Chapter 6

A comparison of two analogical models

Tilburg Memory-Based Learner versus Analogical Modeling*

David Eddington

Introduction

Linguistics in the latter half of the twentieth century has been largely dominated by the rule-based paradigm of generativism. However, in the past few years, a number of non-rule approaches have been proposed and have gained some ground. Interest in non-rule approaches to linguistics may be the result of several different factors: disillusion with the generative paradigm, skepticism regarding the psychological relevance of generative analyses (Eddington 1996), advances in applying computer technology to questions of language (Natural Language Processing), and the heightened interest of psychologists in linguistic issues. Connectionism (see McClelland 1988 for an overview) has surfaced as the most prominent non-rule rival of the rule-driven orthodoxy, and the ongoing debate between connectionists and generativists has been intense (e.g. Clahsen et al. 1992; Daugherty & Seidenberg 1992, 1994; Marcus et al. 1995; Pinker 1991, 1997; Pinker & Prince 1994; Seidenberg 1992; Seidenberg & Bruck 1990).

In spite of its prominence, connectionism is not the sole non-rule model in existence. The present work compares two non-rule models of linguistic cognition, namely Analogical Modeling (AM) (Skousen 1989, 1992, 1995), and the nearest neighbor approach employed by the Tilburg Memory-Based Learner (TIMBL) (Daelemans et al. 1999). Both of these approaches belong to a family of models known as analogy-based, exemplar-based, or instance-based models (e.g. Bod 1998; Medin & Schaffer 1978; Nosofsky 1988, 1990; Riesbeck & Schank 1989; see Shanks 1995 for an overview of exemplar-based models). All of these models assume that previously encountered or processed information is stored in memory and is accessed and used to predict subsequent language behavior. Since each instance-based model employs a different algorithm, it is important to determine
if there are significant empirical differences between the predictions they make. Therefore, the focus of this paper will be to compare AM and TiMBL in terms of their performance on a number of different tasks. I will begin by reviewing the study by Daelemans et al. (1994b) which compares the ability of AM and TiMBL to assign stress to monomorphic Dutch words. Next, I compare TiMBL and AM in terms of their ability to account for Spanish diminutive formation, gender assignment, and stress assignment.

1. The TiMBL algorithm

Before reviewing the evidence from Dutch stress assignment, it is important to understand how TiMBL calculates nearest neighbors. TiMBL is essentially an expansion of the algorithm developed by Aha et al. (1991). It is designed to take an input and find its nearest neighbor(s) in a database of exemplars. During the training session, the model stores in memory series of variables which represent instances of words (or some other entity). The words are stored along with their behavior (e.g., which syllable is stressed, the word’s gender, etc.). In the case that the same word is encountered more than once, a count is kept of how often each word is associated with a given behavior. During the testing phase, when a word is given as input, the model searches for it in memory and applies the behavior that it has been assigned in the majority of cases. If the word is not found in memory, a similarity algorithm is used to find the most similar item in memory – its nearest neighbor. The behavior of the nearest neighbor 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.

The TiMBL algorithm contains several variants. For example, it can be set to determine the behavior on the basis of a single nearest neighbor, or on the basis of several nearest neighbors. In its basic instantiation, called Overlap, all variables are weighted equally. However, two extended algorithms are also available. Information Gain is a variant of Overlap which precalculates how much each variable contributes to determining the correct behavior. These variables are weighted accordingly when calculating similarity and determining nearest neighbors. When a calculation of similarity is carried out using Overlap, the values of a variable are all considered equidistant from each other. However, the Modified Value Difference Metric is also an available option. It is used to precalculate the similarity between the values of a variable, and to adjust the search for nearest neighbors accordingly. In effect, this allows certain values to be regarded as more similar to each other than other values.
2. Dutch stress assignment in TiMBL and AM

Daelemans et al. (1994a) constructed a database consisting of 4860 monomorphic multisyllabic Dutch words. Since stress may fall on any of the final three syllables, the phonemic content of the final three syllables of each word served as the variables in the AM and TiMBL comparisons (Daelemans et al. 1994b). Several ten-fold cross-validation simulations were performed on the database. This involved partitioning the database into ten sets of 486 words, and then running ten simulations for each experimental condition. Each of the ten sets of 486 words had its turn as a test set in one of the ten simulations; the words in the remaining nine sets formed the training sets.

