

DEALING WITH VERBAL INFLECTION IN NATURAL LANGUAGE SELECTION: RESOURCES FOR MINIMAL REQUIREMENT

JUAN RAFAEL ZAMORANO-MANSILLA*

Universität Bremen / Universidad Complutense de Madrid

ABSTRACT. *This article presents an analysis of the inflection of the Spanish verb produced during the development of a generation grammar of Spanish for the generator KPML (<http://www.fb10.uni-bremen.de/anglistik/langpro/kpml>). The claim made in the article is that the analysis is maximally efficient from the point of view of Natural Language Generation, on the grounds that it allows generating the 53 inflected forms associated to any Spanish verb requiring a minimum of information in the lexicon component. The defining characteristic of the analysis is its modularity: instead of establishing models of inflection based on the whole conjugation of a verb, various restricted models were created. These restricted models deal with individual tenses, and they describe the variation patterns to be found in the stem and the endings of verbs within each tense.*

KEYWORDS: *Natural Language Generation, verbal inflection, KPML, Spanish generation grammar.*

RESUMEN. *Este artículo presenta un análisis de la flexión verbal del verbo español creado durante el desarrollo de una gramática de generación de español para el generador automático KPML (<http://www.fb10.uni-bremen.de/anglistik/langpro/kpml>). El análisis aquí expuesto tiene la ventaja de ofrecer máxima eficiencia para su aplicación a la generación automática, debido a que permite generar las 53 formas flexivas asociadas a cualquier verbo español haciendo uso de un mínimo de información en el componente léxico. La característica fundamental del análisis es su modularidad: en vez de definir modelos de inflexión basados en la conjugación completa de un verbo, se han definidos varios modelos restringidos. Dichos modelos restringidos recogen la inflexión de tiempos individuales, y se basan en las pautas de variación presentes en la raíz y las terminaciones de los verbos en cada tiempo.*

PALABRAS CLAVE: *Palabras clave: Generación automática, flexión verbal, KPML, gramática de generación de español.*

1. INTRODUCTION

This article describes the analysis of the inflectional morphology of the Spanish verb produced during the creation of a generation grammar of Spanish for the automatic

generator KPML (Bateman 1997, Dale and Reiter 2000, Matthiessen and Bateman 1991, Mitkov 2003, <http://www.fb10.uni-bremen.de/anglistik/langpro/kpml>). It is an improvement on a previous analysis applied in the early stages of development of the grammar (Zamorano-Mansilla 2002). The analysis presented here not only overcomes some of the inadequacies of the earlier version, but it has also been extended to cover the Subjunctive forms, the Imperative, the Infinitive, the Gerund and the Participle.

The analysis was developed with the aim of providing resources for Natural Language Generation (NLG) that met two basic requirements:

- a) they should be comprehensive, that is, they should cover all the facts connected with the inflectional morphology of the Spanish verb;
- b) they should do so in a maximally efficient way.

The question of what is understood in this article by 'efficient' resources for NLG requires further clarification. NLG resources normally consist of two basic elements: a set of rules and a lexicon. Most linguistic facts may be covered by resources that put more emphasis on one or the other element. For example, in order to cover the plural form of nouns in Spanish, two different strategies can be adopted. One is to include the plural form of every noun in the lexicon. Alternatively, we can design a set of rules that specify which plural form is the right one for a given noun. In principle, this second solution can be said to be more efficient for NLG because the rules can be used with any lexical item, even if it is not known to the generator. By contrast, the introduction of the plural form of every noun of the lexicon is a time-consuming task that gives a huge lexicon file as a result. Generally speaking, rules are more efficient when dealing with the morphology of an open word class, such as nouns, adjectives and verbs. This is even more true of the inflection of the Spanish verb: an average item would require the inclusion of about 53 additional forms in the lexicon to cover the whole inflection. Ideally efficient resources would then consist only of a set of rules that would perform the necessary changes on the entry form of a verb in order to generate all the inflection. Unfortunately, the inflection of the Spanish verb presents so many irregularities that the ideal rules-only approach cannot be adopted. Some information must necessarily be included in the lexicon file.

The claim made in this article is that the analysis of the inflection of the Spanish verb presented here is maximally efficient from the point of view of NLG, on the grounds that the information required in the lexicon file is kept to the minimum that is strictly necessary to allow full coverage of the phenomenon.

2. SOME FACTS ON THE INFLECTION VERBS IN SPANISH

Almost every Spanish verbs has nine simple tenses and three non-personal forms. Compound forms such as the Perfect or the Progressive are ignored here, as they are the result of combining a simple tense and a non-personal form. These tenses have received

different names in the grammatical tradition of Spanish, but perhaps the most widespread labels are the ones shown in table 1.

TENSE	EXAMPLE (verb <i>cantar</i> [sing])	SHORT NAME
Presente de Indicativo (Present Indicative)	<i>canto</i>	PrI
Pretérito Imperfecto de Indicativo (Imperfective Preterite Indicative)	<i>cantaba</i>	P2I
Pretérito Indefinido (Perfective Preterite)	<i>canté</i>	P1
Futuro de Indicativo (Future Indicative)	<i>cantaré</i>	FI
Condicional (Conditional)	<i>cantaría</i>	C
Presente de Subjuntivo (Present Subjunctive)	<i>cante</i>	PrS
Pretérito Imperfecto de Subjuntivo (Imperfective Preterite Subjunctive)	<i>cantara/cantase</i>	P2S
Futuro de Subjuntivo (Future Subjunctive)	<i>cantare</i>	FS
Imperativo (Imperative)	<i>cantad</i>	Im
Infinitivo (Infinitive)	<i>cantar</i>	I
Gerundio (Gerund)	<i>cantando</i>	G
Participio (Past Participle)	<i>cantado</i>	P

Table 1. *The nine simple tenses of Spanish*

For the sake of brevity, I will use the short name assigned to each tense in table 1 in the rest of the article, rather than the full name in Spanish or the translation into English.

Each tense has eight different forms. They are the result of combining three persons (first, second, third), two numbers (singular, plural) and two different forms for the second person, singular or plural, to express meanings along the scales of power/solidarity. These forms reflect the contrast between the pronouns *tú/usted* in the singular and *vosotros/ustedes* in the plural. Such a distinction is also present in languages like French,

German or Russian. Nevertheless, the forms that emphasize power or inequality in Spanish are identical to the third person form (singular or plural), and consequently they are not mentioned in what follows on the assumption that they can be generated applying the same methods that are valid for the generation of the third person forms. In this article the short names shown in table 2 will be employed to refer to each form of a tense.

