Some Aspects Concerning the Automatic Treatment of Adjectives and Adverbs in Spanish: A Pedagogical Application of the NooJ Platform

Andrea Rodrigo[1(✉)], Silvia Reyes[1], and Rodolfo Bonino[2]

[1] Facultad de Humanidades y Artes, Universidad Nacional de Rosario,
Rosario, Argentina
andreafrodrigo@yahoo.com.ar, sisureyes@gmail.com
[2] IES N° 28 "Olga Cossettini", Rosario, Argentina
rodolfobonino@yahoo.com.ar

Abstract. The purpose of this paper is to present some elements of analysis for the automatic treatment of adjectives and adverbs in Spanish with NooJ. The dictionaries of adjectives and adverbs of the Spanish Module Argentina are the result of combining the dictionaries designed by the research team Infosur led by Dr. Zulema Solana (Universidad Nacional de Rosario), with the Spanish Module Spain-UAB (Universidad Autónoma de Barcelona) supervised by Dr. Xavier Blanco. Following [1], we take as a reference the nucleus verb phrase, and assume there are nucleus adjective and adverb phrases as well. A syntactic unit is presented to integrate morphology and syntax in a preliminary way, and to analyse nucleus adjective phrases (SADJN), and nucleus adverb phrases (SADVN). Finally, an exercise thought of as a pedagogical application of NooJ for students to apply the grammar to a given text is proposed.

Keywords: NLP · NooJ · Pedagogy · Spanish adjectives · Spanish adverbs

1 Introduction

This paper is part of the research project "Computer tools in language processing: a pedagogical application for teacher training". The project has been accredited by the Ministry of Science, Technology and Innovative Production of the province of Santa Fe (Argentina), and is carried out by academics and researchers of two Argentine educational institutions: the Instituto de Enseñanza Superior IES N° 28 "Olga Cossettini" and the Universidad Nacional de Rosario.

To be able to work with Spanish adjectives in NooJ, the research group IES_UNR proceeded to adapt the dictionaries created by the research team Infosur, and to integrate the contributions of the Spanish Module Spain-UAB.

2 The Adjective in Spanish

In Spanish, adjectives show gender and number markings. They may be classified into different types according to different semantic or syntactic criteria. For instance, "apocopable" adjectives may be shortened by apocope in certain positions. Our interest lies in determining types of adjectives in relation to how they combine with other word categories, and to the position they occupy in a phrase. Our attention will be focused on the order changes allowed and on the available combinations.

As participles can be used as adjectives, they are grouped into the category of adjectives. Only the masculine singular (past) participle is used in the formation of compound tenses. But all participles can occur in adjective phrases, whether they are masculine, feminine, singular or plural. It should be also noted that some adjectives can have diminutive and superlative degrees.

2.1 The Spanish Adjective in NooJ

The Spanish Module Argentina dictionary of adjectives contains at the moment more than 18,000 entries. Several lexical entries and their inflectional paradigms will be shown below. Each lexical entry was randomly selected, for each represents the other members of its inflectional paradigm:

    gordo, ADJ + FLX = FLACO
    inconsecuente, ADJ + FLX = RECURRENTE

Participles are added at the end of the dictionary, and identified by the tag [ppio]:

    compensado, ADJ + FLX = FLACO + ppio

As regards spelling variants, and taking as a reference the dictionaries of the Royal Academy of Spanish (RAE), each spelling is entered into a different entry, and if applicable, each variant is assigned a different inflectional paradigm.

    antipiquete, ADJ + FLX = RECURRENTE
    antipiquetes, ADJ + FLX = BURDEOS

If an adjective belongs to two inflectional paradigms, it is entered twice in the dictionary:

    antigás, ADJ + FLX = BURDEOS
    antigás, ADJ + FLX = COMÚN

2.2 Inflectional Paradigms

The inflectional paradigms assigned to most Spanish adjectives are shown below:

    FLACO = <E>/masc+sg | s/masc+pl | <B> a/fem+sg | <B> as/fem+pl;

This paradigm comprises a great number of adjectives, and most adjectives inflect in this way. Within this group the adjectives tagged [+ ppio] are included.

    RECURRENTE = <E>/_+sg | s/_+pl;

In this paradigm, adjectives like *inteligente, triunfante*, are included. The dash means that these adjectives agree with masculine as well as with feminine nouns. The majority of these adjectives possess the feature [+ noinicial].

