14 Look Ma, No Rules: Applying Skousen's Analogical Approach to Spanish Nominals in *-ión*

David Eddington
Brigham Young University

Introduction

During the past two decades, a number of language researchers have become disenchanted by models of language which espouse rules and constraints as the principal mechanism of linguistic cognition. Due to the dissatisfaction with such models, several non rule-based models have recently been proposed (e.g., Bybee, 1985, 1988, 1995; Lakoff, 1993; Stemberger, 1985, 1994; Goldsmith, 1993). Other models have been developed under the rubric of connectionism (e.g., Cottrell & Plunkett, 1991; Plunkett & Marchman, 1991; Rumelhart & McClelland, 1986; Seidenberg, 1992; Seidenberg & Bruck, 1990; Seidenberg & McClelland, 1989). Another model that has received less attention in the literature is Skousen's analogical model (1989, 1992, 1995). It is within the framework of this model that I wish to couch the present investigation.

In the first section, I briefly outline the principles behind Skousen's Analogical Modeling of Language (AM). Later, I discuss how AM accounts for the morphophonemic alternations that hold between Spanish verbal forms and their nominal counterparts that ending in *-ión*. Where possible AM is contrasted

with other extant formal analyses of the subject (Harris, 1969; Núñez-Cedeño, 1993).

Analogical Modeling of Language

Traditionally, analogy has been used to account for exceptional outcomes. When an outcome does not obey the general rule a form that is semantically or phonetically similar to the exceptional one is sought which is then said to influence the exceptional form in such a way that it does not develop according to the application of the general processes. What makes this sort of analogy suspicious is that it ultimately serves to patch up the inability of formal mechanisms to derive all forms. In addition, no limits are set regarding what forms can serve as analogs nor on how similar two forms must be in order for analogy to become a factor.

In contrast to the traditional notion of analogy, AM assumes that regular as well as irregular forms are attributable to the analogical influence of other forms. The reader is referred to Skousen (1989, 1992, 1995) for specific details of the model. However, I will briefly outline the tenets of the theory. The idea that language is exemplar–based turns the storage versus processing paradigm on its head. It asserts that people go through life storing all of the linguistic input they receive in all of its redundant glory and with all of its messy, irrelevant detail. This means that speakers do not need to subconsciously find systematic correspondences and generalizations in the data, make rules or constraints out of them, then discard the input they are based on. Instead, the generalizations exist in the linguistic experiences they have stored in their long-term memory. Behavior that appears to be rule-based may be explained by assuming that people have recourse to their past experience that is stored in highly organized matrices in terms of similarity. In short, linguistic cognition entails enormous amounts of storage and little processing. In AM, two things are needed to make predictions: a database of fully specified words[1] and a mechanism for searching and comparing those words. The behavior of the words most similar to the word in question (called the given context) generally predicts the behavior, although

the behavior of less similar words also has a small chance of applying as well. The basic algorithm is the following:

> We first search for actual examples of that context and then move outward in contextual space looking for nearby examples. In working outward away from the given context we systematically eliminate variables, thus creating more general contexts called *supracontexts*. (Skousen 1995, p. 217)

The probability that a word is chosen as an analog for the given context is dependent on three derived properties (Skousen, 1995, p. 217):

> (1) *proximity*: the more similar the example is to the given context, the greater the chances of that example being selected as the analogical model;
> (2) *gang effect*: if the example is surrounded by other examples having the same behavior, then the probability of selecting these similarly behaving examples is substantially increased;
> (3) *heterogeneity*: an example cannot be selected as the analogical model if there are intervening examples, with different behavior, closer to the given context.

These derived properties are important since they constrain what examples can constitute analogs, as well as deciding between competing analogs. These are precisely the factors that traditional appeals to analogy lack.

According to AM, all of the words contained in homogenous supracontexts constitute the analogical set. It is the words from this set that can serve as analogical models for a given context. There are two ways in which the contents of the analogical set can influence the outcome (Skousen, 1989, p. 82). The first is that a word can be randomly selected from among those in the analogical set and the behavior of that word applied to that of the given context. The other possibility is to determine which

outcome is most frequent among the words in the set and assign that outcome to the given context.

One fundamental difference between rule approaches and the analogical approach is the way each divides up contextual space. According to rule and constraint models, contextual space is divided with sharp boundaries between forms that behave differently. A given form either does or does not meet the structural description of a rule or constraint; there is no middle-of-the-road and no shades of gray. From the analogical standpoint, the existence of partial overlap and gradience is a reality that must be dealt with. Consider Figure 1.

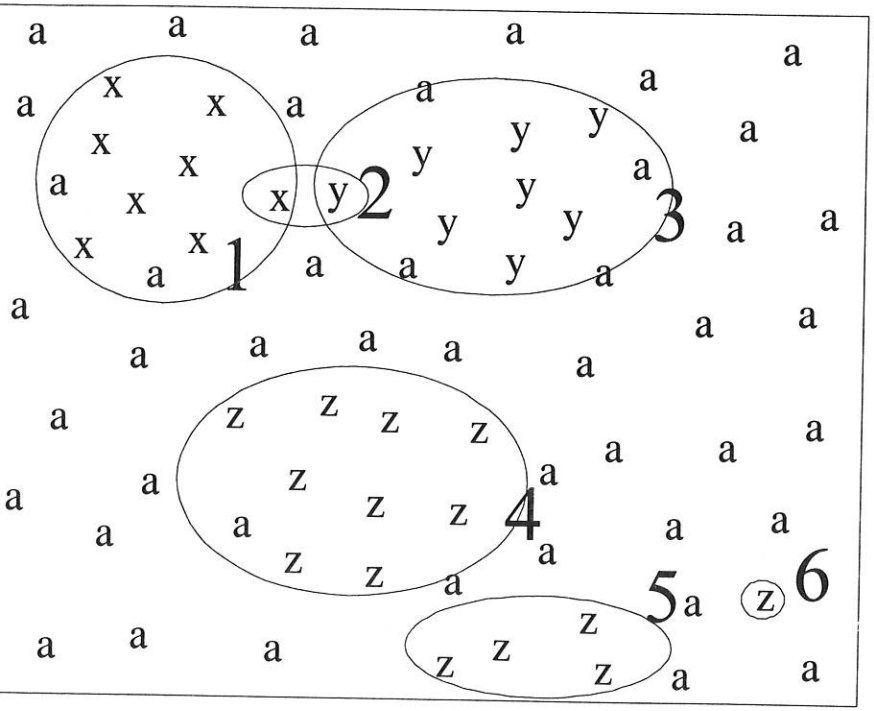

Figure 1. Two-Dimensional Representation of Multi-dimensional Contextual Space.

Each letter represents a word in contextual space that exhibits a certain behavior or relationship to another word. The words in Figure 1 belong to one of four groups based on their behavior: a, x, y, z. The closer any two letters are the more similar the words they represent are.

In general, words that behave the same tend to be similar to each other and fall closer to each other in contextual space. The major premise of AM is that a given form, if unknown or novel to the speaker, will behave in accordance with the neighbors that surround it. Therefore, if it falls in the middle of Circle 1, the majority of its neighbors have the behavior associated with Group X. The chances that one or more members of Group X will be chosen as the analogical model for the given context are very high in this case. Of course, if the word already exists in the mental lexicon the closest analog is the word itself making the probability 100% that the word's own behavior will be chosen, which represents lexical access to a known word.