Daelemans et al. (1994b) applied the basic Overlap algorithm in which the behaviors of one, two, five and ten nearest neighbors were applied to the words in the test sets. In all four conditions, AM’s success rate (80.5%) was statistically superior to those produced by TiBML. However, when varying degrees of noise were added to the four conditions, both models performed equally well (or poorly). When variables were weighted with the Information Gain (IG) algorithm, the Modified Value Difference Metric algorithm (MVDM), and both the IG and MVDM algorithms together, the success rates (81.8%, 79.4%, 81.4% respectively) became statistically equal to that of AM. In short, the findings from the Dutch stress assignment study indicate that TiMBL’s modified algorithms are equally adept at correctly assigning stress as AM. However, it is important to determine if this equivalence will hold true when other data are considered. If not, it is of interest to know which model is empirically superior. To this end, data from Spanish were considered.

3. Spanish gender assignment

The ability to predict gender seemed an apt task for an analogical model. All Spanish nouns belong to either the masculine or feminine gender. In general, words ending in -o are masculine, while those which end in -a are feminine. However, there are many exceptions to this generalization, and it is much more difficult to predict the gender of words ending in other phonemes.

The database for the gender simulation included the 1739 most frequent nouns in the Spanish language taken from LEXESP (Sebastián, Martí, Carreiras, & Cuetos 2002). Each noun was encoded to include the phonemic make-up and syllable structure of the penult rhyme and final syllable. The nouns were also marked as to whether they had masculine or feminine gender (for details, see Eddington 2002). Again, both TiMBL’s and AM’s algorithms were put to the task. AM successfully assigned gender to 94.5% of the database items.
Table 1. Success rate on correctly assigning gender to database items

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>#</th>
<th>%</th>
<th>(\chi^2)</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM</td>
<td>1645</td>
<td>94.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TiBML-no weighting, 3 nn</td>
<td>1563</td>
<td>89.9</td>
<td>2.147</td>
<td>0.25</td>
</tr>
<tr>
<td>TiBML-Information Gain, 3nn</td>
<td>1650</td>
<td>94.9</td>
<td>0.005</td>
<td>0.95</td>
</tr>
<tr>
<td>TiBML-MVDM, 3 nn</td>
<td>1673</td>
<td>96.2</td>
<td>0.220</td>
<td>0.75</td>
</tr>
<tr>
<td>TiBML-MVDM and Info. Gain, 3nn</td>
<td>1668</td>
<td>95.9</td>
<td>0.146</td>
<td>0.75</td>
</tr>
</tbody>
</table>

MVDM = Modified Value Difference Metric; nn = number of nearest neighbors calculated

Given the fact that many nouns have the identical phonological content in their penult rhyme and final syllable, it was necessary to eliminate exact matches between the test item and any items in the database. In the AM simulation, this was done by eliminating any identical given contexts which existed in the database. In order to achieve the same effect in TiBML, it was necessary to set the option to avoid choosing neighbors which are exact matches. This option requires that more than one nearest neighbor be selected, and in order to avoid ties between neighbors with different behaviors, the number of nearest neighbors needs to be odd. For this reason, the analogical influence of three nearest neighbors was considered in the TiBML simulations. Four different TiBML simulations were run using the basic overlap algorithm with no weighting, Information Gain (IG), the Modified Value Difference Metric algorithm (MVDM), and both the IG and MVDM algorithms together. As Table 1 indicates, the success rates of all of the TiBML simulations do not differ significantly from that of AM.

