2001

A Usage-based Simulation of Spanish S-weakening

David Eddington
Brigham Young University, eddington@byu.edu

Original Publication Citation

2001. "A Usage-based Simulation of Spanish S-weakening." Italian Journal of Linguistics 13.191-209.

A Usage-based Simulation of Spanish S-weakening

David Eddington

Version 1-13-2002

Abstract

The pronunciation of Spanish /s/ as an aspirate [h] (or its deletion) is studied in a usage-based simulation that assumes that prior linguistic experiences are stored as memory traces, and that these stored experiences, instead of generalizations abstracted from them, are the source of the knowledge speakers have of their language. This exploratory simulation was carried out using the algorithm of Analogical Modeling of Language (Skousen 1989, 1992).

An initial database of utterances, taken from samples of actual speech containing [s] or one of its weakened variants (n=2353), served as an approximation of past experience with the phoneme /s/. The purpose of the simulation was to analogize on the initial database in order to determine whether new instances of /s/ in a test set would be given a full fricative pronunciation [s], or a weakened pronunciation ([h] or [i]). Analogy to previous experience is assumed to underlie Spanish speakers knowledge of this alternation. Once the phonetic realization of all instances of /s/ in a test set was determined, that test set was added to the initial database since those instances could then be considered part of linguistic experience. This larger database of past experience was then used to determine the phonetic realization of the occurrences of /s/ in the next test set. Ten test sets of 1924 instances of /s/ were used. The simulation mirrors the way in which newly processed exemplars can influence subsequent processing and language change.

The results of the simulation mirror a number of facts about s-weakening. First, it demonstrates the well-known fact that phonetic context plays an important role in the alternation, and it does so without making any sort of rule-like generalization. Second, it models the high degree of variability that has been described in the literature on the subject. Third, it shows how certain lexical items can begin to form associations with certain realizations of /s/ (i.e. lexical diffusion) even while phonetic conditioning of s-weakening continues to be a productive force.

1. *Introduction.* The weakening of Spanish /s/ into [h] or [ɨ], which generally occurs in word final and syllable final position, has been the subject of an extremely large body of linguistic literature, a review of which would occupy an entire paper itself. Nevertheless, various aspects of weakening have been explored including the possible phonetic and perceptual causes for the weakening (e.g. Romero 1994-5; Widdison 1995a, b), and the chronology of its appearance in the Spanish language (e.g. Torreblanca 1987). Other studies have focused on its role in the phonological system of various dialects (e.g. Amastae 1989; Lipski 1983; Terrell 1978, 1980, 1995), as well as on the sociolinguistic variables (e.g. Garcia & Tallon 1995, Valdivieso & Magana 1979, 1988), and functional variables (e.g. Hundley 1987; Poplack 1980, 1984; Ranson 1992) that influence its occurrence. Weakening has also been described using the notational mechanisms of a number of formal approaches (e.g. Goldsmith 1981; Guitart 1981; Harris 1983; Lipski 1999).

It is fairly uncontroversial to state that part of learning a language natively involves acquiring the articulatory patterns and habits of the surrounding community. S-weakening is one of the patterns many Spanish speakers learn, but exactly how such habits and patterns are represented has been hotly debated in past decades. Some argue that generalizations are gleaned from linguistic input and actively used in language processing (e.g. Pinker 1991, 1997, 1999; Pinker & Prince 1988, 1994). This sort of linguistic knowledge[1] has been represented as both rules and rule orderings, and more recently as constraints and constraint rankings (McCarthy & Prince 1993).

The purpose of the present study differs from extant accounts of s-weakening. What I will present is a usage-based simulation of s-weakening. I will argue from the viewpoint that prior linguistic experiences (memory traces, exemplars, or instances) are stored, and that these stored experiences, instead of generalizations abstracted from them, are the source of the knowledge speakers have of their native language. Moreover, linguistic experience is not static, but is continually added to. I will attempt to demonstrate that linguistic processing may be attributed to the dynamic quality of the mind's repository of stored linguistic experiences. To this end I performed an analogical simulation that utilized stored exemplars of past linguistic experience. The results of the simulation mirror a number of facts about s-weakening; it is partially conditioned by phonetic context, but also demonstrates a great deal of variability. In the simulation, weakening can be seen to be at the stage of diachronic development in which phonetic conditioning is still a major factor, but at the same time, certain variants are beginning to become associated with certain lexical items regardless of the phonetic context in which they appear.

2. *Exemplar-based Processing.* Accounting for phonological processes has, for the most part, followed the rule and constraint paradigm. Nevertheless, the psychological implausibility of such analyses, (along with the fact that these analyses are thought to represent the competence of an ideal speaker-hearer, and not the performance of actual speaker-hearers), has lead some researchers toward accounts that more closely model how phonological patterns may be stored and processed (Bybee 1985, 1988, 1998, 1994, 2001; Daelemans, Zavrel van der Sloot, & van den Bosch 1999; McClelland 1988; Pierrehumbert 2001; Stemberger 1994). What these models share is the idea that linguistic knowledge is not rule or constraint-based, but that it emerges through language use. For example, a number of researchers have found that phonotactic

constraints and co-occurrence of phonemes is gradient and probabilistic (Frisch 2000; Frisch, Large, & Pisoni 2000; Frisch, Large, Zawaydeh & Pisoni 2001; Frisch & Zawaydeh 2001; Pierrehumbert 1994; Treiman, Kessler, Knewasser, Tincoff, & Bowman 2000). Such gradience is incompatible with rule-based accounts of phonotactics, but supports usage-based models in which frequency of occurrence is a crucial factor.

