The Spanish Auxiliary Verb System in HPSG

Ivan Meza[1] and Luis Pineda [2]

Dept. Computer Science, IIMAS, UNAM
Cicuito interior s/n, Ciudad Universitaria, Coyoacán, 04510, México D.F.
[1]ivanvladimir@cic1.iimas.unam.mx
[2]luis@leibniz.iimas.unam.mx

Abstract. This paper presents an analysis of the Spanish auxiliary verbs from a syntactic point of view instead a semantic analysis that is proposed by the traditional Spanish grammars. The syntactic context of the Spanish auxiliary verbs is clarified with the definition of five properties, which allow us to determinate if a verb behave as an auxiliary verb or not. The subject raising verb type defined in the formalism HPSG is used in order to capture the behavior of Spanish auxiliary verbs. We conclude with the analysis of a typical auxiliary verb, *poder*, as a case of study.

1. Introduction

The definition and implementation of a robust grammar for Spanish requires a solid analysis of the Spansih auxiliary verb system. Although the verb system for English has been widely studied and implemented in computational grammars since Chomsky´s original analysis in Syntactic Structures, where the crucial role of the study of auxiliary verbs for the development of English grammar is pointed out [1], to our knowledge, no such rigorous computational analysis for the Spanish auxiliaries is available to this date. Even traditional Spanish grammars lack a definite analysis of the grammatical phenomena, as several criteria and classifications, not always consistent, can be found in the literature.

In particular, the need for a useful implementation of this system comes from the DIME project (Diálogos Multimodales Inteligentes en Español) currently developed at IIMAS, UNAM, which has as its main objective the construction of a conversational multimodal agent with spoken input and output facilities in Spanish for helping the interactive solution of simple design task [2]. In particular, the DIME prototype will be able to assist human-users in simple kitchen design tasks. One of the main objectives of this project is the development of a Spanish grammar and parser able to cope with the language of a corpus of Spanish dialogs in this domain that was compiled within the context of this project [3]. For the development of the grammar the HPSG grammatical formalism [4] with its associated environment development was adopted [5].

The need for a full analysis of auxiliary verbs emerged immediately when the first sentences of the DIME corpus were analyzed. Sentence (1) is a typical construction in the corpus:

(1) ¿Me puedes mostrar el catálogo de muebles?
Can you show me the catalog of furniture?

This sentence shows a number of syntactic phenomena that are characteristic of Spanish syntax, which are alien to English: omitted subject, clitic constructions and the "periphrastic conjugation". Also, the sentence is ambiguous because the verb *poder* can be interpreted as signaling ability or possibility: in the latter case it is an auxiliary verb but in the former it is not, as will be explained below in this paper. In addition, (1) is an interrogative form in which the subject-verb inversion of English questions does not take place; furthermore, the word-order for the declarative, interrogative and imperative forms of this kind of sentences can be the same. Each of these syntactic phenomena needs a detailed analysis; however, central to all of them is the analysis of the complex verbal construction and, in particular, of the auxiliary verb construction, which is the skeleton upon which all other phenomena are supported.

2. Spanish Auxiliary Verbs

Intuitively, an auxiliary verb in Spanish is a verb that has lost its original lexical meaning and has acquired a grammatical function or meaning in specific syntactic contexts. Consider the examples in (2) taken from Gili Gaya [6 pp. 108]:

(2) a. Tener que escribir
Have to write
b. Estar escribiendo
To be writing
c. Ir a escribir
Going to write

In all three, there is a reference to the action of writing, but the initial verbs marks a specific mode in which the writing action should be interpreted. In (2.a), *tener* in the context of *tener que* has changed its original meaning of possession to obligation, and *tener que escribir* refers to the obligation of writing. In (2.b), *estar* has lost its meaning of location and *estar escribiendo* alludes to the duration of the act of writing, and in (2.c), the verb *ir* has lost its meaning of physical transfer and marks the incoative notion of the act of writing, and the whole construction means to start to write.

