Procesamiento del Lenguaje Natural, núm. 35 (2005), pp. 253-257 recibido 30-04-2005; aceptado 01-06-2005

Syntax-driven bindings of Spanish clitic pronouns

Ivan V. Meza Ruiz
ICCS
University of Edinburgh
I.V.Meza-Ruiz@sms.ed.ac.uk

Luis A. Pineda Cortes
IIMAS
Universidad Nacional Autónoma de México
luis@leibniz.iimas.unam.mx

Resumen: Este trabajo presenta el análisis del sistema de pronombres clíticos del español y su relación con la teoría de *binding*. Los pronombres clíticos del español presentan un comportamiento dual dado que pueden ser unidades morfológicas o unidades léxicas independientes. Como unidades morfológicas los pronombres son sufijos, mientras que como unidades léxicas independientes son clíticos. En el segundo caso los clíticos son pronombres por lo que es necesario que las estructuras en las que aparecen cumplan con los principios de *binding*. En particular, presentamos los análisis de reflexivos, clíticos pleonásticos y verbos de control de objeto.
Palabras clave: Pronombres clíticos, Teoría de *Binding*, Reflexividad, HPSG

Abstract: In this paper we present an analysis of the relationship between the Spanish clitic pronoun system and the Binding Theory. Spanish clitic pronouns have a dual behaviour; in the proclitic case, they behave like independent lexical units, hence proper clitic, but on the enclitic case they are hence morphological affixes (i.e. inflexions). In the first case, the clitics stand for pronouns, therefore it is necessary that the sentences in which they intervene follow the Binding Principles. We present the analysis for reflexives, double clitic and object control verbs constructions.
Keywords: Clitic pronouns, Binding Theory, Reflexivity, HPSG

1. Introduction

Binding Theory focuses on the coreference relationships between pronouns and their antecedents. It is typically stated in terms of three key aspects: the binding domain, the class of nominals, and a set of syntactic conditions. The binding domain restricts the scope of the possible bindings. The class of nominals defines the syntactic class of words which can participate in binding relations, these are typically pronouns. The syntactic conditions constrain whether a binding relation holds in a sentence. In particular, Head-driven Phrase Structure Grammar (HPSG) defines the binding domains in terms of the $SUBCAT$ attribute (also known as ARG-ST). This attribute specifies the valence of a word, but it also corresponds to a binding domain. A coreference between elements of the same domain is said to be *local* (example 1.a), while one between an element of different domains is *non-local* (example 1.b).

(1) a. Juan dice que [Pedro$_i$ se$_i$ mintió]
Juan saids that Pedro$_i$ lied to himself$_i$

b. Juan$_i$ dice que [Pedro le$_i$ mintió]
Juan$_i$ saids that Pedro lied to him$_i$

As can be seen from the examples the class of nominal which is discussed in this work corresponds to the clitic pronouns of Spanish. A detailed description of them is presented on Section 2. HPSG distinguishes between the anaphoric nominals (i.e $ANA+$) which make a reference to an antecedent, for instance reflexives and reciprocals, and to nominals which do not make it necessarily (i.e $ANA-$), for instance non pronouns and pronominals. Spanish clitics pronouns can be both, anaphoric or not.

Additionally, HPSG states the Binding Principles to constrain the bindings within the sentence as follows [Pollard and Sag, 1994, Everaert, 2001]:

(2) a. A locally *o-commanded* anaphora (i.e. $ANA+$)) must be locally $o-$*bound*.

b. A personal pronoun (i.e. $ANA-$) must be locally *o-free*.

c. A non-pronoun must be *o-free*.

The principles are stated in terms of the relation *o-bind* and *o-free*, where *o* stands for obliqueness, which are defined as:

(3) a. Y (locally) *o-binds* Z just in case Y and Z are coindexed and Y (locally) *o-commands* Z.

b. Z is (locally) *o-free* if Z is not (locally) *o-bound*.

c. Let Y and Z be *synsem* objects with distinct $LOCAL$ values, Y referential. Then

ISSN: 1135-5948

(i) *Y locally o-commands Z* just in case Y is less oblique than Z.

(ii) *Y o-commands Z* just in case Y locally commands X and X dominates Z.

Finally, HPSG defines *obliqueness* in the following way:

(4) (a) A *synsem* object Y is less oblique than a *synsem* object Z just in case it precedes Z on the $SUBCAT$ list of some lexical head.