3.1 Gender assignment task

According to the outcome of the study on the database, no statistically significant difference was found between the two models. Therefore, each model was tested as to its ability to predict native speaker's intuitions about the gender of novel words.

3.1.1 Stimulus materials

118 nouns were extracted from *Diccionario de la lengua española* (Real Academia Española, 1995). Each of these words is considered antiquated and of infrequent use (see Appendix). Therefore, they were highly unlikely to be known by the subjects, which also means that their gender would be unknown. Words were chosen that ended in phonemes other than \(o\) and \(a\). In this way, the more obvious gender-phoneme correspondences were eliminated, and the subjects were obliged to make gender assignments on the more ambiguous cases.
3.1.2  Subjects
31 literate native Spanish speakers from Spain participated in the study, 18 women and 13 men. The average age of the subjects was 33.4.

3.1.3  Procedure
The 118 test items were presented in the form of a written questionnaire. The subjects were asked to circle either the feminine article *la* or the masculine article *el*, which appeared before each test item. They were instructed to choose the article that was most appropriate for the word that followed. Using the database of 1739 words previously described, the 118 words from the study were assigned gender by AM and by TiMBL's most successful algorithm (3nn, MVDM; see Table 1).

3.1.4  Results
TiMBL assigned the same gender as the subjects in the study to 67.8% of the test items. AM scored slightly higher at 71.2%. Nevertheless, the difference is once again not significant ($\chi^2 = 0.055, p < .5$).

3.2  Gender of borrowed words
Another task which analogical models appear to be well suited is predicting the gender of foreign words adopted into Spanish. Zamora (1975) studied borrowings from English into Puerto Rican Spanish. He asked 13 bilingual speakers to determine the gender of 20 English words that are commonly used in Puerto Rican Spanish. He also discusses 67 Native American words which were adopted into Spanish and had to be assigned a gender. Gender predictions were provided by AM and TiBML for these words based on the phonetic make-up of the penult rhyme and final syllable of the words' Spanish adaptation. TiMBL's most successful algorithm (3nn, MVDM) successfully predicted 86.2% of the 87 borrowings considered, while AM attained a success rate of 90.8%. The small difference is not significant ($\chi^2 = 2.157, p < .25$).

As far as the data from gender assignment are concerned, both models perform equally well, and the superiority of one over the other cannot be asserted. Nevertheless, gender is a fairly simple phenomenon since it only entails two possible outcomes. Differences between the models may be found in predicting behaviors with many outcomes. To this end, an experiment with diminutive formation was carried out.
4. Spanish diminutive formation

The formation of diminutive variants of nouns, adjectives and certain adverbs is a highly productive process in Spanish. Several diminutive suffixes exist (-ito, -illo, -zuelo, -ico, -uco), but -ito is the most common, which is why only diminutives ending in -ito/a were considered. All such diminutive forms were extracted from a number of databases.3

With the exception of a handful of highly irregular items, all diminutives fall into one of 13 categories. A circled V or s indicates that that element of the base form does not appear in the diminutive form:

1. -ITO(S): -ito(s) is added to the singular base form, replacing the final vowel: minuto > minutilo, elefante > elefantito.
2. -ITA(S): -ita(s) is added to the singular base form, replacing the final vowel: galleta > galletita, Lupe > Lupita.
3. -ECITO(S): -ecito(s) is added to the singular base form, replacing the final vowel: vidrio > vidricito, quieto > quietecito.
4. -ECITA(S): -ecita(s) is added to the singular base form, replacing the final vowel: yerba > yerbecita, piedra > piedrecita.
5. -CITO(S): -cito(s) is added to the singular base form: traje > trapecito, pastor > pastorcito.
6. -CITA(S): -cita(s) is added to the singular base form: joven > jovencita, llave > llavecita.
7. -ITO(S): -ito(s) is added to the singular base form: normal > normalito, Andrés > Andresito.
8. -ITA(S): -ita(s) is added to the singular base form: nariz > naricita, Isabel > Isabelita.
9. -ECITO(S): -ecito(s) is added to the singular base form: pez > pececito, rey > rejecito.
10. -ECITA(S): -ecita(s) is added to the singular base form: flor > florecita, luz > lucvecita.
11. -ITOS: -itos is added to the singular base form, replacing the vowel and false plural morpheme: lejos > lejitos, Marcos > Marquitos.4
12. -ITAS: -itas is added to the singular base form, replacing the vowel and false plural morpheme: Lucas > Luqutitas, garrapatas > garrapatitas.
13. -CITA(S): -cita(s) is added to the singular base form, replacing the final vowel: jamona > jamoncita, patrona > patroncita.

