

Assessing the processing consequences of segment reduction in Dutch with naive discriminative learning

R. H. Baayen

University of Alberta, Edmonton

Abstract

This study addresses the comprehension of reduced words, taking as point of departure two lexical decision experiments reported in Ernestus (2009). Ernestus discusses the consequences of segment reduction in auditory comprehension in terms of exemplars for reduced forms and generalization processes reconstructing the unreduced form. A different approach is explored in the present study, using a computational model based on discriminative learning to explain the pattern of results in the experimental data. This new modeling approach, which provides the researcher with detailed information into what distributional properties of the language input may drive the observed effects, suggests that the unusual biphones in reduced words are the key to understanding why reduced words can be learned and understood.

In a recent study, Ernestus (2009) investigated the processing of Dutch past participles (e.g., *ge-vraag-d*, ‘asked’), which are formed by simultaneously prefixing *ge-* ([xə]) and suffixing *-d* or *-t*. In Dutch, the schwa in unstressed prefixes such as *ge-* is often omitted. Ernestus presented speakers of Dutch with the past participles of phonotactically legal new monomorphemic verbs. In a familiarization phase, these past participles were combined with pictures in order to ensure that the nonce verbs received an interpretation. A week later, participants were asked to complete an auditory lexical decision experiment in which these past participles were presented. One question addressed in this study was whether reduced forms acquire their own exemplar representations (c.f., e.g., Johnson, 2004) facilitating subsequent processing. A second question was whether the privileged status of unreduced forms (Ernestus, Baayen, & Schreuder, 2002; Gaskell, 2003) would affect the speed of comprehension.

In Experiment 1, participants were exposed either to the reduced or to the full form of the new participles during familiarization, and later performed a lexical decision on the reduced forms. These two conditions will henceforth be referred to as R+R (reduced for training and reduced for testing) and as U+R (unreduced for training, reduced for testing).

I am indebted to Mirjam Ernestus for making the words she used in her experiments available to me, and to both Mirjam Ernestus and Vito Pirelli for their insightful comments on an earlier version of this paper.

Response latencies were significantly shorter for the R+R condition (1271 ms) compared to the U+R condition (1319 ms). In theories framed in terms of form representations, the results of Experiment 1 can be interpreted as evidence for the coming into existence of form representations for reduced words after familiarization. When subsequently encountered in the lexical decision task presenting reduced forms as targets, the newly-formed representations of reduced forms would then provide a better match to the acoustic input, and therefore would give rise to shorter latencies compared to the unreduced words, which would have form representations that do not fully match the input.

Experiment 2 used the same familiarization procedure, but presented unreduced forms (instead of reduced forms) as targets in the lexical decision experiment. The mean latencies for these R+U (reduced training, unreduced testing) and U+U (unreduced training, unreduced testing) conditions were very similar (1328 ms and 1330 ms respectively). Ernestus argued that when listeners encounter a new reduced form, the unreduced form is also reconstructed from the reduced form. Hence, when the unreduced form is presented in lexical decision, it matches a representation irrespective of whether an unreduced or a reduced form was presented during familiarization. As a result, average response latencies for the U+R and U+U conditions are indistinguishable.

This interpretation raises several questions. First, if two representations come into existence upon encountering a novel reduced form, some theories predict competition between these representations (e.g., Luce & Pisoni, 1998). In the R+R and R+U conditions, in which representations for both the reduced and unreduced forms are supposedly available after familiarization, a processing delay would be expected. No such delay is present in the experimental data, however. In fact, the R+R condition leads to shorter instead of longer response latencies than the U+R condition.

Second, the reconstruction of the unreduced form from the reduced form is supposedly driven by a generalization giving canonical status to the unreduced form. In hybrid models combining rules and exemplars (see, e.g., Atallah, Frank, & O'Reilly, 2004; Goldinger, 2007), large-scale generalizations are supposed to be early processes, whereas episodic, exemplar-driven generalization is supposed to be a late process, manifesting itself primarily in elongated processing times. However, when the two experiments of Ernestus (2009) are considered jointly, it is the R+R condition, the condition in which slow exemplar-driven processing should be most prominently involved, for which we observe the shortest latencies, instead of the longest latencies.

Third, positing representations for both unreduced and reduced forms comes with the risk of a proliferation of representations, given the high degrees of variability characterizing speech.

Finally, whereas modeling the processing of acoustic information as an early process and the processing of speaker-specific episodic information as a late process (Goldinger, 2007) may have its advantages. Information about, for instance, a speaker's voice plays no role in phonological or phonetic generalizations about the characteristics discriminating one word from the other words in the lexicon. Yet episodic information about a speaker's voice may co-determine lexical processing at, if Goldinger is correct, later stages in the comprehension process. The present experimental data, however, concern specific and general information that is qualitatively quite similar. Instead of episodic information about a speaker's voice in conjunction with segmental information, the data compare the presence

versus absence of a segment in a lexical representation. Qualitatively similar differences permeate the lexicon, for any pair of words that differ in the presence versus the absence of a segment (e.g., *hand*, *hands* and *hand*, *had*. Although it is possible that the exemplar-driven and generalization-based processes posited to explain the present experimental data are taking place at different sites and at different points in time, this possibility becomes less attractive when both processes concern the same kind of segmental information.

