Frequency Effects in Processing Inflected Dutch Nouns: A Distributed Connectionist Account

Matthew H. Davis, Maarten van Casteren and William D. Marslen-Wilson

1. Introduction

Words that occur more frequently in language are processed more quickly. This is the longest-standing, and most clearly established result in experimental psycholinguistics (for early summaries of research on word frequency, see Howes and Soloman 1951; Broadbent 1967; Morton 1969). This basic result has been replicated in a range of tasks, such as reading aloud, picture naming, semantic or lexical decisions, in a range of languages. The word frequency effect is taken as evidence that the systems involved in language processing respond to basic statistical properties of an individual's linguistic experience. Furthermore, through careful manipulations of different word properties, it has been possible to observe effects of the frequency of individual morphemes in complex words (Taft 1979; Sereno and Jongman 1997; Baayen, Dijkstra, and Schreuder, 1997) or effects of the frequency of different meanings of homonymous words (Borowsky and Masson 1996; Rodd, Gaskell, and Marslen-Wilson 2002) in order to infer which aspects of word frequency are crucial for a particular task.

However, despite widespread agreement that word frequency plays an important role in language processing, a variety of accounts have been proposed of the mechanism by which frequency effects arise. It has been proposed that word frequency effects arise since access to the mental lexicon involves a search through a frequency-ordered word list (in the search model of Forster 1976). Alternatively, it is proposed that changes to the processing properties of the units that represent individual lexical items (such as lowering the threshold for recognition in the logogen model of Morton 1969 or raising the resting activation level in the interactive-activation model of word recognition proposed by McClelland and Rumelhart 1981) allows high frequency words to be recognised more quickly.

In:

Morphological Structure in Language Processing
R. H. Baayen and R. Schreuder (eds.),
Mouton de Gruyter, Berlin, 2003.

Both serial search and logogen accounts share a single common assumption; namely that there is a single lexical unit that uniquely represents each word – that is, they are localist models. Thus effects of the frequency of occurrence can be localised to changes occurring to a lexical unit. In contrast, a recent class of computational account, distributed connectionist models, propose that individual words are represented as a pattern of activation over many active units, with different banks of units representing domains of word knowledge such as orthography, phonology and semantics. Linguistic knowledge in distributed models is represented by the strength or weights of connections that link these banks of units. These connection weights are not pre-determined but are gradually learnt by training the network to translate from one representation to another (e.g. reading aloud involves a translation from orthography to phonology, auditory comprehension involves mapping from phonology to semantics, etc.). During training, a learning algorithm (typically back-propagation – Rumelhart, Hinton, and Williams 1986) adjusts connection strengths to reduce the discrepancy between the network's output and a target representation.[1]

One crucial aspect of distributed connectionist accounts is that they change the interpretation of a distinction that is common to both linguistics and psycholinguistics; the distinction between rule-governed forms and exceptions. In distributed models, a single set of connections is not only able to acquire knowledge of individual lexical items or exceptions, but is also able to extract and apply regularities to new items, showing how these systems can generalise in a seemingly rule-governed way. For example, in generating the past-tense forms of English verbs from their stems, a single system can learn to phonologically translate both regular verbs (jump-jumped, play-played) and irregular verbs (leap-leapt, give-gave) as well as generalising to novel forms (wug-wugged) (Rumelhart and McClelland 1986; Plunkett and Marchman 1991, 1993). Thus the distinction between rule-governed forms and exceptions to these rules may be an accurate description of the structure of linguistic knowledge but need not imply that there are two underlying mechanisms by which these forms are processed.

In the current paper, we explore whether these distributed connectionist models are able to account for the recognition (rather than the phonological transformation) of regularly inflected words. In describing these simulations we will not only consider the capabilities of distributed connectionist networks (that is, whether they can perform the task) but, crucially, we will evaluate the behaviour of these networks by comparison with experimental investigations of the processing of regularly inflected words. In order to conclude that a single-mechanism distribution connectionist model can account for the recognition of inflected words, we require the model to simulate the pattern of behavioural data produced by human participants. Before describing these network simulations, we will therefore review the target empirical phenomena, and their interpretation in more traditional, localist accounts of lexical representation and processing.

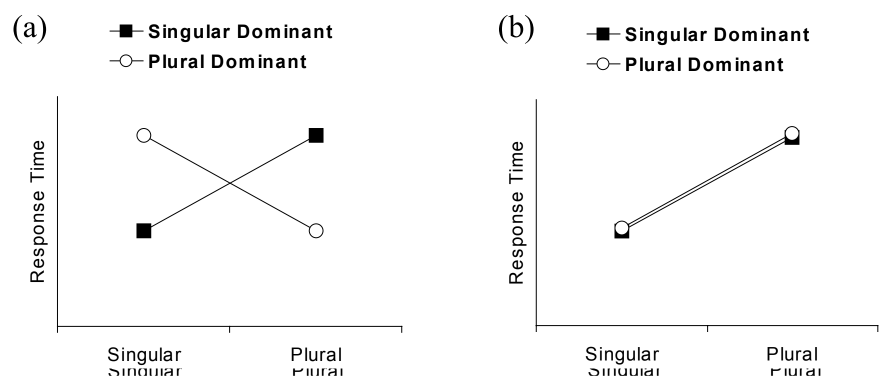
1.1. Frequency effects in the recognition of regularly inflected words

Tasks that involve accessing stored lexical information are sensitive to the frequency of occurrence of these units in the linguistic environment. For example, in the lexical decision task, where participants make a speeded response to indicate whether a letter string or word is a real word or a non-word, a typical response time of 600ms for a word that occurs once in a million words of text, will decrease by approximately 10% for an otherwise matched word that occurs with a frequency of 100 occurrences per million words. In logogen-style models (c.f. Morton 1969), in which words are identified when the activation of the lexical unit exceeds some threshold value, this effect is simulated by assuming that more frequently occurring words have a higher level of resting activation or a lower recognition threshold. In either case, more frequent words have a head-start in the recognition process, and will be identified faster.

One important theoretical question in this framework is whether the critical logogens that are activated during word identification represent whole-words or individual morphemes. Are inflected words like “tables” stored whole or decomposed into the smaller units “table” + “s”? This issue can be addressed using the frequency effect as a diagnostic. For example, if we consider the words “neck/necks” and “lip/lips”, these words are approximately matched on the frequency of occurrence of the lemma[2] {neck} or {lip} at around 80 occurrences/million words in the CELEX database (Baayen, Pipenbrock, and Guilikers, 1995). However, as we might expect (since individuals typically have two lips but only one neck), the frequency of occurrence of singular and plural forms of these two items is very different. The word “neck” occurs many times more frequently than “necks” (word-form frequency, neck = 72/million, necks = 7/million, these frequencies are singular-dominant) whereas the plural “lips” is used much more frequently than the singular “lip” (word-form frequency, lip = 17/million, lips = 61/million, showing that the noun {lip} is plural-dominant).

By an account in which there are individual logogens for the singular and plural form of these words (a whole-word storage account), we would expect substantial differences in response times to “lip” and “neck” or to “lips” and “necks”, with the singular of the singular dominant form being responded to more quickly than the singular of the plural dominant form, and vice-versa for plural forms. This pattern of results is illustrated graphically in Figure 1a. Conversely, if both of these words were decomposed into smaller units such that a single lemma unit (for {neck} or {lip}) was activated for both singular and plural forms, we would not expect a difference in response time between singular- and plural-dominant nouns, so long as those items are matched on lemma frequency. To build a model that functions in this way, we need to invoke an additional process that allows the inflected words “necks” and “lips” to gain access to the appropriate lemma representation. This decomposition process may add to the time required to respond to plural nouns so that although there is no difference between response times to singular- and plural-dominant nouns, responses are generally slowed to plural forms. This predicted pattern is illustrated in Figure 1b.

Figure 1 Predictions of (a) simple storage and (b) full decomposition accounts for the processing of singular-dominant nouns (e.g. neck) and plural-dominant nouns (e.g. lip).