(b) The ordering of *synsem* objects is: subj >obj >sec.obj >obj$_p p$ >verb/pred.comp

This work presents the analysis of the clitic pronouns of Spanish on the light of the Binding Principle. In the next Section, we present the HPSG model of the Spanish clitic pronouns in which this work relies. The following sections analyse the phenomena where anaphoric relations are present and the principles to be satisfied. Section 3 presents the case of reflexives, Section 4 the case of double complements, and Section 5 the case of object control verbs.

2. The Spanish Clitic Pronoun Model

Our analysis is based on the work presented on [Pineda and Meza-Ruiz, 2005]. In our view enclitic pronouns are affixes (e.g. *da-lo*/give it), but proclitics representing expected accusative or dative arguments of a verb are proper clitics (e.g. *lo das*/you give it). If the proclitic augments the basic argument structure of the verb we consider this a lexical idiosyncrasy, and the pronoun is considered an inflexion (e.g. *te vas*/you go).

The previous classification is relevant to the binding of Spanish pronouns. Affixes are not independent lexical units; this is, they are not syntactic units but rather morphological. Therefore, the affix cases of the clitic pronouns of Spanish does not follow the Binding Principles; on the other hand, clitics, particularly the clitic word we introduce below, stand as a pronoun or a set of pronouns. In this case, binding should hold. This differentiate our model from others ones (e.g. [Miller and Sag, 1997, Monachesi, 1999, Schwarze, 2001]) where clitic pronouns are defined as affixes and therefore the binding analysis is not required.

The model is based on the concepts of clitic domain, phonological host, and a set of operations: clitic cancellation, clitic composition, and clitic subsumption. The clitic domain consists of the arguments of the verbs which are realised as clitic pronouns. In the HPSG formulation of this theory, the clitic domain is represented by the $CL-LIST$ attribute which is part of the valence attribute. The phonological host is the verb that has the pronouns attached as affixes in the enclitic case, and the verb that is preceded by the pronouns in the proclitic case. The operation of cancellation deals with the satisfaction of the clitic domain by the clitic host. The operation has two realisations: a set of morphological rules for the inflectional (affixed) case, and the Head Pro-clitic Rule for the proclitic case. In the first case, the morphological rules modify the valence of the clitic taking a complement or complements of the verb and realising them as an affix [Monachesi, 1999, Schwarze, 2001]. This accounts for the following examples:

(5) a. Dáselo
Give-him/her-it
Give it to him/her

b. Puedes dárselo
You-can give-him/her-it
You can give it to him/her

On both cases, the $SUBCAT$ attribute for *dar*/give is: $< NP >$ and the binding principle can not been applied. There is no *obliqueness* relationships.

Furthermore, the Head Pro-clitic Rule (APR) satisfies the clitic domain of a clitic host with a clitic word. A clitic word is a sequence of clitics which represent a sequence of arguments of the verb. For instance *me lo* represents a sequence of two noun phrases arguments where the first one is accusative first person, and the second a masculine dative third person. It is important to point out that the sequence of clitics on the orthography of a clitic word does not necessarily corresponds to the sequence of arguments which they represent. This situation accounts for the strict order of clitics words without stating a complex mechanism to deal with it (a similar account is proposed by Monachesi [1999]). This allows us to deal with sentences as:

(6) a. Me lo das
Me it you-give
You give it to me

b. Me lo puedes dar
Me it you-can give
You can give it to me

c. ¿Lo tomas o dejas?
It you-take or you-leave?
Do you take it or leave it?

Composition is an operation which creates a clitic domain with arguments of two different verbs. This operation accounts for phenomena which involves object control verbs (*ocv*). These verbs have two complements: a dative argument and a verbal phrase. A property of these verbs is that the dative argument is also the subject of the complement verbal phrase (e.g. *oi a Juan decir la noticia*/I hear to Juan say the news). These verbs can have a composed domain which involves its dative argument and its verbal phrase complement. For instance:

(7) Se lo oí decir
Her/him it I-hear say
$< NP > < NP >$
I hear her/him to say it

As can be seen, object control verbs in Spanish compose their clitic domain. The result of the composition on these verbs is a clitic domain which encapsulates its dative argument and the clitic domain of their verbal argument (example 7). After the composition the clitic domain of *oir*/hear is $< NP, NP >$ which is satisfied by the clitic word *se lo*.

