# A Pedagogical Application of NooJ in Language Teaching: the Adjective in Spanish and Italian

Andrea Rodrigo

Universidad Nacional de Rosario Entre Ríos 758 S2000CRN Rosario, Argentina Mario Monteleone Universitá di Salerno Via Giovanni Paolo II, 132 84084 Fisciano SA, Italy

andreafrodrigo@yahoo.com.ar

mmonteleone@unisa.it

Silvia Reyes Universidad Nacional de Rosario Entre Ríos 758 S2000CRN Rosario, Argentina sisureyes@gmail.com

# Abstract

This paper relies on the work developed by the research team IES\_UNR (Argentina) and presents a pedagogical application of NooJ for the teaching and learning of Spanish as a foreign language. However, as this proposal specifically addresses learners of Spanish whose mother tongue is Italian, it also entailed vital collaboration with Mario Monteleone from the University of Salerno, Italy. The adjective was chosen on account of its lower frequency of occurrence in texts written in Spanish, and particularly in the Argentine Rioplatense variety, and with the aim of developing strategies to increase its use. The features that the adjective shares with other grammatical categories render it extremely productive and provide elements that enrich the learner's proficiency. The reference corpus contains the front pages of the Argentinian newspaper *Clarín* related to an emblematic historical moment, whose starting point is March 24, 1976, when a military coup began, and covers a thirty year period until March 24, 2006. The use of the linguistic resources created in NooJ for the automatic processing of texts written in Spanish accounts for the adjective in a relevant historical context for Argentina.

# 1 Introduction

## 1.1 Research Subject

Our subject matter deals with the pedagogical application of NooJ in the teaching and learning of Spanish as a foreign language, which relies on the research project carried out by the IES\_UNR group since 2015 (Rodrigo et al. 2018). An antecedent of this work can be observed in Frigière and Fuentes (2015), which examines the pedagogical application of NooJ for learners of French as a foreign language. In this paper, we focus on a specific group: native speakers of Italian learning Spanish as a foreign language. Our interest lies in a main grammatical category: the adjective. And our work entails the following steps: (1) selection of a reference corpus in Spanish; (2) presentation of the adjective in relation to other categories and inserted in real texts; (3) use of NooJ for automatic text processing (Silberztein 2015); and (4) presentation of a pedagogical proposal specifically directed to learners of Spanish whose mother tongue is Italian (see Appendix B).

## 1.2 Reference corpus

Our approach to Spanish takes the Argentine Rioplatense variety<sup>1</sup> and centres on journalistic texts produced during thirty years in Argentina. These written texts represent an important historical moment since they are concerned with the last dictatorship and the democratic period immediately following it.

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details: http://crea-tivecommons.org/licenses/by/4.0/

<sup>&</sup>lt;sup>1</sup> Although our conclusions may also apply to other Spanish dialects, whether European or American varieties.

Our reference corpus, the *Clarín*\_corpus,<sup>2</sup> spans the years 1976 to 2006, and includes the same date each year, March 24, which was the day the Armed Forces took power by force in 1976 and overthrew the democratic Argentine government in charge of President María Estela Martínez de Perón. This repressive de facto regime known as *la Junta Militar* "the Military Junta" seized power and ruled until December 1983, when a democratic government was again elected by the Argentinian people. The first front page that was published by *Clarín* when the coup d'état began is displayed in Figure 1.



Figure 1. Front page published on March 24, 1976.

The phrase *Nuevo Gobierno* "New Government" is a distant and neutral expression, whereas at the bottom right corner of the last front page of the corpus, which corresponds to March 24, 2006 (see Appendix A), the word *noche* "night," in the expression *A 30 años de la noche más larga* "30 years after the longest night," overtly condemns the beginning of the military regime.

The corpus comprises 2,898 words and always deals with the front pages of the daily newspaper *Clarín*, assuming the front page to be a focus of journalistic and ideological attention.

