The Spanish pronominal clitic system

Luis Pineda and Ivan Meza
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS)
Universidad Nacional Autónoma de México (UNAM)
{luis,ivanvladimir}@leibniz.iimas.unam.mx

Abstract

In this paper we present a model of the Spanish pronominal clitic system. We start with a review of the phenomenon, including simple and complex clitic climbing and the related phenomena of reflexives, and also the impersonal passive-reflexive and active-impersonal forms. Then, we review the status of pronominal clitics and propose that it is a dual phenomenon: on the one hand, enclitics are inflections while proclitics are proper clitics that are realized as independent lexical units. Then a model of the clitic system and its formalization in HPSG is presented. The main result is the statement of the Spanish pronominal clitic principle: Spanish clitic sentences have a single phonological host that has within its scope a single concrete or abstract clitic host per independent clitic domain; abstract clitic hosts are produced by the operations of clitic composition and clitic subsumption, that are also introduced in this paper. We show how this principle accounts for the structure of Spanish pronominal clitic sentences and related phenomena. We also present the implementation of the theory in LKB.

Key words: pronominal clitics, clitic climbing, clitic composition, clitic subsumption, Spanish periphrasis, Spanish reflexives, passive-reflexive, active-impersonal. The Spanish *se*.

1. Clitic pronouns in Spanish

In Spanish, as in many other languages, there is a system of unstressed pronouns that are commonly used in both spoken and written language: the clitic pronoun system. Clitic pronouns normally substitute verbal complements, and from a semantic perspective, represent the corresponding arguments. There are a number of properties of these pronouns that seem to hold of clitic systems in different languages, but there are also some other properties that are language specific. Among the general properties, it seems that in all languages clitics can come attached as inflexions of verbs, the so-called enclitic pronouns, but also, according to traditionally spelling conventions, they can be realized as independent lexical units normally in front (to the left) of their verbal host; pronouns of this latter kind are know as proclitic. Another interesting property is that although their positions in the sentence, and in relation to the verb that they attach to, are quite arbitrary and idiosyncratic, the order in which they appear in relation to other clitic pronouns is fixed. Also, clitics interact with the auxiliary verbal system; in particular, one or more auxiliary or modal verbs can intervene between the realization of the pronouns and the verb that has these pronouns as its arguments; in addition, the auxiliary verbs themselves can have enclitics. However, although the pronouns can be distant from their host, their binding properties can be predicted syntactically, as people can very easily establish their

referents, if the sentence is grammatical. For this reason, clitics are distant but bounded dependencies. Another interesting property of clitics is that they are a locus of dialectal variation within languages, and this suggests that the clitic pronoun system is a very loosely part of the syntax, and the structure of the lexicon may influence the forms permitted in one or another dialect of a language.

Coming to the language specific properties, in Spanish, for instance, infinitives, gerunds and imperatives have enclitic pronouns but verbs in personal forms (i.e. tensed or conjugated) require always proclitics; participles cannot have enclitics also and when they are clitizised, the pronouns must be proclitic. However, in periphrasis the pronouns can appear both in enclitic and proclitic forms. In Spanish, most sentences have a corresponding set of clitic versions, with some or all of the verbal complements clitized, and possibly with some or all these realized pleonastically too. We illustrate the basic phenomena with sentences in (1). For clarity, complements are enclosed in brackets when required, and have the same index in their corresponding clitizised versions; also, the pleonastic complements are co-indexed.

(1)   a.   *Juan muestra* [*el catálogo*]$_i$ [*a María*]$_j$
           Juan shows   the catalog$_i$ to Mary$_j$
      b.   *muéstralo$_i$* [*a María*]$_j$
      c.   *muéstrale$_j$* [*el catálogo*]$_i$
      d.   *muéstrase$_j$lo$_i$*
      e.   *muéstrase$_j$lo$_i$* [*a María*]$_j$
      f.   *lo$_i$ muestra* [*a María*]$_j$
      g.   *se$_j$ lo$_i$ muestra*
      h.   *se$_j$ lo$_i$ muestra* [*a María*]$_i$
      i.   *le$_j$ muestra* [*el catálogo*]$_i$ [*a María*]$_j$

In (1) *lo* is a third person masculine accusative pronoun substituting the direct object *the catalog*, and *le* and *se* are third person dative pronouns substituting *a María*. Sentences (1.b) to (1.e) are all enclitic forms of (1.a) in imperative and (1.f) to (1.i) are proclitic tensed forms. In (1.b) the direct object is clitized with *lo*, and in (1.c) the clitic pronoun *le* represents the indirect object. Sentence (1.d) has the two complements realized as enclitics, but the *se* form is used instead of *le* as the "*le lo*" sequence is not allowed. In general, the personal pronoun *se* is a duplication of the personal *le*, as both come from the personal *illi* of Latin, and when *le* should appear in front of *lo* (or *la*, its feminine form), both in proclitics and enclitics, *se* must be used instead[1]. Sentence (1.e) has, in addition, the indirect object realized pleonastically with *se* and *a María*. Sentence (1.f) has the direct object realized as the proclitic *lo*, and (1.g) has both of the complements realized as proclitics with the *se* form obligatory; in (1.h) both of the complements are also proclitic but the indirect object is realized pleonastically. Finally in (1.i) *le* realizes the indirect complement, that is also pleonastic with *a María*; as there is no other pronoun following

---

[1] The genesis of the personal *se* is illustrated by its evolution in three different stages of the language: *diollelo* (palatal *ll*) → *diogelo* (voiced palatal fricative *ge*) → *diose$_i$lo$_j$* (unvoiced fricative *s*); this last form was consolidated since the XIV century (Gili Gaya, § 176, pp. 235), and corresponds to the enclitic form of *se lo dio* (he/she gave it$_j$ to him$_i$).

the *le* form it does not have to be replaced by *se*. Furthermore, the use of *le* is obligatory because its substitution by s*e* in this context would produce the impersonal form *se muestra el catálogo a María* (someone shows a catalog to Maria), where *se* is an expletive, no longer referential, as will be explained below at the end of this section.

Also, in sentences (1.b) to (1.i) the subject is dropped, as it is contained in the conjugation. In Spanish the subject is only made explicit when it needs to be determined or elaborated; it is a kind of pleonastic construction in relation to the person and number marked in the verbal inflection. Similar, in pleonastic constrictions, the pronoun anticipates the person, gender and number, and the explicit complement occurs only when it needs to be further specified or elaborated.

Unlike Spanish, pleonastic constructions seem to be very restricted in other languages. In French, for instance, the duplication of a clitic with its corresponding complement is never allowed (Miller & Sag, 1995), and in Italian this kind of construction is very restricted (Monachesi, 1993, 1999). The general rule in Spanish is that dative pronouns can be duplicated. However, the accusative is duplicated often also, although there are some restrictions: non-animate direct complements cannot be duplicated. The sentence lo_i *entrega a Juan*$_i$ *a la policía* (someone turns Juan in to the police) is allowed (with the proper intonation), but lo_j *da el coche*$_j$ *a María* (someone gives the car to Maria) is not, as the accusative *lo* duplicates the non-animate direct complement *el coche* (the car). The clitic *le* can duplicate a circumstantial complement too, and in this case it is very far from the normal dative use; for instance, *le compró el libro a María* is ambiguous and means "he/she bought the book for Maria" or "he/she bought the book from Maria"; while *le* duplicates *a Maria* in both interpretations, in the first *a Maria* is the indirect complement of *comprar* and *le* has the normal dative reading, but in the second, *a Maria* is a circumstantial complement, and the preposition *a* marks that Maria is the source of the buying action. In this latter interpretation, *le* is closer to an ablative case.

Clitic pronouns are all enclitic or all proclitic and the same order is preserved in both forms; for instance, *se* and *le* always preceded *lo*. In addition, in the basic form of the phenomenon, there is no lexical material between the clitic pronoun sequence and the clitized verb, and this is a strong constraint that seems to hold not only of Spanish, but of all languages with clitic systems. For this reason, despite that in Spanish the complements can appear in front of the verb, and although alternations of (1.i) in (2.a) and (2.b) are grammatical (with the proper intonation), (2.c) and (2.d) are not:

(2) a. [*el catálogo*]$_i$ *le*$_j$ *muestra* [*a María*]$_j$
     b. [*a María*]$_j$ *le*$_j$ *muestra* [*el catálogo*]$_i$
     c. *le$_j$ [*el catálogo*]$_i$ *muestra* [*a María*]$_j$
     d. *le$_j$ [*a María*]$_j$ *muestra* [*el catálogo*]$_i$

Another property of Spanish clitics is that they can attach to verbs even when they are nominalized, and also to gerunds, forming constituents lacking a verb in a finite form. A nominal phrase with an infinitive (its subject *yo*, first person singular in nominative) is

shown in (3.a) and its clitized version in (3.b). A nominal phrase with a gerund and its clitized version are shown in (4).

(3) a. *El mostrar* [*el cuadro*]$_i$ *yo*
    the to-show   the picture$_i$ I
    my showing of the picture
 b. *El mostrarlo$_i$ yo*
    the to-show-it$_i$ I
    my showing of it

(4) a. *Velazquez pintando* [*el cuadro*]$_i$
    Velazquez painting   the picture$_i$
 b. Velazquez *pintándolo$_i$*
    Velazquez showing it$_i$

In addition to the basic forms, clitics can occur in periphrasis where the pronouns are distant to the verb that has them as its complements; in this situation it is said that the pronouns *climb* over the intermediate lexical material (e.g. auxiliary or modal verbs). In clitic climbing we need to distinguish the *clitic host* from the *phonological host*; the former is verb whose complements are clitizised properly, while the latter is the verb that has the explicit realization of the pronouns attached to it, both as proclitics by its left or as enclitics by its right. Clitic climbing is illustrated in sentences (5) and (6):

(5) a. *Juan puede mostrar* [*el catálogo*]$_i$ [*a María*]$_j$
    Juan can   show   the catalog$_i$ to María$_j$
 b. *puede mostrarlo$_i$* [*a María*]$_j$
 c. *puede mostrarle$_j$* [*el catálogo*]$_i$
 d. *puede mostrárse$_j$lo$_i$*
 e. *puede mostrárse$_j$lo$_i$* [*a María*]$_j$
 f. *se$_j$ lo$_i$ puede mostrar*
 g. *se$_j$ lo$_i$ puede mostrar* [*a María*]$_j$
 h. *le$_j$ puede mostrar* [*el catálogo*]$_i$
 i. *le$_i$ puede mostrar* [*el catálogo*]$_j$ [*a María*]$_j$

(6) a. *Juan puede haber querido mostrar* [*el catálogo*]$_i$ [*a María*]$_j$
    Juan could have wanted show   the catalog$_i$ to María$_j$
    Juan could have wanted to show the catalog$_i$ to María$_j$
 b. *Puede haber querido mostrárse$_j$lo$_i$* [*a María*]$_j$
 c. *Se$_j$ lo$_i$ puede haber querido mostrar* [*a María*]$_j$
 d. *Puede habérse$_j$lo$_i$ querido mostrar* [*a María*]$_j$

In (5) and (6) the clitic host is the verb *mostrar* but the phonological host varies: from (5.b) to (5.e) and in (6.b) the phonological host is also *mostrar*, but from (5.f) to (5.i), and also in (6.c), the phonological host is the auxiliary *puede*. In (6.d), the phonological host is the auxiliary *haber* and the enclitic form is obligatory.

Clitic pronouns can also climb over periphrasis with gerunds, as shown in (7):

(7) a. *El   FBI continúa visitando [a Pedro]$_i$*
The FBI continues visiting Pedro
b. *El FBI continúa visitándolo$_i$*
c. *El FBI lo$_i$ continúa visitando*
d. **El FBI continúalo$_i$ visitando*

In addition, the clitics can climb over the participle in the complex conjugation as in (8):

(8) a. *El   FBI   ha continuado visitando [a Pedro]$_i$*
The FBI has continued   visiting   Pedro
b. *El FBI ha continuado visitándolo$_i$*
c. *El FBI lo$_i$ ha continuado visitando*
d. **El FBI ha continuadolo$_i$ visitando*

In Spanish, auxiliaries are subject-raising verbs and modals are subject-control verbs (Pineda & Meza, 2003, Meza & Pineda, 2002), and both of them take as their arguments the verbal phrase next to them in the periphrasis, and the specifier or subject of their complements is the same (or shared) with their own specifier. Also, auxiliaries select the form of the verb next to them in the periphrasis and, in particular, movement verbs like *continúa* in (7) and (8) select for gerunds always. The sentences (7.d) and (8.d) are ungrammatical because, as was mentioned, tensed forms cannot have enclitics (*continúa* (3rd-singular present) and *ha continuado* (3rd-singular actual perfect-past)).

Periphrasis in (5) to (8) are simple in the sense that all clitic pronouns are arguments of the verb with full semantic content at the end of the periphrasis (i.e. *mostrar* (to show), *visitar* (to visit)). However, there are also complex periphrases in which the clitized arguments belong to different content verbs. In this case, it is more difficult, from the point of view of the linguistic analysis, to establish the binding relation between a clitic pronoun and its corresponding verbal argument. In Spanish this kind of periphrasis includes the so-called object-control verbs, like *ver* (to see), whose direct object is the subject of the next verb in the periphrasis. This kind of periphrasis is illustrated in (9):

(9) a. *Juan ha visto [a María]$_i$ comer [la manzana]$_j$*
Juan has seen to-Maria$_i$ to-eat   the apple$_j$
Juan has seen Maria to eat the apple
b. *Ha visto [a María]$_i$ comerla$_j$*
c. *La$_i$ ha visto comerla$_j$*
d. *Le$_i$ ha visto comerla$_j$*
e. *Se$_i$ la$_j$ ha visto comer*

In (9.a) there are two verbs with full semantic content: *visto,* the past participle of *ver* (to see), and *comer* (to eat), with their corresponding direct objects; in (9.b) the direct object of *comer* is clitized, and in (9.c) the direct objects of both verbs are clitized with the same

pronoun *la*, one proclitic to *visto* and the other enclitic to *comer*. For Spanish speakers there is no ambiguity of the referents of these two instances of the same pronoun; and indeed, clitization is a good test for constituency. In (9.d) the accusative *le* is used instead the first *la*. The substitution of *la* or *lo* by the accusative *le*, normally a dative pronoun, is known as *leísmo* and this phenomenon is widely extended in different dialects of Spanish. In (9.e) *se* is proclitic to *visto* and *la*, the object of *comer*, appears also as proclitic. Although the personal *se* normally appears as dative, it can also appear as accusative, as it is clearly the case in this example. Examples in (10) illustrate similar relations, but with a ditransitive verb, *poner*, instead of the transitive *comer*:

(10)   a.   *Juan ha visto [a María]$_i$ poner [el sombrero]$_j$ sobre la mesa*
             Juan has seen to-Maria$_i$ to-put   the hat   on   the table
             Juan has seen Maria to put the hat on the table
       b.   *Ha visto [a María]$_i$ ponerlo$_j$ sobre la mesa*
       c.   *La$_i$ ha visto ponerlo$_j$*
       d.   *Le$_i$ ha visto [a María]$_i$ ponerlo$_j$ sobre la mesa*

On the light of the examples with pleonastic pronouns, it may appear that a kind of reflexivity is related to clitization. And indeed, the reflexive use of pronouns interacts with the clitic system; in particular, the clitic pronouns *me*, *te* (*os*), *se* (singular first, second and third person) and *nos*, *se* (plural first and third person) appear in reflexive constructions substituting the direct and indirect objects. However, pleonastic constructions and reflexivity are two different phenomena: in pleonastic constructions, the same accusative or dative argument is duplicated and appears twice, with the same case, in different positions; in reflexives, on the other hand, different arguments with their corresponding cases are co-indexed.

The Spanish reflexive is a rich phenomena with a wide range of behavior. In the direct reflexive (11.a), the subject *yo* (in nominative) is co-indexed with the direct object *me* in accusative; in (11.b) the explicit subject is dropped, as it is already contained in the conjugation of the verb *lavo* (1st-singular), and the reflexive relation is indicated with the subscript in the corresponding verbal conjugation. Sentence (11.c) is an instance of the indirect reflexive where the dative *me* stands for the patient.

(11)   a.   *Yo$_i$ me$_i$ lavo*
             I$_i$ myself$_i$ wash
             I wash myself
       b.   *Me$_i$ lavo$_i$*
             myself$_i$ wash
             I wash myself
       c.   *Me$_i$ lavo$_i$ las manos*
             Myself$_i$ wash the hands
             I wash my hands (myself)

In addition to this basic reflexive forms, there is a form, the so-called ethical dative (*dativo ético*), in which the subject is not properly the agent but it only has a participation

or interest in the action, or has an influence in an action that is performed by another agent. There is an ample spectrum for this participation that ranges from the explicit participation to expressions with intransitive verbs, often called pseudo-reflexive (*pseudorreflejas*), and there are also some verbs that have only a reflexive use (Gili Gaya, § 58, pp. 73). The ethical dative is illustrated in (12.a) and (12.b). In (12.a), the reflexive relation is established between the dative *te* (2nd-singular), who is also the subject and the beneficiary of the house that is built, but is not the recipient of the action of building; in (12.b) the dative *se,* co-indexed with the subject *ella*, in nominative, is also the beneficiary of the act of drinking the coffee, but she does not drink herself either directly or indirectly. The pseudo-reflexive is illustrated in (12.c); in *me voy* the verb *voy*, first person singular of *ir* (to leave) has no direct or indirect object, and nevertheless the pronoun *me* appears as a dative pronoun (1st-singular) co-indexed with the implicit subject. In (12.d) the reflexive verb *arrepentirse* (infinitive + *se*) appears in the third person singular form, and *se* (in accusative) is co-indexed with the implicit subject. This latter kind of verbs, like *atraverse* (to dare), *jactarse* (to brag), o *quejarse* (to complain), never appear without the reflexive import, and we refer to them as reflexive verbs.

(12) a. *Tu te$_i$ construiste$_i$ una casa*
        you built-for-yourself   a house
        you built a house for yourself
     b. *Ella$_i$ se$_i$ tomó$_i$ el café*
        She her had   the coffee
        She had a coffee
     c. *Me$_i$ voy$_i$*
        Me leave
        I'm leaving
     d. *Se$_i$ arrepiente$_i$*
        He is-sorry

As can be seen from examples (11) and (12) the reflexive, in all its range of forms, has clitic pronouns, which behave in most respects like standard clitics (i.e. enclitic with non-finite forms, proclitic with finite forms, there is no material between the clitic and the verbal host, etc.) and, indeed, the reflexives seem to be subsumed within the clitic system, with the additional constraint that reflexives reinforce binding relations between the subject in nominative, and the direct or indirect objects in accusative or dative respectively.