FORM	SHORT NAME
First person singular	1s
Second person singular (solidarity)	2s
Second person singular (power)	—
Third person singular	3s
First person plural	1p
Second person plural (solidarity)	2p
Second person plural (power)	—
Third person plural	3p

Table 2. *The eight forms of Spanish tenses*

The Imperative and the Preterite Subjunctive are exceptional for different reasons. The Imperative only contains two forms, one for the second person singular and another for the second person plural. The rest of forms are realized by the Present Subjunctive. The Preterite Subjunctive, on the other hand, contains sixteen forms, since each person and number can be realized by two different forms with no functional contrast in Modern Spanish.

Generally speaking, verbs are grouped in Spanish into three so-called conjugations (the first, the second and the third conjugation). They represent models of inflection that specify the selection of endings in each tense. It is possible to say to which of the three conjugations a verb belongs to by looking at the ending in the Infinitive. Thus, the ending *-ar(se)* is for the first conjugation, the ending *-er(se)* is for the second, and the endings *-ir(se)* or *-ír(se)* are typical of the third conjugation. There are no exceptions to this rule.

Some variations found in the stem of Spanish verbs have a purely orthographic origin: they appear when the rules of Spanish force a change in the spelling of a stem to represent the same phoneme in different phonological contexts (for instance, *cazar* - *cacé*, both representing the phoneme /θ/). These changes are completely regular and predictable and so can be easily handled with an algorithm in the grammar that implements the necessary modifications based on the phonological context. For this reason, these changes in the spelling will not be mentioned in the rest of the article. Table 3 shows the phonemes subject to orthographic variation in Spanish and the context associated to each spelling.

PHONEME IN THE STEM	SPELLING	CONDITION	EXAMPLE
/θ/	<i>c</i> <i>z</i>	When followed by /e/ or /i/ When followed by /a/ or /o/ or in final position	<i>vencer / hacer</i> <i>venza / haz</i>
/x/ʎ	<i>g</i> <i>j</i>	When followed by /e/ or /i/ When followed by /a/ or /o/	<i>coger</i> <i>cojo</i>
/k/	<i>qu</i> <i>c</i>	When followed by /e/ or /i/ When followed by /a/ or /o/	<i>delinquir</i> <i>delinco</i>
/g/	<i>gu</i> <i>g</i>	When followed by /e/ or /i/ When followed by /a/ or /o/	<i>seguir</i> <i>sigamos</i>
/gw/	<i>gü</i> <i>gu</i>	When followed by /e/ or /i/ When followed by /a/ or /o/	<i>averigüé</i> <i>averiguo</i>

Table 3. Conditions for orthographic variations in the stem of Spanish verbs

3. THE TREATMENT OF THE SPANISH VERB IN PREVIOUS GENERATION SYSTEMS

Inflectional morphology has often been neglected in the design of linguistic resources for automatic generation. One reason for this is that the vast majority of generation systems have been created for the English language (Adorni and Zock 1996, Bateman and Zock 2008), which presents a relatively simple inflection for all word classes: the whole conjugation of the most irregular English verb (*be*) can be covered with just eight forms. Consequently, generation systems have often opted for the specification of each inflectional form in the lexicon file, either as independent lexical items with different properties or as variants of the infinitive form. Only completely regular and predictable forms have been generated resorting to rules rather than the lexicon file. This is what we find in the English generation for KPML and other influential systems such as MUG (Reitter 2004.), FUF (Elhadad and Robin 1992) or HALogen (Langkilde-Geary 2002). In the very few cases in which linguistic resources for Spanish have been created for these systems, the morphology of the verb has been treated in a similar way (Aguado *et al.* 1998).

Clearly this approach is not very convenient for the Spanish language. To start with, we have seen that Spanish verbs require many more forms in the lexicon file than English verbs, which renders the task of introducing a verb in the lexicon component lengthy and arduous. To make things worse, the list of Spanish verbs that contain some type of irregularity in the conjugation is considerable; much bigger than the list of irregular verbs in English or German, for instance.

Another problem concerning the treatment of the inflectional morphology of the Spanish verb in generation systems derives from the growing tendency in the area of computational linguistics to diminish the contribution from linguistics. Most computational

linguists concentrate on efficiency in terms of programming, while very often they are surprisingly happy to use obsolete or inadequate linguistic concepts or even do without them altogether, as is the case of statistical methods. As a result of this, an analysis of the conjugation of the Spanish verb specifically designed for minimum lexical requirements has not been felt to be essential. Given the capacity of modern computers, introducing all the inflectional forms of each verb or the numerous paradigms one finds in traditional grammars of Spanish has become a common practice, and this is what we find in the many automatic conjugators of Spanish available on the web. This procedure, however, was dismissed during the creation of a generation grammar of Spanish for KPML due to its lexical demands. Instead, it was deemed important to design resources that would require the minimum of information in the lexicon file in order to facilitate the addition of lexical elements.

4. AN ANALYSIS OF THE INFLECTION OF THE SPANISH VERB FOR THE PURPOSES OF NATURAL LANGUAGE GENERATION

4.1. *General principles behind the analysis*

In order to reduce the amount of information required by the lexicon file and the complexity of the rules that deal with the inflection of the verb, the analysis described here takes a modular approach. This means that the different phenomena of the inflection of the Spanish verb have been treated separately, on the assumption that, unless otherwise stated, every verb is regular. Thus verbs are not assigned to a single model of inflection, as is the case of existing grammatical descriptions. Instead, they are assigned to various models that are restricted to very specific parts of the conjugation. At this point it is important to emphasize that there is nothing inherently 'better' about this modular approach. The claim made in this article is that it simply produces more efficient resources for NLG.

The various restricted models of inflection were established according to three basic principles:

1. Every verbal form can be analysed into two basic components: a stem and an ending (*amas*: $am_{[\text{stem}]}as_{[\text{ending}]}$). The stem is not obtained in this analysis using etymological information, but after isolating the part of the verb that remains unaltered throughout the conjugation of a given tense (e.g., *amo*, *amas*, *ama*, *amamos*, *amáis*, *aman* ⇨ stem = *am*). As a consequence, the stem of a verb need not be the same for every tense. This is an important difference with respect to existing analyses of the conjugation of Spanish verbs, which favour etymology and try to define a stem that remains stable through the whole conjugation.
2. The models of inflection described here are restricted to individual tenses. Contrary to what we find in existing grammatical descriptions (see among others, Bello 1874, Gili Gaya 1961, RAE 1973, Alcina & Blecua 1975, Alarcos Llorach

2000), this analysis does not contain full paradigms of inflection that include all the tenses of the Spanish verb. The advantages of defining models of inflection restricted to individual tenses when dealing with linguistic resources for NLG are clear: since even the most irregular of verbs often have some regular tense, it is possible to reduce the amount of information we must include in the lexicon file if the program only requires additional information for the irregular tenses.