In this study, therefore, a very different approach to understanding these experimental data is pursued, using a computational model first proposed in Baayen, Milin, Filipovic Durjjevic, Hendrix, and Marelli (2010). This model makes use of naive discriminative learning based on the Rescorla-Wagner equations (Wagner & Rescorla, 1972; Danks, 2003), which are well-established in psychology as a mathematical description of learning. In what follows, a brief introduction to this computational model is presented first. Next, the model is illustrated by pitting its predictions for Dutch inflected verb forms against the by-item mean latencies available in the Dutch Lexicon Project, henceforth DLP (Keuleers, 2010). We then zoom in on the Dutch past participle, and the relative importance of prefix and verb stem for lexical access as measured by the lexical decision task. Finally, simulations are presented clarifying that the pattern of results observed by (Ernestus, 2009) follows straightforwardly from discriminative learning.

Naive discriminative learning

The model developed by Baayen et al. (2010) is a two-layer network with symbolic representations for a word’s form and a word’s semantics. A word’s form is coded by means of its unigrams and bigrams (or uniphones and biphones). Its meaning is represented in terms of the semantic units associated with its constituents. Thus, the word *bookcases* is linked with the meanings BOOK, CASE, PLURAL. Each input unit is linked to each meaning unit, and a weight is associated with each link.

When a word is read (or heard), the activations of its unigrams and bigrams are set to 1, and those of all other unigrams and bigrams to 0. Activation is then propagated through the connections. The activation a_i of a meaning i is defined as the sum of the weights on its active incoming links:

$$a_i = \sum_j w_j, \quad (1)$$

where j ranges over all unigrams and bigrams with nonzero activation. The simplest way to estimate the response latency to word i is to assume it is inversely proportional to the activation a_i of that word’s meaning:

$$\text{RT}_i = \log(1/a_i), \quad (2)$$

where we take the logarithm to remove the rightward skew from the distribution of activations. Slightly improved results are often obtained by first calculating the probability of identification P_i of the word’s meaning in the set of its k highest-activated competitors:

$$P_i = \frac{a_i}{a_i + \sum_{j=1}^k a_j}. \quad (3)$$

Reaction times are then estimated using

$$\text{RT}_i = \log(1/P_i). \quad (4)$$

The weights in (1) are estimated by solving the equilibrium equations of the Rescorla-Wagner model developed by Danks (2003). The resulting estimates of the activation of a word's meaning and its probability of identification are a kind of maximum likelihood estimates. They provide an indication of how likely a word's meaning is given the unigrams and bigrams in its input and, importantly, given the conditional co-occurrence probabilities of the unigrams and bigrams as well as the conditional probabilities of a meaning given a specific unigram or bigram.

The model is grounded in *discriminative* learning in the sense that the weights on the links from a given cue to the meanings are optimized for discriminating between these meanings on the basis of this cue.

The model uses *naive* discriminative learning, naive in the sense of naive Bayes classifiers, in that the incoming weights for a given meaning are estimated independently from all other meanings. For further details on the Rescorla-Wagner model and its applications in psychology, the reader is referred to Miller, Barnet, and Grahame (1995); Siegel and Allan (1996); Chater, Tenenbaum, and Yuille (2006); Hsu, Chater, and Vitányi (2010); Ramscar, Yarlett, Dye, Denny, and Thorpe (2010).

Baayen et al. (2010) show that for English simple words, the model generates simulated processing latencies that correlate significantly at the item level with observed latencies as available in the English Lexicon Project (Balota et al., 2007), with correlations up to 0.5. The model also correctly reproduces the priming results of Rastle, Davis, and New (2004), who observed significant priming for *corner-corn* and *dealer-deal*, but not for *brothel-broth*. In addition, the model simulates well the data of Bergen (2004) on phonaesthemes (significant priming for *glimmer-gleam* compared to semantic and orthographic control conditions). Furthermore, the model correctly predicts shorter processing latencies for words with greater morphological families, irrespective of whether these words occur by themselves, or are embedded in derived words or compounds.

The model is also highly sensitive to co-occurrence patterns of words. When applied to Serbian case paradigms, the model correctly predicts inhibition from relative entropy (Milin, Filipović Durdević, & Moscoso del Prado Martín, 2009). The model similarly correctly predicts an inhibitory effect of relative entropy for English prepositional 'paradigms'. The more the relative frequencies with which a given noun co-occurs with spatial prepositions differ from the relative frequencies with which these prepositions are used (aggregating across all nouns), the longer response latencies to that noun are. Both the Serbian case paradigms and the English prepositional paradigms illustrate an exemplar-prototype effect, such that the more an exemplar differs from the prototype, the greater its processing costs are.