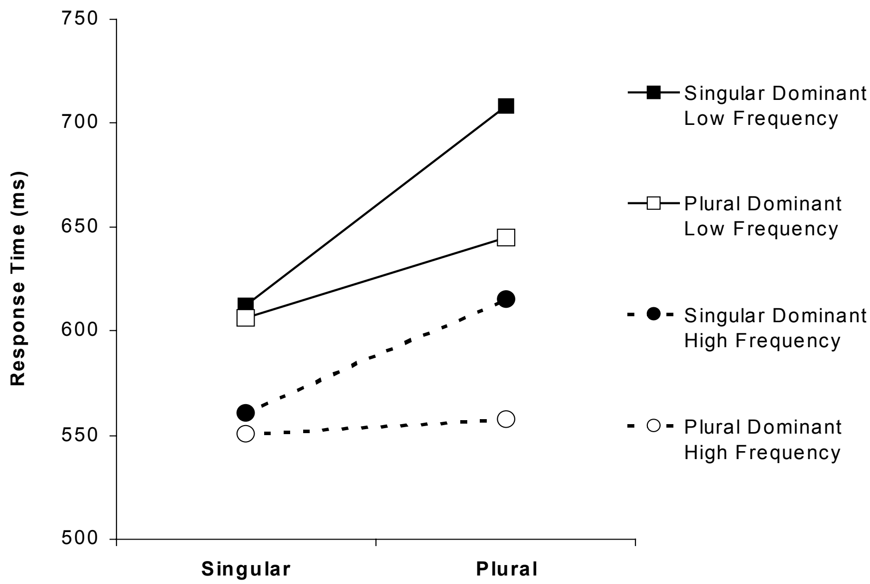
Experiments of this form were first reported in English by Taft (1979). Taft observed (Expt 2) that lexical-decision response times showed an effect of lemma frequency for inflected and uninflected words that were matched on word-form frequency, a pattern suggesting decomposition of inflected forms. However, in a further experiment (Expt 3), Taft also observed an effect of word-form frequency on response times to lemma-frequency matched items; a result that suggests that inflected words are stored whole forms. This pattern was interpreted by Taft as evidence for both whole-word and decomposed lexical representations at different levels of the recognition system.

However, other authors have criticised the experimental materials used by Taft (1979). For instance, Sereno and Jongman (1997) pointed out that Taft's items were not matched for word class, and that the groups of words that contained more nouns would be likely to produce faster responses. Furthermore, it is unclear whether all three of the inflections tested by Taft (-ed, -ing and -s) are equally sensitivity to word-form and lemma frequency. Later in the paper we will review the results of experiments in Dutch and Finnish conducted by Bertram and colleagues (Bertram, Laine, and Karvinen, 1999; Bertram, Schreuder, and Baayen, 2000) which suggest differences in the behaviour of words with different inflectional endings.

In seeking to replicate the Taft (1979) findings, Sereno and Jongman (1997) tested only nouns and used the same set of materials in both singular and plural form. Since word-form frequency effects should arise in opposite directions for the two forms (as depicted by the interaction in Figure 1a), this effect is less likely to arise from a simple confound in the experimental materials. However, Sereno and Jongman conducted separate experiments with different participants on singular and plural nouns (with both words and non-words either all inflected or all uninflected) such that some participants were tested on a word list consisting entirely of inflected items. These participants could have adopted different response strategies from those tested on a mix of inflected and uninflected items – for instance, it is possible that inflectional affixes would be ignored in making a lexical decision response. Furthermore, since the Sereno and Jongman experiments included only a small number of items (12 per condition) it is possible that a lack of statistical power may be responsible for the null effects of lemma frequency that they report.

We will therefore focus on results reported for Dutch nouns by Baayen, Dijkstra, and Schreuder (1997). Their experiment tested lexical decision responses to singular and (-en) plural forms of 93 nouns divided into high and low lemma frequency sets. Each of the 100 participants were tested on a mix of singular and plural nouns with inflected and uninflected non-words to ensure that participants must process inflectional endings in making their lexical decisions. Test items were divided into two lists to ensure that participants did not see both forms of any noun, avoiding priming effects between plurals and singulars. Given the size of the data set we can be reasonably sure that this experiment had sufficient power to detect small behavioural effects. The pattern of response times obtained by Baayen, Dijkstra, and Schreuder is shown in Figure 2.

Figure 2 Results of Experiment 1 from Baayen, Dijkstra, and Schreuder (1997); response times to singular and plural forms of high and low frequency, singular- or plural-dominant nouns.



As can be seen by comparison of Figure 2 with Figure 1a, Baayen, Dijkstra, and Schreuder's (1997) results did not conform to the predictions of a whole-word storage account. Although there was a reliable effect of word-form frequency on response times to the plural nouns, there was no significant difference between response times to the singular form of singular- and plural-dominant nouns. Comparing the results shown in Figure 2 with the predictions of Figure 1b shows that this experiment also failed to confirm the predictions of an account based on decomposition of plural nouns. There was a highly significant effect of dominance on response times to plural forms; an effect that would not be predicted by an account in which plural forms are decomposed and recognised on the basis of a shared lemma representation.

In summary, the results of the experiment reported by Baayen and colleagues shows a pattern that is a mixture of simple storage and full decomposition accounts. Correspondingly, they interpret these results as being in line with the predictions of accounts of word recognition which postulate that both of these processing mechanisms – whole-word storage and morphological decomposition – are involved in the identification of inflected forms.

Dual-route or dual-mechanism models such as these have been proposed by a variety of authors (Carramazza et al. 1985; Pinker 1991; Frauenfelder and Schreuder 1991), each making different assumptions regarding the relative role of storage and decomposition in lexical processing. For example, Pinker[3] (1991) proposes that all regularly-inflected forms are decomposed during processing and accessed via their stems, except for irregular forms which must be stored as whole forms (such as the irregular plural "mice" in English). In the Augmented Addressed Morphology (AAM) model proposed by Carramazza and colleagues (Caramazza et al., 1985; Caramazza, Laudanna, and Romani, 1988), it is suggested that all complex forms that have been recognised previously will have a stored lexical representation and only new or very low frequency plurals (e.g. nouns that have only been seen in the singular form) would be decomposed. A careful consideration of these two models would suggest that despite the presence of two processing mechanisms, neither account would predict the exact pattern that was observed experimentally.

For instance, although the Words and Rules model proposed by Pinker (1991) could account for finding a lemma frequency effect (or an absence of a word-form frequency effect) for singular nouns it would still predict similar results to the full-decomposition account for plurals – i.e. that there would be no dominance effect. Conversely, the AAM model proposed by Carramazza and colleagues can account for the word-form frequency effect for plurals in two different ways; for high-frequency lemmas, plurals of both singular and plural-dominant nouns would be stored and word-form frequency effects would be observed. Secondly, for low-frequency or unfamiliar plurals (such as for low-frequency, singular-dominant nouns) the need to decompose these items would further increase the size of the dominance effect. Nonetheless, for the singular forms (which do not require any decomposition) the AAM model appears to predict a word-form frequency effect that was not observed experimentally.

Part of the problem for both of these accounts is that although they include two processing mechanisms, the recruitment of each of the two routes is a fairly strict 'either/or' based on the familiarity or regularity of the target item. In practice, for the experimental materials used by Baayen and colleagues (which are of reasonable frequency and entirely regular), processing in both of these models would be dominated by one of the two available routes (decomposition in the Pinker model, storage in the AAM model). As is apparent from the comparison of Figure 1 and 2, neither of these single mechanism profiles are appropriate for Baayen and colleagues' data.

The Morphological Race Model (MRM – Baayen, Dijkstra, and Schreuder, 1997; Frauenfelder and Shreuder, 1991; Schreuder and Baayen, 1995) proposed to account for these data again includes mechanisms of both whole-word storage and morphological decomposition. However, rather than selecting one of the two routes for each type of item, the MRM proposes that the two processing routes 'race' against each other, with the output of the winning route (i.e. whichever process is completed faster) accounting for the processing of any particular item. Such a race model provides for a dynamic assignment of items to the two routes, such that whichever process operates more rapidly and efficiently will determine the response time for a particular item. Critically, however, processing is still completed in the non-winning route and the results of the slower route will still influence the behaviour of the model under certain circumstances.

In a series of mathematical simulations, Baayen, Dijkstra, and Schreuder (1997) show that the MRM predicts the correct pattern of results when operating under the following constraints. They propose that the decomposition process operates fairly slowly for –en plurals[4]. For this reason, the majority of the plural nouns are recognised by the faster whole-word route and hence response times for plurals will be primarily determined by word-form frequency (hence the dominance-effect observed for plural forms). Despite the fact that decomposition operates slowly, Baayen and colleagues (1997) still expect that this processing route would correctly analyse plural forms – determining that "lips" is the plural of the singular noun "lip"). They further propose that successful decomposition of plural forms alters the representation of the noun stem by boosting the resting activation of the lexical unit for the stem. In this way, although decomposed processing can not be readily observed in lexical decision responses to plurals (since this route does not win the race and initiate a response), the results of decomposition can be detected in the processing of singulars, since it is the combined frequency of singular and plural forms that determines response times.