## 1.3 The adjective

At a first stage, the adjective is presented in relation to the noun, the verb and the adverb; then, the frequency of adjectives in the *Clarín*\_corpus is detected, in contrast to other categories. At a second stage, our intention is to observe what types of adjectives appear there. For instance, if the term *desaparecido* "disappeared" – which from a historical viewpoint has so many connotations in Argentina – is used, we observe what prefixes and suffixes are allowed with derivations, and in what types of constructions it is included. It is important to recall that the word *desaparecido* is also listed in English dictionaries. The *Oxford Dictionary of English*<sup>3</sup> defines it as, "(especially in South America) a person who has disappeared, presumed killed by members of the armed services or the police." And the *Merriam-Webster's Dictionary*<sup>4</sup> offers the following definition: "an Argentine citizen who has been abducted and usually murdered by right-wing terrorists."

<sup>&</sup>lt;sup>2</sup>*Clarín* Corpus available at http://tapas.clarin.com/tapa.html#19780324

<sup>&</sup>lt;sup>3</sup>Oxford Dictionary of English available at https://en.oxforddictionaries.com

<sup>&</sup>lt;sup>4</sup>*Merriam-Webster's Dictionary* available at https://www.merriam-webster.com

#### 1.4 Use of NooJ

We have been working with the linguistic resources created for the Spanish Module Argentina,<sup>5</sup> specifically built by the IES\_UNR research team to be used with NooJ,<sup>6</sup> a computational tool designed by Max Silberztein. Our linguistic resources include dictionaries and grammars that are now in full development. In this paper we will particularly deal with the dictionary of adjectives and some specific grammars, leaving aside the dictionaries of other grammatical categories, which we use to recognise different types of entries such as nouns, verbs and adverbs.

Similarly, the linguistic resources created for the Italian Module<sup>7</sup> provide an anchor to the Italian learners of Spanish as a foreign language. Notice however that the Italian Module is not our main purpose, for it is only tangentially addressed as a starting point for those learners.

The choice of NooJ entails, above all, a methodological decision, which assumes that it is possible to create language teaching and learning situations by taking into account natural language processing. Then, it is understood that in a given environment and with a given input, learners formulate linguistic hypotheses that may be checked with a computational tool. The NooJ platform provides a favourable means to create this type of situations since: (1) it articulates different levels of analysis (orthography, morphology, syntax, etc.); (2) its connected dictionaries and grammars involve an integral knowledge of language; and (3) a single platform allows a comparison of Italian and Spanish using a common language.

We will try to formulate possible language teaching and learning strategies for learners of Spanish as a foreign language, whose mother tongue is Italian.

### 2 Development

#### 2.1 Comparison between the adjective and other categories in Spanish

From a morphological viewpoint, Spanish nouns and adjectives share the same diminutive suffix *-ito* (e.g., *mesa* "table" *- mesita* "little table" and *claro* "clear" *- clarito* "very clear"). Many adjectives derive from nouns as in: *cultura* "culture" *- cultural* "cultural" and *trabajo* "work" *- trabajador* "worker". From a syntactic viewpoint, an adjective may become the head of a noun phrase, as in the sentence *el militar abrió fuego* "the military (man) opened fire," where the adjective *militar* "military" functions as a noun.

From a morphological viewpoint, it is possible to create adjectives out of certain verbs (e.g., *tonificar* "to tone" - *tonificante* "toning" and *desear* "to desire" - *deseable* "desirable"). From a syntactic viewpoint, past participles and adjectives occur in common environments, as in: *un traje combinado* "a combined suit" and *un traje azul* "a blue suit."

From a morphological viewpoint, adjectives and adverbs admit the same diminutive and superlative suffixes -*ito* and -*ísimo* respectively (e.g., *buenito* "very good" - *tempranito* "very early" and *buenísimo* "great" - *tempranísimo* "very early"). Moreover, adverbs in *-mente* derive from adjectives as in: *lento* "slow" - *lentamente* "slowly." From a syntactic and semantic viewpoint, adjectival adverbs constitute an intersection of adjectives and adverbs, as in the sentence *llovió fuerte* "it rained hard," where *fuerte* "hard" functions as an adjectival adverb modifying the verb *llovió*; whereas in the phrase *un fuerte viento* "a strong wind," *fuerte* "strong" is simply an adjective modifying a noun.