Consider now the members of Group Y. There are members that are not completely surrounded by other Y members. In particular, the member of Group Y in Circle 2 has a neighbor from Group X. AM predicts, that under conditions of imperfect memory, in the formation of neologisms, and in language acquisition or language change, there will be a slight tendency for that member of Y to exhibit the behavior of Group X since it has a close neighbor from X. An example of this is found in English speaking children who occasionally use **brang* as the past tense of *bring* (Bybee & Slobin, 1982). The /ɪ/ of *bring* makes it extremely close to verbs such as *sing/sang, drink/drank,* and *shrink/shrank.* The historical change of the past tense of *sting,* from *stinged* to *stung,* can also be explained in terms of the analogical influence of its neighbors *stick/stuck,* and *shrink/shrunk* (Jespersen, 1942).

Two other situations are illustrated in Figure 1 which I will discuss as they relate to Spanish nominalization. The first is the existence of subgroupings as seen in Circles 4 and 5. The second is the existence of members that demonstrate the behavior of one group, but are more similar to the members of another (Circle 6).

Spanish Nominals in *-ión*

The purpose of this study is to show how analogy accounts for the relationship between Spanish verbal and nominal forms. The relationship between Spanish verbs and their corresponding nominal forms is fairly complex. Not only are there three nominal suffixes, *-ción, -sión,*[2] and *-ión,* but whether the theme vowel appears and how much of the verbal stem appears in the nominal form are also factors. It is the large number of different morphophonemic relationships between verbs and nouns that makes this phenomenon an ideal candidate for analogy.

Construction of the Database

Since AM assumes that all forms are stored in the mental lexicon, it was necessary to construct a database that approximates what exists in the minds of native adult speakers. Therefore, instead of including all nominals found in a standard dictionary the database was derived from the most frequent nominals ending *-ión.* I initially extracted all *ión* nominals that appear in the Alameda and Cuetos (1995) frequency dictionary of Spanish. The least frequent nominals had a frequency of one. Those nominal forms that have no verbal counterparts or that have verbal counterparts that are no longer semantically related were eliminated.[3]

At that point, I divided each pair of verbal and nominal forms into groups according to the morphophonemic relationship that holds between them. There were 59 pairs that did not fit into any groups, and which are considered highly irregular. A relationship is highly irregular if three or fewer verb/noun pairs exhibit the same relationship. For example, no other pairs in the database have the same relationship that *resucitar* 'to resurrect,' bears to *resurrección* 'resurrection.' Similarly, only *compeler/compulsión* 'to compel/compulsion,' and *repeler/repulsión* 'to repel/repulsion' demonstrate the alternation between *-peler* and *-pulsión*, which is why they are considered irregular. The remaining 939 pairs of nominal and verbal forms in the database fit into one of 14 groups. The characteristics of each group will be discussed as they are introduced.

The verbal form of all 939 pairs was encoded in terms of 12 variables. For most words in the database 12 variables are sufficient to encode the phonemic make up of the word and the syllabic structure of the final two syllables. This encoding means that some longer words are not fully specified. For these longer words, some phonemes at the beginning of the word are not included. For example, the first three phonemes of *yuxtaponer,* /yuk/, are missing. This is actually less problematic than it seems since the most relevant part of a verb as far as nominalization is concerned is the final two syllables. This is consistent with the principle of proximity (Skousen, 1989, pp. 52-53).

The infinitive form of the verb was encoded in this manner: The infinitival morpheme *-r* is not included as a variable since it is common to all infinitives. Variables 1-4 incorporate the final syllable of the infinitive. Variables 5-9 incorporate the penult syllable. Variables 10-12 incorporate any phonemes preceding those in the penult syllable, but not the syllable structure of the antepenult syllable.

The algorithm for deriving the variables centers on the syllable nuclei in variables (1) and (7). Starting from the nucleus and working outward the next tautosyllabic phoneme is assigned to the next variable. If there is none the next variable is given the syllable boundary symbol '0.' Any other missing segments are marked with the null symbol '=.'

(1) The theme vowel, which is also the nucleus of the final syllable.
(2) The tautosyllabic phoneme preceding (1) or the syllable boundary symbol '0' if the final syllable has no onset.
(3) The tautosyllabic phoneme preceding (2), or the syllable boundary symbol '0' if there is only one phoneme in the onset, else the non-specification marker '=.'
(4) The tautosyllabic phoneme preceding (3), or the syllable boundary symbol '0' if the onset contains two phonemes, else the non-specification marker '=.'
(5) The second tautosyllabic phoneme following the syllable nucleus of the penult syllable (7), or the syllable boundary marker '0' if only one phoneme follows (7), else '=.'

(6) The first tautosyllabic phoneme following the syllable nucleus of the penult syllable (7), or the syllable boundary marker '0' if the syllable is open, else '='.

(7) The syllable nucleus of the penult syllable, or '0' if the infinitive is monosyllabic.

(8) The tautosyllabic phoneme preceding (7) or the syllable boundary symbol '0' if the penult syllable has no onset, else '='.

(9) The tautosyllabic phoneme preceding (8), or the syllable boundary symbol '0' if there is only one phoneme in the onset, else '='.

(10) The first phoneme preceding the penult syllable, or the syllable boundary marker '0' if the onset of the penult syllable contains two phonemes, else '='.

(11) The first phoneme preceding the penult syllable if the onset of the penult syllable contains two phonemes, or the syllable boundary marker if (10) is the first phoneme of the word, else '='.

(12) The second phoneme preceding the penult syllable if the onset of the penult syllable contains two phonemes, or the syllable boundary marker if (11) is the first phoneme of the word, or the third phoneme preceding the penultimate syllable if the penult syllable has only one phoneme in the onset, else '='.

A number of questions arise regarding the selection of database items and variable selection. Is a database of 939 items too many or too few? Is it a fair approximation of what Spanish speakers know? This database contains types and not tokens. Would a database generated on the basis of token frequency be more representative?[4] The variables in the database represent phonemes and are organized according to syllables. Would it be better to consider phonetic features or acoustic qualities rather than phonemes? Perhaps some alignment of the variables other than according to syllables would be more psychologically plausible. All of these are valid questions that have yet to be answered. Some progress toward answering these important questions has already begun (Eddington, 2004a). Nevertheless, as will be seen,

the variables used in this simulation appear to work well even though they may not be optimal.