Baayen, Dijkstra, and Schreuder (1997) interpret the combination of word-form frequency effects for plurals and lemma frequency effects for singulars as evidence that uniquely favours the dynamic combination of storage and decomposition proposed in their dual-route morphological race model. In further work they show that this model (and identical parameter settings) can readily simulate response time data from experiments in which other manipulations of surface frequency and lemma frequency are made. In conclusion, they suggest that not only do other dual mechanism models fail to predict the correct pattern of results but also that (p. 113):

> "it is difficult to see how these patterns could be understood using monolithic neural nets... modelling in one pass what in our view is a complex multilayered system".

It is this challenge that we address in the current paper, exploring Baayen's claim that this complex pattern of results can not be simulated using a single-mechanism, distributed connectionist model.

1.2. Frequency and regularity in distributed connectionist models

Previous simulation work has shown that distributed connectionist models trained on a variety of mappings are sensitive to the frequency of particular items presented during training as well as the extent to which components of the input-output mapping are consistent across different items (i.e. regularity). For example, in models of the computation of phonology from orthography (e.g. the Seidenberg and McClelland (1989) model of reading aloud), the phonology of words that occur more frequently in the training set is computed with reduced error. Furthermore, error rates are also lower for items with orthographic neighbours that are pronounced in a consistent way (such as "hint", "mint", "splint", "tint" etc.) than for items that are inconsistent with their neighbours (e.g. "pint"). These two effects interact such that effects of frequency are larger for irregular or inconsistent items and that effects of consistency are larger for low frequency items, a pattern reminiscent of that observed in experimental investigations of reading aloud (Taraban and McClelland 1987).

This interaction between frequency and regularity reflects the sensitivity of the network to the frequency of whole forms and the frequency of regular components of those forms. The network learns to associate combinations of letters with speech sounds one word at a time. After each experience of a particular word, network weights are changed so as to capture the relationship between spelling and sound for that word. Items on which the network is trained more often (i.e. words that occur more frequently in the language) will have more opportunity to alter the network's weights and hence a processing advantage is observed for high frequency words. Importantly, the network learns the orthography-phonology mapping by generalising the spelling-sound correspondences found in individual words. For items that have a consistent relationship between spelling and sound (such as the rhyming set of -int words listed before), training on one item will benefit other words that are spelt and pronounced in the same way. Conversely, exception words (such as "pint") do not benefit from the influence of their neighbours and must be learnt as whole forms. The network is therefore much more sensitive to the frequency of presentation of exception words (for a more detailed explanation and mathematical treatment of frequency by regularity interactions, see Plaut, McClelland, Seidenberg & Patterson, 1996).

Having described how frequency and regularity affect models of reading aloud, we can now consider how these properties may be extended to account for the empirical data of Baayen, Dijkstra, and Schreuder (1997). However, previous simulations have shown that networks trained on the task of mapping orthography to phonology are incapable of performing lexical decision at a human-like level of accuracy (Seidenberg and McClelland 1989; Besner et. al. 1990). In order to account for behavioural data obtained from the lexical decision task, we require simulations that map from the spelling (or sound) of words to their meanings (c.f. Plaut 1997; Gaskell and Marslen-Wilson 1997). Since these distributed networks use essentially the same learning algorithms and computational mechanisms we may assume that similar effects of frequency and regularity will be apparent in these mappings between form and meaning.

In the mapping between form and meaning, however, systematic regularities between input and output representations are fewer in number; the relationship between the form and meaning at the single word level is essentially arbitrary. A notable departure from this arbitrariness is provided by the presence of morphological units in the input (such as the English plural affix -s which marks the plurality of nouns, or the similarity in meaning of "lip" and "lips" provided by their shared stem). By extension of the principles found in models of reading aloud we would expect that differences in the frequency of morphological components of the form-to-meaning mapping would be reflected in the behaviour of a trained connectionist network. Indeed prior simulations have shown the effect of these regularities in the learning profile (Rueckl and Raveh 1999), internal representations (Davis, Marslen-Wilson, and Hare 1996; Rueckl and Raveh 1999) and priming behaviour (Plaut and Gonnerman 2000) of distributed connectionist networks. Nonetheless, detailed simulations of the effect of word and lemma frequency on the identification of morphologically complex words have not been conducted. In the current manuscript, we report simulations exploring the behaviour of a distributed connectionist model of the processing of regularly inflected words. Our goal was to simulate the results reported by Baayen, Dijkstra, and Schreuder (1997) on the processing of Dutch singular and plural nouns.

2. Simulation 1 – Modelling Dutch plural morphology

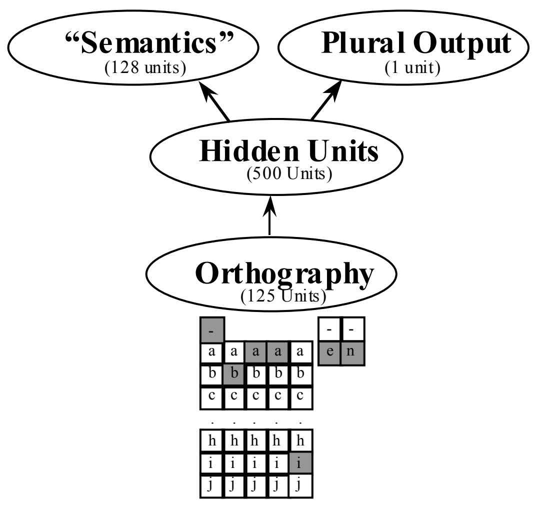
All of the simulations reported in this chapter used a standard, 3-layer feed-forward network of units with a sigmoidal activation function (as used by Seidenberg and McClelland 1989; Plaut and Gonnerman 2000 and others). To simulate the results of visual lexical decision experiments in a distributed network we require a system that maps from a representation of the visual form of a word to a representation of its meaning or semantics. Since many of the critical test items for the network will be inflected with the Dutch plural affix (-en), the network should accommodate nouns that are marked for plurality. For this reason an additional output unit was added to the semantic layer, to be activated in response to nouns presented with the plural affix. As the relationship between word-form and meaning for noun-stems is essentially arbitrary, 500 hidden units were required to map between input and output representations. Despite this large number of hidden units (chosen to improve overall performance and learning time), the network still does not resort to the localist solution of assigning single words to single hidden units (see Bullinaria and Chater 1995, for further discussion). The architecture of the network is depicted in Figure 3.

2.1. Training Procedure

The network was trained on a set of 582 four and five letter, monosyllabic Dutch nouns taken from the CELEX database (Baayen, Pipenbrock, and

Guilikers 1995), all of which took the -en plural affix. This set of nouns included 91 out of 93 nouns from the Baayen, Dijkstra, and Schreuder experiment – two items were excluded for which one form (the plural of singular dominant forms) had a frequency of 0 occurrences in the 42 million CELEX corpus. The network was trained on the singular and plural form of each noun. A further 379 Dutch verbs were added to the training set, though only the verb stem was trained. In selecting the nouns and verbs, we ensured that the two groups of test items had a representative number of orthographic neighbours in the network's training vocabulary (cf. Gaskell and Marslen-Wilson 1997).

Figure 3 Architecture of the network trained in Simulation 1, including an illustration of the input units activated to represent the plural noun "baaien".



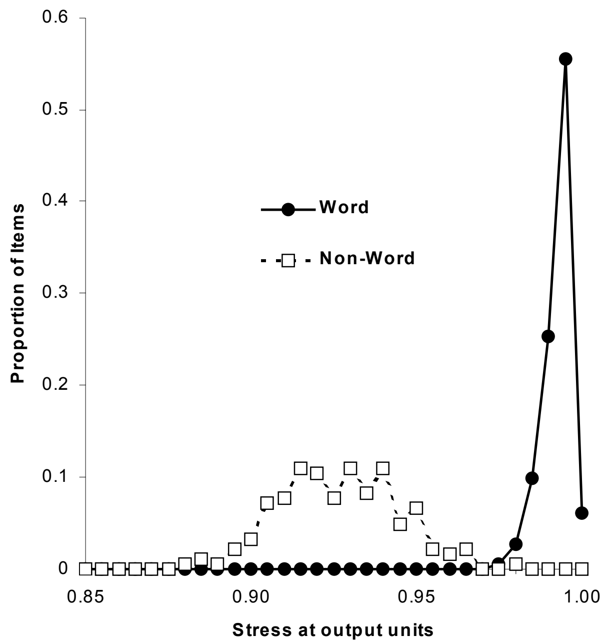
The orthographic input to the network was represented over a bank of units, each of which represented a single letter in a particular position of the word. Thus for the first letter position, there was a bank of 27 input units (representing each of the 26 possible letters, or a blank if no letter was present). This bank of units was duplicated 5 times – once for each letter in the noun or verb stem. Units that were not used by any word in the network's vocabulary were removed for computational convenience. For each of the noun and verb stems, the word was right-aligned in the template (such that the blank space unit was only ever used in the first letter position). For items presented in the plural, the orthographic representation of the stem was always placed over the same bank of units, with the additional letters for the plural affix added to two additional slots – as shown in Figure 3. Although this representation provides the network with a solution to the morphological alignment problem, we are confident that, using appropriate techniques, a distributed connectionist model could discover this (or an equivalent) scheme for itself (cf. Bullinaria 1995). The output representation for the network was a randomly generated binary vector over 128 semantic units, with ten units activated for each noun or verb stem. The semantic vector representing each noun was identical for both the singular and the plural form, except for the activation of the additional unit for plurals.

