

New Paradigms: A rule-and-feature based morpholexical model of the Spanish verbal system

Karen W. Burdette
Tennessee Tech University

1. Purpose of the model

This article introduces a morpholexical, rule-and-feature based model of Spanish inflectional verbal morphology. The purpose of the proposed model is to generalize, condense, and describe as elegantly as possible the patterns of Spanish verbal inflection, so as to produce all attested forms without producing any unattested forms, and without any unnecessary redundancy or superfluous rules. For all three Spanish verb classes (*a*-class, *e*-class, *i*-class), and for all categorical/regular verbs, the basic model presented in the present analysis is capable of producing all forms of the Present Indicative, Present Subjunctive, Imperfect Indicative, and Imperfect Subjunctive.

2. Theoretical basis of the model

The Spanish inflectional verb model proposed here is based in form primarily on Anderson's (1982) extended word-and-paradigm (EWP) model of morphology. In addition, it is theoretically based on Anderson's 'a-morphous morphology' (1995) and Bybee's analogical model of morphology (1985, 1988).

The present analysis corroborates Anderson's EWP model by demonstrating that his proposed blocks of rule-and-feature sets and his proposed ordering system can be successfully applied to the inflection of Spanish verbs, thereby producing all the necessary paradigms for all tenses and moods in the language. The present model also corroborates Anderson's theory of 'a-morphous' morphology, since the proposed model, like Anderson's, is a 'morpheme-less' treatment of morphology. Like Anderson, the present model does not view morphemes as separable, discrete units as in the classical view of the morpheme.

In the present analysis, however, rules are not seen as generative processes operating outside the lexicon as in Anderson's model. Rules and representations are considered as one and the same, as in Bybee's (1985, 1988) analogical model of morphology as lexical organization. Rules and representations are both part of a lexical continuum ranging from the most specific exemplars of lexical forms on one end to the most general exemplars on the other end. For example, an idiosyncratic, completely suppletive form such as *went* for English past tense of *go*, would be at the 'most specific' end of the continuum, while the 'rule' that adds /ed/ to form English past tense would be at the 'most general' end. Both the 'rule' and the suppletive form itself, however, are seen as lexical representations along the same continuum.

Analogical models of language usage emphasize binding form to meaning through analogy, by searching for the exemplar that is closest to the form in question, both semantically and phonologically. Exemplar-based learning systems involve looking for the most similar instances ('nearest neighbors') to predict language behavior (Skousen 1989, 1995). Eddington (2000:282) explains that Analogical Modeling of Language (AML) "is a model of how memory tokens may be used to predict linguistic behavior." AML accepts a whole-word view of the lexicon, with all known words stored as wholes in the mental lexicon.

Similarly, the present proposed model is morpholexical. The word-formation rules in the present model are regarded as taking place in the lexical component of the grammar, although there is multi-

dimensional interaction with other components of the grammar. In the proposed model the lexicon is seen as a dynamic component of the mental grammar with many levels of analogical connections, rather than as a static list of words or morphemes. Anderson (1995), on the other hand, accepts a split-morphology thesis, meaning that derivational morphology takes place in the lexicon, while inflectional morphology takes place in another component (or other components).

The proposed model and analysis support Bybee's (1985, 1988) analogical model of morphology, which emphasizes the role of phonological and semantic connections among words that are similar in form or meaning. The constants in the proposed model (which are, for the most part, the inflectional material of a given verb, or parts thereof) can be regarded as the parts of words that are 'similar in form or meaning' in Bybee's model. For instance, in Bybee's model there would be an analogical link among all words that end in /mos/ (as in *hablamos*, *comemos*, *escribimos*). That link is seen in the proposed model when /mos/ appears as one of the constants in the model, as in the rule /X/ → /X + mos/ for feature sets that contain [+1 Plural]. Similarly, the rule /X/ → /X + βa/ for feature sets that include [+ a-class; + Imperf Indic] shows the link among forms such as *hablaba*, *hablaban*, *amaba*, *amábamos*, *cantaba*, *cantabas*, *amabas*, and other Imperfect Indicative forms for a-class verbs. In other words, there are many verbs in Spanish, each with a different semantic and phonological value for the X variable, that can be mapped onto each of the phonological strings that make up the constants in the proposed model, just as there are many layers of connections among words in Bybee's model.

3. How the model works

Skousen (1995: 213), describes "three basic types of behavior" that a theory of language must account for: categorical, exceptional/regular, and idiosyncratic. 'Categorical' forms are those with completely regular patterns, such as English past forms with /ed/. 'Exceptional/regular' forms, on the other hand, are those with regularities in their irregularity, we might say. Forms in this latter group, although irregular compared to the first group, exhibit patterns and regularities of their own, such as so-called strong English verbs with past forms that can be grouped together based on similarity of form, such as *sang/rang*, *blew/flew/grew*, *bought/fought/brought*. Although the verb model proposed by the present author deals with all of these categories of verbal inflection in Spanish, this article presents only the basic model for categorical/regular Spanish verbs, that is, verbs traditionally viewed as completely 'regular'. Phonologically variant forms and other such stem alternation will not be dealt with here in this basic introduction to the model.