Table 1.
Exemplification of Variables

Examples:	Variables:
	210987654321
percutir	per0ku0==0ti
disminuir	dis0mi0=0nui
evadir	=0e0ba0==0di
concluir	===0kon0klui
poseer	0po0se0===0e
yuxtaponer	sta0po0==0ne

Types of Spanish Nominalization

The most common way to nominalize a Spanish verb is by affixing the suffixes *-ión, -ción* or *-sión* to the verbal stem. The choice of suffix, its exact placement in regard to the stem, and a number of other morphophonemic alternations make nominalization a fairly complex phenomenon. In the course of creating the database of nominals I identified 14 different relationships that a verb may have to its corresponding nominal form.[5] The purpose of this paper is to demonstrate how analogy can handle these 14 categorical relationships. However, it is important to remember that a major premise of AM is that all morphologically related words have individual representation in the mental lexicon. As a result, it does not pretend to be able to correctly derive all outcomes. What is does claim to mirror is actual natural language usage such as acquisition, slips-of-the-tongue, historical changes, dialectal development, etc. I discuss this sort of evidence in a later section. Nevertheless, predicting the behavior of the Spanish nominals with AM gives insight into the workings of the model. Moreover, certain groups of nominals are particularly interesting in

that they allow the analogical approach and formal approaches to be contrasted.

Group A

This is by far the largest group comprising about 75% of the database. The vast majority of these verbs are first conjugation *-ar* verbs and in reality the theme vowel /a/ is the only phonetic trait shared by most members. Only a small number of verbs are of the *-ir* variety and appear with the theme vowel /i/. Nominals in Group A retain the entire verbal stem and theme vowel and take the suffix *-ción:*

Table 2.
Group A: Examples

Infinitive	Gloss	Nominal	Gloss
admirar	'to admire'	*admiración*	'admiration'
competir	'to compete'	*competición*	'competition'
masticar	'to chew'	*masticación*	'chewing'
narrar	'to narrate'	*narración*	'narration'

Due to the size and phonetic diversity of this group its members are scattered throughout contextual space as Figure 1 indicates.

The sheer number of Group A members means that members of most other groups will have several neighbors from Group A. Therefore, if the nominal form of an A-type verb is not remembered or not known its neighbors will in most cases influence it to behave as a member of A. This can be demonstrated by removing a word from the database and predicting its behavior, (i.e. the relationship between the nominal and verbal forms), on the basis of its neighbors. In other words, each verb is treated as if its corresponding nominal form were unknown or novel and is produced for the first time. Table 3 shows the probability that several different Group A verbs will behave as Group A verbs.[6] For example, when *respirar* is eliminated from the database there

is a 100% probability that it will behave as a Group A word and yield the correct nominal *respiración* instead of **respirción, *respirsión,* or **respisión.*

Table 3.
Sampling of Group A Outcomes

Infinitive	Gloss	Prob. of A	Prob. of Other
respirar	'to breathe'	100	0
informar	'to inform'	99.0	1.0 (I)
agregar	'to add'	99.3	0.7 (M)
circular	'to circulate'	99.9	0.1 (N)
narrar	'to narrate'	100	0
planear	'to plan'	100	0
actuar	'to act'	99.8	0.2 (I)
celebrar	'to celebrate'	100	0
**fundar*	'to found'	45.2	54.8 (N)

Results such as these are typical of other samplings of Group A verbs tested with AM. The majority of Group A words are surrounded by other words in this category. However, there are occasional outliers. For example, when *fundar* is eliminated from the database and used as a given context it is found to be very similar to words such as *fundir* 'to melt' and *confundir* 'to confuse' which belong to Group N and whose members exhibit a different morphophonemic relationship between their verbal and nominal forms.

The considerable size of Group A has other consequences as well. Many members on the fringes of other groups have Group A neighbors. This means that leakage from other groups will often be in the direction of Group A. Neologisms also feel the pull of Group A verbs. In Spanish, all new additions to the verbal lexicon take *-ar* morphology. This fact, coupled with the wide dispersal of A words across contextual space means that most nominal neologisms in *-ión* are also swept into the A Group pattern.

Group B

This group includes all verbs whose stem contains the suffix *-ión*. Historically, they are the result of converting nominals into *-ar* verbs:

Table 4.
Group B: Examples

Infinitive	Gloss	Nominal	Gloss
congestionar	'to congest'	*congestión*	'congestion'
erosionar	'to erode'	*erosión*	'erosion'
sancionar	'to sanction'	*sanción*	'sanction'

The close phonemic similarity between the members of this group and any verb ending in *-ionar* results in a strong analogical force. The membership of all 50 Group B words is correctly predicted on the basis of their neighbors. Slight leakage toward Group A is evident in 21 cases.

Group C

All of these verbs end in *-poner*. They form a cohesive group in that *-sición* appears in the nominal in place of the theme vowel and the stem-final /n/:

Table 5.
Group C: Examples

Infinitive	Gloss	Nominal	Gloss
exponer	'to expose'	*exposición*	'exposition'
proponer	'to propose'	*proposición*	'proposition'
suponer	'to suppose'	*suposición*	'supposition'

AM correctly associates all 18 members to this group.

Group D

In this group, *-cción* [kθjón] appears in the nominal in place of the theme vowel and the stem final /θ/.

Table 6.
Group D: Examples

Infinitive	Gloss	Nominal	Gloss
conducir	'to conduce'	*conducción*	'conduction'
deducir	'to deduce'	*deducción*	'deduction'
inducir	'to induce'	*inducción*	'induction'
introducir	'to introduce'	*introducción*	'introduction'
producir	'to produce'	*producción*	'production'
reducir	'to reduce'	*reducción*	'reduction'
reproducir	'to reproduce'	*reproducción*	'reproduction'
seducir	'to seduce'	*seducción*	'seduction'

The nominal form of all 8 verbs is predictable on the basis of their affinity with other Group D verbs.

Group E

All 6 members of this group end in *-primir*. The last two phonemes and theme vowel of the verbal forms are replaced by *-esión* in the nominals:

Table 7.
Group E: Examples

Infinitive	Gloss	Nominal	Gloss
comprimir	'to compress'	*compresión*	'compression'
deprimir	'to depress'	*depresión*	'depression'
oprimir	'to oppress'	*opresión*	'oppression'
reimprimir	'to reprint'	*reimpresión*	'reprint'
reprimir	'to repress'	*represión*	'repression'
suprimir	'to suppress'	*supresión*	'suppression'

AM correctly predicts the behavior of all 6 members.

Group F

All five Group F members end in *-olver*. In the nominal, *-ución* replaces the theme vowel and final stem phoneme of the verbal form:

Table 8.
Group F: Examples

Infinitive	Gloss	Nominal	Gloss
absolver	'to absolve'	*absolución*	'absolution'
devolver	'to return'	*devolución*	'return'
disolver	'to dissolve'	*disolución*	'dissolution'
resolver	'to resolve'	*resolución*	'resolution'
revolver	'to rotate'	*revolución*	'revolution'

The group membership of all 5 members is correctly predicted by their similarity to other members of the group.

Group G

This is the only group that takes the suffix *-ión*. It appears immediately following the stem with no intervening theme vowel. Of the 20 members, 14 end in *-sar* (e.g., *dispersar* 'to disperse,' *expulsar* 'to expel,' *anexar* 'to annex').[7] The 6 odd members are: *unir* 'to unite,' *reunir* 'to meet,' *desunir* 'to divide,' *opinar*, 'to think,' *rebelar* 'to rebel,' and *coercer* 'to coerce':

Table 9.
Group G: Examples

Infinitive	Gloss	Nominal	Gloss
expresar	'to express'	*expresión*	'expresion'
rebelar	'to rebel'	*rebelión*	'rebellion'

With the exception of those words ending in *-unir*, the rest of the members of this group are isolated from each other in conceptual space. As a result, the behavior of a given member cannot be predicted on the basis of any other member. In fact, AM predicts them to be members of Group A. Therefore, errors and language shift are expected to be in the direction of Group A behavior: *anexar* > **anexación* instead of *anexión*, (where orthographic *x* is /ks/).

