



Spanish Inflectional Morphology in DATR^{*}

ANTONIO MORENO-SANDOVAL

Department of Linguistics, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain
E-mail: sandoval@maria.illf.uam.es

JOSÉ MIGUEL GOÑI-MENOYO

Department of Applied Mathematics, Escuela Técnica Superior de Ingenieros de Telecomunicación,
Universidad Politécnica de Madrid, 28040 Madrid, Spain
E-mail: jmg@mat.upm.es

(Received 11 January 1999; in final form 25 October 2000)

Abstract. This paper shows a full description of Spanish inflectional morphology. We have chosen a paradigmatic approach instead of one based on phonological/spelling changes, i.e., the typical two-level model. Such morphological description has been written in the DATR formalism. The result is a network of nodes that makes use of the information inheritance mechanisms – orthogonal node inheritance and default path inheritance – that DATR allows. Some lexical coverage and corpus occurrence figures that support our approach are also given.

Key words: DATR formalism, information inheritance, paradigmatic morphology, Spanish

1. Introduction

We present an explicit treatment of inflection for Spanish based on the following points:

- Languages show a wide variation, both cross-linguistic and intra-linguistic, in the kind of morphologically encoded information. Therefore, constructing both a *universal* morphological model and a *language-specific* model is extremely difficult.
- Spanish is a language with an important inflecting component, which is both paradigmatic and rule governed. In addition, it shows a considerable amount of *syncretism*, and also suppletion and defective forms. In theoretical morphology it was proposed a long time ago (Matthews, 1972) that the best approach to those phenomena is the *Word-and-Paradigm* model (WP) (Carstairs, 1987; Wurzel, 1989).
- Both nominal and verb paradigms, even though they have been described traditionally in a myriad of different models, in fact exhibit a great generalization in their irregularities. This was first shown for Spanish by Bello (1898).

^{*} This work has been supported in part by the Spanish *Plan Nacional de I+D*, through the project TEL96-1367, ATILA: *Telematic Applications of Linguistic Engineering*.

Recently, Pirrelli and Battista (1996) have also shown this fact for Italian, a very close cognate of Spanish, proposing a single schemata for both regular and most of the irregular verbs.

- In the DATR framework, a morphological theory, Network Morphology, has arisen (Corbett and Fraser, 1993; Fraser and Corbett, 1995; Brown et al., 1996). This theory makes use of inheritance (Briscoe et al., 1993) and overwriting in order to describe morphology, and it has been successfully implemented for Russian noun declension, and German adjectives and determiners (Cahill and Gazdar, 1997).

A finite-state model can handle Spanish inflection, as previous work (Tzoukermann and Liberman, 1990; Carulla and Oosterhoff, 1996) has shown. We will show in this paper that a paradigmatic model based on DATR can also handle Spanish inflection.

A previously (known) version of Spanish inflection in DATR has been written by Poch* (Poch, 1992), but we want to stress that our approach is a radically different one: Poch's verb treatment is based on phonological changes, whereas ours is based on word form distribution. On the other hand, Gómez Guinovart and Aguirre Moreno (1998) show a DATR description for Galician (a cognate language of Spanish) that shares several aspects with our approach, for instance, the use of allomorph concatenation. However, their approach is different to ours from the descriptive point of view, since they use *zero allomorph* and split the verb suffix into three segments (thematic vowel, mood-tense and person-number suffixes), however we do not use this type of strategy.

2. The Morphological Framework

We have chosen the WP model for Spanish morphology because it has advantages over a model based on phonological rules (Item-and-Process: IP). Historically, the paradigmatic model was created for inflecting languages (in fact, Greek grammarians developed the model for their language).

First, we will try to show, with quantitative and qualitative arguments, why the paradigmatic model is the most comprehensive one for the treatment of Spanish inflection.

In general, the paradigmatic model competes successfully when dealing with the following phenomena:

1. Morphologically conditioned allomorphy.
2. Suppletion.
3. Syncretism.

* This version is included in the DATR example files distributed with the Brighton/Sussex Prolog implementation (Evans, 1990; Jenkins, 1990a). This DATR material, and more, can be recovered from the URL <ftp://ftp.cogs.sussex.ac.uk/pub/nlp/DATR>.

4. Defectiveness.
5. Discontinuousness.*

The first two affect allomorphy (Section 2.1). The last two will be discussed in separate subsections (2.2 and 2.3).

Finally, in Section 2.4 we will show some concepts of current paradigmatic morphology through the perspective of their treatment in DATR.

2.1. ALLOMORPHY

Allomorphy, or the alternation of different realizations of a given morpheme, is a basic concern for both a descriptive and a computational model for a language. This alternation is usually analysed in terms of the type of conditioning that brings it about:

- *Phonologically conditioned allomorphy*, that is, alternations which can be defined purely in terms of the phonological (or orthographical) context.
- *Morphologically (or grammatically) conditioned allomorphy*, that is, allomorphic variation triggered only by specific morphemes or a particular grammatical element.
- *Lexically conditioned allomorphy*, i.e., the use of a particular allomorph is specific for a given word (e.g., irregular plurals in English such as *ox, oxen* or irregular participles in Spanish such as *imprimir, impreso – to print, printed*). The most extreme case of lexical conditioning is *suppletion*, where allomorphs of a morpheme are phonetically unrelated, as in *good, better*.

The main difference between these types is the difficulty in setting the context of the rule application, either phonological, grammatical or lexical. We show it through some examples from Spanish.

First, the three phonologically conditioned allomorphs for plurals can be described with a rule for each:**

{s}	-->	/s/		[a-u, á-ú] _	e.g.,	casa, casas;
						bambú, bambús
{s}	-->	/es/		[C, á-ú] _	e.g.,	sol, soles;
						bambú, bambúes
{s}	-->	∅		s _	e.g.,	lunes, lunes

where

* For instance, infixation (the splitting of the root by an affix) and transfixation (discontinuous grammatical morphemes). Those phenomena are typical of some fusional and inflecting languages such as Semitic ones. Since Spanish does not present any of these in inflection, we will not comment on them further.

** Words ending in accented vowels usually have two alternative forms, both of which are correct.

Table I. Distribution of verbal entries by type.

Allomorphy type	NL	PL	NC	PC
Regular verbs	5,254	71.21%	19,238	42.18%
Phonological	1,434	19.44%	8,342	18.29%
Morphological	629	8.53%	9,990	21.91%
Lexical	61	0.83%	8,034	17.62%

NL: Entries in lexicon. PL: Percentage in lexicon (types).

NC: Instances in corpus. PC: Percentage in corpus (tokens).

a-u = {a, e, i, o, u}

á-ú = {á, é, í, ó, ú}

C is any consonant except s

In addition, general spelling changes can be easily handled in the same fashion:

[á-ú]n --> [a-u]nes e.g., atún, atunes
 [á-ú]s --> [a-u]ses e.g., revés, reverses
 z --> ces e.g., pez, peces

Since the allomorph selection in these cases is always regular, using a rule-based model for the plural is economical.

Secondly, there are other inflection phenomena where the context is hard to state. Considering again the verb stem allomorphy: some very frequent Spanish verbs have grammatically conditioned irregularity, like *querer* (*to want*). *Querer* has three allomorphs, *quer-*, *quier-* and *quis-*. The proper stem allomorph to be concatenated to the suffix *-o* for the first singular person present is *quier-* to form *quier-o*, instead of the regular stem allomorph *quer-* that would form the agrammatical **quer-o*.

However, there is no phonological context that always changes *quer-* to *quier-* when it is followed by *-o*, since *barquero* (*boatman*) is correct in Spanish. This is a case of grammatically conditioned allomorphy, because the suffix selects the stem allomorph.

Suppletion is the most interesting case of lexically conditioned allomorphy. The Spanish verb *ir* (*to go*) is an example where all the paradigm has to be present in the lexicon because there is no phonological relation between, for instance *voy*, *fui*, *iré* (*I go*, *I went*, *I will go*). Fortunately, there are very few cases like this in Spanish.

