

# The Spanish pronominal clitic system

**Luis Pineda**

Instituto de Investigaciones en Matemáticas  
Aplicadas y en Sistemas (IIMAS)  
Universidad Nacional Autónoma de México  
(UNAM)  
luis@leibniz.iimas.unam.mx

**Ivan Meza**

Division of Informatics  
University of Edinburgh  
I.V.Meza-Ruiz@sms.ed.ac.uk

**Resumen:** En este artículo se presenta un modelo del sistema de pronombres clíticos del español. Se presenta una revisión detallada del fenómeno incluyendo la llamada “subida de clíticos” tanto simple como compleja, así como el fenómeno relacionado de los reflexivos, y también las formas impersonales de la pasiva-refleja y la impersonal-activa. Se presenta también un análisis del estatus representacional de los pronombres clíticos y se postula un modelo dual en el que mientras los enclíticos son inflexiones, los proclíticos se realizan como unidades léxicas independientes y por lo mismo son clíticos propiamente. Se presenta también la formalización del modelo en Head-driven Phrase Structure Grammar (HPSG); para esto la maquinaria estándar de HPSG se extiende con un esquema de combinación sintáctica, la regla Head-Proclitic Rule y el principio de clíticos. Este principio establece que en las oraciones bien formadas que incluyen pronombres clíticos todo dominio clítico se encuentra bajo el alcance de un clítico fonológico. Estas nociones se introducen también en el presente artículo. En particular, los dominios clíticos se forma sobre las operaciones de cliticización, composición y subsumisión de clíticos, que también se presentan en este artículo. La teoría ha sido validada mediante la programación de un sistema prototipo en el ambiente Linguistics Knowledge Building (LKB), el cual también se describe en el presente artículo.

**Palabras clave:** Pronombres clíticos, subida de clíticos, composición y subsumisión de clíticos, perífrasis del español, pasiva refleja, impersonal activa. El clítico *se*.

**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. We also 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 the formalization of the model in Head-driven Phrase Structure Grammar is presented; for this the standard machinery of HPSG is extended with one combination scheme, the Head-Proclitic Rule, and one principle, the clitic principle. This principle states that in well-formed clitic sentences every clitic domain is within the scope of a single phonological host. The notions of clitic domain and clitic scope are also introduced in this paper, in addition to the operations of cliticization, composition and subsumption through which clitic domains are produced. An implementation of the theory in Linguistics Knowledge Building is also presented.

**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 inflections 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 known 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 the structure of the lexicon may influence the forms permitted in one or another dialect of a language.

### 1.1 The structure of Spanish clitic sentences

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) always require proclitics; participles cannot have enclitics also and when they are cliticized, 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 cliticized, and possibly with some or all these realized redundantly too. Next, we review the main properties of clitic sentences and related phenomena<sup>1</sup>; the basic form is illustrated in (1). For clarity, complements are enclosed in brackets when required, and have the same index in their corresponding cliticized versions; also, complements that are realized twice are shown co-indexed<sup>2</sup>.

<sup>1</sup> There is a vast literature on clitics in general (Nevis, 1994), and also about Spanish pronominal clitics and related phenomena, and the present section has the purpose to motivate the computational model presented below in this paper, and also summarize the main facts for unfamiliar readers. With the exception of the notions of clitic domain and scope, and the phenomena of clitic composition and subsumption, which are introduced in this paper, the rest of the material can be found in standard Spanish grammar texts and monographs.

<sup>2</sup> The main examples in this paper are based on a clitic sentence that has been extracted from the Internet, which we consider as our corpus for this purpose. The rest of the sentences in the

- (1) a. *El muchacho da [la copa]<sub>i</sub> [a la Novia]<sub>j</sub>*<sup>3</sup>  
*The groom gives the glass<sub>i</sub> to the bride<sub>j</sub>*  
 b. *dala<sub>i</sub> [a la novia]<sub>j</sub>*  
 c. *dale<sub>j</sub> [la copa]<sub>i</sub>*  
 d. *dase<sub>j</sub>la<sub>i</sub>*  
 e. *dase<sub>j</sub>la<sub>i</sub> [a la novia]<sub>j</sub>*  
 f. *la<sub>i</sub> da [a la novia]<sub>j</sub>*  
 g. *le<sub>j</sub> da [la copa]<sub>i</sub> [a la novia]<sub>j</sub>*  
 h. *se<sub>j</sub> la<sub>i</sub> da*  
 i. *se<sub>j</sub> la<sub>i</sub> da [a la novia]<sub>j</sub>*

In (1) *la* is a third person feminine accusative pronoun substituting the direct object *la copa*, and *le* and *se* are third person dative pronouns substituting *a la novia*. 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 cliticized with *la*, 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 la*” sequence is not allowed. In general, the personal pronoun *se* is a duplication of the personal *le*, as both come from the Latin personal *illi*, and when *le* should appear in front of *la* (or *lo*, its masculine form), both in proclitics and enclitics, *se* must be used instead<sup>4</sup>. Sentence (1.e) has, in addition, the indirect object realized twice with *se* and *la novia*. Sentence (1.f) has the direct object realized as the proclitic *la*, (1.g) the indirect object realized as *le* which also appears duplicated, and in (1.h) both of the complements are realized as proclitics. Finally,

---

corresponding sequence are variations from the reference one which are acceptable and meaningful for native speakers of Spanish, unless otherwise stated.

<sup>3</sup> *El padrino le sirve una copa al muchacho, y este se la da a la novia*  
[http://omega.ilce.edu.mx:3000/sites/litinf/huasteca/html/sec\\_45.htm](http://omega.ilce.edu.mx:3000/sites/litinf/huasteca/html/sec_45.htm)

<sup>4</sup> 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<sub>j</sub>lo<sub>j</sub>* (unvoiced alveolar fricative); 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 to him<sub>i</sub>).

(1.i) is like (1.h) but with the indirect complement realized twice<sup>5</sup>.

Unlike Spanish, duplicated 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 often duplicated too, although there are some restrictions on non-animate direct objects: the sentence *lo<sub>i</sub> entrega a Juan<sub>i</sub> a la policía* (someone turns Juan in to the police) is allowed, but *lo<sub>j</sub> da el coche<sub>j</sub> 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 third complement too (i.e. a prepositional modifier), and in this case it is very far from the normal dative use; for instance, *le compró el libro a María* is ambiguous between “he/she bought the book for Maria” or “he/she bought the book from Maria”; while *le* duplicates *a María* in both interpretations, in the former one *a María* is the beneficiary of the buying action and *le* has the normal dative reading, but in the second, the preposition *a* marks that Maria is the one who sells the book and *le* is closer to an ablative case.

Clitic pronouns are all enclitic or all proclitic and the same order is preserved in both forms; in (1), for instance, *se* always precedes *la*. The general rule is that *se* must be in front always, second persons are in front of first ones and third persons (except *se*) are always last. In addition, in the basic form of the phenomenon (i.e. in non-periphrastic constructions), there is no lexical material between the clitic pronoun sequence and the cliticized verb. For this reason, despite the fact that in Spanish the complements can appear in front of the verb, and although alternations of (1.g) in (2.a) and (2.b) are grammatical (2.c) and (2.d) are not:

- (2) a. [*la copa*]<sub>i</sub> *le<sub>j</sub> da [a la novia]<sub>j</sub>*  
 b. [*a la novia*]<sub>j</sub> *le<sub>j</sub> da [la copa]<sub>i</sub>*

<sup>5</sup> 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 redundant construction in relation to the person and number marked in the verbal inflection.

- c. *\*le<sub>j</sub> [la copa]<sub>i</sub> da [a la novia]<sub>j</sub>*  
 d. *\*le<sub>j</sub> [a la novia]<sub>j</sub> da [la copa]<sub>i</sub>*

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

- (3) a. *El manejar [el carro]<sub>i</sub> es una experiencia única*<sup>6</sup>  
 The to-drive the car is an experience unique  
 To drive the car is an unique experience  
 b. *El manejarlo<sub>i</sub> es una experiencia única*  
 To drive it<sub>i</sub> is an unique experience

- (4) a. *Velazquez pintando [el cuadro]<sub>i</sub>*  
 Velazquez painting the picture;  
 b. *Velazquez pintándolo<sub>i</sub>*  
 Velazquez showing it<sub>i</sub>

In addition to the basic forms, clitics can occur in periphrasis where the pronouns are distant to the verb that has them as its arguments; in this situation it is said that the pronouns *climb* over the intermediate lexical material (e.g. auxiliary or modal verbs<sup>7</sup>). In clitic climbing we need to distinguish the *clitic host* from the *phonological host*; the former is verb whose complements are cliticized 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):

<sup>6</sup> El manejarlo es una experiencia única  
<http://www.solanacadillac.com/nuevo9-3/nuevo9-3.htm>

<sup>7</sup> We adopt Gili Gaya's terminology and call modal verbs to agentive (intentional) verbs appearing in periphrasis: *querer*, *pensar*, etc. This is a conventional choice a no claim about the syntactic status or semantic import of these verbs is intended to be made with this notation.

- (5) a. *No he podido escribir [el post]<sub>i</sub>*<sup>8</sup>  
*Not I-have been-able to-write the post<sub>i</sub>*  
*I have not been able to write the post<sub>i</sub>*  
 b. *No he podido escribirlo<sub>i</sub>*  
 c. *No lo<sub>i</sub> he podido escribir*
- (6) a. *Pude haber querido escribir [el post]<sub>i</sub>*  
*Not I-could to-have wanted to-write the post<sub>i</sub>*  
*I could have wanted to write the post<sub>i</sub>*  
 b. *Pude haber querido escribirlo<sub>i</sub>*  
 c. *Pude haberlo<sub>i</sub> querido escribir*  
 d. *Lo<sub>i</sub> pude haber querido escribir*

In (5) and (6) the clitic host is the verb *escribir* but the phonological host varies: in (5.b) and (6.b) the phonological host is still *escribir*, but in (5.c) and (6.c) the phonological host is the auxiliary *haber* (i.e. *he* and *haber*); in (6.d) the phonological host is the auxiliary *pude*.

Clitic pronouns can also climb over gerunds in periphrasis, as shown in (7), and over the participle in the complex conjugation as in (8)<sup>9</sup>:

- (7) a. *El virus Klez.H continúa liderando [el ranking de este mes]<sub>i</sub>*<sup>10</sup>  
*The virus Klez.H continues leading the ranking of this month*  
*The Klez.H virus continues leading this month's ranking*  
 b. *el virus Klez.H continúa liderándolo*  
 c. *el virus Klez.H lo<sub>i</sub> continúa liderando*
- (8) a. *El virus Klez.H ha continuado liderando [el ranking de este mes]<sub>i</sub>*  
*The Klez.H virus has continued leading this month ranking*  
 b. *The Klez.H virus ha continuado liderándolo<sub>i</sub>*  
 c. *The Klez.H virus lo<sub>i</sub> ha continuado liderando*

<sup>8</sup> el post no lo he podido escribir por la mañana  
<http://blogs.ya.com/vivirsintabaco/>

<sup>9</sup> In Spanish the verb *haber* is used in all finite perfect forms and it is often considered an inflexion, and the composite form *haber* (tensed) + *participle* is known as the complex conjugation (*conjugación compleja*).

<sup>10</sup> *El ranking de este mes lo continúa liderando el Klez.H*

<http://www.vsantivirus.com/virus-report-jun-02.htm>

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.