From a rule-based perspective, AM apparently fails to capture the significant generalization that verbs in *-sar* take the suffix *-ión*. However, a closer examination of the data reveals that no real generalization exists. While there are 14 Group G verbs ending in *-sar*, there are also 10 Group A words which end in *-sar* (e.g., *acusar* 'to accuse,' *conversar* 'to converse,' *cesar* 'to cease'). AM's apparent failure to see the stem-final /s/ and theme vowel /a/ as the unifying variables of the *-sar* subgroup is attributable to the simple fact that they are not unique to Group G. Of course, in a rule approach it is always possible to use a diacritic of some sort to distinguish *-sar* verbs with Group A behavior from *-sar* verbs with Group G behavior. However, in my view, the use of diacritics is

undesirable because allows one to artificially break up contextual space into neat but ad hoc subdivisions.

Group H

Stem-final /x/ (orthographic *g*), is the feature shared by the five members of this group. The /x/ of the verbal form alternates with /k/ in the nominal, and the nominals are formed with the addition of the suffix *-ción*:

Table 10.
Group H: Examples

Infinitive	Gloss	Nominal	Gloss
afligir	'to afflict'	*aflicción*	'affliction'
colegir	'to collect'	*colección*	'collection'
corregir	'to correct'	*corrección*	'correction'
elegir	'to elect'	*elección*	'election'
proteger	'to protect'	*protección*	'protection'

Due to the small size of this group and the small degree of phonemic similarity between its members there is little group cohesiveness. Its members are scattered throughout contextual space as the members of Group G are. Only *corregir* can be predicted on the basis of the other members of the group.

Harris (1969, p. 138) and Núñez-Cedeño (1993, p. 154) account for this alternation with simple rules. However, in order to do so they assume that the underlying form of these words contains an abstract stem-final phoneme /g/ which is absent in all surface forms. Accordingly, the stem of *proteger* is /proteg/ instead of /protex/. This essentially mirrors the diachronic process that transformed Latin /g/ into Spanish /x/ in certain contexts. The underlying /g/ is purely abstract which raises serious doubts about whether it is learnable.

Group I

All 25 members of Group I end in *-tar*.[8] The suffix *-ción* appears in the nominal in place of the theme vowel and the stem final /t/:

Table 11.
Group I: Examples

Infinitive	Gloss	Nominal	Gloss
ejecutar	'to execute'	*ejecución*	'execution'
infectar	'to infect'	*infección*	'infection'
objetar	'to object'	*objeción*	'objection'

Of these 25 members, AM predicts that only 6 will have Group I behavior. The remaining 19 are influenced heavily by members of Group A (see Table 12). This is highly reminiscent of the distribution of Group G.

Table 12.
Sampling of Group I Outcomes

Infinitive	Gloss	Prob. of I	Prob. of A
**adoptar*	'to adopt'	0.2	99.8
**secretar*	'to secrete'	10.9	89.1
**detectar*	'to detect'	2.8	89.4
infectar	'to infect'	81.6	18.1
inyectar	'to inject'	100.0	0.0

The six members that are correctly assigned to this group (*infectar, inyectar, insertar* 'to insert,' *intentar* 'to try,' *inventar* 'to invent,' and *desinfectar* 'to disinfect') are able to influence each other because they have several traits in common in addition to ending in *-tar*. Each begins with the prefix *in-* and has /e/ as the nucleus of a penult syllable that is closed by one consonant and begins with one

With the exception of this subgroup, AM is unable to correctly predict the behavior of words belonging to Group I. On the other hand, obtaining the correct outcome is a simple process in rule-based models (Harris, 1969, pp. 148-52; Núñez-Cedeño, 1993, pp. 151-153). Accordingly, rule application is able derive the correct nominal form of any verb ending in *-tar*, (with the exception of *explotar* 'to explode' whose nominal takes *-sión*). However, the apparent superiority of the rule approach is actually due to the apparent homogeneity of the data.

Harris suggests that *explotar* is the only exception when in reality there is a substantial group of *-tar* verbs whose nominal form would be incorrectly derived by the application of the rules. Besides *explotar*, the nominal of *conectar* 'to connect' is *conexión*. There are also 96 verbs that end in *-tar* (e.g., *habitar* 'to inhabit,' *representar* 'to represent,' *excitar* 'to excite,' *exaltar* 'to exalt') that have Group A behavior. It would surely be unfeasible to have to consider 98 forms as exceptions to a rule devised to account for only about 25 forms. Nevertheless, it is always possible in a rule analysis to mark the words in Group I with an abstract diacritic that has no basis in the morphological, semantic, or phonological characteristics of the words. This is Núñez-Cedeño's solution. He uses the diacritic [±special] which allows *-tar* verbs with Group A behavior to be distinguished from the *-tar* verbs that do not have Group I behavior.

Since such abstract diacritics cannot figure into AM the analogical approach must reflect the reality of the situation, namely that there is no real generalization to be made concerning *-tar* verbs. Furthermore, it makes the empirical prediction that under conditions of imperfect memory many verbs of this type will take the nominal form associated with Group A words.

Group J

The unifying characteristic of this class of words is that the suffix *-ción* immediately follows the stem of the nominal form.[9] This group contains four distinct subgroups (*-uir -tener -venir* and *-scribir*) and four odd members (*asumir* 'to assume,' *presumir* 'to presume,' *concebir* 'to conceive,' and *disecar* 'to disect');

Table 13.
Group J: Examples

Infinitive	Gloss	Nominal	Gloss
atribuir	'to attribute'	*atribución*	'attribution'
concebir	'to conceive'	*concepción*	'conception'
inscribir	'to inscribe'	*inscripción*	'inscription'
intervenir	'to intervene'	*intervención*	'intervention'
retener	'to retain'	*retención*	'retention'

The distribution of this group is illustrated graphically in Figure 2. The different subgroups are represented in the large ellipses and the odd member *concebir* is shown as isolated from other members in contextual space.

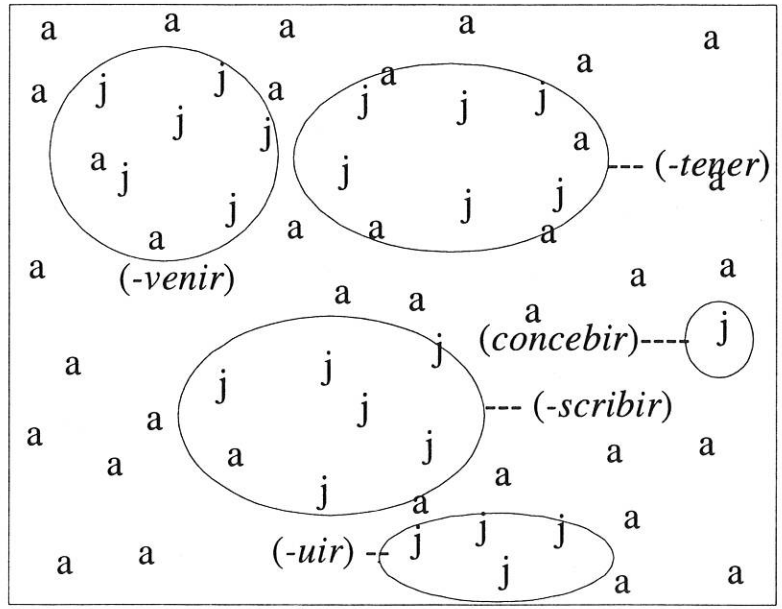

Figure 2. Contextual Space of Group J.