The discriminative learning model indicates that current behavioral data can be understood without special morpho-orthographic parsing operations, without positing representations for (transparent) complex words and multi-word phrases, and without positing that complex words are organized into paradigms. The next section shows that the model is also successful when applied to Dutch inflected verb forms.

Naive discriminative learning of Dutch verb forms

From the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995), all inflected verb forms with non-zero frequency were extracted, subject to the conditions that

the base verb was either monomorphemic, or preceded by a derivational prefix. This resulted in an input lexicon for the model comprising 11608 different word forms, representing 5783141 tokens of monomorphemic verbs (8975 inflectional types) as well as verbs with the prefixes *be-*, *de-*, *des-*, *ex-*, *ge-*, *her-*, *ont-*, *over-*, *pre-*, *re-*, and *ver-* (2633 inflectional types). The total number of different verb stems, and hence the different number of lexical meanings in the model, was 2063.

In order to compare the predictions of the model with actual processing times, we consulted the Dutch Lexicon Project (<http://crr.ugent.be/dlp>), which offers lexical decision latencies for large numbers of Dutch words. For 3368 verb forms in the training lexicon, reaction times are available. Derivational prefixes in the resulting smaller data set turned out to be restricted to *be*, *her*, *ont*, *ver* (276 word forms). The new data set also included 194 different past participles.

The correlation between the observed RTs and the parameter-free simulated RTs was $r = 0.31$ ($t(3364) = 18.6$, $p = 0$). The corresponding correlation for the probability of identification-based simulated RTs (with one free parameter, the number of neighbors taken into consideration, fixed at 20 as in Baayen et al. 2010), was slightly higher at $r = 0.32$ ($t(3364) = 19.57$, $p = 0$).

Although the model is a full-decomposition model taking an input such as *gewandeld* and mapping it onto the meanings WANDEL (‘to walk’) and PAST, PERFECT, it is highly sensitive to the frequencies of the exemplars in its input space. Hence, frequency effects are expected to arise not only for the verb’s lemma (collapsing the frequencies of all its inflectional variants), but also for the frequency of the specific word form that is presented to the model.

As word form frequency and lemma frequency are strongly correlated, we orthogonalized form frequency on lemma frequency by taking the residuals of a model regressing form frequency on lemma frequency. The correlation of the residualized form frequency measure (which represents word form frequency in as far as it cannot be predicted from lemma frequency) with the original form frequency count was 0.63, indicating that the orthogonalized measure is still well-interpretable as a form frequency measure.

RTs	Predictor	Estimate	Std. Error	t-value	p-value
simulated without competitors	Intercept	6.7740	0.0575	117.8723	0.0000
simulated without competitors	Lemma Freq.	-0.6268	0.0077	-81.8879	0.0000
simulated without competitors	Form Freq.	-0.0468	0.0132	-3.5518	0.0004
simulated with competitors	Intercept	7.8599	0.0543	144.6957	0.0000
simulated with competitors	Lemma Freq.	-0.6825	0.0072	-94.3324	0.0000
simulated with competitors	Form Freq.	-0.0803	0.0124	-6.4481	0.0000
observed	Intercept	-1.4145	0.0086	-164.9991	0.0000
observed	Lemma Freq.	-0.0219	0.0011	-19.1744	0.0000
observed	Form Freq.	-0.0168	0.0020	-8.5493	0.0000

Table 1: Lemma and Form Frequency effects for fitted and observed reaction times. Observed RTs were inverse transformed by $-1000/RT$.

Table 1 reports linear regression models fitted to the observed RTs (inverse transformed by $-1000/RT$ to remove the skew from the latency distribution) and to the two

sets of simulated RTs. For all three models, we observe significant effects of both lemma frequency and of wordform frequency. This illustrates that in a model that is fully decompositional at the semantic level, but that takes into account in a principled way the different extents to which the unigrams and bigrams discriminate between meanings, effects of the frequency with which a specific form supports a given meaning can emerge. It is worth noting that, first, form frequency arise in the model without there being any form representations for complex words in the model. Second, form frequency effects arise without a word’s meaning having to be non-compositional. However, as can be seen by comparing the magnitude of the coefficients of the two frequencies listed in Table 1 for the observed and simulated RTs, the model underestimates the form frequency effect. This underestimation is due to the simplifying assumption that all complex words are completely transparent semantically.

Reduction and lexical access to past participles

Given that naive discriminative learning provides a reasonable approximation to the processing costs of Dutch inflected words, correctly predicting a form (exemplar) frequency effect side by side with a lemma frequency effect, we now use the model to study the effects of reduction on lexical access.

Figure 1 summarizes the effect of deleting a segment at positions 1 through 7 of the input word. The top left panel shows that deletion of the first and second segment (the segments of the prefix *ge*) hardly affects the activation of the lexical meaning of the verb. By contrast, as can be seen in the second panel, deletion of the first segment substantially decreases the activation of the affixal meaning (PAST PERFECT). Deletion of the second segment (the schwa) also has a detrimental effect, although smaller.