The resulting database contained 2450 diminutive forms. Each base form was marked as to which of the 13 categories its diminutive belonged to, and the following information about each base form was included: (1) the stressed or unstressed status of the final two syllables; (2) the gender of the word: masculine,
Table 2. Success rate on correctly assigning diminutives to database items

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>n</th>
<th>%</th>
<th>( \chi^2 )</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM</td>
<td>2285</td>
<td>93.27</td>
<td>0.468</td>
<td>0.5</td>
</tr>
<tr>
<td>TiBML-no weighting, 3 nn</td>
<td>2238</td>
<td>91.35</td>
<td>0.063</td>
<td>0.5</td>
</tr>
<tr>
<td>TiBML-no weighting, 5 nn</td>
<td>2136</td>
<td>87.18</td>
<td>0.037</td>
<td>0.9</td>
</tr>
<tr>
<td>TiBML-Information Gain, 3nn</td>
<td>2267</td>
<td>92.53</td>
<td>0.049</td>
<td>0.9</td>
</tr>
<tr>
<td>TiBML-MVDM, 3 nn</td>
<td>2271</td>
<td>92.69</td>
<td>0.049</td>
<td>0.9</td>
</tr>
<tr>
<td>TiBML-MVDM and Info. Gain, 3nn</td>
<td>2269</td>
<td>92.61</td>
<td>0.049</td>
<td>0.9</td>
</tr>
</tbody>
</table>

MVDV = Modified Value Difference Metric; nn = number of nearest neighbors calculated

feminine or none in the case of adverbs and gerunds; (3) the word’s final phoneme; (4) the phonological content of the antepenult rhyme and the final two syllables of the word.

A ten-fold cross-validation simulation was performed using AM’s algorithm, and several of TiBML’s algorithms. In the no weighting conditions using TiBML, and in the AM simulation, the gender variable and the word’s final phoneme were included twice in order to weight them more heavily than any other single variable. This duplication was removed in the Information Gain and Modified Value Difference Metric simulations, since these algorithms are designed to calculate the importance of the variables and values on their own. As Table 2 indicates, the TiBML simulations performed as well as the AM simulation with the exception of the simulation calculated without any of TiBML’s weighting algorithms using five nearest neighbors.

Many of the erroneous diminutives predicted by AM and TiBML’s most successful instantiation appeared to be plausible diminutives. This is evident in the predictions made on the doublets in the database (e.g. cuento, cuentecito). In each case, errors on one member of the doublet always entailed assigning it the diminutive suffix of the other member. This assignment occurred in spite of the fact that, when tested, both members of a doublet were excluded from the database and were unable to serve as analogs for each other. In order to determine if other erroneously predicted forms were actually well-formed diminutives in some dialect of Spanish, the World-Wide Web was consulted. All erroneous forms, were sought on Spanish language pages. Of the 165 errors produced by AM, attested forms of 77 were found, either as an attested doublet in the database or on a Spanish language web page. Therefore, only 88 errors involved truly unattested diminutive forms. In the TiBML simulation, only 84 errors were unattested.