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. *escribir* (to write), *liderar* (to lead)). However, there are also complex periphrases in which the cliticized 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 (e.g. sense verbs like *oir* (to hear)), whose direct object is the subject of the next verb in the periphrasis, as illustrated in (9):

- (9) a. *Oí [a el]<sub>i</sub> decir [el comentario]<sub>j</sub>*<sup>11</sup>  
*I-hear to him<sub>i</sub> to-say the comment<sub>j</sub>*  
*I hear him to say the comment*  
 b. *\*Oílo<sub>i</sub> decirlo<sub>j</sub>*  
 c. *Oyélo<sub>i</sub> decirlo<sub>j</sub>*  
 d. *\*Oyélo<sub>i</sub>lo<sub>j</sub> decir*  
 e. *Oyése<sub>i</sub>lo<sub>j</sub> decir*  
 f. *Le<sub>i</sub> oí decirlo<sub>j</sub>*  
 g. *Se<sub>i</sub> lo<sub>j</sub> oí decir*

In (9.a) there are two verbs with full semantic content: *oír*, the simple past of *oir* (to hear), and *decir* (to tell), with their corresponding direct objects; in (9.b) and (9.c) both of their direct objects are cliticized with the accusative *lo*; in (9.d) the object of *decir* climbs up and appears as enclitic to *oyelo*, which is a cliticized form already. In (9.e) the object of *oir* is realized as *se* instead of *lo*, also in accusative. In (9.f) the accusative object of *oir* is realized as the proclitic *le* and in (9.g) both pronouns are realized as proclitics. The imperative (9.c) illustrates the case in which the direct objects of both of the verbs are cliticized as the accusative *lo* in their

<sup>11</sup> *Se lo oí decir en varios reportajes*  
<http://www.carp.org.ar/eng/idolos.php3>

corresponding base positions; the similar form (9.b) is not grammatical due to the idiosyncratic lexical restriction on phonological hosts, as these cannot have enclitics (although its ungrammaticality is marginal as this form is meaningful and can be heard in colloquial spoken language). In (9.e) the object of *oir* is realized as the accusative *se* and the object of *decir* climbs up and attaches to the clitic form *oyese* also as enclitic, rendering the form *oyéselo*. This form differs from (9.d) in that the accusative complement of *oir* is realized as *se* instead of *lo* and the ungrammaticality of (9.d) is due to the “*lo lo*” sequence. In (9.f) the direct object of *oir* is realized as the proclitic *le* and the complement of *decir* is realized as the accusative *lo* in its base position. The use of *le* as accusative, normally a dative pronoun, illustrates the phenomenon known as *leísmo*, which is widely extended in different dialects of Spanish. Finally, in (9.g) the object of *decir* climbs up and is realized as the accusative *lo* next to the object of *oir* which is realized as the accusative *se*. In this latter case, the use of *se* is obligatory as neither “*le lo*” or “*lo lo*” are allowed. More generally, no in Spanish no sequence of two *l'* pronouns is allowed. Also (9.e) and (9.g) illustrate a case of *leísmo* in which *se*, normally a dative pronoun, is realized as accusative.

The sequence in (9) illustrates that the same pronoun can appear twice (e.g. 9.c) violating the standard restriction of clitic order; a further illustration of order violation would be *Lo<sub>i</sub> oi decirle<sub>j</sub>* (I hear him to tell him/her) in which *lo* appears before *le*. The sequence also shows that there can be lexical material between the pronouns despite that there are no auxiliary or modals in the construction (e.g. 9.c and 9.f); however, these deviations are only apparent because the pronouns violating the standard constraints are arguments of different verbs; to handle this phenomena we say that a cliticization forms a local clitic domain consisting of the arguments of the cliticized verb, and the sentence is grammatical if every local domain is within the scope of its corresponding phonological host, permitting to establish the binding relations between the pronouns and their argument positions, as will show below in the formalization of the model.

The sequence in (9) also illustrates that two pronouns that are arguments of different verbs can be realized next to each other either as enclitics (e.g. 9.e) or proclitics (e.g. 9.g), and

nevertheless sentences including this construction are not ambiguous and the bindings between the pronouns and their argument positions can be established very easily by native speakers of the language; here, we refer to a sequence of two clitic pronouns that are arguments of different verbs as clitic composition. Examples in (10) illustrate similar relations, but with a ditransitive verb, *poner*, instead of the transitive *decir*:

- (10) a. *Oí [a él]<sub>i</sub> poner [las llaves]<sub>j</sub> en la mesa*  
 I heard him<sub>i</sub> to put the keys<sub>j</sub> on the table
- b. *Oí [a él]<sub>i</sub> ponerlas<sub>j</sub> en la mesa*
- c. *Le<sub>i</sub> oí ponerlas<sub>j</sub> en la mesa*
- d. *Se<sub>i</sub> las<sub>j</sub> oí poner en la mesa*

## 1.2 Clitics and reflexives

The Spanish reflexive is a rich phenomena with a wide range of behavior that has complex interactions 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. 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* (1<sup>st</sup>-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 or recipient of the action.

- (11) a. *Yo<sub>i</sub> me<sub>i</sub> lavo*  
 I<sub>i</sub> myself<sub>i</sub> wash  
 I wash myself
- b. *Me<sub>i</sub> lavo<sub>i</sub>*  
 myself<sub>i</sub> wash  
 I wash myself
- c. *Me<sub>i</sub> lavo<sub>i</sub> las manos*  
 Myself<sub>i</sub> 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 agent is not the recipient either directly or indirectly of the action but he or she only has a participation, interest or influence in an action that is performed by another agent. There is an ample spectrum for this phenomenon that ranges from

the expression of explicit participation to the expressions of a passive interest with intransitive verbs (i.e. lacking a recipient of the action, process or state named by the verb), often called pseudo-reflexive (*pseudo reflejas*), 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* (2<sup>nd</sup>-singular), who is also the subject and the beneficiary of the house that is built, but is neither the agent or recipient of the building action; 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. 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 (1<sup>st</sup>-singular) co-indexed with the implicit subject; also *te me voy* (I go for my and your own sake) with two dative clitics<sup>12</sup>. 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. Verbs of this latter kind, 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<sub>i</sub> construiste<sub>i</sub> una casa*  
 you built-for-yourself a house  
 you built a house for yourself
- b. *Ella<sub>i</sub> se<sub>i</sub> tomó<sub>i</sub> el café*  
 She her had the coffee  
 She had a coffee
- c. *Me<sub>i</sub> voy<sub>i</sub>*  
 Me leave  
 I leave (for my own sake)
- d. *Se<sub>i</sub> arrepiente<sub>i</sub>*  
 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

<sup>12</sup> For instance, in a current TV add she says to him “*cuídame porque te me voy*” (look after me otherwise I go for my own for/despite your sake).

nominative, and the direct and indirect objects in accusative or dative respectively, or with other complements with an ablative character that do not belong to the basic argument structure of the verb and are not realized explicitly. Next we consider a further consequence of the interaction between clitics and a beneficiary complement in (13):

- (13) a. *Vi [a usted]<sub>i</sub> comiendo [la cena]<sub>j</sub>*  
*[por/para usted]<sub>i</sub>*<sup>13</sup>  
 see to you<sub>i</sub> eating the dinner<sub>j</sub> for  
 you<sub>i</sub>  
 I see you eating the dinner for you  
 own sake
- b. *Vi [a usted]<sub>i</sub> comiendose<sub>i</sub> [la cena]<sub>j</sub>*
- c. *Vi [a usted]<sub>i</sub> comiendose<sub>i</sub>la<sub>j</sub>*
- d. *\*Vila<sub>i</sub> comiendose<sub>i</sub>la<sub>j</sub>*
- e. *Vela<sub>i</sub> comiendose<sub>i</sub>la<sub>j</sub>*
- f. *La<sub>i</sub> vi comiendose<sub>i</sub>la<sub>j</sub>*
- g. *\*Vela<sub>i</sub>+se<sub>i</sub>la<sub>j</sub>comiendo*
- h. *Vese<sub>i</sub>la<sub>j</sub>comiendo*  
 (i.e. *se<sub>i</sub> = la<sub>i</sub>+se<sub>i</sub>*)
- i. *Se<sub>i</sub> la<sub>j</sub> vi comiendo*

The basic sentences for this sequence are (13.b) and (13.c), and (13.a) does not really occur in the language and it is only presented to illustrate the explicit realization of the ablative pronoun in case it were present. The ethical dative appears as the enclitic *se* in (13.b), and a reflexive relation is established with the direct object of *vi* (the simple past of *ver* (to see)). In (13.c) the direct object of *comiendo* (i.e. *la cena*) is also cliticized as the accusative *la*. The imperative (13.e) illustrates the cliticization of the direct object of *vi* in its base position, and the ungrammaticality of the analogous form (13.d) is due to the restriction on enclitics for tensed phonological hosts. In (13.f) the object of *vi* is realized as a proclitic but the *se la* sequence appears as enclitic to its basic position. In this construction the same pronoun appears twice, but each one occurs in an independent local clitic domain and has a different phonological host, and the construction is grammatical; this form also highlights the reflexive relation between the accusative object of *vi* and the dative of *comer*, the beneficiary of the eating action, and these two pronouns are co-indexed. In (13.g) the enclitics of *comiendo* climb up as enclitics to the already cliticized form *vela* but this sentence is ungrammatical because the

<sup>13</sup> *La vi comiéndose la mesa fría con los ojos*  
<http://www.mundomatero.com/chistes/junio2000.html>

sequence  $la_i+se_jla_j$  is not allowed; however, the accusative pronoun  $la_i$  and the dative  $se_j$  are co-indexed, and only the dative form needs to be realized, as shown in (13.h) which is grammatical. In (13.i) the pronoun sequence climbs up and appears as proclitic to the whole constructions. Here, we refer to the argument reduction in (13.h) and (13.i) as clitic subsumption.

An additional observation is that the (13.i) is ambiguous depending of whether  $se$  has an accusative or a dative reading; this in turns depends on whether the form is the result of clitic composition (non-reflexive) or clitic subsumption (reflexive). In the accusative reading  $se_i la_j vi comiendo$  means “I saw her eating it (e.g. dinner)” but in the dative one, the sentence means “I saw her eating dinner for her own sake”, as the sequence in (13) illustrates. 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.

### 1.3 Clitics and the impersonal forms

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  
f. *Se abre*  
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 three 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*<sup>14</sup>. 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 and the object in the active form becomes the syntactic subject (e.g. *la puerta* in (14.d) and (14.e)), and needs to agree with the verb. Also, as only verbal complements can be cliticized, the subject of the passive reflexive (in nominative) cannot be realized as a clitic pronoun. The form is very flexible and the ellipsis in (14.f) is also allowed.

The active form of transitive verbs 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

Unlike the passive reflexive in which the overt argument (e.g. *la puerta* in 14.e) is the subject, in the impersonal active this argument is a complement; also, the active impersonal has an obligatory omitted subject, similar to impersonal verbs like *llover* (to rain)<sup>15</sup>; however, the  $se$  pronoun has to agree with the verb (i.e. 3<sup>rd</sup>- singular). In (15), for instance, *a los niños* is a complement, and subsequently, needs not to be in agreement with the verb (*se habla al niño* (the child)/*se habla a los niños*

<sup>14</sup> If  $se$  were referential the sentence would have a direct reflexive interpretation where *la puerta* (the door), an inanimate object, would be the agent and the patient of the open action. However, the reflexive import of the pronoun is often present and inanimate objects acquire “agency” given rise to a peculiar “animism” that is very characteristic of the language: *la puerta se abrió solita* (the door was opened all by itself), *la botella se cayó* (the bottle fell down by itself).

<sup>15</sup> The pronouns  $se$  cannot be the subject as subjects never undergo cliticization.

(the children)). Furthermore, it can be cliticized as in (16)<sup>16</sup>:

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

The analogy can be further developed and forms with an inanimate subject like (14.d) can be realized as *se le abre* (with compulsory *leísmo*) but this latter form becomes an impersonal active (no longer a passive-reflexive) as subjects cannot be cliticized and *le* substitutes *la puerta* (the door), which is now an accusative object. The active impersonal can also be used with intransitive, and even with stative verbs, as in (17):

- (17) a. *Se corre rápido*  
one runs fast  
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 the action has no recipient 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* pronoun 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 a passive one, 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

---

<sup>16</sup> Although the sequence *se le* is similar on the surface to the forms produced through composition and subsumption, there can be no confusion between these forms since the referential *se* cannot be followed by the dative *le* because these two are alternative realizations of the same pronoun, as was mentioned, and have the same position in the clitic order (e.g. *se lo oi decir*/\**se le oi decir* and *se la vi comiendo*/\**se le vi comiendo*).