The network was trained using the back-propagation learning algorithm (Rumelhart et al. 1986), and the cross-entropy error function (Hinton 1989) to produce the correct output representation for each input (learning rate = 0.01 throughout, momentum = 0.9, introduced after the first ten epochs). Weight-updates were performed after the presentation of a single word (on-line learning) and training patterns were presented in a random order in each epoch. To simulate the effect of word-frequency, rather than training more often on more frequent words, the magnitude of each weight change was multiplied by the base 10 Log of the frequency of occurrence of the word being trained (cf. Plaut et al. 1996). This training procedure continued until output activations for every training pattern were within 0.4 of the target at all output units, which required just over 100 passes through the training set.

2.2. Testing procedure

In order to simulate the experimental data in the trained network, we first of all need to obtain lexical decision responses. To distinguish between words and non-words we measured the mean 'stress' of all the output units (including the plural unit) using the equation from Plaut (1997)[5]. Output stress is maximal (1.0) when output activations are close to zero or one and is low (0.0) when output activations are close to 0.5. Since the network was trained to produce activations of zero or one for real words this measure can distinguish between real and nonsense words. For the set of 182 Dutch non-words used by Baayen, Dijkstra, and Schreuder (1997) and a stress threshold of 0.98, the network made lexical decision response with more than 95% accuracy (see Figure 4).

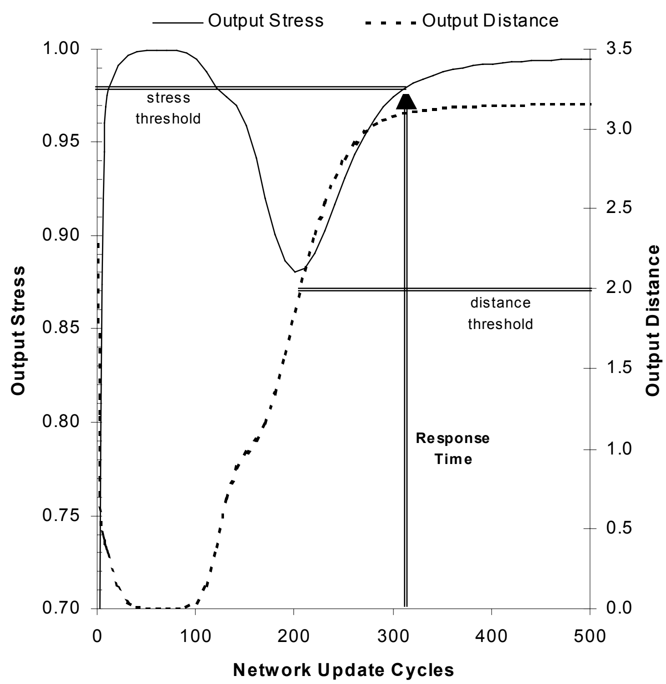
Figure 4 Histogram showing values of the stress measure applied to the output of the network to distinguish between real words and non-words.



As is the case for most feed-forward neural networks trained using back-propagation, output activation is produced in a single forward pass. It is therefore not possible to measure response times directly from the network. Rather than following some previous models by assuming that output error is correlated with response time (Seidenberg and McClelland 1989) we used the equations provided by Cohen, Dunbar, and McClelland (1990), derived from McClelland (1979), to cascade activation through the network[6], allowing us to obtain a direct measure of processing time. In all of the simulations reported, the time-constant τ was set at 0.01, allowing fine-grained differences in response time to be observed.

In order to determine lexical decision response times, we used two criteria based on output activation in the network. The first criterion was that mean output stress should exceed 0.98 – as shown in Figure 4 this allows the network to distinguish between words and non-words. However, this output stress measure is high early on in processing before orthographic information has yet been propagated through the network. In the absence of any input the output of the hidden units is set to 0.5 which is sufficient to set the output to zero since the hidden to output weights are mostly negative[7]. We therefore needed to apply an additional criterion to the output to determine when a response should be made. This second criterion was that the Euclidean distance between the output activation and the origin (i.e. the square root of the summed-squared activation of each output unit) should exceed 2.0. This ensures that the network waits until activation has begun to arrive at the output layer before making a response.

Figure 5 Time-course of output stress (left) and output distance (right scale) during cascaded processing of an example word. The network's response is made when both thresholds are reached.

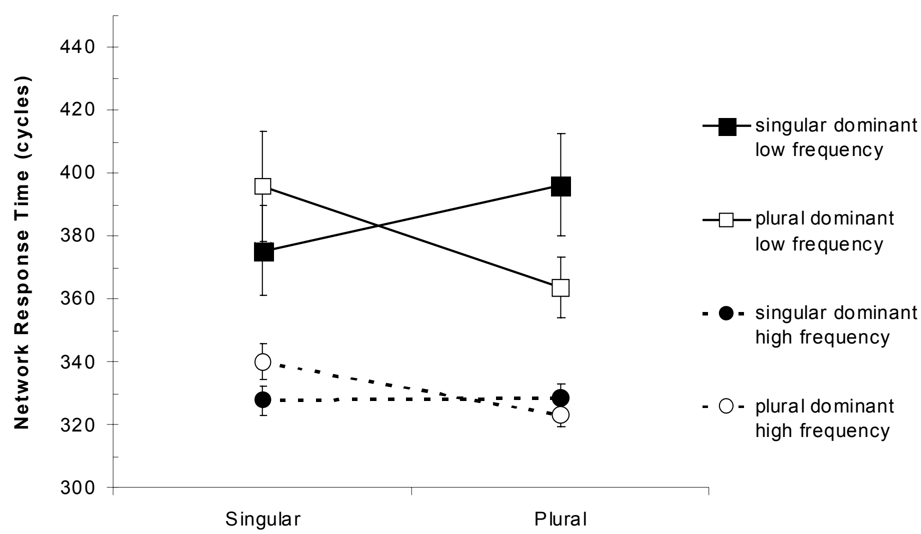


The time-course of a typical processing trial, as reflected in the stress and output distance measure, is illustrated in Figure 5. As shown in the figure, the response time of the network is the update cycle at which the output satisfies both of these criteria. Should the network output fail to satisfy both of these criteria for a test word we assume that the network made a lexical decision error.

2.3. Results

The lexical decision response time of the network for the pairs of singular and plural nouns tested by Baayen, Dijkstra, and Schreuder (1997) is shown in Figure 6. Excluded from this graph are 6 items on which the network made a lexical decision error (3% of the data). By comparison with the experimental data shown in Figure 2, the results of the network simulation do not resemble the experimental data. This failure is confirmed by statistical analysis.

Figure 6 Results of simulation 1. Lexical decision response times to singular and plural forms by dominance and frequency. Error bars show one standard error of the mean over items.



In their analysis of the behavioural data, Baayen and colleagues report significant main effects of number (faster responses for singulars than for plurals) and dominance (faster responses for plural dominant forms) as well as a significant interaction between these two factors (indicating that responses were particularly slowed for the plurals of singular dominant nouns). This pattern was essentially identical for both high and low frequency groups, reflected in a non-significant three way interaction between number, dominance and frequency.

The equivalent three-way analysis of variance was conducted on the results of the network simulation investigating effects of number, dominance and frequency (by items only, since only a single network was tested). This ANOVA showed that the network's lexical decision responses were significantly faster for high frequency items ($F[1,168]=49.53$, $p<.001$). There was no significant main effect of dominance or number, nor any interaction between either of these factors and frequency ($F<1$). However, there was a significant cross-over interaction between number and dominance ($F[1,168]=5.56$, $p<.05$) indicating that dominance had opposite effects for the singular and plural forms. This two way interaction did not differ between the two frequency bands as indicated by the non-significant three-way interaction ($F[1,168]=1.41$, $p>.1$).