### 2.2 Comparison between the adjective and other categories in Italian

In Italian, adjectives have a high percentage of categorical ambiguity with two other grammatical categories, specifically those of the noun and the past participle (Monteleone 2016). A typical example of this ambiguity is the word *anguished*. In the sentence *Giovanni è angosciato* "Giovanni is anguished," *angosciato* is an adjective. And in the sentence *Paolo è un angosciato* "Paul is an anguished man," *angosciato* functions as a noun. However, in the sentence *Paolo è angosciato da Giovanna* "Paolo is anguished because of Giovanna," *angosciato* is a past participle.

<sup>&</sup>lt;sup>5</sup>Spanish Module (Argentina) available at http://www.nooj-association.org/index.php?option= com\_k2&view=item&id=6:spanish-module-argentina&Itemid=611)

<sup>&</sup>lt;sup>6</sup> Software NooJ available at http://nooj-association.org

<sup>&</sup>lt;sup>7</sup>Italian Module available at http://www.nooj-association.org/index.php?option=com\_k2&view= item&task=download&id=80\_cf86247a0b995f68b04b2d484130c7f4&Itemid=611)

However, through specific automatic parsing tools (grammars, automata and finite-state transducers in NooJ), it is possible to solve this ambiguity and automatically tag syntactic sequences containing only adjectives. For reasons of space, those tools will not be discussed here, but developed in a future paper.

### 3 Analysis of the *Clarín*\_corpus

#### 3.1 Category frequency

Within NooJ, we apply the command "Linguistic Analysis" in order to obtain the frequency of the categories contained in the corpus: nouns, verbs, adjectives and adverbs. Then, by applying Locate for each category in turn: <N>, <V>, <ADJ> and <ADV>, the following piece of information is obtained: Nouns: 1,200; Verbs: 658; Adjectives: 353; Adverbs: 137; Other categories: 550 (conjunctions, prepositions, pronouns).

The graph in Figure 2 shows that adjectives and adverbs are the least frequent out of the five categories. This data provides a first conclusion. The low frequency of adjectives points out that, due to its low proportion in the text input, this category is susceptible to be reinforced by the foreign language teacher's intervention.



Figure 2. Frequency of grammatical categories in the *Clarín\_*corpus.

### 3.2 Adjectives

As frequency is our working criterion, those adjectives recording more frequency in the corpus represent the starting point. It is then specified which categories precede these adjectives, which morphological suffixes and prefixes they admit, and which adjectival structures are more recurrent.

In Figure 3, after applying the commands "Locate" and "Statistical Analysis," it is possible to recognise the twenty most frequent adjectives in the corpus: *militar* "military," *político* "political," *política* "political," *desaparecidos* "disappeared," *general* "general," *muertos* "dead," *derechos* "right," *Militar* "Military," *Argentina* "Argentinian, *listas* "ready," *heridos* "injured," *humanos* "human," *Armadas* "Armed, *solo* "alone," *fuerte* "strong," *primer* "first," *exclusivo* "exclusive," *argentino* "Argentine," *central* "central" and *especial* "special."

Some of the adjectives of the above list are selected in order to specify their inflectional and derivational paradigms, their context of occurrence and their syntactic structure.



Figure 3. Frequency of adjectives in the *Clarín*\_corpus.

#### 3.3 Morphological Inflection and Derivation

Next, the inflectional paradigms of some of the most frequent adjectives are shown, as they are recorded in the inflectional grammars (Rodrigo et al. 2018) and in the dictionary of adjectives created in NooJ for the Spanish Module Argentina.

The adjective *militar* "military" inflects like ÁGIL: *militar, militares*. The following words are part of its derivational paradigm as recorded in the dictionary of adjectives: *antimilitar* "anti-military," *anti-militarista* "anti-militarist," *paramilitar* "paramilitary," *desmilitarizado* "demilitarised," *militarizado* "militarised."

The adjective *político* "political" belongs to the paradigm FLACO: *político, política, políticos, políticas*. The words *apolítico* "apolitical," *geopolítico* "geopolitical" and *impolítico* "non-political" are part of its derivational paradigm in the dictionary of adjectives.

