

# Issues in modeling language processing analogically<sup>☆</sup>

David Eddington\*

*Department of Linguistics and English Language, 2129 JKHB, Brigham Young University, Provo, UT 84602-6278, USA*

Received 13 February 2003; received in revised form 1 April 2003; accepted 7 April 2003

---

## Abstract

Exemplar-based models of language assert that linguistic processing involves analogy to past linguistic experiences stored in the mental lexicon. This study explores how three factors influence the predictions made by exemplar-based simulations of linguistic phenomena. Three questions are posed as they relate to such simulations: (1) Is type frequency or token frequency a better predictor of outcomes? (2) What is the optimal way of aligning the variables in the database so that the most relevant analogs are found? (3) Are there significant differences between representing variables as phonemes versus representing them in terms of distinctive features? Spanish stress assignment and English past tense formation served as the linguistic phenomena on which these issues were tested. The results suggest that type frequency is a better predictor of outcomes, although simulations using token frequency were most successful when only middle frequency words were included. Several methods for aligning variables in the analogical database are discussed. The dual-alignment method has advantages for the English past tense task, but not in predicting Spanish stress. In the Spanish task, strict phonemic representation of words demonstrated no advantage over feature representation. However, phonemic representation produced better results than distinctive features in predicting the English past tense.

© 2003 Elsevier B.V. All rights reserved.

*Keywords:* Analogical modeling; Distinctive features versus phonemic representation; English past tense; Spanish stress assignment; Tilburg Memory-based Learner; Token frequency; Type frequency; Variable alignment

---

<sup>☆</sup> I am indebted to Steven Chandler and the anonymous reviewers for their input and critique of this paper.

\* Tel.: +1-801-422-2937; fax: +1-801-422-0474.

*E-mail address:* [david\\_eddington@byu.edu](mailto:david_eddington@byu.edu) (D. Eddington).

## 1. Introduction

In recent decades, as computers have become more accessible, and as their processing speed and memory capacity have increased dramatically, many language researchers have turned their attention to computational methods in order to model linguistic phenomena. The most widely utilized model is arguably connectionism (e.g. McClelland, 1988; Rumelhart and McClelland, 1986). However, connectionism is not the sole computational model linguists have at their disposal. Several other models have been developed that may be classified under the rubric of memory-based, exemplar-based or analogical models, for example, Nosofsky's Generalized Contextual Model (Nosofsky, 1990), Pierrehumbert's exemplar model (Pierrehumbert, 2001), the Tilburg Memory-based Learner (Daelemans et al., 2001), and Analogical Modeling of Language (Skousen, 1989, 1992). Exemplar models have been applied to investigate a wide variety of linguistic phenomena such as word recognition (Goldinger, 1996), Arabic and German plural formation (Nakisa et al., 2000), linking elements in Dutch noun compounds (Krott et al., 2002), phonological alternations in Turkish stems (Rytting, 2000), Dutch stress assignment (Gillis et al., 1993), Italian verb conjugations (Eddington, 2002b), and phonotactic knowledge in Arabic and English (Frisch et al., 2001).

The literature on connectionism is replete with discussions of how different network configurations, training sets, and input variables affect the outcome produced. For example MacWhinney et al. (1989) compared two simulations. In the first, the variables specified the presence or absence of 38 carefully chosen pieces of morphological, semantic, and phonological information about the word the article agrees with (e.g. whether the word contains a specific morpheme, or a phoneme in a certain position). Each of these cues is thought to govern definite article usage in German. In the second simulation, the only variables were the strings of phonemes that comprise the word. That is, no effort was made to include only those elements thought relevant to the task and separate them from those thought to be irrelevant. Nevertheless, this simulation yielded better results than the previous ones that carefully eliminated cues that were considered unimportant to the task of definite article assignment.

The importance of the present study resides in the fact that little attention had been paid to such important issues as they relate to analogical models. One issue that has been addressed in the analogical literature is that of corpus size; larger datasets appear to result in better outcomes (Banko and Brill, 2001; Daelemans et al., 1994a, 1997). However, many other issues have been left virtually undiscussed. Therefore, the principal purpose of the present paper is to fill this gap, and to explore how various factors influence exemplar-based models. In particular, I will attempt to answer three questions as they relate to analogical simulations: (1) Is type frequency or token frequency a better predictor of outcomes?<sup>1</sup> (2) What is the optimal way of aligning the variables in the dataset so that the most relevant analogs are

---

<sup>1</sup> One other measure of frequency which was not considered is that of family size (e.g. De Jong et al., 2000).

found? (3) Are there significant differences between representing variables as phonemes versus representing them in terms of distinctive features?

All of these issues relate to the question of how to develop a database that most closely represents how words are stored in the mental lexicon of language speakers, and how these words affect linguistic processing. Previous research demonstrates that the actual algorithm used to run a simulation does not affect the outcome a great deal (Daelemans et al., 1994b; Daelemans, 2002; Eddington, 2002a; Krott et al., 2002). However, my own experience has shown that altering the contents of the dataset on which the simulation is based has profound consequences. Therefore, it is important to determine how to construct the most optimal datasets for exemplar-modeling. The three issues mentioned above will be discussed as they relate to two phenomenon, English past tense formation and Spanish stress assignment.

The simulations discussed in the remainder of this paper are of two sorts. The first involves using the members of a dataset in order to predict the behavior of each individual member of the dataset (the leave-one-out method). What this essentially measures is the analogical consistency of the dataset, that is, the degree to which each item falls within a gang of similar items with similar behaviors. It is expected that extremely irregular items, such as the past tense of *go*, will not be correctly predicted as *went*. This type of simulation is useful for determining what the best analogical variables are from a corpus-internal standpoint. However, only when the results are correlated with outcomes produced by actual speakers of the language is one entitled to discuss the simulations as relevant to psycholinguistic processing.

## 2. Exemplar-based models

In the traditional rule-based approach, linguistic processing is thought to involve gleaning generalizations from the input data, and codifying these into rules which are then used in subsequent processing. Connectionism, on the other hand, may be classified as a sort of prototype model. The input given to the network results in certain patterns of representation of varying strengths being formed among the interconnected nodes of the network. Once the network is trained, processing relies on the patterns encoded in the network to produce the outcome. In contrast, exemplar- or memory-based models are founded on the idea that no sort of rule or prototypical representation needs to be generalized from the data and stored as a unit or entity separate from the data. Instead, generalizations exist within the stored lexical items themselves. Accordingly, linguistic processing is a matter of lexical access, and analogy to existing patterns found among the lexical items.

Of course, exemplar models require vast amounts of storage space if individual tokens of speech are retained in the mental lexicon. Nevertheless, there is evidence for such massive storage (Alegre and Gordon, 1999; Baayen et al., 1997; Bybee, 1995, 1998; Goldinger, 1997; Manelis and Tharp, 1977; Palmeri et al., 1993; Sereno and Jongman, 1997). It also appears that storage is not limited to the unpredictable features of speech, but that it includes redundant, detailed phonetic information about individual word tokens (Brown and McNeill, 1966; Bybee, 1994, 2000; Frisch,

1996; Pisoni, 1997). Storage may even go beyond individual words and encompass recurrent word combinations as well as entire phrases (Bod, 1998; Bybee, 1998; Pawley and Syder, 1983).

Two exemplar-based algorithms were employed to help answer the questions posed in the introduction, Analogical Modeling of Language (AM) and the Tilburg Memory-based Learner (TiMBL). It must be noted at this point that the accuracy and performance differences between these two algorithms is not the focus of the present paper. Readers who are interested in this topic are may consult the relevant literature (Daelemans et al., 1994b; Daelemans, 2002; Eddington, 2002a; Krott et al., 2002). It should also be noted that the present paper does not, and cannot explore all possible combinations of frequency, variable alignment, and feature versus phonemic representation.