The response times in this simulation closely resemble the predictions of the whole-word storage account (Figure 1a), in which response times are primarily determined by the frequency of the individual word-forms. Lexical decision times in the network simulation were faster to the singular form of singular-dominant nouns and to the plural form of plural-dominant nouns. This suggests that the network is only sensitive to the frequency of occurrence of the whole form during training.

However, before we can conclude that the behaviour of the network is solely determined by word-form frequency and not by lemma frequency we need to conduct a comparison of items that are matched on word-form frequency yet differ on lemma frequency. This comparison was achieved by conducting an additional simulation using a set of verb stems. A selection of these items were assigned word-form frequencies that were matched to the frequency of the test items. Although the verb stems will have the same word-form frequency as the test nouns, since the network is only trained on a single uninflected form they will differ maximally in lemma frequency. Therefore any difference between the response times of these matched verb stems and response times for the test nouns in this supplementary simulation shows the effect of training on the plural forms in the initial simulation (i.e. the magnitude of the lemma frequency effect).

Response times to the singular form of the test nouns and the matched verb stems in this supplementary simulation shows a dramatic effect of lemma frequency (singular test noun RT, low frequency = 378 cycles, high frequency = 331 cycles; verb stem RT, low frequency = 446 cycles, high frequency = 414 cycles). The results of this supplementary simulation therefore show that responses to nouns are speeded by the presence of a morphologically related forms in the training set and does not only depend on word-form frequency. Indeed, this lemma frequency effect appears to be numerically larger than the effect of word-form frequency[8] observed in Figure 6.

We therefore conclude that the network trained in Simulation 1 shows neither of the classic 'single-mechanism' profiles depicted in Figure 1. Although there is some resemblance between the results of Simulation 1, and Figure 1a, the size of the word-form frequency effect is smaller than would be expected if this were the only factor that affected the behaviour of the network. The results of the supplementary simulation shows a substantial effect of the presence of morphologically-related items – indicating that the network is affected by the combined frequency of singular and plural forms (i.e. lemma frequency).

Despite this complex pattern of behaviour, a pattern that corresponds to neither full decomposition of inflected words, nor simple storage of whole words, the network tested in Simulation 1 fails to capture the pattern of results reported by Baayen, Dijkstra, and Schreuder (1997) and shown in Figure 2. The network predicts a word-form frequency effect for singulars that is absent in the experimental data. Perhaps more significantly, as shown by our supplementary simulation, the word-form frequency effect is small by comparison with the experimental data. We will focus on the effect of word-form frequency for plural nouns in explaining the network's failure to simulate the experimental data.

2.4. Discussion

The results of Simulation 1 in combination with the supplementary simulation shows a complex combination of word-form and lemma frequency affects lexical decision RTs in the network. The lemma frequency effect indicates that the network is using the consistent mapping between the orthography and semantics of the stem in singular and plural nouns. This finding replicates results reported by Davis, Marslen-Wilson, and Hare (1996), Gasser (1994), Plaut and Gonnerman (2000), Rueckl and Raveh (1999) and others in showing that distributed networks can extract morphological regularities that are present in the predominantly arbitrary mapping from orthography to semantics. However, since there are differences in response times to singular and plural dominant nouns that are matched on lemma frequency, it is clear that the network's responses are also affected by word-form frequency. Thus the network does not only decompose morphologically complex forms, but also retains knowledge of the individual lexical items on which it was trained.

Nonetheless the combined effect of word-form and lemma frequency shown by the network fails to simulate the reaction time data reported by Baayen, Dijkstra, and Schreuder (1997). Continued training of the network or changes to parameters such as learning rate or number of hidden units appears insufficient to account for the experimental data. Although the magnitude of word-form frequency effects for high frequency nouns is reduced at later stages of training, the network never produces the tilted wedge observed experimentally and statistical analysis of the network's performance does not produce the main effect of number (i.e. slower responses for plural nouns) that is observed experimentally. Indeed the numerical trend often runs in the reverse direction with faster response times for plural than for singular nouns.

The network's failure to simulate the Baayen, Dijkstra, and Schreuder (1997) data shows a discrepancy between the rapid identification of plural nouns in the network and the rather slower identification of plurals that was observed in the behavioural experiment. In the network, the recognition of nouns appears to benefit from training on related forms and the network displays an effect of lemma frequency for inflected forms that is not typically observed experimentally. Word-form frequency is generally a better predictor of response times to morphologically complex forms, not only in the experiments reported by Baayen and colleagues (1997), but also in a number of other studies, for many different types of inflected and derived forms and in a number of different languages including English (Sereno and Jongman 1997; Ford, Davis, and Marslen-Wilson, this volume), Dutch (Bertram, Schreuder, and Baayen 2000) and Finnish (Bertram et al. 1999; Bertram, Hyönä, and Laine 2000). The pattern observed in our simulation – effects of lemma frequency on response times to morphologically complex forms– is rarely observed experimentally and confined to only a restricted set of affixes in these two languages.

Bertram, Schreuder and Baayen (2000) propose that those morphologically complex forms for which response times are best predicted by lemma or stem frequency are those that include affixes with particular properties. For two typologically-distinct languages (Dutch and Finnish), they suggest three factors that determine the presence of word-form or lemma frequency effects for morphologically complex forms – namely: (1) the word-formation type of the affix (whether it is a inflectional, meaning-invariant affix or a derivation, non-invariant affix); (2) the productivity of the affix (whether it is used to form new words); and (3) whether the affix is homonymous (that is, does the orthographic ending serve two or more distinct morphological functions). In the context of a dual-route race model, each of these factors increases the difficulty of the decomposition process. It is only for complex words formed using productive, meaning invariant, non-homonymous affixes that the decomposition process can be completed sufficiently rapidly to allow response times in the lexical decision task to be affected by properties of the stem such that lemma frequency predicts response times for morphologically complex words.

In the case of the Dutch plural affix -en, a critical property of this inflection is that in addition to marking plurals, the -en ending also marks the infinitive form of verbs, as well as the past-participles of certain verb classes (e.g. vang-en, "to catch", ge-vang-en, "caught"). For a homonymous affix such as -en, Bertram, Schreuder and Baayen (2000) propose that word-form frequency will be the best predictor of response times to inflected forms. This possibility is anticipated by Baayen, Dijkstra, and Schreuder (1997: 107) who suggest that "the polyfunctionality of –en might be the critical factor" in explaining why the decomposition process is slow and hence why effects word-form frequency are so pronounced for plural nouns. For this reason a further simulation was conducted to investigate whether the inclusion of an additional (verbal) interpretation of the -en ending similar affects the behaviour of the network.

3. Simulation 2 – Adding affix homonymy

One notable difference between the training vocabulary for Simulation 1 and the properties of the Dutch language is that the network is only taught a single interpretation of the -en affix. In Dutch this ending serves at least two distinct functions, marking the infinitive form of verbs as well as the plural form of nouns. When presented with an -en inflected item in a single-word lexical decision experiment, it will only be possible for participants to interpret the affix correctly (as a plural or infinitive marker) if they know whether the stem is a noun or a verb. This disambiguation process is not required by the network in Simulation 1, however, which is only taught a single interpretation of the -en affix.

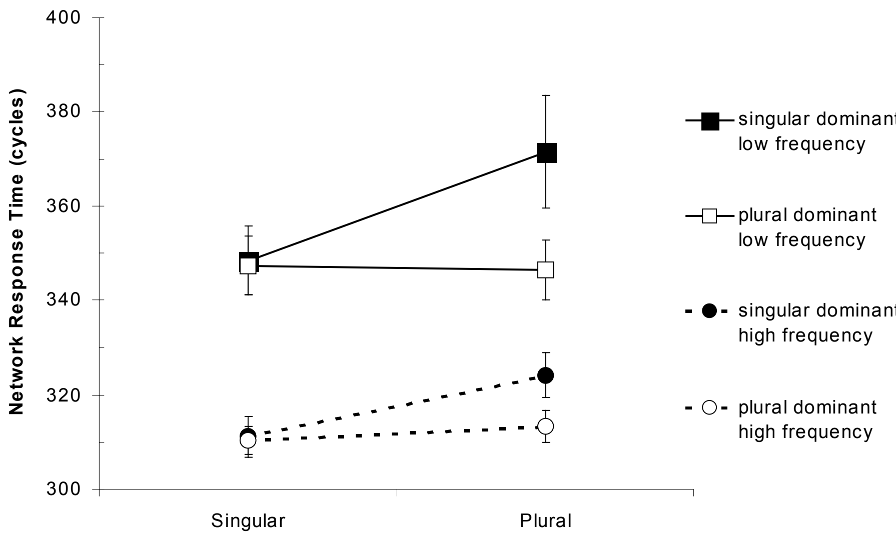
In a second simulation we therefore investigated the effect of adding a second interpretation of the -en affix. We trained a network on the same materials as before but this time included the infinitive (-en) form as well as the stem of the 379 verbs in the training set. To represent these infinitive forms, an additional unit was added to the output, with the network being trained to activate this unit when a verb infinitive was presented. In all other respects, the training set and procedure was as before, the only change made to the network architecture shown in Figure 3 was the addition of an output unit for verb infinitives. This unit was connected to all units in the hidden layer in the same fashion as the plural unit for Simulation 1

The network was once more trained until it reached criterion (all targets within 0.4 of their target), this process taking approximately 40% more epochs than for Simulation 1 – indicating the greater difficulty introduced by the addition of the 379 extra forms into the training set as well as the ambiguity created by including two interpretations of the -en ending. The same testing procedure was used as previously, with activation cascaded through the network, and lexical decision responses determined by a stress and output distance threshold. Perhaps as a consequence of the additional training sweeps that the network had received, the responses of the network were faster and less error prone in this simulation.