In Table I we show figures for each of these types of verb allomorphy in Spanish with respect to:

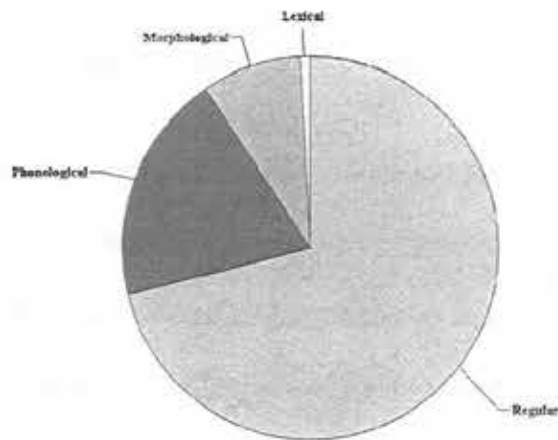


Figure 1. Spanish verb distribution in lexicon.

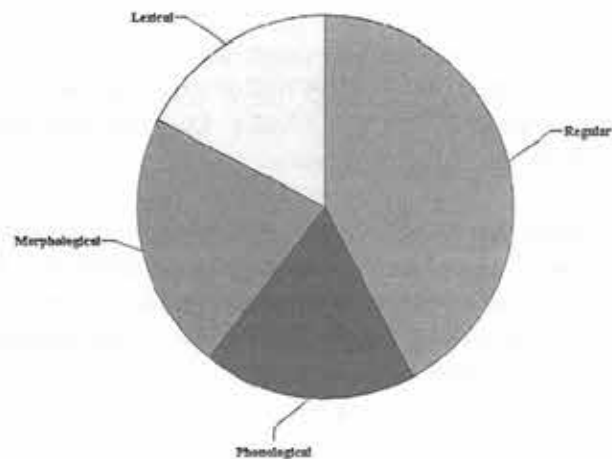


Figure 2. Spanish verb distribution in corpus.

Competence: The first two columns show, respectively, the count of verbal entries in our Spanish lexicon and their percentage* (see also Figure 1).

Performance: The last two columns show, respectively, the count of verbal word forms in a Spanish corpus and their percentage** (see also Figure 2).

* The Spanish ARIES lexicon is based on the morphological model proposed in the Ph.D. thesis of Moreno Sandoval (1991) and is widely described by the one of Goñi Menoyo (1998). It contains 7,378 verb lemmas, 21,386 nouns, and 10,284 adjectives.

** For this experiment we have used a fragment of the Spanish CRATER corpus (McEnery et al., 1997), consisting of 478,828 tokens. It contains 45,604 verb word forms. The corpus used within CRATER contains the International Telecommunications Union CCITT handbook, also known as The Blue Book, in English, French and Spanish versions. Therefore, the corpus is somehow biased and not representative of Spanish, but we think it is useful for our purposes anyway.

Later, in Section 4, we will show our complete set of verbal paradigms with the associated allomorphy type and the competence and performance figures.

From these data we can infer the following:

- Regular verbs have more importance in the lexicon (71%) than in the actual performance (42%).
- Phonologically conditioned allomorphy represent approximately 20 percent of the Spanish verbs both in the lexicon and in the corpus.
- Morphological allomorphy is more relevant in performance (22%) than in competence (9%).
- Lexical allomorphy also produces better scores in performance than in competence (18% vs. 1%).

Although the percentage of verbs with phonological and orthographical allomorphy (19%) doubles the percentage of morphological and lexical types (9%) in the lexicon, we find the contrary in performance terms, where phonological and orthographical allomorphy percentage (18%) is half that of the others (40%).

From the computational perspective, it is very usual to use finite-state models to handle allomorphy.* We think this is for two reasons:

1. The theoretical assumption that “*normally, the distribution of allomorphs is phonologically conditioned*” (Katamba, 1993). Most of the Generative Morphology assumes the IP approach, where allomorphy is handled by one or several morphological rules that convert a underlying structure (the morpheme) into its surface form (the allomorph).
2. The practical issue that the Two-Level model (Koskenniemi, 1983) and its augments and improvements** have been successfully applied to many languages, including Spanish.

In summary, the Two-Level model is generally considered to be the universal computational model for morphology, equally, IP is regarded as the universal descriptive model. Our point is that a true universal model does not as yet exist due to the huge intra and cross-linguistic variation.

As we have shown above with quantitative data, morphologically and lexically conditioned allomorphy is clearly more important in performance than phonological allomorphy. Of course, it is possible to write phonological context rules for the treatment of morphological irregularities. This implies that many rules would have to be written, and the majority of them would be applicable to very few

* As Sproat (1992) clearly states: “*a great deal of the effort in constructing computational models of morphology has been spent on developing techniques for dealing with phonological rules. For some computational models, such as KIMMO, the bulk of the machinery is designed to handle phonological rules.*”

** See Sproat (1992) for a detailed summary.

cases, necessarily marked in the lexicon. This strategy would go against elegance and economy in the grammar, and would probably affect the efficiency of an implementation.

On the other hand, phonological allomorphy can be handled in an economical fashion within a paradigmatic model, since all the phonological changes have a similar distribution in the paradigms.

Consequently, for those types of languages with a significant presence of non-phonological allomorphy, the finite state-IP approach would not be the most appropriate in terms of elegance and simplicity. Of course, this does not imply that the paradigmatic model, which we defend for Spanish verb inflection, is universal.

2.2. SYNCRETISM

Syncretism is the fusion of more than one morpheme into the same morpho. While allomorphy affects both the lexeme and the inflecting morphemes, syncretism is only applied to the latter.

Non-paradigmatic models typically segment words into discrete morphemes that assume a symmetrical relationship, that is, a morpheme (or lexical representation) maps onto a morpho (or surface representation). The problem arises when asymmetrical relationships take place, such as syncretism, where the same signifier expresses several signifieds (meanings). In those cases, a clear segmentation is not possible.

An excellent example in Spanish is the morpho /-o/:

1. am-o, com-o, part-o
2. quis-o, pus-o
3. muert-o, frit-o
4. gat-o, bonit-o

In (1), the /-o/ means first singular person present indicative; in (2), third singular past indicative; in (3), a past participle; and in (4), it is an allomorph for masculine.

In (1) and (2), for example, the same morpho (/ -o/) stand for four morphemes (person, number, tense and mood). Obviously, we cannot split such a morpho in four segments.

Usually, morphological descriptions based on discrete units need a special zero morpho in order to preserve coherence. In this case the question is how to determine which morpheme the zero morpho corresponds to.*

Summing up, when a language has a high degree of syncretism then it is a good candidate for a paradigmatic description.

* The theoretical soundness of the zero morpho has always been extensively discussed, because of the difficulty of demonstrating empirically the existence of something that has no phonic substance. Again, this is an asymmetrical relation: one-to-zero, one meaning without signifier.

2.3. DEFECTIVENESS

Defectiveness occurs when some forms in the paradigm are missing.* As suppletion, defectiveness is an exceptional phenomenon, because it needs special storage. The speakers have to learn individually which verbs lack particular forms. Some verbs are semantically predictable, such as ones that express meteorological phenomena (e.g., *llover*, *to rain*); whereas some other are hardly predictable, such as *abolir* (*to abolish*), that only allows inflecting forms whose suffix starts with /-i/, e.g., *abol-í* (first singular past indicative) but **abol-o* (first singular present indicative).

Defectiveness is also a problem for a model that claims to be reversible, that is, suitable both for analysis and generation. If a given word form does not exist, it must be neither generated nor recognized. In a paradigmatic model, this idea is easily captured: simply leaving holes in the paradigm. In other words, defective verbs are treated as another paradigm: they are not an exception. In a rule system, it is customary to write "negative" rules that inhibit the generation of non-existent word forms.

2.4. PARADIGMATIC MORPHOLOGY

For the phenomena described in the previous sections, typical of the inflecting or fusional languages, it is better to use a paradigm rather than a morpheme concatenation model (Item-and-Arrangement: IA) or the application of processes to an underlying form (IP model). This idea is generally accepted in theoretical morphology,** since it seems the simplest and most comprehensive approach to the description of asymmetric relationships between a morpheme and its allomorphs. The preference for the paradigmatic model in these cases is so despite the fact that it is probably the less universal model compared to the other two (Beard, 1995), and on the other hand, its formal expressiveness is also too powerful, thus allowing non-existent languages to be described (Spencer, 1991; Bauer, 1988). As a result, for both typological and formal reasons, the paradigmatic approach has had less success than the others, since for simple and regular phenomena, such as phonologically conditioned allomorphy or morpheme agglutination, it is very expensive to maintain such a redundant model.