For segment deletions in the base (at positions 3–5), the pattern reverses. These deletions decrease the activation of the verbal meaning, but hardly affect the activation of the inflectional meaning. Deletion of later segments does not affect the activation of the verbal meaning as much as deletions at positions 3–5. This is because deletions at later positions tend to involve the deletion of the final dental of the Dutch past participle (*ge-wandel-d*, ‘walked’; *ge-pak-t*, ‘grasped’), thereby making the stem better identifiable. As the position of deletion increases, the likelihood that the final inflectional suffix is removed rather than a segment that is part of the verbal stem increases as well, leading to attenuation of the negative consequences of segment deletion.

The lower panel of Figure 1 graphs the consequences of segment deletion at different positions in the written form for the correlations of simulated RTs (calculated for input with the missing segment) and the observed visual lexical decisions in the Dutch Lexicon Project. In this way, we can gauge the importance of different segments for learning a word’s meaning. If a segment is removed from the model’s input, whereas that segment plays a crucial role in reading, the correlation of the model’s predicted RTs with the actually observed RTs should decrease. Such a decrease is visible in the lower panel of Figure 1 for segments of the base. The resulting pattern mirrors that in the upper left, indicating that the observed lexical decisions are driven primarily by accessing the meaning of the verb, and not by accessing the meaning of the inflectional circumfix (PAST PERFECT).

In summary, naive discriminative learning provides good predictions for lexical decision latencies to inflected Dutch verbs, with only a single free parameter, the number of

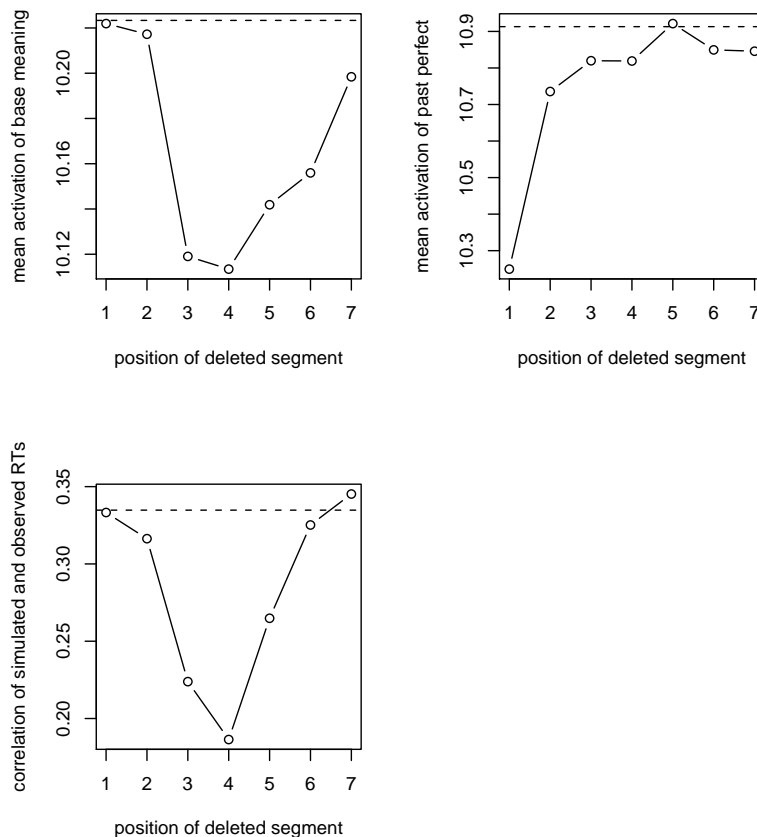


Figure 1. Effect of deletion of a segment at varying positions in the input. Top left: effect on mean activation of the meaning of the verb. Top right: effect on the mean activation of the meaning of the inflectional suffix. Bottom: Effect of segment deletion on the correlation of the simulated RTs with the observed RTs. Dashed horizontal lines represent the case with no segment deletion.

neighbors taken into consideration for calculating the probability of identification of a verb’s meaning. The model indicates that participants may have based their decisions mainly on the activation of the lexical meaning of the verb, ignoring the grammatical meaning of the inflectional circumfix. For a proper evaluation of the role of the inflected form, including its inflectional ending, processing measures gauging sentential reading will be essential.

Accessing participial neologisms

The experiments of Ernestus (2009) examined the effect of acoustic reduction on the learning and subsequent processing of novel verbs. Simulated lexical decision latencies for these experiments were obtained as follows. Starting from the input lexicon derived from CELEX, two new lexica were constructed. The first lexicon combined the input lexicon with the reduced experimental words. The second lexicon combined the input lexicon with the

unreduced experimental words. Each novel experimental word was assigned unit frequency, and was linked with a unique lexical meaning as well as with the grammatical meanings for PAST and PERFECT. From these lexicons, two weight matrices were derived, one for familiarization with the reduced neologisms, and one for familiarization with the unreduced neologisms. These two weight matrices represent the state of the participants after the familiarization phase, but before the lexical decision experiment was administered.