3.1. Results

The results of this simulation on the test items from Baayen, Dijkstra, and Schreuder (1997) is shown in Figure 7. The behavioural profile of the network is clearly a much better match to the experimental data in Figure 2. This increased similarity between model and data is confirmed statistically. A three-way ANOVA investigated the effect of number, frequency and dominance on the lexical decision response times generated by the network, producing results comparable to those obtained in the behavioural study. The network showed significant main effects of all three variables, frequency ($F[1,173]=71.76$, $p<.001$), number ($F[1,173]=4.389$, $p<.05$) and dominance (F[1,173]=4.28, p<.05). The source of the main effect of number and dominance is clarified by the presence of a marginally significant interaction between number and dominance (F[1,173]=3.48, p<.1) showing that responses were slowed specifically for the plurals of singular-dominant nouns – consistent with the pattern shown in the empirical data. All other two- and three-way interactions were non-significant in analysis of the networks behaviour (all F<1).

Figure 7 Results of simulation 2. Lexical decision response times to singular and plural forms by dominance and frequency. Error



bars show one standard error of the mean over items.

We will explore the reasons behind this apparent difference in the effect of frequency dominance on singular and plural forms in the general discussion. However, we will first report an analysis to determine the significance of these differences in the behaviour of these two simulations. These comparisons were conducted using a four-way analysis of variance, with responses of the two different simulations coded as a repeated measure over items. For data from the two simulations, there was a main effect of frequency (F[1,168]=70.35, p<.001) and an interaction between number and dominance (F[1,168]=6.02, p<.05). More interesting, though, this analysis also included main effects and interactions indicating significant differences between the two simulations. There was a main effect of simulation, reflecting faster responses in Simulation 2 ($F[1,168]=72.23$, $p<.001$). This finding is likely to reflect the greater number of weight updates that were required for the network in Simulation 2 to reach the criterion for stopping training. As might be expected (since low-frequency items will be slowest to train), this effect interacted with frequency ($F[1,168]=10.49$, $p<.001$) such that the advantage for Simulation 2 was greater for low frequency items. Of greater interest was a significant interaction between simulation and number ($F[1,169]=7.71$, $p<.01$). In Simulation 1, there was no overall difference in response time between singulars and plurals, whereas Simulation 2 showed a response time advantage for singular nouns as observed in the experimental data. Finally, there was a marginally significant interaction between simulation, number and dominance ($F[1,168]=2.99$, $p<.1$) reflecting a change from the cross-over interaction between number and dominance in Simulation 1 (Figure 4), to the wedge-shaped pattern significant main-effect and interaction produced by Simulation 2 (Figure 6). All other main effects and interactions in this analysis failed to reach significance (all $p>.1$).

4. General Discussion

We have presented two neural network simulations of the recognition of singular and plural Dutch nouns. Both networks map orthographic forms onto a representation of the meaning and plurality of a large number of nouns. They also provide a realistic simulation of the processes involved in making speeded discriminations between real words and pseudo-words (i.e. items that do not appear in the network's training set) – and can therefore be compared to reaction-time data obtained using the lexical decision task. Both simulations were trained using the same procedures and to the same criteria. They were tested on the same set of items using the same parameters for cascading activation and measuring responses. Despite these similarities, we observed significant changes to the behavioural profile observed in simulation experiments investigating effects of word-form and lemma frequency in processing singular and plural dominant nouns. In Simulation 1, response times reflected both word-form and lemma frequency for singular and plural forms of the test nouns. In Simulation 2, responses to plural nouns were slowed and showed more pronounced effects of word-form frequency, consistent with experimental data obtained in Dutch.

The critical difference between the two networks is the presence of a homonymous affix in the training set. Whereas in Simulation 1 the -en ending was unique to noun plurals, the second simulation more accurately captured the dual function of this affix which also marks the infinitive and other forms of Dutch verbs. In comparing the results of the two simulations, we have seen that it is only when this second interpretation of the inflectional ending is included that the network can simulate the pattern of behavioural results observed for Dutch nouns. This finding illustrates an important point; the behaviour of a distributed connectionist network not only depends on details of the network architecture and processing mechanisms, but also on the statistical structure of the linguistic environment on which it is trained (see Elman et al., 1996 for further discussion). We will describe in some detail how and why affix homonymy plays such an important role in determining the behaviour of the network, in particular comparing the processing properties and internal representations of the two simulations.

4.1. Effects of Affix Homonymy

In the behavioural literature on frequency effects in morphological processing, Bertram and colleagues have noted that complex words formed using homonymous affixes show effects of word-form rather than lemma frequency (Bertram, Laine and Karvinen, 1999; Bertram, Schreuder and Baayen, 2000). In explaining this finding it is proposed that the decomposition of these items is slowed by homonymy; in particular the need to resolve the ambiguity introduced by the affix. For this reason, the recognition of items with homonymous affixes (in speeded tasks such as lexical decision) is driven primarily by stored, whole-word representations. As we have seen, affix homonymy seems to have a similar effect on the network trained in Simulation 2, though our explanation will differ since there is no clear distinction between whole-word and decomposed lexical representation and processing in the network.

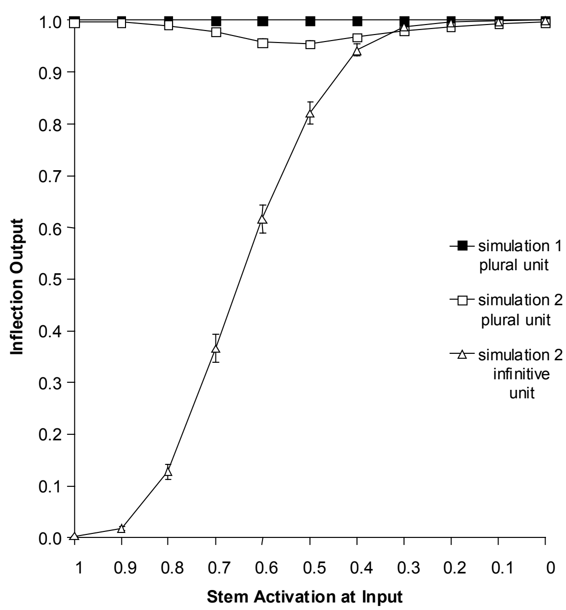
In Simulation 1, the network is trained to identify nouns with and without the -en plural ending. Since this affix has only a single interpretation, the network can activate the plural output solely based on the presence or absence of the -en ending. In Simulation 2, however, the -en ending has a different interpretation for nouns and verbs (i.e. two different output units may be activated). In order to activate the appropriate output unit for affixed forms, the network must identify the stem, since this determines the correct interpretation of the -en affix. The network in Simulation 2 must combine information from the stem and affix, and this suggests that the identification of the affix is 'non-componential' in his network; that is, morphologically complex words are not being processed as two independent components. We will report two analyses that allow us to assess in a quantative fashion, the componentiality of the networks' processing and internal representations – boundary analysis and contribution correlation[9].

One technique for assessing the componentiality of the networks in our two simulations is to observe the processing of the affix as information concerning the stem is removed from the input. In a truly componential system the identification of the affix would be essentially unaffected by the presence or absence of information concerning the stem, whereas for a non-componential system the processing of the affix would be affected by information from the stem. Similar analyses of the mappings learnt by distributed networks ('boundary analyses') have been reported for models of reading aloud by Plaut et al. (1996) and for models of morphological processing by Rueckl and Raveh (1999). We will apply this technique to the two simulations reported here as a means of comparing the componentiality of the affix in the two systems.