As far as diminutivization in Spanish is concerned, TiBML and AM are able to correctly produce the great majority of the tested forms correctly. An inspection of the errors made by both models does not yield any insight that allows one model to be declared superior to the other.
5. Spanish stress assignment

Stress in Spanish generally falls on one of the last three syllables. The database chosen for the present study essentially includes the 4970 most frequent words, and word plus clitic pronoun combinations, from the Alameda and Cuertos frequency dictionary (1995). (Details about the database and variables are found in Eddington 2000.) As in the Dutch study, the phonemic content of the final three syllables of each word was used as variables. However, unlike the study on Dutch, both monomorphemic and polymorphemic words appeared in the Spanish database. This is a crucial difference since in Spanish stress is often contrastive, especially in polymorphemic and verbal forms: encontrará 's/he found, imperfect subjunctive,' encontrará 's/he will find'; busco 'I seek,' buscó 's/he sought.'

Therefore, in addition to the phonemic information, morphological variables were included. For verbal forms, one variable indicated the person, and three identical variables indicated the tense form of the verb. Repeating a variable more than once is the only way to manipulate the weight of one variable or another prior to running the AM program. In essence, what this implies is that the tense form of the verb is considered three times more important than any single onset, nucleus or coda. In the AM simulation, the only significant difference that weighting this variable made was in the number of errors that occurred on preterit verbs with final stress. Fifty errors occurred without the weighting, in comparison to 27 when it was included three times.

Each word was encoded as a series of 13 variables. In Table 3, hyphens represent empty categories, '0' indicates that the entry is a non-verb, 'pt' designates the verb is in the preterit tense, and '6' defines the verb as third person plural.

Given the fact that the database contained several inflectional variants of many words, a possible confound exists. If one of the test items is the adjective rójas, the chances are quite high that its inflectional variants rója, rója, and rójos will be chosen as nearest neighbors and influence it to receive penult stress. This is an undesirable state of affairs since the purpose of the study is to determine how successfully the model can assign stress to words that it is unfamiliar with. A simple way of controlling for this unwanted effect was to alphabetize the database prior to partitioning it for the ten-fold study. In this way, inflectional variants were grouped together in the same test set, and were unable to serve as analogs for each other.

Table 3. Examples of variable assignment

<table>
<thead>
<tr>
<th>Examples</th>
<th>Stress</th>
<th>Morphological variables</th>
<th>Phonemic variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>personal</td>
<td>Final</td>
<td>-- 0</td>
<td>personal</td>
</tr>
<tr>
<td>hablaron</td>
<td>Penult</td>
<td>6 pt pt pt</td>
<td>a-blancon</td>
</tr>
</tbody>
</table>
Table 4. Success rate on correctly assigning stress to database items

| Algorithm                        | #    | %    | $\chi^2$ | P <  
|----------------------------------|------|------|----------|------
| AM                              | 4693 | 94.4 |          |      
| TiBML-no weighting, 1 nn        | 4628 | 93.1 | 0.439    | 0.25 
| TiBML-no weighting, 2 nn        | 4565 | 91.8 | 1.742    | 0.25 
| TiBML-no weighting, 5 nn        | 4019 | 80.8 | 51.989   | 0.001
| TiBML-no weighting, 10 nn       | 3675 | 73.9 | 123.600  | 0.001
| TiBML-Information Gain, 1nn     | 4643 | 93.4 | 0.257    | 0.75 
| TiBML-MVDM, 1 nn                | 4688 | 94.3 | 0.002    | 0.9  
| TiBML-MVDM and Info. Gain, 1nn  | 4657 | 93.7 | 0.131    | 0.75 

MVDV = Modified Value Difference Metric; nn = number of nearest neighbors calculated

Once the database was partitioned, the stress placement of each word was determined in a ten-fold cross-validation. AM successfully assigned stress to 94.4% of the words in the database. This success rate is compared with those produced by TiMBL under the same experimental conditions tested in the study on Dutch stress assignment. Note that in the no weighting conditions using TiBML, and in the AM simulation, the tense form variable was included three times. It was only included once in the Information Gain and Modified Value Difference Metric simulations, since these algorithms are designed to calculate the importance of the variables and values on their own. In five of the seven experimental conditions, the success rates for the AM and TiMBL algorithms were statistically equivalent.