ÁGIL = <E>/_+sg |es/_+pl;

It includes adjectives like *accidental, gremial*. The dash refers to masculine as well as to feminine nouns. A large number of these adjectives have the feature [+noinicial]

OPOSITOR = <E>/masc+sg | es/masc+pl | a/fem+sg | as/fem+pl;

This paradigm consists of adjectives like *aclarador, deshollinador, desgarrador*.

HARAGÁN = <E>/masc+sg | <A> es/masc+pl | <A> a/fem+sg | <A> as/fem+pl;

Adjectives like *albanés, besucón, bocón*, belong to this paradigm.

The inflectional paradigms for the Spanish adjective are thirteen, and for reasons of space, the paradigms that comprise few adjectives will not be discussed here.

To know the frequency of adjectives comprised in the dictionary in terms of the inflectional paradigms to which they belong, Standard Score is used. Saving the whole dictionary as a text, adjective frequency is analysed with Locate, and then, a statistical analysis of all occurrences is done (see Fig. 1).
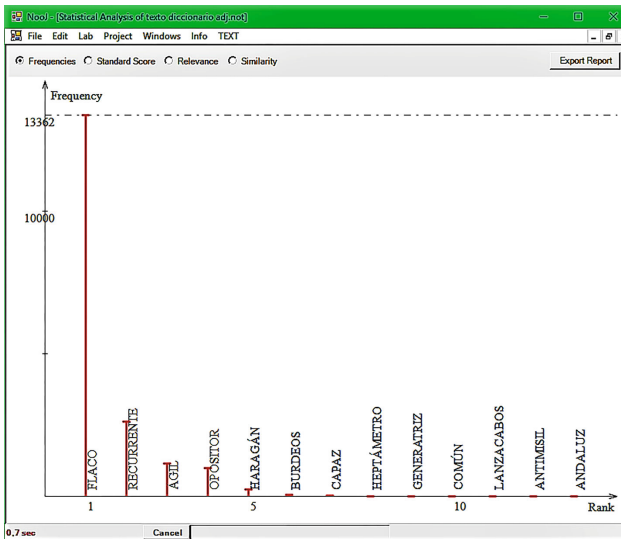

Fig. 1. Frequency of adjective inflectional paradigms in Spanish

2.3 Apocope and Order Constraints

Adjectives like *grande, bueno*, and *malo* are "apocopable", i.e. they are shortened by apocope when they precede a masculine singular noun. So, it is possible to say:

un buen hombre/un hombre bueno (a good man), but not **un bueno hombre*.
estos hombres buenos/estos buenos hombres (these good men)

As it can be clearly seen, constraints occur only in the singular. To mark these constraints, we use the features [apócope] and [apocopable]. Adjectives are entered in the dictionary with one of these two features:

buen, ADJ+masc+sg+apócope
bueno, ADJ+FLX=FLACO+apocopable

Likewise, there are a large number of adjectives that can never precede nouns. Some of them are relational adjectives, as the following:

sede judicial (judicial venue), but not **judicial sede*
rasgo bilabial (bilabial feature), but not **bilabial rasgo*

Many other adjectives, even though they are not relational, cannot either precede nouns. To all these adjectives, we add the feature [noinicial]:

republicano, ADJ+FLX=FLACO+noinicial

2.4 A Morphological Grammar

A morphological grammar to recognize the superlative degree of adjectives is created, specifically for the superlatives constructed with the suffix *-ísimo*. Superlatives ending in *-érrimo* are not included herein, for they are only a few, and are entered in the dictionary with the feature [superl]:

misérrimo, ADJ+FLX=FLACO+superl

This grammar can also recognize diminutives. It is a non-specific grammar able to recognize any sequence of letters, and suitable for adjectives and adverbs as well (see Fig. 2). The adverb will be discussed in the second part of our study.

When this grammar is applied to a specific sample of adjectives and adverbs, it can recognize both superlatives and diminutives. Being a non-specific grammar, it would not be appropriate to generate words, but it can indeed recognize superlative and diminutive degrees.