Finally, there is a case where two arguments of a verb are fused in one clitic. This case is handled by the subsumption operation. On the example 8.a, the clitic domain of *ver*/see encapsulates its nominal argument (i.e. him/her) and the pseudo-reflexive (i.e. for him/her own sake) as one clitic (i.e. *se*). On example 8.b, it can be appreciated the case where this does not happen and the arguments are realised separately.

(8) a. Se_i lo vi comer
Him/her it I-see eat
I saw $\text{him}_i/\text{her}_i$ to eat it for $\text{him}_i/\text{her}_i$ own sake

b. Le_i vi comérselo
Him/her I-see eat-her/him-it
I saw $\text{him}_i/\text{her}_i$ to eat it for $\text{him}_i/\text{her}_i$ own sake

The previous concepts are put together in the clitic principle:

> Spanish pronominal clitic sentences can be accounted for in terms of the operations of clitic cancellation, clitic composition and clitic subsumption; alternatively, a clitic host, simple of composite, must be within the scope of its phonological host.

3. Reflexives

The reflexivity phenomena in Spanish uses exhaustively the clitic pronouns. Most of the cases, these are preferred over the non-clitic structures (e.g. *me lavo*/I wash myself vs **lavo a mí*). This phenomena is captured by the model of clitic pronouns defined in the previous section. In the affix case, a lexical rule realises the complements of the verb as affixes. In the case of proclitics, a lexical rule cliticizes the verb (9). The rule constrains the realisation of the arguments; however it does not modify the argument structure of the $SUBCAT$ attribute. This situation makes compatible the approach with the Binding Theory as stated in Pollard and Sag [1994][1].

(9) $\begin{bmatrix} word \\ \text{HEAD} & \text{verb} \\ \text{SUBCAT} & \langle \boxed{1} \rangle \oplus \boxed{a} \end{bmatrix} \mapsto \begin{bmatrix} \text{VAL} \begin{bmatrix} \text{COMPS} & \langle\rangle \\ \text{CL-LIST} & \boxed{a} \end{bmatrix} \end{bmatrix}$

This accounts for the following examples[2]:

(10) a. Te_i lavas_i
You_i you-wash_i
You_i wash yourself_i

b. Te_i puedes_i lavar
You_i you-can_i wash
You_i can wash yourself_i

Both of the cases have the following $SUBCAT$ attribute for the verb *lavar*/wash:

$$< NP[ANA-]_i, NP[ANA+]_i >$$

As can be appreciated the structure follows the Binding Principles. The reflexive clitic is *o-bound* by the subject. In 10.b, the fact that *puedes* and *lavar* share the clitic domain does not affect the binding, the $SUBCAT$ attribute for *lavar*/wash remains the same. In case the clitic pronoun is not the reflexive case (i.e. the clitic is ANA -), the principle B of binding theory establishes that that it should be locally *o-free*. This account for the examples:

(11) a. *Juan_i le_i lava
Juan_i him_i lava
*Juan_i wash him_i

b. Juan_i le_j lava
Juan_i him_j lava
Juan_i wash him_j

[1] The $SUBCAT$ attribute is a list of the arguments of the lexical unit. In case of verbs, the first argument usually represents the subject of the action (e.i. $SUBJ$), while the rest the complements (e.i. $COMPS$). Particular to Spanish, the rest of arguments also can be realised as clitics (e.i. $CL-LIST$). The Argument Realisation Principle constrains the possible combination of these arguments into the $SUBJ$, $COMPS$, and $CL-LIST$ lists.

[2] The subindex i on *lavar*/wash or *puedes*/can verbs refers to its subject which is implicit to the conjugation.