The adjective *desaparecido* "disappeared" also follows the paradigm FLACO: *desaparecido, desaparecida, desaparecidas, desaparecidas*. It is important to note that there are no words derived from *desaparecido*.

The adjective *general* "general" inflects like ÁGIL: *general, generales*. The words *generalista* "generalist," *generalizable* "generalisable" and *generalizado* "generalised" are part of its derivational paradigm as it is recorded in the dictionary of adjectives.

### 3.4 Context of occurrence

The context of occurrence allows us to know which categories can accompany the adjective. The adjective *militar* "military" is preceded by nouns, as in the phrases *reacción militar* "military reaction," *autocrítica militar* "military self-criticism," *caudillo militar* "military leader," *uniforme militar* "military uniform," *golpe militar* "military coup," *colegio militar* "military college."

The adjectives *político*, *política* "political" are preceded by nouns, as in: *sector político* "political sector," *Reforma política* "political Reform," *favoritismo político* "political favouritism," *enemigo político* "political enemy" and *terror político* "political terror." However, they may also be preceded by determiners: *su política* "his/her/their/your politics," *la política* "(the) politics."

The adjective *desaparecidos* "disappeared" is preceded by verbs (e.g., *son desaparecidos* "(they) are disappeared"), by the construction adjective+preposition (e.g., *miles de desaparecidos* "thousands of disappeared"), by numeral adjectives (e.g., *catorce desaparecidos* "fourteen disappeared"), and by nouns (e.g., *alumnos desaparecidos* "disappeared students").

#### 3.5 Syntactic structure

The analysis of two expressions containing adjectives and selected from the *Clarín\_*corpus, *Hubo miles de desaparecidos* "There were thousands of disappeared" and *último golpe militar* "last military coup" will be shown. The first example is an impersonal sentence and the second one is a noun phrase.

The basis of our analysis are the concepts developed by Bès (1999) and our approach is related to context-sensitive grammars. Our NooJ grammars take into account two types of constructions: nucleus phrases, such as a nucleus verb phrase (SVN), a nucleus adjective phrase (SADJN), a nucleus preposition phrase (SPN), a nucleus noun phrase (SNN), and larger phrases, such as the noun phrase (SINNOM), which contain nucleus phrases. These expressions will be analysed with the different syntactic grammars available in the Spanish Module Argentina.

In Spanish, the expression *Hubo miles de desaparecidos* "There were thousands of disappeared" is an impersonal sentence, i.e., without a subject. In the above example, we distinguish a nucleus adjective phrase (SADJN), *miles* "thousands" and a nucleus preposition phrase (SPN), *de desaparecidos* "of disappeared," both inside the noun phrase (SINNOM). On the other hand, the verb *hubo* "were," which is the past indicative of *haber* "to be," constitutes a nucleus verb phrase (SVN). These two structures make up the impersonal sentence, as it can be seen in Figure 4.



Figure 4. Analysis of Hubo miles de desaparecidos.

The corresponding main grammar for impersonal sentences is displayed in Figure 5.



Figure 5. Grammar for impersonal sentences.

As regards the second expression, the noun phrase *último golpe militar* "last military coup", instead of analysing it with the "Linguistic Analysis" command, we use the NooJ "Show Debug" feature instead. By using "Show Debug," it is possible to analyse whether the phrase entered can be considered acceptable, when the noun phrase grammar is applied. As can be seen in Figure 6, the result is a Perfect Match.



Figure 6. Noun Phrase Grammar.

"Show Debug" checks to see if the grammar designed for the noun phrase identifies that *último* "last" and *militar* "military" are two nucleus adjective phrases (SADJN), inside a larger unit, the noun phrase (SINNOM). The intermediate unit is the nucleus noun phrase (SNN), *último golpe* "last coup."

The analyses of both expressions, *Hubo miles de desaparecidos* "There were thousands of disappeared" and *último golpe militar* "last military coup," show the adjective in two different syntactic constructions. In the first example, the adjective *miles* "thousands" is the nucleus of the noun phrase (SINNOM) *miles de desaparecidos* "thousands of disappeared" and is part of an impersonal sentence; in the second example, two adjectives occur in different positions: *último* "last" is a nucleus adjective phrase (SADJN) inside the nucleus noun phrase (SNN) *último golpe* "last coup," and, at the same time, *último* "last" is part of the noun phrase (SINNOM) *último golpe militar* "last military coup," and *militar* "military" is a nucleus adjective phrase (SADJN) inside a noun phrase (SINNOM), peripheral to the nucleus noun phrase (SNN).