Consider also sentence (13.a) which is a variation of (9.a), and its clitic versions in (13.b) and (13.c); although the meaning of both (13.a) and (9.a) above is almost the same (13.a) sounds better as (9.a) seems to be somehow incomplete:

(13) a. *Juan ha visto* [*a María*]$_i$ *comerse$_i$*   [*la manzana*]$_j$
        Juan has seen to Maria$_i$ to-eat-herself $_i$ the apple$_j$
        Juan has seen Maria to eat the apple for herself
     b. *La$_i$ ha visto comerse$_i$la$_j$*
     c. *Se$_i$ la$_j$ ha visto comer*

The pronoun *se* of *comerse* in (13.a) is an ethical dative, co-indexed with the *a María*, and has the function of indicating that its referent, Maria, is the beneficiary of the eating action. The role of the arguments represented by these pronouns can be better appreciated in the corresponding clitic version in (13.b) where the object of *visto* is realized by the proclitic la_i, and the direct and indirect objects of *comer* by the clitics la_j and se_i respectively, and the reflexive import of the ethical dative is explicit in the co-indexing relation between the direct object of *visto*, and the indirect object of *comer*. In sentence (13.c) the ethical dative *se* and the direct object of comer, la_j, are both realized as proclitics; particularly interesting is the fact that the accusative la_i and the dative se_i of (13.b) seem to be subsumed in the dative se_i of (13.c). Furthermore, the form of *se la ha visto comer* in (9.e) and (13.c) is ambiguous: in (13.c) *se* is a dative-reflexive while the *se* in (9.e) is, as was mentioned, an alternative realization of *la* or *le* in accusative. The ambiguity is lexical as the form *se* realizes to different pronouns: unlike the personal *se* which comes from the Latin *illi* (see above), the reflexive *se* comes from the Latin reflexive *se* (Gili Gaya, § 104, pp. 126), and it is the only form of the reflexive for third persons, singular and plural, and it expresses direct and indirect complements.

In addition to the personal and reflexive use, the form *se* appears in two additional constructions in Spanish that are similar on the surface form to clitic constructions, but have a very different syntactic and semantic behavior. These are the so-called passive-reflexive (*pasiva refleja*) and the impersonal-active (*activa impersonal*). The passive-reflexive is illustrated in sentences (14):

(14) a. *Juan abre la puerta*   (Juan opens the door)
     b. *La puerta es abierta por Juan*   (The door is opened by Juan)
     c. *La puerta es abierta*   (The door is opened)
     d. *La puerta se abre*   (The door is opened)
     e. *Se abre la puerta*   (The door is opened)

The original active sentence in (14.a) has its first and second of passive versions in (14.b) and (14.c); however, these forms are seldom used in Spanish, and the forms (14.d) and (14.e) are very commonly used instead. In these latter two sentences the verb appears in active form, and the *se* particle, which derives from the reflexive *se*, is non-referential, very much like the English expletive *it*. The purpose of this *se* is to mark the passive interpretation of the active form of the verb, and the passive-reflexive is used when the object is unknown or has no interest for the speaker. This form can only occur with transitive verbs, as it is fundamentally a passive construction in which the patient of the action is the object in the active form, and becomes the syntactic subject in the passive-reflexive (e.g. *la puerta* in (14.d) and (14.e)), and needs to agree with the verb. Also, as only accusative and dative complements can be clitized, the subject of the passive reflexive (in nominative) cannot be realized as a clitic pronoun.

The active form of the verb in the passive-reflexive permitted the evolution by analogy of an active sentence, the impersonal-active, with a similar structure; however, in the active interpretation *se* is not a syntactic mark, but a personal pronoun whose reference is

underdetermined, like the English *one*, either because the referent needs to remain anonymous, or there is no interest in its identity, or simply because it is not known; for this reason, the pronoun is not resolved and the construction remains impersonal. In (15) someone, not specified, is the one who talks to the children:

    (15) *Se habla a los niños*   (one talks to the children)

Also, unlike the passive reflexive in which the overt argument (e.g. *la puerta* in 14.e) is the subject, the active impersonal has an obligatory omitted subject, similar to impersonal verbs like *llover* (to rain); however, the *se* pronoun has to agree with the verb (i.e. 3rd-singular). In (15), for instance, *a los niños* is the direct object, and subsequently, needs not to be in agreement with the verb. On the other hand, it can be clitized as in (16), with *leísmo* obligatory in this case:

    (16) *Se les habla*   (one speaks to them)

The active impersonal can also be used with intransitive, and even with stative, verbs, as in (17):

    (17) a. *Se duerme mal*   (one sleeps badly)
           b. *Se vive tranquilo en el campo*   (one lives quietly in the country side)

These latter forms cannot be confused with the passive or the reflexive, as there is no recipient of the action, or there is no action at all, and the active impersonal provides a very expressive resource to refer to events, processes and states whose agent or patient does not need to be mentioned.

The interaction between the passive-reflexive, the impersonal-active, the use of the *se* form in these forms, and the interaction of these constructions with the clitic system is quite complex. According to Gili Gaya (§ 61, pp. 76-77), in old Spanish there was an ambiguity between the reciprocal and the passive voice: sentence (18) could mean that the students help each other or that someone, not mentioned, helps them:

    (18) *Se ayudan los estudiantes*

In this form *los estudiantes* was the subject if the sentence was interpreted as passive, but it was the object if it received the reciprocal (reflexive), active, interpretation. The ambiguity between the subject and the object could also occur due to the Spanish flexible order of constituents, as the object can be placed in front of the verb. To solve this ambiguity, the language developed the use of the preposition *a* to mark syntactic objects when they were animate or highly determined abstract entities (since the XV century and this use is compulsory in the current state of the language). With this expressive resource, the ambiguity of (18) could be resolved by using the preposition *a* in (19) where *los estudiantes* is the object; if there is no *a* preposition, on the other hand, *los estudiantes* is necessarily the subject and the sentence is passive.

(19) *Se ayudan a los estudiantes*   (some people help the students)

From this latter form, and fixing the verb in singular, the sentence becomes an active one, with an undetermined subject as shown in (20).

(20) *Se ayuda a los estudiantes*   (one helps the students)

The final construction is the modern impersonal-active, and it can be used not only with personal but with all kind of objects; however, without the *a* preposition if the direct object is not animate, as in (21.a). In the plural form, the subject needs to agree with the verb, and sentence (21.b) is clearly a passive one. These forms are both present in the current state of the language and for all practical purposes their meaning is the same.

(21) a. *Se vende libros*   (one sells books)
      b. *Se venden libros*   (books are sold)

With this we conclude the description of the Spanish clitic system and its interactions with reflexives and the impersonal forms. In summary, the clitic system provides a very direct and expressive device of the language, and clitic pronouns occur very often in both spoken and written language. From the syntactic point of view it is a system of distant but bounded dependencies subject of a number of constraints that on the surface seem to be quite idiosyncratic; however, this system as a whole has an underlying systematic behavior. In addition, it interacts with other syntactic systems, like the reflexives, both direct, indirect, reciprocals and ethical dative, and also with the pseudo-reflexives; all of these forms can also be thought of as a part of the clitic system. Clitic sentences are also very similarly and related, by analogy through the reflexive *se*, to the passive reflexive and the impersonal active, which are also very common and useful expressive devices of modern Spanish. For all this, a model of clitics posses not only an interesting theoretical challenge to the syntactic analysis, but also constitutes a very important component of a computational grammar of this language. In the rest of the paper we present a computational model for the clitic system of Spanish, robust enough to cope with the properties of clitics described in this section, and the models of the related phenomena, that interact with the clitic system, sometimes producing apparent ambiguities, as in *se muestra el catálogo a María*, where *se* is necessarily a personal pronoun, but sometimes producing complex lexical and syntactic ambiguities, as in *se muestra el catálogo*, where *se* can be interpreted as a reflexive pronoun or as an expletive, and the whole expression can be reflexive, impersonal-active and even passive-reflexive.

2. Representation level of clitic pronouns

The notion of clitic, in its more general sense, refers to morphological material that has no independent prosodic status and depends on another unit, the host, which is a proper prosodic word (Nevis et. al., 1994, § xii). In this sense, unstressed particles or function words, that are used always in conjunction with other stressed lexical material, are candidates to be clitics. Traditionally, these units are called *proclitic* or *enclitic* according

to whether they appear to the left or to the right of their host respectively. Clitics need to be distinguished from inflexions and derivations which also consist of morphological material that depends on a prosodic word, but in this latter case, the particle is attached to its host in the morphology, and the whole unit is realized as ready made unit in syntactic structures. However, it seems that there is an agreement that the phenomenon of clitics is mainly syntactic:

> "…word-clitic combinality is largely governed by SYNTACTIC considerations. The conditions governing the combinability of stems with affixes are of quite a different sort: they are MORPHOLOGICAL and/or LEXICAL in character, being concerned with the substructure of a finite set of words" (Zwicky and Pullum, pp. 503, 1983).

It is interesting that although traditionally "clitic" pronouns are considered clitics, there is an ambiguity about their level of representation, as they can also be considered inflexions. In traditional orthographic conventions of Spanish and many other languages, for instance, proclitic pronouns are represented as independent units, and are clitics, but enclitic pronouns are spelled attached to their hosts, and can be considered inflexions. More generally, clitic pronouns occur in the interface between the syntax and the morphology, and a model of clitic pronouns must address the question of what is the proper level of representation of these objects. This dilemma is reflected in formal study of clitic pronouns and it seems that there is an underlying assumption that all clitic pronouns must be dealt with at the same level of linguistic representation. In transformational approaches, as reported by Miller and Sag (1997), clitics were regarded as full *NP*s in their usual argument position in deep structure, and their behavior was analyzed in terms of syntactic movement. In lexical approaches, couched in the *HPSG* formalism (Pollard & Sag, 1994), on the other hand, clitic pronouns, both proclitic and enclitic, are considered affixes that have to be dealt with in the morphology; in this latter line of work, clitic pronouns are no longer clitics and, indeed, in Miller and Sag (*ibid.*), these pronouns are explicitly renamed as pronominal affixes

The case for treating clitics as inflectional affixes is based on criteria (A) to (F) below, advanced by Zwicky and Pullum (1983), to distinguish clitics from inflexions in English. These criteria are based on the analysis of the English clitics *'s* 'is', *'s* 'has' and *'ve* "have" and the ways these clear clitics differ from the English inflexions for plurals (*knights*), past tense (*arrived*) and the superlative for adjectives (*fastest*); on the basis of (A) to (F) they concluded that the English contracted negative (e.g. *she hasn't gone*) is not a clitic but an inflection. The criteria advanced by Zwicky and Pullum are as follows:

> (A)   The degree of selection between the dependent morpheme and its host: clitics have low degree of selection and can attach to words of different categories, while inflexions are highly selective and attach only to hosts of a specific categories.

(B)   There are no arbitrary gaps in the set of host-clitic combination, while inflectional paradigms do have gaps: when a particular host combines with one of the three English clitic forms, it combines always with the other two.

(C)   There are no host-clitic combinations showing an unexpected phonological form (e.g. the English clitics '*s* and '*ve* always have the same form). Inflectional formations do exhibit idiosyncrasies (e.g. the English forms *dice*, *oxen*, *feet* for the plural affix).

(D)   There are no semantic idiosyncrasies in host-clitic combinations. The meaning of full forms and clitic forms is the same (e.g. *she is gone* means the same than *she's gone*). Inflectional forms, on the other hand, do exhibit semantic idiosyncrasies and the meaning a whole word is not always composed regularly of the meaning of its parts.

(E)   Syntactic rules can affect affixed words, but cannot affect clitic groups. No syntactic operation, for instance, treat a clitized word like *I've* as a unit; however, inflected words are regularly treated by syntactic operations.

(F)   Clitics can attach to words already clitized, while inflexions cannot. For instance, *I'd've done it if you'd asked me*.

These criteria permit to distinguish clitics from inflexion in English. However, English has no clitic pronouns, and despite that Miller and Sag (1995) and Abeillé et. al's (1996) have adopted these criteria for French, and Monachesi (1999) for Italian, it is not clear whether (A) to (F) can be used to classify Spanish clitic pronouns as affixes. Clearly, on the basis of (D), Spanish clitic pronouns do count as clitics, as the meaning of a full sentence and all its clitized versions is the same; similarly, on the basis of (F) they count as clitics, as these pronouns attach to verbs already clitized (e.g. *muestra-me-lo* (show it to me)); the other criteria are also problematic, and can be argued to classify the pronouns as inflexions, but also as clitics.

According to (E) Spanish clitic pronouns seem to be clitics also; in Spanish, and unlike French and Italian, Spanish proclitic pronouns can have a wider scope over coordination; for instance, (22.b), a version of (22.a), is grammatical:

(22) a. *Lo llevó   y lo puso sobre la mesa*
           it took-he and it put-he on the table
           He took it and put it on the table

     b. *Lo llevó   y puso sobre la mesa*
           it took-he and put-he on the table
           He took it and put it on the table

However, there are contexts where the factorization of the proclitic is clearly ungrammatical, as shown in (23):

(23) a. *Te vas o te quedas*   (you leave or you stay)
      b. **Te vas o quedas*

In relation to (A), although Spanish pronouns select verbs as their host most of the time, they also attach to infinitives, gerunds and participles (e.g. (6) to (8)), that are normally considered verbal nouns, adverbs and adjectives respectively (Gili Gaya, 1991), and clitic pronouns can occur in constructions without a finite or personal verb, as in (3) and (4). Consider also the attributive sentence (24.a) whose meaning is that being faithful to Maria is a property that Juan has; in its clitized version in (24.b), the host of the dative *le* seems to be the adjectival predicate *es-fiel* (is-faithful). An alternative hypothesis is that the clitic host is the copulative *es* (to be); however, unlike predicative sentences with active verbs, that have direct and direct objects, attributives have no complements to be clitized.

(24) a. *Juan es fiel a María*   (Juan is faithful to María)
      b. *Le es fiel a María*   (he is faithful to María)

In relation to (B), clitization is a very systematic behavior, and all sentences can clitizise with very few restrictions and, in this sense, the pronouns behave like clitics. However, there are gaps in the set of possible clitic combinations; for instance, a sequence of two *l'* pronouns never occurs, and in the case "*le la*" is needed, as was mentioned, "*se la*" must be used instead; another idiosyncrasy is the clitic rigid order, with one qualification, as will be explained below; consequently, according to (C), the pronouns do behave more like inflections.

From this discussion, it seems that clitic pronouns in Spanish behave like clitics, but they also have inflectional properties. An additional source of relevant information is speech repairs in task oriented conversation. A common clitic repair occurring in the DIME corpus (Villaseñor et. al., 2001) is "*me...muéstrame otra vez los muebles*" (to me … show-me again the furniture). Here, the speaker pronounces the dative *me* that would be proclitic to *mostrar* (to show), with the clear intention to say "*me muestras...*" (to me you-show…); however, there is a repair at this point and, after a brief pause, the enclitic form is used instead. If the pronoun were an inflexion, this repair could not occur, as the form "*me muestras*" would have to be ready made at the morphological level, before lexical insertion and phonological realization.

For all these reasons, we adopt the traditional intuitions reflected in the orthographic conventions of Spanish, and consider that clitics pronouns have a dual behavior; on the one hand, proclitic pronouns are really clitics, but enclitic pronouns, on the other, are inflexions. Consequently, the combination of a proclitic pronoun with its host is a syntactic operation. However, there are severe constraints on the application of this operation, as there is never lexical material between the clitic sequence and the verbal phrase that has the clitics as its arguments. Also, a sequence of pronouns respects a fixed order, and there is never lexical material between pronouns in the sequence; in addition, there are no repairs interrupting a pronoun sequence. All this suggests that proclitic

sequences are realized as units in the lexicon, and as such, they can undergo syntactic operations; on the other hand, although it is possible that enclitics originated as clitics, in modern Spanish they are inflexions; also, as the forms following the verbs in enclitics are exactly the same as the ones that can appear independently as proclitics, we suggest that the morphological-lexical operation that forms enclitics takes a verbal lexeme and a clitic sequence, present in the lexicon, as its input for the realization of the word.

3. A Model for Spanish Clitic pronouns

According to the previous discussion, we postulate a model for the clitic pronouns, proclitics, and pronominal affixes, enclitics, with a morphological/lexical and a syntactic component. For clarity, and following traditional terminology, we refer to both proclitics and enclitics as clitics. The model is based on a lexical operation that changes the combinatorial potential or valence of the clitic host, and allows the realization of clitics instead the explicit arguments. We refer to this operation on the valence of the host as *clitization*. For this reason, although the model has a syntactic operation, it has a strong lexicalist orientation, and the theory as a whole is formalized in HPSG, as will be shown in Section 4.

To allow pleonastic constructions, the clitization operation can optionally, in addition to select one or more arguments for clitization, preserve the clitized arguments in the valence of the host; in this case, clitic and pleonastic argument are co-indexed by the same operation. In this way, the host undergoes two orthogonal operations: on the one hand, it consumes its arguments through standard syntactic combination schemes; on the other, combines with its clitic arguments through the scheme shown below.

We also postulate that there is a clitic lexicon consisting of a set of clitic pronoun sequences, including case information; clitic pronouns are also marked for number and gender (masculine, feminine and neuter) and these features are also specified in the lexical entry. There is an entry in the clitic lexicon for each pronoun sequence with a combination of these features. The entries with their case value for one element sequences used in Mexican Spanish are shown in (25), the entries for two singular pronouns sequences in (26) and the combination of singular and plurals in (27).