It is well known that languages have a mixture of morphological mechanisms. Consequently, it is not realistic to assume that there is only one universal model

* Defectiveness is hard for a theoretical description, since it is not possible to reach an agreement between grammarians on the existence of some particular forms. On the other hand, a performance based description is not possible either: the lack of a particular form in a corpus can be due to non existence or to incompleteness of the corpus.

** Spencer (1991) writes: "There remains a class of phenomena which neither IA or IP seem well equipped to handle and that is the fusional nature of inflectional systems. The problem is that both IA and IP are fundamentally agglutinating theories." Bubenik (1999) also says: "Another aspect of the WP model which is totally absent from the IA model is its preoccupation with irregular and suppletive morphology."

for morphology, at least with the theoretical tools we currently have. On the other hand, it is more feasible to defend that a language can be better described with a particular model, using both qualitative and quantitative arguments.

For Spanish, we believe that the arguments given in this section are enough to justify a paradigmatic inflectional morphology model.

3. The Descriptive Framework. An Introduction to DATR

DATR was designed as a formal declarative language for representing a restricted class of inheritance networks, permitting both default and multiple* inheritance. Its main intended area of application is the representation of lexical knowledge (Gazdar, 1990; Evans and Gazdar, 1996).

The main features of DATR (Evans and Gazdar, 1990) are: (i) it has enough expressive power for encoding the entries needed by contemporary unification-based grammars; (ii) it can express the linguistic generalizations on the implicit information present in these entries; (iii) it has an explicit theory of inference**; (iv) it is computationally tractable; and (v) it has explicit declarative semantics.†

A DATR description (or theory) is a set of nodes, and each one has a list of assignments of values to various attribute paths. Values may be atoms, inheritance specifications or sequences of both.

For example:

```
C123:
  <des reg ind pres 1 sing>      == o
  <des pret1 ind indf 1 sing>    == e
  <des pret1 ind indf 3 sing>    == o.
```

describes the node C123, that encodes the fact that some verbal suffixes are shared between the three verbal conjugations (see later). Each sequence of labels‡ enclosed in <> represents a path associated to the node, and the value on the right hand side of the == is the value associated to such path. (In our example, an atom in each path). The next example§ shows how an inheritance specification is encoded:

```
C3:
  <des>                          == C23:<des>
  <des reg cond 1 sing>          == 'iría'
  <des reg cond 3 sing>          == <des reg cond 1 sing>.
```

* As will be shown, the language design enforces orthogonal inheritance, that is, no conflicts arise when inheriting information from multiple sources. Evans et al. (1993) show that multiple prioritized inheritance can be simulated, however.

** DATR has a set of seven rules of inference and a general principle of default inference. For details, see Evans and Gazdar (1989a).

† Initially proposed by Evans and Gazdar (1989b), and later re-elaborated and enhanced by Keller (1995).

‡ Appendix A shows the meanings of all these path labels.

§ It is a partial encoding of the actual C3.

The first path specification shows that the value of the <des> path in C3 is inherited from the homonymous path in the node C23. The second path specification is a regular assignation path, and the third one states that the path on the left hand side inherits the value of the path on the right hand side of the same node.*

With only the non-local inheritance feature, DATR nodes can be organized in inheritance hierarchies. Such hierarchies can be classified into the so-called *monotonic multiple orthogonal inheritance hierarchies* (Daelemans et al., 1992; Erbach, 1994). They are (i) *monotonic* since adding new node-path definitions to a DATR description will not retract previously inferred assignments to paths; (ii) *multiple* since a node can inherit information from different nodes for different paths; and (iii) *orthogonal*, since DATR syntax requires that different attributes be inherited from different sources, thus avoiding conflicts.

Apart from this monotonic inheritance device, DATR has another inheritance one, for introducing non-monotonicity into the descriptions. This feature is called *path extension*. To show how it works, let us recall the C3 node example above. Whether the value of the <des> path is inherited from another node or not, all the extensions of such a path inherit their value from the <des> path, *unless* a more specific value is provided in the node.

In our example, paths like <des reg ind pres 1 sing> inherits the value of the <des> one, whatever it is. But the explicitly declared paths <des reg cond 1 sing> and <des reg cond 3 sing> do not, since they are assigned the explicit declared values in C3 node definition.

This feature is a very powerful one, and allows us to describe some linguistic facts like subregularity, irregularity and exceptions, typical of inflecting paradigms, and introduces non-monotonicity in DATR descriptions.**

The last DATR feature that will be shown in this brief introduction to the formalism is the so-called, in DATR jargon, *global reference*. In the following example:

```
REGULAR-V:
<>                == VERB
<sfx>              == "<des reg>"
<root>            == "<stem 1>"
<wf>              == "<root>" "<sfx>"
<wf imper pol sing> == "<wf subj pres 3 sing>"
<wf imper pol plu> == "<wf subj pres 3 plu>" .
```

the <sfx> and <root> paths are inherited from the value of <des reg> and <stem 1> paths, but not in the local node. The global reference mechanism (invoked with the double quotes) causes such values to be inherited from the node

* The first and the third line of the C3 node description are inheritance descriptions. The latter is a *local inheritance* description (when the node name prefix is missing), and the former is a *non-local inheritance* description, that permits inheritance from a different node.

** This is because of the fact that if a DATR sentence is added to a description, the value inferred for a particular path could be no longer valid, provided that it is explicitly stated in the new sentence.

where a DATR query starts. It permits a particular node, such as the one shown in the example, to inherit information from a node that is in the leaf of a hierarchy. The example also shows that a path can be assigned with a sequence of two or more values, i.e., the <wf> (word form) path is built by means of the concatenation of <root> (root) and <sf> (suffix) values. This example shows the actual morphological rule for verbs in our approach. The path extension mechanism permits word formation for the different inflecting forms with just one concatenating rule.

There are several DATR implementations,* apart from the original from Sussex University (Evans, 1990; Jenkins, 1990a). DATR has been used for the construction of relevant lexical material for different applications (Cahill and Evans, 1990; Andry et al., 1992; Cahill, 1993b), for the description of several morphophonological (Gibbon, 1992; Cahill, 1993a), inflecting (Jenkins, 1990b; Gazdar, 1992; Fraser and Corbett, 1995) and derivative (Evans, 1992; Kilbury, 1992) phenomena; for lexical semantics (Kilgarriff, 1993; Kilgarriff and Gazdar, 1995) and even for concisely encoding Lexicalized Tree Adjoining Grammars (Evans et al., 1994).

With respect to its formal power, DATR is equivalent to a Turing machine, as Moser (1992d) shows, and this can lead to intractability in the DATR descriptions (Moser, 1992b). This property permits the use of DATR for simulating special operations, like negation, disjunction or equality (Moser, 1992c), multiple prioritized inheritance (Evans et al., 1993), or even interpreting languages such as PROLOG (Moser, 1992a).

4. Verbal Paradigms

The (synthetic) conjugation of the Spanish verb has 55 inflected forms. (We discard the obsolete future imperfect subjunctive, but include the two alternative realizations for the past imperfect of subjunctive and two *courtesy* imperative forms.) We will not consider here the analytic conjugation for the perfective, i.e., auxiliary *haber* inflected form plus the participle of the main verb, since it is fully predictable and regular.

Taking the suffix as the parameter for comparison, the verb conjugation is classified into three paradigms, C1, C2, and C3, traditionally known as first, second and third conjugations. They refer to verbs whose infinitive form ends in -ar, -er and -ir/ír, respectively. Paradigmatic examples, which are fully inflected in Table II, are *amar* - *to love*, *temer* - *to fear*, and *partir* - *to go away*.

From the beginning we can observe several homonymies:**

1. 1 sing present indicative suffix is the same for the three conjugations: -o. (e.g., *am-o*, *tem-o*, and *part-o*.)