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 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* was necessarily the subject and the sentence was passive:

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

From this latter form, and fixing the verb in singular, evolved the corresponding 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 animate but with all kind of objects; however, if the direct object is not animate the constructions does not have the *a* preposition 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

The introduction of the *a* preposition for animate direct objects interacts also with the phenomenon of *leísmo*; in old Spanish direct objects were always substituted by the accusative *lo* or *la* and indirect objects, always with the *a* (to) or *para* (for) prepositions, were substituted by the datives *le* or its equivalent realization *se*, and these substitutions were independent phenomena; however, the new *a* preposition in animate direct objects as in *Juan ayuda a Pedro* (Juan helps Pedro) made these direct objects analogous in form to indirect ones, like *Juan habla al niño* (Juan talks to the child), and direct objects could then be substituted by *le* or *se* (e.g. *Juan le ayuda* and *Juan le habla*). This analogy was further developed and inanimate direct objects could also be replaced by dative pronouns, contributing to the animism of inanimate objects: *se abre la puerta*, *se le abre* but not \**se lo abre*, as in this last form the reflexive connotation of *se* conflicts with the inanimate import of *lo*.

With this we conclude the description of the Spanish clitic system and its interactions with reflexives and the impersonal forms. From the syntactic point of view it is a system of distant but bounded dependencies subject to a number of constraints that on the surface seem to be quite idiosyncratic; in addition, it interacts with other syntactic systems, like the reflexives, both direct, indirect, reciprocals and ethical dative, and also with the pseudo-reflexives. Clitic sentences are also related, by analogy through the reflexive *se*, to the passive reflexive and the impersonal active. For all this, a model of clitics poses 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 lo oi decir*, where *se* is a referential or a reflexive 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 (she is showing the catalogue to herself), impersonal-active (someone shows the catalog) and even passive-reflexive (that catalogue is being shown).

## 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 combinability 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).

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 studies 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 *NPs* 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 inflexion. 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 cliticized word like *I've* as a unit; however, inflected words are regularly treated by syntactic operations.
- (F) Clitics can attach to words already cliticized, 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, it is not clear whether (A) to (F) can be used to classify all Spanish clitic pronouns as affixes<sup>17</sup>. In relation to (A), although Spanish pronouns select verbs as their host most of the time, they also attach to infinitives, gerunds and participles, that are normally considered verbal nouns, adverbs and adjectives respectively (Gili Gaya, 1991), and clitic pronouns can occur in

<sup>17</sup> For a reply to Zwicky and Pullum, see Klavans (1985).

constructions without a finite or personal verb, as in (3) and (4). In relation to (B), cliticization is a very systematic behavior, and all sentences can cliticize 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; consequently, according to (C), the pronouns do behave more like inflexions. On the basis of (D), most Spanish clitic pronouns do count as clitics, as the meaning of a full sentence and all its cliticized versions is the same. However, there are situations in which clitics constructions do present idiosyncrasies, as when the pronouns stand for verbal arguments, that extend the basic argument structure of the verb (e.g. forms with the ethical dative (e.g. *comerse*), reflexives attached to intransitive verbs (e.g. *me voy*) or even with two complements (e.g. *te me voy*) and also reflexive verbs (e.g. *arrepentirse* (to be sorry)) and, in this regard, they do behave like inflexions. Another idiosyncrasy is exhibited in attributives, were the phonological and clitic host is the copulative verb, and the clitic is attached to it as an inflexion, as shown in (22).

- (22) a. *Todo es posible [al que cree]<sub>i</sub>*<sup>18</sup>  
 Everything is possible to the believer  
 b. *Todo le<sub>i</sub> es posible*  
 Everything is possible to him

Criterion (E) and (F) follow directly from the explicit assumption that inflexions are combined with their stems at the morphological and lexical levels, and these atomic units cannot be altered by syntactic operations after lexical insertion. For this reason if clitics are inflexions, the pronouns cannot have a wider scope over coordination, as it is the case for French and Italian. However, in Spanish the pronouns can appear either attached to their corresponding phonological and clitic hosts as in (23.a), or factored out as a single lexical unit with a wider scope over the coordination as in (23.b):

- (23) a. *como lo llevaron y lo pusieron en alto*<sup>19</sup>

<sup>18</sup> *Al que Cree, Todo le es Posible*  
<http://www.maran-ata.net/Temas/cree.htm>

<sup>19</sup> *como lo llevaron y lo pusieron en alto*

as it they-took and it they-put in high  
As they took it and raised it

- b. *lo lleva y pone en práctica*<sup>20</sup>  
it he-takes and he-puts in practice  
he takes it and puts it into practice

On the other hand, there are contexts where the factorization of the proclitic is clearly ungrammatical, as shown in (24):

- (24) a. *Te vas o te quedas*  
you leave or you stay  
b. \**Te vas o quedas*

Examples in (23) and (24) suggest that clitic pronouns exhibit a dual behaviour; if the pronouns substitute normal accusative or dative complements (e.g. in the basic argument structure of the verb) there is some freedom in the construction and they can appear as next to their corresponding phonological hosts, or alternatively the pronoun can be factored out from the coordination as a single realization. On the other hand, if the basic argument structure of the verb is extended with one or two arguments (i.e. there is an lexical idiosyncrasy) it seems to follow the constraints on inflexions and the clitic cannot be factored out from the coordination.

Finally, on the basis of (F) accusative and dative pronouns do count as clitics, as these attach to verbs already cliticized (e.g. *óye-se-lo* (hear him/her say it), as was shown in sequences (9) and (13).

From this discussion we adopt the position that clitics are a dual phenomena and they behave both as clitics and also as inflexions; the rule seems to be as follows: in the basic form of the phenomenon, when the cliticized pronouns represent complements required by the basic argument structure of the verb, enclitics are inflexions but proclitics are independent lexical units and hence real clitics. In this case the pronouns reproduce full syntactic constituents and replace them in the argument structure of the verb directly. On the other hand, if the argument structure of the verb is extended with idiosyncratic arguments at the

lexical level of representation, clitics substitute one or even two additional complements with a dative character (sometimes even closer to an ablative one) which is/are not realized explicitly, and the clitic-host combination is a lexical or a morphological operation: pseudo-reflexives (e.g. *me voy* (I go for my own sake) or *te vas/te quedas* (you go/stay for your own sake, or even *te me voy* (I go for my own and for your sake) with two dative clitics) and reflexive verbs (e.g. *arrepentirse* (to be sorry him/herself)); also, attributives have no complements and clitics attach to copulatives as inflexions (*todo le es posible*).

### 3. A Model for Spanish Clitic pronouns

According to the previous discussion, we postulate a model for the clitic pronouns, proclitics and 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 of the explicit arguments. We refer to this operation on the valence of the clitic host as *cliticization*. 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 the duplication of complements, the cliticization operation can optionally, in addition to selecting one or more arguments for cliticization, preserve the cliticized arguments in the valence of the host; in this case, the pronoun and the explicit 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 and, on the other, it 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

---

<http://www.globovision.com/documentos/discursos.transcripciones/2002.08/agasajoatletas/index.shtml>

<sup>20</sup> *lo lleva y pone en práctica*

<http://216.239.59.104/search?q=cache:K3x6pYF PKCgJ:www.juntadeandalucia.es/averroes/iespuerta dearenas/instituto/proyectocentro.pdf+%22lo+lleva+y+pone%22+&hl=en>

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<sub>DAT</sub>*”, “*se<sub>ACC</sub>*”, “*te<sub>DAT</sub>*”, “*te<sub>ACC</sub>*”,  
 “*me<sub>DAT</sub>*”, “*me<sub>ACC</sub>*”, “*lo<sub>ACC</sub>*”,  
 “*la<sub>ACC</sub>*”, “*le<sub>DAT</sub>*”, “*le<sub>ACC</sub>*”,  
 “*los<sub>ACC</sub>*”, “*las<sub>ACC</sub>*”, “*les<sub>DAT</sub>*”,  
 “*les<sub>ACC</sub>*”, “*nos<sub>ACC</sub>*”, “*nos<sub>DAT</sub>*”

(26) “*se<sub>DAT</sub> te<sub>ACC</sub>*”, “*se<sub>DAT</sub> me<sub>ACC</sub>*”,  
 “*se<sub>DAT</sub> le<sub>ACC</sub>*”, “*se<sub>DAT</sub> la<sub>ACC</sub>*”,  
 “*se<sub>DAT</sub> lo<sub>ACC</sub>*”, “*se<sub>ACC</sub> te<sub>ACC</sub>*”,  
 “*se<sub>ACC</sub> me<sub>ACC</sub>*”, “*se<sub>ACC</sub> le<sub>ACC</sub>*”,  
 “*se<sub>ACC</sub> la<sub>ACC</sub>*”, “*se<sub>ACC</sub> lo<sub>ACC</sub>*”,  
 “*se<sub>ACC</sub> te<sub>DAT</sub>*”, “*se<sub>ACC</sub> me<sub>DAT</sub>*”,  
 “*se<sub>ACC</sub> le<sub>DAT</sub>*”, “*te<sub>ACC</sub> me<sub>ACC</sub>*”,  
 “*te<sub>ACC</sub> la<sub>ACC</sub>*”, “*te<sub>ACC</sub> lo<sub>ACC</sub>*”,  
 “*te<sub>DAT</sub> me<sub>ACC</sub>*”, “*te<sub>DAT</sub> la<sub>ACC</sub>*”,  
 “*te<sub>DAT</sub> lo<sub>ACC</sub>*”, “*me<sub>ACC</sub> le<sub>ACC</sub>*”,  
 “*me<sub>ACC</sub> la<sub>ACC</sub>*”, “*me<sub>ACC</sub> lo<sub>ACC</sub>*”,  
 “*me<sub>DAT</sub> le<sub>ACC</sub>*”, “*me<sub>DAT</sub> la<sub>ACC</sub>*”,  
 “*me<sub>DAT</sub> lo<sub>ACC</sub>*”

(27) “*se<sub>DAT</sub> nos<sub>ACC</sub>*”, “*se<sub>DAT</sub> les<sub>ACC</sub>*”,  
 “*se<sub>DAT</sub> las<sub>ACC</sub>*”, “*se<sub>DAT</sub> los<sub>ACC</sub>*”,  
 “*se<sub>ACC</sub> nos<sub>ACC</sub>*”, “*se<sub>ACC</sub> les<sub>ACC</sub>*”,  
 “*se<sub>ACC</sub> las<sub>ACC</sub>*”, “*se<sub>ACC</sub> los<sub>ACC</sub>*”,  
 “*se<sub>ACC</sub> nos<sub>DAT</sub>*”, “*se<sub>ACC</sub> les<sub>DAT</sub>*”,  
 “*te<sub>ACC</sub> nos<sub>ACC</sub>*”, “*te<sub>ACC</sub> las<sub>ACC</sub>*”,  
 “*te<sub>ACC</sub> los<sub>ACC</sub>*”, “*te<sub>DAT</sub> nos<sub>ACC</sub>*”,  
 “*te<sub>DAT</sub> las<sub>ACC</sub>*”, “*te<sub>DAT</sub> los<sub>ACC</sub>*”,  
 “*nos<sub>ACC</sub> la<sub>ACC</sub>*”, “*nos<sub>ACC</sub> lo<sub>ACC</sub>*”,  
 “*nos<sub>ACC</sub> las<sub>ACC</sub>*”, “*nos<sub>ACC</sub> los<sub>ACC</sub>*”,  
 “*nos<sub>DAT</sub> las<sub>ACC</sub>*”, “*nos<sub>DAT</sub> los<sub>ACC</sub>*”

The number of pronoun sequences is rather small as not all possible combinations are allowed: *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*. 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: *se* is always in front, second persons are in front of first ones and third persons (except *se*) are always last. However, some dialects have a different order rule and allow 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.