One way to account for such effects is to assume that each time a word is processed, it leave a memory trace in the mind. There is evidence to support the idea that words are stored with a great deal of detailed redundant phonetic information (Brown & McNeill 1966; Bybee 1994; Palmeri, Goldinger & Pisoni, 1993; Goldinger 1997; Pisoni 1997). This contrasts with the traditional conception of a lexicon containing only a single entry for each word which includes only unpredictable phonetic characteristics. In addition to including detailed phonetic information, words, especially high frequency words, appear to be stored as morphologically undecomposed wholes, as some have suggested (Baayen 1994; Butterworth 1983; Bybee 1985, 1988, 1998, 2001; Halle 1973; Jackendoff 1975 Stemberger 1994). The psychological literature is not without empirical evidence to support the kind of massive lexical storage such models entail (e.g. Alegre & Gordon 1999; Baayen, Dijkstra, & Schreuder 1997; Bybee 1995; Manelis & Tharp 1977; Sereno & Jongman 1997).

2.1. *Evidence for Exemplar Processing.* The existence of stored exemplars opens up the possibility that past linguistic experience may play an active role in language processing. Such exemplar-based processing may be termed analogy. The ability of analogy to model language processing has been demonstrated in simulations of several different linguistic phenomena, such as predicting linking elements in Dutch noun compounds (Krott in press), accounting for a particular phonological alternation in Turkish stems (Rytting 2000), and assigning stress to Dutch words (Daelemans et al. 1994). However, a more crucial test of the psychological import of analogy is that it mirrors actual language usage phenomenon. For example, an analogical simulation of Spanish stress assignment (Eddington 2000b) not only accounts for the phenomenon better than an empirically testable rule approach, but makes the same sort of errors as children do. In addition, it helps explain the different patterns of errors made by children of different ages.

Evidence from analogy has also come to bear on the debate between those who assume that a single system can handle both regular and irregular inflections (e.g. Daugherty & Seidenberg 1994), and those who espouse separate mechanisms for regular and irregular inflection (e.g. Pinker & Prince 1994). For instance, in a study of English past tense formation (Prasada & Pinker 1993), subjects were asked to inflect nonce words. The results of this experiment were originally interpreted as support of the dual-route model. However, an analogical simulation of the same nonce items (Eddington 2000a) mirrored the subjects responses closely, demonstrating that a single-route model is equally capable of accounting for the test subjects' intuitions.[2] In a similar study, Say & Clahsen (2001) asked Italian speakers to determine what past participle, (and hence conjugation), they would assign nonce verbs. Say & Clahsen claim that their findings favor a dual-route model for Italian verbal stem formation. However, the nonce words were assigned conjugational class in an analogical simulation of the data, which again closely mirrored the responses of the test subjects (Eddington 2002).

3. *Analogical Algorithm.* The analogical algorithm that will be used in the present study is the one incorporated in Analogical Modeling of Language or AML (Skousen 1989, 1992, 1995, 1998). AML is designed to account for linguistic processing by referring to memory tokens of past linguistic experience. In this way, it is similar to several other exemplar-based models (Aha, Kibler, & Albert 1991; Medin & Schaffer 1978; Pierrehumbert 2001; Riesbeck & Schank 1989; see Shanks 1995 for an overview of exemplar models; see Daelemans, Gillis, & Durieux 1994 for a comparison of AML & Aha et al.). Readers who are interested in the specifics of AML's algorithm are invited to consult the works by Skousen cited above.[3]

In essence, AML works by searching a database that represents a speaker's past experience with some phenomenon. In the present study, the database represents cases in which /s/ is given its full fricative pronunciation, and cases in which it is given a weakened pronunciation ([h] or [i]). These possible outcomes are termed behaviors. Given a test case, that is, a set of variables representing /s/ in a particular word, and in a specific morphological and phonetic context, AML searches the database to find similarities between the test case and the items in the database. When an exact match is found, AML calculates a 100% probability that the behavior of the item found in the database will be applied to the test case. However, this does not mean that the exact match from the database is necessarily the only item affecting the outcome. If three exact matches are found, for example, and two demonstrate the same s-weakening behavior, while one appears with [s], the probability that the test case will be given a weakened pronunciation is 66%. There would only be a 33% that the test case would be [s].

When no exact matches are encountered, the algorithm calculates the behavior based on items that share characteristics with the test case. More similar items will generally affect the behavior more than less similar ones. In AML, the probability that a behavior will apply is given as a percentage. Under certain conditions, items that share characteristics with the test item may be eliminated from consideration as possible analogs, or their influence on the behavior assigned to the test item may be diminished. This may occur when two or more database items bear the same similarities to the test item, but have different behaviors (i.e. category variables, see section 4.2). Note that the algorithm proceeds in the same fashion whether or not exact matches are found.

3.1. *Criteria for Selection of Behavior.* There are essentially two types of behavior: deterministic and variational. Deterministic behavior involves a choice between an essentially correct and an incorrect outcome. For example, if the task entails choosing a definite article given the noun *casas* 'houses', the correct choice is *las* while *el, los*, or *la* would be incorrect by any definition. However, s-weakening is not deterministic, but variational. A weakened variant can appear in any position, including word and syllable inital position in some dialects. Exactly where and under what conditions weakening occurs varies a great deal, but its occurrence, of lack thereof, cannot be deemed correct or incorrect.