(25) "se_{DAT}", "se_{ACC}", "te_{DAT}", "te_{ACC}", "me_{DAT}", "me_{ACC}", "lo_{ACC}", "la_{ACC}", "le_{DAT}", "le_{ACC}", "los_{ACC}", "las_{ACC}", "les_{DAT}", "les_{ACC}"

(26) "$se_{DAT}\ te_{ACC}$", "$se_{DAT}\ me_{ACC}$", "$se_{DAT}\ le_{ACC}$", "$se_{DAT}\ la_{ACC}$", "$se_{DAT}\ lo_{ACC}$", "$se_{ACC}\ te_{ACC}$", "$se_{ACC}\ me_{ACC}$", "$se_{ACC}\ le_{ACC}$", "$se_{ACC}\ la_{ACC}$", "$se_{ACC}\ lo_{ACC}$", "$se_{ACC}\ te_{DAT}$", "$se_{ACC}\ me_{DAT}$", "$se_{ACC}\ le_{DAT}$" "$te_{ACC}\ me_{ACC}$", "$te_{ACC}\ la_{ACC}$", "$te_{ACC}\ lo_{ACC}$", "$te_{DAT}\ me_{ACC}$", "$te_{DAT}\ la_{ACC}$", "$te_{DAT}\ lo_{ACC}$", "$me_{ACC}\ le_{ACC}$", "$me_{ACC}\ la_{ACC}$", "$me_{ACC}\ lo_{ACC}$" "$me_{DAT}\ le_{ACC}$", "$me_{DAT}\ la_{ACC}$", "$me_{DAT}\ lo_{ACC}$"

(27) "se_{DAT} nos_{ACC}", "se_{DAT} les_{ACC}", "se_{DAT} las_{ACC}", "se_{DAT} los_{ACC}",
       "se_{ACC} nos_{ACC}", "se_{ACC} les_{ACC}", "se_{ACC} las_{ACC}", "se_{ACC} los_{ACC}",
       "se_{ACC} nos_{DAT}", "se_{ACC} les_{DAT}"
       "te_{ACC} nos_{ACC}", "te_{ACC} las_{ACC}", "te_{ACC} los_{ACC}",
       "te_{DAT} nos_{ACC}", "te_{DAT} las_{ACC}", "te_{DAT} los_{ACC}",
       "nos_{ACC} las_{ACC}", "nos_{ACC} los_{ACC}", "nos_{DAT} las_{ACC}", "nos_{DAT} los_{ACC}"

The number of pronoun sequences is rather small as not all possible combination exist: we have observed no instances of a two dative pronouns sequence, *se* is the only form for the third person, singular and plural, there is no pronoun for the plural second person, no sequence with two *l*'s pronouns (i.e. *la*, *las*, *lo*, *los*, *le*, *les*) is allowed and *te* can never be followed by *le* or *les*. The clitic sequences in (25) to (27) are the forms used in Mexican Spanish, although there may be some exceptions no included in the list, as the use of *le* in substitution of a circumstantial complement, as was mentioned in Section 1. However, in common language, a small subset of the most common forms is all that is needed. It is also interesting that locus of dialectal variation in Spanish depends on the clitics sequences used in different dialects. In several regions of Spain, *lo* and *la* are used as dative (*loísmo* and *laísmo*) but these uses are never found in Mexico, for instance.

The clitic lexicon captures the clitic idiosyncratic order; the rule in Spanish is that *se* must be in front always, second persons are in front of first persons and third persons (but *se*) are always last. It is also interesting that some dialects can violate this standard order allowing the sequences "*me se*" and "*te se*", as in Caribbean Spanish (Rivero-Castillo, 1997). In our model, an order is possible if the corresponding clitic sequence is present in the clitic lexicon, and this kind of dialectal variation can be explained in terms of the presence or absence of specific clitic sequences and its use in the dialect in question.

Clitic realization in our model is a dual phenomena: proclitics are realized as lexical units or words directly and are subject of a syntactic operation; enclitics, on the other hand, are realized as inflexions, but clitic sequences in the clitic lexicon are channeled as inputs to the morphological operation that produces the inflected form, and the whole enclitic is lexically realized as a ready made inflected word. In this way, the clitic order is preserved and no clitic sequence can be realized either as proclitic or enclitic if it is not present as an entry in the clitic lexicon. We refer to clitic realization as *clitic insertion* for both proclitics and enclitics. While clitization is an operation on the valence of the clitic host, in clitic insertion the pronouns are realized attached to or next to the phonological host.

The explicit realization of a clitic combines with its corresponding clitizations in the production of a clitic sentence. We refer to this clitic operation as *clitic cancellation*. Through this cancellation the bindings between the clitic pronouns and the corresponding arguments of the host are established. Cancellation can be both lexical and syntactic. In the production of an enclitic, the host is subject to clitization, clitic insertion and clitic cancellation, and the three operations can be implemented through a single lexical rule, and we refer to this process as *Simple Lexical Cancellation* (*S-LC*); however, if the pronouns appear as enclitics but the clitic host is different from the phonological host, clitic insertion and clitic cancellation can be implemented by the same lexical rule, which

is different from the lexical rule that performs the clitization of the host; we refer to this latter operation as *Complex Lexical Cancellation* (*C-LC*). Also, according to the Spanish grammar, personal or finite forms and also participles, cannot have enclitics; consequently, clitic insertion cannot be performed on these forms, and this is a morphological constraint on the corresponding operations.

Proclitics, on the other hand, are inserted lexically, and combine with the host by a syntactic operation that we refer to as *syntactic cancellation* (*SC*). However, this is a restricted operation as the complement list of the verbal phrase must be saturated before clitic cancellation takes place; as a consequence of this restriction, there is never lexical material between the clitic sequence and the phonological host.

With these elements and definitions, we formulate the following principle for clitic constructions:

(28) *Principle of clitic cancellation* (version 1): Clitic sentences undergo a single clitization, a single clitic insertion and a single clitic cancellation operation.

Principle in (28) allows us to account for sentences in (1) which are repeated here in (29) for clarity. In the right column, the corresponding kind of cancellation operation, and the clitic sequence involved are shown:

(29) j. *Juan muestra* [*el catálogo*]$_i$ [*a María*]$_j$
   Juan shows   the catalog$_i$ to Mary$_j$
   k. *muéstralo$_i$* [*a María*]$_j$ (*S-LC*: "*lo*$_{ACC}$")
   l. *muéstrale$_j$* [*el catálogo*]$_i$ (*S-LC*: "*le*$_{DAT}$")
   m. *muéstrase$_j$lo$_i$*   (*S-LC*: "*se*$_{DAT}$ *lo*$_{ACC}$")
   n. *muéstrase$_j$lo$_i$* [*a María*]$_j$ (*S-LC*: "*se*$_{DAT}$ *lo*$_{ACC}$")
   o. *lo$_i$ muestra* [*a María*]$_j$ (*SC*: "*lo*$_{ACC}$")
   p. *se$_j$ lo$_i$ muestra*   (*SC*: "*se*$_{DAT}$ *lo*$_{ACC}$")
   q. *se$_j$ lo$_i$ muestra* [*a María*]$_i$ (*SC*: "*se*$_{DAT}$ *lo*$_{ACC}$")
   a. *le$_j$ muestra* [*el catálogo*]$_i$ [*a María*]$_j$ (*SC*: "*le*$_{DAT}$")

The principle also accounts for the noun phrases in (3) and (4) which are analyzed through simple lexical cancellation; also, sentence in (2) is ruled out as there is material between the clitic word and the clitizised verb, preventing clitic syntactic cancellation. In (30), we show how the principle of clitic cancellation accounts for the sentences with simple clitic climbing (i.e. all clitic pronouns are arguments of the content verb at the end of the periphrasis) in (5):

(30) a. *Juan puede mostrar* [*el catálogo*]$_i$ [*a María*]$_j$
Juan can show the catalog$_i$ to María$_j$
b. *puede mostrarlo$_i$* [*a María*]$_j$ (*S-LC:* "*lo*$_{ACC}$")
c. *puede mostrarle$_j$* [*el catálogo*]$_i$ (*S-LC:* "*le*$_{DAT}$")
d. *puede mostrárse$_j$lo$_i$*   (*S-LC:* "*se*$_{DAT}$ *lo*$_{ACC}$")
e. *puede mostrárse$_j$lo$_i$* [*a María*]$_j$ (*S-LC:* "*se*$_{DAT}$ *lo*$_{ACC}$")
f. *se$_j$ lo$_i$ puede mostrar*   (*SC:* "*se*$_{DAT}$ *lo*$_{ACC}$")
g. *se$_j$ lo$_i$ puede mostrar* [*a María*]$_j$ (*SC:* "*se*$_{DAT}$ *lo*$_{ACC}$")
h. *le$_j$ puede mostrar* [*el catálogo*]$_i$ (*SC:* "*le*$_{DAT}$")
i. *le$_i$ puede mostrar* [*el catálogo*]$_j$ [*a María*]$_j$ (*SC:* "*le*$_{DAT}$")

In (31), we show how the principle of clitic cancellation accounts for the sentences in (6) including clitic insertion in auxiliaries and complex lexical cancellation:

(31) a. *Juan puede haber querido mostrar* [*el catálogo*]$_i$ [*a María*]$_j$
Juan could have wanted show the catalog$_i$ to María$_j$
Juan could have wanted to show the catalog$_i$ to María$_j$
b. *Puede haber querido mostrarse$_j$lo$_i$* [*a María*]$_j$ (*S-LC:* "*se*$_{DAT}$ *lo*$_{ACC}$")
c. *Se$_j$ lo$_i$ puede haber querido mostrar* [*a María*]$_j$ (*SC:* "*se*$_{DAT}$ *lo*$_{ACC}$")
d. *Puede habérse$_j$lo$_i$ querido mostrar* [*a María*]$_j$ (*C-LC:* "*se*$_{DAT}$ *lo*$_{ACC}$")

Sentences in (7) and (8) with gerunds and participles are also accounted for through the principle of clitic cancellation, taking into account the restrictions on clitic insertion on tensed forms and participles.

The principle of clitic cancellation in (28) can handle the basic clitic phenomenon including simple clitic climbing in a simple and natural way; however, complex clitic climbing (i.e. periphrasis where the clitic pronouns in the sequence are arguments of different verbs) present additional problems. Consider sentence (32.b) (a simplified form of (9)), where the direct objects of both *ve* (sees) and *comer* (to eat) are clitized, but with two different instances of the same pronoun *la*. As *ve* is a conjugated form (i.e. 3rd-singular) its clitized direct object needs to be proclitic, and the whole construction seems to undergo two different clitizations, two clitic insertions, and two clitic cancellation operations (i.e. *SC* by the left and *S-LC* by the right) violating the principle of clitic cancellation in (28). Also, it appears that (32.c) contains the sequence "*la la*" which is not in the clitic sequence, and there is also lexical material between the two instances of the pronoun.

(32) a. *Juan ve* [*a María*]$_i$ *comer* [*la manzana*]$_j$
Juan sees to-María$_i$ to-eat the apple$_j$
Juan sees María to eat the apple
b. *La$_i$ ve comerla$_j$*

The situation looks even more confusing if we consider (33), that corresponds to (13); in addition to all the problems mentioned for (32.b), (33.b) seems to violate the clitic order,

as there is a *la* preceding the *se*. Also, the clitic sequence "*la se la*", interrupted with lexical material, is not in the clitic lexicon.

(33) a. *Juan ve* [*a María*]$_i$ *comerse*$_i$ [*la manzana*]$_j$
        Juan sees to-Maria$_i$ to-eat-herself   the apple$_j$
        Juan sees Maria to eat the apple for herself
    b. *La$_i$ ve comerse$_i$la$_j$*

However, all these problems are only apparent, as in each of (32.b) and (33.b) there are two independent clitic sequences attached to two different phonological hosts that are cancelled with their corresponding clitizations. To capture this phenomenon, we define the *local clitic domain* of a clitic host as the list of its clitized arguments. Also, for interpreting a pronoun successfully, its corresponding clitic domain must be within the scope of its phonological host; otherwise the binding relation between the pronoun and its corresponding argument position cannot be established. As the pronouns are combined directly with the phonological host, if the clitic host is within the scope of the phonological host, it is also accessible to the pronouns. In (32.b) and (33.b) *la$_i$* has access to its argument position (i.e. the direct object or *ve*), but the clitized arguments of *comer* are hidden from it within the verbal phrase headed by *comer*, as these arguments are already cancelled by their corresponding pronouns. Intuitively, a local clitic domain is within the scope of a phonological host if there is a binding path between both of them. This is the case if there are no clitic pronouns interrupting the binding path (inducing a referential ambiguity) between the pronouns and their corresponding argument position.

With these observations in mind, we modify the principle of clitic cancellation as follows:

(34) *Principle of clitic cancellation* (version 2): There is a single clitic insertion and a single cancellation for each for each clitic host. Cancellation is allowed if every clitic host is within the scope of its corresponding phonological host.

In simple and complex periphrasis, the verbs with semantic content are the ones that clitizise and have proper local clitic domains, as other verbs in the periphrasis (i.e. auxiliaries and modals) only allow the creation of binding paths between clitization and the phonological host. For this, we define a local clitic domain for every periphrastic verb, and say that two local domains are accessible if there is a binding path between the them; also accessible local clitic domains form a *clitic domain*; although we will present the formal definition of the accessibility relation between local clitic domains in Section 4, we can anticipate intuitively that the clitic host is within the scope of the phonological host if their local clitic domains are accessible, or alternatively, if they belong to the same clitic domain. In all three cancellation modes, *S-LC*, *C-LC* and *SC*, the phonological host and its corresponding clitic host belong to the same clitic domain. Consider sentences in (35):

(35) a. *Juan pudiera haber visto* [*a María*]$_i$ *comer* [*la manzana*]$_j$
Juan could   to-have seen to-Maria$_i$ to-eat   the apple$_j$
Juan could had seen Maria to eat the apple
   b. *La$_i$ pudiera haber visto comerla$_j$*
   c. *Le$_i$ pudiera haber visto comerla$_j$*
   d. *Se$_i$ la$_j$ pudiera haber visto comer*

In (35.b) the local clitic domain of *comer* is also its clitic domain; however, the proclitic *la$_i$* has climbed up to the front of the sentence; in this latter case, lexical insertion and cancellation is performed outside the local clitic domain of *visto*, but this latter verb and its phonological host *pudiera* are within the same clitic domain. In this example there are two independent clitic domains and the principle that there is one phonological host and one cancellation for each clitization holds; in addition, there are no pronouns blocking the binding path between phonological and clitic host in neither of the two clitic domains; similarly in (35.c).

Version 2 of the principle can account for the (32.b), (33.b), (35.b) and (35.c). However, it cannot account for sentence (35.d) as the pronoun *la$_j$* is placed between the pronoun *se$_i$* and its clitic domain, blocking the binding path between *se$_i$* and its clitic host. In this situation, it seems that the clitic domain of *visto* is no longer within the scope of the first pronoun, preventing clitic cancellation. However, the clitic sequence "*se$_i$ la$_j$*" appears in front of the whole construction, and this unit can be cancelled with an abstract clitic host that is formed by the composition of the clitizations of *visto* and *comer*; in this case, both of the pronouns seem to have climbed from their original local domains, composing a clitic word (i.e. "*se$_{ACC}$ la$_{ACC}$*"). The process is illustrated schematically in (36):

(36) a. *Juan pudiera haber visto* [*a María*]$_i$ *comer* [*la manzana*]$_j$
Juan could to-have seen to Maria$_i$ to-eat   the apple$_j$
Juan could had seen Maria to eat the apple
   b. **Pudiera haber vístole$_i$ comerla$_j$*
   c. **pudiera haber vístose$_i$la$_j$ comer*   (*visto* + "*se$_i$ + la$_j$*")
   d. *pudiera habérse$_i$la$_j$ visto comer*   (*haber* + "*se$_i$ + la$_j$*")
   e. **pudiérase$_i$la$_j$ haber visto comer*   (*pudiera* + "*se$_i$ + la$_j$*")
   f. *Se$_i$ la$_j$ pudiera haber visto comer*

In (36.b) the direct objects of *visto* and *comer* are clitized (with leísmo for *visto*) in their corresponding local clitic domains; in (36.c) the direct object of *comer* climbs up to a position next to the clitized direct object of *visto*, with the compulsory substitution of *le* by *se* but preserving the accusative case, forming a clitic sequence present in the clitic lexicon; in (36.d) the whole of the composed clitic sequence climbs up, and it is realized as an enclitic of *haber*; the clitic composite climbs up again in (36.e) and, finally, to (36.f) where it is realized as proclitic to the whole of the periphrasis. In the sequence, sentences (36.b), (36.c) and (36.e) are ungrammatical because participles and finite forms cannot have enclitics (i.e. *pudiera* is the pluscuamperfect of the subjunctive), but the ungrammaticality of these sentences is unrelated to clitic climbing, as in Spanish this seems to be a very unconstrained phenomenon. Furthermore, Spanish speakers have no

problem for establishing the bindings required to interpret all sentences in (36) successfully, and the ungrammaticality of enclitics with participles and finite forms seems to be a lexical constraint on these forms, that is contingent to the current linguistic use. These sentences sound bad, but they can be perfectly understood. And, indeed, some of these can be used in pedantic or sarcastic discourse.

The sequence in (36) is only schematic and there is no notion of movement, empty constituents or traces involved. The point is only to illustrate how the clitization of two different host combine in the climbing process, forming an *abstract clitic host*, or a local but composite clitic domain. From the point of view of our clitic model and its formalization, as will be shown in Section 4, sentences (36.d) and (36.f) are analyzed in terms of two clitizations in two different local clitic domains that are *composed* into one abstract clitic host which requires only one clitic insertion and one clitic cancellation. Also, the phonological host and the abstract clitic host belong to the same clitic domain. Sentence (36.d) is analyzed in terms of complex lexical cancellation, and (36.f) by syntactic cancellation. On the basis of this intuition we introduce the following principle:

> (37) *Principle of clitic composition*: clitizations of two accessible local clitic domains can be composed to form an abstract clitic domain.

We also reformulate the principle of clitic cancellation to account for the composition of clitizations:

> (38) *Principle of clitic cancellation* (version 3): Clitic sentences undergo a single clitic insertion and a single cancellation for each independent clitic domain. Cancellation is allowed if every clitic host is within the scope of its corresponding phonological host.

Principles (37) and (38) account for the grammaticality of (35.d) and also (36.f), where the clitic word used is "se_{ACC} la_{ACC}"; however, the sentence (35.c) is ambiguous and can also be produced by the clitization of (39.a); in particular, the proclitic word in (39.c) is "se_{DAT} la_{ACC}" instead, and this latter sentences has a reflexive connotation due to the ethical dative that is absent in (35), as shown in the clitization sequence in (39):

> (39) a. *Juan pudiera haber visto* [*a María*]$_i$ *comerse*$_i$ [*la manzana*]$_j$
>   Juan could to-have seen to Maria$_i$ eat-herself$_i$ the apple$_j$
>   Juan could had seen Maria to eat the apple for herself
>   b. *La$_i$ pudiera haber visto comerse$_i$la$_j$*
>   c. *Se$_i$ la$_j$ pudiera haber visto comer*

In fact, (39.a) is already a clitic version of a sentence that is never explicitly realized, but we show in (40.a) to illustrate the meaning of the ethical dative. We also show a possible schematic genesis of (39.c):

(40) a. *Juan pudiera haber visto* [*a María*]$_i$ *comer* [*la manzana*]$_j$ [*para María*]$_i$
　　　Juan　could　to-have seen　to-Maria$_i$　to-eat　　the apple$_j$　　for-Maria$_i$
　　　Juan could had seen Maria to eat the apple for Maria
　　b. *Juan pudiera haber visto* [*a María*]$_i$ *comerse*$_i$　[*la manzana*]$_j$
　　c. **pudiera haber vístola*$_i$ *comerse*$_i$*la*$_j$
　　d. **pudiera haber vístola*$_i$+*se*$_i$*la*$_j$ *comer*
　　e. **pudiera haber vístose*$_i$*la*$_j$ *comer*　　　　　　(i.e. *se*$_i$ = *la*$_i$+*se*$_i$)
　　f. *pudiera haberse*$_i$*la*$_j$ *visto comer*
　　g. **pudierase*$_i$*la*$_j$ *haber visto comer*
　　h. *Se*$_i$ *la*$_j$ *pudiera haber visto comer*

As can be seen, (40.b) the dative *se* of *comerse* substitutes the indirect object of *comer* and this argument is also the direct object of *visto*, and also the subject of *comer*. Sentence (40.c) consists on the clitization of these three arguments, where *la*$_i$ and *se*$_i$ correfer, despite that *la*$_i$ is accusative and *se*$_i$ is clearly a dative, but this difference of case of co-indexed terms is what distinguish reflexives from simple duplications. Very interesting is the fact that (40.d) has one more argument than it corresponding version (40.e), as in this latter sentence the dative *se*$_i$ seems to subsume the accusative *la*$_i$. On the basis of these observations, we introduce the following principle:

(41) *Principle of clitic subsumption*: an abstract clitic domain can be created by the subsumption of two clitizations of accessible clitic domains, if the subsumed clitization is co-indexed with an element of the subsuming clitization.