For this boundary analysis we tested networks from the two simulations on each of the 182 test items from the Baayen, Dijkstra, and Schreuder (1997) experiment in non-cascaded mode. For each item, we recorded the network's output for input patterns in which the activation of the orthographic units for the stem was reduced from 1.0 to 0.9, 0.8, etc. Changes in the activation of the affix output as the stem input is reduced are shown in Figure 8. As can be seen in the graph, reducing stem input had no effect on the activation of the plural unit in Simulation 1; the plural output remains fully active even with no stem input. However in Simulation 2, reducing the input to the stem produces a decrease in the activation of the plural unit. Furthermore, as information on the stem is lost, the network begins to activate both the verb infinitive and the noun plural output units. Without the information provided by the stem, the network is unable to correctly identify the form of the affix that is presented at the input, and produces an ambiguous output. This ambiguity will affect the output stress measure during cascaded processing and therefore slow down the network's lexical decision responses for plurals.

This inability to identify an inflectional affix separately from its stem illustrates the non-compositional processing of the -en affix, forced by its homonymy. When information to identify the stem is absent in Simulation 2, the best that the network can do is to activate both interpretations of the affix. In order to activate the correct output unit, the network had to learn which stems that take the -en affix are nouns and which are verbs. This can only be learnt from experience of the inflected form of each stem (unmarked noun and verb stems are processed in the same way by the network). For the network to resolve the ambiguity of the -en affix therefore requires training on specific inflected forms and will be affected by the word-form frequency of the inflected form.

Figure 8 Boundary analysis comparing activation of the affix units in simulation 1 and 2 as the orthographic input to the stem is progressively reduced. Error bars show one standard error of the mean over the test items.



The effect of the ambiguity of the -en affix can also be observed by investigating the representations that the network develops in the hidden units for the noun stems and affixes. One index of componentiality in these representations can be derived using the 'contribution correlation' technique developed by Plaut et al., (1996), and used by Rueckl and Raveh (1999) for assessing morphological representations. This analysis measures the similarity of the hidden-unit representation of a particular morphological unit both in and out of context. An item that is represented componentially will produce a very similar pattern of hidden unit representations in a particular context and when presented out of context. For example, if an affix is represented in the same way when presented to the network without a stem as when presented with a stem then we might consider this affix to be represented componentially.

To calculate the contribution correlation for the -en affix, we took the hidden-unit representation of an inflected form (e.g. 'ambten') and subtracting the representation of the stem ('ambt') when presented without input at the affix units. The resulting vector provides a representation of the -en affix in the context of the stem 'ambt'. We then calculated the hidden-unit representation of the affix outside of any context by taking the hidden activation for the -en affix alone (i.e. when presented without a stem) and subtracting the hidden unit response to a blank input. If the hidden unit representations of the affix in and out of context are identical or directly proportional to each other, then the affix is represented consistently in and out of context, and can be considered to be componential. If the two vectors are orthogonal (i.e. dissimilar) then the representation of the affix depends on the context in which it is presented and is therefore non-componential. The similarity of the two vectors is assessed by calculating the correlation coefficient between the corresponding components of the two vectors. The components of similar vectors will be highly correlated, with a correlation coefficient approaching one, while two orthogonal vectors will return a similarity tending towards zero.

The average correlation calculated for the -en affix averaged over all 91 test lemmas was 0.23 for Simulation 1, and 0.19 for Simulation 2, these values differed significantly ($t(90)=26.60$, $p<.001$), indicating that the -en affix was represented in a more componential fashion in Simulation 1 when there was only a single interpretation of the affix than in Simulation 2 when two interpretations are possible.

The same procedure was also used to calculate correlation coefficients for the stems. Since there are two independent contexts for each stem (with and without the -en ending), hidden unit representations of the stem in these two contexts can be compared directly (rather than using the subtraction with a null context as for the affix). When the contribution correlation is calculated in this way, the average contribution correlation for the 91 test stems on which the network was trained was 0.88 for Simulation 1 and 0.89 for Simulation 2. The first thing to note is that these values are substantially higher for the stem than for the affix – the stem has a more consistent hidden unit representation in the network than the affix. This indicates that activity in the hidden units is affected more by changes to the stem than by changes to the affix (perhaps unsurprising given the greater number of input and output units that represent the stem). Furthermore, although small, the difference between the contribution correlation for the stem in the two simulations is statistically reliable ($t(90)=2.03$, $p<.05$), thus although the hidden-unit representation of the affix is more componential in Simulation 1 than in Simulation 2, this is not also the case for the representation of the stem.

This finding reflects an important property of the mapping learnt by the network – that is, despite differences in the processing of the affix, the orthography-to-semantics mapping for the stem is consistent for singular and plural nouns in both simulations. For this reason, training experience with inflected forms can still benefit the network when presented with a bare stem. In comparing the two simulations, it does appear that the word-form frequency effect for singular nouns is larger in Simulation 1 than in Simulation 2. The exact cause of this change between the two simulations is unclear. This may be a side-effect of the greater training required for the network in Simulation 2 to reach criterion, which is itself a consequence of the ambiguity of the -en affix. Alternatively it may be that some further reorganisation of the internal representations for noun stems is promoted by the presence of an ambiguous affix.

4.2. Storage and decomposition in morphological processing

The two simulations that we have reported here have demonstrated that a single-mechanism distributed connectionist model can display a complex pattern of behaviour in responding to singular and plural nouns of varying frequencies. Our simulations demonstrate the success of models that do not include the distinct mechanisms of whole-word storage and morphological decomposition that are required in traditional localist accounts. Despite lacking explicit procedures for decomposition (segmentation, licensing, checking) as proposed by Schreuder and Baayen (1995), we observe componential processing of stems and affixes in Simulation 1. Nonetheless, despite this componentiality we still observe effects of word-form frequency in this simulation.

The introduction of a homonymous affix in Simulation 2 further discourages the network from processing inflected items compositionally (i.e. identifying the affix separately from the stem). Non-compositional processing of the plural affix seems necessary for the network to simulate experimental data from Dutch plurals reported by Baayen, Dijkstra, and Schreuder (1997). Our simulations therefore support the observation of Bertram, Schreuder and Baayen (2000) that affix homonymy affects the processing of morphologically complex words, and illustrates how differences in the structure of the linguistic environment can impact on lexical processing.

Distributed models in other domains have demonstrated the non-necessity of qualitative processing distinctions between rule-governed and exceptional forms in the generation of regular and irregular pronunciations in reading aloud (Seidenberg and McClelland, 1989; Plaut et al., 1996). We similarly propose that distinct processing mechanisms for whole-word storage and morphological decomposition need not be included in models of lexical processing. Our simulations are therefore in broad agreement with other authors (e.g. Ruckel and Raveh 1999; Plaut and Gonnerman 2000) in suggesting a distributed connectionist alternative to localist accounts of morphological processing. Fundamental to these distributed accounts is the suggestion that although terms like 'storage' and 'decomposition' may be used to describe the behavioural profiles observed in experimental investigations, that these terms need not reflect either explicitly-implemented computational procedures or separable components of the human lexical processing system.

Despite our arguments in favour of this single-mechanism, distributed connectionist model, there is always more than one way in which to simulate any given set of behavioural data. Baayen, Dijkstra, and Schreuder (1997) report mathematical simulations of a localist race model which achieves a more precise numerical simulation of experimental data on the processing of singular and plural nouns. Clearly the various styles of model have different strengths and weaknesses – in particular it is often difficult to use a distributed model to achieve precise numerical fits to empirical data. However, there may still be reasons to favour distributed models. For instance, they provide a basis for understanding how a system could self-organise in order to produce a particular form of functional organisation, rather than requiring external design to produce an implemented model.

Further research will be required if we are to adjudicate between localist and distributed accounts of lexical processing. From a behavioural perspective, one testable prediction of our distributed account would be that since competition between verb and noun interpretations of the -en affix play an important role in the behaviour of the network, that other affects of competition may also be observed experimentally (in priming studies, for example). Alternatively, linguistic or experimental contexts that favour one interpretation of otherwise ambiguous affixes might reduce competition, thereby shifting the balance between word-form and lemma frequency effects (cf. Bertram, Hyönä and Laine, 2000).

Despite these suggestions for future behavioural research, it is likely that an extended research programme, including neuropsychology, neuroimaging, and a consideration of the underlying neurophysiology will do more to advance the debate between localist and distributed accounts of lexical processing. On the basis of the current work though, we encourage researchers to be cautious before assuming that patterns of behavioural data that suggest the 'storage' or 'decomposition' of morphologically complex words can be mapped in a one-to-one fashion onto underlying computational mechanisms.