In a simulation of a deterministic phenomenon, the behavior whose occurrence AML predicts at the highest level of probability is normally thought to apply. For a variational behavior, on the other hand, the calculated probabilities have a different interpretation. This is especially true in the case of s-weakening since its behavior is dependent on a number of sociolinguistic factors which have been intentionally omitted from the simulation. In eliminating sociolinguistic factors, I in no way want to give the impression that they are not an extremely important factor in governing the alternation between [s] and its weakened varieties. It is simply that the goal of the study is not to evaluate the role of sociolinguistic factors, but to present a highly simplified simulation of the progression of s-weakening. Therefore, in the speculative simulation I describe below, I assume that whatever combination of sociolinguistic factors leads to the greatest degree of weakening is what applies in the simulation. This a priori assumption is important because it is one way of artificially accelerating the rate of change in a simulation designed to evaluate how change may progress.

Some way of accelerating the rate of change is essential for another reason also. Language changes typically take place over extended periods of time. That is, a phonetic evolution occurs as speakers process hundreds of thousands of instances of the phone in question. A simulation of that magnitude is neither possible nor practical. In reality, little change would be expected when only a few thousand test items are processed and added to the mental storehouse of past experience, as in the present study.

Therefore, in order to represent the most weakening prone sociolinguistic state of affairs and accelerate the rate of change for the purposes of the simulation, a probability of 30% was established for weakening. That is, whenever AML calculated a 30% or greater the probability that s-weakening would occur, it was assumed to occur. In this way, an instance of the test item *comes mucho* whose /s/ has a 30% or greater probability of being given a weakened pronunciation, is added to the database from which analogs will be drawn for the next test set, with a category variable [weakened]. In actuality, a cut-off point of 30% is not a undue manipulation of the data given the fact that the simulation involves a variational and not a deterministic behavior.

4. *Overview of the Simulation.* The present simulation was inspired by Kirchner's (no date) analogical simulation of Tigrinya spirantization. It begins with two databases. The first is an initial database of stored exemplars that represents past linguistic experience. These exemplars were drawn from a phonetic transcription of actually occurring natural language data. This initial database contains information about the specific phonological and morphological contexts in which many different occurrences of /s/ or its weakened variants appeared in natural discourse. The other database contains ten test sets, which were also taken from a corpus of utterances, and that also gives the same contextual information about /s/ as the in the initial database. The test sets differ from the initial database in that the actual phonetic realization of the instances of /s/ are not known.

The purpose of the simulation was to use the information in the initial database to determine whether the instances of /s/ in the test sets would be realized as /s/ or as a weakened variant. Once this determination was made, the first test set, whose pronunciations of /s/ were determined by the algorithm, was added to the initial database since the members of the first test set could then be considered part of linguistic experience. This larger database of past experience was then used to determine the phonetic realization of the occurrences of /s/ in the next test set. A total of ten test sets, containing the same number of instances of /s/, was applied to an ever growing database of past experience.

4.1. *Databases and Variables*. The initial database for the simulation was taken from phonetically transcribed interviews of two native speakers of the Spanish dialect of Lima Peru (Carvedo 1989). This particular dialect was chosen because, in contrast to Carribean dialects, for example, the rates of deletion and aspiration are relatively low. One speaker was a 26 year old female university student, and the other a 43 year old male lawyer. A total of 2353 instances of [s], or one of its variants [h] or [i] were taken from the two interviews.

Information about the context in which [s], [h], or [i] appeared was converted into sets of variables. The category variable had one of two specifications: whether /s/ was given its full unreduced pronunciation, or whether it had been given a weakened pronunciation ([h] or [i]).[4] The two elements to the left and right of the phone in question were included as contextual variables. These variables contained either the phonemes in these positions, or indicated that a word boundary or pause appeared. Additional variables indicated whether each phoneme was a consonant or a vowel, the part of speech of the word, and whether the /s/ was morphemic or not. Another variable indicated the word that /s/ or its weakened counterparts appeared in. For example, *come[h] mucho* ‘you eat a lot’ would yield the following category variable and 11 contextual variables:

Category Variable: /s/ is weakened

Contextual Variables: 1-An /m/ appears two positions to the left of the weakened /s/.

2-The /m/ is a consonant.

3-An /e/ appears to the immediate left of the weakened /s/.

4-The /e/ is a vowel.

5-A word boundary appears to the immediate right of the weakened /s/

6-The word boundary is not a phoneme so it is not specified as being either a vowel or a consonant.

7-An /m/ appears two slots to the right of the weakened /s/.

8-The /m/ is a consonant.

9-The word *comes* is a verb.

10-The weakened /s/ is a morpheme (2nd person singular).

11-The phone in question appears in the word *comes*.

Ten test sets of 1924 instances of orthographic /s/ were constructed. The test sets were taken from a corpus of orthographically transcribed oral speech (Marcos Marín no date). These instances were encoded using the same 11 contextual variables as in the initial database. However, they were left unspecified as far as the category variable is concerned for two reasons. First, the orthographic transcription does not allow the phonetic realization of /s/ to be specified. Second, the purpose of the simulation is to predict the category variable ([s] or weakened variant) of the items in the test set.

4.2. Results of the Simulation. S-weakening was considered in six different contexts:

1-Word-initial (#s)

2-Word-internal followed by a consonant (sC)

3-Word-internal followed by a vowel (sV)

4-Word-final followed by a pause[5] (s##)

5-Word-final followed by a vowel (s#V)

6-Word-final followed by a consonant (s#C)

For each of the six contexts, Figure 1 indicates the percentages of /s/ (n=2,353) that were given weakened pronunciations in the initial database, and the percentages of /s/ (n=21,593) that were given weakened pronunciations in the final database, once the ten sets of test items had been run.