The syntactic structure shows how the adjective can integrate noun phrases and even become the nucleus of a noun phrase.

#### 4 Conclusions

In this paper, we have proposed a pedagogical application of the NooJ approach to a specific situation by choosing a significant corpus (the *Clarín*\_corpus), which can provide an idea of the use of the Spanish language in a particular historical and cultural context. Our intention has been to identify adjectives and to do a global reading of them within the corpus. Our focus was mainly placed on the adjective *desaparecido* "disappeared," because of its ideological importance. As we have already stated in the Introduction, our subject matter deals with the pedagogical application of NooJ in the teaching-learning of Spanish as a foreign language. And we have concentrated on a specific group: native speakers of Italian learning Spanish as a foreign language. Our strategy has been to make explicit some contact points between Spanish and Italian by using NooJ as a common platform. Our purpose was not to compare the Italian and Spanish resources, but to highlight that both languages can be formalised using a common platform, since the structure of dictionaries and grammars is the same. This provides unity and coherence to our pedagogical proposal, which is the goal of the research project carried out by the IES\_UNR team, since both languages can be analysed with NooJ (Appendix B).

#### References

Gabriel G. Bès, 1999. La phrase verbale noyau en français. *Recherches sur le français parlé*, volume 15: 273–358. Université de Provence.

- Andrea Rodrigo, Silvia Reyes and Rodolfo Bonino. 2018. Some Aspects Concerning the Automatic Treatment of Adjectives and Adverbs in Spanish: A Pedagogical Application of the NooJ Platform. In: Mbarki S., Mourchid M., Silberztein M. (eds) Formalizing Natural Languages with NooJ and Its Natural Language Processing Applications. NooJ 2017. Springer, Cham. <u>https://doi.org/10.1007/978-3-319-73420-0\_11</u>. *Communications in Computer and Information Science*, volume 811: 130-140.
- Julia Frigière and Sandrine Fuentes. 2015. Pedagogical Use of NooJ dealing with French as a Foreign Language. In J. Monti, M. Silberztein, M. Monteleone, and M. P. di Buono (eds.) *Formalizing Natural Languages with NooJ*, Cambridge Scholars Publishing, London, 186-197.
- Mario Monteleone. 2016. Local Grammars and Formal Semantics: Past Participles vs. Adjectives in Italian. In T. Okrut, Y. Hetsevich, M. Silberztein, and H. Stanislavenka, (eds) Automatic Processing of Natural-Language Electronic Texts with NooJ. *Communications in Computer and Information Science*, volume 607:83-95, Springer, Cham.

Max Silberztein, 2015. La formalization des langues, l'approche de NooJ. Iste Ediciones, London.

## Appendices

#### A. Last front page of the Clarín\_corpus.

The last front page that was published by Clarín (2006) is displayed in Figure 7.



Figure 7. Last front page published on March 24, 2006.

## B. A pedagogical proposal.

Our work will be carried out in parallel to the NooJ Italian module, since the starting point is the same platform and our proposal includes the activities detailed below.

- Open the text in the Italian Module: Ringraziamenti (Dan Brown, The Da Vinci Code, 2003)<sup>8</sup>.
- Apply Locate <A> in order to extract all adjectes, as stated in the properties\_definiton file (Figure 8):