It should come as no surprise to find that the four odd members are found by AM to have more in common with members of Group A than they do with Group J. On the other hand, the model correctly assigns Group J behavior to 30 of the remaining 32 Group J words.

The incorrect assignment of *abstener* and *contravenir* to Group A is puzzling. The model finds *abstener* and *contravenir* slightly similar to a large number of Group A verbs that have a similar syllable structure, (they all end in -CV.CVr), and that share a phoneme or two and thus assigns them Group A behavior. In contrast, all other words in *-tener* and *-venir* feel the largest analogical pull from other Group J words ending in *-tener* and *-venir* respectively. It is difficult to conceptualize *abstener* and *contravenir* as being isolated from the other members of Group J. In my view this is a clear case in which an apparently obvious analogy is missed.

Group K

The suffix *-sión* appears after the stem of the nominals of this category. Five of the verbal forms end in *-uir* and the remaining three in *-Ver:*

Table 14.
Group K: Examples

Infinitive	Gloss	Nominal	Gloss
corroer	'to corrode'	*corrosión*	'corrosion'
desposeer	'to dispossess'	*desposesión*	'dispossession'
poseer	'to possess'	*posesión*	'possession'
concluir	'to conclude'	*conclusión*	'conclusion'
excluir	'to exclude'	*exclusión*	'exclusion'
incluir	'to include'	*inclusión*	'inclusion'
ocluir	'to obstruct'	*oclusión*	'obstruction'
recluir	'to imprison'	*reclusión*	'imprisonment'

All of the forms in *-uir* are correctly predicted by AM to belong to Group K. *Desposeer* and *poseer* serve as the major sources of analogy for each other which allows them to be successfully linked to Group K as well. *Corroer* is the only member that is sufficiently distant from the others that it cannot be assigned Group K behavior based on analogy to other members.

Group L

The unifying factor in this group is the appearance of *-ción* after the stem of the nominal with no intervening theme vowel. There are two subgroups, words ending in *-uir* and words ending in *-traer*:

Table 15.
Group L: Examples

Infinitive	Gloss	Nominal	Gloss
construir	'to construct'	*construcción*	'construction'
destruir	'to destroy'	*destrucción*	'destruction'
instruir	'to instruct'	*instrucción*	'instruction'
reconstruir	'to reconstruct'	*reconstrucción*	'reconstruction'
atraer	'to attract'	*atracción*	'attraction'
contraer	'to contract'	*contracción*	'contraction'
distraer	'to distract'	*distracción*	'distraction'
extraer	'to extract'	*extracción*	'extraction'
sustraer	'to subtract'	*sustracción*	'subtraction'

All of the *-traer* words are predictable on the basis of the other similar verbs in the group. However, only two of the four *-uir* verbs (*construir, instruir)* are successfully associated with Group L. The major analog of *reconstruir* is *construir,* but enough similarities are found with other Group A verbs that 50.5% of the analogical force is from Group A and only 42.6% from Group L.

Destruir appears to be even more isolated (2.7% L, 90.7% A), a fact that I will return to below.

Overlap between -uir verbs. Groups J, K, and L each contain subgroups of words whose verbal form ends in *-uir*. In addition, *fruir* 'to enjoy,' and *intuir* 'to sense' belong to Group A. Therefore, an important aspect of any account of how nominal and verbal forms are related is that it can distinguish between these similar forms. The *-uir* members of Group K have a great deal in common. Since they end in *-cluir* their behavior may be predicted on the basis of their similarity to the other members of the group. The same is true of the members of Group L, all of which end in *-struir*.

The members of Group J bear less similarity to each other since they do not all end in the same word final phonemes. Nevertheless, they are all predictable on the basis of the similarities they have with other members of the group. Although they do not share exactly the same word final phonemes an analysis of the analogical set constructed for each Group J word reveals a group similarity which is not immediately apparent. Not only do they all end in *-uir* but they have a common syllable structure. The Group J *-uir* verbs have an open penult syllable with the nucleus /i/; that is they end in *-i.Cuir*. These similarities are responsible for the cohesiveness of Group J. AM also correctly associates *intuir* with Group A. Therefore, the only case of true misassignment made by AM is that it groups *fruir* with Group L instead of Group A and predicts **frucción.*

The nominalization of the different *-uir* verbs can be handled in a rule-based approach as well although it requires a bit of abstraction. Harris (1969) and Núñez-Cedeño (1993) propose rules that derive the nominals from the verbal forms. They effectively separate the *-uir* verbs in each category by positing abstract underlying forms containing /d/ and /g/ for the *-uir* verbs of Groups K and L respectively. (They make no mention of the two Group A forms):

Table 16.
Groups J-L: Underlying Forms

Group	Example	Underlying Form
Group J	*instituir*	/instituir/
Group K	*incluir*	/inklu**d**ir/
Group L	*destruir*	/destru**g**ir/

Their inclusion of abstract phonemes in the underlying representation is directly traceable to the perceived need in generative analyses to derive all surface forms from unique underliers. For example, their rules are meant to derive not only the nominal *destrucción* from the verb *destruir,* but the agentive *destructor* and the adjective *destructivo.* The same rules are applied to derive *lector* 'reader' and *lectura* 'reading' from *leer* /leger/ 'to read' even though *leer* does not end in *-struir.* The point here is that instead of simply treating *leer* as an exception to the generalization about verbs ending in *-struir*, rule-based analyses are compelled to derive the morphological relatives of *leer* with the same rules that derive the morphological relatives of *-struir* verbs. For this reason, they do not entertain the possibility that *-struir* could constitute the context for the rules, hence the need to posit an abstract /g/ in the stem.

In contrast to the rule-based approach, distinguishing between the *-uir* verbs in each group is a straightforward task in the analogical approach. All Group L verbs end in *-struir.* This high level of similarity enables AM to predict each member's behavior on the basis of the other members of Group L. As mentioned above, all Group J verbs end in *-i.Cuir.* This sort of similarity is seen by analogy but poses a problem for a rule-based analysis since there is no unique context for all these verbs. For this reason, the rule-based accounts must resort to abstract diacritics.

Group M

All 38 Group M verbs are of the *-er* or *-ir* conjugation and characterized by a stem-final /d/ or /t/. The corresponding nominals take *-sión* and do not include the stem-final /d/ or /t/:

Table 17.
Group M: Examples

Infinitive	Gloss	Nominal	Gloss
aludir	'to allude'	*alusión*	'allusion'
emitir	'to emit'	*emisión*	'emission'
persuadir	'to pursuade'	*persuasión*	'pursuasion'
transmitir	'to transmit'	*transimisión*	'transmission'

Only 21 of these forms find a significant amount of analogical pull from other members of the group so that their behavior is predicted to be that of Group M. The strongest pull that the remaining members feel is from their Group A neighbors.