The grammaticalization of the verbs is not a phenomenon particular to Spanish, as can be seen from the translations, but it does give rise to some kind of ambiguities that are very peculiar to Spanish. Auxiliaries in English are fully grammaticallized: when a verb becomes an auxiliary, it preserves this function most of the time; however, in Spanish this is not the case, as shown in (3). In (3.a), *debo* preserves its original meaning of obligation of paying a debt, but to express the same idea in English requires the use of the verb *owe* which is not an auxiliary. On the other hand, in the verbal complex *debo ir* in (3.b), *debo* is an auxiliary marking a general kind of obligation, and its expression in English requires the auxiliary verb *must*. In this regard, English is better behaved than Spanish.

(3) a. Debo cien pesos
I owe one hundred pesos
b. Debo ir al banco
I must go to the bank

Verbal complex constructions are known as periphrastic conjugation (*conjugación perifrásica*) in which the first verb is conjugated (*forma personal*) and the rest of the verbs in the complex are in a non-finite form (*formas no personales*) which are the infinitive, the gerund and the participle, as can be seen in Table 1.

a. Iba a decir *He/she was going to tell*	Infinitive
b. Estab comiendo *He/she/it was eating*	Gerund
c. He caminado *I have walked*	Participle

Table 1. Different non-finite forms taken by auxiliars.

In Spanish, all auxiliaries appear in periphrastic constructions but there are periphrastic constructions in which the conjugated verbs are not auxiliaries; for this reason it is not trivial to state when a verb has such a function. Even traditional grammars of Spanish have different criteria. Gili Gaya [6], for instance, adopts the strongest position and defines an auxiliary as a verb that has lost his meaning and has taken a new one, as illustrated in (3). However, most Spanish grammars adopt a less restricted position, and propose a hierarchy of auxiliaries: while *ser*, *estar* and *haber* are classified as full auxiliaries, all modal verbs function as semi-auxiliaries [7]. Nevertheless, the indecision prevails, and the full set of auxiliares is not well-defined.

To clarify the distinction, it can be noticed that changes of meaning are accompanied with changes in syntactic behavior (e.g., in (3.b) the verb *debo* is n longer a transitive). Accordingly, we adopt the view that to understand the behavior of auxiliaries not only semantic criteria but also syntactic properties must be taken into account.

2.1 Behavior of periphrastic constructions

For the study of periphrastic constructions we focus on the behavior of agents and patients. The agent is individual who executes the action named by the verb and the patient is the one how receives such an action; in general, the agent corresponds to syntactic subject and the patient to the direct object in active sentences. In Spanish periphrastic constructions agents and verbs are related through the following syntactic properties:

1. The conjugation of the auxiliary verb contains the syntactic subject, which is normally omitted; however, auxiliaries do not require agents, as the main semantic import of the periphrasis is marked by the verb in a non-finite form. For this reason, the syntactic subject marked by the conjugation helps to identify the agent of the non-finite verb. In (4), for instance, the verb *voy* (1st-sing) marking the incoative action helps to identify that the agent of *comer* is me (*yo*).
   (4) (Yo) voy a comer
   I am going to eat
2. It is possible to construct periphrasis with verbs that do not need an agent. These verbs are known as impersonal (*impersonales*) and they represent agentless actions like raining and snowing. In addition, impersonal constructions can also be formed dropping the agent, either because it is not known or just because it is not important. In (5.a), *va* (3rd-sing) marks the subject of the sentence, but *llover*

requires no agent, and the information provided by the conjugation is simply not used. In (5.b), *van* (3rd-pl) marks that someone (perhaps more than one) are going to knock the door, but we don't know how is it. Again, the information provided by the subject is not used, because the agent needs not to be determined.

(5) a. Va a llover
It is going to rain
b. Van a tocar la puerta
Someone is going to knock the door

3. It is not possible construct an interrogative sentence asking for the direct object of the periphrases using the auxiliary verb only. For instance, in *Voy a estudiar matemáticas* the direct object of whole periphrasis (i.e., *matemáticas*) is an argument of *estudiar*; this is the case because the auxiliary verb *voy a* has no semantic import, but only that the action will take place in the future. Question (5.b) ask for the direct object of *estudiar* and is well-formed, but (5.a) is not well-formed because it asks for the direct object of the auxiliary, which does not exist.

(6) a. *¿Qué (tú) vas a?
**What are you going to?*
b. ¿Qué (tú) vas a estudiar?
What are you going to study?

4. It is not possible construct an interrogative sentence asking for the action named by the non-finite verb using the auxiliary verb only; to form this kind of questions a wildcard verb is required. The action of the sentence *Voy a estudiar* can be inquired with question (6.b) where the wild-card verb *hacer* substitutes the action *estudiar*; however, question (6.a) is not well formed.