Notes

* We would like to thank members of the Language group at the Cognition and Brain Sciences Unit, notably Mike Ford, Dennis Norris and Tim Rogers for useful discussions of this work. We would like to thank Harold Baayen and Jay Rueckl for comments and suggestions on a previous draft of this manuscript. Email should be addressed to matt.davis@mrc-cbu.cam.ac.uk. This work was supported by the Medical Research Council of the United Kingdom.

1. For a general introduction to distributed connectionist modelling, the interested reader is directed to MacLeod, Plunkett, and Rolls (1998) or Bechtel and Abrahams (1991). For a discussion of distributed connectionist models of linguistic processing see Christiansen and Chater (2001) or Plaut and Gonnerman (2000). For a critical appraisal of distributed and localist models of cognition see Page (2000) and associated commentaries.
2. In English, the lemma frequency of a noun refers to the summed frequency of singular and plural forms. Other languages, such as Dutch, also include a diminutive inflection and lemma frequencies are therefore not a simple sum of the singular and plural frequency of the noun.

3. In later work, Pinker (1999) has revised his account to include some storage of high-frequency inflected forms. However, for clarity of exposition we will focus on an early version of this account in which all regularly inflected items are decomposed.
4. Schreuder and Baayen (1995) describe the decomposition process as involving three components, segmentation (splitting the word into stem and affix), licensing (confirming that syntactic constraints allow the stem and affix to be combined) and checking (determining that the resulting combination is semantically interpretable). It is possible that decomposition may be particularly slowed by one or more aspect of this process.
5. Output stress is as follows (where a_j is output activation for unit j):
   $$\text{Stress} = \Sigma a_j \log2(a_j) + (1- a_j) \log2(1- a_j) + 1$$
6. Cascaded activation function for hidden and output units is calculated using the following equation, where τ is the time-constant for integration, net-in(t) is the weighted sum of the input from the previous layer at time t (current) or t-1 (previous time step):
   $$\text{activation (t)} = \frac{1}{1 + e^{-(\tau\, \text{net-in(t)} + (1 - \tau)\, \text{net-in(t - 1)})}}$$
7. The average magnitude of the hidden-to-semantic weights is negative because there are only 10 out of 128 semantic units active for any given input.
8. This comparison is an extreme form of the typical manipulation of lemma frequency used experimentally; the verb stems we are using for comparison do not have any morphologically related forms – rather than having items with a low frequency relative. Nonetheless the results of this comparison still indicate that response times in the network are not only determined by word-form frequency.
9. We would like to thank Jay Rueckl for suggesting these analyses.

References

Baayen, R. H., Dijkstra, T., and Schreuder, R.
1997 Singulars and plurals in Dutch: Evidence for a parallel dual-route modes. *Journal of Memory and Language,* 37(1), 94-117.

Baayen, R. H., Pipenbrock, R., and Guilikers, L.
1995 *The CELEX Lexical Database (CD-ROM)* (Version v2.5). Philadelphia, PA: Linguistic Data Consortium,University of Pennsylvania.

Bechtel, W., and Abrahamsen, A.
1991 *Connectionism and the Mind*. Oxford: Blackwell.

Bertram R, Hyönä J, Laine M
2000 The role of context in morphological processing: Evidence from Finnish. *Language and Cognitive Processes,* 15(4/5), 367-388.

Bertram, R., Laine, M., Baayen, R. H., Schreuder, R., and Hyona, J.
2000 Affixal Homonymy triggers full-form storage, even with inflected words, even in a morphologically rich language. *Cognition,* 74(2), B13-B25.

Bertram, R., Laine, M., and Karvinen, K.
1999 The interplay of word formation type, affixal homonymy, and productivity in lexical processing: Evidence from a morphologically rich language. *Journal of Psycholinguistic Research*, 28(3), 213-226.

Bertram, R., Schreuder, R., and Baayen, R. H.
2000 The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology-Learning Memory and Cognition*, 26(2), 489-511.

Besner, D., Twilley, L., McCann, R. S., and Seergobin, K.
1990 On the connection between connectionism and data: are a few words necessary? *Psychological Review*, 97, 432-446.

Borowsky, R., and Masson, M. E. J.
1996 Semantic ambiguity effects in word identification. *Journal of Experimental Psychology-Learning Memory and Cognition*, 22(1), 63-85.

Broadbent, D. E.
1967 Word-Frequency effect and response bias. *Psychological Review*, 74(1), 1-15.

Bullinaria, J. A.
1995 Neural networks learning from ambiguous training data. *Connection Science*, 7(2), 99-122.

Bullinaria, J. A., and Chater, N.
1995 Connectionist modelling: Implications for cognitive neuropsychology. *Language and Cognitive Processes*, 10, 227-264.

Caramazza, A., Laudanna, A., and Romani, C.
1988 Lexical access and inflectional morphology. *Cognition*, 28(3), 297-332.

Caramazza, A., Miceli, G., Silveri, M. C., and Laudanna, A.
1985 Reading mechanisms and the organisation of the lexicon: Evidence from acquired dyslexia. *Cognitive Neuropsychology*, 2(1), 81-114.

Christiansen, M. H., and Chater, N.
2001 Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, 5, 82-88.

Cohen, J. D., Dunbar, K., and McClelland, J. L.
1990 On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97(3), 332-361.

Davis, M. H., Marslen-Wilson, W. D., and Hare, M.
1996 Representing regularity: The English past tense, in G. W. Cottrell (Ed), *Proceedings of the 18th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Elman, J., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K.
1996 *Rethinking Innateness*. Camb, MA: MIT Press.

Ford, M. A., Marslen-Wilson, W. D. and Davis, M. H.
2003 Morphology and frequency: Contrasting methodologies. In R. H. Baayen (Ed.), *Aspects of Morphological Processing*, Berlin: Mouton de Gruyter

Forster, K. I.
1976 Accessing the mental lexicon. In R. J. Wales and E. Walker (Eds.), *New Approaches to Language Mechanisms* (pp. 257-287). Amsterdam: North-Holland.

Frauenfelder, U. H., and Schreuder, R.
1991 Constraining psycholinguistic models of morphological processing and representation: the role of productivity. In G. Booij and J. v. Marle (Eds.), *Yearbook of Morphology* (pp. 165-183).

Gaskell, M. G., and Marslen-Wilson, W. D.
1997 Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes*, 12, 613-656.

Gasser, M.
1994 Acquiring receptive morphology: a connectionist model. *Proceedings of the International Conference on Computational Linguistics*, 15, 214-220.

Hinton, G. E.
1989 Connectionist learning procedures. *Artificial Intelligence*, 40, 185-234.

Howes, D. H., and Solomon, R. L.
1951 Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 41, 401-410.

MacLeod, P., Plunkett, K., and Rolls, E. T.
1998 *Introduction to connectionist modelling of cognitive processes*. Oxford, UK: Oxford University Press.

McClelland, J. L.
1979 On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86(4), 287-330.

McClelland, J. L., and Rumelhart, D. E.
1981 An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88(375-407).

Morton, J.
1969 Interaction of information in word recognition. *Psychological Review*, 76(2), 165-178.

Page, M.
2000 Connectionist modelling in psychology: A localist manifesto. *Behavioural and Brain Sciences*, 23(4), 443-512.

Pinker, S.
1991 Rules of language. *Science*, 253(August), 530-535.

Pinker, S.
1999 *Words and Rules: The Ingredients of Language*. New York: Basic Books.

Plaut, D. C.
1997 Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12(5-6), 765-805.

Plaut, D. C., and Gonnerman, L. M.
2000 Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, 15(4-5), 445-485.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K.
1996 Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56-115.

Plunkett, K., and Marchman, V.
1991 U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38, 43-102.

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

Rodd, J. M., Gaskell, M. G., and Marslen-Wilson, W. D.
2002 Making sense out of ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 245-266.

Rueckl, J. G., and Raveh, M.
1999 The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language*, 68(1-2), 110-117.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.
1986 Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1). Cambridge, MA: MIT Press.

Rumelhart, D. E., and McClelland, J. L.
1986 On learning the past tenses of English verbs. In J. L. McClelland and D. E. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2). Cambridge: MA: MIT Press.

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

Seidenberg, M. S., and McClelland, J. L.
1989 A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.

Sereno, J. A., and Jongman, A.
1997 Processing of English inflectional morphology. *Memory and Cognition*, 25(4), 425-437.

Taft, M.
1979 Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7(4), 263-272.

Taraban, R., and McClelland, J. L.
1987 Conspiracy effects in word pronunciation. *Journal of Memory and Language*, 26(6), 608-631.