* Evans and Gazdar (1996) affirm that they know a dozen different implementations.

** This supports Carstairs' Systematic Homonymy Claim: "a homonymy introduced in a fusional [language] paradigm can be treated as systematic" (Carstairs, 1987).

Table II. Regular paradigms.

Tense/Mood	1 sing	2 sing	3 sing	1 plur	2 plur	3 plur
Present indicative	am-o	am-as	am-a	am-amos	am-áis	am-an
	tem-o	tem-es	tem-e	tem-emos	tem-éis	tem-en
	part-o	part-es	part-e	part-imos	part-ís	part-en
Imperfect indicative	am-aba	am-abas	am-aba	am-ábamos	am-abais	am-aban
	tem-fa	tem-fas	tem-fa	tem-famos	tem-fais	tem-fan
	part-fa	part-fas	part-fa	part-famos	part-fais	part-fan
Past indicative	am-é	am-aste	am-ó	am-amos	am-asteis	am-aron
	tem-f	tem-iste	tem-ió	tem-imos	tem-isteis	tem-ieron
	part-f	part-iste	part-ió	part-imos	part-isteis	part-ieron
Future indicative	am-aré	am-arás	am-ará	am-aremos	am-aréis	am-arán
	tem-cré	tem-erás	tem-crá	tem-eremos	tem-eréis	tem-erán
	part-iré	part-irás	part-iré	part-iremos	part-iréis	part-irán
Present subjunctive	am-e	am-es	am-e	am-emos	am-éis	am-en
	tem-a	tem-as	tem-a	tem-amos	tem-áis	tem-an
	part-a	part-as	part-a	part-amos	part-áis	part-an
Imperfect subjunctive (alternatives)	am-ara	am-aras	am-ara	am-áramos	am-arais	am-aran
	tem-iera	tem-ieras	tem-iera	tem-iéramos	tem-ierais	tem-ieran
	part-iera	part-ieras	part-iera	part-iéramos	part-ierais	part-ieran
	am-ase	am-ases	am-ase	am-ásemos	am-aseis	am-asen
	tem-iese	tem-ieses	tem-iese	tem-iésemos	tem-ieseis	tem-iesen
part-iese	part-ieses	part-iese	part-iésemos	part-ieseis	part-iesen	
Conditional	am-aría	am-arías	am-aría	am-aríamos	am-aríais	am-arían
	tem-ería	tem-erías	tem-ería	tem-eríamos	tem-eríais	tem-erían
	part-iría	part-irías	part-iría	part-iríamos	part-iríais	part-irían
Imperative		am-a	am-e		am-ad	am-en
		tem-e	tem-a		tem-ed	tem-an
		part-e	part-a		part-id	part-an
	Infinitive	Gerund	Participle			
Non-finite	am-ar	am-ando	am-ado			
	tem-er	tem-iendo	tem-ido			
	part-ir	part-iendo	part-ido			

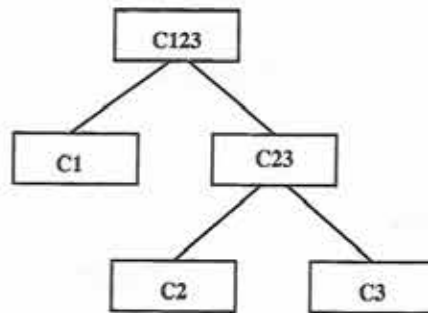


Figure 3. Conjugation node network.

2. 2 sing, 3 sing and 3 plur present indicative suffixes are the same for C2 and C3: -es, -e, -en, respectively (e.g., tem-es, part-es; tem-e, part-e; and tem-en, part-en).
3. All the word forms for imperfect indicative, past indicative, present subjunctive, and imperfect subjunctive are the same for C2 and C3 (e.g., for 1 sing present subjunctive tem-a, part-a).
4. 2 sing imperative is the same for C2 and C3 (e.g., tem-e, part-e).
5. Gerund and participle forms are the same for C2 and C3 (e.g., tem-iendo, part-iendo; and tem-ido, part-ido).

These generalizations can be captured in a network of inheritance, as shown in Figure 3. For instance, in DATR, the first homonymy is expressed as follows:

C123:
 <des reg ind pres 1 sing> == o.

C23 and C1 DATR nodes inherit this information, and analogously C2 and C3 from C23. All DATR definitions of such homonymies can be found in the actual DATR encoding (cf. Appendix A). There the reader can see that although the three conjugations have the same number of word forms, the specifications of C2 and C3 are much shorter than the one for C1, since the former ones inherit from C23.

The hierarchy shown in Figure 3 describes suffix distribution in regular paradigms. It tries to minimize redundancy exploiting homonymies. However, the Spanish language exhibits a great variety of irregularities both in stems and suffixes. For example, The American Heritage Larousse Spanish Dictionary counts 80 paradigms while Seco (1986) considers only 62.

It is important to emphasize that this large number of irregular paradigms is due to the combination of two different types of irregularities:

Stems: Many verbs present two or more allomorphs for the stem.

Suffixes: Some verbs have a few word forms with a non-regular suffix allomorph.

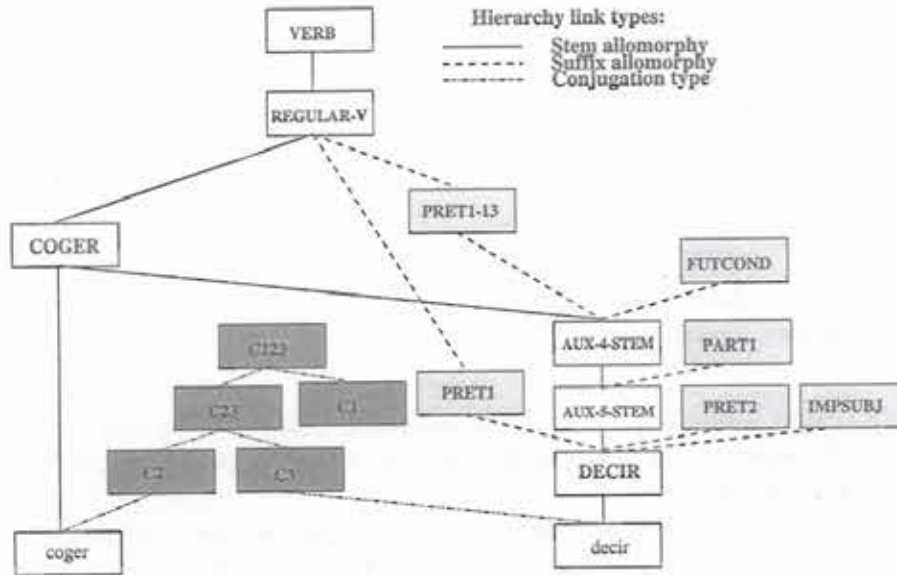


Figure 4. Fragment of verb hierarchy.

Therefore, we need two orthogonal* inheritance networks. Our proposal makes use of very few default paradigms (*macro-paradigms*) for each type of irregularity. From those *macro-paradigms* other *sub-paradigms* inherit. The complete verb paradigm is ordered into three networks of inheritance, which captures redundancy: stem allomorphy hierarchy, suffix allomorphy hierarchy and conjugation type hierarchy. The first two hierarchies merge in complex forms, while the conjugation type hierarchy meets at the lower level where verbs directly inherit information. Figure 4 shows part of such complex hierarchies.

Our approach has further advantages, in comparison to non-paradigmatic models, in handling some phenomena:

- *Duplicate alternative forms* are easily encoded by means of a special path extension. For instance, all imperfect subjunctive forms always have two alternative forms in Spanish that are encoded into the paths. For instance, paths <wf subj impf 1 sing> and <wf subj impf 1 sing alt> will render the forms *tuviera* and *tuviese* for the lemma *tener*. The appropriate endings are encoded in the conjugation models for these alternate forms. In addition, this strategy is also suitable for duplicate participles: *impreso* and *imprimido* for *imprimir*.
- *Defective verbs*, i.e., lacking some word forms in their paradigm, are also easily encoded, not allowing the definition of the missing paths in their corresponding paradigmatic model. The model *ABOLIR* is a good example, since

* In the sense explained above.

it does not allow the following word forms: indicative present (all singular and third plural), indicative future, subjunctive present and second plural imperative).