3. The models of inflection for each individual tense of the conjugation are divided into two types: those that deal with the variation we can identify in the stem of the verb (*vengo, vienes, venís*) and those associated with the variation to be found in the ending (*supieron, dijeron*). As a rule, the variations in the stem can only be handled with extra information in the lexicon file. As for the variation affecting the ending, this is normally phonologically conditioned and so can be generated by simple algorithms in the grammar that perform the appropriate changes automatically. Completely irregular forms – be it due to an unpredictable stem or ending – must also be covered with additional information in the lexicon file.

The various models of inflection defined for each tense in the design of the Spanish generation grammar are described in the following sections.

4.2. *Presente de Indicativo (PrI)*

If we attend to the changes affecting the stem of the verb, it is possible to discover four patterns of variation in the inflection of the PrI:

1. The 11111-model. Most Spanish verbs use a stem for all the forms of the PrI that is obtained after removing the last two letters from the Infinitive form, as table 4 shows. I will refer to this stem obtained from the Infinitive after removing the last two letters as ‘stem1’. The verbs that follow this model do not require additional information in the lexicon file, as stem1 can be automatically extracted from the Infinitive.

Infinitive	<i>amar</i>	<i>comer</i>	<i>partir</i>
1s	<i>amo</i>	<i>como</i>	<i>parto</i>
2s	<i>amas</i>	<i>comes</i>	<i>partes</i>
3s	<i>ama</i>	<i>come</i>	<i>parte</i>
1p	<i>amamos</i>	<i>comemos</i>	<i>partimos</i>
2p	<i>amáis</i>	<i>coméis</i>	<i>partís</i>
3p	<i>aman</i>	<i>comen</i>	<i>parten</i>

Table 4. *Examples of the 11111-model in the PrI tense*

2. The 222112-model. The verbs belonging to this group employ stem1 for the 1p and 2p forms only. For the rest of forms, they use a stem which is slightly different from stem1, as shown in table 5. This stem that is used in the 1s, 2s, 3s and 3p forms of the PrI will be labelled ‘stem2’. Verbs that follow this model must incorporate the information regarding stem2 in the lexicon file, as shown in figure 1.

Infinitive	<i>poder</i>	<i>huir</i>	<i>reír</i>
1s	<i>puedo</i>	<i>huyo</i>	<i>río</i>
2s	<i>puedes</i>	<i>huyes</i>	<i>ríes</i>
3s	<i>puede</i>	<i>huye</i>	<i>ríe</i>
1p	<i>podemos</i>	<i>huimos</i>	<i>reimos</i>
2p	<i>podéis</i>	<i>huís</i>	<i>reís</i>
3p	<i>pueden</i>	<i>huyen</i>	<i>ríen</i>

Table 5. Examples of the 222112-model in the PrI tense

(LEXICAL-ITEM :NAME poder :SPELLING “poder” :FEATURES (222112PrI) :PROPERTIES ((stem2 “pued”)))	(LEXICAL-ITEM :NAME huir :SPELLING “huir” :FEATURES (222112PrI) :PROPERTIES ((stem2 “huy”)))	(LEXICAL-ITEM :NAME reir :SPELLING “reír” :FEATURES (222112PrI) :PROPERTIES ((stem2 “rír”)))
--	---	---

Figure 1. Computational specification of verbs following the 222112-model in the PrI tense

3. The 311111-model. Some verbs employ a stem in the 1s form which is slightly different from stem1. I will refer to this stem that appears only in the 1s form of the PrI tense as ‘stem3’. The rest of forms of the PrI are completely regular, being formed with stem1 (table 6). The information concerning stem3 must be added to the lexicon file in the way shown in figure 2.

Infinitive	<i>poner</i>	<i>hacer</i>	<i>nacer</i>
1s	<i>pongo</i>	<i>hago</i>	<i>nazco</i>
2s	<i>pones</i>	<i>haces</i>	<i>naces</i>
3s	<i>pone</i>	<i>hace</i>	<i>nace</i>
1p	<i>ponemos</i>	<i>hacemos</i>	<i>nacemos</i>
2p	<i>ponéis</i>	<i>hacéis</i>	<i>nacéis</i>
3p	<i>ponen</i>	<i>hacen</i>	<i>nacen</i>

Table 6. Examples of the 311111-model in the PrI tense

(LEXICAL-ITEM :NAME poner :SPELLING “poner” :FEATURES (311111PrI) :PROPERTIES ((stem3 “pong”)))	(LEXICAL-ITEM :NAME hacer :SPELLING “hacer” :FEATURES (311111PrI) :PROPERTIES ((stem3 “hag”)))	(LEXICAL-ITEM :NAME nacer :SPELLING “nacer” :FEATURES (311111PrI) :PROPERTIES ((stem3 “nacz”)))
--	---	--

Figure 2. Computational specification of verbs following the 311111-model in the PrI tense

4. The 322112-model. Finally some verbs combine the 222112 and 311111 models, thus producing a model in which three different stems are employed in the inflection of the PrI tense. Stem3 appears in the 1s form, stem2 is used in the 2s, 3s and 3p forms, while stem1 is restricted to the 1p and 2p forms (table 7). Figure 3 shows how verbs that belong to this group are specified in the lexicon file.

Infinitive	<i>decir</i>	<i>tener</i>	<i>oír</i>
1s	<i>digo</i>	<i>tengo</i>	<i>oigo</i>
2s	<i>dices</i>	<i>tienes</i>	<i>oyes</i>
3s	<i>dice</i>	<i>tiene</i>	<i>oye</i>
1p	<i>decimos</i>	<i>tenemos</i>	<i>oímos</i>
2p	<i>decís</i>	<i>tenéis</i>	<i>oís</i>
3p	<i>dicen</i>	<i>tienen</i>	<i>oyen</i>

Table 7. Examples of the 322112-model in the PrI tense

(LEXICAL-ITEM :NAME decir :SPELLING “decir” :FEATURES (322112PrI) :PROPERTIES ((stem2 “dic”) (stem3 “dig”)))	(LEXICAL-ITEM :NAME tener :SPELLING “tener” :FEATURES (322112PrI) :PROPERTIES ((stem2 “tien”) (stem3 “teng”)))	(LEXICAL-ITEM :NAME oír :SPELLING “oír” :FEATURES (322112PrI) :PROPERTIES ((stem2 “oy”) (stem3 “oig”)))
---	---	--

Figure 3. Computational specification of verbs following the 322112-model in the PrI tense

The variation patterns of the ending in the PrI are mainly based on the conjugation each verb belongs to, as can be seen in table 8.