One crucial element of all exemplar-based models is a database of words (or other linguistic variables) that serves as a kind of approximation of the mental lexicon of language speakers. If one's goal is to study consonant spirantization, the database would contain instances of words or word combinations in which spirantization has or has not occurred. One of the variables would indicate whether the database item is an example of spirantization or not. A study designed to predict the part of speech of words based on their phonological structure would contain phonological representations of nouns, verbs, adjectives, etc., along with a variable specifying the part of speech of each entry. Of course, the database should always be based on naturally occurring language data. The goal of exemplar modeling is to use a database representing a speaker's prior language experience, to predict linguistic behavior. For example, to predict the part of speech of a word whose part of speech is unknown, the task of the algorithm would be to compare the test word to the words in the database whose part of speech is known, and to extrapolate or analogize the word's part of speech based on database items that bear similarities to the word in question.

### *2.1. Memory-based language processing*

One memory-based algorithm used in the present study is found in the family of algorithms incorporated in the Tilburg Memory-based Learner (henceforth TiMBL). TiMBL is an expansion of the algorithm developed by Aha et al. (1991), and a detailed description of the algorithm is found in Daelemans et al. (2001). In essence, TiMBL takes an input and determines which items in a database of exemplars are the most similar to the input form. These are known as the nearest neighbors of the input. During the training session, the model stores in memory series of variables which represent instances of words. The words are stored along with their behavior (e.g. the type of past tense taken, or an indication of which syllable is stressed). In the case that the same word is encountered more than once in the database, a count is kept of how often each word is associated with a given behavior. In addition storing instances and counting duplicates, the training stage may also involve calculating the overall extent to which each variable helps predict the correct outcomes. This is called information gain.

During the testing phase, when an input is presented, the model searches for it in the database and applies the behavior that it has been assigned in the majority of cases. If the word is not found in the database, a similarity algorithm is used to find the most similar item(s)—its nearest neighbor(s). The behavior of the nearest neighbor(s) is then applied to the word in question. If two or more items are equidistant from the word in question, the most frequent behavior of the tied items is applied to the word in question. In the algorithm utilized in the majority of the present simulations (the modified value difference metric), the similarity between the values of a variable is precalculated and used to adjust the search for nearest neighbors accordingly. This precalculation allows certain values to be regarded as more similar to each other than other values.

## 2.2. Analogical modeling of language processes

Another memory-based model is found in Analogical Modeling of Language (hereafter AM; see Skousen, 1992, for an in depth explanation of the AM algorithm, and Eddington, 2000a, for a succinct description of AM's functioning). Unlike TiMBL, AM does not process the data beforehand in order to predetermine and weight which factors are most globally relevant to the task, nor does it precalculate variable value similarity. Instead, AM conducts a search of a database looking for words similar to the input word. In AM, the search begins with the entries most similar to the input word whose behavior is being predicted, and then extends to less similar entries. The members of the database are grouped into sets called subcontexts whose members share similarities with the input form. For example, in determining the past tense behavior of the English nonce verb *kive*, one subcontext would be comprised of all database items ending in /v/, another would contain those that end in /ajv/, another all items whose final syllable begins with /k/, another all items whose final syllable begins with /k/ and ends in /v/, and so forth until all possible combinations of all variables are explored.

One derived property that results from dividing the database in this manner is that of proximity. Database items that share more features with the nonce input *kive* will appear in more subcontexts and will therefore have a higher likelihood of influencing the probability that *kive* will be assigned a given past tense form. Gang effects also fall out of this architecture. Groups of similar items that display the same behavior will increase their chances of influencing the input form.

Heterogeneity is another important property of AM. It suggests that a word in the database cannot be chosen as an analog if there are intervening words, with a different behavior that are more similar to the input item. Calculating heterogeneity involves determining disagreements. A disagreement occurs when one member of a subcontext has a behavior that is different from the behavior of another member of the same subcontext. For example, *drive* and *thrive* share a final /aiv/, but form their past tense in a different manner (*drove*, *thrived*). As a result, when they appear in the same subcontext, they disagree in terms of what type of past tense they take. Under certain conditions, the analogical influence of the members of a subcontext that contains disagreements will be reduced or eliminated. AM's output is given in term of the statistical probability that one or more behaviors will apply to the input word.

Having introduced the two algorithms, the remainder of the paper will be dedicated to applying these algorithms to answer the questions posed in the introduction.

### 3. Spanish stress assignment

Spanish stress assignment is a fairly predictable phenomenon. With the exception of gerunds followed by two clitic pronouns (*haciéndoselo* ‘doing it for him/her’) stress falls on one of the final three syllables of a word. Words ending in a vowel or /s/ are more commonly stressed on the penultimate syllable, while those that end in a consonant other than /s/ usually receive final stress. This generalization holds for about 87% of Spanish words (Eddington, 2000b).

The databases for the Spanish stress assignment simulations are based on the 4970 most frequent Spanish word taken from a frequency dictionary (Alameda and Cuetos, 1995). Two databases were created. The type database contained only one entry for each word, while the token database contained multiple instances of words according to the word’s frequency in the dictionary. For example, one instance of the word *abuela* ‘grandmother’ appears in the type database, while 158 instances appear in the token database since the frequency of *abuela* is 158 in the frequency dictionary. The phonemic information in the final three syllables of each word were included as variables. Variables indicating the tense and person of each verbal form were also included. For example, in Table 1, ‘3’ indicates third person singular, and ‘pt’ indicates preterit tense. Dashes are used in place of spaces to keep the syllables aligned, and ‘0’ indicates the absence of a syllable when in the nucleus column, and a non-verb when in the tense column.

#### 3.1. Spanish stress: type and token frequency

There is a plethora of psycholinguistic evidence demonstrating that both type and token frequency play a part in language processing (e.g. Allen et al., 1992; Kelliher and Henderson, 1990; MacKay, 1982; Scarborough et al., 1977). However, one issue that has been left unanswered in the exemplar literature is the question of whether analogies are made on the basis of type or token frequency. The type and token databases of Spanish were designed to help answer this question, but some measure of performance was needed. One measure of generalization performance is cross-validation (Breiman et al., 1984). This consists of dividing a database into ten

Table 1  
Examples of variable encoding

Word	Stress	Variables										
		11	10	9	8	7	6	5	4	3	2	1
monstruo	Penult	–	0	–	0	–	m	o	ns	trw	o	–
resbaló	Final	3	pt	rr	e	s	b	a	–	l	o	–

groups. Each group is extracted from the remaining nine and its members serve as the test cases, while the members of the remaining nine groups comprise the data set from which analogs are chosen. In this way, each item in the database serves as a test case once, and as a possible source of analogical influence nine times. One advantage of cross-validation is that it does not allow exact matches to be found. That is, no word in the test set will find the exact same word in the database, therefore, the behavior of the words must be determined by analogy with similar words. If every test word finds the identical word in the database, that is the psychological equivalent of remembering the word (along with its stress pattern). In that case, the success rate would be an uninteresting 100%.