The essence of the proposed model is the link between the two parts of the linguistic sign as defined by Ferdinand de Saussure (trans. 1959:66): the *signifié*, which is the mental concept, the meaning, or the semantic content of a given word, and the *signifiant*, which is the mental imprint of the physical form, that is, the sound, that represents the given meaning of the word. In the present model of Spanish verbal morphology, as in Anderson's extended word-and-paradigm model, the *signifié* is represented by semantic/grammatical feature sets, and the *signifiant* is represented by rules that transform a given stem into a fully inflected word. The proposed model, with its two-part representations of verb forms, shows the semantic and phonological connections among forms and paradigms.

The proposed model is rule-and-feature based. The individual squares with the 'formulas' in them are the rule-and-feature sets of the model. The top half of each rule-and-feature set contains the set of semantic/grammatical/collocation features that are associated with the lexical form represented by the given rule. Individual features in the feature sets are enclosed in brackets and represented by the notation [+ Grammatical Function]. Anderson (1982) uses a binary notation system, using [+] features and [-] features in the feature sets. It is important to note that the [+] notation is not used anywhere in the present model to mean 'plus', in the sense of adding something onto a word, such as an affix. The [+] or [-] binary notation is used here in the same way it is used in syntactic and phonological theory, that is, to designate a positive or negative value for a given feature. In the present model, the features are grammatical features, and the [+] notation means that a given form has a given feature, such as a particular tense, as opposed to any other tense, for which the form would have a [-] value.

The present model uses only the [+] notation whenever possible for the sake of simplicity, with the assumption that, for instance, a [+ Pret] notation in a given feature set implies a [-] (negative)

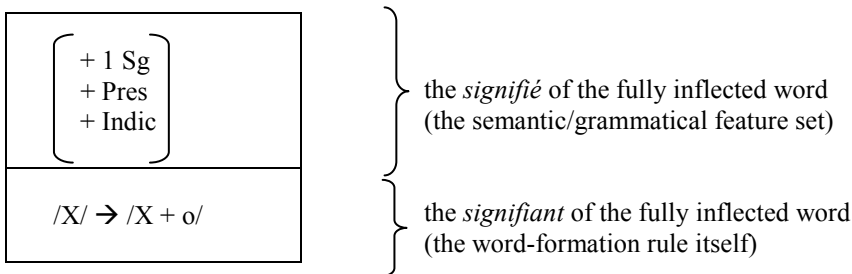
value for tenses other than Preterite. Although negative notations are not necessary in the basic model presented in this introduction to the model, it will be necessary in some cases later in the model to specify a [-] value, such as in one of the rules in the Preterite model. The negative notation [- Pret] for /X/ → /X + s/ is necessary to block the production of ungrammatical forms such as *hablastes or *comistes for [2 Sg Pret]. (In nonstandard dialects that use such regularized forms, however, such a notation would not be necessary.)

The following abbreviations are used in the present model: Pres, Indic, Subj, Imperf, as well as the numerals 1, 2, and 3 for first, second, and third person designation and Sg and Pl for singular and plural number designation. It should be noted that the grammatical labels and notations used in the proposed model are really just linguistic shorthand for the complex mental concepts they represent, as is a word itself. An ordinary speaker of the language does not typically possess such metalinguistic knowledge and is not necessarily aware of grammatical labels. Any speaker of a given language does, however, have tacit knowledge of the grammatical concepts represented in the given language. A Spanish speaker, therefore, might not be aware of the notation [2Sg Pres Indic], for example, but he or she is aware of the concepts that notation represents.

The rules in the bottom half (representing the *signifiant*), such as the rule /X/ → /X + o/ in Figure 1 below, describe an operation that transforms a stem into a fully inflected word form; thus the symbol → means ‘becomes’ or ‘is realized as’. The constants in the rules of the model are the phonological material that many words have in common: the phonological strings commonly referred to as grammatical morphemes, which in Spanish are affixes attached to the end of a given lexical stem. On the other hand, the variables /X/ that enter the model, enclosed between slashes as phonemes are conventionally notated, are the phonological strings that carry the semantic content of the word. Unlike the constants in the model, which belong to a closed class, the variables that initially enter the model belong to an open class, since this class of semantic variables can be augmented by adding new words to the language. This alternate view of constants and variables is important to the way the proposed model functions, since the variables add more phonological material as they progress through the blocks of the model, as will be seen below.

Figure 1 below shows one rule-and-feature set from one of the blocks of word-formation rules in the proposed model. This particular rule would yield a form such as *compro* or *hablo*, both of which are First Person Present Indicative forms.

FIGURE 1. Sample rule and feature set from the blocks of word-formation rules in the proposed model.