4. Doubled complements

A property of the Spanish clitic pronouns is that one of the clitic in the clitic word can appear in the sentence along with the dative complement it represents. This phenomena is called double complement. Consider the following examples:

(12) a. Me$_i$ da el lápiz a mí$_i$
Me$_i$ gives the pencil to me$_i$
he gives to me$_i$ the pencil

b. *Lo$_i$ entrego el carro a Juan$_i$
Him$_i$ give the car to Juan$_i$

These examples do not involve a reflexive clitic pronouns (i.e. $ANA-$); this implies that the binding should be on different domains to satisfy principle B. In order to capture this we propose the following general lexical rule:

(13) $$\begin{bmatrix} word \\ \text{HEAD} \quad \text{verb} \\ \text{SUBCAT} \quad \boxed{a} \oplus \langle \boxed{1} \rangle \end{bmatrix} \mapsto \begin{bmatrix} \text{VAL} \begin{bmatrix} \text{COMPS} & \langle \ldots \rangle \oplus \langle \boxed{1} \rangle \\ \text{CL-LIST} & \langle \boxed{1} \rangle \end{bmatrix} \end{bmatrix}$$

This rule constrains the new structure to realise its last argument as a clitic and complement. The argument satisfaction takes care of this specific combination. Additionally, it can be seen that the SUBCAT attribute remains unmodified, but not the clitic domain. As the locally domain captured by SUBCAT attribute has not access to the clitic domain, the binding on $SUBCAT$ is locally *o-free*. This can be appreciated on the following $SUBCAT$ and clitic domain lists which correspond to the previous examples.

(14) a. $$\begin{bmatrix} \text{VAL} \mid \text{CL-LIST} & \left\langle \text{NP}\left[\text{ANA -}\right]_i \right\rangle \\ \text{SUBCAT} & \left\langle \text{NP,NP,NP}\left[\text{ANA -}\right]_i \right\rangle \end{bmatrix}$$

b. $$^{*}\begin{bmatrix} \text{VAL} \mid \text{CL-LIST} & \left\langle \text{NP}\left[\text{CASE dat}\right]_i \right\rangle \\ \text{SUBCAT} & \left\langle \text{NP,NP,NP}\left[\text{CASE acc}\right]_i \right\rangle \end{bmatrix}$$

Finally, we review the double complement with reflexives clitics. For these cases we propose the following lexical rule:

(15) $$\begin{bmatrix} word \\ \text{HEAD} \quad \text{verb} \\ \text{SUBCAT} \quad \langle \boxed{a} \oplus \boxed{1} \rangle \end{bmatrix} \mapsto \begin{bmatrix} \text{HEAD verb} \\ \text{VAL} \mid \text{CL-LIST} \left\langle \boxed{2} \text{ NP } \left[\text{ANA +}\right] \right\rangle \\ \text{SUBCAT} \left\langle \boxed{a} \oplus \boxed{1} \oplus \boxed{2} \right\rangle \end{bmatrix}$$

In this case, the SUBCAT attribute is modified by adding the reflexive clitic pronoun (i.e. [ANA +]). This pronoun is *o-commanded* by the subject of the sentence and the doubled pronoun, therefore it can be bound by both of them. This rule accounts for the following example:

(16) me$_i$ entregó$_i$ el carro a mí$_i$
me$_i$ I$_i$-give the car to myself$_i$
I$_i$ give the car to myself$_i$

The corresponding $SUBCAT$ attribute has the value:

$$< NP_i, NP, NP[ANA-]_i, NP[ANA+]_i >$$

This should not be understood as one binding involving three elements, but rather two bindings, one between the reflexive pronoun and the subject, the other one between the reflexive and the non-reflexive pronouns.

5. Object control subsumption

Clitic composition allows to join two different clitic domains in one. This operation takes place on subject control and raising verbs, and object control verbs. As the subject control and raising verbs have an empty clitic domain the result of the composition is the clitic domain of its complement which makes the binding similar to the cases previously analysed. However, the case of object control verb requires a review. In this case, the clitic domain of the object control verb can be not empty. However, this situation does not modify the Binding principles, since the $SUBCAT$ attribute is not modified. For instance:

(17) a. Te$_i$ lo oigo decir$_i$
You$_i$ it I-listen say$_i$
$< NP > < NP >$
I listen you to say it

b. Me$_i$ lo oigo$_i$ decir$_i$
Myself$_i$ it I$_i$-listen say$_i$
$< NP > < NP >$
I listen myself to say it

have the following $SUBCAT$ attributes:

(18) a. $\left\langle \text{NP,NP}\left[\text{ANA -}\right]_i\text{,VP} \right\rangle$ and
$\left\langle \text{NP}_i\text{,NP}\left[\text{ANA -}\right] \right\rangle$

b. $\left\langle \text{NP}_i\text{,NP}\left[\text{ANA +}\right]_i\text{,VP} \right\rangle$ and
$\left\langle \text{NP}_i\text{,NP}\left[\text{ANA -}\right] \right\rangle$

In both cases, the coreferred index between the domains is the result of the behaviour of the subject control verb. This coreference is not the result of the Binding Principles. However, in the second example the reflexive

pronoun is *o-bound* to the subject. In both cases, composition takes place and as a result the clitic domain $< NP, NP >$ is created for *oir*/listen verb.