	1s	2s	3s	1p	2p	3p
1st conjugation	<i>-o</i>	<i>-as</i>	<i>-a</i>	<i>-amos</i>	<i>-áis</i>	<i>-an</i>
2nd conjugation	<i>-o</i>	<i>-es</i>	<i>-e</i>	<i>-emos</i>	<i>-éis</i>	<i>-en</i>
3rd conjugation	<i>-o</i>	<i>-es</i>	<i>-e</i>	<i>-imos</i>	<i>-ís</i>	<i>-en</i>

Table 8. Regular endings in the PrI tense

The only variation on the regular endings shown in table 8 is to be found in the 1s form. Here some verbs add an -y to the normal ending -o (e.g. *doy, estoy, voy*) or replace it with -e (e.g. *he, sé*). Nevertheless these verbs tend to present other anomalies in the rest of forms of the tense, such as:

- a) irregular stems (*soy, eres, es, somos, sois, son*).
- b) unusual position of the word stress with consequences on the spelling: *estoy, estás, está, estamos, estáis, están*.
- c) unusual presence or absence of accents: *sé* (instead of *se*), *sabes, sabe, sabemos, sabéis, saben // doy, das, da, damos, daís* (instead of *dáis, dan*).

Consequently, these verbs are better treated as completely irregular cases. As for the irregular form *hay*, it must be treated as a separate lexical item, since it is unaffected by the categories of person and number when expressing existentiality and so is not a result of the inflectional verb system of Spanish. Figure 4 shows how verbs with a completely irregular PrI tense were specified in the lexicon file.

(LEXICAL-ITEM :NAME ser :SPELLING “ser” :FEATURES (IrrPrI) :PROPERTIES ((PrI1s “soy”)(PrI2s “eres”) (PrI3s “es”) (PrI1p “somos”) (PrI2p “sois”) (PrI3p “son”)))	(LEXICAL-ITEM :NAME haber :SPELLING “haber” :FEATURES (IrrPrI) :PROPERTIES ((PrI1s “he”) (PrI2s “has”) (PrI3s “ha”) (PrI1p “hemos”) (PrI2p “habéis”) (PrI3p “han”)))
---	--

Figure 4. Computational specification of verbs with an irregular PrI tense

4.3. Pretérito Imperfecto de Indicativo (P2I)

The inflection of the P2I is rather regular: the only pattern of stem variation to be found is the 111111-model, that is, only stem1 is employed in all the forms of the tense.

The regular endings attached to the stem in the P2I vary according to the conjugation each verb belongs to, the second and third conjugations being identical. These regular endings are shown in table 9.

	1s	2s	3s	1p	2p	3p
1st conjugation	-aba	-abas	-aba	-ábamos	-abais	-aban
2nd conjugation	-ía	-ías	-ía	-íamos	-íais	-ían
3rd conjugation						

Table 9. Regular endings in the P2I tense

The P2I tense presents no irregularities in the ending, and only two verbs, *ser* and *ir*, have a completely irregular inflection for the P2I tense, which must consequently be fully specified in the lexicon file (figure 5).

<pre>(LEXICAL-ITEM :NAME ser :SPELLING "ser" :FEATURES (IrrP2) :PROPERTIES ((P21s "era") (P22s "eras") (P23s "era") (P21p "éramos") (P22p "erais") (P23p "eran")))</pre>	<pre>(LEXICAL-ITEM :NAME ir :SPELLING "ir" :FEATURES (IrrP2) :PROPERTIES ((P21s "iba") (P22s "ibas") (P23s "iba") (P21p "íbamos") (P22p "ibais") (P23p "iban")))</pre>
---	---

Figure 5. Computational specification of verbs with an irregular P2I tense

4.4. Pretérito Indefinido (P1)

There are three patterns of variation on the stem of the P1:

1. The 111111-model. The majority of regular verbs use stem1 for all the forms of the P1.
2. The 114114-model. Some regular verbs employ a stem in the 3s and 3p forms which is slightly different from stem1 (table 10). This stem that appears in the 3s and 3p forms of the P1 tense will be labelled 'root4'. The information regarding stem4 must be added to the lexicon file, as shown in figure 6.

Infinitive	<i>sentir</i>	<i>dormir</i>
1s	<i>sentí</i>	<i>dormí</i>
2s	<i>sentiste</i>	<i>dormiste</i>
3s	<i>sintió</i>	<i>durmió</i>
1p	<i>sentimos</i>	<i>dormimos</i>
2p	<i>sentisteis</i>	<i>dormisteis</i>
3p	<i>sintieron</i>	<i>durmieron</i>

Table 10. Examples of the 114114-model in the P1 tense

<pre>(LEXICAL-ITEM :NAME sentir :SPELLING "sentir" :FEATURES (114114P1) :PROPERTIES ((stem4 "sint")))</pre>	<pre>(LEXICAL-ITEM :NAME dormir :SPELLING "dormir" :FEATURES (114114P1) :PROPERTIES ((stem4 "durm")))</pre>
--	--

Fig. 6. Computational specification of verbs following the 114114-model in the P1 tense

3. The 55555-model. A group of verbs employ a stem for all the forms of the P1 that is completely different from stem1(table 11). This stem will receive the label 'stem5'. Figure 7 shows how the information regarding stem5 is specified in the lexicon file.

Infinitive	<i>saber</i>	<i>conducir</i>	<i>estar</i>
1s	<i>supe</i>	<i>conduje</i>	<i>estuvo</i>
2s	<i>supiste</i>	<i>condujiste</i>	<i>estuviste</i>
3s	<i>supo</i>	<i>condujo</i>	<i>estuvo</i>
1p	<i>supimos</i>	<i>condujimos</i>	<i>estuvimos</i>
2p	<i>supisteis</i>	<i>condujisteis</i>	<i>estuvisteis</i>
3p	<i>supieron</i>	<i>condujeron</i>	<i>estuvieron</i>

Table 11. Examples of the 55555-model in the P1 tense

(LEXICAL-ITEM :NAME <i>saber</i> :SPELLING “saber” :FEATURES (55555P1) :PROPERTIES ((stem5 “sup”)))	(LEXICAL-ITEM :NAME <i>conducir</i> :SPELLING “conducir” :FEATURES (55555P1) :PROPERTIES ((stem5 “conduj”)))	(LEXICAL-ITEM :NAME <i>estar</i> :SPELLING “estar” :FEATURES (55555P1) :PROPERTIES ((stem5 “estuv”)))
--	---	--

Figure 7. Computational specification of verbs following the 55555-model in the P1 tense

As table 12 shows, the endings of the P1 tense vary with the stem taken by the verb. The presence of root5 is associated with an exclusive set of endings, which are independent of the conjugation to which the verb belongs. Verbs that do not take root5 employ a set of endings which is dependent on the conjugation type. The endings of the second and third conjugations are the same.