The grammaticality of (40.f) and (40.h) in its reflexive interpretation is explained in terms of (41) and the clitic cancellation principle (version 3). In particular, the (40.f) is produced through a complex lexical-cancellation, and (40.h) through syntactic cancellation. In both of these sentences the phonological host and the abstract clitic host belong to the same clitic domain, which is produced by subsumption, and there is a single clitic cancellation. As before, the ungrammaticality of sentences in (40), but (40.d), is due to lexical constraints, but if these were relaxed, these sentences would be grammatical, as Spanish speakers have no problem to establishing the corresponding bindings for achieving successful interpretation. Sentence (40.d) is ungrammatical, in addition, because the sequence "*la se la*" is not in the clitic lexicon.

With this elements, we conclude the presentation of the elements for the clitic model, and we present the following principle for well-formed clitic sentences:

(42) *Clitic principle*: The structured of Spanish pronominal clitic sentences can be accounted for in terms of the principles of clitic cancellation (version 3), clitic composition and clitic subsumption.

At this point it can be observed that the creation of abstract clitic hosts supports the case that proclitics are independent lexical items; if they were inflections it would not be possible to define clitic composition and subsumption, because clitizations could not be factored out to form abstract clitic domains: the clitic arguments could be seen from

outside the local domain, but they could not participate in clitic composition or subsumption. Clitic insertion could be defined as a lexical operation, as in Miller and Sag (1995) and Monachesi (1999), but clitic cancellation could not be defined for periphrasis involving the formation of abstract clitic hosts.

The notions of abstract clitization and abstract clitic host allows us to account for constructions in which the clitic has a wider scope over coordination as in (22.b). In *Lo llevó y puso sobre la mesa*, the two hosts *llevó* and *puso* undergo independent clitizations; however, as the clitizations are co-indexed, an abstract clitic host can be formed by subsumption, and the clitic principle accounts for the whole construction: one clitic insertion and one cancellation. The theory also explains how the word *llevó* is the phonological host of the clitization of *puso*. In the case the proclitic is not factorized out of the coordination, there are two independent clitic domains, and the sentence is accounted for by the clitic principle. The restriction in constructions like (23.b) (**Te vas o quedas*) may be accounted for in terms of other interacting phenomena; in (23.b), for instance, both of the verbs *vas* and *quedar* are reflexive and, as was mentioned in Section 1, this use is lexical in the current state of the language (Gili Gaya, § 58, pp. 74); subsequently, the two instance of *te* in (23.a) (*Te vas o te quedas*) are inflexions.

With the clitic principle at hand, the analysis of Spanish reflexives is straightforward. Reflexives are clitic sentences with an additional constraint marking the reflexive relation between the agent and the recipient of the action; also, their respective representations as subject and complement (direct or indirect) must agree in number and gender. Consequently, the clitization operation for reflexive sentences, in addition of changing the valence of the host, co-indexes the subject with the clitized argument. With this additional constraints, reflexives are accounted for in terms of the clitic principle. In (11.a) and (11.b), for instance, the direct object is clitized and this argument is co-indexed with the subject; in (11.c) the clitized argument is the indirect object. In all three examples, the sentence undergoes one clitization, one clitic insertion and one clitic cancellation. Also, in order that the reflexive relation holds, the phonological and clitic host must belong to the same clitic domain, in the same sense than in traditional reflexive theories, the reflexive relation is established between arguments of the same verb. The case of the ethical or participation dative in (12.a) and (12.b) have a similar analysis. In the case of pseudo-reflexives with intransitive verbs, like (12.c), the clitization rule clitizises an additional argument, not included in original the valence of the verb, which is also co-indexed with the subject. In the reflexive verbs, the additional argument is already lexicallized.

The clitic principle also applies to attributes sentences with predicative adjectives; in *le$_i$ es fiel a María$_i$* or simply, *le es fiel*, (he is faithful to Maria), the complement of the adjective (*a María*) is replaced by the dative *le* that is realized as proclitic (the enclitic form is ungrammatical); here, the local clitic domain of the copulative *es* (is), which is also a subject-raising verb and the local clitic domain of the adjective itself are accessible and form a clitic domain. The relation also holds for the complex conjugation with a more complex periphrasis: in *le hubiera sido fiel* (he would have been faithful to her), *le*

still substitutes the adjectival complement and the local clitic domains of the auxiliaries and the adjective are accessible and form a single clitic domain.

In summary, the definition of the clitic principle accounts for three basic forms of clitic sentences; in the first, the sentence is partitioned in one or more independent clitic domains, and each of these domains is subject of the three clitic operations. For this kind of sentences to be well-formed each phonological and clitic host must belong to the same clitic domain; in particular, there must be no lexical or structural ambiguity breaking up the clitic domain. Sentences violating this constraint are ambiguous and it is not possible to establish the binding relation between the pronouns and their corresponding clitization. In *la pudiera haberla visto comer*, for instance, it is not possible to determine unambiguously the binding relations between the two *la*'s and their corresponding clitic domains, as both seem to be attracted to Maria, the direct object o *visto,* and the sentence is rendered ungrammatical, as predicted by the clitic principle. In the second kind of sentence, an abstract clitic host with its corresponding local clitic domain is created by the composition of two clitizations, and this abstract host is the one which undergoes clitic cancellation; in the third kind, an abstract clitic host is also created, but by subsumption of one clitic domain into another. All clitic sentences that we have observed can be accounted for by these three simple clitic operations, and by the principles of clitic cancellation, clitic composition and clitic subsumption, or simply, by the clitic principle.

The theory also explain the ungrammaticality of sentences with clitic sequences absent from the clitic lexicon: a clitization, either basic or abstract, needs to be cancelled with a valid clitic form. So, if clitic sequences not included in the lexicon are formed through clitic composition, the sentence is rendered ungrammatical as the clitic composite cannot be cancelled. Also, some idiosyncrasies of the words included in the lexicon can be explained on the basis of the clitic principle. For instance, sequence of two *l'* pronouns are absent from the lexicon because their inclusion would create binding ambiguities: if the sequence "*la la*" were used instead of "*se la*" in (36.f), (i.e. *la_i la_j pudiera haber visto comer*), the binding relations between the two instances of *la* and their corresponding hosts would be too ambiguous; this kind of pressure (independently of the phonological factors) could have prevented the inclusion of this kind of sequences in the clitic lexicon. With these remarks, we conclude the presentation of the model, and proceed to its formalization in HPSG in Section 4. The analysis of the impersonal constructions, which also use the clitic machinery, is presented in Section 5.

4. The HPSG Model

Clitization is a lexical operation on the valence of content verbs; as almost every Spanish sentence has a number of clitic forms, we define a *clitic-list* attribute that, together with the subject and complements, constitute the valence of the host. The value of this attribute is the local clitic domain of the clitic host. In particular, in the absence of clitization, the clitic-list is empty. Clitic insertion, on the other hand, consists on the realization of a clitic sequence, included in the clitic lexicon, either as an inflexion on enclitics and auxiliary verbs, or as an independent lexical unit in the case of proclitics. For these reasons, the basic elements for the clitic model are the set of clitic words in the

clitic lexicon, and the definition of a clitic-list attribute in the valence of verbs with full semantic content.

For the formalization we adopt the HPSG formalism, as in Pollar and Sag (1994). Clitic words are defined as pairs consisting of a form (i.e. the orthography of the clitic sequence) and a typed feature structure. The type of a clitic word is *clitic* and it has a *clitic-list* attribute with the specification of a nominal object including case information for each clitized complement. There is an entry of this type for each possible clitic sequence, and case information can be sub-specified. The entry for the dative-accusative clitic sequence "*me lo*", for instance, is shown in Figure 1.

$$
\begin{bmatrix}
me \quad lo \\
SS\,|\,LOC \begin{bmatrix}
CAT \begin{bmatrix}
HEAD & clitic \\
VAL\,|\,CL-LIST & \langle NP-dat_i, \quad NP-acc_j \rangle
\end{bmatrix} \\
CONT\,|\,RESTR & \{[speaker-rel]_i, [male-rel]_j\}
\end{bmatrix}
\end{bmatrix}
$$

Figure 1. Clitic word

The entry for the ditransitive verb *mostrar* (to show), including the basic definition of its local clitic domain, is shown in Figure 2.

$$
\begin{bmatrix}
mostrar \\
SS\,|\,LOC \begin{bmatrix}
CAT \begin{bmatrix}
HEAD & verb \\
VAL\,|\,CL-LIST & \langle \ \rangle \\
ARG-ST & \langle NP_i, NP_j, PP_k \rangle
\end{bmatrix} \\
CONT\,|\,RESTR \left\{ \begin{bmatrix}
show-rel \\
ACT & i \\
UND & j \\
THM & k
\end{bmatrix} \right\}
\end{bmatrix}
\end{bmatrix}
$$

Figure 2. Lexeme entry for the verb *mostrar*

With these elements we can define the basic model for simple lexical cancellation (*S-LC*) and syntactic cancellation (*SC*). For *S-LC*, clitization, insertion and cancellation are defined with a single lexical rule, as shown in Figure 3. The *HEAD|FORM* feature's value *non-part* is a subtype of the type *fin* for finite forms that are not participles (i.e. infinitive and gerunds). The cancellation operation removes the clitizised complements from the valence of the verb, perform clitic insertion in its morphology, and establishes the corresponding bindings in the semantics.

$$
\begin{bmatrix}
dtrv-lxm \\
ORTH & \#1 \\
HEAD\,|\,FORM & non-part \\
VAL\,|\,CL-LIST & \langle \ \rangle \\
ARG-ST & \langle \#2,[],[] \rangle
\end{bmatrix} \mapsto \begin{bmatrix}
ORTH & F_{-melo}(\#1) \\
VAL\,|\,CL-LIST & \langle \ \rangle \\
ARG-ST & \langle \#2 \rangle
\end{bmatrix}
$$

Figure 3. *S-LC* Lexical rule

The analysis of the imperative sentence *muéstramelo* (show it to me) is shown in Figure 4, where the word resulting from lexical realization by means of the *S-LC* rule, is promoted into a phrase by the *Head-Complement Rule* (*HCR*), taking an empty

complements list in this case, and into a sentence by mean of the Spanish *Head PRO-Drop Rule* (*HPDR*) that consumes the omitted subject (i.e. 2nd. person singular).

$$
\begin{bmatrix} HEAD & \#1 \\ VAL & \begin{bmatrix} SUBJ & \langle\ \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix} \end{bmatrix} \quad (=S)
$$

$$
\begin{bmatrix} HEAD & \#1\ \&\ verb \\ VAL & \begin{bmatrix} SUBJ & \langle[I]\rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix} \end{bmatrix}
$$

muéstramelo
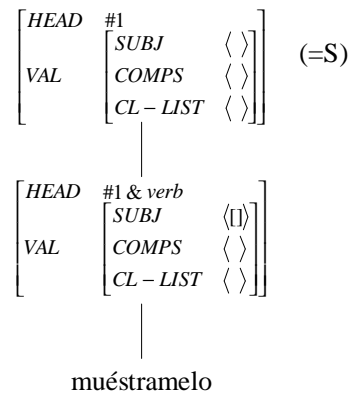
Figure 4. Analysis of enclitic sentences

The first approach to the analysis of syntactic cancellation is equally simple. In this case, clitization is performed by a lexical rule that modifies the valence of the verb by consuming the clitic arguments from the complement list, moving them into the clitic-list attribute, defining in this way a non-empty local clitic domain. The clitization rule is defined in Figure 5.

$$
\begin{bmatrix} ORTH & \#1 \\ HEAD & \#2\ \&\ \begin{bmatrix} verb \\ FORM & fin \end{bmatrix} \\ VAL|CL-LIST & \langle\ \rangle \\ ARG-ST & \langle\#3,\#4,\#5\rangle \end{bmatrix} \mapsto \begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \\ VAL|CL-LIST & \langle\#5,\#4\rangle \\ ARG-ST & \langle\#3\rangle \end{bmatrix}
$$
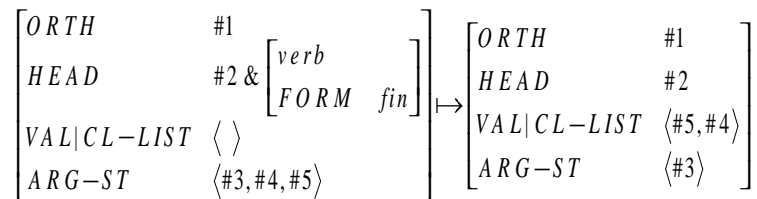
Figure 5. Simple clitization rule

In *SC,* clitic insertion is a process of lexical realization, and syntactic cancellation is performed by the *Head-Proclitic Rule* (*HPR*). This rule combines a head verbal phrase (with a saturated complement list) with a clitic word, provided the clitic-list attributes of the head and the clitic word are the same. The definition of the *HPR* is shown Figure 6.

$$
\begin{bmatrix} HEAD|CLIT & true \\ VAL|CL-LIST & \langle\ \rangle \end{bmatrix} \rightarrow \begin{bmatrix} HEAD & clitic \\ VAL|CL-LIST & \#1 \end{bmatrix} H \begin{bmatrix} VAL & \begin{bmatrix} SUBJ & \langle[I]\rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \#1 \end{bmatrix} \end{bmatrix}
$$

Figure 6. Head Proclitic Rule

The value *true* in the head attribute *CLIT* of the left-side of the rule marks that the periphrasis has been combined with its clitic word, and the rule can be applied only once.

The analysis of *me lo muestras* (show it to me) is shown in Figure 7. The clitic word is realized lexically and *SC* is performed by the *HPR*. The final sentence is produced by means of the *HPDR*.

$$
\begin{bmatrix} HEAD & \#1 \\ VAL & \begin{bmatrix} SUBJ & \langle\ \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix} \end{bmatrix} \quad (=S)
$$

$$
\begin{bmatrix} HEAD & \#1\ \&\ [CLIT\quad true] \\ VAL & \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix} \end{bmatrix} \quad (=VP)
$$

$$
\begin{bmatrix} HEAD & clitic \\ VAL\,|\,CL-LIST & \#a \end{bmatrix}
\qquad
\begin{bmatrix} HEAD & \#1\ \&\ verb \\ VAL & \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \#a\langle NP-dat, NP-acc\rangle \end{bmatrix} \end{bmatrix}
$$

me lo   muestras
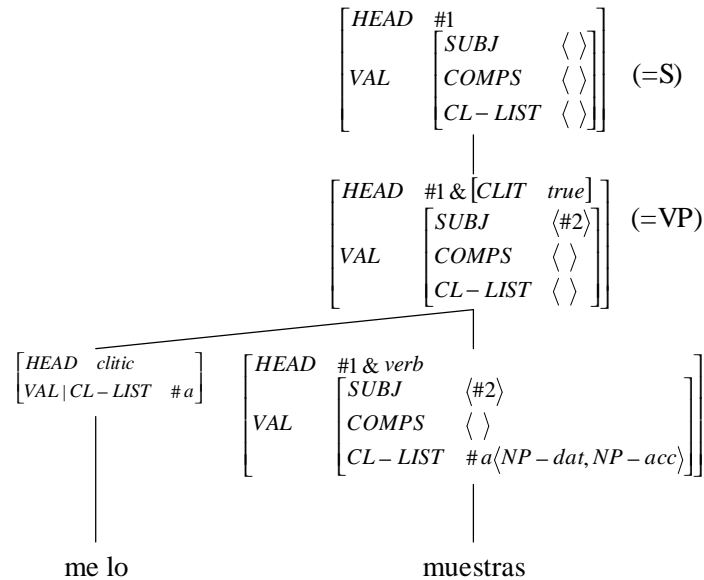
Figure 7. Analysis of proclitic sentence[1]

We turn now to the analysis of simple clitic climbing in which all clitics are arguments of a single content verb, normally at the end of the periphrasis. Although the study of the Spanish periphrasis is a complex topic that we cannot deal in detail here, we provide a summary of its most salient properties, which we use for our clitic model. First, auxiliary verbs in Spanish are subject-raising (Pineda & Meza, 2003, Meza & Pineda, 2002). However, these verbal forms can also appear with their original semantic content, and in this latter reading behave like subject-control verbs. For instance, the sentences *me lo puedes mostrar* and *puedes mostrármelo* (can you show it to me) are ambiguous according to whether *puedes* is interpreted as an auxiliary, implying the possibility of showing something (it is possible to show something?), or as a verb of capacity (are you able to show me something?), in which there needs be an agent capable of performing the showing action. In Spanish, this kind of ambiguity appears often, and must be resolved at a pragmatic level.

Auxiliary verbs can also subcategorize for an infinitive, a gerund or a participle VP. Auxiliaries with incoative meaning (*ir a*), approximative (*llegar a*), terminative (*venir a*), possibility (*poder*), reiterative (*volver a*), obligative (*deber de*), hypothetical (*haber de*) and perfective (*acabar de*, *alcanzar a*) normally require an infinitive complement; auxiliaries with a stative (*estar*) or continuative meaning (*ir*, *venir*, *seguir*, *andar*) normally subcategorize for a gerund; finally, the copulative verbs *ser* and *estar* (to be) require a participle to form the passive voice, as well as the verb *haber* (to have), which is used in the complex conjugation. However, from the syntactic point of view, all of these verbs are subject-raising, and they only differ in the kind of complement they require, and in the semantic-aspectual information that modify their complements.

Auxiliaries cannot be clitizised as they do not have nominal complements, but their lexeme definitions include a clitic-list attribute that constitutes their local clitic domains. The definition of subject-raising (and also subject-control verbs) specifies that these verbs share their clitic-list with their VP-complements. In particular, we define the accessibility relation between local clitic domains as follows:

(43) *Clitic Accessibility relation*: Local clitic domains are accessible if they share their clitic-list attribute.

Consequently, a clitic domain is created by all periphrastic verbs that share their clitic list attribute. We also define the scope of a clitic form as follows:

(44) *Clitic scope*: A local clitic domain is within the scope of a phonological host if their clitic domains are accessible.