++Insert Figure 1 here++

Figure 1. Percent of weakening in initial and final databases.

As can be seen, in word initial position no weakening occurred. The major gains in weakening are found in word-final position when followed by a consonant. The largest decrease in weakening occurs in word-internal position followed by a consonant. Changes in the remainder of the contexts are small and it is difficult to determine if they are significant. Slight increases are seen in word-final position followed by a pause, and word-internal position followed by a vowel. In the latter case, however, the increase is due entirely to several instances of the word-medial /s/ in *nosotros* being weakened. In word-final position followed by a vowel, a small decrease in weakening is registered.

At this point it is important to place the results of the simulation in its proper context. Do the results of the simulation purport to predict the linguistic future of the Spanish of Lima? There are several reasons why the answer must be no. First, the simulation was performed with a few thousand exemplars, which falls fall short of the millions of exemplars that would have to represent the language experience of a native speaker. Second, the initial database was derived from only two speakers, which is hardly a representative sample of the *limeño* dialect. Third, in the simulation no attempt was made to include sociolinguistic variables that have been shown to be extremely influential in Spanish s-weakening.

In that case, what do the results of such a speculative simulation indicate? The purpose of performing this simulation is to demonstrate how an exemplar-based model of language use is able to account for certain facts about s-weakening. Perhaps the most consistent finding in the literature is that weakening is partially conditioned by phonetic context. The effect of phonetic context is evident in the outcome of the simulation; words processed during the simulation show a strong tendency to weaken before consonants and to retain the full fricative [s] before vowels. Before a pause there is also a tendency toward [s]. This sort of phonetic conditioning is found whether or not the initial database contained instances of the word in question (see Table 1). That is, analogy to existing cases of words in the initial database cannot account for all of the phonetic conditioning observed.

++Insert Table 1 here++

Table 1. Examples of Phonetic Conditioning.

Although phonetic conditioning is evident in the vast majority of the words processed in the simulation, it is by no means categorical. In fact, another characteristic of s-weakening is that, in spite of phonetic conditioning, there is a good deal of variability regarding whether [s], or [h]/[i] appear in a given word in a given context. The simulation also demonstrates the existence of this sort of variability; Table 2 contains some examples of words ending in /s/, that show variability in the phonetic realization of /s/.

++Insert Table 2 here++

Table 2. Examples of Variability between [s] and Weakened Pronunciations.

Perhaps the most interesting outcome of this simulation is that it demonstrates that even while phonetic conditioning continues to productively govern the alternation between /s/ and its weakenend variants, this does not preclude the possibility of lexical diffusion, that is, that specific lexical items may, at the same time, be in the process of becoming uniquely associated with one member of the alternation. For example, consider the words *desde, después, esfera, España,* and *estoy* from Table 3.

++Insert Table 3 here++

Table 3. Examples of Specific Realizations of Word-internal /s/ Becoming Lexicalized.

It is not surprising to find that /s/ is weakened before the following consonant in these words. Although not all of the instances of these words in the initial database contained weakened varieties of /s/, the final database indicates that what once was variability has settled into stability. More interesting are the cases of lexical diffusion that run counter to the general tendencies predicted by phonetic context. The word internal /s/ in *estamos* and *estás,* for example, has moved in the opposite direction; these words are in the process of becoming uniquely associated with an unweakened /s/.

Nosotros presents another interesting case in the present study. In the initial database, four instances of [nosotros] alternated with two instances of [nohotros], but all 47 of the new test cases of *nosotros* that were processed during the simulation yielded a weakened pronunciation of the word-internal /s/. In other words, the weakened /s/ version is beginning to dominate in the lexical representation of the word *nosotros*.

Words containing an /s/ in word-final position provide the best evidence for lexical diffusion, or the emerging lexical association of one pronunciation over another. This is true because the phonetic context in which [s], [h], or [i] appears is not constant as it is in words with word internal /s/ such as *estatua*. Therefore, words whose final /s/ demonstrates little variation, regardless of what phonetic segment follows, are words in which a specific pronunciation is becoming an integral part of its lexical representation, or in other words, is becoming normative.

++Insert Table 4 here++

Table 4. Examples of Lexically Associated [s].

In the present simulation, *años, dejas, dices, eres, otros,* and *veces* (see Table 4) appear with final [s] regardless of phonetic context. Table5 on the other hand, shows that the lexical representation of *muchas, todos, todas,* and *unas* is moving toward a weakened pronunciation of /s/.

++Insert Table 5 here++

Table 5. Examples of Lexically Associated [h, i].

5. *Conclusions.* I will once again reiterate that the simulation is based on an extremely small data set, and on a very limited number of variables and test cases. For this reason, the outcome of the simulation serves only to demonstrate, in principle, how an exemplar-based model may account for language variability and change. It does not purport to make specific predictions about the exact way individual words or dialects may evolve in the future. Nevertheless, the simulation proves adept at capturing certain generalizations about s-weakening. First, it captures the fact that phonetic context plays an important role in the alternation. Second, it shows the high degree of variability that has been described in the literature on the subject. Third, it demonstrates how lexical associations may form for individual lexical items even while phonetic conditioning of s-weakening continues to be a productive force.