		1s	2s	3s	1p	2p	3p
Stem 5		<i>-e</i>	<i>-iste</i>	<i>-o</i>	<i>-imos</i>	<i>-isteis</i>	<i>-ieron</i>
Stem1/Stem4	1st conjugation	<i>-é</i>	<i>-aste</i>	<i>-ó</i>	<i>-amos</i>	<i>-asteis</i>	<i>-aron</i>
	2nd conjugation	<i>-í</i>	<i>-iste</i>	<i>-ió</i>	<i>-imos</i>	<i>-isteis</i>	<i>-ieron</i>
	3rd conjugation						

Table 12. Regular endings in the P1 tense

The endings of the P1 present two types of variation on the models shown in table 12:

- a) reinforcement. It affects the 3s and 3p endings that begin with *-i*, and it takes place when the last letter in the stem is a vowel. The result is that the vowel *-i* of the ending changes into the consonant *-y* (e.g. *hu-ió* > *huyó*; *ca-ieron* > *cayeron*)
- b) deletion. It also affects the 3s and 3p endings that begin with *-i*, and it takes place when the last letter in the stem is an *-ñ*, *-ll*, *-y* or *-i*, plus *-j* if it is part of stem5, not when it is part of stem1. The result is the deletion of the vowel *-i* in the ending (e.g. *tiñ-ió* > *tiñó*; *bull-ieron* > *bulleron*; *ri-ió* > *rió*; *dij_[stem5]-ieron* > *dijeron*, but *tej_[stem1]-ieron* > *tejieron*).

The only completely irregular verbs in the P1 are *ser* and *dar*, which must consequently be fully specified in the lexicon file (figure 8).

<pre>(LEXICAL-ITEM :NAME ser :SPELLING "ser" :FEATURES (IrrP1) :PROPERTIES ((P11s "fui")(P12s "fuiste") (P13s "fue") (P11p "fuimos") (P12p "fuisteis") (P13p "fueron")))</pre>	<pre>(LEXICAL-ITEM :NAME dar :SPELLING "dar" :FEATURES (IrrP1) :PROPERTIES ((P11s "di") (P12s "diste") (P13s "dio") (P11p "dimos") (P12p "disteis") (P13p "dieron")))</pre>
---	--

Figure 8. Computational specification of verbs with an irregular P1 tense

4.5. Futuro de Indicativo (FI)

With a couple of exceptions, all Spanish verbs take stem1 for all the forms of the FI. But unlike all the tenses seen so far, this tense takes stem1 without removing the last two letters. A few verbs however delete the last but one letter from stem1, thus producing contracted Future forms. This phenomenon is unpredictable, so it must be specified in the lexicon file (figure 9) through the inclusion of the feature ‘ContractedFI’. In addition, when the contraction of stem1 leads to the sequences *-nr-* or *-lr-*, an epenthetic *-d-* is inserted. This phenomenon is completely predictable from the phonological context, so it does not require the inclusion of further information in the lexicon file (figure 9). Table 13 shows examples of the different alterations of stem1 in the Future tense.

Infinitive	<i>amar</i>	<i>haber</i>	<i>salir</i>
1s	<i>amaré</i>	<i>habré</i>	<i>saldré</i>
2s	<i>amarás</i>	<i>habrás</i>	<i>saldrás</i>
3s	<i>amará</i>	<i>habrá</i>	<i>saldrá</i>
1p	<i>amaremos</i>	<i>habremos</i>	<i>saldremos</i>
2p	<i>amaréis</i>	<i>habréis</i>	<i>saldréis</i>
3p	<i>amarán</i>	<i>habrán</i>	<i>saldrán</i>

Table 13. Examples of variations on the 111111-model in the FI tense

(LEXICAL-ITEM :NAME haber :SPELLING “haber” :FEATURES (ContractedFI))	(LEXICAL-ITEM :NAME salir :SPELLING “salir” :FEATURES (ContractedFI))
--	--

Figure 9. Computational specification of verbs with a contracted FI tense

Only two verbs, *hacer* and *decir*, take a stem different from stem1 for the formation of the FI tense in Spanish. This stem that appears in the FI will be labelled ‘stem6’. Figure 10 shows how the information related to stem6 is specified in the lexicon file.

(LEXICAL-ITEM :NAME hacer :SPELLING “hacer” :FEATURES (IrrF) :PROPERTIES ((stem6 “har”)))	(LEXICAL-ITEM :NAME decir :SPELLING “decir” :FEATURES (IrrF) :PROPERTIES ((stem6 “dir”)))
---	---

Figure 10. Computational specification of verbs following the 666666-model in the FI tense

Compounds of *hacer*, even if they are etymological (*satisfacer*), also have a stem6. But compounds of *decir* (*bendecir*, *contradecir*) are normally treated as completely regular verbs, although it is possible to find some personal variation here.

The set of endings attached to the stem in the FI tense are the same for the three conjugations and they present no variations. This set of endings is shown in Table 14.

	1s	2s	3s	1p	2p	3p
1st conjugation						
2nd conjugation	-é	-ás	-á	-emos	-éis	-án
3rd conjugation						

Table 14. Regular endings in the FI tense

There are no irregular patterns in the inflection of the FI tense.

4.6. *Condicional (C)*

The inflection of the C tense in Spanish follows exactly the same principles as for the FI tense. The only difference is to be found in the endings, shown in table 15.

	1s	2s	3s	1p	2p	3p
1st conjugation						
2nd conjugation	<i>-ía</i>	<i>-ías</i>	<i>-ía</i>	<i>-íamos</i>	<i>-íais</i>	<i>-ían</i>
3rd conjugation						

Table 15. *Regular endings in the C tense*

There are no irregular patterns in the inflection of the C tense.

4.7. *Presente de Subjuntivo (PrS)*

Attending to the changes in the stem of the verb, we can identify six patterns of variation in the inflection of the PrS:

1. The 111111-model. Most Spanish verbs employ stem1 (after removing the last two letters) for all the forms of the PrS. In fact, we can predict that this is the case whenever a verb follows the 111111-model in the PrI tense, so no specific information must be added to the lexicon file for the generation of the PrS. As we will see below, this is also true of most inflection models of the Subjunctive mode, since they can be predicted from the presence or absence of certain stems in the Indicative.
2. The 222112-model. Some verbs use stem2 for the 1s, 2s, 3s and 3p forms and stem1 for the 1p and 2p forms in the inflection of the PrS (Table 16). We can predict that this will be the case whenever a verb follows the 222112-model in the PrI tense and it has no stem4.

Infinitive	<i>poder</i>	<i>querer</i>
1s	<i>pueda</i>	<i>quiera</i>
2s	<i>puedas</i>	<i>quieras</i>
3s	<i>pueda</i>	<i>quiera</i>
1p	<i>podamos</i>	<i>queramos</i>
2p	<i>podáis</i>	<i>queráis</i>
3p	<i>puedan</i>	<i>quieran</i>

Table 16. *Examples of the 222112-model in the PrS tense*

3. The 222222-model. There is in Spanish a small group of verbs that employ stem2 in all the forms of the PrS (Table 17). These verbs have the characteristics that they follow the 222112-model in the PrI, they have no stem4 and they end in the sequence *-uir* or *-üir* in the Infinitive.