Clitic subsumption allows the fusion of two clitics from a different clitic domain. In particular, clitic subsumption takes places with object control verbs and pseudo reflexives. For instance:

(19) a. Se_i la ve_i comer_i
Herself_i it she_i-see eat_i
$< NP > < NP, NP >$
She saw herself eat it for her own sake

b. *Se_i la veo comer_i
Her_i it I-see eat_i
$< NP > < NP, NP >$

c. Se_i la veo_j comer_i
Her_i it I_j-see eat_i
$< NP > < NP >$
I saw her eat it

have the $SUBCAT$ following attributes:

(20) a. $\left\langle \text{NP}_i,\text{NP}\left[\text{ANA} \quad +\right]_i,\text{VP} \right\rangle$ and
$\left\langle \text{NP}_i,\text{NP},\text{NP}\left[\text{ANA} \quad +\right]_i \right\rangle$

b. $^{*}\left\langle \text{NP},\text{NP}\left[\text{ANA} \quad -\right],\text{VP} \right\rangle$ and
$\left\langle \text{NP}_i,\text{NP},\text{NP}\left[\text{ANA} \quad +\right]_i \right\rangle$

c. $\left\langle \text{NP}_i,\text{NP}\left[\text{ANA} \quad +\right]_i,\text{VP} \right\rangle$ and
$\left\langle \text{NP}_i,\text{NP} \right\rangle$

Example 19.a shows the case when subsumption takes place. This operation requires to share the clitic realisation (i.e. *se*) for the second argument of *ver/ve* and the third argument of *comer*. These arguments are realised as clitics, however in this case the subsumption creates the clitic domain $< NP, NP >$ where the second NP fusions two NPs. As can be seen on both local domains the Binding principles are satisfied. The coreference between argument of *ver* and subject of *comer* is the result of object control verb behaviour. Example 19.b shows the case when there is not reflexivity on the object control verb. As can been seen, this is not possible because subsumption implies both pronouns should be reflexive (i.e. $ANA+$). The correct reading for this sentence is presented on example 19.c which involves composition and not pseudo-reflexives.

6. Conclusions

We have presented the analysis of bindings on constructions with clitics pronouns. In our model proclitic pronouns which represent arguments of the verb are proper clitics. This situation implies they should follow the Binding Principles. In particular, we have analysed the reflexive, the double clitic, and the object control verb constructions. In all these cases, the $SUBCAT$ attribute satisfies the constraints of Binding Principles while the verbal arguments were satisfied the Clitic Principle. In the affix case of the clitic pronouns the coreferences are established through the morphology therefore they do not follow Binding Principles. This shows that the current state of our model interacts parsimoniously with Binding Principles as stated in HPSG and not additional consideration has to be done. An implementation of the model is available in the LKB system [Copestake, 2002] which deals with the Spanish clitic phenomena.

Referencias

Ann Copestake. *Implementing Typed Feature Structure Grammars*. CSLI Publications, 2002.

Martin Everaert. Binding theories: A comparison of grammatical models. In M. Oostendorp and E. Anagnostopoulou, editors, *Progress in grammar: Comparison of Grammatical Models Group: Roquade*, 2001. URL `http://www.roquade.nl/meertens/progressingrammar/toc.html`.

Phillip Miller and Ivan A. Sag. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory*, pages 573–639, 1997.

Paola Monachesi. A lexical analysis of italian clitics. In *Proceedings of Vertex*, pages 57–65, Venezia, 1999.

Luis A. Pineda and Ivan V. Meza-Ruiz. The spanish pronominal clitic system. *Procesamiento del Lenguaje Natural*, 34:34–71, 2005. to appear.

Carl Pollard and Ivan A. Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press, 1994.

Christoph Schwarze. On the representation of french and italian clitics. In Miriam Butt and Tracy Holloway King, editors, *The Proceedings of the LFG'01 Conference*, University of Hong Kong, 2001. CSLI Publications.