A 10-fold cross-validation was performed on the type database by dividing it into 10 sets of 497 items. TiMBL's modified value difference metric was used to determine the nearest neighbor<sup>2</sup> of each test item. The stress of the nearest neighbor was used to predict the stress of each test item. In this simulation, stress was successfully predicted on an average of 95.72% across the 10 type databases (range: 93.56–97.18%). The 10 token databases were then formed by multiplying the instances of the words in each type database by the word's frequency in Alameda and Cuetos (1995). However, even though the databases were augmented to reflect token frequency, the stress placement of each of the 4970 words was predicted only once, not multiple times according to their token frequency. Again, a 10-fold cross-validation was performed yielding an average success rate of 93.58% (range: 91.35–95.77%). These data suggest that type frequency is a better predictor of stress assignment than token frequency [ $\chi^2(1) = 21.12, P < 0.001$ ].

The fact that type frequency outperforms token frequency is not unique to Spanish stress assignment. I have found similar effects in simulations of Italian conjugation classes and Spanish gender assignment (Eddington, 2002b,c). Processing of Dutch morphology also appears to involve type frequency over token frequency (Bertram et al., 2000). Bybee (1985, 2001) suggests that type frequency is more important than token frequency in cases involving productivity. It appears that the computational method of treating each test item as if were a new and previously unencountered word is tantamount to testing the productivity of the Spanish stress patterns.

Another question of frequency which needs to be examined involves the frequencies at which the optimal analogs may be found. In a study of the phonotactics of English, Bailey and Hahn (2001) observed that subjects' ratings of wordlikeness appeared to be influenced more by the token frequency of medium frequency words; extremely high and extremely low token frequency words did not exert much influence. This finding contradicts Anshen and Aronoff (1988) who suggest that high frequency words are more likely to be chosen as analogs. However, it supports Bybee's (2001) claim that high frequency words are more autonomous from other

---

<sup>2</sup> A number of simulations were run in which either one, three, five, or ten nearest neighbors were used to determine behavior. However, in each case the simulation with one nearest neighbor always produced slightly better results than with three and five, which is why the results with three, five, and 10 are not reported.

words and are actually less likely to exert analogical influence. The present token database of Spanish words only represents high to medium frequency Spanish words. An extremely large database of Spanish words would need to be constructed in order to cover the entire frequency range. Nevertheless, if the token database is divided into two equal parts based on frequency, the most frequent part would contain high frequency words while the less frequent part could be considered representative of middle frequency words.

The most desirable way of testing the high and medium token frequency databases is to use all 4970 words as test items, and to test these against the high and middle frequency databases. In this way, it can be seen if there are differences in the results produced with the two databases of differing frequencies. The difficulty with this is structuring the simulation so that no test item finds its exact counterpart in the database. This required a bit of manipulation. First, the entire high frequency database was used as the database on which the stress placement of the middle frequency words was determined, and vice versa. Next, a 5-fold cross validation was performed on each database so that the high frequency words could be tested against the high frequency database, and the middle frequency words could be tested against the middle frequency database without encountering undesirable exact matches.

The average success rate for the simulations was 92.43% using the high frequency database, and 94.06% using the middle frequency database.<sup>3</sup> This demonstrates that the middle frequency database is a better set on which to analogize [ $\chi^2(1) = 9.778$ ,  $P < 0.005$ ]. Exactly why this is so is somewhat puzzling. One could imagine that the most frequent words contain more irregularly stressed words. However, this is not the case; 12.1% of the most frequent words have irregular stress, while 15.7% of the middle frequency words are irregular.<sup>4</sup> This finding is consistent with Bybee's (2001) hypothesis that high frequency words have a higher degree of autonomy from other words, and are therefore less likely to influence other words analogically. Lexical autonomy of this sort may affect the way speakers assign stress, however, in the simulation all words are equally accessible to the computer algorithm which means that the autonomy of high frequency items is not a factor. The relationship between this computer-generated data and the high degree of autonomy that high frequency words have for actual speakers is unclear.

### 3.2. Spanish stress: variable alignment

One question that arises when converting language data into computer-readable format is how to correctly codify the data. Consider the way the words in the stress assignment simulation were encoded in Table 1. *Monstruo* 'monster' contains eight phonemes, but these are compressed into only five variables. In this encoding, all

<sup>3</sup> In calculating success rates, care was taken to weight the results of each test set in proportion to the number of test cases it contained.

<sup>4</sup> Penult stress on vowel-final words and those ending in -s is regular, while other consonant-final words with final stress are considered regular. Antepenult stress is always regarded as irregular.

phonemes that fall into a syllable onset or coda combine to form a single variable; *ns* and *trw* form one variable. This means that the *ns* in the coda of the penultimate syllable of *mons.truo* will be counted as similar to the *ns* in *cons.truir* ‘to build’, but no similarity will be found to the *s* of *ras.go* ‘trait’, nor to the *n* of *can.to* ‘chant’. In order for this to occur, the members of the onsets and codas must be counted as separate variables. An additional question involves where glides belong. The current encoding places them in the onset or coda, but it could be argued that they belong in the nucleus. Accordingly, the final nucleus of *monstruo* should contain *wo* instead of a simple *o*.

To answer these questions, I modified the original database in several ways that are described below, and compared the performance of each encoding. The comparison of these encodings was done using a leave-one-out method, which is another measure of performance (Weiss and Kulikowski, 1991). This consists of removing each word from the database one at a time. The word that has been extracted becomes the test item, while the remainder of the items serve as the data set from which analogies are drawn. In this way, the stress placement of each word is calculated only once. A leave-one-out simulation was not possible with the token database because it would have resulted in multiple predictions being made for the same word. That, in turn, would not have allowed the results of the type and token simulations to be evaluated on equal grounds. Kohavi (1995) reports that cross-validation may have advantages over the leave-one-out method when one’s goal is to determine the superiority of one computational model over another. However, the focus of the present paper is not model selection, but the evaluation of databases with differing characteristics.

The question of whether glides should be included in the nucleus may be answered by reencoding the data to reflect this. Therefore, the original alignment of the phonemes in *monstruo* (m/o/ns/trw/o) was changed to (m/o/ns/tr/wo). I will refer to the latter as the glide-in-the-nucleus alignment, and to the former as the no-glide-in-the-nucleus alignment. The same changes were made in every word in the type database containing a glide. The success rates are as follows:

No-glide-in-the-nucleus (cross-validation method)	95.72%
No-glide-in-the-nucleus (leave-one-out method)	95.96%
Glide-in-the-nucleus (leave-one-out method)	95.89%

The first thing that must be noted is that there is no statistical difference between the cross validation and leave-one-out methods of measuring performance [ $\chi^2(1) = 0.348$ ,  $P < 0.25$ ]. Second, whether the glide is placed in the nucleus or in the onset or coda does not appear to be a factor that influences stress assignment [ $\chi^2(1) = 0.022$ ,  $P < 0.75$ ].

The next question to address is whether any benefit results from considering the individual members of a consonant cluster in an onset or coda as separate variables. This entails recoding words such as *monstruo* from (m/o/ns/trw/o) into something along the lines of (m/o/ n/s/t/r/w/o). However, this encoding is not adequate because it does not address the problem of how to correctly align the members of the onsets

and codas. Consider the final syllables of the words *monstruo*, *filtro* ‘filter’, and *continuo* ‘continuum’ ([trwo], [tro] and [nwo]). The ideal alignment would show that *continuo* shares *w* and *o* with *monstruo*, and *filtro* shares *t*, *r*, and *o* with *monstruo*. If we arrange the variables so that the first consonants in the onsets are aligned, the phonemes that *monstruo* and *filtro* have in common are correctly identified, in that they appear as variables in the same column:

(1)	t	r	w	o	<i>monstruo</i>
	t	r	-	o	<i>filtro</i>
	n	w	-	o	<i>continuo</i>

However, the *w* of *continuo* and *monstruo* belong to different variables, which means that the similarity between the words will not be identified by the analogical algorithm. The other possibility is to align the phonemes starting from the nucleus and working toward the left.