In the model, proclitics are realized as independent words from the clitic lexicon directly and enclitics are channeled as inputs to the morphological operation that produces the inflected form, and the whole enclitic is lexically realized as a ready made word. In this way, the clitic order is preserved and no clitic sequence can be realized 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 cliticization 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 cliticizations 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 cliticization, clitic insertion and clitic cancellation, and the three operations are implemented through a single lexical rule, and we refer to this process as *Simple Lexical Cancellation (SLC)*; however, if the pronouns appear as enclitics but the clitic host is different from the phonological host, clitic insertion and clitic cancellation are implemented by the same lexical rule, but a different lexical rule is needed for the cliticization of the host; we refer to this latter operation as *Complex Lexical Cancellation (CLC)*. Also, according to the Spanish grammar, tensed forms and participles cannot have enclitics and clitic insertion cannot be performed on these forms; this is a morphological constraint on the corresponding operations.

Most 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 its phonological host. With these elements and definitions, we

formulate the cancellation operation for clitic constructions as follows:

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

Definition (28) allows us to account for sentences in (1) which are repeated here in (29) for clarity: we also show the corresponding kind of cancellation operation and the word of the clitic lexicon that is used in clitic insertion:

- (29) a. *El muchacho da*  
           [*la copa*]<sub>i</sub> [*a la novia*]<sub>j</sub>  
           *The groom gives*  
           *the glass<sub>i</sub> to the bride<sub>j</sub>*
- b. *dala<sub>i</sub> [a la novia]<sub>j</sub>*  
           SLC: “*la<sub>ACC</sub>*”
- c. *dale<sub>j</sub> [la copa]<sub>i</sub>*  
           SLC: “*le<sub>DAT</sub>*”
- d. *dase<sub>j</sub>la<sub>i</sub>*  
           SLC: “*se<sub>DAT</sub> la<sub>ACC</sub>*”
- e. *dase<sub>j</sub>la<sub>i</sub> [a la novia]<sub>j</sub>*  
           SLC: “*se<sub>DAT</sub> la<sub>ACC</sub>*”
- f. *la<sub>i</sub> da [a la novia]<sub>j</sub>*  
           SC: “*la<sub>ACC</sub>*”
- g. *le<sub>j</sub> da [la copa]<sub>i</sub> [a la novia]<sub>i</sub>*  
           SC: “*le<sub>DAT</sub>*”
- h. *se<sub>j</sub> la<sub>i</sub> da*  
           SC: “*se<sub>DAT</sub> la<sub>ACC</sub>*”
- i. *se<sub>j</sub> la<sub>i</sub> da [a la novia]<sub>i</sub>*  
           SC: “*se<sub>DAT</sub> la<sub>ACC</sub>*”

Definition (28) 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 cliticized verb, preventing clitic syntactic cancellation. In (30), we show how the operation of clitic cancellation accounts for the sentences with simple clitic climbing in (5):

- (30) a. *No he podido escribir [el post]<sub>i</sub>*  
           *Not I-have been-able to-write the post<sub>i</sub>*  
           *I have not been able to write the post<sub>i</sub>*
- b. *No he podido escribirlo<sub>i</sub>*
- c. *No lo<sub>i</sub> he podido escribir*

In (31), we show how clitic cancellation accounts for the sentences with auxiliaries and modals:

- (31) a. *Pude haber querido escribir [el post]<sub>i</sub>*

*I-could to-have wanted to-write the post<sub>i</sub>*

*I could have wanted to write the post<sub>i</sub>*

b. *Pude haber querido escribirlo<sub>i</sub>*  
           SLC: “*lo<sub>ACC</sub>*”

c. *Pude haberlo<sub>i</sub> querido escribir*  
           CLC: “*lo<sub>ACC</sub>*”

b. *Lo<sub>i</sub> pude haber querido escribir*  
           SC: “*se<sub>DAT</sub> lo<sub>ACC</sub>*”

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

### 3.1 Local clitic domain and clitic scope

The operation of clitic cancellation in (28) can handle the basic clitic phenomenon including simple clitic climbing; however, complex clitic climbing (i.e. periphrasis where the clitic pronouns in the sequence are arguments of different verbs) present additional problems. Consider (9.a) to (9.c) repeated here as (32). Sentences (32.b) and (32.c) undergo two different cliticizations, two clitic insertions, and two clitic cancellation operations violating the restriction of a single clitic cancellation per clitic sentence.

- (32) a. *Oí [a el]<sub>i</sub> decir [el comentario]<sub>j</sub>*  
           *I-hear to him<sub>i</sub> to-say the comment<sub>j</sub>*  
           *I hear him to say the comment*

b. *Oyélo<sub>i</sub> decirlo<sub>j</sub>*     SLC + SLC  
           *Lo<sub>i</sub> oi decirlo<sub>j</sub>*     SC + SLC

However, these problems are only apparent, as in (32.b) and (32.c) there are two independent clitic sequences attached to two different phonological hosts that are cancelled with their corresponding cliticizations. To capture this phenomenon, we define the *local clitic domain* of a clitic host as the list of its cliticized arguments. Also, for interpreting a pronoun successfully, its corresponding local 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. In (32.b) *lo<sub>i</sub>* and *lo<sub>j</sub>* are in their base positions already and their phonological and clitic hosts are the same; in (32.c) both local clitic domains are within the scope of their corresponding phonological host. However, if a clitic domain “captures” a pronoun, and at the same time blocks the binding path between another pronoun and a local clitic domain, not all binding relations can be established and the

sentence is rendered ungrammatical; for instance *\*lo óyelo decir*.

In simple and complex periphrasis, the verbs with semantic content are the ones that cliticize 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 a local clitic domain and its corresponding phonological host. For this, we define a local clitic domain for every periphrastic verb, and say that two local domains define a *clitic domain* if their corresponding local clitic domains are accessible and, subsequently, there is a binding path between them; the accessibility relation is transitive and a clitic domain can be formed by more than two local clitic domains, and in the limiting case, all local clitic domains of a periphrasis can form a single 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 a 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, *SLC*, *CLC* and *SC*, the phonological host and its corresponding clitic host must belong to the same clitic domain. With these observations in mind, we modify the operation of clitic cancellation as follows:

(33) *Clitic cancellation* (version 2):

There is a single clitic insertion and a single cancellation for each for each independent clitic domain. Cancellation is allowed if every clitic host is within the scope of its corresponding phonological host.

As an illustration of (33) consider (34):

- (34) a. *Pude haber oído [a el]<sub>i</sub> decir [el comentario]<sub>j</sub>*  
 I-could have heard to him<sub>i</sub>  
 to-say the comment<sub>j</sub>  
 I could have heard him to  
 say the comment
- b. *Lo<sub>i</sub> pude haber oído decirlo<sub>j</sub>*

In (34.b) the local clitic domain of *decir* is also its clitic domain; however, the proclitic *lo<sub>i</sub>* 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 *oído*, but the the verbs *puede*, *haber* and *oído* belong to the same clitic domain and clitic cancellation is allowed. In this example there are two independent clitic domains and

the restriction that there is one phonological host and one cancellation for each cliticization holds; in addition, there are no pronouns blocking the binding path between phonological and clitic host in any of the two clitic domains.

### 3.2 Clitic composition

Next, we consider the clitic composition in (35) which cannot be accounted for by clitic cancellation as stated in (33); in (35.b) the pronoun *lo<sub>j</sub>* climbs up and moves out from its local clitic domain, forming the sequence *óyese<sub>i</sub> + lo<sub>j</sub>*, and although there are two cliticizations forming two different local clitic domains, there is a single clitic insertion (i.e. of the word “*se<sub>ACC</sub> lo<sub>ACC</sub>*”) both as enclitic in (35.b) and as proclitic in (35.c), violating the restriction of a single cancellation per independent clitic domain. However, the verbal sequence *oi decir* can be thought of as a composite predicate with two accusative objects: the one who is heard to say something and the thing that this individual says; as the local clitic domains of these two verbs are accessible, they can be combined forming a composite clitic domain; alternatively, we say that the two verbs form a composite clitic host. Then, the clitic word can be inserted attached to its phonological host and cancelled with its corresponding cliticization in a single operation by complex lexical cancellation, as shown in (35.b) or by syntactic cancellation as in (35.c).

- (35) a. *Oí [a el]<sub>i</sub> decir [el comentario]<sub>j</sub>*  
 I-hear to him<sub>i</sub> to-say the comment<sub>j</sub>  
 I hear him to say the comment
- b. *Oyése<sub>i</sub>lo<sub>j</sub> decir CSC + Comp*
- c. *Se<sub>i</sub> lo<sub>j</sub> oí decir SC + Comp*

Clitic compositions can also climb up as atomic units in periphrasis with auxiliary and modals through a clitic domain, as illustrated in (36).

- (36) a. *Pude haber oído [a el]<sub>i</sub> decir [el comentario]<sub>j</sub>*  
 I-could have heard to him<sub>i</sub>  
 to-say the comment<sub>j</sub>  
 I could have heard him to  
 say the comment
- b. *\*Pude haber oído<sub>i</sub> decirlo<sub>j</sub>*
- c. *\*pude haber oído<sub>i</sub>lo<sub>j</sub> decir*  
*Comp: se<sub>i</sub> + lo<sub>j</sub>*
- d. *pude habérse<sub>i</sub>lo<sub>j</sub> oído decir*  
*CLC*
- e. *\*pudése<sub>i</sub>lo<sub>j</sub> haber oído decir*  
*SC*
- f. *Se<sub>i</sub> lo<sub>j</sub> pude haber oído decir*

In (36.b) the direct objects of *oído* and *decir* are cliticized (with *leísmo* for *oído*) in their corresponding local clitic domains; in (36.c) the direct object of *decir* climbs up to a position next to the cliticized direct object of *oído*, with the compulsory substitution of *le* by *se* but preserving the accusative case, forming a clitic composition present in the clitic lexicon; in (36.d) the clitic composition climbs up and it is realized as enclitic to *haber*; in (36.e) the composition climbs up again as an enclitic to *puede* and, finally, in (36.f) it climbs up to the front of the periphrasis and is realized as a proclitic to *pude*. The grammatical sentences in the sequence (36.d) and (36.f) are analyzed in terms complex lexical cancellation and syntactic cancellation respectively. On the other hand, sentences (36.b), (36.c) and (36.e) are ungrammatical due to the restrictions on phonological host; however, the ungrammatically is marginal as these sentences can be perfectly understood, and can be used in pedantic discourse, as in previous examples. On the basis of these observations we introduce the operation of clitic composition as follows:

(37) *Clitic composition*: a sequence of two clitic pronouns next to each other that represent arguments of different verbs forms a clitic composition. The sequence of cliticized verbs forms a composite clitic host that has a composite local clitic domain.

We also reformulate the definition of clitic cancellation to account for the composition of cliticizations:

(38) *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, simple or composite, is within the scope of its corresponding phonological host.

### 3.3 Clitic subsumption

Next, we consider clitic compositions that unlike the previous examples have a beneficiary or a reflexive connotation. This is the case when the clitic word has a pronoun with dative case (i.e. “*se<sub>DAT</sub> la<sub>ACC</sub>*”), as it is the case in (39.b) and (39.c):

(39) a. *Pude haber visto [a usted]<sub>i</sub>  
comiendose<sub>i</sub> [la cena]<sub>j</sub>*  
I-could have seen to you<sub>i</sub>  
eating the dinner<sub>j</sub>

I could have seen you eating  
the dinner

- b. *La<sub>i</sub> pude haber visto comerse<sub>i</sub>la<sub>j</sub>*  
c. *Se<sub>i</sub> la<sub>j</sub> pude haber visto comer*

In the base sentence of this sequence the verb *comiendo* has already the dative *se* as enclitic, and the cliticization of its direct object produces the clitic word *se<sub>i</sub>la<sub>j</sub>* which is realized as enclitic to *comiendose*; however, according to our terminology, this sequence is not a composition because both of the pronouns represent arguments of the same verb. Furthermore, the object of *visto* appears as proclitic to the whole of the periphrasis, and (39.b) is analysed with two clitic cancellations, one for each clitic domain. Now, consider (39.c) in which the clitic word has climbed up to the front of the periphrasis, but the accusative object of *visto* is not explicitly realized. Here, we notice that the dative *se* refers also to the subject agent of *comer*, who is the one benefited by the eating action, and *comerse<sub>i</sub>la<sub>j</sub>* is a reflexive construction; in addition, this argument is also object of *visto*, and these three argument positions are co-indexed. However, in (32.c) only the dative realization is present and we say that the accusative *la<sub>i</sub>* is subsumed by the dative *se<sub>i</sub>*. We illustrate this argument reduction operation with the sequence in (40).