The analysis of *abrigadísimo*, the superlative degree of the adjective *abrigado*, is as follows:

abrigadísimo, ADJ+superl+masc+sg
abrigadísimo, ADV+superl

2.5 A Syntactic Grammar for the Nucleus Adjective Phrase

Following [1], it is assumed that the adjective forms a nucleus adjective phrase (SADJN) that can either be inside the nucleus noun phrase (SNN), or adjacent to it, as part of a greater unit, the noun phrase (SINNOM).

In this way, a nucleus noun phrase (SNN) is identified as follows: *[la casa]*, where the nucleus is at the end of the phrase, i.e. *casa*. In *[la [hermosa] casa]*, "the beautiful house," the adjective precedes the nucleus, and so the nucleus adjective phrase
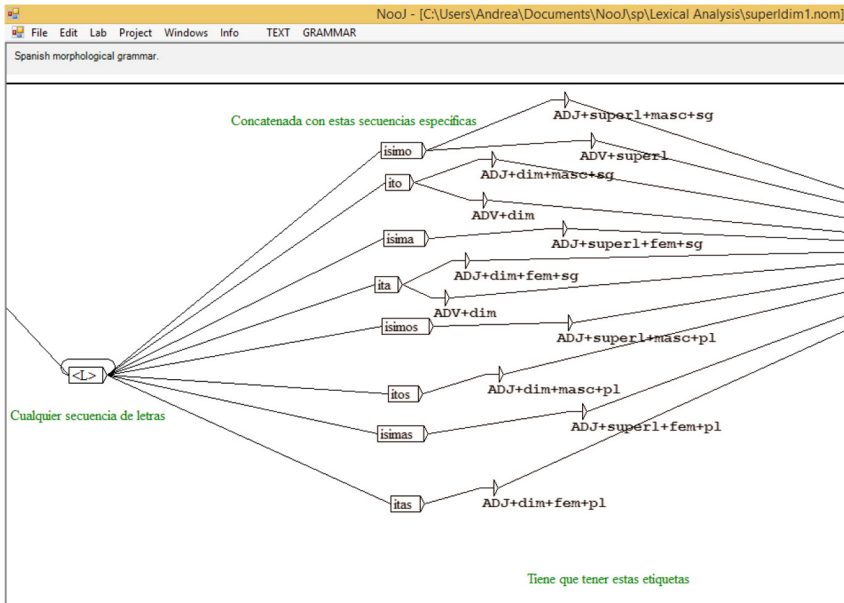
Fig. 2. A morphological grammar for adjectives and adverbs in Spanish

(SADJN), *hermosa*, is inside the nucleus noun phrase (SNN), *la casa*. On the other hand, in a noun phrase (SINNOM) like *[[la casa][hermosa]]*, "the beautiful house," *hermosa* is a nucleus adjective phrase (SADJN) inside the noun phrase (SINNOM), but adjacent to the nucleus noun phrase (SNN). The syntactic grammar recognizes different adjectives inside the noun phrase (SINNOM), or inside the nucleus noun phrase (SNN).

Some nucleus adjective phrases (SADJN) inside the SNN, e.g. *bellas*, en *[las [bellas] mañanas]*, are illustrated in the following graph, <ADJ+fem+pl-noinicial> (see Fig. 3):
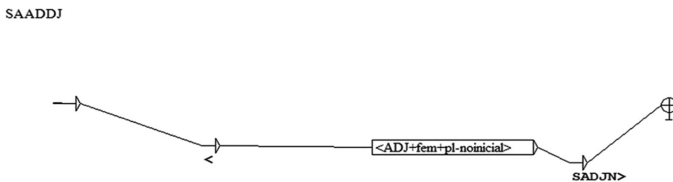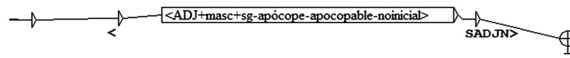

Fig. 3. Embedded graph for feminine plural adjectives inside the SNN

Adjectives shortened by apocope, like *buen*, in *[un [buen] hombre]* "a good man," are represented by the following graph, <ADJ+masc+sg-apócope-apocopable-noinicial> (see Fig. 4):
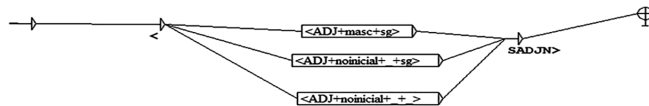
SADDJ