Bibliographical References

Aha, David W., Dennis Kibler, & Marc K. Albert (1991), “Instance-based learning algorithms”, *Machine Learning* 6:37-66.

Alegre, Maria, & Peter Gordon (1999), “Frequency effects and the representational status of regular inflections”, *Journal of Memory and Language* 40:41-61.

Amastae, Jon (1989), “The intersection of s-aspiration/deletion and spirantization in Honduran Spanish”, *Language Variation and Change* 1:169-183.

Baayen, R. Harald (1994), “Productivity in language production”, *Language and Cognitive Processes* 9:447-469.

Baayen, R. Harald (1995), Review of *Analogy and Structure*, by Royal Skousen. *Language*71: 390-396.

Baayen, R. Harald, Ton Dijkstra, & Robert Schreuder (1997), “Singulars and plurals in Dutch: Evidence for a parallel dual-route model”, *Journal of Memory and Language* 37:94-117.

Broe, Michael B. & Janet. P. Pierrehumbert, eds. (2000), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, Cambridge, Cambridge University Press.

Brown, Roger and David Mc Neill (1966), “The ‘tip of the tongue’ phenomenon”, *Journal of Verbal Learning and Verbal Behavior* 5:325-337.

Brown, Esther L., & Rena Torres Cacoullos (2002), “¿Qué le vamoh aher: Taking the syllable out of Spanish /s/-reduction”, Ten Working Papers in Linguistics (PWPL) 8.3: Selected Papers from NWAVE 30, in press.

Burton-Roberts, Noel, Philip Carr, & Gerard Docherty, eds. (2000), *Phonological Knowledge: Conceptual and Empirical Issues*, Oxford, Oxford UP.

Butterworth, B., ed, (1983), *Language production*, vol. 2, London, Academic Press.

Butterworth, B. (1983), "Lexical representation", in Butterworth (1983:257-294).

Bybee, Joan (1985), *Morphology*. Amsterdam, John Benjamins.

Bybee, Joan (1988), "Morphology as lexical organization", in Hammond and Noonan (1988:119-141).

Bybee, Joan (1994), "A view of phonology from a cognitive and functional perspective", *Cognitive Linguistics* 5:285-305.

Bybee, Joan (1995), " Regular morphology and the lexicon", *Language and Cognitive Processes* 10:425-55.

Bybee, Joan (1998), "The emergent lexicon", in Gruber, Higgins, Olson & Wysocki (1998:421-435).

Bybee, Joan (2001), *Phonology and Language Use*. Cambridge, Cambridge UP.

Bybee, Joan, & Paul Hopper, eds (2001), *Frequency and the Emergence of Linguistic Structure*, Amsterdam, Benjamins.

Caravedo, Rocío (1989), *El español de Lima: Materiales para el estudio del habla culta*, Lima, Pontificia Universidad Católica del Perú.

Carr, Philip (2000), "Scientific realism, sociophonetic variation, and innate endowment in phonology", in Burton-Roberts, Carr, & Docherty (2000:67-104).

Chomsky Noam, & Morris Halle (1968), *The Sound Pattern of English*. New York, Harper and Row.

Cohen, David & Jessica R. Wirth, eds. (1975), *Testing Linguistic Hypotheses*, Washington D. C., Hemisphere.

Cressey, William W. & Donna Jo Napoli, eds. (1981), *Linguistic Symposium on Romance Languages*, Washington D. C., Georgetown UP.

Daugherty, Kim, & Mark S. Seidenberg (1994), “Beyond rules and exceptions: A Connectionist approach to inflectional morphology”, in Lima, Corrigan, & Iverson (1994:353-388).

Daelemans, Walter, Jakub Zavrel, Ko van der Sloot, & Antal van den Bosch (1999), TiMBL: Tilburg memory based learner, version 2.0, reference guide. Induction of Linguistic Knowledge Technical Report. Tilburg, Netherlands: ILK Research Group, Tilburg University. (Http://ilk.kub.nl/~ilk/papers/ilk9901.ps.gz).

Daelemans, Walter, Steven Gillis, & Gert Durieux (1994), “Skousen's analogical modeling algorithm: A comparison with lazy learning”, in Jones (1994:1-7).

Eddington, David (1996), “The psychological status of phonological analyses”, *Linguistica* 31:17-37.

Eddington, David (2000a), “Analogy and the dual-route model of morphology”, *Lingua* 110:281-298.

Eddington, David (2000b), “Spanish stress assignment within the Analogical Modeling of Language”, *Language* 76:92-109.

Eddington, David (2002), “Dissociation in Italian conjugations: A single-route account”, *Brain and Language*. In press.

Frisch, Stefan (2000), “Temporally organized lexical representations as phonological units”, in Broe and Pierrehumbert (2000:283-298).

Frisch, Stefan. A., Nathan R. Large, & David B. Pisoni (2000), “Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords”, *Journal of*

Memory and Language 42:481-496.

Frisch, Stefan A., Nathan R. Large, Bushra Zawaydeh, & David B. Pisoni (2001), "Emergent phonotactic generalizations in English and Arabic", in Bybee and Hooper (2001:159-179).

Frisch, Stefan A., & Bushra Zawaydeh (2001), "The psychological reality of OCP-place in Arabic", *Language* 77:91-106.

García, Mary Ellen, & Michael Tallon (1995), "Postnuclear /-s/ in San Antonio Spanish: Nohotros no aspiramos", *The Georgetown Journal of Languages and Linguistics* 3,139-162.