At the outset, upon entering the blocks, /X/ = the content-bearing stem, which for most Spanish verbs is the infinitive form minus the /ar/, /er/ or /ir/ ending. . The variable for any given word, at least at the outset of the model, is typically the phonological string that carries the semantic content of a given word, such as /aβl/ for ‘speak’, as in *hablo*, *hablas*, *habla*. The term ‘content-bearing’ refers to the part of a word that conveys the word’s semantic content, that is the variable in the proposed model, as opposed to the part or parts commonly called ‘grammatical morphemes’, although so-called grammatical morphemes could actually be considered as part of the overall semantic value of a given word. For example, there is a semantic difference between the mental image of one apple and the mental image of a dozen apples. Likewise, there is a difference in meaning between a verb that designates action occurring at the present moment and a verb that designates an action that took place a hundred years ago. Just because it is possible to separate some words into discrete, identifiable parts,

this does not necessarily mean that words are inherently separable in terms of their semantic and grammatical functions, nor does it necessarily imply that grammatical functions in a word are psychologically different from ‘true’ semantic functions.

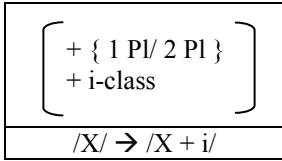
In the present model, all parts of a word are just parts, just phonological strings, as in Anderson’s (1995) a-morphous, morpheme-less morphology. Semantic features and grammatical features in the present model are all seen as having semantic value, and the phonological strings are just seen as variables or constants that are collocated with certain specified other phonological strings. As more phonological material is added to the word, more semantic/grammatical/collocation features are also added, but these features may not be entirely separable either in the top (*signifié*) half of the set or in the bottom (*signifiant*) half of the given rule-and-feature set, as they would be in a hypothetical, strictly agglutinative language.

In the production or comprehension of a spoken word, the speaker or listener begins with an initial sound or sound cluster. For most verbs in Spanish, this initial segment is the so-called content-bearing part of the word, such as /kant/ in a form such as *cantábamos* ‘sing’ [1 Pl Imper Indic]. In the model proposed in the present study, the true content-bearing segment for any given verb is the variable that enters the first block of the model, corresponding with the first segment of Bull’s (1965) and Politzer’s (1965) three-part analysis of Spanish verbs. However, this X variable is not viewed as a true morpheme in the classical sense as in Bull and Politzer, since the initial variable in the present model collects more phonological material as it progresses through the blocks of rules and features sets. Thus the variable entering successive blocks is usually altered in some way from the original variable, and the resulting final form of the word is the entire string of phonological material that exits the last block after picking up one or more of the constants, which in most cases is the phonological material commonly referred to as grammatical morphemes. We might say that the stem ‘grows’ as it progresses through the blocks of word-formation rules, but it is no longer a stem in the traditional sense after it adds additional phonological material and proceeds as the new variable to the next block of rules. It is for this reason that it is advantageous to use the term ‘variable’ in the model instead of ‘stem’. The proposed model is thus ‘morphemeless’. For example, the X variable in Rule 5 of Block I is /ye/, which is not a morpheme in the classical sense.

Unlike Anderson and the present proposed model, traditional analyses of Spanish verbs treat the various phonological segments as separable morphemes in the classical sense. Traditional analyses such as Bull’s (1965) and Politzer’s (1965) may serve well for many Spanish verb forms, but some forms, such as the preterite, cause problems. The present indicative can be viewed as stem plus class vowel plus person/number marker, as in: *habl + a + s* for *hablar* ‘speak’ [2S Pres Indic], although even in this instance there is no separable marker for tense, mood, and aspect. For a preterite form, however, there is even more fusion of the segments, thus making them less separable and thus more compatible with a morphemeless treatment. For example, for *hablaste* ‘speak’ [2Sg Pret Indic], we still have the stem *habl* plus the class vowel *a*, but the *ste* segment seems to signify tense, mood, aspect, person, and number all at once. And in irregular preterite forms such as *tuviste* [2Sg Pret Indic] from *tener* ‘have’, the stem *tuv* itself seems to signify tense. Furthermore, the class vowel segment in Bull (1965) and Politzer (1965) poses a problem for traditional analyses, since the class vowel has no semantic content and thus is not consistent with the definition of the morpheme as a minimal unit of meaning.

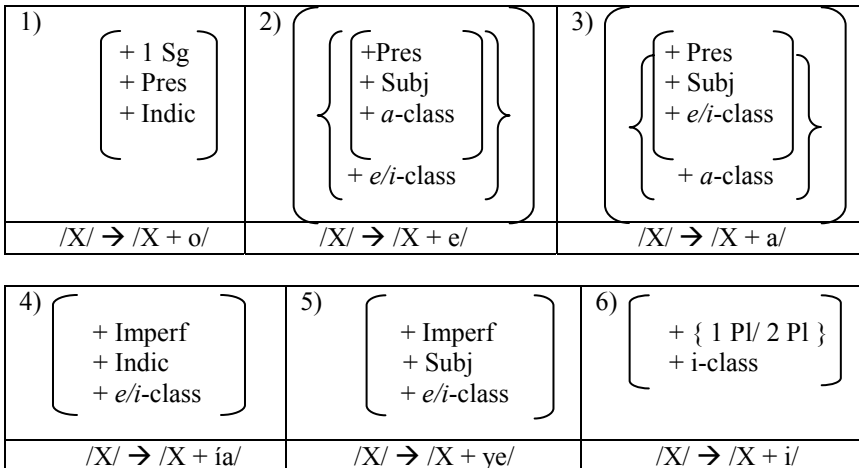
The distinction between square brackets [] and curly brackets { } is important for understanding the model. When a feature or a group of features is enclosed within regular brackets, the brackets mean that all of the features within the brackets must apply in the given rule. For example, if a rule includes the feature set [+ Pres, + Indic, + 1 Sg], then the resulting form must have all of these features; thus the resulting form will be a First Person Singular Present Indicative form such as *hablo* or *escribo*. When more than one feature is enclosed within ‘curly’ brackets, however, the enclosed features are mutually exclusive, meaning that only one of them can apply for any given form. For example, if a rule includes the feature set {1 Pl; 2 Pl}, then only one of the enclosed features can apply for any given form. There may be a combination of regular and curly brackets, as in the sample rule in Figure 2 below: [+ {1 Pl; 2 Pl}; + *i*-class]. This feature set means that the resulting form will be either 1 Plural or 2 Plural, and, in addition, it must also belong to the *i*-class of verbs.