(2)	t	r	w	o	<i>monstruo</i>
	-	t	r	o	<i>filtro</i>
	-	n	w	o	<i>continuo</i>

This yields an alignment that highlights the fact that *continuo* shares *w* and *o* with *monstruo*, but fails to capture the *t* and *r* that *monstruo* and *filtro* have in common. The best resolution to this paradox, in my view, is to encode the data so that both alignments are represented at the same time. I will refer to this as dual-alignment:

(1)	t	r	w	t	r	w	o	<i>monstruo</i>
	t	r	-	-	t	r	o	<i>filtro</i>
	n	w	-	-	n	w	o	<i>continuo</i>

The dual-alignment database was created by separating the members of each onset and coda into separate variables and aligning them as in (3). Although the dual-alignment appeared to be an intuitively more correct way to encode the words, it produced no significant change in the outcome when a leave-one-out simulation was performed [ $\chi^2(1) = 0.690$ ,  $P < 0.5$ ].

No-glide-in-the-nucleus	95.96%
Dual-alignment	95.61%

The three different alignments experimented with to this point failed to demonstrate any real differing effect on the performance of the model. The reason why this is so may be found by inspecting the feature permutation. In the course of running a simulation, TiMBL ranks the variables in terms of how much each one contributes to making the predictions. This is called the feature permutation. In the no-glide-in-the-nucleus simulation, the five most important variables were: (1) the phoneme or absence of phoneme in the coda of the final syllable; few Spanish words have

complex codas in the final syllable; (2) the variable indicating the tense of verbal forms, (3) the variable indicating the person of verbal forms; none of the three alignments altered the morphological variables; (4) the nucleus of the final syllable; few Spanish words contain glides in the final syllable; (5) the nucleus of the penultimate syllable. What becomes obvious is that the different alignments manipulate the variables that are least relevant to stress assignment.

### 3.3. *Spanish stress: features versus phonemes*

One objection that could be made to the databases used in the simulations is that they use phonemes as variables. Phonemes may not capture similarities between variables in the same way that features can. In reality, the TiMBL algorithm used to this point (the modified value difference metric) compares all pairs of values in terms of their similarity as far as making analogical predictions is concerned. These similarity values play a part in the calculations. However, a pair of variable values may be calculated to be very similar as far as the task is concerned, but may not necessarily be phonetically similar. A case could be made that the use of distinctive features would improve the results of the simulations.

To clearly test the efficiency of phonemes versus phonetic features, I modified the dual-alignment database so that the only variables were the phonemes in the onset, nucleus, and rime of the final syllable. No morphological variables were included. The algorithm used in TiMBL for this simulation was overlap which does not calculate the similarity of variable values. Under these conditions, a leave-one-out analysis was performed which yielded a success rate of 88.62%. The phonemes in this database were then converted into series of 17 binary features.<sup>5</sup> Features that are irrelevant for consonants were marked with a '0' in the database, and features that are not pertinent to vowels were marked in the same fashion. The resulting success rate was an identical 88.62% to the simulation using phonemic representation.

One objection that could be made to comparing feature and phonemic representations is that the two databases are radically different, as a result, the task of predicting stress is essentially redefined. It may be that one algorithm (either AM's or one of TiMBL's) produces better results when the database contains phonemic representations, but another algorithm may prove more adept at processing binary features. While this may be true, if both the algorithm and the database are modified, it becomes impossible to tell whether the differing outcomes are due changes in the algorithm, changes in the database configuration, or a combination of both.

To summarize thus far, the different variable alignments, and the use of binary features in place of phonemes, produced no difference in the predictive ability of the exemplar model. This is most likely due to the fact that the specific consonants and consonant clusters which appear in onsets and codas play an insignificant role in stress assignment. As far as type versus token frequency is concerned, however, a

---

<sup>5</sup> Sonorant, consonantal, syllabic, continuant, voiced, aspirated, nasal, labial, anterior, coronal, strident, lateral, high, low, back, rounded, tense.

database reflecting type frequency is a much better predictor of stress assignment than one based on token frequency, although a token database consisting of middle frequency items is not ineffective.

#### 4. English past tense

The remainder of the present study describes a number of simulations that predict the English past tense. The goal of these simulations is, again, to evaluate the role of frequency, variable alignment, and phonemic representation as they relate to exemplar models of language processing. The English past tense has occupied a central role in the debate on language processing at least since Rumelhart and McClelland's controversial study in 1986 (e.g. Chandler and Skousen, 1997; Cottrell and Plunkett, 1994; Daugherty and Hare, 1993; Daugherty and Seidenberg, 1992, 1994; Eddington, 2000a; Jaeger et al., 1996; Lachter and Bever, 1988; MacWhinney and Leinbach, 1991; Marcus, 1995, 1998; Marcus et al., 1995; Pinker, 1999; Pinker and Prince, 1988; Plunkett and Marchman, 1991; Prasada et al., 1990; Seidenberg and Hoeffner, 1998). A survey of the literature on this debate, besides being extremely lengthy, would fall beyond the scope of the present paper. Instead of entering the debate, the past tense is simply used as a test case against which analogical databases with differing characteristics may be tested.

##### 4.1. *English past tense: type and token frequency*

The verbs utilized in this study were the same 2179 English verbs and their corresponding past tense forms used in the study by MacWhinney and Leinbach (1991). These include all verbs from Francis and Kuèera's English frequency dictionary (1982) as well as many extant verbs not found in that sample. Several English verbs allow two past tense forms (dived/dove), and in these cases, each alternative was included in the database. In the initial database, the present tense forms were encoded with the same variables used in Derwing and Skousen (1994). This includes the phonemes of the final two syllables, along with an indication of whether the final syllable is stressed or stressless.

The first few past tense simulations were carried out with AM's algorithm. AM was chosen because it gives the outcome in terms of the probability that one outcome or another will be applied.<sup>6</sup> This sort of output has the advantage of being interpreted in two different ways. One interpretation, termed 'selection by plurality' (Skousen, 1989), involves considering the behavior with the highest probability to apply. This sort of winner-take-all output is the produced by connectionist networks, as well as the TiMBL simulations discussed previously. With AM's 'random

---

<sup>6</sup> TiMBL does have another algorithm whose output is not deterministic. However, TiMBL's predictions are made based on a small sampling of possible analogs, while AM generally allows a large number of database items to affect the outcome to varying degrees. For this reason, I believe AM's results are more similar to those produced by actual speakers.

Table 2  
Success rates for simulations using different frequency measures

	Token frequency	Type frequency
All 2179 verbs	0.814	0.903
Most frequent 1090 verbs	0.777	0.877
Least frequent 1089 verbs	0.884	0.892
Middle frequency 1090 verbs	0.899	0.896

selection', on the other hand, one considers the degree to which two or more outcomes are predicted. This more fine-grained output is important when comparing the model to the results of psycholinguistic experiments which usually entail some degree of variability.

A leave-one-out simulation was run using all 2179 items in the database. Selection by plurality yielded a success rate of 90.32%. A token database was constructed by multiplying each item in the type database by the number of times they occur in Francis and Kučera (1982).<sup>7</sup> In this way, a series of variables representing an item with a frequency of 15 appeared 15 times in the token database. When the 2179 verbs were tested against this database a success rate of 81.37% resulted. The token simulation performed significantly poorer than the simulation using types [ $\chi^2(1) = 61.628, P < 0.001$ ].