Goldinger, Stephen D. (1997), "Words and Voices: Perception and production in an episodic lexicon", in Johnson and Mullennix (1997:33-65).

Goldsmith, John (1981), "Subsegmentals in Spanish phonology", in Cressey and Napoli (1981:1-16).

Gruber, M. Catherine, Derrick Higgins, Kenneth S. Olson, & Tamra Wysocki, eds. (1998), *Proceedings of the Chicago Linguistic Society*, vol. 34, Chicago, Chicago Linguistic Society.

Guitart, Jorge M. (1981) "En torno a la sílaba como entidad fonemática en los dialectos del caribe hispano" *Thesaurus* 36:457-463.

Gutiérrez-Rexach, Javier & Fernando Martínez-Gil, eds. (1999), *Advances in Hispanic Linguistics*, Somerville, MA, Cascadilla Press.

Halle, Morris. (1973), "Prolegomena to a theory of word formation", *Linguistic Inquiry* 4:3-16.

Hammond, Michael, & Michael Noonan, eds. (1988), *Theoretical Approaches to Morphology*, San Diego, Academic Press.

Hare, Mary., Elman, Jeff L., & Daugherty, Kim D. (1995), “Default generalization in connectionist networks”, *Language and Cognitive Processes* 10:601-630.

Harris, James (1983), *Syllabic Structure and Stress in Spanish*, Cambridge, MA, MIT Press.

Hashemipour, Peggy, Ricardo Maldonado, & Margaret van Naerssen, eds. (1995), *Studies in Language Learning and Spanish Linguistics,* New York, McGraw-Hill.

Hundley, James E. (1987), “ Functional constraints on plural marker deletion in peruvian Spanish”, *Hispania* 70:891-895.

Jackendoff, Ray (1975), “Morphological and semantic regularities in the lexicon”, *Language* 51:639-671.

Johnson, Keith & John W. Mullennix, eds. (1997), *Talker Variability in Speech Processing*, San Diego, Academic.

Jones, D. ed. (1994), *Proceedings of the International Conference on New Methods in Language Processing*, Manchester, MIST.

Keating, Patricia A., ed. (1994), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, Cambridge, Cambridge University Press.

Kager, René (1999), *Optimality Theory*, Cambridge, Cambridge UP.

Kiparsky, Paul (1975), “ What are phonological theories about?”, in Cohen and Wirth (1975:187-209).

Kiparsky, Paul (1982), “Lexical morphology and phonology”, in The Linguistic Society of Korea, (1982:3-89).

Kirchner, Robert (no date), “Preliminary thoughts on “phonologization” within an exemplar-based speech processing system", Rutgers Optimality Archive #320 (http://ruccs.rutgers.edu/roa.html).

Krott, Andrea (in press), "Analogical hierarchy: Exemplar-based modeling of linkers in Dutch noun-noun compounds", in Skousen, Lonsdale, & Parkinson.

Lima, Susan D., Roberta L. Corrigan, & Gregory K. Iverson, eds. (1994), *The Reality of Linguistic Rules*, Amsterdam, Benjamins.

Linguistic Society of Korea (1982), *Linguistics in the Morning Calm*, Seoul, Hanshin Publishing.

Lipski, John M. (1983), "Reducción de /s/ en el español de Honduras", *Nueva Revista de Filología Hispánica* 32:272-288.

Lipski, John (1999), "The many faces of Spanish /s/-weakening: (Re) alignment and Ambisyllabicity", in Gutiérrez-Rexach & Martínez-Gil, (1999:198-213).

Manelis, Leon & David A. Tharp (1977), "The processing of affixed words", *Memory and Cognition* 5:690-695.

Marcos Marín, Francisco (no date), *Corpus oral de referencia del español contemporáneo.* Textual corpus, Universidad Autónoma de Madrid. Http://elvira.lllf.uam.es/docs_es/corpus/corpus.html.

McCarthy, John, & Alan Prince (1993), *Prosodic morphology I*. Technical reports of the Rutgers University Center for Cognitive Science.

McClelland, James L. (1988), "Connectionist models and psychological evidence", *Journal of Memory and Language* 27:107-123.

Medin, Douglas L. & Marguerite M. Schaffer (1978), "Context theory of classification learning", *Psychological Review* 85:207-238.

Melby, Alan K. & Arle R. Lommel, eds. (2000), *Lacus Forum XVII: The Lexicon*, Fullerton, CA, Linguistic Association of Canada and the United States.

Nooteboom, Sieb, Fred Weerman, & Frank Wijnen, eds. (2001), *Storage and Computation in the Language Faculty*, Dordrecht, Kluwer.

Palmeri, Thomas J., Stephen D. Goldinger, & David B. Pisoni (1993), "Episodic encoding of voice attributes and recognition memory for spoken words", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:309-28.

Pierrehumbert, Janet (1994), "Syllable structure and word structure: A study of triconsonantal clusters in English", in Keating, (1994:168-188).

Pierrehumbert, Janet (2001), "Exemplar dynamics: Word frequency, lenition and contrast", in Bybee and Hopper (2001:137-157).

Pinker, Steven (1991), "Rules of language", *Science* 253:530-534.

Pinker, Steven (1997), "Words and rules in the human brain", *Nature* 387:547-548.

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

Pinker, Steven , & Alan Prince (1988), "On language and connectionism: Analysis of a parallel distributed processing model of language acquisition", *Cognition* 28:73-193.

Pinker, Steven, & Alan Prince, (1994), "Regular and irregular morphology and the psychological status of rules of grammar", in Lima, Corrigan, & Iverson, (1994:321-351).