FIGURE 2. Sample rule-and-feature set with mutually exclusive features enclosed in curly brackets.



Let’s look at the first block in the basic model, shown in Figure 3. Any stem, or, in a more general sense, any word-initial set of phonemes, may enter the block, but the given stem (or phonological string) may only proceed further through the blocks if its own semantic and grammatical specifications do not contradict one of the feature sets in the block. The initial phonemes of the verb *hablar*, for example, are /aβl/, and its associated feature set would specify a semantic value of ‘speak’, as well as being specified as *a*-class. Since *hablar* is a categorical/regular verb and has no alternate stems as some verbs do, there would be no other specifications in its feature set that would bar the application of the rules in this block, except for the *a*-class specification of *hablar*, which would bar the application of any set that includes *e/i*-class in its feature set.

FIGURE 3. First block of the proposed Basic Model for Spanish verbs.



For any categorical/regular verb, there is one and only one stem available to enter the blocks for the given verb: the infinitive form minus the /ar/, /er/, or /ir/ ending. It should be noted, however, that the infinitive form is not necessarily considered here as a ‘base’ form from a psychological standpoint, and is, in fact, generated in the proposed model by rules presented later in the model (but not in the present introduction to the model). Bull comments that “Spanish verb forms are always made up of three parts, a stem plus two suffixes,” adding that “This is readily observable in the infinitives and both participles” (1965: 112). Thus, the infinitive itself begins with a stem. Furthermore, Alvar and Pottier (1983: 215) maintain that the Infinitive is not really a verb form in the strict sense of the term ‘verb’, since the Infinitive is unspecified for tense and person and actually functions as a noun.

The Infinitive form is used here only as a reference for identifying the default stem, meaning the stem that applies if there is no alternant stem specified. For the purposes of the present analysis, it is convenient to refer to the Infinitive stem as a reference point, since the Infinitive form is the form typically listed in dictionaries and vocabulary lists. This default stem is the initial variable for any given verb in the proposed model unless there is a rule that specifies another stem, hence another variable. For categorical/regular verbs, such as *hablar*, *comer*, or *vivir*, however, we can proceed through the first block with no problems. For the first block of the proposed basic verb model shown in Figure 3, any ‘regular’ stem may enter.

Although the investigative process in this analysis involves looking for generalizations and finding the lowest common denominators in the system and working from the most general to the most

specific, the model itself is essentially organized from the most specific to the most general, as in Anderson's (1992) model. A careful examination of the sample block shown in Figure 3 shows that all of the rules in the block are mutually exclusive. Rule 1, and only Rule 1, yields the 1 Singular Present Indicative form, such as *hablo* or *vivo*. This form ending with /o/ is the most specific in the block, since, unlike the other forms in the block, its feature set contains person and number (PN) as well as tense and mood (TM). The rule-and-feature sets do not necessarily apply in the order in which they appear in a given block, but rather, from most specific to least specific. No other rule can apply within this sample block; hence, we say that the rule-and-feature sets in the model are disjunctively ordered within each block, as they are also in Anderson's (1992) model. This ordering of rules from most-specific to least specific (or least-general to most-general) is consistent with Bybee's lexical continuum as discussed above.

The rules are mutually exclusive within a given block; that is, only one rule within a block is possible. Once a rule applies in a given block, the resulting form, or phonological string, exits the block. This string of phonological material is the new /X/ variable that proceeds to the next block. After a rule in Block I applies, the resulting form exits the block and proceeds to the next block, and this exiting form is the new variable for the next block. However, in the case of Rule 1 of Block I, the resulting form is the final form for that particular rule-and-feature set, such as *hablo*, *canto*, *bebo*, *vivo*, since, for a 1 Singular Present Indicative form, no other rule in any successive block has a matching feature set.