Model RTs	Experiment	Predictor	Estimate	Std. Error	t-value	p-value
without competitors	II	U+U (Intercept)	6.2087	0.2101	29.5470	0.0000
without competitors	II	R+U	-0.0000	0.2972	-0.0000	1.0000
without competitors	I	U+R	0.1327	0.2972	0.4467	0.6558
without competitors	I	R+R	-2.7713	0.4159	-6.6641	0.0000
with competitors	II	U+U (Intercept)	7.9820	0.3891	20.5128	0.0000
with competitors	II	R+U	-0.0000	0.5503	-0.0000	1.0000
with competitors	I	U+R	1.4142	0.5464	2.5885	0.0107
with competitors	I	R+R	-5.2587	0.7700	-6.8294	0.0000

Table 2: Effects of training and testing for simulated latencies. Coefficients represent intercept and treatment contrasts. The reference level for the factors Training and Testing was Unreduced

Next, each weight matrix was applied twice to calculate the activation of the lexical meanings, once for targets presented to the model in unreduced form, and once for targets presented in reduced form. Figure 2 plots the observed group means as reported by (Ernestus, 2009) (left panel), the group means for the simulated latencies without considering neighbors (center panel), and the group means when the 20 strongest competitors are taken into account (right panel). Table 2 summarizes the intercept and treatment contrasts for a linear regression model fitted to the simulated latencies obtained with and without the competitors included.

The best fit to the data is provided by the model that does not include neighbors in the calculation of the simulated RTs, as can be seen by comparing the first two panels of Figure 2. When neighbors are included, the group mean for the U+U condition is overestimated compared to the observed group mean. This difference between the observed and simulated latencies is open to two interpretations.

First, if a little noise, representing response execution, is added to the latencies simulated for Experiment 2, then the only significant effect in the model will be restricted to the advantage of R+R forms. Under this interpretation, both types of simulated latencies support the same interpretation.

Second, however, increasing the number of neighbors taken into account when estimating response latencies increases the average latency in the U+R condition. (It is in this condition, in which a shorter form that is more similar to a monomorphemic word is presented, that the summed activation of the highest-frequency neighbors is substantially higher than for the other conditions.) Since for existing words latencies simulated by including neighbors outperform latencies simulated without neighbors in terms of the correlation with the observed by-item RTs, it remains a possibility that participants in the actual experiments with neologisms did not take neighbor competitors into account when making their lexical decisions. Under this interpretation, the cumulative evidence for lexicality pro-

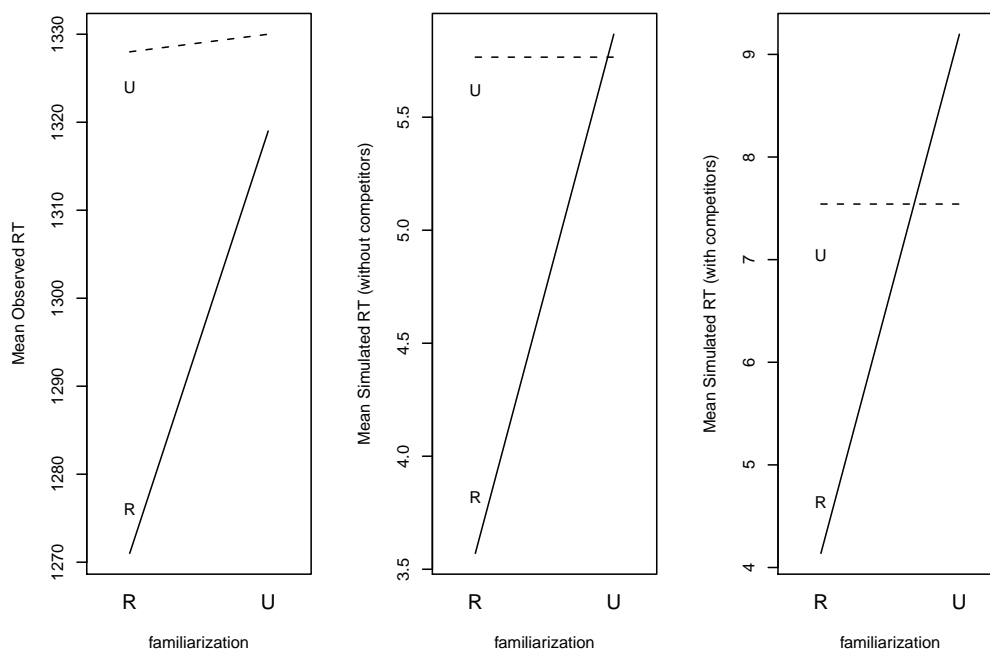


Figure 2. Mean reaction times for the four conditions of Familiarization by Testing with Unreduced (U) and Reduced (R) neologisms. Solid lines represent Experiment 1, dashed lines Experiment 2.

vided by lexical competitors is not a useful source of information for lexical decisions, as such cumulative evidence is similar for both the neologisms and nonwords. In this scenario, decisions would be based solely on the memory traces of a word’s meaning as presented during the familiarization phase.