di questo libro. E l' E l'incomparabile Heide Lange, l'incomparabile Heide Lange, instancabile difensore del Codice da Vinci, del Codice da Vinci, agente Codice da Vinci, agente straordinaria, da Vinci, agente straordinaria, amica pienamente la gratitudine per l' squadra della Doubleday, per la la fede e la guida Bill Thomas e Steve Rubin, Steve Rubin, che hanno creduto 'inizio. Grazie anche al gruppo anche al gruppo iniziale di gruppo iniziale di sostenitori all' sostenitori all'interno della casa all'interno della casa editrice.	incomparabile instancabile difensore agente straordinaria amica fidata eccezionale loro illuminata che in iniziale sostenitori interno editrice guidato	Heide Lange, instancabile difensore del difensore del Codice da Vinci del Codice da Vinci, agente straordinaria, amica fidata. Non riuscirei , amica fidata. Non riuscirei mai fidata. Non riuscirei mai fidata. Non riuscirei mai a. Non riuscirei mai a squadra della Doubleday, per la generosità, la fede e la . Un grazie soprattutto a Bill hanno creduto in questo libro questo libro fin dall'inizio di sostenitori all'interno della all'interno della casa editrice della casa editrice, guidato da , guidato da Michael Palgon, Suzanne da Michael Palgon, Suzanne Herz
• • •		

Figure 8. Locate <A> in the Italian text.

• Apply Locate and search a string of characters for the adjective *buon* "good." Note how concordance is done. Figure 9 presents a sample of some of the sequences.

dall'interno. «Il dolore è	buon	o, Monsieur» disse l'albino. Poi
a Dio. 'Il dolore è	buon	o.' I suoi occhi dalle iridi
di purificazione. 'Il dolore è	buon	o' sussurrò fra sé, ripetendo le
che non sarebbe stata una	buon	a idea. Mentre si muoveva verso
spalline dei nostri generali.» 'Con	buon	a pace della dea dell'amore
meno di pensare che le	buon	e grazie di cui godeva a
dagli altri agenti per mantenere	buon	e relazioni con la comunità. Quando
'altra per avere infangato il	buon	nome della Chiesa cattolica.' Collet
momento infelice, era probabilmente una	buon	a notizia. Fache, dopo avere scoperto
«L'ho spento per un	buon	motivo» disse Fache, soffiando come
ucciso Jacques Saunière?» «Ha alcune	buon	e ragioni per sospettare di lei
salto. «Con un po' di	buon	a mira, lei può uscire di
solo dopo averne percorso un	buon	tratto si riusciva a scorgere
una zona alberata, a una	buon	a distanza dalla strada. Vernet comparve
«Quest'uomo è veramente un	buon	amico?» chiese Sophie. Langdon dubitava

Figure 9. Locate *buon* "good" in the Italian text.

<sup>&</sup>lt;sup>8</sup>Ringraziamenti (Dan Brown The Da Vinci Code, 2003) available at

https://archive.org/stream/IlCodiceDaVinciDanBrown/Il%20codice%20Da%20Vinci%20-%20Dan%20Brown\_djvu.txt

- Identify the inflectional features of some of the nouns following the previous adjective: e.g., *il buon nome* "the good name," *un buon amico* "a good friend," etc.
- Return to the Spanish Module Argentina, and perform Linguistic Analysis on the following sequence: *un hombre bueno* "a good man" (Figure 10).



Figure 10. Linguistic Analysis of an adjectival sequence

- Detect concordance features in the sequence: masculine (masc) and singular (sg).
- State which variations this adjective may adopt, since according to the corresponding lexical grammar, *bueno* "good" inflects like FLACO (FLACO = <E>/masc+sg | s/masc+pl | <B> a/fem+sg | <B> as/fem+pl;).
- Look up, in the dictionary of adjectives, another adjective inflecting like *desaparecido* and complete the following expressions:

```
Un hombre… . "A man…"
Unos hombres… . "Some men…"
Mi libro… . ."My book…"
La señora… . "The lady…"
```

- Check through Linguistic Analysis how it is possible, by virtue of concordance, to build noun phrases.
- Do an online search to complete the analysis with its historical implications of the term *desapare-cido* in expressions such *desaparecidos Argentina* and *disappeared Argentina*<sup>9</sup>.

<sup>&</sup>lt;sup>9</sup>The searches 'desaparecidos Argentina' and 'disappeared Argentina' are available at https://es.wikipedia.org/wiki/Desaparecidos\_durante\_el\_terrorismo\_de\_Estado\_en\_Argentina and https://en.wikipedia.org/wiki/Dirty\_War.