Another rule in the block, Rule 4 (and only Rule 4), yields the basic Imperfect Indicative form for *e/i*-class verbs, such as *bebía* or *vivía*. Person and number are not specified at this point, since the model seeks the lowest common denominators and aims for as many generalizations as possible. The phonological material that signifies PN will be added later in another block (as shown in Figure 4 below), except in the case of the 1 Singular and 3 Singular forms. The 1 Singular and 3 Singular forms (for the Imperfect Indicative for *e/i*-class verbs) do not need to be specified as such in any of the rule-and-feature sets, since the form (such as *bebía* or *vivía*) produced by Block 1 is also the final form in the case of 1 Singular and 3 Singular Imperfect Indicative for *e/i*-class verbs. So if no further phonological material is added on, as specified by further rule-and-feature sets in subsequent blocks, then the form that exits from Block I (such as *bebía* or *vivía*) then exits the final block as is, as a sort of default form, with a 'zero morph', if we wish to use traditional terminology. We might formulate a rule for such 1 Singular and 3 Singular forms, specifying a zero morph, such as /X/ → /X + Ø/ for the feature set [+ 1 Singular] or [3 Singular]. However, in the interest of simplifying the rules and not including any superfluous rules, it is more economical not to include such a rule, but to allow the 1 Singular and 3 Singular forms to exit as default forms, unspecified for PN.

This zero-morph, default form is interesting, given the fact that a child learning his or her first language is likely to hear and use the first and third persons singular more than any other. In addition, according to Bybee (1985:50;59-60), the 3 Singular Present Indicative form seems to be the default verb form in first language acquisition for speakers of Spanish. According to Bybee, it is this form that Spanish-speaking children learn first, and it is this 3 Singular Present Indicative form that children use in place of all other verb forms until they acquire the other forms. As noted earlier, Spencer (1997:218) also comments that the 3 Singular form is the default PN specification "in many inflectional systems."

While the rules within each block are disjunctively ordered, the blocks themselves are conjunctively ordered, meaning that if a rule can apply in a given block, then it must apply, and then the form proceeds to the next sequentially ordered block. For example, Rule 4 of Block 1 above must apply for a feature set that includes [+Imperf, + Indic, + *e/i*-class], producing Imperfect Indicative forms such as *bebía*, *vivía*, *escribía*, not yet specified for person and number. Once this rule has applied, no other rules in that block can apply, and the resulting form then must exit the block and proceed to the next sequentially (conjunctively) ordered block, and so on through all of the blocks in the model.

The Elsewhere Principle applies within all of the blocks, meaning that application of a more specific rule blocks that of a later more general one (Anderson 1995: 132), thus assuring mutually exclusive affixation, as with the 1 Singular Present Indicative form of Rule 1 of Block 1 above. Since, for example, the specification of the feature set [+1Sg, + Pres, + Indic] is more specific than just the

specification of [+ *a*-class], the rule /X/ → /X + o/ takes precedence over a more general rule anywhere in the model that specifies [+ *a*-class]. Therefore, only the more specific form (in this case, a 1 Singular Present Indicative form such as *hablo, canto, bebo, vivo*) can be produced. This more specific rule/form blocks application of the more general rule stating /X/ → /X + a/ for *a*-class verbs.

4. Further clarifications

Each of the feature sets given in the blocks is part of the semantic feature set conceptually associated with the *signifié* of a whole inflected word. A feature set for an actual word would actually include all of the semantic features associated with the given word as well as the so-called ‘grammatical’ features. In this morphological analysis, however, only the grammatical features (including verb class) are listed in the feature sets, since the phonological material corresponding to these ‘grammatical’ features are the constants in the model, while the ‘content’ part of the word is the variable in this model. It should be noted, however, that in the model proposed here, these grammatical features are seen as part of the whole set of semantic features, not as something separate. This stipulation is primarily for the purpose of emphasizing that there is not always a one-to-one correspondence between function and form, and because the proposed model accepts a whole-word view of the lexicon. Just as we are not accustomed to breaking down the so-called ‘content’ part of a word into discrete morphemes, and in fact it is not possible in some words (such as *dog* or *casa* ‘house’), it is not always possible to isolate and break down inflectional material in this way either. For example, as mentioned above, Spanish Preterite affixes are not easily separable in terms of person, number, tense, mood, and aspect, but rather, the indicators for PN and TMA are fused together.

It should also be noted that the designation of verbs as *a*-class, *e*-class, or *i*-class is considered here to be part of the semantic/grammatical/collocation feature set of a given verb, and would be considered part of a native speaker’s linguistic competence, although the average speaker himself probably wouldn’t consciously divide verbs into classes in this way. Spencer (1997:217) comments that “deciding which ... formatives are to be regarded as inflections proper” is “a theory-internal choice.” In discussing Italian verbal morphology, he says that

The theme elements [that is, vowels, as in Spanish verb class based on the theme vowels *a, e, i* (present author’s note)] are often thus regarded as derivational, stem-forming suffixes, which would suggest that their presence should be accounted for by other rules (namely, derivational rules in the lexicon). However, in a certain sense these theme vowels are part of the paradigm and interact with other, genuinely inflectional, suffixes in complex ways. Therefore, we shall assume that it is our morpholexical rules which introduce the theme vowels (217).