Fig. 4. Embedded graph for adjectives shortened by apocope inside the SNN

The first node <ADJ+masc+sg> of the following graph is suitable for masculine singular adjectives outside the SNN, e.g., *negro*, in *[[el perro][negro]]* "the black dog" (see Fig. 5):

s

Fig. 5. Embedded graph for masculine singular adjectives outside the SNN

The third node <ADJ+noinicial+_+_> of the following graph is applicable to gender-number invariable adjectives outside the nucleus noun phrase (SNN), like *burdeos*, *in [[los vinos][burdeos]]* "the Bordeaux wines" (see Fig. 6):
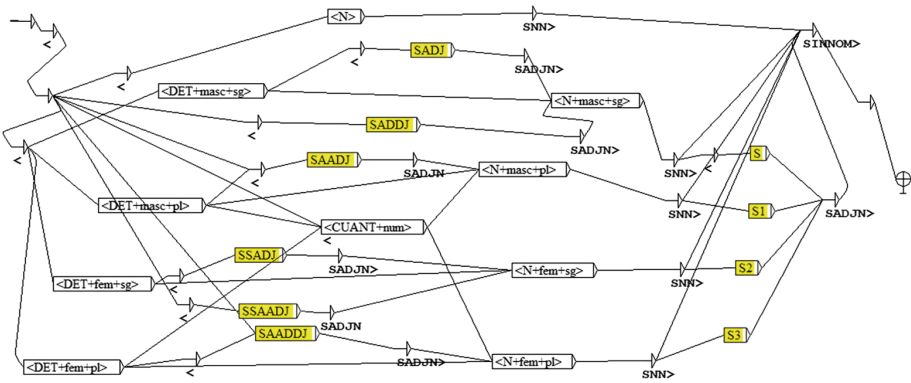
S1

Fig. 6. Embedded graph for masculine plural adjectives, gender invariable adjectives, and number-gender invariable adjectives outside the SNN, the latter nodes with feature [noinicial]

The following graph applies to feminine plural adjectives <ADJ+fem+pl>, e.g. *rosas*, in *[[las sandalias][rosas]]* "the pink sandals," to gender invariable adjectives <ADJ+noinicial+_+pl> , e.g. malvas, in *[[las blusas][malvas]]* "the mauve blouses," and to number-gender invariable adjectives (see Fig. 7):

S3

Fig. 7. Embedded graph for feminine plural adjectives outside the SNN

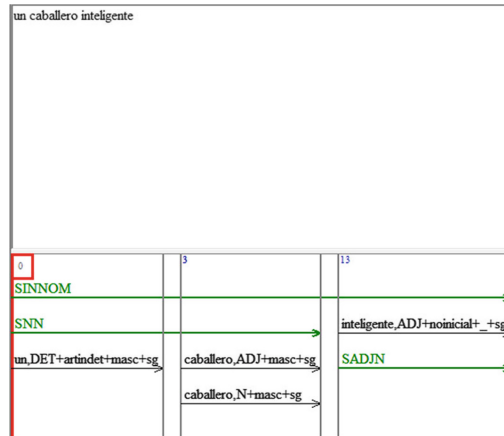Finally, here is the main grammar (see Fig. 8):

Fig. 8. A syntactic grammar for the noun phrase (SINNOM) in Spanish

Next, it can be clearly observed how the grammar is applied to several phrases, and the analysis it performs (see Figs. 9 and 10):

Fig. 9. Analysis of *un gran dictador* (a great dictator): a nucleus adjective phrase (SADJN) inside a nucleus noun phrase (SNN)

Fig. 10. Analysis of *un caballero inteligente* (an intelligent knight): a nucleus adjective phrase (SADJN) inside the noun phrase (SINNOM)

3   The Adverb in Spanish

Unlike adjectives, adverbs do not present gender and number markings. However, like adjectives, adverbs may have diminutive and superlative degrees. Different types of adverbs are identified and defined according to the combinations they can assume inside the nucleus adverb phrase (SADVN) [2]:

ADV+1: These adverbs always appear alone inside the SADVN, and do not admit modifiers like *equívocamente, finalmente, físicamente*.

ADV+2: These adverbs may either appear alone, or admit modifiers inside the SADVN. Three types of adverbs can be distinguished herein:

ADV+2a (advnuc): they are always the nucleus of the SADVN: *concisamente, abajo, adentro*, but can also be modified by another adverb, like a quantifier: *muy abajo, bien adentro*.