- (40) a. *Pude haber visto [a usted]<sub>i</sub>  
comiendose<sub>i</sub> [la cena]<sub>j</sub>  
[por/para usted]<sub>i</sub>*  
I-could have seen to you<sub>i</sub>  
eating the dinner<sub>j</sub> for you  
I could have seen you eating  
the dinner for you own sake  
b. *\*pude haber vistola<sub>i</sub>  
comiendose<sub>i</sub>la<sub>j</sub>*  
c. *\*pude haber vistola<sub>i</sub>+se<sub>i</sub>la<sub>j</sub>  
comiendo*  
d. *\*pude haber vistose<sub>i</sub>la<sub>j</sub>  
comiendo (i.e. se<sub>i</sub> = la<sub>i</sub>+se<sub>i</sub>)*  
e. *pude haberse<sub>i</sub>la<sub>j</sub> visto comiendo*  
f. *\*pudese<sub>i</sub>la<sub>j</sub> haber visto  
comiendo*  
g. *Se<sub>i</sub> la<sub>j</sub> pude haber visto  
comiendo*

Sentence (40.a) is never explicitly realized and it is only shown to illustrate the meaning of *se* in *comerse*. (40.b) shows the cliticization of the object of *comer* and also of the additional argument, and the clitic word *se<sub>i</sub>la<sub>j</sub>* appears as enclitic to *comiendo*. In (40.c), the clitic word climbs up next to the already cliticized *vistola<sub>i</sub>*

and the sequence  $la_i+se_i la_j$  of three clitics is formed; however,  $la_i$  and  $se_i$  co-refer and are subsumed into one, which is realized as the dative  $se_i$  in (40.d) forming the clitic composition  $se_i la_j$ ; this composition climbs up again to (40.e) as enclitic to *haber*, and also to (40.g) where is realized as a proclitic to the whole periphrasis. The grammatical forms (40.e) and (40.g) are analyzed in terms *CLC* and *SC* respectively and both cancellations operations use the clitic word “ $se_{DAT} la_{ACC}$ ”. As before, the ungrammaticality of (40.b) to (40.d) and (40.f) depends on the restrictions on phonological hosts. On the basis of these observations, we introduce the operation of clitic subsumption:

- (41) *Clitic subsumption*: co-indexed pronouns in a clitic composition are realized as one, with the dative case prevailing.

We summarize these relations in following principle for well-formed clitic sentences:

- (42) *Clitic principle*: Spanish pronominal clitic sentences can be accounted for in terms of the operations of clitic cancellation (version 3), clitic composition and clitic subsumption; alternatively, a clitic host, simple or composite, must be within the scope of its phonological host.

The notions of composite cliticization and clitic host allows us to account for constructions in which the clitic has a wider scope over coordination as in (23.b). In *Lo lleva y pone en práctica*, the two transitive verbs *lleva* and *pone* undergo independent cliticizations; however, as the cliticizations are co-indexed, a composite clitic host including both of the coordinated verbs can be formed by subsumption, and the clitic principle accounts for the whole construction: one clitic insertion and one cancellation. In the case the proclitic is not factored out of the coordination, there are two independent clitic domains, and the sentence is accounted for by the clitic principle. However, if the clitic is an inflexion lexically attached to its host the clitic cannot have a wider scope; in (24.b) (\**Te vas o quedas*), for instance, both of the verbs are intransitive and the clitics represent additional complements which come attached as inflexions to their hosts and the form cannot be split off by a syntactic operation; in this case the explicit realization of both pronouns is required (i.e. *te vas o te quedas*).

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 cliticization operation for reflexive sentences, in addition of changing the valence of the host, co-indexes the subject with the cliticized argument. In (11.a) and (11.b), for instance, the direct object is cliticized and this argument is co-indexed with the subject; in (11.c) the cliticized argument is the indirect object. In all three examples, the sentence undergoes one cliticization, 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 dative in (12.a) and (12.b) has a similar analysis. In the case of pseudo-reflexives with intransitive verbs, like (12.c), the cliticization rule cliticized 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 lexicalized. The clitic principle also handles the cliticization of attributives, which are lexical operations, and there is one cancellation per clitic domain; in *todo le es posible*, for instance, the copulative is cliticized and its attributive adjective is removed from its complement list, and the structure of the form is accounted for by the clitic principle. 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

Cliticization 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 cliticization, the clitic-list is empty. Clitic

insertion, on the other hand, consists in 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 Pollard and Sag (1994). Clitic words are defined as pairs consisting of a form (i.e. the morphology of the clitic word) 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 cliticized 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 “*se la*”, for instance, is shown in Figure 1. The entry for the ditransitive verb *dar* (to give), including the basic definition of its local clitic domain, is shown in Figure 2.

$$\left[ \begin{array}{l} se \ la \\ SS | LOC \left[ \begin{array}{l} CAT \\ CONT | RESTR \end{array} \left[ \begin{array}{l} HEAD \quad clitic \\ VAL | CL-LIST \quad \langle NP-dat_i, NP-acc_j \rangle \\ \{ [person-rel]_i, [female-rel]_j \} \end{array} \right] \end{array} \right] \end{array} \right]$$

Figure 1. Clitic word

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

Figure 2. Lexeme entry for the verb *dar*

With these elements we can define the basic model for simple lexical cancellation (SLC) and syntactic cancellation (SC). For SLC, cliticization, clitic 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. The cancellation operation removes the cliticized complements from the valence of the verb, performs clitic insertion, and

establishes the corresponding bindings in the semantics.

The analysis of the imperative sentence *dásela* (give it to her) is shown in Figure 4, where the word resulting from lexical realization by means of the SLC 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. 2<sup>nd</sup>. person singular).

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

Figure 3. SLC Lexical rule

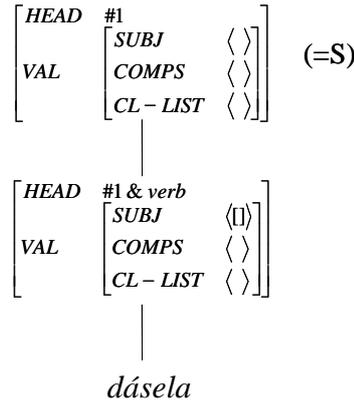


Figure 4. Analysis of enclitic sentences

The first approach to the analysis of syntactic cancellation is equally simple. In this case, cliticization 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, and defining in this way a non-empty local clitic domain. The cliticization rule is defined in Figure 5.

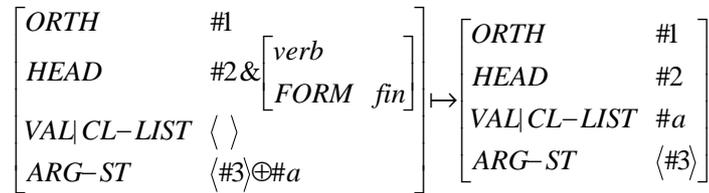


Figure 5. Simple cliticization rule

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. 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 *se la da* (he/she gives it to him/her) 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*.

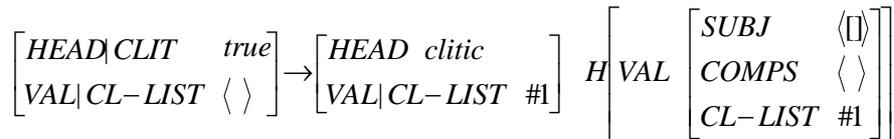


Figure 6. Head Proclitic Rule

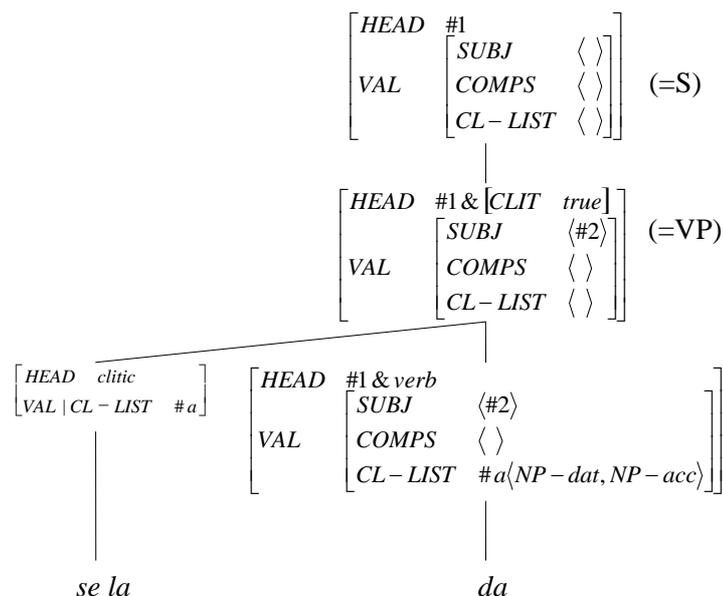


Figure 7. Analysis of proclitic sentence

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. 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 dar* and *puedes dármelo* (you can you give it to me) are ambiguous according to whether *puedes* is interpreted as an auxiliary, implying the possibility of showing something, or as a verb of capacity, in which there needs be an agent capable of performing the showing action.

Auxiliary verbs can also subcategorize for an infinitive, a gerund or a participle VP. Auxiliaries with inchoative 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*)

(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

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 cliticized 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

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.

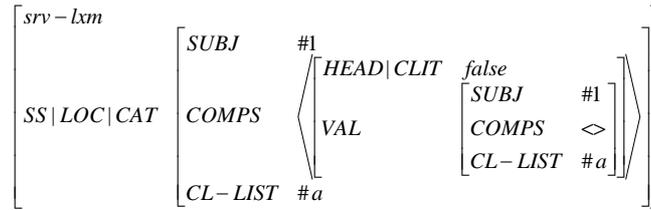


Figure 8. Lexical entry for auxiliary verbs

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 *se la puede dar* 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 dar*. A similar mechanism will be used for subject-control and object-control verbs.

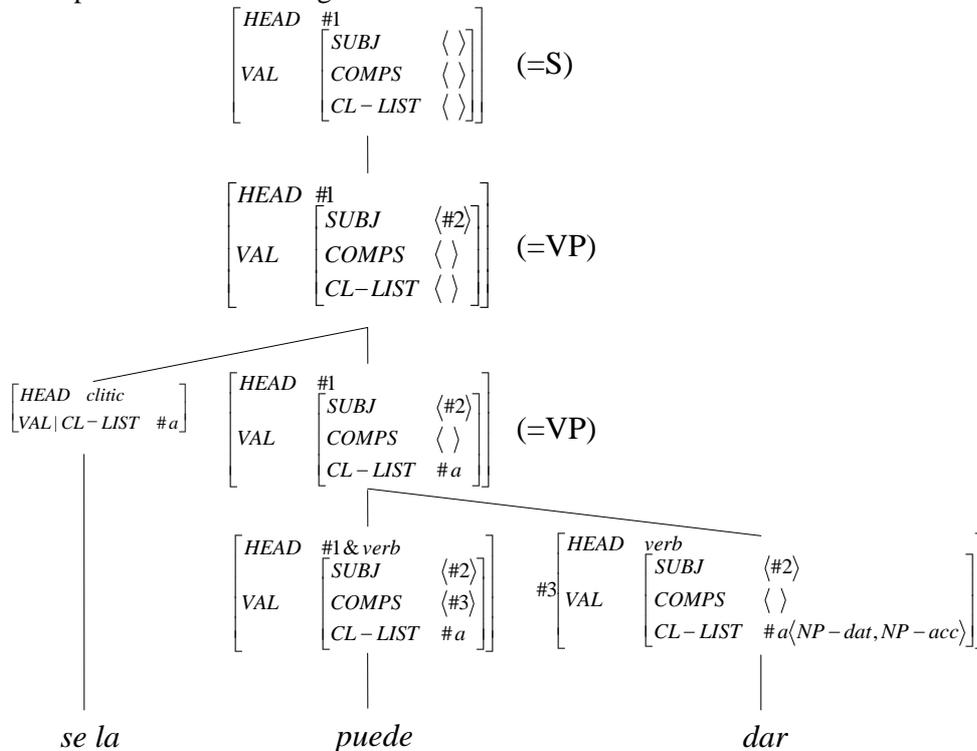


Figure 9. Analysis of simple clitic climbing

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 subcategorization 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 cliticized 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 cliticization. The definition of subject-control verbs is shown in Figure 10.