One advantage of the past tense database is that, although it does not contain every English verb, it does represent a continuum spanning high frequency verbs (e.g. *go, have*) to very low frequency verbs (*slay, flog*). Therefore, that ability of an analogical database comprised of middle frequency words to outperform high and low frequency words could be assessed. In order to test this, the token and type frequency databases were divided in three ways. The first was to divide them in half according to frequency so that the verbs with the highest token frequency appeared in one database and those with the lowest token frequency in another. Another division eliminated the most frequent and least frequent fourths of the database, so that only the middle frequency items remained. The 2179 words were tested against these type and token databases. An option was set in the AM algorithm so that when a test item encountered an exact match in the data set, the influence of that item on the outcome was eliminated. The success rates are given in Table 2.

As far as token frequency is concerned, the middle frequency verbs appear to constitute a better set of items on which to analogize. Nevertheless, the highest success rate for the middle frequency token simulation (89.86%) is statistically equivalent to the success rate obtained using the entire type frequency database [90.32%;  $\chi^2(1) = 0.080, P < 0.9$ ]. Of course, it should not be surprising that the most frequent past tense verbs are the poorest set from which to choose analogs; they contain the most verbs with irregular past tense forms. In like manner, the low frequency database contains few irregular items from which verbs with irregular past tenses forms

<sup>7</sup> The frequency of words not found in Francis and Kučera was set at one for the purposes of the present studies.

Table 3  
Correlations with Prasada and Pinker's nonce verbs

	Token frequency	Type frequency
All 2179 verbs	0.896	0.996
Most frequent 1090 verbs	0.938	0.982
Least frequent 1089 verbs	0.987	0.989
Middle frequency 1090 verbs	0.987	0.993

can find correct analogies to other irregular items. When the highest frequency words are eliminated, the remaining words do not exhibit such radical frequency differences among themselves, which means that to a certain extent, the resulting token database is more similar to a type database. This may be one of the reasons that the middle frequency token database provides the best success rate on a computer simulation. However, other factors come into play when actual speakers process the English past tense as discussed below.

The use of the same database to draw both test items and data set items from is a common practice in natural language processing tasks of the sort reported on herein. However, most theories of language processing hold that irregular past tense forms must be stored in memory. Therefore, treating them as novel items in a simulation may be problematic (Ling and Marinov, 1993). One way to avoid this potential problem is by utilizing nonce words in place of existing words. Prasada and Pinker (1993) conducted a study in which they elicited the past tense form of 60 nonce words. 10 of the nonce words were designed to be highly similar to extant regular verbs (prototypical), ten were somewhat similar to regular verbs (intermediate), and 10 were very dissimilar to regular verbs (distant). The remaining 30 nonce verbs were arranged according to how similar they were to extant verbs with irregular past tense forms. In the experiment, they measured how often the subjects produced one of the regular past tense suffixes [-d, -t, -id] and how often an irregular form involving some sort of vowel change was produced. A number of simulations were run in order to predict the type of past tense form that would be given to Prasada and Pinker's nonce verbs. They were modified to reflect the sort of output reported by Prasada and Pinker. The simulations differed according to which database was used to draw analogies from, as seen in Table 2.

The results of the leave-one-out simulations and the nonce word study produced very similar outcomes (compares Table 2 and 3). In each case, the type databases outperformed the token databases, and in both cases, the most successful simulation utilized the type database that contained all 2179 verbs (see Fig. 1).<sup>8</sup> Simulations done with the highest frequency words underperformed all others. In the leave-one-out simulations using token databases, the middle frequency items appear to provide the best pool of possible analogs. In the nonce simulations, the token databases

<sup>8</sup> The reason that the results of this simulation is more highly correlated with the subjects' responses, when compared to the simulation reported on in Eddington (2000a) is most likely due to the larger database (2179 items versus 848 items in the previous study).

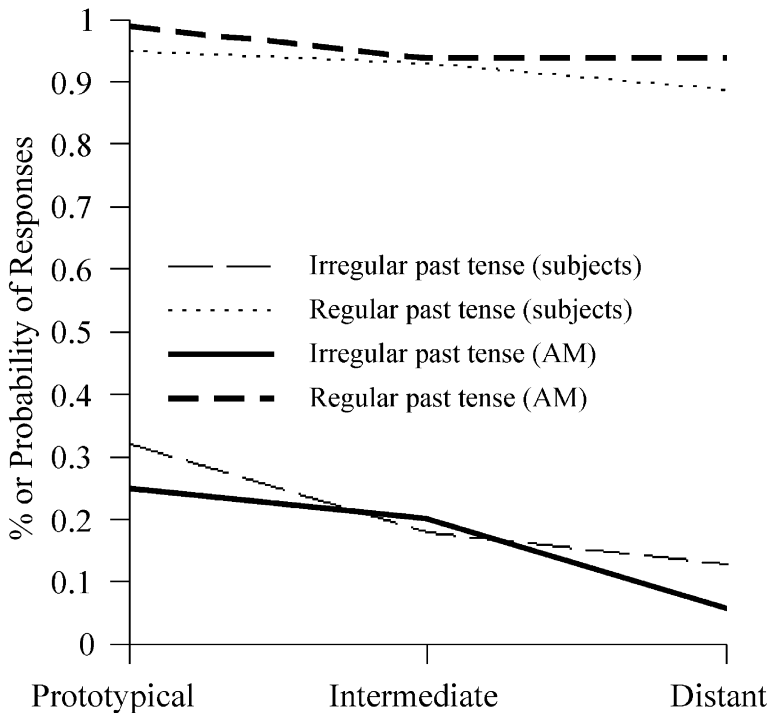


Fig. 1. Mean number/probability of past tense forms.

containing the middle frequency words also outperformed the high frequency database and the database containing all 2179 items. However, the low frequency database performed as well as the middle frequency database.

The fact that speakers appear to avoid high frequency items when forming analogies is consistent with Bybee's (2001) hypothesis that high frequency words are more autonomous from other words and are not as likely to be chosen as analogs. In a separate experiment on the English past tense, Moder (1992) elicited the past tense of nonce words designed to be similar to real English verbs. However, one group of subjects was primed by having to provide a past tense form for extant high frequency irregular verbs before responding to the test items. The other group was primed with medium frequency extant irregular verbs. Fewer regular forms were produced when following medium frequency primes compared to high frequency primes. Priming by the medium frequency verbs also resulted in more of the nonce words being given the same type of irregular past tense form as they were patterned after. Once again, high frequency forms are shown to be less likely to be used as analogs.

A number of things may be concluded from these simulations. (1) Since the outcomes of the nonce study and the leave-one-out study differ little, the concern that it is problematic to use the same database items as both test and data sets appears unwarranted; (2) Simulations using type frequency outperform those that use

Table 4  
Examples of the no-syllable-boundary alignment

Verb	12	11	10	9	8	7	6	5	4	3	2	1
<i>transform</i>	t	r	æ	n	s	s	f	o	r	m	F	m
<i>distribute</i>	t	r	I	b	j	b	j	u	t	0	N	t
<i>tally</i>	0	t	æ	l	i	æ	l	i	0	–	N	i

tokens; (3) Eliminating the most frequent items leads to better performance when token frequency is used. Whether this holds true for modeling phenomena besides the English past tense is a matter that needs to be investigated further.

#### 4.2. *English past tense: features versus phonemes*

In the Spanish stress simulation, it was seen that a feature representation of the phonemes provided no significant advantage over a purely phonemic representation, but this does not necessarily mean that the use of features may not yield more optimal results in modeling other phenomena. For example, in a study of English phonotactics, Ohala and Ohala (1986) presented nonce words to subjects and asked them to judge how English-sounding they were. Differences when measured in terms of phonemes rather than features best explained the experimental outcome.