Pisoni, David (1997), "Some thoughts on 'normalization' in speech perception", in Johnson & Mullennix, (1997:9-32).

Poplack, Shana (1980), "Deletion and disambiguation in Puerto Rican Spanish", *Language*

56:371-385.

Poplack, Shana (1984), “Variable concord and sentential plural marking in Puerto Rican Spanish”, *Hispanic Review* 52:205-222.

Prasada, Sandeep, & Steven Pinker (1993), “Generalisation of regular and irregular morphological patterns”, *Language and Cognitive Processes* 8:1-56.

Ranson, Diana L. (1992), “Nominal number marking in Andalusian Spanish in the wake of /s/ deletion”, *Hispanic Linguistics* 4:301-327.

Riesbeck, Chris K., & Roger S. Schank (1989), *Inside Case-based Reasoning* Hillsdale, N.J., Erlbaum.

Rytting, C. Anton (2000), “An empirical test of analogical modeling: The /k/ ~ i alternation”, in Melby & Lommel, (2000:73-84).

Romero, Joaquín (1994-1995), “An articulatory view of historical s-aspiration in Spanish”, *Haskins Laboratories Status Report on Speech Research* 119-120:255-266.

Say, Tessa, & Harald Clahsen (2001), “Words, rules, and stems in the Italian mental lexicon”, in Nooteboom, Weerman, & Wijnen, (2001:75-108).

Sereno, Joan A., & Allard Jongman (1997), “Processing of English inflectional morphology”, *Memory and Cognition* 25:425-37.

Shanks, David R. (1995), *The Psychology of Associative Learning*, Cambridge, Cambridge UP.

Skousen, Royal. (1989), *Analogical Modeling of Language*, Dordrecht, Kluwer Academic.

Skousen, Royal (1992), *Analogy and Structure*, Dordrecht, Kluwer Academic.

Skousen, Royal (1995), “Analogy: A non-rule alternative to neural networks”, *Rivista di Linguistica* 7:213-232.

Skousen, Royal (1998), “Natural statistics in language modeling”, *Journal of Quantitative Linguistics* 5:246-255.

Skousen, Royal, Deryle Lonsdale & Dilworth Parkinson, eds. (In press), *Analogical Modeling: An Exemplar-based Approach to Language*, Amsterdam, John Benjamins.

Stemberger, Joseph P. (1994). Rule-less morphology at the phonology-lexicon interface. The reality of linguistic rules, ed. by Susan D. Lima, Roberta L. Corrigan, & Gregory K. Iverson, 147-169. Amsterdam: Benjamins.

Terrell, Tracy D. (1978), “Sobre la aspiración y elisión de /s/ implosiva y final en el español de Puerto Rico”, *Nueva Revista de Filología Hispánica* 27:24-38.

Terrell, Tracy D. (1995), “Constraints on the aspiration and deletion of final /s/ in Cuban and Puerto Rican Spanish”, in Hashemipour, Maldonado, & van Naerssen, (1995:261-280).

Terrell, Tracy (1980), “Final /s/ in Cuban Spanish”, *Hispania* 62:599-612.

Torreblanca, Máximo (1987), “Sobre la evolución de sibilantes implosivas en español”, *Journal of Hispanic Philology* 11;151-173, 223-249.

Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff, & Margo Bowman (2000), “English speakers′ sensitivity to phonotactic patterns”, in Broe & Pierrehumbert, (2000:269-282).

Valdivieso, Humberto, & Juanita Magana (1988), “Variación lingüística: La /s/ implosiva en Concepción”, *Revista de Lingüística Teórica y Aplicada* 26:91-103.

Valdivieso, Humberto, & Juanita Magana (1979), “Final /s/ in Cuban Spanish”, *Hispania* 62:599-612.

Westermann, Gert (1997), “A constructivist neural network learns the past tense of English verbs”, *Proceedings of the GALA '97 Conference on Language Acquisition*, pp. 393-398. Edinburg: HCRC.

Widdison, Kirk A. (1995a), “An acoustic and perceptual study of the Spanish sound change s > h”, *Rivista di Linguistica* 7:175-190.

Widdison, Kirk A. (1995b), “Two models of Spanish s-aspiration”, *Language Sciences* 17:329-343.

Notes

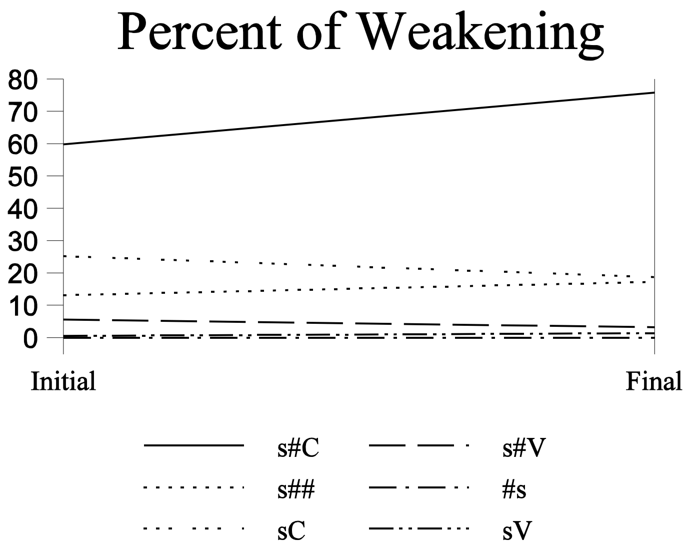
1. Many linguists would agree that formal representations in the form of rules and constraints do not relate to the actual processing of language (performance) by only to competence (e.g. Chomsky & Halle 1968:117; Kiparsky 1975:198; Kiparsky 1982:34). On the other hand, the issue is clouded because formal representations are spoken of as if they do explain processing (Carr 2000; Eddington 1996). For example, Kager (1999:26) states that "explaining the actual processing of linguistic knowledge by the human mind is not the goal of the formal theory of grammar, . . . [and] that a grammatical model should not be equated with its computational implementation." Nevertheless, a few pages later (33) he discusses how the formal mechanisms of optimality theory relate to actual language acquisition, and acquisition by any definition entails actual mental processing of linguistic material.