Infinitive	<i>huir</i>	<i>argüir</i>
1s	<i>huya</i>	<i>arguya</i>
2s	<i>huyas</i>	<i>arguyas</i>
3s	<i>huya</i>	<i>arguya</i>
1p	<i>huyamos</i>	<i>arguyamos</i>
2p	<i>huyáis</i>	<i>arguyáis</i>
3p	<i>huyan</i>	<i>arguyan</i>

Table 17. Examples of the 222222-model in the PrS tense

4. The 222442-model. If a verb follows the 222112-model in the inflection of the PrI and it has a stem4, then this verb will use stem2 for the 1s, 2s, 3s and 3p forms of the PrS and stem4 for the 1p and 2p forms (Table 18).

Infinitive	<i>pedir</i>	<i>dormir</i>
1s	<i>pidá</i>	<i>duerma</i>
2s	<i>pidas</i>	<i>duermas</i>
3s	<i>pidá</i>	<i>duerma</i>
1p	<i>pidamos</i>	<i>durmamos</i>
2p	<i>pidáis</i>	<i>durmáis</i>
3p	<i>pidan</i>	<i>duerman</i>

Table 18. Examples of the 222442-model in the PrS tense

5. The 333333-model. Finally some verbs only use stem3 throughout the forms of the PrS (Table 19). This is the case whenever the verb has a stem3.

Infinitive	<i>nacer</i>	<i>tener</i>	<i>oír</i>
1s	<i>nazca</i>	<i>tenga</i>	<i>oiga</i>
2s	<i>nazcas</i>	<i>tengas</i>	<i>oigas</i>
3s	<i>nazca</i>	<i>tenga</i>	<i>oiga</i>
1p	<i>nazcamos</i>	<i>tengamos</i>	<i>oigamos</i>
2p	<i>nazcáis</i>	<i>tengáis</i>	<i>oigáis</i>
3p	<i>nazcan</i>	<i>tengan</i>	<i>oigan</i>

Table 19. Examples of the 333333-model in the PrS tense

6. The 777777-model. A few Spanish verbs have an irregular inflection of the PrS mainly due to the use of an unexpected stem (Table 20). This stem, which will be labelled ‘stem7’, must be specified in the lexicon file (figure 11).

Infinitive	<i>ver</i>	<i>ir</i>
1s	<i>vea</i>	<i>vaya</i>
2s	<i>veas</i>	<i>vayas</i>
3s	<i>vea</i>	<i>vaya</i>
1p	<i>veamos</i>	<i>vayamos</i>
2p	<i>veáis</i>	<i>vayáis</i>
3p	<i>vean</i>	<i>vayan</i>

Table 20. Examples of the 777777-model in the PrS tense

(LEXICAL-ITEM :NAME <i>ver</i> :SPELLING “ver” :FEATURES (777777PrS) :PROPERTIES ((stem7 “ve”)))	(LEXICAL-ITEM :NAME <i>ir</i> :SPELLING “ir” :FEATURES (777777PrS) :PROPERTIES ((stem7 “vay”)))
--	---

Figure 11. Computational specification of verbs following the 777777-model in the PrS tense

The set of endings attached to the stem in the inflection of the PrS vary according to the conjugations each verb belongs to. This set of endings is shown in Table 21. There are no variations on these regular patterns.

	1s	2s	3s	1p	2p	3p
1st conjugation	<i>-e</i>	<i>-es</i>	<i>-e</i>	<i>-emos</i>	<i>-éis</i>	<i>-en</i>
2nd conjugation	<i>-a</i>	<i>-as</i>	<i>-a</i>	<i>-amos</i>	<i>-áis</i>	<i>-an</i>
3rd conjugation						

Table 21. Regular endings in the PrS tense

The only verbs that must be classified as completely irregular in their inflection of the PrS are those that present some unusual behaviour due to the irregular presence or absence of accents, such as *dé* (instead of *de*), *deís* (instead of *déis*), *esté* (instead of *este*), *estés* (instead of *estes*) and *estén* (instead of *estén*). Such irregular cases must be fully specified in the lexicon file (figure 12).

(LEXICAL-ITEM

:NAME *estar*

:SPELLING “*estar*”

:FEATURES (IrrPrS)

:PROPERTIES ((PrS1s “*esté*”)(PrS2s “*estés*”) (PrS3s “*esté*”) (PrS1p “*estemos*”) (PrS2p “*estéis*”) (PrS3p “*estén*”))

Figure 12. Computational specification of verbs with an irregular PrS tense

4.8. Pretérito Imperfecto de Subjuntivo (P2S)

Every Spanish verb follows exactly the same variation patterns that are observed in the formation of the 3p form of the P1 tense when building the P2S. That is, all verbs take the stem found in the 3p form of the P1 for all the forms of the P2S. As a consequence, no specific information must be added to the linguistic resources for the generation of the stem of the P2S.

The endings attached to the stem in the formation of the P2S are shown in Table 22. It will be noticed that the endings attached to the stem of verbs that belong to the first conjugation but contain a stem5 or a completely irregular inflection of the P1 tense are identical to those of verbs that belong to the second or third conjugations. This covers cases such as *anduviera*, *estuviera* and *diera*. In addition, this tense presents the peculiarity of possessing two complete sets of endings with no functional contrast.

	1s	2s	3s	1p	2p	3p
1st conjugation	<i>-ara</i> <i>-ase</i>	<i>-aras</i> <i>-ases</i>	<i>-ara</i> <i>-ase</i>	<i>-áramos</i> <i>-ásemos</i>	<i>-arais</i> <i>-aseis</i>	<i>-aran</i> <i>-asen</i>
2nd conjugation	<i>-iera</i> <i>-iese</i>	<i>-ieras</i> <i>-ieses</i>	<i>-iera</i> <i>-iese</i>	<i>-iéramos</i> <i>-iésemos</i>	<i>-ierais</i> <i>-ieseis</i>	<i>-ieran</i> <i>-iesen</i>
3rd conjugation						
Stem5/IrrP1						

Table 22. Regular endings in the P2S tense

All verbs repeat the alteration of the ending found in the P1 (reinforcement or deletion of *-i*) in the P2S.

4.9. Futuro de Subjuntivo (FS)

The variation patterns of the stem involved in the formation of the FS are exactly the same described for the P2S, so they will not be repeated here.

The endings attached to the stem in the formation of the FS (table 23) are very similar to those involved in the formation of the P2S (Table 22), and in fact they are

organized according to the same principles. They are also affected by the same phenomena of reinforcement and deletion of the *-i* under the same conditions.