Alternatively, a clitic host is within the scope of the phonological host if they belong to the same clitic domain. Also, the local clitic domain of the clitic host is within the scope of the pronoun, as cancellation (either lexical or syntactic) combines the pronouns with the phonological host directly (i.e. pronoun and phonological host are accessible).

With these definitions we proceed to the analysis of simple clitic climbing. The definition of auxiliary verbs is shown in Figure 8, and the analysis of a clitic sentence *se lo puedes mostrar* in Figure 9. In order to prevent that the complement of the auxiliary has already been combined with a clitic word, the value of the *CLIT* attribute of the complement is specified as *false*, ruling out expressions like *puedes me lo mostrar*. A similar mechanism will be used for subject-control and object-control verbs.

$$
\begin{bmatrix}
srv-lxm \\
\\
SS\,|\,LOC\,|\,CAT \begin{bmatrix} SUBJ & \#1 \\ \\ COMPS & \left\langle \begin{bmatrix} HEAD\,|\,CLIT & false \\ \\ VAL & \begin{bmatrix} SUBJ & \#1 \\ COMPS & <> \\ CL-LIST & \#a \end{bmatrix} \end{bmatrix} \right\rangle \\ \\ CL-LIST & \#a \end{bmatrix}
\end{bmatrix}
$$

Figure 8. Lexical entry for auxiliary verbs

$$\begin{bmatrix} HEAD & \#1 \\ VAL & \begin{bmatrix} SUBJ & \langle\,\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \end{bmatrix} \quad (=\text{S})$$

|

$$\begin{bmatrix} HEAD & \#1 \\ VAL & \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \end{bmatrix} \quad (=\text{VP})$$

$$\begin{bmatrix} HEAD & clitic \\ VAL|CL-LIST & \#a \end{bmatrix} \qquad \begin{bmatrix} HEAD & \#1 \\ VAL & \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \#a \end{bmatrix} \end{bmatrix} \quad (=\text{VP})$$

$$\begin{bmatrix} HEAD & \#1\,\&\,verb \\ VAL & \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\#3\rangle \\ CL-LIST & \#a \end{bmatrix} \end{bmatrix} \qquad \#3 \begin{bmatrix} HEAD & verb \\ VAL & \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \#a\langle NP-dat, NP-acc\rangle \end{bmatrix} \end{bmatrix}$$

se lo   puedes   mostrar

Figure 9. Analysis of simple clitic climbing

Similarly, modals like *poder* (be able to), *querer* (to want), *deber* (must), *osar* (to dare), *pensar* (to think), *esperar* (to expect), *lograr* (to manage), *saber* (to know), appear in periphrasis. The study of the sucategorization requirements for this kind of verbs is quite complex and beyond the scope of this paper; however, they behave as transitives: *quiere un pastel* (she wants a cake), subcategorize for a direct object infinitival complement: *quiere comer* (she wants to eat), or for a nominal relative phrase, which can substitute the nominal argument, like *quiere que le den de comer* (she wants to be feed). This direct complement can normally be clitized as well: *lo quiere* (the cake); also, in the question *¿que es lo que quiere*? (what's does he/she want?), the clitic substitutes *comer* and also *que le den de comer*, respectively. Modals can also be followed by an infinitive VP like *poder comer* in *quiere poder comer* (she wants to be able to eat); in this case the agentive import of the modal selects the capacity meaning of *poder*, removing the potential ambiguity.

Here we do not address the case that the modal subcategorize for a relative complement, and concentrate on the case where it requires an explicit nominal complement, which is a simple transitive verb, and the case where it subcategorizes for a VP complement; in this latter case, a modal behaves as a subject-control verb, takes an infinitival verbal phrase as its complement, and the verb and complement share their subject-agent. In addition, the subject-control verb shares its clitic-list with its complement, establishing the accessibility path between the phonological host and the host's clitization. The definition of subject-control verbs is shown in Figure 10.

$$
\begin{bmatrix}
scv-lxm \\[6pt]
SS\,|\,LOC\,|\,CAT\,|\,VAL
\begin{bmatrix}
SUBJ & \#1_i \\[4pt]
COMPS & \left\langle
\begin{bmatrix}
HEAD\,|\,CLIT & false \\[4pt]
VAL &
\begin{bmatrix}
SUBJ & \#1_i \\
COMPS & \langle\ \rangle \\
CL-LIST & \#a
\end{bmatrix}
\end{bmatrix}
\right\rangle \\[4pt]
CL-LIST & \#a
\end{bmatrix}
\end{bmatrix}
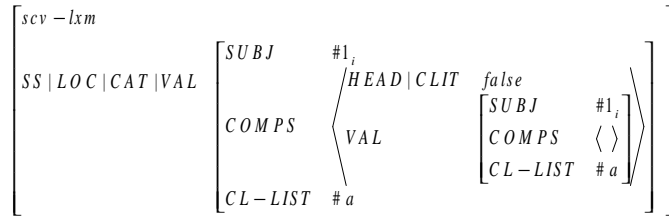$$

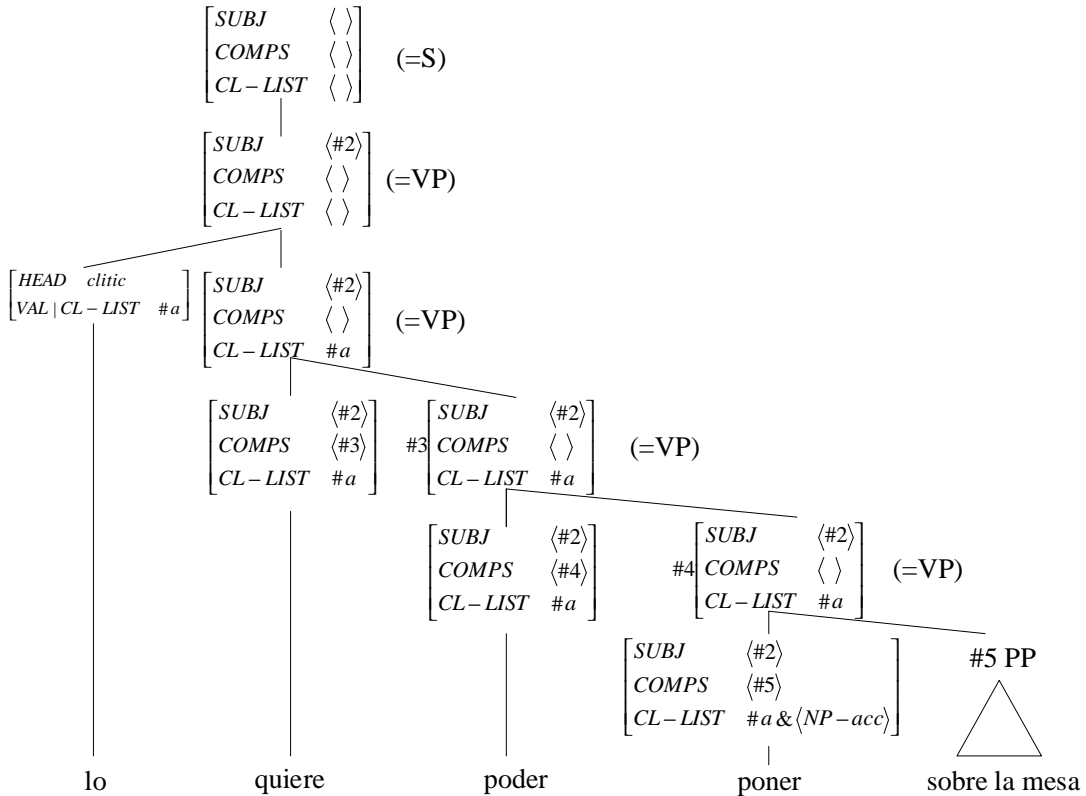Figure 10. Lexical entry for modal verbs

Figure 11. Simple periphrasis involving a sequence of subject-control verbs

In addition to auxiliary verbs and modal verbs, most forms that are realized as auxiliary verbs are also realized as agentive verbs preserving their original semantic content, as the case of *poder* already mentioned[2]. In this case, these verbs become subject-control, as they require an agent, which is shared with the subject of their VP-complements, and the semantics of these verbs specifies the relation between the subject-agent, and the action undertaken by such agent. Similarly, their clitic-list attributes is shared with their VP-complements. With these elements at hand, it is possible to analyze simple periphrasis involving a sequence of verbs; an instance of this kind of periphrasis is the sentence

---

[2] The verb *haber* is an exception in this respect, as it used for the complex conjugation in the indicative: *he amado* (actual-perfect past: I have loved), *había amado* (pluscuamperfect: I had loved), *hube amado* (past perfect-past (antepresente): I have had loved), *habré amado* (future perfect-past (antefuturo): I will have loved), *habría amado* (hypothetical future: I would had loved) and also in the subjunctive: *haya amado* (perfect past: I had loved), *hubiera o hubiese amado* (pluscuamperfect: I would had loved), *hubiere amado* (hypothetical future perfect-past: I would had loved). When the original possessive meaning is needed the verb *tener* is used instead.

quiere poder poner el sombrero sobre la mesa (he wants to be able to put the hat on the table) and its clitic version *lo quiere poder poner sobre la mesa* (he wants to be able to put it on the table). In this case, the clitic-list of *quiere* is accessible to the clitic-list of *poner* as they share their clitic-list attribute through the clitic-list of *poder*, which establishes the required binding path. The analysis of these sentences are shown in Figure 11.

We can now provide the analysis for composite lexical cancellation (*C-LC*). In this case, a simple clitization rule establishes the basic clitic domain, but clitic insertion and cancellation are performed by a second lexical rule that operates on the phonological host. This rule is subject to the lexical constraint that tensed verbs and participles cannot have enclitics, and the pronouns must be realized attached to an infinitive or a gerund as enclitics. The *C-LC* lexical rule for the infinitive case is shown in Figure 12.

$$
\begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \,\&\, [FORM \quad inf] \\ ARG-ST & \langle \#3,\#4 \rangle \end{bmatrix} \mapsto \begin{bmatrix} ORTH & F_{-selo}(\#1) \\ HEAD & \#2 \\ CL-LIST & \langle\ \rangle \\ ARG-ST & \langle \#3,\#4 \,\&\, [CL-LIST \quad \langle NP-dat, NP-acc \rangle] \rangle \end{bmatrix}
$$

Figure 12. Lexical rule for *C-LC*

$$
\begin{bmatrix} SUBJ & \langle\ \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix}\ (=S)
$$

$$
\begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix}\ (=VP)
$$

$$
\begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle \#3 \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix}\ \#3 \begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix}\ (=VP)
$$

$$
\begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle \#4 \,\&\, [CL-LIST \quad \#a] \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix}\ \#4 \begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \#a \end{bmatrix}\ (=VP)
$$

$$
\begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle \#5 \rangle \\ CL-LIST & \#a \end{bmatrix}\ \#5 \begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \#a \langle NP-dat, NP-acc \rangle \end{bmatrix}
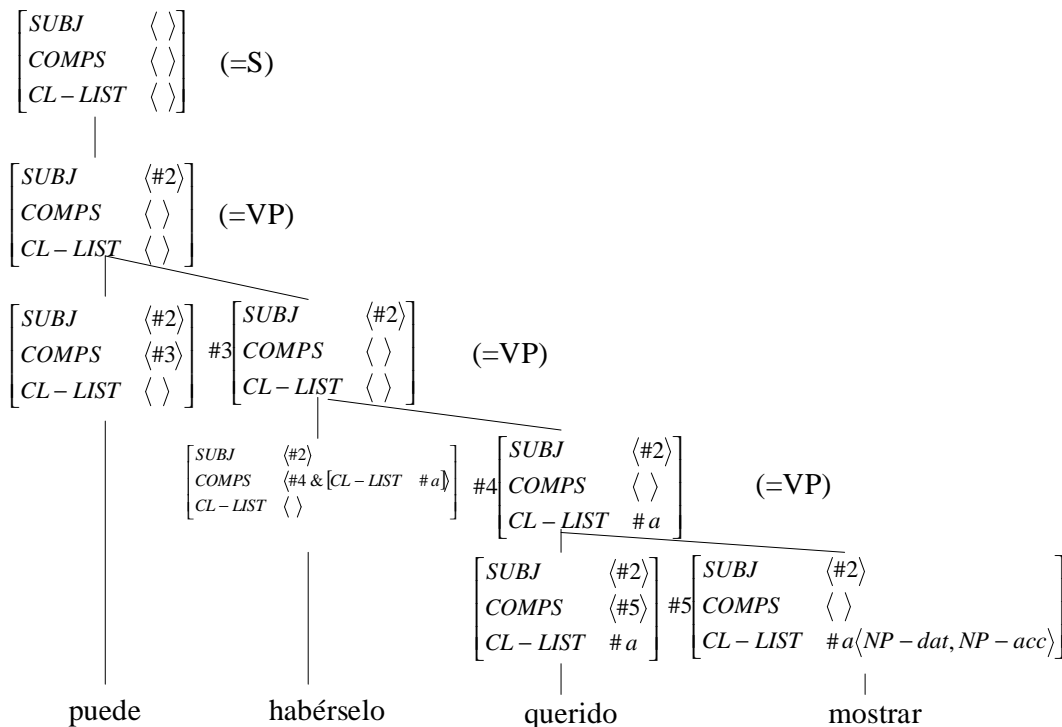$$

puede   habérselo   querido   mostrar

Figure 13. Analysis of sentence with *C-LC*

With these elements we can proceed to the analysis of simple periphrasis with several verbs, where the pronouns appear as enclitic in the middle of the periphrasis. In Figure 13, the analysis of *puede habérselo querido mostrar* (she could have wanted to show it) is

illustrated. The lexical restriction on tensed forms and participles rules out *puédelo haber querido mostrar* and *puede haber querídoselo mostrar* which are ungrammatical.

With this we complete the analysis of simple clitic climbing. As can be seen, in all the examples shown for *S-LC*, *C-LC* and *SC* the clitic principle is satisfied: one clitization, one clitic insertion and one cancellation per independent clitic domain. Also, there are no other constraints on clitic climbing but the lexical constraints on tensed forms and participles. Notice also that clitic cancellation requires that the complement list is saturated (this is implicit in *S-LC* and explicit in the definition of *C-LC* and *SC*), so all overt complements must be combined with the clitic host before clitic insertion; as a direct consequence, no cancellation mode allows the presence of lexical material between the pronouns and the phonological host, and in this respect, the model is consistent with the observations that motivate the lexical approaches to clitization. In our model, clitic insertion in *S-LC* and *C-LC* are also realized at the morphological level, and although in *SC* there is a syntactic operation involved, this is hidden within the analysis, and there are no traces of it from the external point of view.

We turn now to the analysis of complex periphrasis in which abstract local clitic domains are formed by clitization of different verbs, and to the definition of the clitic composition and subsumption operations. The first observation is that for the formation of abstract clitic domains there must be more than one verb with a full semantic content in the periphrasis, with clitizisable complements. In our analysis we have observed that this phenomenon occurs when the so-called object-control verbs appear in the periphrasis. We have identified three main cases: the verbs of the senses, like *ver* (to see) and *oir* (to hear), the verbs of will, like *mandar* (to command), *permitir* (to allow) and *prohibir* (to forbid) and the factitive verb *hacer* (to make). In *ve comer la manzana a María* (he sees Maria to eat the apple) and *oye cantar a los pájaros* (she hears the birds to sing), the subject of the control verb and the subject of its infinitival complement are not the same: someone sees Maria to eat, and someone hears the birds to sing; but the direct object of the control verb is the subject of its VP-complement. Similarly in *manda marchar a los soldados* (he commands the soldiers to march), the one who commands and the one who marches are different, but the object of *mandar* is the subject of *marchar*. In *hace reir a la gente* (makes the people laugh) the direct object of *hace*, *la gente*, is also the subject of *reir* (to laugh). According to these observations and in a manner consistent with Sag and Wasow (1999), the object control verbs have two complements: a nominal direct object, and a verbal phrase headed by the verb at the end of the periphrasis. In addition, the subject of the verbal phrase (i.e. the second complement) is shared with the direct object of the control verb (i.e. the first complement). The definition of this kind of verbs is as shown in Figure 14[3]:

---

[3] From a semantic perspective there is an alternative approach where these verbs have a single VP-complement. In this latter view, the object of *ver* is the act of Maria eaten the apples and the direct object of *oir* is the action of the birds singing, which is probably more intuitive from the one that we present here; however, our approach seems to be the appropriate one for verbs of the will in which the action of the control verb is received by an individual (the one is commanded, or allowed or forbidden); but for the factitive verb *hacer*, the second alternative seem perhaps more appropriate: the sense in which people laughing is the complement of factitive verbs seems a much more intuitive analysis than the sense in which

$$\begin{bmatrix} ocv-lxm \\ SS\,|\,LOC\,|\,CAT\,|\,VAL \quad \begin{bmatrix} SUBJ & \langle[\ ]\rangle \\ COMPS & \left\langle \#1\ \&\ NP-acc_i,\ \begin{bmatrix} SUBJ & \#1_i \\ COMPS & \langle\ \rangle \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Figure 14. Definition of object control-verbs

To capture clitization a clitic-list attribute is also included in the valence of both of the verbs. In particular, the direct object of the object-control verb and the complements of the verb at the end of the periphrasis can be subject to clitization. Also, the clitizations of different local clitic domains can be composed and climb as shown in (36). Here, we define the clitic composition as a lexical property of the object-control verb in such a way that its clitic-list is the concatenation of its own clitizations with the clitic-list of its VP-complement. The basic definition of the object-control verb lexeme including clitization is shown in Figure 15, where #a is an empty list.

$$\begin{bmatrix} ocv-lxm \\ SS\,|\,LOC\,|\,CAT\,|\,VAL \quad \begin{bmatrix} SUBJ & \langle[\ ]\rangle \\ COMPS & \left\langle \#1_i,\ \begin{bmatrix} HEAD\,|\,CLIT & true \\ VAL & \begin{bmatrix} SUBJ & \#1_i \\ COMPS & \langle\ \rangle \\ CL-LIST & \#a \end{bmatrix} \end{bmatrix} \right\rangle \\ CL-LIST & \#a \end{bmatrix} \end{bmatrix}$$

Figure 15. Definition of object control-verbs (version 2)

With this definition we can analyze the clitic versions of *Juan pudiera haber visto a María comer la manzana* as in (36) above. In *Se la pudiera haber visto comer* in (36.f), for instance, the accusative *se* stands for *María*, the direct object of *visto* and the accusative *la* stands for *la manzana*, the direct object of *comer*. The clitization of the *comer* places the direct object in its clitic-list attribute as usual; however, for the clitization of the object-control verb, a lexical rule that inserts the clitic argument at the front of its clitic-list attribute is required. This lexical rule also states that the clitic list of the verb is the concatenation of the clitization of its direct object with the clitic list of its second complement. This is the definition of the clitic composition operation. The lexical rule is shown in Figure 16.

$$\begin{bmatrix} ocv-lxm \\ ORTH & \#1 \\ HEAD & \#2 \\ ARG-ST & \langle\#3,[\ ]_i,\#4\rangle \end{bmatrix} \mapsto \begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \\ VAL\,|\,CL-LIST & \langle NP-acc_i\rangle \oplus \#a \\ ARG-ST & \langle\#3,\#4\ \&\ [CL-LIST\ \#a]\rangle \end{bmatrix}$$

Figure 16. Clitization lexical rule for object-control verbs

---

a clown makes the people, who happens to laugh. However, for uniformity of analysis, we adopt the two complements approach for the three kinds of verbs, and we leave the issue for further research.