In a sense, then, it is somewhat arbitrary and a theory-internal choice to regard vowel class as part of the word-formation rules for Spanish inflectional morphology. For the present proposed verb model, it is stipulated at the outset that verb class is considered part of the semantic feature set, thereby perhaps slightly relaxing the definition of ‘feature set’.

If there is not some sort of mental grouping of verbs according to vowel class, then it is hard to explain how the ‘correct’ vowel comes to be used in the verb endings. In Bybee’s (1985, 1988) analogical model, verbs would be grouped together according to vowel class, or theme vowel, by virtue of their phonological similarities. In Bybee’s model, it does not matter whether there is any semantic significance to the vowels, since her model is based on semantic and/or phonological connections among forms. However, connections are stronger where there is semantic and phonological identity. Furthermore, Bybee’s (1985) gradient view of inflection vs. derivation means that both of these morphological processes occur on a continuum between lexical expression on the one end and syntactic expression on the other, with both inflection and derivation somewhere in between. In this view, then, it is not so important to distinguish between derivation and inflection regarding the theme vowels.

The reason the term ‘semantic/grammatical/collocation features’ is used here in reference to the feature sets of the word-formation rules is that although the designation of verb class has no descriptive content and does not convey any real semantic information, verb class does implicate certain specified forms that must be collocated with the given form in question. The notion of ‘collocation’ used in the present analysis extends to neighboring phonological material that may be

part of a given word, as well as to the standard notion of whole words that are neighbors of a given form or word, such as a particular preposition that typically follows a given verb. Collocations, at least in this broad sense of the word, are relevant to the present morphological analysis in that they represent regular patterns of association among forms in the lexicon. For example, the segment /fa/ in Imperfect Indicative forms (in Rule 4 of Block I in Figures 3 and 4) requires that it be collocated only with forms that fit in with the *e/i*-class verb patterns.

Although all of the feature sets in a given block are mutually exclusive, meaning that only one can apply, it is not necessary that one of the sets apply in a given block. If none of the feature sets in a block applies, then the phonological form entering the block exits the block unchanged. If there is never a specification in the feature sets for 1 Singular or 3 Singular after the block in which *hablaba*, *bebía* or *vivía* are produced, then these forms will be the final forms for the respective verbs, as discussed above. When one of the rules does apply in any given block, however, then the form that exits that block enters the next block as the new variable /X/ and then it undergoes some changes, adding phonological material, so that a different form exits the final block. In the case of the above three forms, *hablaba*, *bebía* and *vivía*, later blocks have other PN specifications in their feature sets, such as the one that adds /s/ for 2 Singular and the one that adds /mos/ for 1 Plural, thus producing *hablabas*, *bebías* and *vivías* for 2 Singular Imperfect Indicative and *hablábamos*, *bebíamos* and *vivíamos* for 1 Plural Imperfect Indicative.

5. The Basic Model: Present Indicative, Present Subjunctive, Imperfect Indicative, and Imperfect Subjunctive

For all three Spanish verb classes (*a*-class, *e*-class, *i*-class), the basic model presented in Figure 4 below is capable of producing the following forms for all categorical/regular verbs: Present Indicative, Present Subjunctive, Imperfect Indicative, and Imperfect Subjunctive. At the outset, upon entering the blocks, /X/ = the content-bearing stem, which for most Spanish verbs is the infinitive form minus the /ar/, /er/ or /ir/ ending. Because the present analysis is just a basic introduction to the proposed Spanish verb model, stem allomorphy and suppletion are not dealt with here.

For the sake of easy recognition of words by the reader and to avoid confusion, conventional spelling (with no slashes, of course) is used when giving an orthographic representation in the text, as opposed to using slashes for the phonological representation of a word or string of phonemes. In the rule-and-feature sets in the model itself, however, phonological representations are used, including the use of /y/ for the high front glide instead of the orthographic <i>. This will be important later in the discussion of the possible influence of *yod* (the semi-vowel or glide [y]) in producing some phonological stem variants of certain verbs, although such phonological variants are not discussed here in this basic introduction to the model.

FIGURE 4. Basic Model: Present Indicative, Present Subjunctive, Imperfect Indicative, and Imperfect Subjunctive.

Block I

1) $\left[\begin{array}{l} +1 \text{ Sg} \\ + \text{ Pres} \\ + \text{ Indic} \end{array} \right]$	2) $\left\{ \left[\begin{array}{l} + \text{ Pres} \\ + \text{ Subj} \\ + a\text{-class} \end{array} \right] \right\}$ + <i>e/i</i> -class	3) $\left\{ \left[\begin{array}{l} + \text{ Pres} \\ + \text{ Subj} \\ + e/i\text{-class} \end{array} \right] \right\}$ + <i>a</i> -class
/X/ → /X + o/	/X/ → /X + e/	/X/ → /X + a/

4) $\left[\begin{array}{l} + \text{Imperf} \\ + \text{Indic} \\ + e/i\text{-class} \end{array} \right]$	5) $\left[\begin{array}{l} + \text{Imperf} \\ + \text{Subj} \\ + e/i\text{-class} \end{array} \right]$	6) $\left[\begin{array}{l} + \{ 1 \text{ Pl} / 2 \text{ Pl} \} \\ + i\text{-class} \end{array} \right]$
/X/ → /X + ía/	/X/ → /X + ye/	/X/ → /X + i/