Irrespective of which of these two interpretations is correct, it is clear that naive discriminative learning captures the pattern of results observed by Ernestus (2009). It is remarkable that the best fit to the data is obtained with a completely parameter-free model that is driven entirely by the frequential and distributional properties of the input.

Naive discriminative learning also makes it possible to trace the source of the processing advantage of R+R neologisms by examining which unigrams and bigrams contribute most to the activation of a neologism’s meaning. A boxplot summarizing the distribution of the activations of these meanings for the four conditions of the two experiments is shown in the upper left panel of Figure 3. The higher activations for the R+R condition correspond to the faster simulated latencies for this condition. The remaining panels visualize the contributions of different parts of the words to these activations.

The top second panel clarifies that the first segment, the velar fricative of the inflectional prefix, is not contributing at all to the activation of the lexical meaning of the neologism. The panel to its right shows that when the schwa of the prefix is presented (in conditions U+U and R+U), it does not contribute to the lexical meaning either. (In the U+R and R+R conditions, there is no schwa, so the contribution of the non-existent second

segment of the prefix is zero.)

The bottom left panel indicates that the initial bigram (the transition from silence into the initial segment of the prefix) makes a negative contribution to the meaning activations in the case of familiarization with the reduced form of the participle.

The next panel considers the boundary bigram for the transition of the prefix into the verb stem. For the unreduced participles presented at testing, this bigram begins with the schwa of the prefix, for the reduced participles, it begins with the velar fricative. When familiarized with the reduced form, presentation of the unreduced form makes the meaning of the word somewhat less likely.

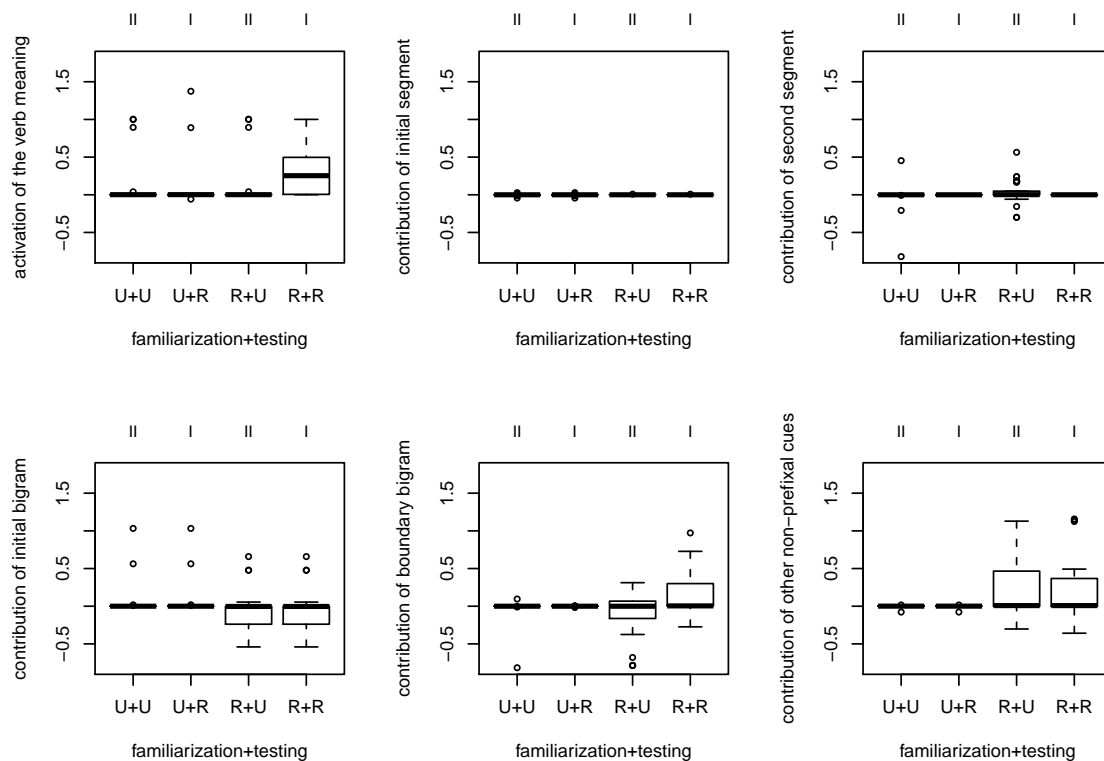


Figure 3. Activation of the verbal meaning, and contributions to this activation, for the combinations of Training and Testing with Unreduced (U) and Reduced (R) neologisms.

In the R+R condition, by contrast, we see that the boundary bigram contributes positively to the activation of the meaning of the neologism. Since the stimuli used by Ernestus (2009) were chosen such that removal of the schwa resulted in an initial bigram that does not occur word-internally in Dutch words, the only experience the model has with these bigrams is in the neologisms, allowing these bigrams to become excellent cues for discriminating between the meanings of the neologisms.

