, 605), dier (503, 556), duin (603, 589), eend (516, 581), fout (568, 535), gast (516, 570), heup (565, 548), kaars (562, 547), klant (531, 550), long (551, 537), maand (513, 538), mens (525, 508), mouw (580, 614), norm (553, 571), plank (580, 538), rots (587, 577), term (590, 622), voet (554, 524), wand (515,

546), wang (555, 537), wolk (545, 533), woord (521, 525), zenuw (538, 562), zuil (591, 574).

Stems used in Experiment 2. Words followed by (*) were not included in the analysis due to high error rates.

Singular-dominant Stems: beer(528, 551), bink(611, 647), erts(695, 668), fuik(598, 673), geit(561, 542), kruin(613, 643), laan(540, 623), lont(612, 661), mand(521, 576), part(566, 630), pion(582, 741), pond(588, 673), prooi(535, 698), rand(573, 608), sein(628, 646), sloep(632, 715), snoek(539, 608), soort(504, 592), stoep(549, 691), stolp(681, 744), teil(643, 768), terp(634, 683), trog(687, 788), veld(507, 536), vlies(536, 646).
Dominance-neutral Stems: baat (673, 681), buis (617, 571), darm (514, 543), dier (483, 543), dwerg (534, 504), eend (530, 574), geul (645, 714), klont (562, 608), lint (546, 646), mouw (526, 587), naad (560, 615), piek (588, 641), plank (551, 571), prent (561, 589), sjerp (*), spar (577, 801), steen (526, 549), troep (509, 576), vink (519, 558), vlek (521, 560), wesp (537, 564), wrat (670, 544), zerk (*), zwaan (511, 522).
Plural-dominant Stems: been (532, 531), bloem (496, 512), buur (635, 591), ding (531, 572), duin (538, 557), flard (*), hiel (577, 630), kers (561, 559), klauw (551, 607), kruid (506, 517), krul (584, 661), laars (512, 495), long (560, 549), mier (522, 554), norm (520, 572), spier (521, 610), ster (598, 692), tand (495, 504), teen (492, 515), traan (485, 552), tros (526, 606), wiek (652, 718), zenuw (528, 535).

Stems used in Experiment 3.

Nouns: arts (504, 552), boot (595, 649), burcht (581, 794), cel (566, 518), dorp (548, 612), dwerg (556, 558), fornuis (560, 592), gang (499, 624), geest (481, 539), gitaar (488, 564), held (684, 722), hemd (534, 618), hert (500, 584), hut (543, 710), jurk (478, 514), kasteel (504, 515), kraan (553, 615), kruik (621, 623), lamp (489, 560), mast (500, 670), mos (621, 624), orkest (529, 597), paus (600, 789), sigaar (525, 543), stronk(589, 693), vork (548, 598).
Verbs: blonk (597, 708), droeg (520, 556), dwong (547, 518), floot (620, 739), gleed (613, 605), gold (606, 659), hief (762, 812), klom (599, 629), klonk (545, 646), kromp (613, 625),

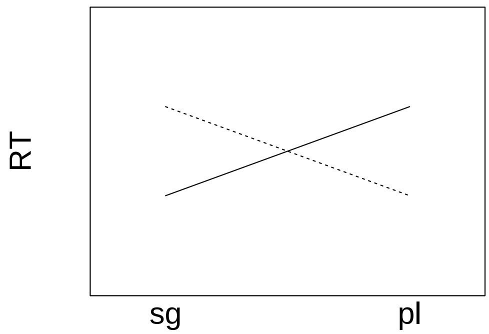
placht (788, 610), rees (638, 690), sloeg (571, 573), slonk (658, 695), smeet (603, 632), snoof (586, 596), snoot (696, 712), ving (622, 636), vlocht (586, 569), vloog (520, 603), vroor (690, 684), wierp (609, 532), zocht (539, 519), zong (528, 478), zonk (559, 511), zwierf (589, 557).

Figure captions.

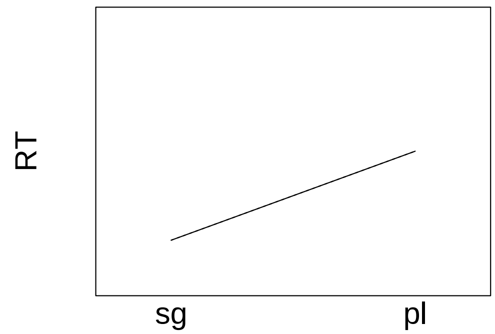
Figure 1. Predicted pattern of reaction times (RT) for singulars (sg) and plurals (pl) for four models of morphological processing. Dotted lines represent plural-dominant pairs, solid lines singular-dominant pairs.

Figure 2. Modeling results for Experiment 1. Observed reaction times of plural-dominant singulars (sg) and plurals (pl) are represented by solid lines, observed reaction times of singular-dominant singulars and plurals by dashed lines. The reaction times generated by the model are plotted with dotted lines.

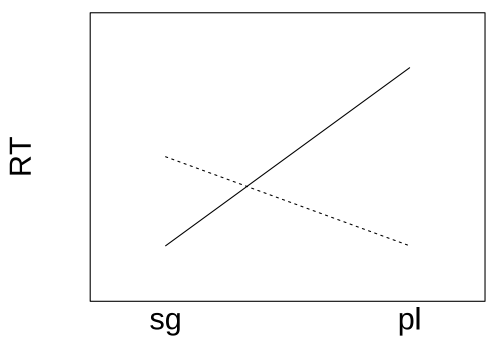
Figure 3. Observed (left) and predicted (right) pattern of reaction times for singulars (sg) and plurals (pl) in Experiment 2. Solid lines denote singular-dominant nouns, dashed lines denote plural-dominant nouns, and dotted lines denote dominance-neutral nouns.