2. Hare, Elman, & Daugherty (1995) and Westermann (1997) also replicated Prasada & Pinker's study using other single-route models that were not exemplar-based.

3. The AML algorithm is freely available for download at: http://humanities.byu.edu/am/. AML's algorithm has been criticized on mathematical grounds (Baayen 1995). Nevertheless, the mathematical adequacy of AML is beyond the scope of the present study. A more relevant test of the algorithm is how well it mirrors language use.

4. One problematic aspect of conflating [h] and [i] in the simulation is that while both occur commonly syllable finally, [h] is much more common than [i] syllable initially (Brown and Torres Cacoullos (2002).

5. While pauses were indicated in the transcription on which the initial database was based, they were not marked in the orthographic transcription from which the test items were drawn. Pauses in the test items were assumed wherever there was a punctuation mark.


Percent of Weakening
80
70
60
50
40
30
20
10
0
Initial
Final
s#C
s#V
s##
#s
sC
sV

	Initial database						Final database					
	before ##		before C		before V		before ##		before C		before V	
Word	**h, i**	**s**	**h, i**	**s**	**h, i**	**s**	**h, i**	**s**	**h, i**	**s**	**h, i**	**s**
antes	0	0	0	0	1	1	0	10	26	1	1	8
dos	0	2	4	5	0	10	2	24	58	22	0	56
es	2	15	18	24	2	27	0	127	856	78	3	452
formas	0	0	0	0	0	0	1	4	4	0	0	1
fotos	0	0	0	0	0	0	0	10	2	0	0	3
haces	0	0	0	0	0	0	0	3	4	1	0	9
has	0	0	1	0	0	0	0	3	26	2	0	18
kilómetros	0	0	1	0	0	0	0	8	8	0	0	1
menos	0	3	5	7	0	1	4	20	39	9	0	27
pues	0	6	4	3	0	0	0	98	230	94	0	183
tres	0	1	5	4	0	2	0	27	33	9	0	26
vamos	1	1	0	0	0	2	8	47	8	0	0	101

	Initial database						Final database					
	before ##		before C		before V		before ##		before C		before V	
Word	**h, i**	**s**	**h, i**	**s**	**h, i**	**s**	**h, i**	**s**	**h, i**	**s**	**h, i**	**s**
atrás	0	0	0	0	0	0	0	4	2	1	0	1
detrás	0	0	1	0	0	0	0	2	4	0	0	1
días	1	6	0	2	0	0	2	13	3	5	0	4
ellos	0	1	1	2	0	0	0	18	25	5	4	4
fijaros	0	0	0	0	0	0	14	27	11	3	1	3
grandes	0	0	0	0	0	0	1	4	2	3	0	6
manchas	0	0	0	0	0	0	1	0	4	1	0	4
nutrientes	0	0	0	0	0	0	5	9	8	5	0	17
podías	0	0	0	0	0	1	0	0	3	1	0	3
podríamos	0	0	0	0	0	0	0	0	5	2	0	1
varios	0	0	0	0	0	0	0	2	5	3	0	2

	Initial Database		Final Database	
Word	**s**	**h, i**	**s**	**h, i**
desde	4	8	4	71
después	3	17	3	55
esfera	0	0	0	23
España	2	7	2	23
estoy	2	3	2	59
estamos	1	2	33	10
estás	0	0	29	10
nosotros	4	2	4	49

	Initial database						Final database					
	##		C		V		##		C		V	
Word	**[h, i]**	**[s]**	**[h, i]**	**[s]**	**[h, i]**	**[s]**	**[h, i]**	**[s]**	**[h, i]**	**[s]**	**[h, i]**	**[s]**
años	0	8	0	1	0	1	0	49	0	17	0	15
dejas	0	0	0	0	0	0	0	2	0	1	0	4
dices	0	0	0	0	0	0	0	30	0	9	0	1
eres	0	1	0	0	0	0	0	3	0	6	0	10
otros	0	1	0	2	0	0	0	6	0	47	0	4
veces	0	3	0	4	0	2	0	17	0	30	0	13

	Initial database						Final database					
	##		C		V		##		C		V	
Word	**[h, i]**	**[s]**	**[h, i]**	**[s]**	**[h, i]**	**[s]**	**[h, i]**	**[s]**	**[h, i]**	**[s]**	**[h, i]**	**[s]**
muchas	0	0	2	0	0	0	3	0	59	0	5	0
todas	0	0	7	1	1	1	1	2	54	1	8	3
todos	2	1	7	7	1	1	14	8	88	14	17	8
unas	0	0	2	0	0	0	5	0	40	0	7	12

David Eddington

Ortega Hall 235

University of New Mexico

Albuquerque, NM 87131

U.S.A.

davee@unm.edu