- *Suppletive forms* in the highly irregular verbs (e.g., *ser*, *ir*, *dar*, *haber*, *estar*, *andar*, and *ver*). Each of these verbs has a unique paradigm, since they override most defaults. As an example, we provide the entry for *dar* in the actual DATR description (cf. Appendix A).

To finish this section, we present in Table III* the number of verbs, present in our ARIES lexicon and in a Spanish corpus, that belong to each of our verbal paradigms.

5. Nominal Paradigms

The Spanish nominal morphology consists of two classes: gender and number. Both classes can be expressed either by a suffix (what we call *morphological gender/number*) or be lexically *inherent* or *homonymous* (that is the same word form is *gender/number invariant*). Those phenomena can be combined in several ways:

- Morphological gender and number. For instance *chic-o*, *chic-a*, *chic-o-s* and *chic-a-s* (*boy/girl*).
- Inherent gender and morphological number: *pez*, singular masculine, and *pec-es*, plural, (*fish*).
- Invariant gender and morphological number: *artista*, singular masculine and feminine, *artista-s*, plural, (*artist*).
- Inherent gender and invariant number: *crisis*, both feminine singular and plural.
- Invariant gender and number: *rubiales*, the four combinations of gender and number, (*blonde*).
- Inherent gender and number: *estrés*, only masculine singular, (*stress*). The words in this case are also known as *singularia tantum* or *pluralia tantum*.

Morphological gender is marked with three different suffixes: */-o/* and */-e/* for masculine, and */-a/* for feminine. In some cases, the masculine form does not have suffix while the feminine is marked (e.g., *doctor*, *doctor-a*), and some others also need a different stem allomorph for the feminine due to spelling changes (e.g., *león*, *leon-a*: *lion*).

Morphological plural suffix has two allomorphs: */-s/* and */-es/*, whereas singular number has no mark, that is, no morpheme is needed. In some cases stem allomorphy occurs also, because of spelling changes, for example *raíz* and

* LEXICAL and DEPECT rows in such a table do not correspond to particular paradigms, but they group the set of very irregular verbs and defectives, respectively.

Table III. Distribution of verbal entries by paradigm in lexicon (types) and corpus (tokens).

Paradigm	AT	NLEX	PLEX	NCOR	PCOR
REGULAR-V	–	5,254	71.21%	19,238	42.18%
BULLIR	P	24	0.33%	7	0.02%
LIAR	P	6	0.08%		0.00%
CAZAR	P	1,136	15.40%	6,411	14.06%
CREER	P	12	0.16%	58	0.13%
COGER	P	213	2.89%	1,189	2.61%
HUIR	P	43	0.58%	677	1.48%
REÑIR	M	10	0.14%	421	0.92%
CERRAR	M	341	4.62%	2,266	4.97%
SEGUIR	M	12	0.16%	521	1.14%
NEGAR	M	49	0.66%	184	0.40%
DORMIR	M	44	0.60%	611	1.34%
CAER	M	6	0.08%	16	0.04%
VALER	M	3	0.04%	3	0.01%
SALIR	M	3	0.04%	30	0.07%
QUERER	M	4	0.05%	5	0.01%
PODER	M	1	0.01%	2,436	5.34%
CONducIR	M	14	0.19%	542	1.19%
CABER	M	2	0.03%	11	0.02%
FREIR	M	7	0.09%		0.00%
OIR	M	4	0.05%	14	0.03%
COCER	M	7	0.09%		0.00%
MORIR	M	2	0.03%		0.00%
TRAER	M	14	0.19%	18	0.04%
SABER	M	3	0.04%	28	0.06%
TENER	M	11	0.15%	1,460	3.20%
VENIR	M	17	0.23%	214	0.47%
PONER	M	26	0.35%	632	1.39%
DECIR	M	8	0.11%	12	0.03%
HACER	M	6	0.08%	361	0.79%
VESTIR	M	27	0.37%		0.00%
VOLVER	M	8	0.11%	205	0.45%
ROMPER	L	27	0.37%	427	0.94%
LEXICAL	L	17	0.23%	7,562	16.58%
DEFECT	L	17	0.23%	45	0.10%
TOTAL		7,378		45,604	

AT: Allomorphy type. P: Phonological. M: Morphological. L: Lexical.

NLEX: Number of instances in lexicon. PLEX: Percentage in lexicon (types).

NCOR: Number of instances in corpus. PCOR: Percentage in corpus (tokens).

Table IV. Distribution of nominal entries by paradigm in lexicon (types) and corpus (tokens).

GT	NT	MODEL	NOUNS				ADJECTIVES			
			LEXICON		CORPUS		LEXICON		CORPUS	
M	M	CHICO	580	2.71%	849	0.81%	6,301	61.27%	11,734	50.51%
M	M	NENE	7	0.03%	0	0.00%	3	0.03%	0	0.00%
M	M	DOCTOR	302	1.41%	1,020	0.97%	442	4.30%	81	0.35%
M	M	LEON	45	0.21%	0	0.00%	287	2.79%	34	0.15%
M	M	ANDALUZ	2	0.01%	0	0.00%	1	0.01%	0	0.00%
IM	M	POLO	7,330	34.27%	36,391	34.78%	7	0.07%	25	0.11%
IM	M	SOL	1,251	5.85%	6,334	6.05%	0	0.00%	0	0.00%
IM	M	PEZ	953	4.46%	2,092	2.00%	0	0.00%	0	0.00%
IM	M	BAMBU	70	0.33%	0	0.00%	0	0.00%	0	0.00%
IM	IV	VIRUS	477	2.23%	1,007	0.96%	0	0.00%	0	0.00%
IM	IS	ESTRES	106	0.50%	671	0.64%	6	0.06%	2	0.01%
IM	IP	ALICATES	17	0.08%	0	0.00%	1	0.01%	0	0.00%
IF	M	LUNA	6,261	29.28%	18,660	17.83%	47	0.46%	0	0.00%
IF	M	MALDAD	826	3.86%	8,789	8.40%	34	0.33%	3	0.01%
IF	M	ILUSION	1,937	9.06%	27,840	26.61%	0	0.00%	0	0.00%
IF	M	GACHI	3	0.01%	0	0.00%	0	0.00%	0	0.00%
IF	IV	CRISIS	153	0.72%	108	0.10%	0	0.00%	0	0.00%
IF	IS	ENOLOGIA	300	1.40%	375	0.36%	1	0.01%	0	0.00%
IF	IP	TJERAS	45	0.21%	0	0.00%	0	0.00%	0	0.00%
IV	M	ARTISTA	636	2.97%	294	0.28%	2,164	21.04%	4,546	19.57%
IV	M	MARTIR	11	0.05%	1	0.00%	878	8.54%	6,610	28.45%
IV	M	REHEN	6	0.03%	205	0.20%	45	0.44%	198	0.85%
IV	M	SAUDI	2	0.01%	2	0.00%	33	0.32%	0	0.00%
IV	IV	RUBIALES	66	0.31%	0	0.00%	34	0.33%	0	0.00%
TOTAL			21,386		104,638		10,284		23,233	

GT: Gender type. NT: Number type. M: Morphological.
 IM: Inherent masculine. IF: Inherent feminine.
 IS: Inherent singular. IP: Inherent plural. IV: Invariant.

raic-es (*root*). In addition to this, a few nominals can accept the two plural suffixes: *bambú-s* and *bambú-es* (*bamboo*).

These facts mean that several paradigm models are needed to encode all the actual combinations of these phenomena. In Table IV we show the paradigm names used in our DATR implementation with their gender and number types, as well as the count and the percentages of occurrences in our lexicon and the CRATER corpus.

Tables V and VI summarize the distribution of the gender and number types, respectively. The figures are significant: Spanish marks lexically gender in nouns overwhelmingly, both in competence and performance. In adjectives, however, gender is marked with a suffix in two thirds of the lexicon; the other third is gender invariant. With respect to performance, half of the adjective occurrences are gender

Table V. Distribution of nominal entries by gender type in lexicon (types) and corpus (tokens).