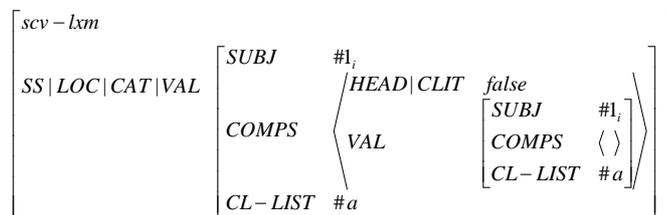


Figure 10. Lexical entry for modal verbs

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<sup>21</sup>. 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 attribute 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 *quiere poder dar la llave a María* (he wants to be able give Maria the key) and its clitic version *se la quiere poder dar* (he wants to be able to give it to her). In this case, the clitic-list of *quiere* is accessible to the clitic-list of *dar* as they share their clitic-list attribute through the clitic-list of *poder*, which establishes the required binding

path. The analysis of this sentence is shown in Figure 11.

We can now provide the analysis for composite lexical cancellation (CLC). In this case, a simple cliticization 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 CLC lexical rule for the infinitive case is shown in Figure 12.

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 habérsela querido dar* (she could have wanted to give it to her) is illustrated. The lexical restriction on tensed forms and participles rules out *puédelo haber querido dar* and *puede haber queridoselo dar* which are ungrammatical.

<sup>21</sup> 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.



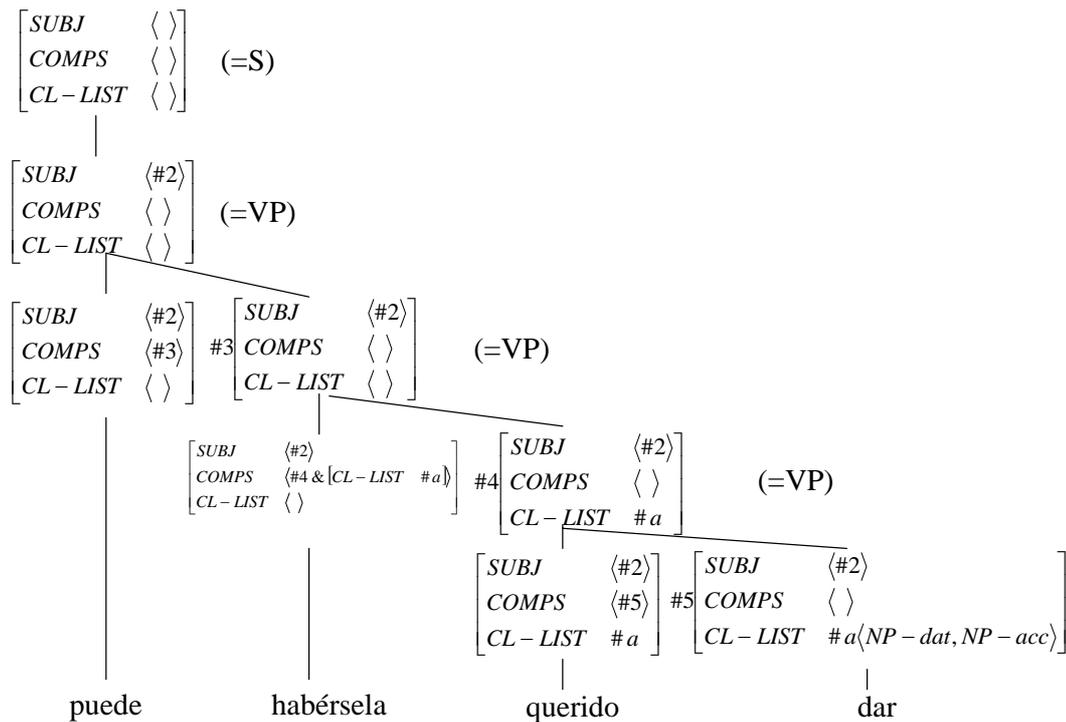


Figure 13. Analysis of sentence with *CLC*

We turn now to the analysis of complex periphrasis with composition and subsumption. The first observation is that for the formation of composite clitic domains there must be more than one verb with a full semantic content in the periphrasis, with cliticizable 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 reír a la gente* (makes the people laugh) the direct object of *hace*, *la gente*, is also the subject of *reír* (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<sup>22</sup>.

To capture cliticization 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

<sup>22</sup> 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; for the factitive verb *hacer*, this alternative seem perhaps more appropriate too: the sense in which people laughing is the complement of factitive verbs seems a much more intuitive analysis than the sense in which a clown makes the people, who happens to laugh. 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); here, for uniformity of analysis, we adopt the two complements approach for the three kinds of verbs, and we leave the issue for further research.

verb at the end of the periphrasis can be subject to cliticization. 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 cliticizations with the clitic-list of its VP-complement. The basic definition of the object-control verb lexeme including cliticization is shown in Figure 15, where  $\#a$  is an empty list.

With this definition we can analyze the clitic versions of *pude haber oido a Juan decir el comentario* as in (36) above. In *Se lo pude haber oido decir* in (36.f), for instance, the accusative *se* stands for *Juan*, the direct object of *oido* and the accusative *lo* stands for *el comentario*, the direct object of *oido*. The cliticization of the *decir* places the direct object in its clitic-list attribute as usual; however, for the cliticization 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 cliticization 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. When this rule is applied, the lexeme definition in Figure 15 is modified as shown in Figure 17.

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 lo pude haber oido decir* in Figure 18. As can be seen, the local clitic domains of both the cliticized verbs are composed in the definition of the local clitic domain of the verbal phrase *oido decir* by the clitic composition operation. As this composite clitic domain is accessible to the phonological host, syntactic cancellation is permitted.

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

Figure 14. Definition of object control-verbs

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

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

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

Figure 16. Cliticization lexical rule for object-control verbs

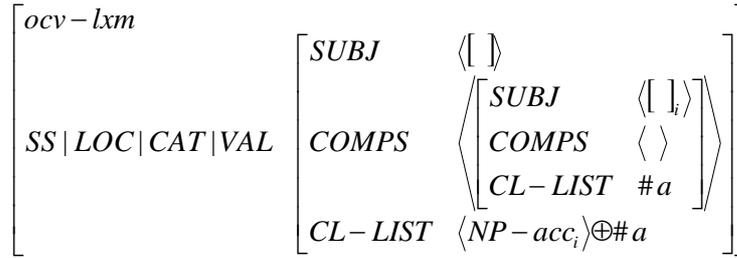


Figure 17. Object-control verb after cliticization

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.

The composition operation is also used when the verb at the end of the periphrasis has its complements as enclitics, as in *le ha oido*

*decirlo*, although in this latter case the clitic list appended to the cliticization of the control verb is empty. In addition, this kind of constructions have two independent clitic domains; the analysis of this sentence, in particular, requires one *SLC* for *decirlo* and one *SC* for *le ha oido* as shown in Figure 20.