In order to test phonemic versus feature representation, TiBML's overlap with gain ratio weighting algorithm was used simply because the AM algorithm could not handle the number of variables required. A leave-one-out simulation on the above mentioned database yielded a success rate of 91.55% when the variables were phonemic.<sup>9</sup> The phonemes were then converted into series of 17 binary features.<sup>10</sup> With binary feature representation, the leave-one-out analysis significantly underperformed the phonemic representation by correctly assigning the past tense to only 90.13% of the verbs [ $\chi^2(1) = 5.533, P < 0.025$ ].<sup>11</sup>

#### 4.3. *English past tense: variable alignment*

To this point, the databases utilized in the past tense simulations have been encoded using the variable alignment exemplified in Table 4, which I will refer to as the no-syllable-boundary alignment.

<sup>9</sup> The overlap algorithm was used so that similarities among phonemes were not precalculated. With this same database AM correctly predicted 90.32% of the past tense forms which is somewhat less successful than TiBML's 91.55% success rate [ $\chi^2(1) = 4.130, P < 0.005$ ].

<sup>10</sup> Sonorant, consonantal, syllabic, continuant, voiced, aspirated, nasal, labial, anterior, coronal, strident, lateral, high, low, back, rounded, tense. Consonantal features not relevant to vowels were marked with a '0', as were vocalic features not relevant to consonants.

<sup>11</sup> A number of other simulations were run in TiBML using various algorithms (overlap, no weighting,  $k = 1, 3, 5$ ; overlap, gain ration weighting,  $k = 1, 3, 5$ ; overlap, chi-squared weighting,  $k = 1, 3, 5$ ; overlap, shared variance weighting,  $k = 1, 3, 5$ ; modified value difference metric, information gain weighting,  $k = 1, 3, 5$ ). In all of these simulations, the phonemic and the feature databases were used, and in all simulations the phonemic database outperformed the feature database.

This alignment, which was taken from Derwing and Skousen's (1994) study of the English past tense, centers on the nuclei of the final two syllables (variables 5 and 10). Any phonemes appearing two slots before and after the nuclei are included as phonemes, regardless of whether these phonemes belong to the same syllable as the nucleus. Cases in which no phoneme appears are marked with '0'. In all of the alignments discussed below, variable 1 indicates the verb's final phoneme, and variable 2 whether the verb's stress falls on the final syllable (F) or not (N). One possible objection to the no-syllable-boundary alignment is that it encodes some phonemes twice (as in the case of the /s/ of *transform* and the /bj/ of *distribute*), and only once in other verbs.

The lumped-cluster alignment provides an encoding that respects syllable boundaries and syllable constituents. For example, in Table 5 variable 8 contains the onset of the penult syllable, variable 7 the nucleus, and variable 6 the coda. Empty syllable positions appear with '0'. Although the lumped-cluster alignment addresses the problem of respecting syllable constituents and boundaries, it may be problematic in that the members of a consonant cluster are not separate variables. For example, variable 8 contains *tr* for *transform* and *str* for *distribute*. Although both words share *tr*, the algorithms will treat the two variables as having nothing in common. It also fails to represent the fact that both *tally* and *transform* both begin with the same phoneme/variable *t*.

One way of separating the consonant clusters in to separate variables appears in Table 6. Here assignment of variables begins with the nucleus and moves outward incorporating phonemes belonging to the onset and coda of the syllable. This alignment faithfully represents the fact that *distribute* and *transform* share *tr*, but does not demonstrate that both *tally* and *transform* share word initial *t*. In order to represent both of these similarities, all of the consonants in an onset and codas must be encoded twice: once left-justified and once right-justified as in Table 7. This dual-alignment is the only way of encoding the variables so that all possible similarities between the members of codas and onsets are made.

Table 5  
Examples of the lumped-consonant-cluster alignment

Verb	8	7	6	5	4	3	2	1
<i>transform</i>	tr	æ	ns	f	o	rm	F	m
<i>distribute</i>	str	I	0	bj	u	t	N	t
<i>tally</i>	t	æ	0	l	i	0	N	I

Table 6  
Examples of the separated-consonant-cluster alignment

Verb	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
<i>transform</i>	0	t	r	æ	n	s	0	–	0	f	o	r	m	0	F	m
<i>distribute</i>	s	t	r	I	0	–	–	0	b	j	u	t	0	–	N	t
<i>tally</i>	–	0	t	æ	0	–	–	–	0	l	i	0	–	–	N	i

Table 7  
Examples of dual-alignment

Verb	8	7	6	5	4	3	2	1	0	9	8	7	6	5	4	3	2	1	0	9	8	7	6	5	4	3	2	1
<i>transform</i>	t	r	-	-	t	r	æ	n	s	-	-	n	s	f	-	-	-	-	f	o	r	m	-	-	r	m	F	m
<i>distribute</i>	s	t	r	s	t	r	I	0	-	-	-	-	0	b	j	-	-	b	j	u	t	-	-	-	-	t	N	t
<i>tally</i>	t	-	-	-	-	t	æ	0	-	-	-	-	0	l	-	-	-	-	l	i	0	-	-	-	-	0	N	i

In order to test the adequacy of each of these four alignments, four leave-one-out simulations were performed with type frequency represented. AM's algorithm was applied to this task with the following success rates, none of which differ significantly from another [ $\chi^2(1) = 1.450, P < 0.75$ ]:

No-syllable-boundary alignment	90.31%
Lumped-consonant-cluster alignment	91.28%
Separate-consonant-cluster alignment	91.19%
Dual-alignment	91.14%

These results are reminiscent of the insignificant differences that the alignments in the Spanish stress assignment produced. Nevertheless, it is important to establish, not only whether these alignments are equally optimal from a database-internal perspective, but whether they are equal when measured in terms of actual language processing. To this end, the four alignments were used to predict the past test forms of nonce words devised by Albright and Hayes (2001).

In two different experiments, Albright and Hayes asked subjects to provide past tenses for 58 nonce words. They calculated the percentage of responses in which a particular past tense form was provided. This involved determining how often a regular past tense form, or one or more irregular past tense forms, was given. For example the past tense of *spling* was given as *splinged* by 51.4% of the subjects, as *splung* by 32.4%, and as *splang* by 10.8% of the subjects. These data were correlated with AM's predicted probability that each past tense form would occur. Analogies were made with databases using each of the four variable alignments with the following results, all of which demonstrate a significant positive correlation with the subjects' responses ( $P < 0.005$  level, two-tailed):

No-syllable-boundary alignment	0.856
Lumped-consonant-cluster alignment	0.848
Separate-consonant-cluster alignment	0.877
Dual-alignment	0.886

It is clear that the least successful alignment is the one in which consonant clusters in codas and onsets are not treated as separate variables, but lumped together. The dual-alignment achieves the highest correlation since it allows more similarities to be found between words. That is, it is the only alignment that shows that *sing* and *string* both begin with *s*, and at the same time highlights the *r* that *string* and *ring* have in common.

The fact that the best alignment occurs when all syllabic constituents are represented and when syllable boundaries are respected may indicate that words are encoded syllabically in the mental lexicon. There is evidence to support the notion that syllable structure plays a role in language processing (Carreiras et al., 1993; Costa and Sebastián, 1998; Levelt and Wheeldon, 1995; Perea and Carreiras, 1998). However, such evidence is based on languages such as Dutch, French, and Spanish. In other studies, syllables do not appear to be a significant factor in processing English (Cutler et al., 1983, 1986).

## 5. Conclusions

The purpose of this study was to evaluate the role of frequency, variable alignment, and phonemic representation in analogical simulations. In the simulations which were carried out, databases based on type frequency yielded better results than those based on token frequency in predicting both Spanish stress assignment and English past tense forms. However, the simulations also suggest that when token frequency is used, the middle frequency items provide the best set from which to draw analogies.