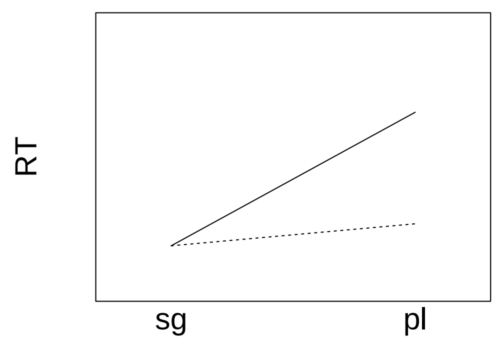
Full Listing Model



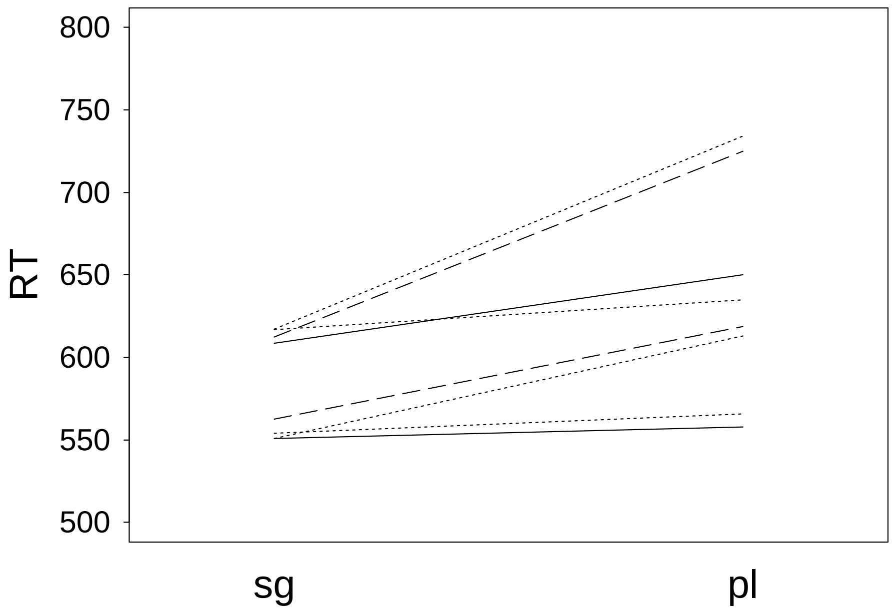
Full Parsing Model

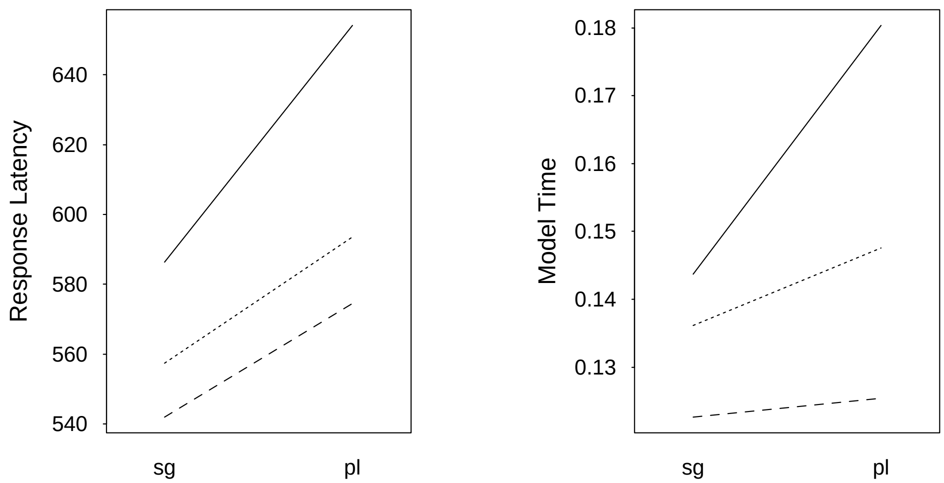


AAM Model



Schreuder and Baayen (1995)


800
750
700
650
600
550
500
RT
sg
pl



Observed Reaction Times

Predictions Race Model

Table 1

Mean Latencies (in Milliseconds) and Percentages of Errors for -en Plurals in Experiment 1.

Stem Frequency	Dominance	Singular		Plural	
High	SgDom	561	(2)	615	(6)
High	PlDom	551	(2)	558	(2)
Low	SgDom	612	(6)	708	(19)
Low	PlDom	606	(7)	645	(9)

Table 2

Mean Latencies (in Milliseconds) and Percentages of Errors

for -en Noun Plurals and corresponding Singulars in Experiment 2.

Dominance	Singular		Plural	
PlDom	542	(4)	575	(3)
Neutral	557	(4)	593	(6)
SgDom	586	(7)	654	(16)

Table 3

Mean Latencies (in Milliseconds) and Percentages of Errors for -en Noun and Verb Singulars and Plurals in Experiment 3.

Word Category	Singular		Plural	
Noun	545	(2)	611	(6)
Verb	603	(7)	612	(10)