(7) a. *¿Qué vas a?
**What are you going to?*
b. ¿Qué vas a hacer?
**What are you going to do?*

5. It is not possible construct periphrasis with the auxiliary in passive voice; if the periphrasis in *voy a entregar la carta* is expressed as a passive construction, the participle needs to be non-finite verb.

(8) *La carta es ida a entregar por mi
**The letter is gone deliver by me*
(9) La carta va a ser entregada por mi
The letter is going to be delivered by me

These five properties define the auxiliary verbs in Spanish. This analysis contrasts with most traditional accounts of Spanish grammar where the sole presence of a periphrastic construction is normally taken to signal auxiliary verbs. As there are periphrastic construction in which these five properties do not hold, our analysis permits to distinguish finite verbs occurring in periphrastic construction which do function as auxiliaries, from verbs also occurring in this kind of construction which, nevertheless, are not auxiliaries. Sentences (10) and (11) are examples of this latter phenomenon.

(10) Quiero comer una manzana
I want to eat an apple
(11) Tengo estudiada la materia
I have studied the subject

Gili Gaya classifies auxiliaries according to the non-finite verb occurring in the periphrasis as follows [6 pp. 100]:

- With infinitives: *ir a, pasar a, echar a, venir a, volver a, haber de, haber que, tener que, deber de, llegar a, acabar de* and *alcanzar a*;
- With gerunds: *estar, ir, venir, seguir* and *andar*;
- With participles: *llevar, tener, traer, quedar, dejar, esta* and *ser.*

The auxiliary verb *haber* used in all tenses of the Spanish conjugation is taken as fix desinence and its given an independent treatment.

We have tested all five properties in the set and only sixteen passed the test. In particular *echar a, pasar a* and *haber que*, and also all five that are followed by participle, do not count as auxiliaries in our criteria. On the other hand, the verb *haber* in the conjugation of composite tenses do counts as a normal auxiliary verb.

As all five properties also occur in English, our analysis also shows that despite traditional perception, the auxiliary verbal systems of both of these languages are quite similar. However, English auxiliaries have, in addition, the so-called NICE properties (negation, inversion, contraction and ellipsis) [4 pp. 302] making their behavior more systematic than the corresponding Spanish constructions. On the other hand, unlike auxiliary verbs in English, which seem to be fully grammaticalized, the Spanish verbs that function as auxiliaries, can preserve its original function, even in periphrastic constructions, producing a number of ambiguities that need to be addressed. Next, we turn to the formal analysis of periphrasis in Spanish both when auxiliaries are involved, and also when they are not.

3. Auxiliary verbs in HPSG

In HPSG all lexemes are related in a lattice of types [4]. In particular, verb-lexemes (*verb-lxm*) have the so-called subject raising verb (*srv-lxm*) and subject control verbs (*scv-lxm*) as subtypes. This distinction can be found original in Chomsky's Extended Transformational Grammar [4 pp. 280], and it is widely used in HPSG and related grammatical formalisms. Auxiliary verbs (*auxv-lxm*) are also subtypes of the type *srv-lx* . In particular, objects of type *auxv-lxm* are objects of type *srv-lxm* that also have the NICE properties. Here, we will claim that auxiliary verbs in Spanish are *srv-lx* . But in addition, we will claim that the same verbs can also occur as *scv-lx* and even as transitive or intransitive verbs. In particular, the verb *poder* occurs at least in three different types, with different syntactic properties, and different semantic import.

3.1 Subject-Raising and subject-control verbs

The type *srv-lxm* is defined in HPSG as a attribute-value matrix (AVM) in (12), where the symbol *srv-lxm* is the type identifier, and the argument structure (ARG-ST) has two arguments, where the first argument must also be the specifier of the second. The principal characteristic of this AVM is that nothing is specified about the agent of sentences in which the head of a verbal phrase is of this type (*srv-lxm*).