One question that was also addressed was whether phonemic or feature representation is optimal. In the Spanish stress assignment task, both representations produced statistically similar results, while a feature representation of English verbs performed significantly poorer than did a straight phonemic representation.

A number of different variable alignments were tried on the Spanish data, none of which performed significantly better than the other. This may be due to the fact that the variables which are most important to Spanish stress assignment were not affected much by the different alignments. As far as predicting the English past tense, however, comparisons with the nonce word task suggest that it is better to consider the individual members of onsets and codas as separate variables. The dual-alignment proved most adept and may have some advantages over the others.

The results of the simulations reported in the present paper must be construed as being relevant only for the tasks to which they were applied. For instance, the fact that token frequency underperformed type frequency in simulations of Spanish stress and English past tense formation does not necessarily indicate that token frequency will never play an important role in predicting other linguistic behaviors analogically. Further investigation into other linguistic phenomena as well as into other languages are warranted before any broad generalizations may be made.

## References

- Aha, D.W., Kibler, D., Albert, M.K., 1991. Instance-based learning algorithms. *Machine Learning* 6, 37–66.
- Alameda, J.R., Cuetos, F., 1995. *Diccionario de Frecuencias de las Unidades Lingüísticas del Castellano*. University of Oviedo Press, Oviedo, Spain.
- Albright, A., Hayes, B., 2001. Rules vs. analogy in English past tenses: a computational/experimental study. Manuscript, UCLA. Available from: <http://www.linguistics.ucla.edu/people/hayes/rulesvsanalogy>.

- Alegre, M., Gordon, P., 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40, 41–61.
- Allen, P., McNeal, M., Kvak, D., 1992. Perhaps the lexicon is coded as a function of word frequency. *Journal of Memory and Language* 31, 826–844.
- Anshen, F., Aronoff, M., 1988. Producing morphologically complex words. *Linguistics* 26, 641–655.
- Baayen, H.R., Dijkstra, T., Schreuder, R., 1997. Singulars and plurals in Dutch: evidence for a parallel dual-route model. *Journal of Memory and Language* 37, 94–117.
- Bailey, T., Hahn, U., 2001. Determinants of wordlikeness: phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44, 568–591.
- Banko, M., Brill, E., 2001. Mitigating the paucity-of-data problem: exploring the effect of training corpus size on classifier performance for natural language processing. Paper presented at the Conference on Human Language Technology, San Diego.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth, Belmont, CA.
- Bertram, R., Baayen, R.H., Schreuder, R., 2000. Effects of family size for complex words. *Journal of Memory and Language* 42, 390–405.
- Brown, R., McNeill, D., 1966. The ‘tip of the tongue’ phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5, 325–337.
- Bod, R., 1998. *Beyond Grammar*. CSLI, Stanford, CA.
- Bybee, J., 1985. *Morphology*. John Benjamins, Amsterdam.
- Bybee, J., 1994. A view of phonology from a cognitive and functional perspective. *Cognitive Linguistics* 5, 285–305.
- Bybee, J., 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425–455.
- Bybee, J., 1998. The emergent lexicon. In: Gruber, M., Higgins, C.D., Olson, K.S., Wysocki, T. (Eds.), *Proceedings of the Chicago Linguistic Society*, Vol. 34. Chicago Linguistic Society, Chicago, pp. 421–435.
- Bybee, J., 2000. The phonology of the lexicon: evidence from lexical diffusion. In: Barlow, Michael, Kemmer, Suzanne (Eds.), *Usage-based Models of Language*. CLSI Publications, Stanford, CA, pp. 65–85.
- Bybee, J., 2001. *Phonology and Language Use*. Cambridge UP, Cambridge.
- Carreiras, M., Alvarez, C.J., de Vega, M., 1993. Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language* 32, 766–780.
- Chandler, S., Skousen, R., 1997. Analogical modeling and the English past tense: a reply to Jaeger et al. 1996. Manuscript, available from <http://humanities.byu.edu/aml/jaeger.html>.
- Costa, A., Sebastián, N., 1988. Abstract phonological structure in language production: evidence from Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, 886–903.
- Cottrell, G.W., Plunkett, K., 1994. Acquiring the mapping from meaning to sounds. *Connection Science* 6, 379–412.
- Cutler, A., Mehler, J., Norris, D., Seguí, J., 1983. A language-specific comprehension strategy. *Nature* 304, 159–160.
- Cutler, A., Mehler, J., Norris, D., Seguí, J., 1986. The syllable’s differing role in the segmentation of French and English. *Journal of Memory and Language* 25, 385–400.
- Daelemans, W., 2002. A comparison of analogical modeling to memory-based language processing. In: Skousen, R., Lonsdale, D., Parkinson, B.D. (Eds.), *Analogical Modeling: An Exemplar-based Approach to Language*. John Benjamins, Amsterdam, pp. 157–179.
- Daelemans, W., Berck, P., Gillis, S.S., 1997. Data mining as a method for linguistics analysis: Dutch diminutives. *Folia Linguistica* 31, 57–75.
- Daelemans, W., Gillis, S., Durieux, G., 1994a. The acquisition of stress: a data-oriented approach. *Computational Linguistics* 20, 421–451.
- Daelemans, W., Gillis, S., Durieux, G., 1994b. Skousen’s analogical modeling algorithm: a comparison with lazy learning. In: Jones, D. (Ed.), *Proceedings of the International Conference on New Methods in Language Processing*. UMIST, Manchester, pp. 1–7.
- Daelemans, W., Zavrel, J., van der Sloot, K., van den Bosch, A.M., 2001. TiMBL: Tilburg Memory-based