When this rule is applied, the lexeme definition in Figure 15 is modified as shown in Figure 17.

$$
\begin{bmatrix}
ocv-lxm \\[4pt]
SS\,|\,LOC\,|\,CAT\,|\,VAL
\begin{bmatrix}
SUBJ & \langle[\ ]\rangle \\[4pt]
COMPS &
\left\langle
\begin{bmatrix}
SUBJ & \langle[\ ]_i\rangle \\
COMPS & \langle\ \rangle \\
CL-LIST & \#a
\end{bmatrix}
\right\rangle \\[10pt]
CL-LIST & \langle NP-acc_i\rangle \oplus \#a
\end{bmatrix}
\end{bmatrix}
$$

Figure 17. Object-control verb after clitization

The combination of the object-control verb with its complements by means of the *Head-Complement Rule* implements the clitic composition operation. This is illustrated with the analysis of sentence *Se la pudiera haber visto comer* in Figure 18. As can be seen, the local clitic domains of both the clitized verbs are summed up in the definition of the local clitic domain of the verbal phrase *visto comer* by the clitic composition operation. As this abstract clitic domain is accessible to the phonological host, syntactic cancellation is permitted.
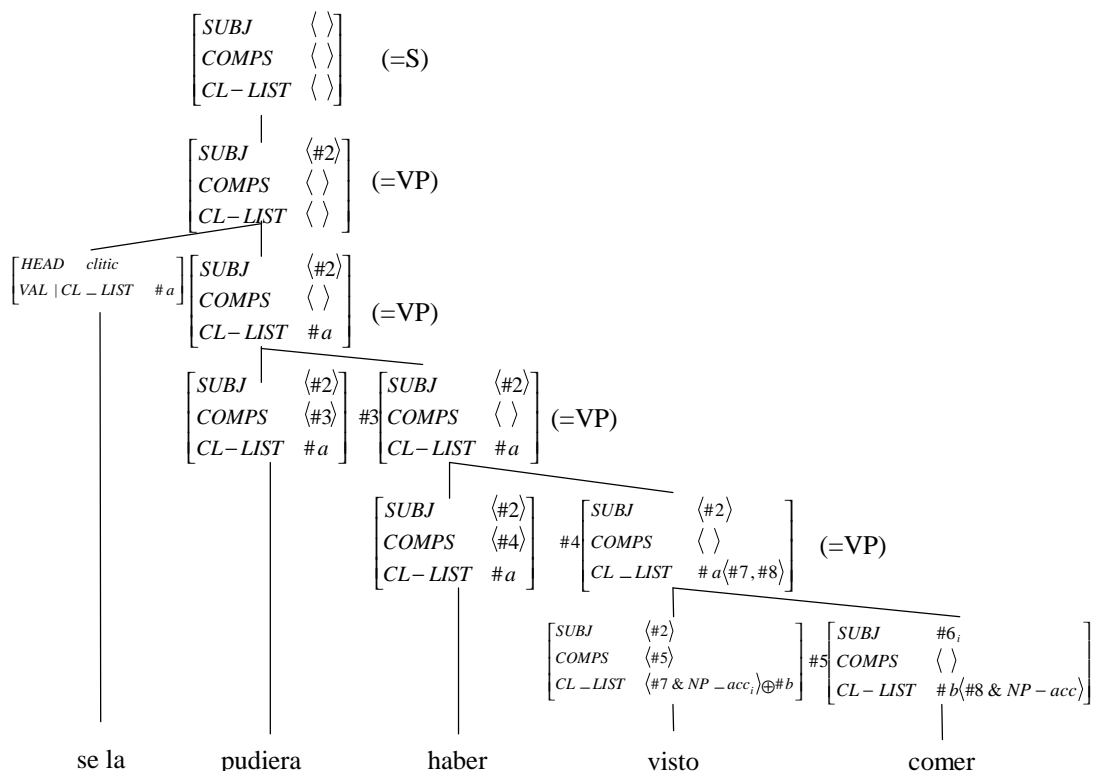

Figure 18. Analysis of sentence with clitic composition and *SC*

Similarly, if the pronouns are enclitic to the auxiliary, as in (36.d), the sentence is analyzed through complex lexical cancellation, as shown in Figure 19.
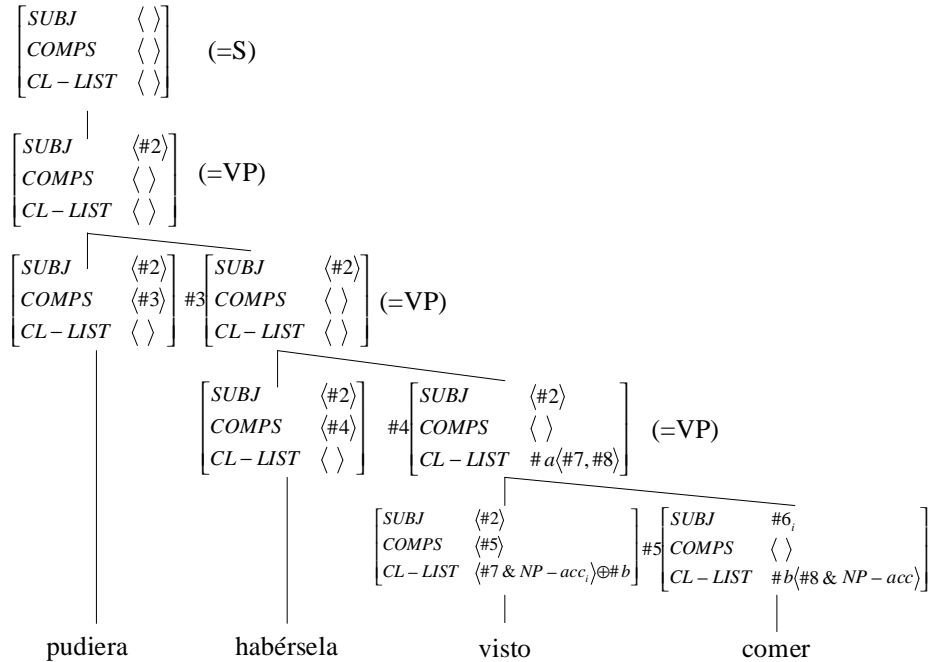
$$
\begin{bmatrix} SUBJ & \langle\,\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad (=\text{S})
$$

$$
\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad (=\text{VP})
$$

$$
\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\#3\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad \#3 \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad (=\text{VP})
$$

$$
\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\#4\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad \#4 \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \#a\langle\#7,\#8\rangle \end{bmatrix} \quad (=\text{VP})
$$

$$
\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\#5\rangle \\ CL-LIST & \langle\#7 \& NP-acc_i\rangle\oplus\#b \end{bmatrix} \quad \#5 \begin{bmatrix} SUBJ & \#6_i \\ COMPS & \langle\,\rangle \\ CL-LIST & \#b\langle\#8 \& NP-acc\rangle \end{bmatrix}
$$

pudiera   habérsela   visto   comer

Figure 19. Analysis of sentence with clitic composition and C-*LC*

The composition operation is also used when the verb at the end of the periphrasis has its complements as enclitics, as in *la ha visto comerla* in (9.c), although in this latter case the clitic list appended to the clitization of the control verb is empty. In addition, this kind of constructions have two independent clitic domains; the analysis of (9.c), in particular, requires one *S-LC* for *comerla* and one SC for *la ha visto*. The syntactic structure of this sentence is shown in Figure 20.

$$
\begin{bmatrix} SUBJ & \langle\,\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad (=\text{S})
$$

$$
\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad (=\text{VP})
$$

$$
\begin{bmatrix} HEAD & clitic \\ VAL\,|\,CL-LIST & \#a \end{bmatrix} \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \#a \end{bmatrix} \quad (=\text{VP})
$$

$$
\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\#3\rangle \\ CL-LIST & \#a \end{bmatrix} \quad \#3 \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \#a\langle\#5\rangle \end{bmatrix} \quad (=\text{VP})
$$

$$
\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\#4\rangle \\ CL-LIST & \langle\#5 \& NP-acc_i\rangle\oplus\#b \end{bmatrix} \quad \#4 \begin{bmatrix} SUBJ & \langle[]_i\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \#b\langle\,\rangle \end{bmatrix}
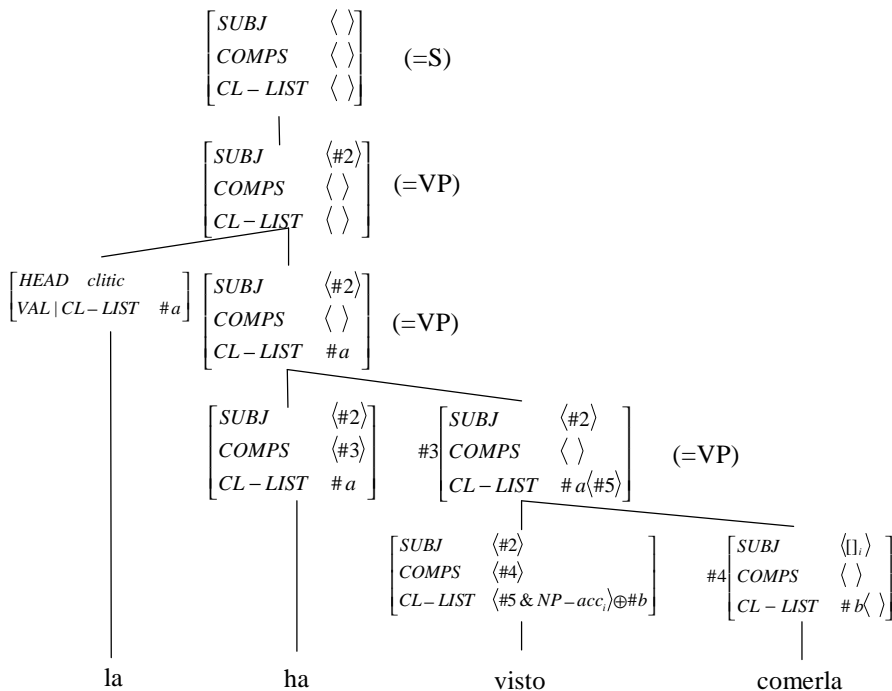$$

la   ha   visto   comerla

Figure 20. Sentence with two independent clitic domains

The analysis of clitic subsumption is developed along similar lines. In *pudiera haber visto a María comerse la manzana* and its clitic version *se la pudiera haber visto comer* in (40.h), for instance, *a María* is not only the object of *visto* and the subject of *comer* but also she is the beneficiary of the eating action, given rise to the ethical dative. As can be seen in (40.c) and (40.d) the two clitizations form two accessible clitic domains that can be combined to form an abstract clitic domain; however, unlike the composition combination in (36), the sequence in (40) involves two correferential instances with a different case that are subsumed into one, with the dative case prevailing. We define the subsumption operation as a lexical property of object-control verbs as well. For the analysis we use the lexeme definition in Figure 15, which is also the subject of the subsumption lexical rule. This rule creates an abstract local clitic domain by subsuming the accusative clitic of object-control verb (e.g. *ver*) into the dative clitic of second verb (e.g. *comer*) if these two arguments are co-indexed; through this effect, the clitic-list of the abstract clitic domain is simply the clitic list of the VP-complement of the object-control verb. The definition of the subsumption lexical rule is presented in Figure 21. The analysis of (40.h) is shown in Figure 22.
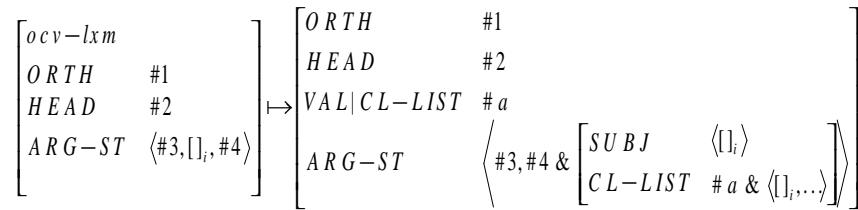

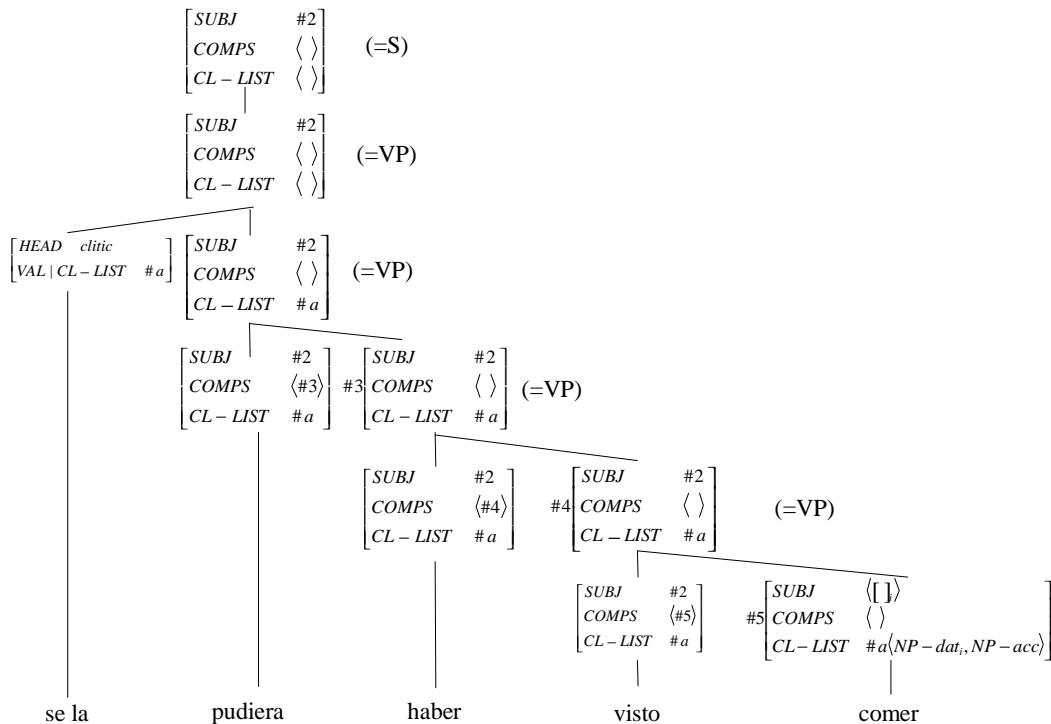Figure 21. Subsumption lexical rule

Figure 22. Analysis of sentence with clitic subsumption

We turn now to the analysis of construction in which the clitic pronouns have a wider scope over coordination. In *lo llevó y puso sobre la mesa*, the two verbs undergo independent clitizations and form two different local clitic domains; however, the arguments are co-indexed and can be combined by the subsumption operation, which is also defined in the *Coordination Rule*, forming an abstract local clitic domain for the whole conjunction; the combination of the pronoun and the abstract clitic host is produced by *SC*. Although the analysis of coordinated structures including clitic pronouns is a very complex phenomenon that is beyond the scope of this paper, we suggest the Coordination Rule for verbal phrases in Figure 23 as an starting point; the analysis of a conjunctive clitic sentence in terms of this rule is shown in Figure 24.

$$\begin{bmatrix} HEAD & \#1 \& verb \\ VAL & \#2 \end{bmatrix} \rightarrow \begin{bmatrix} HEAD & \#1 \\ VAL & \#2 \end{bmatrix} \begin{bmatrix} HEAD & \#1 \\ VAL & \#2 \end{bmatrix} \dots \begin{bmatrix} HEAD & conj \end{bmatrix} \begin{bmatrix} HEAD & \#1 \\ VAL & \#2 \end{bmatrix}$$

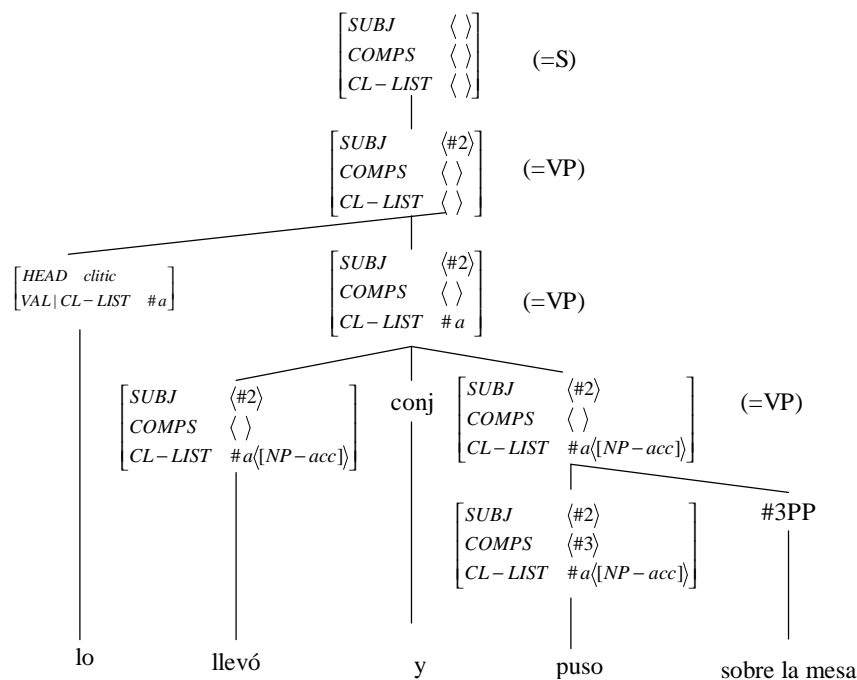Figure 23. Coordination Rule with clitic subsumption

Figure 24. Analysis of clitic coordinated sentence

With the presentation of the three cancellation modes and the clitic composition and subsumption operations we conclude the presentation of the model. The analysis for pleonastic constructions and reflexives follows straightforwardly.

In case of pleonastic constructions, like *se$_j$ lo$_i$ muestra a María$_i$* in (1.g), dative arguments are not removed from the complement list, and the corresponding instances in the clitic and complement lists are co-indexed. Accusative arguments can also be duplicated, unless they are subject to specific constraints (i.e. non-animate direct objects). As the complement list must be saturated for clitic cancellation to take place, pleonastic complements are consumed before clitic insertion, and they can never appear between a

clitic pronoun and its phonological host. In Figure 25, a lexical rule for pleonastic clitization is presented, and the pleonastic clitization of the *muestra* is shown in Figure 26; the analysis of (1.g) is shown in Figure 27.