(12)

$$\begin{bmatrix} srv-lxm \\ ARG-ST & \left\langle \#1, \begin{bmatrix} phrase \\ SPR & \langle \#1 \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

In the phrase *poder mostrar*, for instance, *poder* is a verb of type *srv-lxm* and it second argument has *mostrar* as its head. Consequently, according to (12), the first argument of *poder*, the subject of the sentence, is also the specifier (the first argument) of *mostrar*. As there are no further constraints in this specification, the subject of the sentence is the agent of the action denoted by *mostrar* (the one who is doing the showing), regardless whether the agent is mentioned explictly, or it is absent, as it is the case in impersonal constructions. Here, the verb *poder* is an auxiliary verb marking the possibility of the act of showing.

Next, we consider the subject-control verbs. The type *srv-lxm* is defined with the following AVM:

(13)

$$\begin{bmatrix} \textit{scv–lxm} & \\ \textit{ARG–ST} & \left\langle NP_i, \begin{bmatrix} \textit{phrase} & \\ \textit{FORM} & \text{inf} \\ \textit{SPR} & \langle NP_i \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

This structure is similar to the definition of *srv-lx* , but it forces a restriction in the type of the specifier. If *poder*, in *poder mostrar*, is of this latter type, its first argument is marked as a noun-phrase that is also the first argument of *mostrar*. In this case, the subject of *poder* becomes the agent of *mostrar*, and this agent must necessarily be there. The meaning of *poder* in this interpretation is that an agent has the capability of showing or is able to show something. English prefers the form *be able* for this function.

3.2 Definition of auxiliary verbs in Spanish

Unlike English where auxiliary verbs have the NICE properties, and the type *auxv-lxm* can be defined as a subtype of *srv-lx* , in Spanish there no such distinction and we identify the type of auxiliaries with the *srv-lxm* type, as follows:

Definition (1): auxiliary verbs in Spanish are of type *srv-lxm*.

4. *Poder*

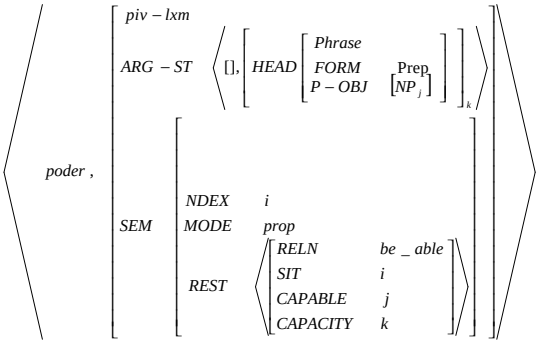
Next, we present an analysis of the verb *poder* in the three different contexts as shown in (14.a) to (14.c), as follows:

(14) a. Puedes con las matemáticas
You are capable in mathematics
b. Puedes mostrar el catálogo
You are able to show the catalog
c. Puedes mostrar el catálogo
It is possible that you show the catalog

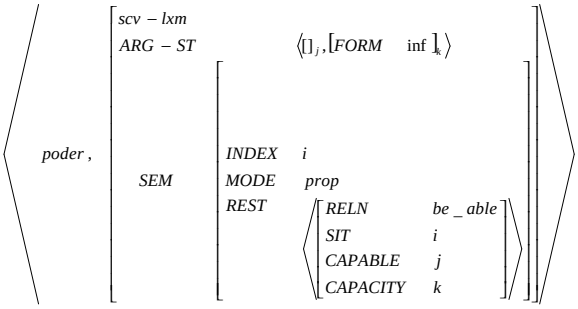
In the sentence (14.a) the verb *poder* means ability; this is perhaps its original lexical meaning. In this context, *poder* is a prepositional intransitive verb (*piv-lxm*).

In (14.b) it has the meaning of capability, and in (14.c) functions as an auxiliary and signals the possibility of showing. Next, we show the corresponding lexical entries in HPSG in (15).