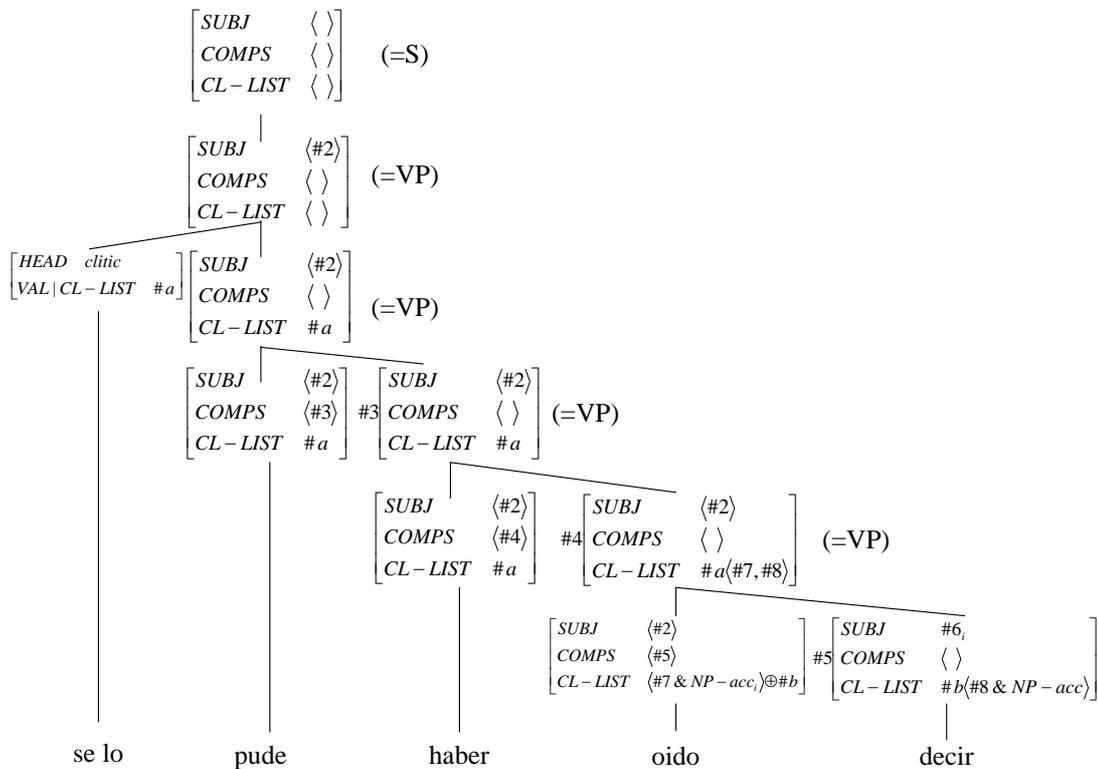


Figure 18. Analysis of sentence with clitic composition and SC

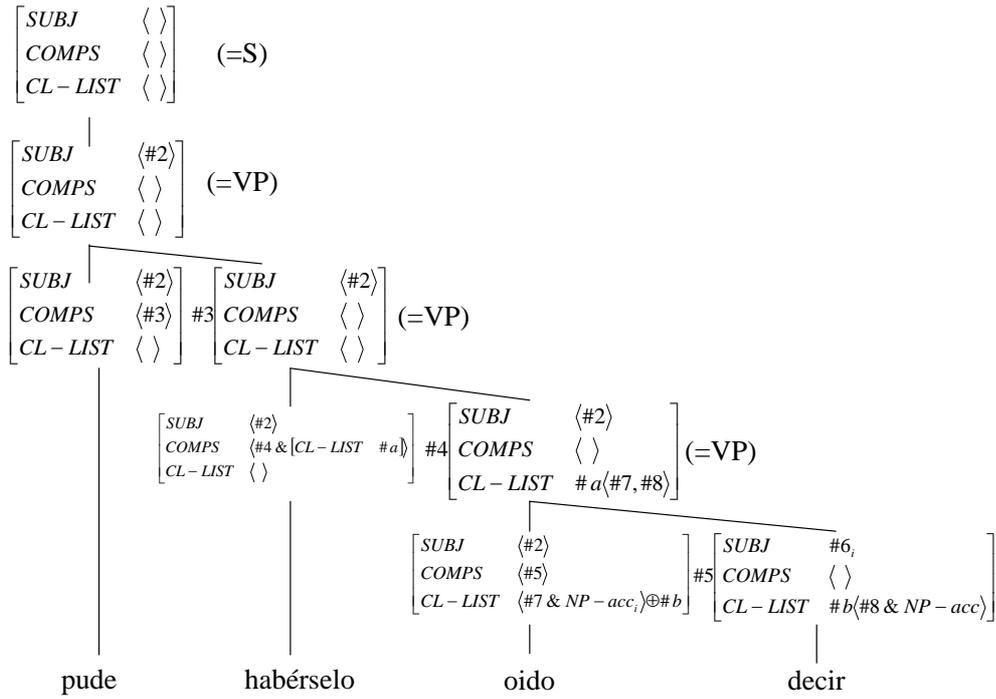


Figure 19. Analysis of sentence with clitic composition and CLC

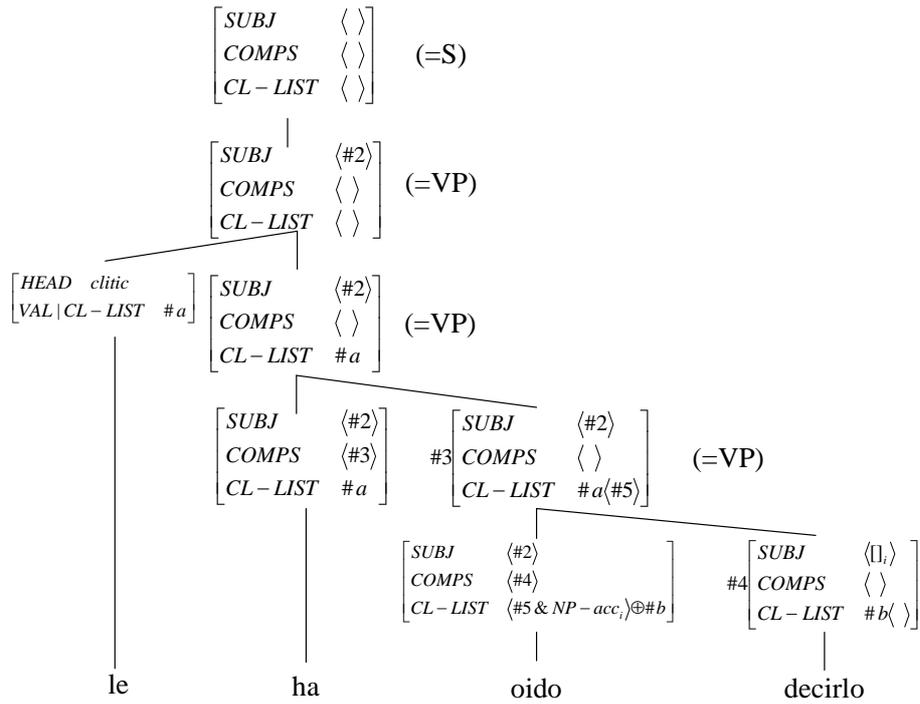


Figure 20. Sentence with two independent clitic domains

The analysis of clitic subsumption is developed along similar lines. Consider the sentence *pude haber visto a María comiéndose la cena* and its clitic version *se la pude haber visto comiendo* in (40.g). As was mentioned in (40.c) and (40.d) the two cliticizations form two accessible clitic domains that can be combined to form an composite clitic domain with co-referential pronouns that can be subsumed. 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 a composite clitic domain by subsuming the accusative clitic of the 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 composite clitic domain is



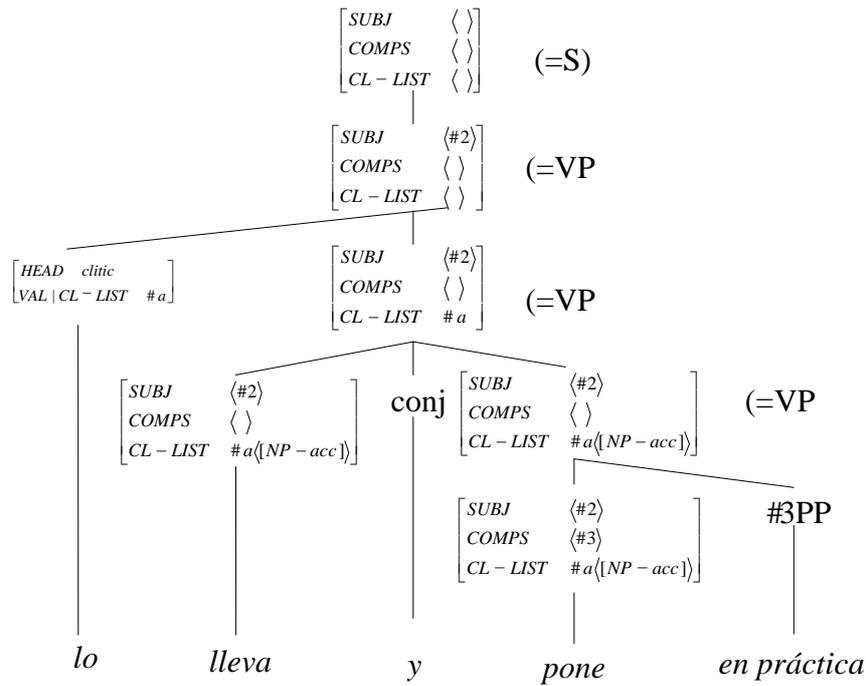


Figure 24. Analysis of clitic coordinated sentence

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 the duplication of arguments and reflexives follows straightforwardly.

In case of duplications, like  $se_j la_i da a la novia_i$  in (1.i), dative arguments are not removed from the complement list, and the corresponding instances in the clitic and complement lists are co-indexed. Accusative complements can also be duplicated, unless

they are subject to specific constraints (i.e. inanimate direct objects). As the complement list must be saturated for clitic cancellation to take place, duplicated complements are consumed before clitic insertion, and they can never appear between a clitic pronoun and its phonological host. A lexical rule for the explicit realization of the clitic and the complements it represents is shown in Figure 25, and the corresponding cliticized entry for the verb *da* is shown in Figure 26; the analysis of (1.i) is shown in Figure 27.

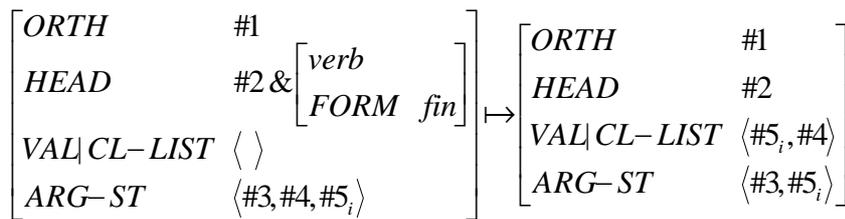


Figure 25. Lexical rule for duplication of the cliticized complement

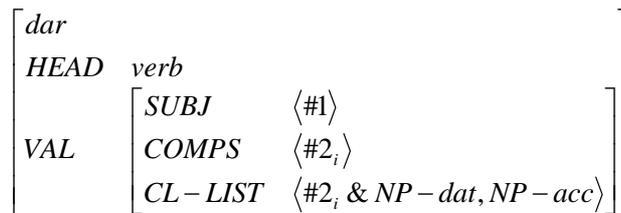


Figure 26. Doubled cliticized lexeme

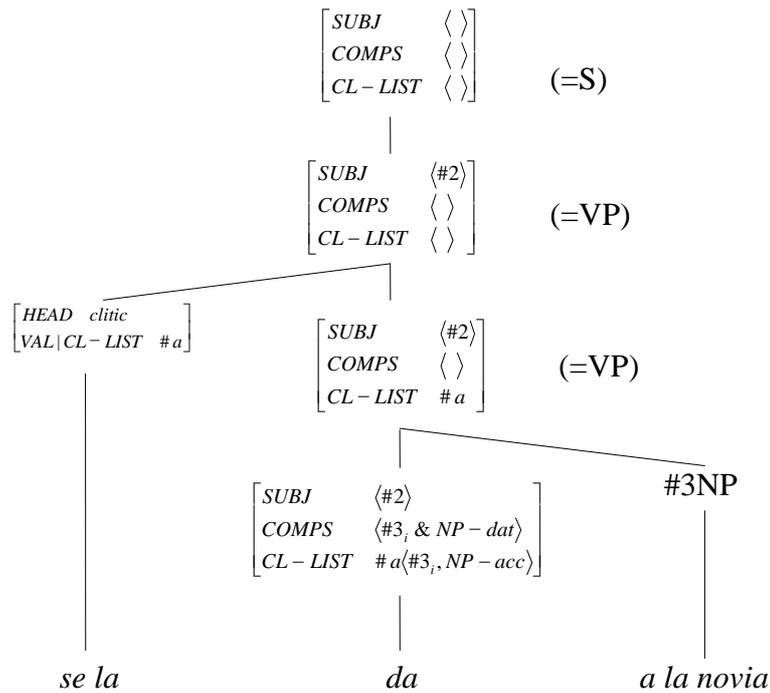


Figure 27. Analysis of clitic sentence with duplication of arguments

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)  $Me_i \text{ lavo}_i$  (I wash myself), the cliticization 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 reflexive cliticization lexical rule is shown in Figure 28, and the analysis of  $me \text{ lavo}$  is shown in Figure 29. The analysis of the indirect reflexive in (11.c)  $me \text{ lavo las manos}$  (I wash my hands) is carried on along similar lines.

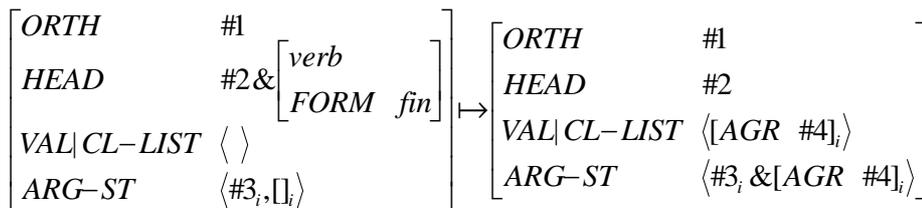


Figure 28. Lexical rule for the direct reflexive

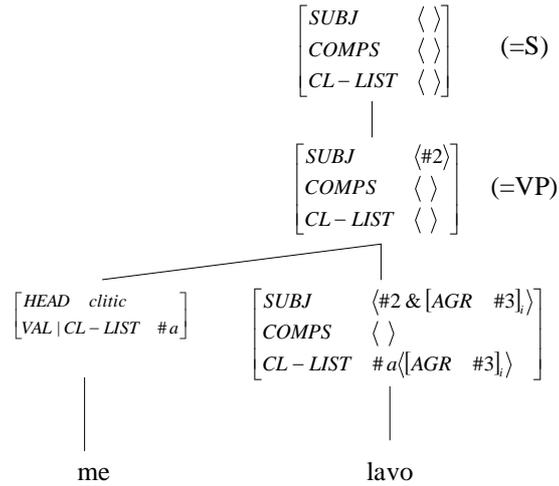


Figure 29. Analysis of direct reflexive sentence

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 cliticization 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 cliticization 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, cliticization is specified in the basic lexeme definition of the verb. Attribute sentences with copulative cliticized (e.g. *le es posible*) receive an analysis along similar lines.

### 5. Impersonal forms

In this section the analysis of the passive-reflexive and the impersonal-active is

presented. The passive-reflexive is handled with a lexical rule that changes the valence attribute of the host. This 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 a placeholder for 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. The passive-reflexive is defined for third-singular persons only, and the *se* word is marked for agreement too. Finally, the lexical rule defines the morphology of the corresponding inflexion of the host. The basic form of 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 syntactic cancellation. Finally, in Figure 33, the semantics of the sentence is presented.