As previously mentioned, the database contained only the most frequently occurring Spanish words. It may be that extremely infrequent words have different stress patterns. To test this, a set of 497 words was assembled from among the items in Alameda and Cueto (1995) that had a frequency of 0.2 per million. The words in this test set were assigned stress in a ten-fold cross-validation study according to AM and TiMBL (Modified Value Difference Metric, one nearest neighbor). The resulting success rates were 91.8% and 90.2% respectively ($\chi^2 = .0603, p < .9$). It again appears that neither model may claim superiority over the other.

5.1 Error analysis

Given the similar success rates of both models, an analysis of the errors made by each model was performed in order to uncover any telling differences. The analysis compares AM with TiBML's most successful simulation (namely, MVDM) and calculates only one nearest neighbor. Table 5 specifies the number of errors made in each category, as well as the percentage of database items on which errors were made.
Table 5. Errors per category

<table>
<thead>
<tr>
<th></th>
<th>AM</th>
<th></th>
<th>TiBML-MVDM, 1 nm</th>
<th></th>
<th>(\chi^2)</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>%</td>
<td>#</td>
<td>%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penult</td>
<td>41</td>
<td>1.2</td>
<td>122</td>
<td>3.4</td>
<td>39.2638</td>
<td>0.001</td>
</tr>
<tr>
<td>Final</td>
<td>72</td>
<td>6.4</td>
<td>59</td>
<td>5.2</td>
<td>1.0992</td>
<td>0.5</td>
</tr>
<tr>
<td>Antepenult</td>
<td>164</td>
<td>59.9</td>
<td>101</td>
<td>36.9</td>
<td>14.5056</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Both models fare equally well on predicting final stress. However, TiBML proves more consistent in predicting antepenult stress, while AM is more adept at predicting penult stress. In terms of the percentage of errors per category, both models show the same hierarchy of difficulty: penult < final < antepenult. This is consistent with the hierarchy of difficulty that native Spanish speaking children demonstrated on a number of stress placement tasks (Hochberg 1988), and provides further evidence that the models' predictions have empirical value.

One test of the models' accuracy is the extent to which they have captured the classes of regularity and irregularity in the Spanish stress system. In Spanish, penult stress is regular (or unmarked) for words ending in a vowel or /s/; final stress is regular for words ending in any consonant except /s/; antepenult stress is always marked (see Eddington 2000). A model which captures this stress system would be expected to make most of its errors on words with irregular stress. Of the 282 errors made by TiBML, 156 (or 56%) occurred on irregularly stressed words. On the other hand, 80.1% (222 of 277) of the errors made by AM were made on irregularly stressed words.

The percentages just cited are interesting, but not indicative of a true difference between the models. It is important to ascertain, not only how many errors are made on irregular items, but the direction of the errors. That is, do the errors on the irregular items move stress onto the syllable which regularizes stress, or onto a syllable that keeps the word stress irregular. A model that correctly captures Spanish stress should also be expected to commit few errors that assign irregular stress to a word that is regularly stressed. In Hochberg's study (1988), children made more errors that regularized irregularly stressed words compared to errors that gave regularly stressed words an irregular stress.

Table 6 summarizes the rates of regularization and regularization produced by each model. The database contains 649 irregularly stressed words and 4177 words with regular stress. In calculating rates of regularization and irregularization, all 144 monosyllabic words were excluded.