Block II

1) $\left[\begin{array}{l} + \text{Imperf} \\ + \text{Indic} \\ + a\text{-class} \end{array} \right]$	2) $\left[\begin{array}{l} + \text{Imperf} \\ + \text{Subj} \end{array} \right]$
/X/ → /X + ía/	/X/ → /X + ra/

Block III

1) $\left[+ 2 \text{ Sg} \right]$	2) $\left[+ 1 \text{ Pl} \right]$	3) $\left[+ 2 \text{ Pl} \right]$	4) $\left[+ 3 \text{ Pl} \right]$
/X/ → /X + s/	/X/ → /X + mos/	/X/ → /X + is/	/X/ → /X + n/

Let's look at an example of a form produced by these blocks of rule-and-feature sets. For a feature set specifying the semantic notion of 'dance', for example, with a TMA value of Present Subjunctive and a PN value of 2 Singular, the stem entering the first block would be /bayl/, the only stem in the lexical stem set for the verb *bailar*. This stem would be specified as [+ a-class] and would enter Block I as the X variable, where the only set that matches is the set in Rule 2, stating that /X/ → /X + e/. Thus the form exiting Block I is /bayle/. This form now enters Block II as the new X variable, where we see that there are no matching feature sets. The form /bayle/ then exits the block with no changes and then enters Block III as the X variable for that block, where Rule 1 has the matching feature set of [2 Sg]. Therefore, the final form exiting the last block is /bayles/, conventionally spelled *bailes*.

Block III of Figure 2 shows the lack of need for specification of zero-morphs in the model. There is no feature set given for 1 Singular or 3 Singular, since the X variable exiting Block II and entering Block III is itself the 1 Singular and 3 Singular form for all of the above tenses and moods, except for the 1 Singular Present Indicative, which dropped out in the first block as the most specific rule in the block. For example, the [3 Sg Pres Indic] form for *bailar* is *baila*, which resulted from Block I but added nothing in Blocks II or III. There is not a more general rule that could have applied in a later block to the 1 Singular and 3 Singular, because there are no later feature sets that include either of these two features. Hence, the form exiting the final block as a sort of default form is itself the 1 Singular and 3 Singular form for the other tenses and moods specified above.

An astute Spanish-speaking observer might think at this point that the model has produced an unattested form in the 2 Plural Present Indicative of *i*-class verbs, since according to the rules in the above blocks, the 2 Plural Present Indicative form for the verb *vivir* ('live') would be *viviis*, with a double <i>. Although this is not the correct spelling of the form, it is in fact the correct pronunciation if we regard each <i> as a separate syllable. If so, then the model has not produced an ungrammatical form after all, because the penultimate syllable, meaning the next-to-the last /i/ in this case, would be stressed according to normal Spanish stress rules, yielding *viví-is*. (The hyphen is inserted here just to show hiatus, not to indicate spelling.) If this is true, then what happens is that the unstressed second /i/ in the affix is deleted in speech, following a natural language tendency toward lenition or deletion of unstressed vowels in final syllable position (Lathrop 1980:66), and is thus not represented in the traditional orthographic representation of the form. Therefore, the form must be spelled with a written accent over the <i> in the affix, since with only one <i>, instead of two treated as separate syllables, the normal stress rules are broken.

If we were to create an extra rule which would generate the form with just one /i/ instead of two, then the stress would be wrong. According to the natural stress rules of Spanish, the penultimate syllable is stressed if the final consonant is /s/, contrary to the pronunciation of the actual form *vivís*. An additional rule to create a form with only one /i/ in the affix would yield */βiβiβis/, with the first syllable (the stem syllable) stressed as per the normal Spanish stress rule, instead of the actual form /βiβiβis/ with the last syllable stressed. However, if we leave the two /i/'s there in the form generated by the model, as in /βiβiβis/, and assume hiatus rather than diphthong (since each /i/ is affixed independently in the model), then, according to normal Spanish stress rules, we do indeed get stress on the penultimate syllable, which is the first affixed /i/. Then if the unstressed second /i/ in the affix is deleted in speech due to lenition, then the model does yield the form orthographically represented as *vivís*, stressed on what is at this point the last syllable. According to conventional orthography, this written form must then be spelled with a written accent over the <i> in the last syllable, since it does not conform to normal Spanish stress patterns.

6. Conclusion and significance of the model

The model introduced in the present analysis is a viable representation of the entire system of inflectional verbal morphology in Spanish. Although the introduction presented here represents only part of the model, the entire model does succeed in condensing and representing the patterns of Spanish verbal inflection. It is capable of producing all attested forms without producing any unattested forms, and without any unnecessary redundancy or superfluous rules. The above rules in the basic model produce all of the categorical/regular forms, that is, all of the forms for verbs traditionally viewed as completely 'regular', for all verbs in the Present Indicative, Present Subjunctive, Imperfect Indicative, and Imperfect Subjunctive. Further expansion of the model presents word-formation rules for other tenses and moods, including phonologically variant and suppletive forms, by specifying in the feature sets which stem variant belongs with which feature set.