	1s	2s	3s	1p	2p	3p
1st conjugation	<i>-are</i>	<i>-ares</i>	<i>-are</i>	<i>-áremos</i>	<i>-areis</i>	<i>-aren</i>
2nd conjugation	<i>-iere</i>	<i>-ieres</i>	<i>-iere</i>	<i>-iéremos</i>	<i>-ereis</i>	<i>-ieren</i>
3rd conjugation						
Stem5/IrrP1						

Table 23. *Regular endings in the FS tense*

4.10. Imperativo (Im)

The changes in the stem of the Im forms produce two patterns of variation:

1. The 11-model. Most Spanish verbs take stem1 for the formation of the Im. This is the case when the verb follows the 111111-model in the formation of the PrI or when it has a stem3, which blocks the use of stem2 in the Im, as Table 24 shows.

Infinitive	<i>temer</i>	<i>tener</i>
2s	<i>teme</i>	<i>ten</i>
2p	<i>temed</i>	<i>tened</i>

Table 24. *Examples of the 11-model in the Im tense*

2. The 21-model. Verbs that have a stem2 but not a stem3 always employ stem2 with the 2s form of the Im, as table 25 shows. The only exceptions are *oír* and the compounds of *decir* (*bendecir*, *maldecir*), which employ stem2 in spite of the fact that they have a stem3.

Infinitive	<i>sentir</i>	<i>volver</i>	<i>bendecir</i>	<i>oír</i>
2s	<i>siente</i>	<i>vuelve</i>	<i>bendice</i>	<i>oye</i>
2p	<i>sentid</i>	<i>volved</i>	<i>bendecid</i>	<i>oíd</i>

Table 25. *Examples of the 21-model in the Im tense*

The endings attached to the stem to form the Im vary according to the conjugation each verb belongs to in the 2s form, but is the same for the 2p form, as Table 26 shows. In addition, the stem employed in the 2s form is obtained after removing the last two letters from stem1, while only one letter must be removed to build the 2p form. This procedure enables us to specify a single ending for the 2p form (*-d*) and also avoids the

problems derived from the presence of an accent in some verbs (*oíd*), as this accent is consistently present in the Infinitive (*oír*) too.

	2s	2p
1st conjugation	-a	-d
2nd conjugation	-e	
3rd conjugation	-e	

Table 26. Regular endings in the *Im* tense

The endings shown in Table 26 remain rather stable in the inflection of verbs. The only variation is that the *-e* of the 2s form is dropped with some verbs. These verbs form a closed set: *poner* and its compounds (*proponer*, *disponer*, *componer*, etc), *venir* and its compounds (*provenir*, *prevenir*, *convenir*, etc), *tener* and its compounds (*retener*, *contener*, *mantener*, etc), *salir* and its compounds (*sobresalir*, etc), and *hacer* and its compounds (*deshacer*, *rehacer*, etc). Historical compounds of *hacer* (*satisfacer*) are used both with the dropping of final *-e* and without it.

In addition the dropping of the final *-e* provokes changes in the spelling of the compounds of *venir*, *tener* and *poner*, as the vowel preceding the consonant *-n-* receives an accent (*propón*, *contén*, *prevén*)

The completely irregular forms of the *Im* we can find in Spanish have two main sources:

- a) the stem employed is unpredictable (*di* instead of ***dice*** or ***diz***);
- b) the stress falls on the ending rather than the stem, sometimes as a result of a vowel contraction (*está* instead of *esta*, *sé* instead of *see*, *ve* instead of *vee*, *prevé* instead of *prevee*). These irregular patterns however only affect the formation of the 2s form, so this is the only form that must be included in the lexicon file.

4.11. *Infinitivo (I)*

The I form needs not be generated, as it is the entry form included in the lexicon file for every verb.

4.12. *Gerundio (G)*

The vast majority of verbs take stem1 (after removing the last two letters) for the formation of the G. The only exception are the verbs that have a stem4, which always use it to build the G (*pedir*: ***pidiendo***; *dormir*: ***durmiendo***). In addition, some verbs with no stem4 also use its equivalent in the G (*decir*: ***diciendo***; *poder*: ***pudiendo***). The verbs *decir* and *poder* should also display this stem4 in the 3s and 3p forms of the P1 tense, but this fact is not visible because the employ a stem5 for that tense. As a result, the

presence of stem4 in the G cannot be predicted and must be specified in the lexicon file as an irregular form.

The endings attached to the stem in the Gerund form vary according to the conjugation each verb belongs to, as Table 27 shows.

	1 st Conjugation	2 nd Conjugation	3 rd Conjugation
Gerund ending	<i>-ando</i>	<i>-iendo</i>	<i>-iendo</i>

Table 27. *Regular endings in the G form*

The endings shown in Table 27 can be altered in certain phonological contexts. Thus the form *-iendo* is reduced to *-endo* if preceded by *-i*, *-ñ*, or *-ll*, and reinforced to *-yendo* if preceded by a vowel or no sound at all (Table 28).

Infinitive	<i>reír</i>	<i>teñir</i>	<i>bullir</i>	<i>caer</i>	<i>huir</i>	<i>ir</i>
Gerund	<i>riendo</i>	<i>tiñendo</i>	<i>bullendo</i>	<i>cayendo</i>	<i>huyendo</i>	<i>yendo</i>

Table 28. *Variations on the regular endings in the G form*

4.13. *Participio (P)*

The P form is built with stem1 (after removing the last two letters). Spanish verbs have either a completely regular P or a completely irregular P form which must be specified in the lexicon file (figure 13). Regular P forms add an ending to the stem which varies according to the conjugation (table 29). There is however a variation on the ending *-ido*: the vowel *i* is accented if the last letter in the stem is the vowel *a*, *e* or *o* (*traído*, *leído*, *roído*).