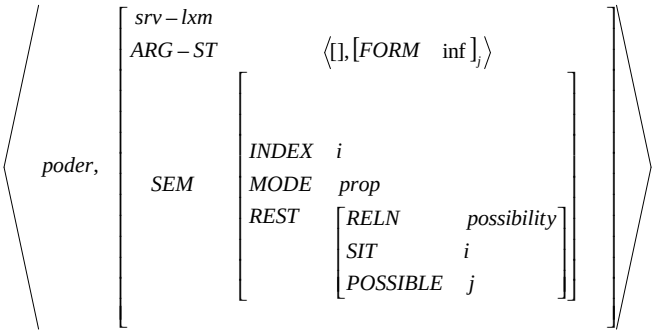
(15) a. *poder* as a *piv-lxm*



b. *Poder* as a *scv-lxm*

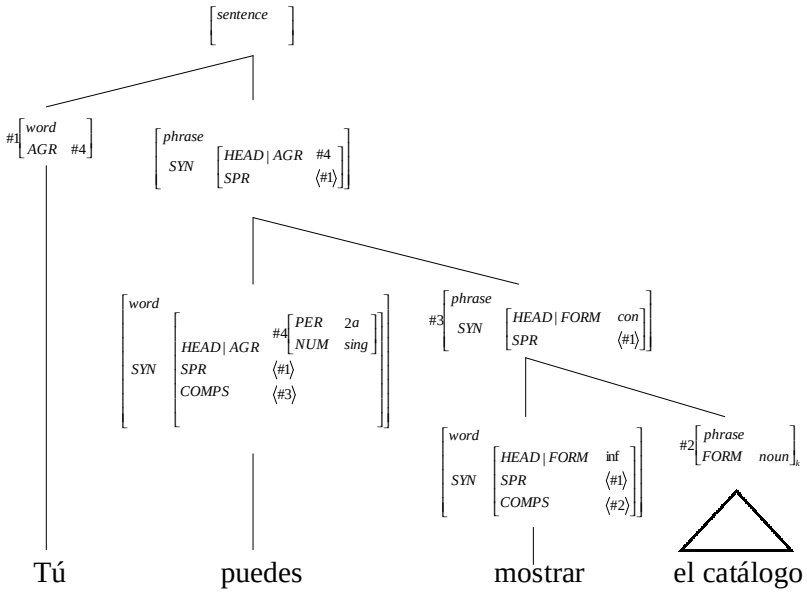


c. *Poder* as a *srv-lxm*



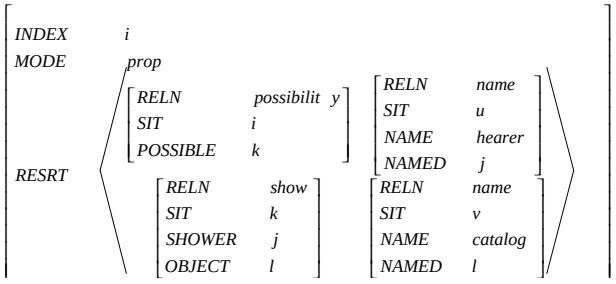
As can be seen in the argument structure of (15.a) and (15.b), the difference between these lexical entries is that while the intransitive verb takes a prepositional phrase as a complement, the subject control verb takes a non-finite form as its complement. On the other hand, the semantics in both of these entries is the same, and states the relation between an individual and something that he or she is capable of. (15.c) differs in the syntax as was explained above. The semantics is also different as the relation marked is one of possibility, and the agent needs not to be present. Next, we show the syntactic structure of (15.c).

(16)



This example shows how the first structure that is formed is the verbal phrase *mostrar el catalogo*, when the verb *mostrar* takes his complements in the application of the Head Complement Rule. After, this structure unifies with the first argument of the COMPS of the verb *poder*, the Head Complement Rule is applied again and the verbal phras *puedes mostrar el catálogo* is obtained. Finally, the SPR of *mostrar* unifies with the SPR of the verb *poder*; this last phrase can take his subject *tú*(2nd-sing), which helps to define the agent of the second verb. The syntactic structure of (14.a) and (14.b) is obtained by a similar process. The final semantics for the sentences in (14.c) are:

(17) *Poder* meaning possibility:



This semantic refers to the situation *i* that signals the possibility of an another situation; here the possible situation is mark by index *k* which corresponds to the relation of showing, which needs two arguments: one who shows that is the agent of showing act and the object showed; the shower corresponds to the hearer because the syntactic subject of the sentence was *tú* (2nd-sing), and the object is the catalog. Notice that the value that corresponds to the agent is only used by the relation of show and it does not apear in the relation of possibility.

