River Plate Spanish Clitic Packages: An OpenCCG Account of Order Constraints

Víctor M. Castel

Consejo Nacional de Investigaciones Científicas y Técnicas

Universidad Nacional de Cuyo

Abstract

This paper presents a micro-grammar of River Plate Spanish clitics designed within the OpenCCG framework described in Bozşahin *et al* (2007). It addresses the word order constraints underlying the combinatory potential of clitics with other clitics, and clitics with their governing verbs. Clitics are conceived as functor signs that forward seek for arguments that are either verbs or other clitics. In the former case, the function delivers verbal signs with some of theirs arguments appropriately consumed. In the latter case, the function delivers clitic packages that in turn can be sought for by infinitive, gerund and imperative endings to deliver verbal signs so that again the relevant argument consumption takes place, except that now more than one argument is involved at the same time. The proposed mechanism consists of a single lexical specification for clitics, which is not only independent of the meaning of the clitics but also of the position they occupy in relation to the governing verb. This solution is better than other categorial treatments like, for example, Baschung *et al* (1987) for French, which declare different order constraints depending on whether the clitics occur on the left or right side of their governing verbs.

Key words: River Plate Spanish clitics, Clitic packages, Word order, Combinatory Categorial Grammar, OpenCCG, Parsing and Generation

River Plate Spanish Clitic Packages: An OpenCCG Account of Order Constraints

1 Introduction*

Spanish clitics constitute a thoroughly studied area of research in Hispanic Linguistics from both formal and non-formal perspectives (among various others, Grimshaw 2001, Jaegli 1992, Suñer 1980, Luján 1979, García 1975, Bastida 1974, Perlmutter 1971). However, there are no publications on Spanish clitics, let alone River Plate clitics, from the point of view of OpenCCG.[1] Particularly, there are no studies that model the order patterns among clitics and between clitics and the governing verb.

The strategy adopted here is to design, within the OpenCCG environment, a micro-grammar that is capable of accounting adequately and elegantly for the order constraints governing clitic distribution. Since an OpenCCG grammar is an XML declaration of a combinatory categorial grammar, and, thus, it is a complex and extensive object, this paper presents only a simplified, largely graphic, version of the most distinctive traits of the micro-grammar.[2] For the instructions giving access to the complete version of the micro-grammar, namely, the *.xml* files, contact vcastel@lab.cricyt.edu.ar.

Categorial grammars are simple and yet powerful formal systems. OpenCCG is, thus, simple and powerful in this respect. However, since the descriptive burden lies in the specification of lexical entries that are sensitive to the interaction of a "couple" of rules of application and composition of functions, a single set of data can be described, in principle, in more than one adequate manner. This paper does not offer alternative solutions; rather, it advances just one proposal with the idea of triggering a fruitful discussion.

The underlying categorial conceptualization is based on Castel (1994, 1996, 1997), which in turn draws on Zeevat *et al* (1987), and Baschung *et al* (1987). The generalizations on order constraints are, basically, Perlmutter's (1971). The distinctive characteristic of the micro-grammar presented here is that it accounts for the order patterns among clitics and between clitics and their governing verbs. Clitics are conceived as forward functions capable of taking verbs or other clitics as arguments, according to context, to deliver, respectively, verbs with clitics adjoined to their left or clitic packages. Clitic packages so derived are, in turn, arguments that can be taken by imperative, infinitive and gerund endings to deliver verbs with clitics adjoined to their right. A positive consequence of this treatment, and eventually a comparative advantage, lies in the fact that the data addressed can be accounted for with a single set of lexical entries for clitics.[3]

The paper is written under the assumption that potential readers at least know the basics of CCG (Baldridge 2002, Steedman & Baldridge 2003), are familiar with OpenCCG

(Bozşahin *et al* 2007), and are willing to "run" the micro-grammar interactively with the text so that it can be thoroughly evaluated.[4]

2 Order constraints

The order constraints underlying the distribution of clitics in River Plate Spanish can be abbreviated by the following schemas (Perlmutter 1971):

(1) *se* II I III(2) *se* II I III + V[non-imperative personal form](3) V[imperative or infinitive or gerund] + *se* II I III

where '*se* II I III' is a growing monotonic sequence in which the roman numbers stand for clitic person values, *se* refers to the clitic itself, 'V' stands for a verb, the expressions within square brackets refer to forms of 'V', and '+' marks the relative order between the sequence '*se* II I III' and 'V'. The function of these schemas is to capture in an abbreviated way two aspects of order: (i) the constraints controlling the co-occurrence of clitics with other clitics (cf. (1)), and (ii) the constraints controlling the co-occurrence of clitics with their governing verbs (cf. (2)-(3)).