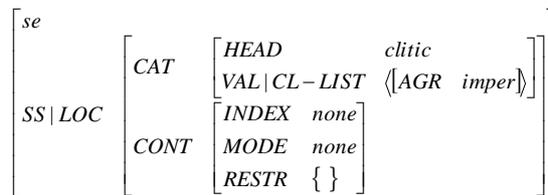


Figure 30. Lexeme definition of the expletive *se*

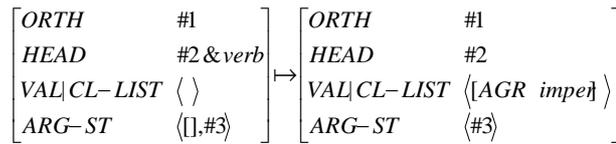


Figure 31. Lexical rule for the passive-reflexive

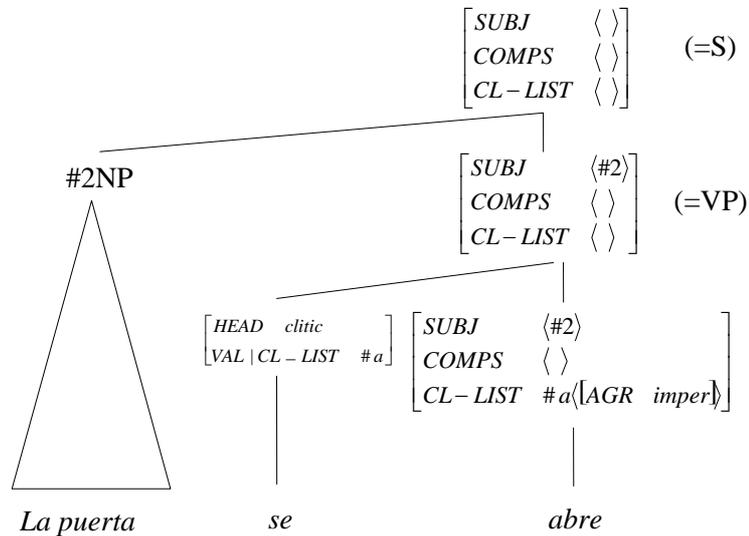


Figure 32. Analysis of the passive-reflexive sentence

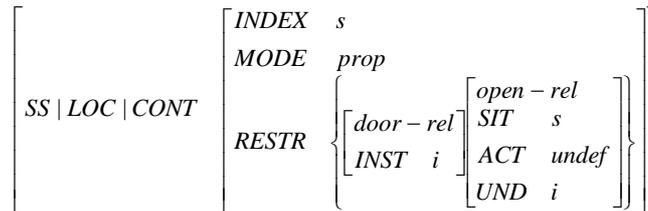


Figure 33. Semantics of the passive-reflexive

The impersonal-active is analyzed along the same lines. This latter form is defined for both transitive and intransitive verbs. The form is also impersonal and the expletive *se* in Figure 30 is used too. Although the forms looks very similar to the passive-reflexive, specially when the verb is transitive and a sentences can be ambiguous between these two (e.g. *se abre la puerta*), this latter form is active, and its explicit argument is not the subject but a complement. 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 (3<sup>rd</sup>-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 a complement cliticized is shown in 35, and the analysis of *se les habla* in 36.



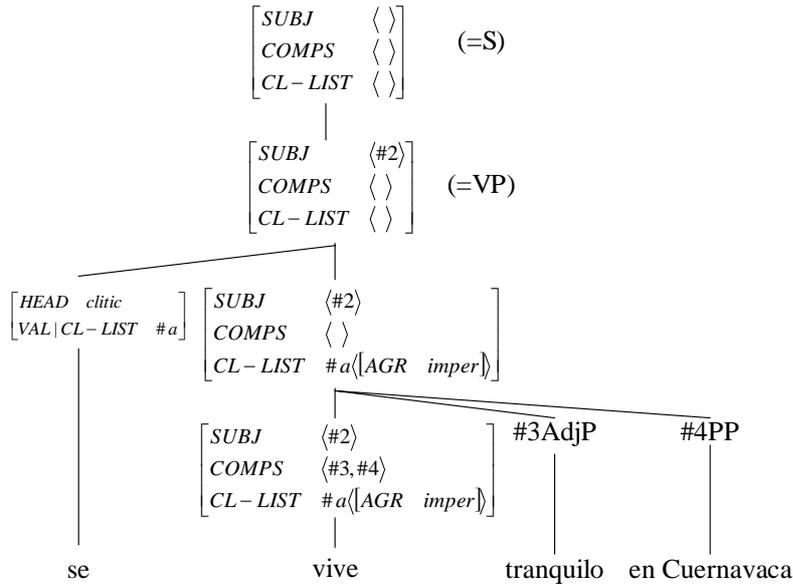


Figure 38. Analysis of the impersonal active with an stative verb

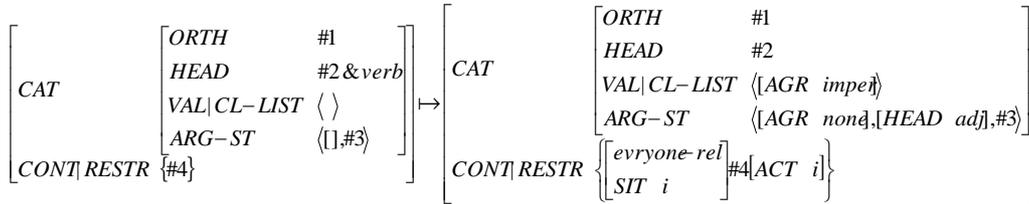


Figure 37. Lexical rule for the impersonal active with intransitive verb

## 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 starting 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 (Villaseñor *et al.*, 2001) and the examples in this paper. *SLC* is implemented through a single lexical rule and *CLC* requires one rule for cliticization 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 cliticization 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, most predicative sentences have a number of clitic versions in which the accusative or dative complement, or both, are substituted by unstressed accusative and dative pronouns respectively, with the 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 that attach to a phonological host and the clitic host that has the pronouns as its arguments, and the form is a case of distant but bounded dependencies.

We have reviewed the main forms of the phenomenon, including the basic form including simple clitic climbing, where all pronouns are arguments of a single verb with full semantic content. 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 presented its implementation in LKB.

In this paper, we have argued that the phenomenon presents a dual behavior, and

although enclitics are considered inflections, proclitics that represent expected verbal complements are treated as independent lexical units and hence proper clitics. On the other hand, pronouns representing arguments that extend the basic argument structure of the verb are considered inflexions, regardless whether they appear as proclitics or enclitics. These forms include verbs naming actions with an additional dative complemente (e.g. *comerse*), verbs of actions that have no recipient (e.g. *irse*, *quedarte*, etc.), pseudo-reflexives (e.g. *te vas*), reflexive verbs (e.g. *disculparse*), and the copulative *ser* and *estar* (to be) in attributives (e.g. *le es fiel*). In summary, the proclitic pronouns are words only when they represent normal complements that are a part of the basic form of the argument structure of the verb; in other cases, these pronouns are also inflexions.

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 concluded that some proclitic pronouns are really clitics. The most important source of evidence is the creation of composite clitic host, that are formed 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 a composite clitic host with their corresponding clitic domains is due to the syntactic scheme that combines the verb with its complements (i.e. The Head-Complement Rule in the HPSG formalization); as a consequence, the subsumption operation interacts with the syntactic coordination rule permitting a single realization of two (or more) proclitics representing complements of different coordinated verbs, which has a wider scope over the coordination. If these pronouns were inflected they could not be split off from their host after lexical realization. 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 were realized fully in the morphology. 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 and distinguish between the notions of host cliticization, 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 cliticized, and the 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 (*SLC*), complex lexical cancellation (*CLC*) and syntactic cancellation (*SC*). All of this is summarized in the operation of clitic cancellation: all Spanish clitic sentences undergo one cliticization and one cancellation per independent clitic domain; for cancellation to take place, the clitic host must be accessible to the phonological host. 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 basic or composite clitic host per independent clitic domain; composite clitic hosts are produced by the operations of clitic composition and clitic subsumption.

The clitic principle captures an economical phonological constraint that results on a structural property of the language: the phonological host attracts all clitic pronouns as a single unit, whether these are local or distant from their host. Clitic words, on the other hand, seem to be ordered sequences of pronouns, strengthened by the local linguistic use. These sequences can have one or two elements, or even three in some dialects, and the set of possible combinations, taken into account case, gender and number, is rather small. 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. 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 could 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 a composite 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 on the phonological host only: imperatives and tensed forms have enclitics but participles and tensed forms need proclitics always. 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 cliticized: *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 cliticization, as no structural

constraint seems to block the accessibility relation between the phonological and clitic host. Every independent cliticization, 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 cliticizations 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. In our model, the different uses of *leísmo* and *laísmo* 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 phonological and prosodic structure of clitic sentences and its interaction with the argument structure of verbs must be addressed explicitly. In particular, the question of whether a pronoun represents a normal complement or a complements that extends the basic argument structure of the verb can only be answered in relation to predefined lexicon, and knowledge of the conditions under which verbal arguments can be extended. Another interesting question for further research is whether there is a single phonological host per independent clitic domain in other languages, and whether the constraints on clitic climbing can be identified as lexical in nature.

### **Acknowledgements**

We thank the support of CONACyT grants C092 and 39380-U.

### **References**

Abeillé, A, Godard, D., Miller, P. and Sag, Ivan. 1998. ‘French Bounded Dependencies’ in Luca Dini and Sergio Balari (eds.), *Romance in HPSG*, Stanford: CSLI Publications.

Copetake, Ann. 2002. The LKB System, Stanford University, <http://www-csli.stanford.edu/~aac/lkb.html>

Gili Gaya, Samuel. 1991. *Curso Superior de Sintaxis Española*. Bibliograf, S. A., Barcelona.

Green, John. 1988. ‘Spanish’, in Martin Harris and Nigel Vincent (eds.), *The Romance Languages*, Oxford University Press, New York, pp. 79–130.

Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 61, 95-120.

Meza, Ivan and Pineda, Luis. 2002. ‘The Spanish auxiliary verb system in HPSG’, *Proceedings of CICling-2002*, Alexander Gelbuck (ed.), Lecture Notes in Computer Science 2276, Springer-Verlag, pp. 200-209.

Miller P. H. and Sag, Ivan. 1995. ‘French Clitic Movement Without Clitics or Movement’, *Natural Language and Linguistic Theory* 15, 573–639.

Monachesi, Paula. 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, Paula. 1999. ‘A Lexical Approach to Italian Cliticization’, Lecture Notes series No. 84, CSLI, Stanford, Cambridge University Press.

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, Luis, Massé, Antonio, Meza, Ivan, Salas, Miguel, Schwarz, Erik, Uruga, Esmeralda and Villaseñor, Luis. 2002. ‘The Dime project’, *Proceedings of MICAI-2002*, Lectures Notes in Artificial Intelligence 2313, pp.166–175.

Pineda, Luis and Meza, Ivan. 2003. *Una gramática básica del español en HPSG*, Reporte interno, Depto. De Ciencias de la Computación, IIMAS, UNAM, Mex.

Pollard, Carl and Sag, Ivan. 1994. *Head-Driven Phrase Structure Grammar*, CSLI, Stanford. The University of Chicago Press, Chicago & London.

Quirino, Ana. 2001, *Spanish clitics: A computational model*, Unpublished MSc. Thesis, Universität Bielefeld.

Rivero-Castillo, Y. 1997. 'Proclitic groups and word order in Caribbean Spanish', in J. R. Black and V. Motapanyana (eds.), *Clitics, Pronouns and Movement, Current issues in*

*Linguistic Theory*, 140. John Benjamins Pub. Co. Amsterdam/Philadelphia. pp. 237–272.

Sag, Ivan and Wasow, Thomas. 1999. *Syntactic Theory: A Formal Introduction*, CSLI Publications, Stanford.

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, A, & Pullum, G. 1983. Cliticization vs. Inflection: English *N'T*, *Language* 59, 502–513.