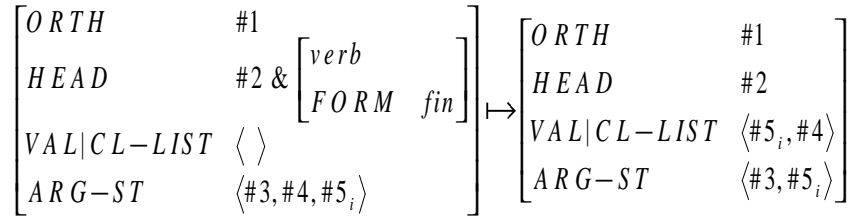$$\begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \ \& \begin{bmatrix} verb \\ FORM & fin \end{bmatrix} \\ VAL|CL-LIST & \langle\ \rangle \\ ARG-ST & \langle \#3, \#4, \#5_i \rangle \end{bmatrix} \mapsto \begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \\ VAL|CL-LIST & \langle \#5_i, \#4 \rangle \\ ARG-ST & \langle \#3, \#5_i \rangle \end{bmatrix}$$

Figure 25. Lexical rule for pleonastic clitization

$$\begin{bmatrix} mostrar \\ HEAD & verb \\ VAL & \begin{bmatrix} SUBJ & \langle \#1 \rangle \\ COMPS & \langle \#2_i \rangle \\ CL-LIST & \langle \#2_i \ \& \ NP-dat, NP-acc \rangle \end{bmatrix} \end{bmatrix}$$

Figure 26. Pleonastic clitizised lexeme

$$\begin{bmatrix} SUBJ & \langle\ \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix} \quad (=S)$$

$$\begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix} \quad (=VP)$$

$$\begin{bmatrix} HEAD & clitic \\ VAL\,|\,CL-LIST & \#a \end{bmatrix} \quad \begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \#a \end{bmatrix} \quad (=VP)$$

$$\begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle \#3_i \ \& \ NP-dat \rangle \\ CL-LIST & \#a \langle \#3_i, NP-acc \rangle \end{bmatrix} \qquad \#3NP$$

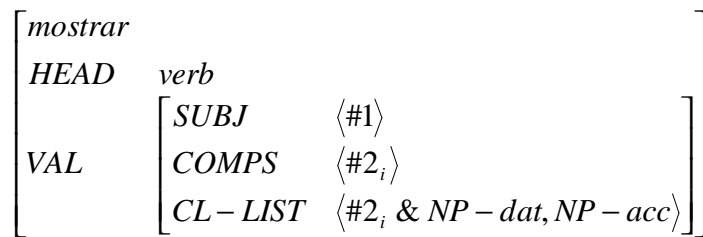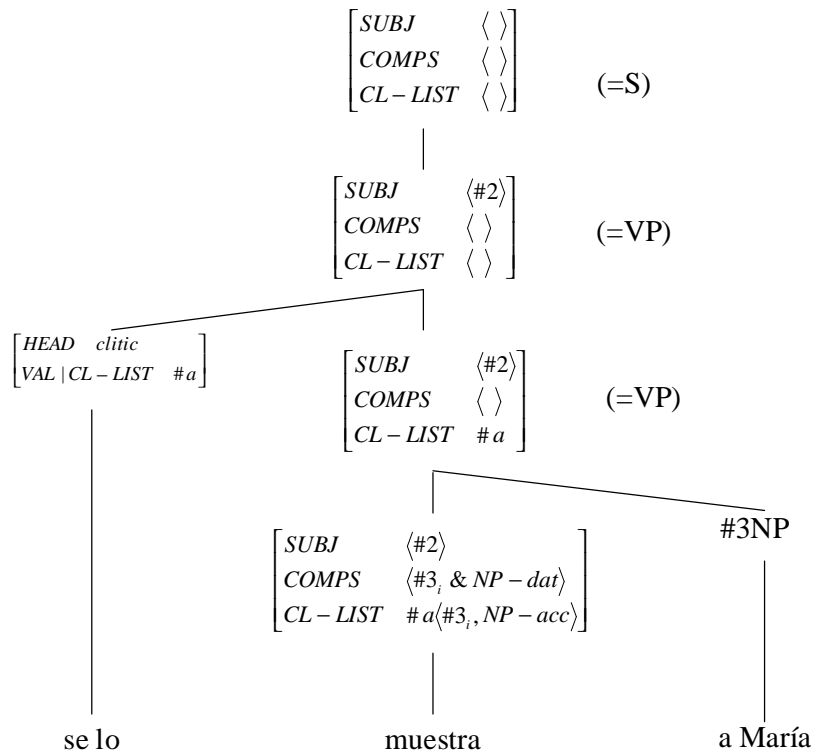se lo　　　　　　　　　muestra　　　　　　a María

Figure 27. Analysis of pleonastic clitic sentence

For the analysis of reflexives we use the clitic machinery developed so far. The analysis of the ethical dative by means of clitic subsumption has already been shown. For the direct reflexive in (11.b) *Me$_i$ lavo$_i$* (I wash myself), a the clitization rule co-indexes, in addition, the accusative complement with the nominative subject (implicit in the

conjugation), and imposes the additional constraint that the number, person and gender of subject and object must agree. Also, the pronouns must be proclitic, and cancellation is syntactic. This lexical rule is shown in Figure 28, and the analysis of *me lavo* is shown in Figure 29. The analysis of the indirect reflexive in (11.c) *me lavo las manos* (I wash my hands) is carried on along similar lines.

$$
\begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \, \& \, \begin{bmatrix} verb \\ FORM & fin \end{bmatrix} \\ VAL|CL-LIST & \langle \, \rangle \\ ARG-ST & \langle \#3_i, [\,]_i \rangle \end{bmatrix} \mapsto \begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \\ VAL|CL-LIST & \langle [AGR \quad \#4]_i \rangle \\ ARG-ST & \langle \#3_i \, \& \, [AGR \quad \#4]_i \rangle \end{bmatrix}
$$

Figure 28. Lexical rule for accusative reflexive

$$
\begin{bmatrix} SUBJ & \langle \, \rangle \\ COMPS & \langle \, \rangle \\ CL-LIST & \langle \, \rangle \end{bmatrix} \quad (=S)
$$

$$
\begin{bmatrix} SUBJ & \langle \#2 \rangle \\ COMPS & \langle \, \rangle \\ CL-LIST & \langle \, \rangle \end{bmatrix} \quad (=VP)
$$

$$
\begin{bmatrix} HEAD & clitic \\ VAL|CL-LIST & \#a \end{bmatrix} \qquad \begin{bmatrix} SUBJ & \langle \#2 \, \& \, [AGR \quad \#3] \rangle \\ COMPS & \langle \, \rangle \\ CL-LIST & \#a \langle [AGR \quad \#3] \rangle \end{bmatrix}
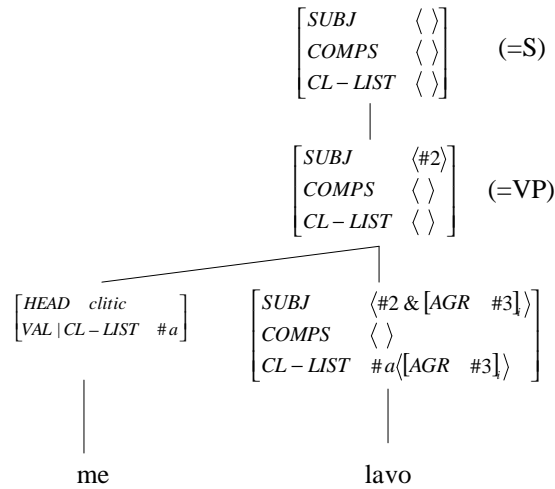$$

me   lavo

Figure 29. Analysis of accusative reflexive sentence

The ethical datives in *te construiste una casa* in (12.a) and *se tomó un café* in (12.b) are also analyzed in similar lines, as the pronouns *te* and *se* substitute the indirect complements of *construir* (to build) and *tomar* (to drink) respectively. As these verbs are normally transitives, the clitization rule adds the extra dative argument, and marks the beneficiary interpretation in the semantics. In the case of pseudo-reflexives like *me voy* in (12.c), the clitization rule simply adds an extra argument in the clitic list of the intransitive *voy*, unmarked for case and without semantic content but co-indexed with the subject, and the sentence receives an analysis similar to the direct reflexive. In verbs like *arrepentirse* in (12.b), that cannot be used without the reflexive import, clitization is specified in the basic lexeme definition of the verb. Attribute sentences with the complements of predicate adjectives clitized (e.g. *le es fiel*) receive an analysis along similar lines.

With this we conclude the presentation of the model. All clitic sentences that we have observed satisfy the clitic principle. As a general observation from the analysis, its seems that the single unifying factor structuring the whole of the phenomena is that there is only one phonological host for every independent clitic domain. Also, clitic climbing is

allowed as long as the clitic host is accessible to or within the scope of its corresponding phonological host. This seems to be a very economical constraint imposed by the phonological level of representation upon the syntactic one. In addition to this structural constraint, the inclusion of clitic words in the clitic lexicon seems to depend on common clitizations that can attach to phonological hosts, and the holes in the clitic paradigm seem to depend on phonological factors, and also on the possible referential conflicts that the presence of certain clitic words would permit. Clitic climbing is prevented only by lexical constraints that seem to have a contingent character and are related to the current linguistic use. Finally, the set of clitic words used in a dialect, and the lexical constraint on clitic realization, can help to explain why the clitic system is the locus of much dialectal variation in Spanish, and also in other romance languages that present the phenomena of pronominal clitics.

5. Impersonal forms

In this section the analysis of the passive-reflexive and the impersonal-active is presented. The analysis of these forms also uses the clitic machinery and conforms to the clitic principle. However, in both of these impersonal constructions, the *se* becomes a simple syntactic marker and has no referential import, very much like the expletive English *it*; for this reason, the semantic mode in the semantics of this word is *none*,

The passive-reflexive is also handle with a lexical rule that changes the valence attribute of the verb. The form is a passive sentence, and it can only appear with transitive verbs with a patient subject. The passive-reflexive lexical rule removes the direct object from the complement list and places it as the subject, with the corresponding agreement attributes. In this operation the original subject is discharged, as the construction is fully impersonal. The lexical rule also includes the syntactic marker *se* in the clitic list of the host; however, this clitic word is not specified for case, as it is not referential. As the passive-reflexive is only defined for third-singular persons, the *se* word is marked for agreement too. Finally, the lexical rule defines the morphology of the corresponding inflexion of the host. The clitic word for the expletive *se* is shown in Figure 30, the lexical rule for the passive-reflexive in 31, and the analysis of the sentence *la puerta se abre* in Figure 32. As can be seen, the expletive *se* is consumed by a process of syntactic cancellation. Finally, in Figure 33, the semantics of the sentence is presented.

$$\begin{bmatrix} se \\ \\ SS \mid LOC \begin{bmatrix} CAT & \begin{bmatrix} HEAD & clitic \\ VAL \mid CL-LIST & \langle [AGR \quad imper] \rangle \end{bmatrix} \\ CONT & \begin{bmatrix} INDEX & none \\ MODE & none \\ RESTR & \{\} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 30. Lexeme definition of the expletive *se*

$$
\begin{bmatrix} ORTH & \#1 \\ HEAD & \#2\ \&\ verb \\ VAL|CL-LIST & \langle\ \rangle \\ ARG-ST & \langle[\,],\#3\rangle \end{bmatrix} \mapsto \begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \\ VAL|CL-LIST & \langle[AGR\ \ imper]\ \rangle \\ ARG-ST & \langle\#3\rangle \end{bmatrix}
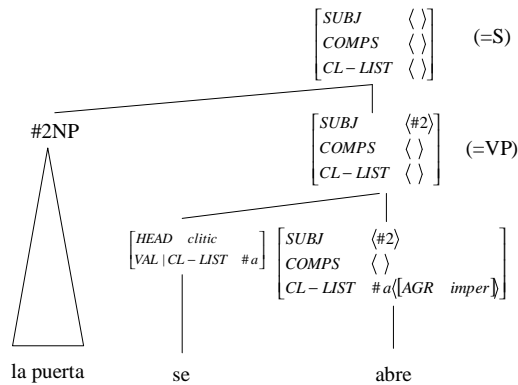$$

Figure 31. Lexical rule for the passive-reflexive

$$
\begin{bmatrix} SUBJ & \langle\ \rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix}\quad (=S)
$$

#2NP

$$
\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \langle\ \rangle \end{bmatrix}\quad (=VP)
$$

$$
\begin{bmatrix} HEAD & clitic \\ VAL\,|\,CL-LIST & \#a \end{bmatrix} \quad \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\ \rangle \\ CL-LIST & \#a\langle[AGR\ \ imper]\rangle \end{bmatrix}
$$

la puerta   se   abre

Figure 32. Analysis of the passive-reflexive sentence

$$
\begin{bmatrix} SS\,|\,LOC\,|\,CONT & \begin{bmatrix} INDEX & s \\ MODE & prop \\ RESTR & \left\{\begin{bmatrix} door-rel \\ INST & i \end{bmatrix}\begin{bmatrix} open-rel \\ SIT & s \\ ACT & undef \\ UND & i \end{bmatrix}\right\} \end{bmatrix} \end{bmatrix}
$$

Figure 33. Semantics of the passive-reflexive

The impersonal-active is analyzed along the same lines. This latter form is defined for transitive, intransitive and also stative verbs. The form is also fully impersonal and the expletive *se* in Figure 30 is also used. Although the forms looks very similar to the passive-reflexive, specially when the verb is transitive, this latter form is active, and its explicit argument is the direct object. In *se habla a los niños* (one talks to the children), the nominal *a los niños* is the direct object and can be clitizised (i.e. *se les habla*), even pleonastically as in *se les habla a los niños*. However, this duplication is not possible if the object is not animate: *se les$_i$ vende libros$_i$*, the pleonastic form of *se vende libros* and *se les vende* (one sells books), is not grammatical. The combination of the clitic with the verbal phrase is produced by the *HPR* as usual. The lexical rule for this form also drops the subject, and the subject implicit in the conjugation (3rd-singular) is consumed by the Subject-PRO-Drop rule. The lexical entry for the clitic word "*se les*" is shown in Figure 34, the lexical rule for the impersonal-active for transitive verbs with the direct object clitized is shown in 35, and the analysis of *se les habla* in 36.
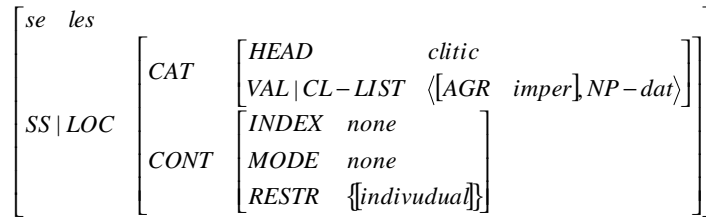
$$\begin{bmatrix} se & les \\ & \\ SS\,|\,LOC & \begin{bmatrix} CAT & \begin{bmatrix} HEAD & clitic \\ VAL\,|\,CL-LIST & \langle [AGR \quad imper],NP-dat\rangle \end{bmatrix} \\ & \\ CONT & \begin{bmatrix} INDEX & none \\ MODE & none \\ RESTR & \{[indivudual]\} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 34. Lexical entry for the clitic word

$$\begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \,\&\, verb \\ VAL\,|\,CL-LIST & \langle\,\rangle \\ ARG-ST & \langle[],\#3,\#4\rangle \end{bmatrix} \mapsto \begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \\ VAL\,|\,CL-LIST & \langle[AGR \quad imper],\#4\rangle \\ ARG-ST & \langle\#3\rangle \end{bmatrix}$$
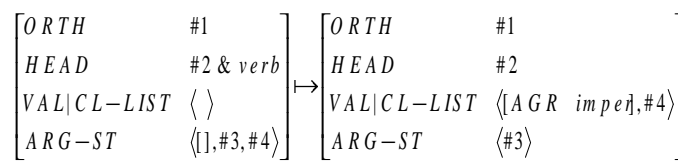
Figure 35. Lexical rule for the impersonal active with clitized object

$$\begin{bmatrix} SUBJ & \langle\,\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad (=S)$$

|

$$\begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \langle\,\rangle \end{bmatrix} \quad (=VP)$$

$$\begin{bmatrix} HEAD & clitic \\ VAL\,|\,CL-LIST & \#a \end{bmatrix} \qquad \begin{bmatrix} SUBJ & \langle\#2\rangle \\ COMPS & \langle\,\rangle \\ CL-LIST & \#a\langle[AGR \quad imper],NP-dat\rangle \end{bmatrix}$$

se les   habla
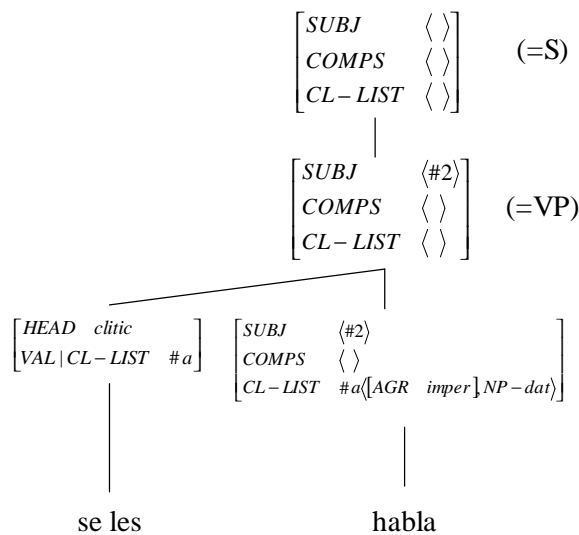
Figure 36. Analysis of the impersonal active

When the verb is intransitive or stative, the sentence requires one or two circumstantial complements or modifiers to complete its meaning, although these complements can be absent and the form becomes close to an interjection with some semantic content like, for instance, *se vive* (one lives). However, the active-impersonal lexical rule we propose here for intransitive and stative verbs does include a number of arguments to complement the meaning of the verbal phrase, as shown in Figure 37. The analysis of the sentence *se vive tranquilo en el campo* (one lives quietly in the country side) in Figure 38.