ADV+2b (advmodif): adjectives like *excesivamente* can modify the nucleus, and the nucleus may be an adv2a (advnuc), *abajo*, or an adv2c (adververs), *bien*: *excesivamente abajo, excesivamente bien*.

ADV+2c (advrevers): adjectives like *completamente* may be the nucleus of the SADVN; but when they are not the nucleus, unlike adv2b, they can either modify an adv2a, *completamente adentro*, or be modified by an adv2b, *casi completamente*.

3.1   The Spanish Adverb in NooJ

The Spanish Module Argentina dictionary of adverbs contains around 400 entries. The distinction between adv1, whose morphological ending is *-mente* (bearing the number 1), and the other adverbs (identified with the number 2) must be included in the dictionary. This distinction is necessary because the adverbs ending in *-mente* cannot appear together consecutively. So, when any combination is proposed, for example,

adv2b with adv2a, which is very common, this distinction prevents an ungrammatical combination like *estupendamente divinamente*. Therefore, adverbs are described as follows:

finalmente, ADV + 1
concisamente, ADV + 2a1

Since the morphological grammar used for the adjective is non-specific, it can be modified in order to analyse superlative adverbs (see Fig. 11).
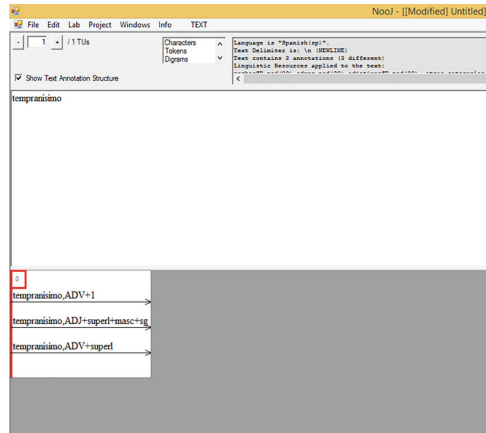

Fig. 11. Analysis of *tempranísimo* (very early), the superlative of the adverb *temprano* (early)

3.2   A Syntactic Grammar for the Nucleus Adverb Phrase

Following [1], the existence of nucleus adverb phrases (SADVN) is assumed. They can either be inside the nucleus verb phrase (SVN), or adjacent to it, inside the verb phrase
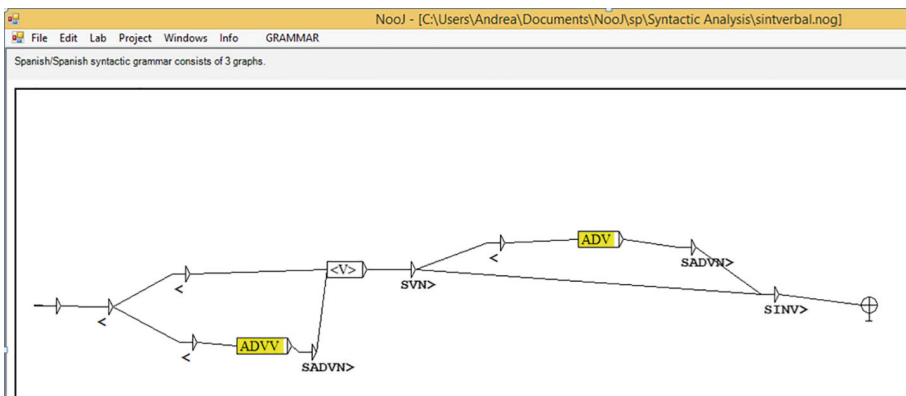

Fig. 12. A grammar for the SINV: the SADVN may be inside SVN, or adjacent to it.

(SINV). Thus, inside the verb phrase (SINV), e.g. *[comen [rápidamente]]*, "they eat quickly," there is a nucleus adverb phrase (SADVN), *[rápidamente].* It is possible to add a quantifier to the adverb and say: *[comen [muy rápidamente]]*, "they eat very quickly." The quantifier *muy* is used here as a modifier of the nucleus in the nucleus adverb phrase (SADVN).

The main grammar is shown in the figure above (Fig. 12).