GENDER TYPE	NOUNS		ADJECTIVES	
	LEXICON	CORPUS	LEXICON	CORPUS
MORPHOLOGICAL	4.38%	1.79%	68.40%	51.00%
INHERENT	92.25%	97.73%	0.93%	0.13%
INVARIANT	3.37%	0.48%	30.67%	48.87%

Table VI. Distribution of nominal entries by plural type in lexicon (types) and corpus (tokens).

NUMBER TYPE	NOUNS		ADJECTIVES	
	LEXICON	CORPUS	LEXICON	CORPUS
-s	69.62%	53.71%	83.19%	70.18%
-es	25.29%	44.23%	16.73%	29.81%
IV	3.25%	1.07%	0.33%	0.00%
IH	2.19%	1.00%	0.08%	0.01%

IV: Invariant. No plural morpheme.

IH: Inherent. *Singularia and pluralia tantum*.

invariant and the other half are marked by a suffix. Inherent gender adjectives are marginal.

However, the plural suffix is generalized in both syntactic categories. The most frequent allomorph suffix is /-s/. Inherent and invariant number are very rare in the corpus for both categories.

Therefore, the gender is mainly lexically conditioned in Spanish, especially in nouns. This implies that every morphological description needs to encode much of the gender information directly in the lexical entry. On the other hand, morphological gender is also important, mainly in adjectives. The stem allomorphy is phonologically conditioned but the suffix allomorphy /-o/, /-e/ in masculine is morphologically conditioned. As a consequence a model based on phonological rules (IP), such a two-level description, does not take advantage of its full descriptive power, since these entries have to be marked.

The nominal number allomorphs are always phonologically conditioned when number is morphologically marked, which happens in more than a 95% of the lemmas. In this phenomenon, phonological rule-based models have clear advantages over paradigmatic models.

Our DATR implementation of the paradigm models shown above makes use of three different hierarchies, a fragment of which can be viewed in Figure 5:

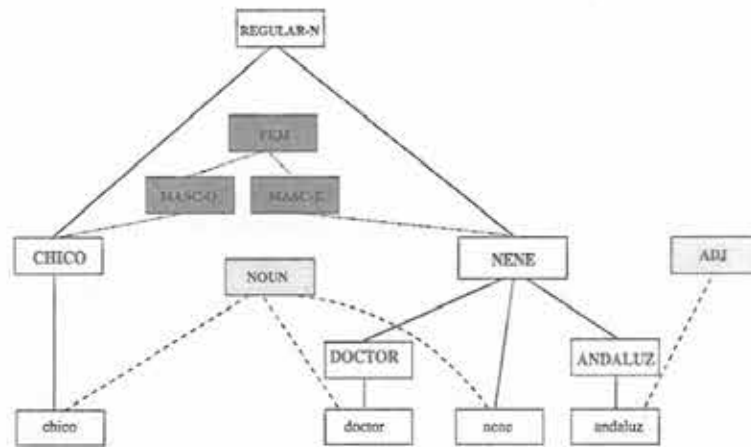


Figure 5. A fragment of nominal hierarchy.

- Paradigmatic models: Paradigms that reflect the different distributional behaviour, with respect to stem allomorphy, of lemmas summarized in Table IV. We have defined a default paradigm, REGULAR-N, for the whole nominal class (noun, adjective)* followed by a network of paradigm inheritance.
- Morphological suffix: Paradigmatic models for lemmas which have gender and/or plural suffixes inherit from these the appropriate suffix allomorph. Our implementation defines that the two masculine DATR models (for allomorphs /-e/ and /-o/) inherits information for the unique feminine DATR model, since both masculines share the same information for the feminine.
- Part-of-speech: Each lemma inherits syntactic information from the relevant category paradigm. In our implementation, only adjectives and nouns do.

Some nominal paradigms in Table IV, which are defined following descriptive criteria, are actually grouped together taking advantage of DATR features that permit us to capture parsimony, in a particular path extension. For example, DOCTOR and LEON are grouped in the model DOCTOR-LEON that assumes that two allomorphs are needed for its lemmas: Since the LEON paradigm needs two allomorph stems, the paths <stem> and <stem 2> need to be supplied in the entry, whereas the DOCTOR paradigm only needs to define <stem>, and <stem 2> is inherited by the path extension mechanism (both paths are needed in the DOCTOR-LEON DATR model). *Singularia and pluralia tantum* models (ESTRES, ENOLOGIA, ALICATES and TIJERAS) are not defined in our DATR implementation: each lemma is lexicalized. For instance, this lemma (*enology*) only can appear in masculine singular form:

* Determiners, pronouns and some quantifiers take gender and number suffixes, but since they are closed-class categories, i.e., non-productive, we prefer to treat each of them as unique paradigms, in the same way as we did with the strong irregular verbs.

Enologia:

<wf masc sing> == 'enología'.

6. Conclusions

In the case of Spanish, inflection is both rule-governed and paradigmatic. For paradigmatic inflection (e.g., most part of the verb morphology), an inheritance model such as DATR is best-suitable, while for phonological and orthographical allomorphy (e.g., typically the nominal morphology), a finite-state model seems to be more appropriate.

We have shown that Spanish nominal inflection is mostly regular, agglutinative and phonologically conditioned, while verb inflection presents a varied catalogue of lexical exceptions, the realization of the morphemes is basically syncretic, and the morphologically conditioned allomorphy is important.

How can it be determined which one is more representative of the Spanish morphology? From a pure quantitative approach, it seems clear that both in competence and performance nominals exceed verbs (see total verbal and nominal entries in the lexicon and corpus from Tables III and IV). However, from a pure morphological perspective, verb inflection is significantly much richer: just compare 55 verb word forms against the four nominal word forms.

On the other hand, some qualitative arguments could support the hypothesis that verbal morphology is more representative of the complexities of the Spanish inflection. For instance, Spanish native speakers clearly show more variation and insecurity in their well-formation judgments of a particular word form if it is a verb. The same is true for the learners of Spanish as a second language. Similarly, grammatical descriptions dedicate more space to verb inflection than to the nominal one. In general, speakers and grammarians feel that verb inflection is more important and crucial.

If this hypothesis is accepted, then the decision as to which model is best-suited seems to be clear: the one that can express the complexities of the verb inflection more elegantly. Assuming that a rule-governed, finite-state model is able to deal with the problem but at the high cost of introducing many exceptional rules and many lexicalized entries, we have tried to show that a paradigmatic approach based on inheritance can explain the apparent subregularities and exceptions within a single formalism and with less effort.

The drawback of the paradigmatic approach, if it is taken as the only model, is that it makes use of an excessive formal power for the nominal inflection. Comparing both phenomena and approaches, one could claim that it is more realistic to split the inflection treatment into different submodels. However, neither in theoretical works nor in computational implementations has this splitting been adopted. Future work should address how to combine different approaches in the description of varied and complex morphological phenomena in natural languages.

As can be seen, our DATR implementation needs each lemma entry node to have every stem allomorph. This is so since no morphophonological aspects are reflected in our model. Nevertheless, some morphophonological aspects from IP models can be incorporated into our description: the allomorph stems can be computed off-line by means of regular-expression rules. The details have been shown in the authors' previous work (Goñi and González, 1995; Goñi et al., 1995, 1997; Goñi Menoyo, 1998).

As currently defined, DATR is biased to evaluate only one-direction queries (in the example shown in this paper, word generation can be achieved, but not analysis), although some proposals have been reported for evaluating reverse queries (Langer, 1994). In addition to this, the programming interface with other components of a processing system has to be specially designed* (Kilbury et al., 1991). Current implementations lack indexing mechanisms for storage and fast retrieval of huge lexicons, and rely heavily on the usually inefficient host language** (Prolog).

We hope to have shown that the DATR formalism is able to capture the significant generalizations on the inflectional (sub)regularities of Spanish, providing a rather compact and elegant description. In spite of some of its good points, some criticisms can be made about DATR. Although Evans and Gazdar (1996) claim that DATR should be taken as a kind of *programming language*, not a theoretical framework for the lexicon, we seriously doubt that it can be successfully used for real-scale language engineering applications, although it has proven to be an excellent prototyping and descriptive tool.