With this we conclude the analysis of the Spanish impersonal forms. As was shown, the analysis of these two forms is straightforward given the clitic system machinery. This analysis supports also the case that proclitic pronouns are really clitics and not inflections as expletives are normally considered words.
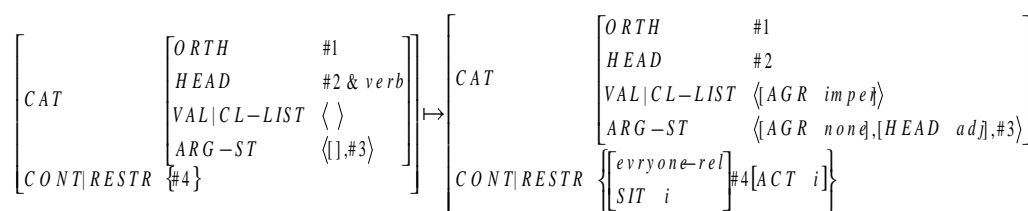
$$\begin{bmatrix} CAT & \begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \;\&\; verb \\ VAL|CL-LIST & \langle\,\rangle \\ ARG-ST & \langle[\,],\#3\rangle \end{bmatrix} \\ CONT|RESTR & \{\#4\} \end{bmatrix} \mapsto \begin{bmatrix} CAT & \begin{bmatrix} ORTH & \#1 \\ HEAD & \#2 \\ VAL|CL-LIST & \langle[AGR\;\;imper]\rangle \\ ARG-ST & \langle[AGR\;\;none],[HEAD\;\;adj],\#3\rangle \end{bmatrix} \\ CONT|RESTR & \left\{\begin{bmatrix} evryone-rel \\ SIT\;\;i \end{bmatrix}\#4[ACT\;\;i]\right\} \end{bmatrix}$$

Figure 37. Lexical rule for the impersonal active with intransitive verb
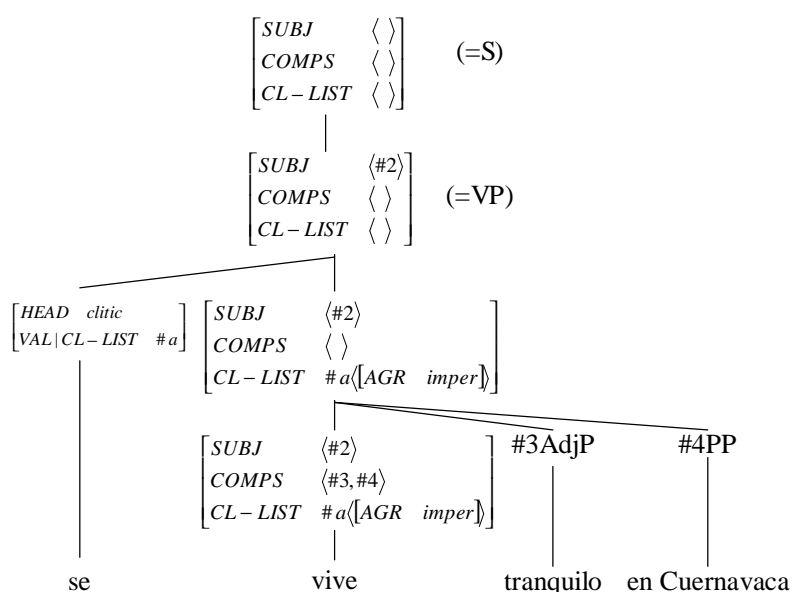
Figure 38. Analysis of the impersonal active with intransitive verb

6. Implementation

The model has been fully implemented in the context of our basic grammar for Spanish (Pineda & Meza, 2003) in the LKB grammar development environment (Copestake, 2002). As a staring point we used the grammar implemented by Quirino (2001), which covers the Spanish phenomena corresponding roughly the HPSG English textbook grammar as presented by Sag and Wasow (1999). We refined and augmented Quirino's environment according to our model for the Spanish periphrasis, and the clitic system. Our implementation follows closely the definition of the HPSG model, as presented in Section 4, and covers all phenomena and examples presented in this paper, with a few exceptions due to the limitations of LKB to handle complex morphology. In the current state our grammar has 15 syntactic combination schemes: Two versions of the Head-Specifier Rule (for subject-verb and postponed subject orders), three version of the Head-Complement Rule (for non, one and two complements), two versions of the Coordination Rule (for noun-phrases and verb-phrases), two versions of the Subject-Pro Drop Rule (one for normal omitted subjects, and one for unipersonal verbs), two versions of the

Head-Filler Rule (for one or two gaps), two versions of the Head-Modifier Rule (for pre and post modifiers), the Head-Proclitic Rule (only one version), and one additional rule for handling marking prepositions in verbal complements. The lexicon includes about 180 entries. The clitic lexicon includes 15 clitic words, that are enough to handle all the cases appearing in our corpus, and in most of the examples in this paper. *S-LC* is implemented through a single lexical rule and *C-LC* requires one rule for clitization and another for clitic insertion and cancellation. The grammar has 120 lexical rules; from these, 56 correspond to the clitic system, including reflexives with the corresponding binding relations, and the operations of clitic composition and subsumption. The clitization operation is implemented through 18 lexical rules and enclitics are handled through 29 rules. To distinguish reflexive from non reflexive instances, 9 lexical rules were implemented. The rest of the lexical rules handle other phenomena, like inflexions, the passive voice, plurals, nominalizations, etc. In particular, 5 rules were required for the passive-reflexive and the impersonal-active. The lexical rules and syntactic schemes included to handle the clitic system and related phenomena interact well with the rest of the grammar.

7. Conclusions and further work

In this paper we have presented a model for the Spanish pronominal clitic system. In Spanish, almost every predicative sentence has a number of clitic versions in which the direct or the indirect complement, or both, can be substituted by unstressed accusative and dative pronouns respectively, with the pleonastic repetition of the dative case allowed always, and the repetition of the accusative allowed often. These pronouns can appear both as enclitic attached at the end of their verbal host, or as proclitics in front of the host or periphrastic verbal phrases. The construction exhibits the phenomenon known as clitic climbing in which there is periphrastic lexical material between the explicit realization of the pronouns and the verbal host that has the pronouns as its arguments, and the form is a case of distant but bounded dependencies. The form provides an economical and direct expressive device of the language and clitic sentences are very commonly used in both spoken and written language.

We have reviewed the main forms of the phenomenon, including the basic form, the clitization of nominal phrases, and the case of simple clitic climbing, where all pronouns are arguments of a single verb with full semantic content, normally at the end of the periphrasis. We have also studied the case of complex clitic climbing where a periphrasis contains more than one content verb, and the clitic pronouns are arguments of different verbs; we have accounted for this phenomenon in terms of the clitic composition and subsumption operations; we have also studied the case of pronouns that have a wider scope over coordination, that are also accounted for in terms of clitic subsumption. We have also analyzed the different forms of the reflexive, which interacts and specializes the clitic system. Finally, we have studied the passive-reflexive and the impersonal-active, two impersonal forms commonly used in both spoken and written Spanish. These forms also use the clitic system, although the pronoun *se* becomes an expletive without referential import. We have proposed a model for all these forms in HPSG, and also showed its implementation in LKB.

Unlike lexical approaches to clitization that consider clitic pronouns, both enclitic and proclitic, as inflectional affixes, we have argued that the phenomenon presents a dual behavior, and although enclitics are also considered inflections, proclitics are treated as independent lexical units. In this respect, our model is consistent with the traditional intuitions and the spelling conventions of the language. We base our case in different sorts of evidence. First we reviewed the criteria proposed by Zwicky and Pullum (1983) to distinguish clitics from inflections, and conclude that proclitic pronouns are really clitics; in particular, clitic pronouns in Spanish can be subject of syntactic operations, as these can have a wider scope over coordination; in addition, we have noticed in our corpus a common speech repair consisting in the interruption of a proclitic sentence, and the use of an enclitic form instead, a phenomenon that could not occur if the form where realized fully in the morphology. However, the most important source of evidence is the creation of abstract clitic host, that are formed out of more than one concrete clitic host by the operations of clitic composition and clitic subsumption; although these operations are defined as lexical properties of object-control verbs, the actual creation of an abstract clitic domain is due to the syntactic scheme that combines the verb with its complements (i.e. The Head-Complement Rule in the HPSG formalization); the subsumption operation is also defined in the syntactic coordination rule. In particular, the analysis of constructions involving a clitic with a wider scope over coordination involves the creation of an abstract clitic domain by clitic subsumption. A final piece of evidence is that the reflexive system, and also the different forms of the expletive *se*, are defined upon the clitic machinery, and the combination of the reflexive pronouns and the expletive *se* with their corresponding verbal phrases requires a syntactic operation; furthermore, reflexive pronouns and expletives are normally considered words.

We also introduce novel terminology that clarifies the structure of the phenomenon. Unlike previous approaches, we distinguish between the notions of host clitization, clitic insertion and clitic cancellation explicitly. These notions highlight the distinction between clitic host, the verb with full semantic content whose direct and indirect objects are clitized, and phonological host, the word that the pronouns attach phonologically. We also introduce the notion of local clitic domain, the accessibility relation between local clitic domains to form clitic domains, and also the notion of clitic scope: the local clitic domain is simply the list of clitic arguments of a clitic host and a clitic host is within the scope of a phonological host if their local domains are accessible; alternatively, clitic and phonological host are accessible if they belong the same clitic domain. In our model there are three modes of clitic cancellation: simple lexical cancellation (*S-LC*), complex lexical cancellation (*C-LC*) and syntactic cancellation (*SC*). All of this is summarized in the principle of clitic cancellation: all Spanish clitic sentences undergo one clitization and one cancellation per independent clitic domain; for cancellation to take place, the clitic host must be accessible to the phonological host. This principle has full generality when abstract clitic hosts, produced by clitic composition and subsumption, are considered, as every clitic sentence undergoes only one clitic cancellation operation per independent clitic domain, even if the clitic word is formed by pronouns belonging to different concrete clitizations. On the basis of this, we postulate the Spanish pronominal clitic principle: Spanish clitic sentences have a single phonological host that has within its

scope a single concrete or abstract clitic host per independent clitic domain; abstract clitic hosts are produced by the operations of clitic composition and clitic subsumption.

Our approach is also consistent with the observations about the atomicity of clitic compounds that motivate the treatment of clitics as inflections. On the surface form, there is never lexical material between the pronouns and the phonological host, not even when the pronouns are realized pleonastically. In our theory, this is a consequence of both lexical and syntactic factors: in *S-LC* the clitic and phonological host is the same word, and the pronouns are attached to the host directly; in *C-LC* the clitic host and the phonological host are different words, but the pronouns attach to the phonological host in the morphology too; in *SC*, on the other hand, the inclusion of lexical material between the pronouns and the phonological host is prevented by the requirement that the complement list of the verbal head phrase must be saturated, and the pronouns attach to the phonological host directly, despite that the *Head-Proclitic Rule* is a syntactic operation. In fact, in all three forms of clitic cancellation, the complement list must be saturated, although in the HPSG formalization and the LKB implementation, this is hidden within the lexical rules that implement *S-CL* and *C-LC*. The net result is that the surface form does look like an indivisible unit, but one that is rendered by a "hidden" syntactic operation. In addition, *SC* offers a unified mechanism to deal with reflexive pronouns and expletives, which cannot be thought of as inflexions that easily. Also, our approach overcomes the shortcomings of lexical approaches that cannot deal with constructions exhibiting clitic composition and subsumption, like the ethical dative, or constructions in which the clitic has a wider scope over coordination, that involve the creation of an abstract clitic domain.

The clitic principle seems to capture a very general structural property of languages with clitic system. The phenomenon seems to follow from an economical constraint on the phonology: the fact that there is a single phonological host per independent clitic domain. The phonological host attracts all clitic pronouns as a single unit, whether these are local or distant from the original local clitic domain. Clitic words, on the other hand, seem to be common ordered sequences of pronouns, strengthen by the local linguistic use. These sequences can have one or two elements, and the set of possible combinations, taken into account case, gender and number, is rather small. In Mexican Spanish we have identified no more than 60 words, including the clitic words used in the impersonal constructions, where *se* is an expletive. The clitic words included in the clitic lexicon impose a number of lexical constraints on the clitic sentences: these are the only sequences that can appear both as enclitics or proclitics. This explains the idiosyncratic order of the pronouns. Also, the absence of certain clitic combinations in the clitic lexicon (i.e. the absence of a clitic paradigm) can be explained in terms of phonological factors (e.g. the evolution of the personal *se* as a duplication of *le*), and also in terms of the possible referential ambiguities that the presence of some clitic sequence would cause (e.g. "*la la*"), or the interaction of these two factors (e.g. "*se lo*" evolved because referential constraints prevented the formation of sequences of two *l'* pronouns, like "*le lo*"). Although both the inflectional paradigm and the clitic system may exhibit idiosyncratic exceptions, the holes in the "paradigms" may have very different causes. Also, the only constraint

blocking clitic composition and subsumption depends on the clitic lexicon, as the abstract clitic domain formed by these operations must be cancelled with a clitic word.

More generally, clitic climbing seems to be a very unconstrained phenomenon that can be prevented by lexical constraints only: participles and tensed forms cannot have enclitics. In this respect, it has been argued that modal verbs prevent clitic climbing in some contexts; according to Green (1988), for instance, the enclitic forms *creo haberlo dicho* (I think I said it) and *siento haberlo hecho* (I'm sorry I did it) are grammatical, but their proclitic counterparts *lo creo haber dicho* and *lo siento haber hecho* are not. However, at least in Mexican Spanish, the first of these proclitic forms is acceptable; the second sounds less natural, but this can be due to the fact that the expression *lo siento* (I'm sorry) is a ready made idiom, so often used that the full form *lo siento haber hecho* is never used. Another exception pointed out by Green is that modal verbs can also subcategorize for a prepositional phrase where the preposition's object is a nominalized verb, that can also be clitized: *insisto en hacerlo* (I insist on doing it) and *sueño con verlo* (I dream of seeing it). Here, the preposition do blocks clitic climbing and the corresponding proclitic sentences *lo insisto en hacer* and *lo sueño con ver* are ungrammatical, and the clitic must stay within its local clitic domain. However, in these examples there is not really a periphrasis involved as the whole of the prepositional phrase is a simple verbal modifier. Besides these constraints, the periphrastic verbs are transparent to clitization, as no structural constraint seems to block the accessibility relation between the phonological and clitic host. Every independent clitization, in turn, depends on the presence of a phonological host. The clitic principle applies to all Spanish phenomena that we have studied; our preliminary observations show that it also applies to clitizations of or within relative clauses, although we leave the detail study of these forms for further research.

The clitic principle can help to explain why the clitic system is a focus of dialectal and diachronic variation. The fact that a phonological host has within its scope the clitic host, and attracts all clitic pronouns as a single unit remain constant; however, lexical constraints due to local use, or in a given language state, may render some sentences ungrammatical. The different uses of *leísmo* and *laismo* in Mexico, Spain, and the rest of Hispano America, can be explained in terms of the particular set of entries (e.g. *le* accusative or *lo* and *la* dative) in the clitic lexicon of the dialect, and also in the strength of use of those entries; also, the uncommon proclitic order of the Caribbean Spanish can be explained by the inclusion of words like "*me se*" and "*te se*" in the clitic lexicon. However, we leave the detailed study of how this model can be applied to other Spanish dialects for further research.

The study of Spanish pronominal clitics presented in this paper is mainly restricted to morphological and syntactic considerations; however, in a more comprehensive theory, the prosodic structure of clitic sentences must be addressed explicitly, and we leave for further research the question of whether the clitic principle, as stated in this paper, has a role to play at the prosodic level. We also leave for further research whether the clitic principle applies to other languages with pronominal clitic systems. This would be the case if there are other languages that favor the presence of one phonological host per

independent clitic domain, and the constraints on clitic climbing are identified as lexical in nature.

Acknowledgements

References

[Abeillé et. al. 1998] Abeillé, A, Godard, D., Miller, P. And Sag, I. (1998). French Bounded Dependencies. In Luca Dini and Sergio Balari, (eds.), Romance in HPSG, Standford: CSLI Publications.

[Copestake 2002] A. Copestake. (2002) The LKB System, Stanford University, http://www-csli.stanford.edu/\symbol/~aac/lkb.html

[Gili Gaya 1991] Gili Gaya, S. (1991). Curso Superior de Sintaxis Española, Biblograf, S. A., 103-153.

[Green 1988] Green, N. J. (1988). Spanish, in The Romance Languages, Chapter 3., Martin Harris and Nigel Vincent (eds.), Oxford University Press, New York.

[Meza & Pineda 2002]   Meza, I. & Pineda, L. (2002). The Spanish auxiliary verb system in HPSG. In proceedings of CICling-2002 Alexander Gelbuck (ed.), Lecture Notes in Computer Science 2276, Springer-Verlag, pp. 200-209.

[Miller & Sag 1995] Miller P. H. and Sag, I. A. (1995). French Clitic Movement Without Clitics or Movement. Natural Language and Linguistic Theory, 15: 573—639.

[Monachesi 1993] Monachesi, P, (1993). Object clitics and clitic climbing in Italian HPSG grammar, Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics. ACL Conferences, Utrecht, ACL.

[Monachesi 1999] Monachesi, P, (1999). A Lexical Approach to Italian Clitization. Lecture Notes series No. 84, CSLI, Stanford, Cambridge University Press.

[Navis et. al. 1994] Nevis, J. A., Joseph, B. D., Wanner, D. and Zwicky, A. M. (1994). Clitics, A Comprehensive Bibliography 1892-1991. Library and Information Sources in Linguistics, 22. John Benjamins Pub. Co., Amsterdam/Philadelphia.

[Pineda et. al. 2002] Pineda, L. A., Massé, A., Meza, I., Salas, M., Schwarz, E., Uraga, E and Villaseñor, L. (2002). The Dime project. In proceedings of MICAI2002, Lectures Notes in Artificial Intelligence 2313, pp.166-175.

[Pineda & Meza, 2003]   Pineda, L. & Meza, I. (2003). Una gramática básica del español en HPSG. Reporte interno, Depto. De Ciencias de la Computación, IIMAS, UNAM, Mex.

[Pollard & Sag, 1994] Pollard, C. & Sag, I. (1994). Head-Driven Phrase Structure Grammar, CSLI, Stanford, The University of Chicago Press, Chicago & London.

[Quirino 2001] Quirino, A. P. (2001), *Spanish clitics: A computational model*, Universität Bielefeld, Master's thesis, "http://coral.lili.uni-bielefeld.de/~aquirino/thesis.html

[Rivero-Castillo 1997]. Rivero-Castillo, Y. (1997). Proclitic groups and word order in Caribbean Spanish. In Clitics, Pronouns and Movement, Current issues in Linguistic Theory, 140. James R. Black and Virginia Motapanyana (eds.). John Benjamins Pub. Co. Amsterdam/Philadelphica. pp. 237-272.

[Sag & Wasow 1999] Sag I. & Wasow T. (1999). Syntactic Theory: A Formal Introduction, CSLI Publications, Stanford.

[Villaseñor 2001]Villaseñor, L., Massé, A. & Pineda, L. A. (2001). The DIME Corpus, Memorias 3º. Encuentro Internacional de Ciencias de la Computación ENC01, Tomo II, C. Zozaya, M. Mejía, P. Noriega y A. Sánchez (eds.), SMCC, Aguascalientes, Ags. México, Septiembre, 2001.

[Zwicky & Pullum 1983] Zwicky, A, & Pullum, G. (1983). Clitization vs. Inflection: English *N'T*", *Language* 59, 502-513.