- Learner, Version 4.1 Reference Guide. Induction of Linguistic Knowledge Technical Report, ILK 01-04. ILK Research Group, Tilburg University, Tilburg, Netherlands. Available from <http://ilk.kub.nl/>.
- Daugherty, K., Hare, M., 1993. The past tense by some other name might be called a connectionist net. In: Mozer, M., Smolensky, P., Touretzky, D., Elman, J. (Eds.), *Proceedings of the 1993 Connectionist Models Summer School*. Erlbaum, Hillsdale, NJ.
- Daugherty, K., Seidenberg, M.S., 1992. Rules or connections? The past tense revised. In: *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Erlbaum, Hillsdale, NJ, pp. 259–264.
- Daugherty, K., Seidenberg, M.S., 1994. Beyond rules and exceptions: a Connectionist approach to inflectional morphology. In: Lima, S.D., Corrigan, R.L., Iverson, G.K. (Eds.), *The Reality of Linguistic Rules*. Benjamins, Amsterdam, pp. 353–388.
- De Jong, N., Schreuder, R., Baayen, H., 2000. The morphological family size effect and morphology. *Language and Cognitive Processes* 15, 329–365.
- Derwing, B.L., Skousen, R., 1994. Productivity and the English past tense: testing Skousen's analogy model. In: Lima, S.D., Corrigan, R.L., Iverson, G.K. (Eds.), *The Reality of Linguistic Rules*. Benjamins, Amsterdam, pp. 193–218.
- Eddington, D., 2000a. Analogy and the dual-route model of morphology. *Lingua* 110, 281–298.
- Eddington, D., 2000b. Spanish stress assignment within the analogical modeling of language. *Language* 76, 92–109.
- Eddington, D., 2002a. A comparison of two models: Tilburg memory-based learner versus analogical modeling of language. In: Skousen, R., Lonsdale, D., Parkinson, D.B. (Eds.), *Analogical Modeling: An Exemplar-based Approach to Language*. John Benjamins, Amsterdam, pp. 141–155.
- Eddington, D., 2002b. Dissociation in Italian conjugations: a single-route account. *Brain and Language* 81, 291–302.
- Eddington, D., 2002c. Spanish gender assignment in an analogical framework. *Journal of Quantitative Linguistics* 9, 49–75.
- Francis, N.W., Kuèera, H., 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin, Boston.
- Frisch, Stefan, 1996. *Similarity and frequency in phonology*, Dissertation, Northwestern University.
- Frisch, S.A., Large, N.R., Zawaydeh, B., Pisoni, D.B., 2001. Emergent phonotactic generalizations in English and Arabic. In: Bybee, J., Hooper, P. (Eds.), *Frequency and the Emergence of Linguistic Structure*. John Benjamins, Amsterdam, pp. 159–179.
- Gillis, S., Daelemans, W., Durieux, G., van den Bosch, A., 1993. Learnability and markedness: Dutch stress assignment. In: *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, NJ, pp. 452–457.
- Goldinger, S.D., 1996. Words and voices: episodic traces in spoken word identification and recognition in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 1166–1183.
- Goldinger, S.D., 1997. Words and voices: perception and production in an episodic lexicon. In: Johnson, K., Mullennix, J.W. (Eds.), *Talker Variability in Speech Processing*. Academic Press, San Diego, pp. 33–65.
- Jaeger, J.J., Lockwood, A.H., Kemmerer, D.L., Van Valin Jr., R.D., Murphy, B.W., Khalak, H.G., 1996. A positron emission tomographic study of regular and irregular verb morphology in English. *Language* 72, 451–497.
- Kelliher, S., Henderson, L., 1990. Morphologically based frequency effects in the recognition of irregularly inflected verbs. *British Journal of Psychology* 81, 527–539.
- Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, vol. 2, Morgan Kaufmann, San Mateo, CA., pp. 1137–1145.
- Krott, A., Schreuder, R., Baayen, R.H., 2002. Analogical hierarchy: exemplar-based modeling of linkers in Dutch noun-noun compounds. In: Skousen, R., Lonsdale, D., Parkinson, D.B. (Eds.), *Analogical Modeling: An Exemplar-based Approach to Language*. John Benjamins, Amsterdam, pp. 181–206.
- Lachter, J., Bever, T.G., 1988. The relation between linguistic structure and associative theories of language learning: a constructive critique of some Connectionist learning models. *Cognition* 28, 195–247.

- Levelt, W., Wheeldon, L., 1995. Do speakers have access to a mental syllabary? In: Mehler, J., Franck, S. (Eds.), *Cognition on Cognition*. MIT Press, Cambridge, MA, pp. 301–331.
- Ling, C.X., Marinov, M., 1993. Answering the connectionist challenge: a symbolic model of learning the past tenses of English verbs. *Cognition* 49, 235–290.
- MacKay, D.G., 1982. The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review* 89, 60–94.
- MacWhinney, B., Leinbach, J., 1991. Implementations are not conceptualisations: revising the verb learning model. *Cognition* 29, 121–157.
- MacWhinney, B., Leinbach, J., Taraban, R., McDonald, J., 1989. Language learning: rules or cues? *Journal of Memory and Language* 28, 255–277.
- Manelis, L., Tharp, D.A., 1977. The processing of affixed words. *Memory and Cognition* 5, 690–695.
- Marcus, G., 1995. The acquisition of the English past tense in children and multilayered connectionist networks. *Cognition* 56, 271–279.
- Marcus, G., 1998. *The Algebraic Mind: Reflections on Connectionism and Cognitive Science*. MIT Press, Cambridge, MA.
- Marcus, G.F., Brinkmann, U., Clahsen, H., Wiese, R., Woest, A., Pinker, S., 1995. German inflection: the exception that proves the rule. *Cognitive Psychology* 29, 189–256.
- McClelland, J.L., 1988. Connectionist models and psychological evidence. *Journal of Memory and Language* 27, 107–123.
- Moder, C.L., 1992. Productivity and categorization in morphological classes. Dissertation, State University of New York, Buffalo.
- Nakisa, R.C., Plunkett, K., Hahn, U., 2000. Single- and dual-route models of inflectional morphology. In: Broeder, P., Murre, J. (Eds.), *Models of Language Acquisition: Inductive and Deductive Approaches*. Oxford UP, Oxford, pp. 201–222.
- Nosofsky, R.M., 1990. Relations between exemplar similarity and likelihood models of classification. *Journal of Mathematical Psychology* 34, 393–418.
- Ohala, J.J., Ohala, M., 1986. Testing hypotheses regarding the psychological manifestation of morpheme structure constraints. In: Ohala, J.J., Jaeger, J.J. (Eds.), *Experimental Phonology*. Academic Press, Orlando, pp. 239–252.
- Palmeri, T.J., Goldinger, S.D., Pisoni, D.B., 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19, 309–328.
- Pawley, A., Hodgetts Syder, F., 1993. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: Richards, J.C., Smith, R.W. (Eds.), *Language and Communication*. Longman, London, pp. 191–225.
- Perea, M., Carreiras, M., 1998. Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 24, 134–144.
- Pierrehumbert, J., 2001. Exemplar dynamics: word frequency, lenition and contrast. In: Bybee, J., Hopper, P. (Eds.), *Frequency and the Emergence of Linguistic Structure*. John Benjamins, Amsterdam, pp. 137–158.
- Pinker, S., 1999. *Words and Rules: The Ingredients of Language*. Basic Books, New York.
- Pinker, S., Prince, A., 1988. On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition* 28, 73–193.
- Pisoni, D., 1997. Some thoughts on ‘normalization’ in speech perception. In: Johnson, K., Mullenix, J.W. (Eds.), *Talker Variability in Speech Processing*. Academic Press, San Diego, pp. 9–32.
- Plunkett, K., Marchman, V., 1991. U-shaped learning and frequency effects in a multi-layered perception. *Cognition* 38, 43–102.
- Prasada, S., Pinker, S., 1993. Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes* 8, 1–56.
- Prasada, S., Pinker, S., Snyder, W., 1990. Some evidence that irregular forms are retrieved from memory but regular forms are rule-generated. Paper presented at the 31st Annual Meeting of the Psychonomic Society, November, New Orleans.

- Rytting, C.A., 2000. An empirical test of analogical modeling: the /k/~t alternation. In: Melby, A.K., Lommel, A.R. (Eds.), *Lacus Forum XVII: The Lexicon*. Linguistic Association of Canada and the United States, Fullerton, CA, pp. 73–84.
- Rumelhart, D.E., McClelland, J.L., 1986. On learning the past tense of English verbs. In: McClelland, J.L., Rumelhart, D.E. and the PDP Research Group (Eds.), *Parallel Distributed Processing*, Vol. 2. MIT Press, Cambridge, MA.
- Scarborough, D.L., Cortese, C., Scarborough, H.S., 1977. Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology* 3, 1–17.
- Seidenberg, M.S., Hoeffner, J.H., 1998. Evaluating behavioral and neuroimaging data on past tense processing. *Language* 74, 104–122.
- Sereno, J.A., Jongman, A., 1997. Processing of English inflectional morphology. *Memory and Cognition* 25, 425–437.
- Skousen, R., 1989. *Analogical Modeling of Language*. Kluwer, Dordrecht.
- Skousen, R., 1992. *Analogy and Structure*. Kluwer, Dordrecht.
- Weiss, S.M., Kulikowski, C.A., 1991. *Computer Systems that Learn*. Morgan Kaufmann, San Mateo, CA.