Next, we show the final semantic of *poder* for (14.a) and (14.c):

(18) a. *Poder* meaning ability

$$\begin{bmatrix} INDEX & i \\ MODE & prop \\ RESRT & \left\langle \begin{bmatrix} RELN & be_able \\ SIT & i \\ CAPABLE & j \\ CAPACITY & k \end{bmatrix} \begin{bmatrix} RELN & name \\ SIT & u \\ NAME & math \\ NAMED & k \end{bmatrix} \begin{bmatrix} RELN & name \\ SIT & v \\ NAME & listener \\ NAMED & j \end{bmatrix} \right\rangle \end{bmatrix}$$

b. *Poder* meaning capability

$$\begin{bmatrix} INDEX & i \\ MODE & prop \\ RESRT & \left\langle \begin{bmatrix} RELN & be_able \\ SIT & i \\ CAPABLE & j \\ CAPACITY & k \end{bmatrix} \begin{bmatrix} RELN & name \\ SIT & u \\ NAME & hearer \\ NAMED & j \end{bmatrix} \begin{bmatrix} RELN & show \\ SIT & k \\ SHOWER & j \\ OBJECT & l \end{bmatrix} \begin{bmatrix} RELN & name \\ SIT & v \\ NAME & catalog \\ NAMED & l \end{bmatrix} \right\rangle \end{bmatrix}$$

Structure (18.a) and (18.b) are similar as the firts argument in their restriction list i the same, but while (18.a) states that the agent *j* has the capability *k* (mathematics), (18.b) states that the agent who is able is also the agent of the showing action. In adition, (18.a) has a restiction which names the patient of the showing action (i.e., the catalog). On the other hand, (18.b) is also similar to (17), as both of them have the same relation in their corresponding restriction list, but while (18.b) has an agent for the capability (the index *j*), the restiction of posibility has not agent; it just marks a situation in which it is possible that the hearer performs a showing action.

5. Conclusions

In this paper a computational analysis of Spanish auxiliary verbs has been presented. The analysis was centered on the syntactic properties of auxiliaries, in opposition to traditional analysis that are mainly semantically oriented. Five syntactic properties of

auxiliaries were identified which allowed us to separate the notions of periphrastic conjugation and auxiliary verbs that are usually taken together in traditional analysis. In particular, auxiliaries in Spanish were identified as subject raising verbs. Our analysis allow us also to identify a number of ambiguities arise as a consequence of the use of the same verbs in different context, whether they function as auxiliaries or not in such context. We are able to identify a new set of auxiliaries that can be formally distinguished, and is also intuitively appropriated. In particular, a set of sixteen verbs that function as auxiliaries has been identified, as follows: *poder, ir a, venir a, volver a, haber de, tener que, deber de, llegar a, acabar de* and *alcanzar a* which take an infinitive complement; *estar, ir, venir, seguir* and *andar* which take a gerund as a complement, and *haber* which accepts past participles as complements. This verbs have been implemented in the LKB develop environment for HPSG, and they will be used in the interpretation of natural languages Spanish sentences on the prototype of the DIME project.

Acknowledgments

Discussions and feedback were very much appreciated, specially from the following people: Erick Schwarz, Antonio Massé, Esmeralda Uraga, Miguel Salas, Luis Villaseñor and Celina López, and also from James Allen, Eduard Hovy and the NL-group at the University of Rochester. The authors also gratefully acknowledge the support of Conacyt's grant 499316-5-C092A

References

1. Chomsky, Noam : Syntactic Structures. Mouton & Co. 'S- Gravenhage. (1962)
2. Pineda, L. A., Massé J. A., Uraga E., y Villaseñor, L., Salas, M., Schwarz, E. y Meza, I.: El Proyecto DIME. In Procedings of Second International Workshop on Spanish Language Processsing and Language Technologies. Jaén, Spain. pages 41-45 (2001)
3. Villaseñor, L., Massé J. A., Pineda, L. A.: The DIME Corpus. In Memorias del *3er. Encuentro Internacional de Ciencias de la Computacion.* Aguascalientes, México (2001) 591-600
4. Sag, I. y Wasow, T.: Syntactic Theory A Formal Introduction. CSLI Publications. Stanford, California (1999)
5. Copestake, A.: The LKB System. Technical report. Stanford University (2001) http://www-csli.stanford.edu/~acc/lkb.html
6. Gili Gaya, S.: Curso superior de sintaxis española. Decimoquinta edición. Vox. Barcelona. (1991) 102-119
7. Alonso, A. y Henríquez, P.: Gramática castellana, segundo curso. 22ª. Edición. Buenos Aires. (1967)