The proposed model demonstrates the relationships among forms in the lexicon, such as *hablamos, comemos, escribimos*, all of which share /mos/ and the semantic/grammatical notion of [1PI], or *vivía, vivían, comía, comíamos, escribíamos, escribías, bebías*, which share the phonological string /ía/ and the semantic/grammatical concept of [e/i-class Imperf Indic]. Such lexical relationships are consistent with Bybee's (1985, 1988) analogical model of morphology as lexical organization. Furthermore, the ordering of rules from most specific to most general is consistent with Bybee's (1985) continuum of lexical representations, with generalized rules at the 'least specific/most general' end of the continuum.

The proposed model uses the basic structure of Anderson's (1985, 1988, 1995) rule-based generative model, but at the same time, it demonstrates that it is consistent with Bybee's (1985, 1988) analogical model. Thus, the proposed model reconciles a rule-based approach with an analogical, exemplar-based approach to morphology.

References

- Alvar, Manuel, and Bernard Pottier. 1983. *Morfología histórica del español*. Madrid: Biblioteca Románica Hispánica, Editorial Gredos.
- Anderson, Stephen R. 1985. Inflectional morphology. [In: Shopen, T. (ed.). *Language typology and syntactic description*, v III, *Grammatical categories and the lexicon*. Cambridge: University Press. 150-201.]
- Anderson, Stephen R. 1986. Disjunctive ordering in inflectional morphology. *Natural Language and Linguistic Theory* 4. 1-32.
- Anderson, Stephen R. 1988a. Inflection. [In: Hammond, Michael, and Michael Noonan (eds.). *Theoretical Morphology*. San Diego: Academic Press. 23-43.]
- Anderson, Stephen R. 1988b. Morphological theory. [In: Newmeyer (ed.). *Linguistics: The Cambridge Survey*, v I, *Linguistic Theory: Foundations*. Cambridge: Cambridge University Press. 146-191.]
- Anderson, Stephen R. 1995. *A-Morphous Morphology*. Cambridge: University Press.
- Bull, William. 1965. *Spanish for Teachers*. New York: Ronald Press Co.
- Bybee, Joan. 1985. *Morphology: a Study of the Relation between Meaning and Form*. Amsterdam: Benjamins.

- Bybee, Joan. 1988. Morphology as lexical organization. [In: Hammond, Michael, and Michael Noonan (eds.). *Theoretical Morphology*. San Diego: Academic Press. 119-142.]
- Eddington, David. 1999. The role of experiments in phonological investigation. *Journal of Quantitative Linguistics* 1999, 6: 14-28.
- Eddington, David. 2000. Analogy and the dual-route model of morphology. *Lingua* 2000, 110: 281-298.
- Hammond, Michael, and Michael Noonan (eds). 1988. *Theoretical Morphology*. San Diego: Academic Press.
- Lathrop, Thomas A. 1980. *The Evolution of Spanish: An Introductory Historical Grammar*. Newark, DE: Juan de la Cuesta.
- Newmeyer, Frederick J. (ed). 1988. *Linguistics: The Cambridge Survey*, v I, *Linguistic Theory: Foundations*. Cambridge: Cambridge University Press.
- Politzer, Robert L. 1965. *Teaching Spanish: A Linguistic Orientation*. Waltham, MA: Blaisdell.
- Real Academia Español. 2002. Formas verbales. [Online.] Available: <http://dlc.rae.es> [accessed August 4, 2002].
- Saussure, Ferdinand de. 1916. *Course in General Linguistics*. [Bally, Charles, and Albert Sechehaye (eds.). Baskin, Wade (trans). New York and Toronto: McGraw-Hill. 1959.]
- Skousen, Royal. 1989. *Analogical Modeling of Language*. Dordrecht, The Netherlands: Kluwer.
- Skousen, Royal. 1995. Analogy: A non-rule alternative to neural networks. *Rivista di Linguistica* 7: 213-232.
- Spencer, Andrew. 1997. *Morphological Theory*. Oxford: Blackwell.
- Weinrich, Harald. 1964. *Estructura y función de los tiempos en el lenguaje*. [Versión española de Federico Latorre. 1968. Madrid: Editorial Gredos.] (*Tempus, Besprochene und Erzählte Welt*. Stuttgart: W. Kohlhammer Verlag.)

Selected Proceedings of the 7th Hispanic Linguistics Symposium

edited by David Eddington

Cascadilla Proceedings Project Somerville, MA 2005

Copyright information

Selected Proceedings of the 7th Hispanic Linguistics Symposium
© 2005 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 1-57473-403-2 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Burdette, Karen W. 2005. New Paradigms: A Rule-and-Feature Based Morpholexical Model of the Spanish Verbal System. In *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, ed. David Eddington, 158-168. Somerville, MA: Cascadilla Proceedings Project.

or:

Burdette, Karen W. 2005. New Paradigms: A Rule-and-Feature Based Morpholexical Model of the Spanish Verbal System. In *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, ed. David Eddington, 158-168. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1095.