Rule analyses of Group M. As far as Group M is concerned, the correct nominal form is predicted by AM for only about half of the verbs. This suggests that the stem-final /t/ or /d/ is apparently not a sufficiently unifying trait. However, according to Harris (1969, pp. 143-153) and Núñez-Cedeño (1993, pp. 160-170), a stem-final /d/ or /t/ in an *-ir* or *-er* verb triggers the application of rules that derive all of the nominal forms in Group M from their respective verbal forms. Harris cites only a handful of exceptions to his rule. It would appear that the success rate of the rule-approach far outweighs that of AM in accounting for the behavior of Group M. However, appearances are deceiving since there are actually more than a mere handful of exceptions to the rule. Besides the *-tender* verbs Harris mentions, (*atender* 'to attend,' *contender* 'to contend') the nominal of *deglutir* 'to swallow' is *deglución* and not **deglusión* as the rules would derive. *Competir* 'to compete,' *partir* 'to divide,' *repetir* 'to repeat,' and *repartir* 'to distribute,' exhibit Group A behavior as do

medir 'to measure,' and *rendir* 'to render.' Several other verbs with Group A behavior which are not included in the database would also be incorrectly affixed by the rules: *expedir* 'to expedite,' *vender* 'to sell,' *hundir* 'to sink,' '*perder* 'to lose,' *comedir* 'to exercise restraint.' In addition to these exceptions, 7 additional forms are discussed in the following section that would be incorrectly derived by the proposed rules.

Since verbs ending in *-dir -der -tir* and *-ter* are not consistently tied to the same morphophonemic alternation between their verbal and nominal forms, the analogical approach does not find a relationship between all verbs with stem-final /d/ or /t/ and Group M behavior. In other words, AM's inability to consistently predict the behavior of Group M words reflects the fact that no single unique variable unites the group and that no broad generalization exists. In the face of such evidence the only way to salvage a rule-based analysis would again be to with the use of some sort of diacritic to separate the t-stem and d-stem words that undergo the rule from those that do not.

Group N

Six of the 7 verbs in this group have a stem ending in *-fundir,* the exception being *escindir*. The nominal forms take the suffix *-sión* and the stem final consonants /nd/ are absent. The behavior of 5 members of this group is predictable on the basis of their similarity to other members of the group. Given its different phonological makeup it is not surprising that the nominal of *escindir* is not predictable on the basis of other Group N words. However, AM associates *difundir* more closely with Group A than it does with Group N in spite of the fact that it shares *-fundir* with 5 other Group N words. In this instance, AM seems to miss an obvious analogy. Although Harris' and Núñez-Cedeño's analyses do not include this group of verbs their nominal forms could of course be derived by rules. However, some sort of mechanism would be required so that the rules that apply to d-stem verbs in Group M do not apply to the d-stem verbs in Group N.

Table 18.
Group N: Examples

Infinitive	Gloss	Nominal	Gloss
confundir	'to confuse'	*confusión*	'confusion'
difundir	'to disseminate'	*difusión*	'dissemination'
efundir	'to effuse'	*efusión*	'effusion'
escindir	'to divide'	*escisión*	'division'
fundir	'to fuse'	*fusión*	'fusion'
infundir	'to instill'	*infusión*	'instilling'
transfundir	'to transfuse'	*transfusión*	'transfusion'

Of course, the elsewhere condition (Kiparsky, 1973) could be invoked to handle this situation. In this case, the rules that derive the nominals from the stem *-fundir*[10] would take precedence over those that derive the nominals from the simple d-stem verbs since the former context is more restrictive than the latter.

However, from the analogical perspective it appears that the elsewhere condition is merely an epiphenomenon. In fact, this same point was made by Daugherty and Seidenberg (1994) for connectionist models. There are no instances in which AM assigns a Group M behavior to a Group N word. This is because most Group N words have a great deal in common. The fact that they share the stem *-fundir* allows them to exert a large analogical force on each other. The fact that they share stem-final /d/ with verbs from Group M, (as well as with some verbs in other categories), is simply not enough similarity for them to be drawn in to displaying Group M behavior. There is no need to invoke the elsewhere condition; it simply falls out as a natural result of making analogies based on similarities. The more similarities that exist the greater the possibility of analogical influence between members of the same group.

Summary of Results

The purpose of this section has been to investigate the morphophonemic alternations that hold between Spanish verbal forms and their related nominals ending in *-ión* in terms of an explicit model of analogy. To do this, the verbal forms which were tested were individually removed from the database and the corresponding nominal forms were predicted on the basis of the verbs' phonological similarity to other verbs in the database. 91% of the nominal forms were correctly predicted. Several conclusions may be drawn from this exercise:

1) AM works well when true generalizations exist. That is, when the members of a group have similar characteristics which are not found in other groups AM correctly assigns the group behavior to the individual members of the group. In addition, the more traits the members of a group share the higher the probability that the behavior of any one member will be influenced by other members of the group. This situation is found in Groups B, C, D, E, F and with a few minor exceptions in Groups J, K, L and N as well.

2) Members of a group which are less phonologically similar to the other members often find themselves isolated in contextual space. As a result they receive little analogical influence from the other members of their group and are assigned exceptional behaviors. This is the case for a few members of Groups J, K and L.

3) If a group is large enough, such as Group A, a high degree of analogical influence may occur even if there is little overall group cohesiveness. For example, in Group A the only unifying group trait is that most, but not all members have the theme vowel /a/. A high degree of group influence is achieved because of sheer numerical dominance. The large number of members assures that any given member of the group will have at least some neighbors from that group.

4) Members of very small groups which share few traits are most likely to be isolated in contextual space and unable to exert much influence on each other analogically (Group H).

5) Groups G, I and M exemplify a situation in which the members have a trait in common, but the trait is not unique to the members of the group, and therefore does not constitute a true

generalization. This lack of group unity means that some members find same-group analogs based on other shared traits while others feel a stronger analogical pull from the members of other groups.
6) Occasionally, a core member of a group is found to be more similar to members of other groups and what appears to be a straightforward analogy is missed (*abstener* and *contravenir* in Group J and *difundir* in Group N).

AM is able to predict the nominal form of a verb on the basis of the similarities the verb bears to other verbs in most cases. It does so without recourse to rule systems or constraint systems.

Evidence from Language Usage

As already mentioned, the real test of AM is not its ability to correctly assign behavior to all forms. Its acceptance of gradience and fuzzy boundaries necessarily entails that some members of a group are separated in contextual space from other members of the same group. This presents no problem for speakers whose mental lexicon contains the forms and who are able to recall them. However, the above exercise consisted of removing the verbal form from the lexicon and attempting to predict its nominal on the basis of other verbal forms. Under these circumstances, outlying members will be less influenced by the other members of their group. The true test of AM is its ability to account for evidence arising from natural language use, to which I will now turn my attention.