The final panel shows that unigrams and bigrams that are not associated with the prefix also contribute to the activation of the new meanings, but only when the neologisms were presented in reduced form. The reason is that the weights are estimated not only from

the probability of a meaning given a unigram or bigram, but also from the cooccurrence probabilities of the unigrams and bigrams. The boundary bigram, which is a rare and hence highly informative bigram for familiarization with the reduced form, therefore allows stem unigrams and bigrams to also become more strongly associated with the new lexical meanings. In summary, what this analysis indicates is that the boundary bigram of the reduced forms is the crucial factor driving the pattern of results.

As mentioned above, the boundary bigrams were explicitly selected by Ernestus to be phonotactically illegal in order to facilitate the reconstruction by the listener of the underlying form of the prefix as containing a schwa. The idea here is that unlikely boundary bigrams provide the segmentation process with an excellent cue about the location of the morphological boundary (cf. Hay, 2003). However, within the context of naive discriminative learning, exactly the low-probability bigrams create unique opportunities for discriminating between the meanings of a word. High-frequency boundary bigrams are typically used word-internally in many different words, and therefore provide weak cues for the meanings of these words. Low-frequency boundary bigrams, by contrast, are much more distinctive and hence more informative about which meaning is at issue.

In the simulation studies discussed thus far, the CELEX-derived lexicon contained only unreduced forms. To probe the consequences of the occurrence of reduced forms in natural speech for the model's estimates, new lexica were created in which, for each word form type with any of the prefixes *be-*, *ver-*, *ge-*, half of its tokens were reduced. The overall pattern of results is the same, with the R+R condition as the only condition emerging with a processing advantage. Interestingly, the weights on the links to a word's meaning for the initial bigram, as well as for the non-prefixal orthographic cues, are now very much reduced. The only contributor to the higher activation of word meanings for the R+R condition is the boundary bigram . This boundary bigram has become a less reliable cue for the meaning of a new participle, as it is now instantiated in many other words. The co-occurrence of this bigram with other unigrams and bigrams is now no longer unique or nearly unique to the neologisms. As a consequence, the indirect contribution of the boundary bigram through the initial bigram and the non-prefixal unigrams and bigrams becomes very small, and only the contribution of the boundary bigram itself remains to drive the group difference.

General Discussion

The experiments of Ernestus (2009) show that Dutch participles that are initially learned in reduced form and subsequently encountered in reduced form, have a processing advantage over participles that are either learned or encountered in unreduced form.

The explanation offered by Ernestus argues that during learning, reduced forms receive two representations, one exemplar representation for the reduced form, and one representation for the unreduced form reconstructed from the reduced form on the basis of the morphophonological regularities in the lexicon. Unreduced forms, by contrast, receive only one representation, namely, one for the word's full form. When a word is subsequently encountered in reduced form, a fully matching representation is available only for words learned earlier in reduced form. This matching exemplar would then afford faster processing compared to words previously learned in unreduced form. When a word is encountered in unreduced form, a matching representation is available, irrespective of whether the word was learned as reduced or unreduced, thanks to morphophonological generalization giving

rise to a full word representation also for words initially learned in reduced form. This would explain the very similar mean reaction times in Experiment 2.

The explanation pursued in the present study argues that the pattern of results observed by Ernestus can arise under discriminative learning. During learning, the reduced participles come with boundary bigrams that are attested in language use only when words with the prefix *ge-* are reduced. These boundary bigrams are not attested word-internally in monomorphemic (unreduced) forms. As a consequence, these bigrams provide excellent cues for discriminating the meaning of the new word during familiarization from the meanings of other words. When subsequently during testing a reduced form is presented, the effects of exemplar-like learning emerge straightforwardly. But unlike in the theory of Ernestus, which assumes separate exemplars for reduced forms side by side with representations for unreduced forms, the role of exemplar knowledge in the present discriminative learning approach is focused around the weights associated with the boundary bigram. No separate exemplar representations are required for reduced forms. Furthermore, it is not necessary to posit that speakers also reconstruct and store the unreduced representation when hearing a reduced pronunciation. In short, naive discriminative learning offers a simpler explanation of the data.

The present approach may help explain why some words display a wide array of reduced variants. An example from Dutch is *natuurlijk*, which can mean both ‘naturally’ and ‘of course’. In the latter sense, it is pronounced as [natyrlək, natylək, ntylək, ntyk, tyrlək, tylək, tylk], and [tuk]. In the light of the present modeling results, it is striking that in many of these forms biphones and triphones appear that are either rare or phonotactically illegal. For instance, [nt] in [ntylək, ntyk] does not occur word-initially, and [ylk] in [tylk] is not attested word-finally in normal Dutch words. Although [yk] is attested word-finally in other words, there are only three such words in the CELEX lexical database: *stuc*, *truc*, and *kaduuk*. Word-initial [ty] occurs only in *tuberculose*, *tumor*, *tumult*, *tuniek*, *tureluurs*, and *tutoyeren* and a few derived words and compounds derived from these words. The highest frequency word in this set is *truc* (13 per million), followed by *tumult* (5 per million). It seems, therefore, that the reduction variants of *natuurlijk* make use of segment combinations that are not frequent in the language. Precisely their low frequency (both in terms of types and tokens) makes them excellent cues for discriminative learning. The connection weights from these biphones and triphones to the meaning ‘of course’ will be relatively high. As a consequence, as soon as the reduced form has been heard and interpreted correctly, discriminative learning predicts it will be relatively easy to understand this form at subsequent encounters.