After applying Show Debug, two possibilities are observed: adverbs inside the nucleus verb phrase (SVN), and adverbs inside the verb phrase (SINV) (see Fig. 13).
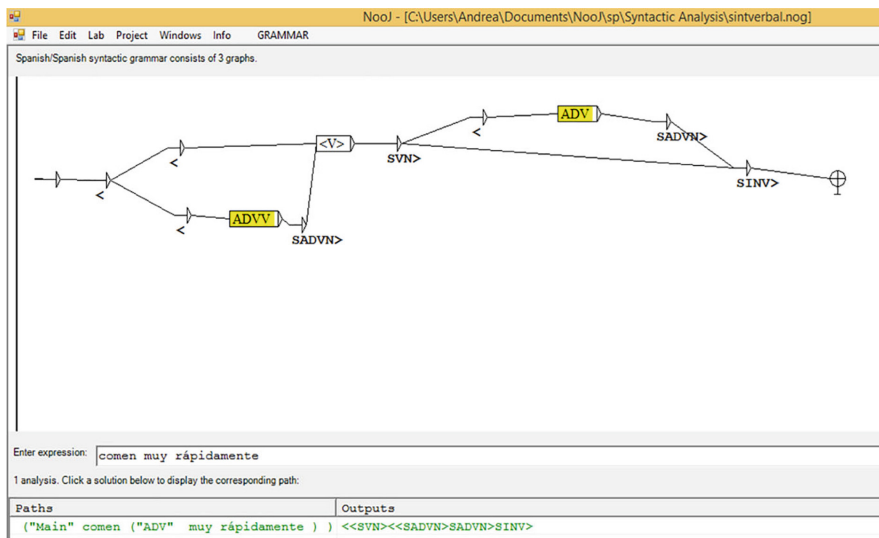

Fig. 13. The expression *comen muy rápidamente* (they eat very quickly) is considered correct.

4   A Pedagogical Application of NooJ

An exercise was proposed for students to apply the grammar to a given text.

4.1   Purposes

– To observe how adjectives that are shortened by apocope appear in real texts.
– To draw conclusions about grammatical constraints in real texts.

4.2   Activity

The text to be analysed is a fragment of *El conde Lucanor: "Lo que sucedió a un hombre bueno con su hijo."* [3] Students were asked to do a search of <ADJ> with Locate. Attention was drawn towards the adjective *bueno* and its particular behaviour. Students had to answer two questions: What happens to the adjective *bueno* in the two

occurrences found by Locate? What happens when these occurrences, which are in the singular, are turned into the plural? Then, students were asked to analyse the dictionary entries by pressing CTRL F:

buen, ADJ + masc + sg + apócope
bueno, ADJ + FLX = FLACO + apocopable

4.3   Evaluation

Students were asked to perform a search and look for other adjectives behaving like bueno, and to write a short text using this type of adjectives.

This exercise will be included in the didactic material folder under preparation.

5   Conclusions and Future Perspectives

We have illustrated some elements related to the automatic treatment of adjectives and adverbs in Spanish, by taking advantage of the possibilities offered by NooJ, in order to improve and enhance our research horizon, which is the teaching of Spanish both as a first (L1) and as a second language (L2). An inflectional grammar for the adjective in particular, a morphological grammar that can either be used for the adjective as well as for adverb phrases, a syntactic grammar for the adjective, and another for the adverb were shown. Our work constitutes a first step towards the formalization of the study of adjectives and adverbs, since many aspects have not been yet considered. However, the continuous aspiration of our ongoing research project is to deal with new issues as we continue to enlarge and improve the dictionaries and grammars of the IES_UNR Spanish Module Argentina, now in full development. Our next step will be to study numeral adjectives, on the one hand, and adjectival adverbs, like *llovió fuerte*, "it rained hard," on the other, and to share our progress in future papers.

References

1. Bès, G.G.: La phrase verbal noyau en français. Recherches sur le français parlé **15**, 273–358 (1999). Université de Provence
2. Rodrigo, A.: Tratamiento automático de textos: el sintagma adverbial núcleo. Tesis doctoral. Facultad de Humanidades y Artes. UNR. Ediciones Juglaría, Rosario (2011)
3. Lucanor, C.: Lo que sucedió a un hombre bueno con su hijo. http://www.guillermoromero.es/cuentospopulares/index.php?section=47&page=25