As is evident in the data, AM appears to have more correctly captured the Spanish stress system. It imposes regularity on irregular items to a greater extent than TiBML. In addition, it assigns irregular stress to fewer regular items.
Table 6. Rates of regularization and irregularization

|                | AM  | TiMBL | $\chi^2$  | p <  
|----------------|-----|-------|-----------|------
| # Regularized  | 220 | 155   | 10.9226   | 0.001
| % Regularized  | 33.9| 23.9  |           |      
| # Irregularized| 54  | 127   | 28.6408   | 0.001
| % Irregularized| 1.3 | 3.0   |           |      

6. Conclusions

The purpose of this study was to compare TiMBL and AM on a number of different tasks. Neither model significantly outperformed the other in the gender assignment and diminutive assignment tasks. In the previous comparison by Daelemans et al. (1994b), AM outperformed TiMBL on a Dutch stress assignment task, except when noise was added to the system they performed equally well. The present study pitted the two models against each other in terms of their ability to assign stress to Spanish words. Both models were able to correctly assign stress to the most frequent 4970 Spanish words with about a 94% degree of accuracy. Their performance on highly infrequent words was slightly lower, but neither model was able to statistically outperform the other on either of these tasks. The only differences were evident in the error analysis. AM applied the regular stress patterns to irregularly stressed words to a greater extent than TiMBL. TiBML, on the other hand, had a higher incidence of misapplying irregular stress patterns to regularly stressed items. This indicates that AM more successfully captured patterns of regularity and irregularity in the Spanish stress system.

Notes

* This study was carried out with the help of a grant from the National Science Foundation (#00821950).
1. The current study is based on an earlier pre-print version of LEXESP, a morphologically tagged frequency dictionary of Spanish of about 3 million words. A more recent printed version is based on a 5 million word corpus (Sebastián, Martí, Carreiras, & Cuetos 2002).
2. I am most indebted to Milagros Malo Fernández and Elías Álvarez Ortigosa who generously gave of their time to administer the questionnaires.
3. Alameda and Cuetos (1995); Sebastián, Martí, Carreiras, and Cuetos (2002); Marcos Marín (no date a, no date b). In addition to these sources, Mark Davies of Illinois State University graciously provided me with the diminutive forms from his corpus project totaling 39.8 million words: <http://mdavies.fort.ilstu.edu/personal/texts.htm>.
4. In some words from groups 11 and 12, \( s \) represents what seems to be the plural morpheme since it appears word finally and follows a stressless vowel. In other cases, such as cumpleaños, the word ends in the plural morpheme derivationally speaking (cumple + años 'complete + years'), but is used to denote both the plural and singular.

References


Appendix

Stimulus materials

abarraz    canez    estruz
acates     carauz    evagación
acemite    cección    fabledad
acordación celtre     fenestraje
acumen     cifaque    fluición
afer       cipión     folguin
afice      cobil      fosal
alancel     coce      gañez
alcaduz    cocadriz    gagate
alcalafaje compage    gariñalte
alcamiz    consuetud    grasor
aliinde     consultaje    guiblete
alioj      copanete    guiage
alizace    cotrofe     ingre
amarillor  crenche     jusente
anascote    criażón     lailán
arrafiz     crochel     lande
asperez    chivil        lavajal
atarfe     delate      lerdez
azaríentez desdón      linamen
azción      deslante    mandrial
azoche      destín      mansuetud
balizaje    disfrez     másticis
barrunte    egestión    menge
beudez     elebor      meridión
bitumen    emiente     merode
bocacín    entalle     nacre
botor      entenzón    orebce
broznedad  epiglosis    palude
-cación    escambrón    panol
cabrial    escorche    perale
cafiz      escrocón    pernicie
calicud    esgambete    polex
calonge    esguarda    primaz
cambil     esledor     }
cancelor     estipe
<table>
<thead>
<tr>
<th>pujés</th>
<th>senojil</th>
<th>trascol</th>
</tr>
</thead>
<tbody>
<tr>
<td>realme</td>
<td>sorce</td>
<td>velambre</td>
</tr>
<tr>
<td>rebalaj</td>
<td>sozprior</td>
<td>venadriz</td>
</tr>
<tr>
<td>riste</td>
<td>tabelión</td>
<td>venderache</td>
</tr>
</tbody>
</table>