Previous research has established that heavily reduced forms are often incomprehensible out of context (Ernestus et al., 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004). It follows that the cue strength of the unusual diphones and triphones is not sufficient to fully activate the appropriate meaning out of context. Clearly, contextual information must be taken into account during learning. Interestingly, naive discriminative learning has been shown to be feasible when training proceeds on short phrases instead of isolated words. Moreover, exemplar-versus-prototype effects at the phrasal level (for prepositional paradigms in English) have been modeled successfully in this framework (see Baayen et al., 2010). Crucially, these phrasal paradigmatic effects, as well as phrasal frequency effects (cf., e.g. Tremblay & Baayen, 2010), arise in a model that has no explicit representations

for phrases. The present model therefore promises to be able to learn to link phrases with strongly reduced words to the correct meanings, without having to assume different representations for the different segmental realizations in these phrases. If this promise is indeed realized, the long-standing riddle of why reduced words are so context-dependent for their interpretation will have been solved. At the core, the answer to the riddle will be the same as for why *arp* is unintelligible without the context *c_et*.

This approach to the comprehension of reduced words raises the question of whether reduced forms are driven by listener-related or speaker-related processing constraints. It seems most likely that the primary constraint originates with the speaker minimizing articulatory effort. Consider, for instance, the reduced and unreduced forms of the past participle of [platsə], ‘to place’. The reduced form, [xplatst] will tend to have a shorter acoustic duration than the unreduced form, [xəplatst] (‘placed’). The principle of least effort will therefore favor the production of the reduced form above the unreduced form. Here, it is assumed that in this case speed of articulatory production takes precedence over ease of articulation of an unfamiliar bigram ([xp]).

However, the speaker cannot reduce at random. Dropping the [p] in [prat] (‘talk’) will result in a different word (‘honey comb’) and confusion on the part of the listener. Unsurprisingly, the reduction processes considered in the present study never lead to mergers with existing words. Interestingly, although reduction processes probably originate with the speaker, the hypothesis proposed here is that reduced forms will be more successful to become established alternative pronunciation variants if they are easy to learn and understand for the listener. From this perspective, the emergence of reduced forms is the result of synergetic processes involving both the speaker and the listener, with ease of articulation on the part of the speaker being, over time, reinforced by the sharper acoustic distinctness of the reduced output. The speaker, for ease of articulation, produces variants with low-probability segment sequences. These low-probability segment sequences in turn are easy for the listener to learn as pronunciation variants. In other words, established reduced forms, such as [ntyk] for [natyrlək] are likely to be local optima satisfying both the minimization of the speaker’s production effort and the minimization of the listener’s learning effort.

References

- Atallah, H. E., Frank, M. J., & O’Reilly, R. (2004). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Neurobiology of Learning and Memory*, 82, 253–267.
- Baayen, R. H., Milin, P., Filipovic Durdjevic, D., Hendrix, P., & Marelli, M. (2010). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Submitted*.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (cd-rom)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445–459.
- Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 80, 290–311.
- Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Science*, 10(7), 287–291.
- Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47(2), 109–121.

- Ernestus, M. (2009). The roles of reconstruction and lexical storage in the comprehension of regular pronunciation variants. In *Interspeech 2009* (pp. 1875–1878). Brighton.
- Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*, 162–173.
- Gaskell, M. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, *31*(3–4), 447–463.
- Goldinger, S. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 49–54). Saarbruecken.
- Hay, J. B. (2003). *Causes and Consequences of Word Structure*. New York and London: Routledge.
- Hsu, A. S., Chater, N., & Vitányi, P. (2010). The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Manuscript submitted for publication*.
- Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: data and analysis. proceedings of the 1st session of the 10th international symposium* (p. 29–54). Tokyo, Japan.
- Kemps, R., Ernestus, M., Schreuder, R., & Baayen, R. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *19*, 117–127.
- Keuleers, E. (2010). *The Dutch Lexicon Project*. Available from <http://crr.ugent.be/dlp>
- Luce, P., & Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1–36.
- Milin, P., Filipović Durdević, D., & Moscoso del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, *50*–64.
- Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the rescorla-wagner model. *Psychological Bulletin*, *117*(3), 363–386.
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(7), in press.
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*, 1090–1098.
- Siegel, S., & Allan, L. G. (1996). The widespread influence of the rescorla-wagner model. *Psychonomic Bulletin & Review*, *3*(3), 314–321.
- Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and erp study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.
- Wagner, A., & Rescorla, R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii* (pp. 64–99). New York: Appleton-Century-Crofts.