Núñez-Cedeño's Experimental Study

The purpose of Núñez-Cedeño's study (1993, pp. 183-190) was to determine the psychological reality of the rules he suggests are responsible for alternation between stem-final /d, t/ and /s/ in forms such as *dividir divisor división*. In his first experiment, he asked 8 subjects to add the suffixes *-ión -or -ivo* and *-ble* to eight extant but uncommon Spanish verbs (*exordir efundir exaudir enfurtir cohonder despender luir*[11] *derruir*) and to one nonce form (*catir*). In less that 10% of the cases did the subjects appear to obey the rules when adding *-ión* onto the verbal stem (e.g., *enfurtir* > *enfursión,* Group M behavior). In over 60% of the cases, the rule appears not to have been applied and the stem was left intact (e.g., *enfurtir* > *enfurtición*, Group A behavior). The remaining cases involve various other odd changes.

In a second study, he presented the subjects with the verbal forms and a list of corresponding nominal forms with different morphophonemic alternations (e.g., *enfurtir* > *enfursión, enfurtión, enfurtición, enfusión).* The task of the subjects was to rate the forms on a scale of acceptability ranging from highly acceptable to highly unacceptable. In only 25% of the cases did subjects accept forms ending in *-ión* that demonstrated the change *t, d* > *s* as his rules suggest (e.g., *enfursión enfusión*). While in 50% of the cases the subjects accepted forms ending in *-ión* that maintained the stem-final /d, t/ (e.g., *enfurtición, enfurtión*).

In many regards, it is difficult to give these results a precise interpretation. The extremely small number of subjects and test items admit the possibility that one item or subject may have severely skewed the overall results. It also means that the results do not lend themselves to reliable statistical analysis. Moreover, the results are conflated by test item as well as by subject. Since the results for individual items are not presented many questions are left unanswered. For example, half of the forms which maintained the /d, t/ in the stem were rejected by the subjects. I suspect that the half that was rejected consisted of nominals such as *enfurtión* whose phonological shape is odd in Spanish. The half that was judged as acceptable was most likely of the *enfurtición* type. However, given the paucity of data presented I cannot support or refute my intuition about this matter.

Methodological problems aside, Núñez-Cedeño is left to grapple with the fact that in more cases that not the subjects did not apply the rules he posits. From the analogical perspective, the results of these experiments are much less surprising. In the section on Group M words, I discussed how stem-final /d, t/ is not a sufficiently unique trait to allow for strong analogy between the members of Group M. As a result, only 55% of the words of this category received strong analogical influence from other members of the group which would yield *t, d* > *s*. Most of the remaining 45% have closer neighbors from Group A. This means that under

conditions of imperfect memory or when dealing with nominal forms that are new to the speaker 45% of the words would receive Group A influence in which the stem-final /d, t/ is left untouched. This is generally consistent with the outcome of Núñez-Cedeño's experiment.

As already mentioned, Núñez-Cedeño does not discuss the outcomes for each individual test word. Nevertheless, I applied AM to the nine test items and found that four of them are given nominal forms in which *-sión* takes the place of stem-final /d, t/. The remaining five are assigned nominal forms in which the stem remains intact. In other words, the rule approach expects *t, d* > *s* to occur in all cases while AM only predicts it in about half the cases. Therefore, it seems safe to conclude that Núñez-Cedeño's experimental results support analogy to a greater extent than they do the rule approach.

Error Analysis

As a whole, the simulation was able to correctly predict the nominal form of 91% of the verbs when they were treated as novel items. While this number is impressive the errors made are also telling since AM claims to be able to predict language change, neologisms, and slips of the tongue. To test this I conducted a search of the internet using Google for all of the 81 'erroneous' forms predicted. Using the internet for linguistic analysis is becoming more common (e.g., Rainer, 2003) and it has some advantages and disadvantages. On the on hand, it allows for rapid inspection of massive amounts of data. On the other hand, information about the author (e.g., age, sex, country of origin, etc.) is not always readily apparent. Nevertheless, all of the nominals cited below were found on web pages that appear to be written by native Spanish speakers or at least highly proficient non-native speakers. None of the uses appear to be jocular nor are they discussed as examples of improper usage or dialectal peculiarities. Of the 81 erroneous forms, 60% (n= 49) are actually attested on Spanish-language web pages.

Table 19.
Attested Erroneous Nominals

absteneción	*decidición*	*exentación*	*revisación+*
adapción	*desertación*	*expandición*	*secretación*
adoptación+	*destruición+*	*expresación*	*sujetación*
afectación+	*detectación*	*frucción*	*televisación*
anexación+	*difractación*	*fundición+*	*transmitición*
asumición	*difundición*	*intervenición*	*unición*
cesión+	*disecación+*	*invadición*	*proyectación*
colegición	*dispersación*	*objectación*	*rebelación*[12]
concebición	*editación*	*opinación*	*recolectación*
conjuntación	*ejecutación*	*optación+*	*retransmitición*
construición	*excentación*	*presumición*	
convertición	*exceptación+*	*redactación*	
corregición	*excretación*	*impulsación+*	

The 38 items not marked with a cross can be considered either slips of the keyboard or neologisms. A word such as *unición,* which appeared in the context *Unición Europea* 'European Union,' could represent an attempt by the writers who used it to nominalize the verb *unir* when the word *unión* was either missing from the mental lexicon or temporarily unavailable from memory. This demonstrates that *unir* is not a central member of its group but falls closer to Group A (see Figure 1, example 6). This explains why it takes Group A morphology. Exemplar-based models predict the direction that this sort of slippage takes in actual usage.

The 11 words marked with a cross are interesting for another reason. They appear as entries in an online dictionary of the Real Academia Española.[13] These words may be considered to be doublets, that is, two nominals related to the same verb but with differing morphology. What makes these words interesting is the century in which they are first attested. I found this by searching for all possible spelling variants of these words on *Corpus del*

Español.[14] As Table 20 indicates, AM predicts the newer nominal form related to the verb, hence, it predict the direction of the language change.

Table 20.

Attested Neologisms Predicted by Analogy

Word predicted by AM	Century	Word in original database	Century
fundición	16th	*fusión*	14th
adoptación	19	*adopción*	15th
impulsación	RAE	*impulsión*	15th
afectación	16th	*afección*	15th
revisación	20th	*revisión*	16th
anexación	RAE	*anexión*	17th
disecación	RAE	*disección*	18th
cesión [15]	13th	*cesación*	16th

Note. RAE = appear in the online dictionary, but not in *Corpus del Español*.

One interesting case is *destruición*. The verb *destruir* originally existed alongside its nominal *destrucción*. However, analogy shows that *destruir* is closer to words with Group A behavior than it is to other members of its group *construir* and *instruir* because its nominal is predicted to be *destruición*. Sometime in the 15th century, the influence of Group A on *destruir* gave it this new nominal. The analogically-based *destruición* coexisted alongside *destrucción* until sometime in the 19th century when the former disappeared in careful writing. However, the analogical influence has never disappeared completely which is why *destruición* may be found on the internet in contemporary contexts.