Appendix A: DATR Description of Spanish

A.1. THE DATR IMPLEMENTATION

A.1.1. *Verb Query Interface*

Inflecting forms of the lexical verb nodes can be obtained by means of a DATR query such as:

Lemma:<wf \$mood \$tense \$pers \$num>

where

- \$mood is one of: ind, subj, cond, imper, inf, ger, part.
- \$tense is one of: pres, impf, indf, fut, (when mood is ind); pres, impf (when mood is subj); and it is not needed for other mood values.

* Although its name seems to sound similar to PATR, the similarities stop there. Unification, that is essential to PATR-II plays no role in DATR. On the contrary, inheritance is essential here, while marginal in PATR-II.

** A direct Prolog-encoding of a flat version of our ARIES lexicon and morphological rules, both for analysis and generation, as reported in Moreno and Goñi (1995), also had such efficiency drawbacks.

- \$pers is one of: 1, 2, 3, pol (for polite forms).
- \$num is one of: sing, plu.

Some inflecting verb forms, that are alternative forms of others, can be obtained by means of a query in the form of:

Lemma:<wf subj impf 1 sing alt>
 Valer:<wf imper 2 sing alt>

A.1.2. *Nominal Query Interface*

Inflecting forms of the lexical nominal nodes can be obtained by means of the DATR query:

Lemma:<wf \$gend \$num>

where

- \$gend is one of: masc, fem.
- \$num is one of: sing, plu.

Alternative forms for the plural of some lemmas are accessible by means of the query:

Lemma:<wf masc plu alt>
 Lemma:<wf fem plu alt>

A.1.3. *Verb Morphology Implementation*

The default morphological rule that builds a word form from a stem and a suffix is included in the definition of REGULAR-V model:

```
REGULAR-V:
  <>                == VERB
  <sfx>              == "<des reg>"
  <root>             == "<stem 1>"
  <wf>              == "<root>" "<sfx>"
  <wf imper pol sing> == "<wf subj pres 3 sing>"
  <wf imper pol plu> == "<wf subj pres 3 plu>" .
```

The <wf> (word form) value is built from the concatenation of <root> (root) and <sfx> (suffix) values, which are inherited from the value of <stem 1> and <des reg> paths from the node where the DATR query started (by means of the global reference mechanism invoked with the double quotes). The path extension mechanism allows word formation for the different inflecting forms with just one concatenating rule.

root values are obtained by default from first stem. If a particular word form (reflected as a particular extension of the root path) needs a different stem allomorph, the paradigmatic model has to state it explicitly:

```
ROMPER:
  <>          == REGULAR-V
  <sfx part>   == PART1
  <root part> == "<stem 2>".
```

stem values are encoded in the leaf nodes (lexical entries). Each lexical node has to have all its allomorph stems needed to build all the possible word forms:

```
Romper:
  <>      == ROMPER
  <des>   == C2
  <stem>  == romp
  <stem 2> == rot.
```

des values are obtained from the conjugation hierarchy (cf. Figure 4) for each lemma verb. Regular suffixes are obtained from extensions of the path <des reg>, whereas irregular endings* are obtained from one of:

- <des pres>, <des pret1>, <des pret2>.
- <des fut_cond>, <des imp_subj>, <des imper>.
- <des infin>, <des ger>, <des part1>, <des part2>.

that reflects the fact that an irregular allomorph suffix is used for particular word forms. Such a fact is captured in the allomorph suffix hierarchy (cf. Figure 4). In the ROMPER example above, it is shown how it uses an irregular allomorph suffix from PART1 node, which gets it from conjugation hierarchy as follows:

```
PART1:
  <sfx part>          == "<des part1 part>".
```

A.1.4. *Nominal Morphology Implementation*

Similarly as in the case of verbs, the default morphological rule that builds a word form from a stem and a suffix is included in the definition of REGULAR-N model:

```
REGULAR-N:
  <root>          == "<stem 1>"
  <wf>           == "<root>" "<sfx>".
```

* Irregular allomorph suffixes for indicative present, preterite indicative (2 suffixes), future conditional, imperfect subjunctive, imperative, infinitive, gerund and past participle (2 suffixes).

The <wf> value is built from the concatenation of <root> (root) and <sfx> values. The former is inherited from the value of <stem 1> paths from the node where the DATR query started (by means of the global reference mechanism invoked with the double quotes). The path extension mechanism allows word formation for the different inflecting forms with just one concatenating rule.

sfx values are inherited from the morphological suffixes hierarchy. For instance, the NENE model inherits suffixes from MASC-E, which inherits feminine suffixes from FEM:

```
NENE:
  <>           == REGULAR-N
  <sfx>        == MASC-E.
```

```
MASC-E:
  <sfx>        == FEM
  <sfx masc sing> == e
  <sfx masc plu> == es.
```

As can be seen in the nominals section, there are some cases in which lexical nodes directly encode word forms (*singularia* and *pluralia tantum*), or several allomorph stems are required (e.g., PEZ).

Spanish diacritical mark encoding is done by means of ISO-8859-1 (latin1) encoding (e.g., "ñ," "ú," "ü"). In DATR nodes such characters are not used, since DATR implementations usually do not allow such codes.

Our implementation detects non-existent forms, making such paths non-evaluable. For instance, some forms for defective verbs or feminine forms for inherent masculine nominals cannot be generated.

For space reasons, the actual DATR implementation is not reproduced here. It can be found on the following URLs instead:

```
http://www.mat.upm.es/aries/aries-verbs.dtr
http://www.mat.upm.es/aries/aries-nominals.dtr
```

References

- Andry, F., Fraser, N.M., McGlashan, S., Thornton, S., and Youd, N.J., 1992, "Making DATR work for speech: Lexicon compilation in SUNDIAL," *Computational Linguistics* 18, 245-267.
- Bauer, L., 1988, *Introducing Linguistics Morphology*, Edinburgh: Edinburg University Press.
- Beard, R., 1995, *Lexeme-Morpheme Based Morphology*, New York: State University of New York Press.
- Bello, A., 1898, *Gramática de la lengua castellana*, Madrid: Edaf (reimp. 1984).
- Briscoe, T., Paiva, V., and Copestake, A., eds., 1993, *Inheritance, Defaults, and the Lexicon*, Studies in Natural Language Processing, Cambridge: Cambridge University Press.
- Brown, D., Corbett, G., Fraser, N., Hippiusley, A., and Timberlake, A., 1996, "Russian noun stress and network morphology," *Linguistics* 34, 53-107.