	1 st Conjugation	2 nd Conjugation	3 rd Conjugation
Participle ending	<i>-ado</i>	<i>-ido / -ído</i>	

Table 29. *Regular endings in the P form*

(LEXICAL-ITEM :NAME ver :SPELLING “ver” :FEATURES (IrrP) :PROPERTIES ((Participle “visto”)))		(LEXICAL-ITEM :NAME decir :SPELLING “decir” :FEATURES (IrrP) :PROPERTIES ((Participle “dicho”)))
--	--	--

Figure 13. *Computational specification of verbs with irregular P form*

5. EXAMPLES OF SPECIFICATION OF SPANISH VERBS FOR GENERATION

In this section I offer some examples of how Spanish verbs must be specified in the lexicon file of the generation system KMPL. It can be seen that the resources described in the preceding sections require a minimum of information concerning the morphological inflection of verbs. Completely regular verbs, for instance, only require the entry form or infinitive (figure 14).

```
(LEXICAL-ITEM
:NAME partir
:SPELLING "partir")
```

Figure 14. Computational specification of the regular verb *partir*

However, the greatest reduction in information is observed in verbs with local irregularities, such as *poder*, *sentir*, *tener* or *hacer*. Regardless of the origin of the irregularity (a change in the vowel of the stem, a change in the consonant, an addition of a consonant, the diphthongization of a vowel), most verbs can be fully generated utilizing an average of three features and three stems (figure 15). This is in sharp contrast to the paradigm-based model found in traditional grammars, which would require the specification of all the forms of the whole conjugation of these verbs.

```
(LEXICAL-ITEM
:NAME poder
:SPELLING "poder"
:FEATURES (222112PrI 555555P1
ContractedFI IrrG)
:PROPERTIES ((stem2 "pued") (stem5
"pud") (gerund "pudiendo")))
```

```
(LEXICAL-ITEM
:NAME sentir
:SPELLING "sentir"
:FEATURES (222112PrI 114114P1)
:PROPERTIES ((stem2 "sient")
(stem4 "sint")))
```

```
(LEXICAL-ITEM
:NAME tener
:SPELLING "tener"
:FEATURES (322112PrI 555555P1
ContractedFI)
:PROPERTIES ((stem2 "tien") (stem3
"teng") (stem5 "tuv")))
```

```
(LEXICAL-ITEM
:NAME hacer
:SPELLING "hacer"
:FEATURES (311111PrI 555555P1
IrrF IrrP)
:PROPERTIES ((stem3 "hag") (stem5
"hic") (stem6 "har") (Participle "hecho")))
```

Figure 15. Computational specification of some verbs with local irregularities

Futhermore, even fully irregular verbs such as *ser* or *haber* demand a considerably smaller amount of information in the lexicon file. Thanks to the modular approach of the analysis, it is possible to omit unnecessary information for the regular parts of the

inflection of these verbs. The information included in the lexicon file is the minimum necessary to generate the irregular parts of the conjugation (figure 16).

(LEXICAL-ITEM

:NAME ser

:SPELLING “ser”

:FEATURES (IrrPrI IrrP2 IrrP1 777777PrS IrrIm)

:PROPERTIES ((PrI1s “soy”)(PrI2s “eres”)(PrI3s “es”)(PrI1p “somos”)(PrI2p “sois”)(PrI3p “son”)(P21s “era”)(P22s “eras”)(P23s “era”)(P21p “éramos”)(P22p “erais”)(P23p “eran”)(P11s “fui”)(P12s “fuiste”)(P13s “fue”)(P11p “fuimos”)(P12p “fuisteis”)(P13p “fueron”)(stem7 “se”)(Imperative “sé”))

(LEXICAL-ITEM

:NAME haber

:SPELLING “haber”

:FEATURES (IrrPrI 555555P1 ContractedFI 777777PrS)

:PROPERTIES ((PrI1s “he”)(PrI2s “has”)(PrI3s “ha”)(PrI1p “hemos”)(PrI2p “habéis”)(PrI3p “han”)(stem5 “huv”)(stem7 “hay”))

Figure 16. *Computational specification of some highly irregular verbs*

6. CONCLUSIONS

This article has presented an analysis of the inflection of the Spanish verb motivated by the needs inherent to the design of linguistic resources for NLG. Establishing models of inflection based on the individual tenses rather than the whole conjugation, as is the norm in existing grammatical descriptions, it is possible to reduce the amount of information that must be added to the linguistic resources. Furthermore, this strategy has revealed the existence of morphological interdependencies between Spanish tenses which enable us to make predictions about the model of inflection followed by a verb utilizing a minimum of information even for the most irregular of Spanish verbs. The principles behind the analysis presented here are probably equally suitable for languages with complex inflection in which former regular patterns have been distorted by phonological changes, as is the case in Italian, French, Portuguese or Romanian.

NOTES

* Correspondence to: Juan Rafael Zamorano-Mansilla, Universidad Complutense de Madrid, Dpto. Filología Inglesa I, Facultad de Filología, Edificio A, Ciudad Universitaria, 28040 Madrid. E-mail: juanrafaelzm@yahoo.es.

1. A few verbs retain the spelling with *j* regardless of the following vowel. These verbs can be recognized by the presence of the letter *j* in the spelling of the Infinitive (*tejer*, *crujir*).

REFERENCES

- Adorni, G. and M. Zock 1996. *Trends in Natural Language Generation—An Artificial Intelligence Perspective*. Berlin: Springer.
- Aguado, G. *et al.* 1998. “ONTOGENERATION: Reusing domain and linguistic ontologies for Spanish text generation.” *Workshop on Applications of Ontologies and Problem Solving Methods, European Conference on Artificial Intelligence*, Brighton, U.K.
- Alarcos Llorach, E. 2000. *Gramática de la lengua española*. Madrid: Espasa Calpe.
- Alcina, J. and J.M. Blecua. 1975. *Gramática española*. Madrid: Gredos.
- Bateman, J. 1997. “Enabling technology for multilingual natural language generation: the KPML development environment”. *Journal of Natural Language Engineering* 3: 15-55.
- Bateman, J. and M. Zock. 2008. *The John Bateman and Michael Zock’s list of Natural Language Generation Systems*. [Documento en Internet disponible en <http://www.fb10.uni-bremen.de/anglistik/langpro/NLG-table/NLG-table-root.htm>].
- Bello, A. 1847. *Gramática de la lengua castellana destinada a uso de los americanos*. Santiago de Chile, Madrid: EDAF.
- Dale, R. and E. Reiter. 2000. *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.
- Elhadad, M. and J. Robin. 1992. “Controlling Content Realization with Functional Unification Grammars.” *Aspects of Automated Natural Language Generation*. Eds. R. Dale, E. Hovy, D. Rosner, and O. Stock. Berlin: Springer Verlag. 89-104.
- Gili Gaya, S. 1961. *Curso superior de sintaxis española*. Barcelona: Spes.
- Langkilde-Geary, I. 2002. “An Empirical Verification of Coverage and Correctness for a General-Purpose Sentence Generator.” *International Natural Language Generation Conference*. Philadelphia, USA.
- Matthiessen, C. and J. Bateman. 1991. *Text Generation and Systemic-Functional Linguistics. Experiences from English and Japanese*. London: Pinter Publishers.
- Mitkov, R. 2003. *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.
- Real Academia Española. 1973. *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa Calpe.
- Reitter, D. 2004. “A development environment for multimodal functional unification generation grammars.” *Third International Conference on Natural Language Generation*. Brighton, U.K.
- Zamorano-Mansilla, J.R. 2002. “La morfología verbal del español y la generación automática.”. *SEPLN* 28: 35-45.