Conclusions

In general, AM does a respectable job of assigning the correct nominal behaviors, especially in those cases in which true generalizations exist. The data from natural language use, although sparse, also argues in favor of AM. It is hoped that more evidence of this type will be found in the future. Nevertheless, AM appears to be a viable alternative to formal analyses in explaining actual language use.

References

Alameda, J.R., & Cuètos F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano.* Oviedo, Spain: University of Oviedo Press.

Baayen, H. & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, *29*(5), 801-843.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* (Vol. 2, pp. 257-294). London: Academic Press.

Bybee, J. (1985). *Morphology*. Amsterdam: John Benjamins.

Bybee, J. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical approaches to morphology* (pp. 119-141). San Diego: Academic Press.

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

Bybee, J.L., & Slobin, D.I. (1982). Rules and schemas in the development and use of the English past tense. *Language*, *58*, 265-289.

Cottrell, G.W., & Plunkett, K. (1991). *Learning the past tense in a recurrent network: Acquiring the mapping from meaning to sounds.* In Proceeding of the thirteenth annual conference of the Cognitive Science Society. Hillsdale N.J.: Erlbaum.

Daugherty, K., & Seidenberg, M.S. (1994). Beyond rules and exceptions: A connectionist approach to inflectional morphology. In S.D. Lima, R.L. Corrigan, & G.K. Iverson

(Eds.), *The reality of linguistic rules* (pp. 353-388). Amsterdam: John Benjamins.

Derwing, B.L., and Skousen, R. (1994). Productivity and the English past tense: Testing Skousen's analogy model. In S.D. Lima, R.L. Corrigan, & G.K. Iverson (Eds.), *The reality of linguistic rules* (pp. 193-218). Amsterdam: John Benjamins.

Eddington, D. (2004a). Issues in modeling language processing analogically. *Lingua, 114*, 849-871.

Eddington, D. (2004b). *Spanish phonology and morphology.* Amsterdam: John Benjamins.

Goldsmith, J. (1993). Harmonic phonology. In J. Goldsmith (Ed.), *The last phonological rule* (pp. 21-60). Chicago: University of Chicago Press.

Harris, J.W. (1969). *Spanish phonology.* Cambridge MA: MIT Press.

Jespersen, O. (1942). *A modern English grammar on historical principles VI: Morphology.* Copenhagen: Munskgaard.

Kiparsky, P. (1973). Elsewhere in phonology. In S. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 93-106). New York: Holt Rinehart and Winston.

Lakoff, G. (1993). Cognitive phonology. In J. Goldsmith (Ed.), *The last phonological rule* (pp. 117-145). Chicago: University of Chicago Press.

Núñez-Cedeño, R.A. (1993). *Morfología de la sufijación española.* Santo Domingo: Universidad Nacional Pedro Henríquez Ureña.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception. *Cognition, 38*, 43-102.

Rainer, F. (2003). Studying restrictions on patterns of word formation by means of the Internet. *Italian Journal of Linguistics, 15*, 131-140.

Real Academia Española. (1995). *Diccionario de la lengua española*, CD-Rom Version. Madrid: Espasa Calpe.

Rumelhart, D.E., & McClelland, J.L. (1986). On learning the past tense of English verbs. In J.L. McClelland, D.E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing* (Vol. 2, pp. 216-271). Cambridge MA: MIT Press.

Seidenberg, M.S. (1992). Connectionism without tears. In S. Davis (Ed.), *Connectionism: Theory and practice* (pp. 84-122). New York: Oxford Univeristy Press.

Seidenberg, M., & Bruck, M. (1990). Consistency effects in the generation of past tense morphology. Annual meeting of the American Psychonomic Society, New Orleans. LA.

Seidenberg, M., & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523-68.

Skousen, R. (1989). *Analogical modeling of language*. Dordrecht: Kluwer Academic.

Skousen, R. (1992). *Analogy and structure*. Dordrecht: Kluwer Academic.

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

Stemberger, J.P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143-186). London: Erlbaum.

Stemberger, J.P. (1994). Rule-less morphology at the phonology-lexicon interface. In S.D. Lima, R.L. Corrigan, & G.K. Iverson (Eds.), *The reality of linguistic rules* (pp. 147-169). Amsterdam: John Benjamins.

Wang, H.S., & Derwing, B.L. (1994). Some vowel schemas in three English morphological classes: Experimental evidence. In M. Chen & O. Tzeng (Eds.), *In Honor of Professor William S-Y. Wang: Interdisciplinary studies on language and language change* (pp. 561-575). Taipei, Taiwan: Pyramid Publishers.

Author Note

An abbreviated discussion of this topic is found in Eddington (2004b).

Endnotes

1 In this study, the phonemic attributes of words are assumed to be the relevant variables. However, AM can also incorporate other variables such as sociolinguistic variables: age, sex, social class etc. (Skousen, 1989, pp. 97-100).

2 In the present study, the distinction between /c/ and /s/ was maintained. This means that the pronunciation of *-ción* and *-sión* is assumed to be [- jón] and [-sjón] respectively. In other dialects, both would receive the same pronunciation and one could only speak of two suffixes at the phonemic level.

3 Semantic relatedness was determined by comparing the definitions of each verbal and nominal form in the *Diccionario de la lengua española* (1995). To be considered related the nominal form must entail the process or act of the verbal form.

4 Making predictions as if the word were unknown is tantamount to testing the productivity of the process. Evidence suggests that productivity is based on type frequency (Baayen & Lieber, 1991; Bybee, 1985, 1995; Wang & Derwing, 1994).

5 Not all 14 groups I discuss are treated in other analyses of the topic (Harris, 1969; Núñez-Cedeño, 1993) nor do my groupings always correspond exactly to those of these other researchers.

6 For a detailed explanation of the algorithm AM uses to calculate probability and the amount of influence exerted by individual members of the analogical set the reader is referred to Skousen (1989, chapter 2).

7 *anexar, confesar, dispersar, expresar, expulsar, impulsar, intrusar, precisar, profesar, progresar, regresar, revisar, televisar, tensar.*

8 *adoptar, afectar, cantar, conjuntar, desertar, desinfectar, detectar, difractar, editar, ejecutar, exceptar, excretar, exentar, infectar, insertar, intentar, inventar, inyectar, objetar, optar, proyectar, recolectar, redactar, secretar, sujetar.*

9 In some cases, voicing or nasal assimilation between the stem-final consonant and *-ción* is represented graphemically. However, such assimilation between contiguous consonants is an automatic and unconstrained process in native Spanish words whether it receives orthographic representation or not. Therefore, it is irrelevant to the present study which focuses on nominal morphology in terms of phonemic representations.

10 Of course *escindir* would need to be marked as an exception.

11 The words *luir* and *derruir* presumably have a stem-final /d/ in underlying representation from which the /s/ of the suffix *-sión* is derived. For this reason, they were included in the study. This mirrors a defunct historical process by which /d/ > /s/.

12 In the context, the noun is clearly related to *rebelar* 'to rebel' and not to *revelar* 'to reveal.'

13 www.rae.es

14 www.corpusdelespanol.org

15 The only exception may be *cesión.* This word appears once in the 13th century, but the context it appears in, coupled with the orthographic uncertainty of the period, makes it possible that the word is actually *sesión.*