- Bubenik, V., 1999, *An Introduction to the Study of Morphology*, Lincom Europa.
- Cahill, L.J., 1993a, "Morphonology in the lexicon," pp. 87–96 in *Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics (EACL'93)*.
- Cahill, L.J., 1993b, "Some reflections on the conversion of the TIC lexicon into DATR," pp. 47–57 in *Inheritance, Defaults, and the Lexicon*, Briscoe, T., Paiva, V., and Copestake, A., eds., Studies in Natural Language Processing, Cambridge: Cambridge University Press.
- Cahill, L.J. and Evans, R., 1990, "An application of DATR: The TIC lexicon," pp. 120–125 in *Proceedings of the 9th European Conference on Artificial Intelligence (ECAI'90)*. Also in Evans and Gazdar (1990), pp. 31–40.
- Cahill, L. and Gazdar, G., 1997, "The inflectional phonology of German adjectives, determiners and pronouns," *Linguistics* 35, 211–245.
- Carstairs, A.D., 1987, *Allomorphy in Inflection*, London: Croom Helm.
- Carulla, M. and Oosterhoff, A., 1996, "El tratamiento de la morfología flexiva del castellano mediante reglas de dos niveles," pp. 72–80 in *Actas del XII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN'96)*, Sevilla.
- Corbett, G.G. and Fraser, N.M., 1993, "Network morphology: A DATR account of Russian inflectional morphology," *Journal of Linguistics* 29, 113–142.
- Daelemans, W., De Smedt, K., and Gazdar, G., 1992, "Inheritance in natural language processing," *Computational Linguistics* 18, 205–218.
- Erbach, G., 1994, "Multi-dimensional inheritance," pp. 102–111 in *Proceedings of KONVENS'94*, Vienna, H. Trost, ed. Also on <http://xxx.lanl.gov/ps/cmp-ig/9411025>
- Evans, R., 1990, "An introduction to the Sussex Prolog DATR system," pp. 63–71 in *The DATR Papers*, R. Evans and G. Gazdar, Brighton: University of Sussex.
- Evans, R., 1992, "Derivational morphology in DATR," pp. 55–69 in *Sussex Papers in General and Computational Linguistics*, L.J. Cahill and R. Coates, eds., Brighton: University of Sussex.
- Evans, R. and Gazdar, G., 1989a, "Inference in DATR," pp. 66–71 in *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics (EACL'89)*. Also pp. 15–20 in *The DATR Papers*, R. Evans and G. Gazdar, Brighton: University of Sussex.
- Evans, R. and Gazdar, G., 1989b, "The semantics of DATR," pp. 79–87 in *Proceedings of the Seventh Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*, A.G. Cohn, ed. Also pp. 21–30 in *The DATR Papers*, R. Evans and G. Gazdar, Brighton: University of Sussex.
- Evans, R. and Gazdar, G., 1990, *The DATR Papers*, CSRP 139, Brighton: University of Sussex.
- Evans, R. and Gazdar, G., 1996, "DATR: A language for lexical knowledge representation," *Computational Linguistics* 22, 167–216.
- Evans, R., Gazdar, G., and Moser, L., 1993, "Prioritised multiple inheritance in DATR," pp. 38–46 in *Inheritance, Defaults, and the Lexicon*, Briscoe, T., Paiva, V., and Copestake, A., eds., Studies in Natural Language Processing, Cambridge: Cambridge University Press.
- Evans, R., Gazdar, G., and Weir, D., 1994, "Using default inheritance to describe LTAG," in *Proceedings of 3e Colloque International sur les Grammaires d'Arbres Adjoints (TAG'3)*. Also on <http://xxx.lanl.gov/ps/cmp-ig/9501001>
- Fraser, N.M. and Corbett, G.G., 1995, "Gender, animacy, and declensional class assignment: A unified account for Russian," pp. 123–150 in *Yearbook of Morphology 1994*, G. Booij and J. van Marle, eds., Dordrecht: Kluwer Academic Publishers.
- Gazdar, G., 1990, "An introduction to DATR," pp. 1–14 in *The DATR Papers*, R. Evans and G. Gazdar, Brighton: University of Sussex.
- Gazdar, G., 1992, "Paradigm function morphology in DATR," pp. 43–53 in *Sussex Papers in General and Computational Linguistics*, L.J. Cahill and R. Coates, eds., Brighton: University of Sussex.
- Gibbon, D., 1992, "ILEX: A linguistic approach to computational lexica," pp. 32–53 in *Computatio Linguae: Aufsätze zur algorithmischen und quantitativen Analyse der Sprache (Zeitschrift für Dialektologie und Linguistik, Beiheft 73)*, U. Klenk, ed., Stuttgart: Franz Steiner Verlag.

- Gómez Guinovart, J. and Aguirre Moreno, J.L., 1998, "Alternancias morfológicas y paradigmas irregulares en la morfología flexiva verbal del gallego," pp. 177–184 in *Actas del XIV Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN'98)*, Alicante, Spain.
- Goñi, J.M. and González, J.C., 1995, "A framework for lexical representation," pp. 243–252 in *Proceedings of AI'95: Fifteenth International Conference. Language Engineering'95*, Montpellier, France. Also on <http://xxx.lanl.gov/ps/cmp-ig/9507002>
- Goñi, J.M., González, J.C., and Moreno, A., 1995, "A lexical platform for Spanish," pp. 61–65 in *Proceedings of the Computational Lexicon (ESSLLI'95 Workshop)*, Barcelona, Spain, M.F. Verdejo, ed.
- Goñi, J.M., González, J.C., and Moreno, A., 1997, "ARIES: A lexical platform for engineering Spanish processing tools," *Natural Language Engineering* 3, 317–345.
- Goñi Menoyo, J.M., 1998, "Arquitectura para representación del conocimiento léxico en sistemas de procesamiento de lenguaje natural," Ph.D. Thesis, Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid.
- Jenkins, L., 1990a, "Enhancements to the Sussex Prolog DATR implementation," pp. 41–61 in *The DATR Papers*, R. Evans and G. Gazdar, Brighton: University of Sussex.
- Jenkins, L., 1990b, "Japanese verbs in DATR," pp. 73–78 in *The DATR Papers*, R. Evans and G. Gazdar, Brighton: University of Sussex.
- Katamba, F., 1993, *Morphology*, New York: St. Martin's Press.
- Keller, B., 1995, "DATR theories and DATR models," pp. 55–62 in *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL'95)*.
- Kilbury, J., 1992, "Paradigm-based derivational morphology," in *Proceedings of KONVENS'92*, G. Görz, ed.
- Kilbury, J., Naerger, P., and Renz, I., 1991, "DATR as a lexical component for PATR," pp. 137–142 in *Proceedings of the 5th Conference of the European Chapter of the Association for Computational Linguistics (EACL'91)*.
- Kilgarriff, A., 1993, "Inheriting verb alternations," pp. 213–221 in *Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics (EACL'93)*.
- Kilgarriff, A. and Gazdar, G., 1995, "Polysemous relations," pp. 1–25 in *Grammar and Meaning: Essays in honour of Sir John Lyons*, F.R. Palmer, ed., Cambridge: Cambridge University Press.
- Koskenniemi, K., 1983, "Two-level morphology: A general computational model for word-form recognition and production," Ph.D. Thesis, Department of Linguistics, University of Helsinki.
- Langer, H., 1994, "Reverse queries in DATR," pp. 1089–1095 in *Proceedings of the Fifteenth International Conference on Computational Linguistics (COLING'94)*. Also on <http://xxx.lanl.gov/ps/cmp-ig/9411024>
- Matthews, P.H., 1972, *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*, Cambridge: Cambridge University Press.
- McEnery, T., Wilson, A., Sánchez, F., and Nieto, A.F., 1997, "Multilingual resources for European languages: Contributions to the CRATER project," *Literary and Linguistic Computing* 12, 119–122.
- Moreno, A. and Goñi, J.M., 1995, "GRAMPAL: A morphological model and processor for Spanish implemented in Prolog," pp. 321–331 in *Proceedings of the Joint Conference on Declarative Programming (GULP-PRODE'95)*, Marina di Vietri, Italy, M. Sessa and M. Alpuente, eds. Also on <http://xxx.lanl.gov/ps/cmp-ig/9507004>
- Moreno Sandoval, A., 1991, "Un modelo computacional basado en la unificación para el análisis y la generación de la morfología del español," Ph.D. Thesis, Universidad Autónoma de Madrid, Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia.
- Moser, L., 1992a, *DATR Paths as Arguments*, CSRP 215, Brighton: University of Sussex.
- Moser, L., 1992b, *Evaluation in DATR is co-NP-Hard*, CSRP 240, Brighton: University of Sussex.
- Moser, L., 1992c, *Lexical Constraints in DATR*, CSRP 216, Brighton: University of Sussex.

- Moser, L., 1992d, *Simulating Turing Machines in DATR*, CSRP 241, Brighton: University of Sussex.
- Pirrelli, V. and Battista, M., 1996, "Monotonic paradigmatic schemata in Italian verb inflection," pp. 77-82 in *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96)*.
- Poch, A., 1992, "Representación del conocimiento léxico: Un análisis con DATR," Ph.D. Thesis, Universidad de Barcelona.
- Seco, M., 1986, *Diccionario de dudas y dificultades de la lengua española*, Madrid: Espasa Calpe.
- Spencer, A., 1991, *Morphological Theory*, Oxford: Blackwell Publishing.
- Sproat, R., 1992, *Morphology and Computation*, Cambridge, MA: MIT Press.
- Tzoukermann, E. and Liberman, M.Y., 1990, "A finite-state morphological processor for Spanish," pp. 277-281 in *Proceedings of the 13th International Conference on Computational Linguistics (COLING'90)*.
- Wurzel, W.U., 1989, *Inflectional Morphology and Naturalness*, Dordrecht: D. Reidel.