While schema (1) makes full sense in the context of schemas (2) and (3), it is important to make its scope explicit also as a stipulation of the relative order of clitics among themselves. Thus, in sentences containing a single clitic, the problem of the relative order among clitics does not even arise, but the schema stipulates that the clitics to be ordered are those in (4), which, by the way, are the clitics occurring in sentences containing just one clitic. In sentences containing two clitics, the clitic sequences are those in (5); in sentences containing three clitics, the sequences are those in (6); and, in sentences containing four clitics, the sequences are those in (7).

(4) {se, te, me, nos, le(s), lo/a(s)}

(5) {se te, se me, se le(s), se lo/a(s), te me, te le(s), te lo/a(s), me le(s), me lo/a(s)}

(6) {se te me, se te le(s), se te lo/(s), se me le(s), se me lo/(s), te me le(s), te me lo/a(s)}

(7) {se te me le(s), se te me lo/a(s)}

Any clitic sequence differing from those in (5)-(7) is ill-formed. For example, the sequences in (8)-(10), among many others, are ill-formed, i.e. there are no well-formed sentences containing such sequences:

(8) *lo/a(s) me, *lo/a(s) te, *le(s) me, *me te, *te se (And many other sequences of two clitics.)

(9) *lo/a(s) te me, *lo/a(s) me te, *le(s) se te, *lo/a(s) le(s) te (And many other sequences of three clitics.)

(10) *te se me lo/a(s), *te me se le(s) (And many other sequences of four clitics.)

The joint effect of schemas (1)-(3) is to express that clitics, whether separately (cf. (4)) or in co-occurrence with other clitics (cf. (5)-(7)), immediately precede non-imperative personal verb forms (cf. (2)), and immediately follow imperative, infinitival and gerundive verb forms (cf. (3)). The following examples illustrate this distribution:

(11a) *Me lo regalaron*.	(12a) *Quieren regalármelo*.	(13a) Regalámelo.	(14a) Estaba regalándotelos.
(11b) **Lo me regalaron*.	(12b) *Me lo quieren regalar*.	(13b) *Regalálome.	(14b) Te los estaba regalando.
(11c) **Regaláronmelo*.	(12c) **Quieren regalárlome*.	(13c) *Me lo regalá.	(14c) *Estaba regalándoloste.
(11d) **Regaláronlome*.	(12d) **Quieren me lo regalar*.	(13d)*Lo me regalá.	(14d) *Estaba te los regalando.
	(12e) **Quieren lo me regalar*.		(14e) *Estaba los te regalando.

The well-formedness contrast of the sentences in (11) follows from the schemas (1)-(3) as described immediately. Sentence (11a) meets the constraints of schema (1) because the sequence *me lo* satisfies the order 'I III', and the relative order between the sequence *me lo* and the verb *regalaron* is in accordance with the stipulation in (2). However, sentence (11b) is ill-formed because, although the clitics making it up precede the governing verb, the relative order between the clitics *lo* and *me* violates the order constraint 'I III' of schema (1). Sentence (11c) is ill-formed, despite the inherent well-formedness of the clitic sequence *me lo*, because the relative order between such a sequence and *regalaron* violates the stipulation given in schema (2). In the case of sentence (11d), the violation is twofold: the sequence *lo me* violates the constraint 'I III' of schema (1), *and* the relative order between such a sequence and *regalaron* violates the constraint stipulated by schema (2).

The well-formedness contrast of the sentences in (12) also follows from the schemas (1)-(3). Sentences (12a-b) are well-formed because not only do they satisfy the constraint 'I III' of schema (1), but also the constraints imposed by the schemas (3) and (2), respectively. Sentence (12a) meets the order constraint between the sequence *me lo* and *regalar* stipulated by (3), while sentence (12b) meets the order constraint between the sequence *me lo* and the verb *quieren* stipulated by (1). However, sentences (12c-e) are ill-formed. Sentence (12c) violates the constraint 'I III' of schema (1), even though the clitic sequence *lo me* occurs on the right side of *regalar*, as stipulated by schema (3); sentence (12d) meets the constraint 'I III', but it violates both the order constraint between the clitic sequence *me lo* and the verb *regalar* of schema (3), and the order constraint between the clitic sequence *me lo* and the verb *quieren* of schema (2). Sentence (12e) violates both the constraint 'I III' of schema (1), and the order constraint between the clitics and the verbs *quieren* and *regalar*, in accordance with the schemas (2) and (3), respectively.

Similarly, *mutatis mutandis*, the well-formedness contrasts of sentences (13) and (14) are captured by the schemas (1)-(3). The reader is encouraged to try and test the scope of the schemas (1)-(3) with any verb and any clitic combination.

Notice that clitic sequences (cf. (5)-(10)) must obey the order pattern stipulated by schema (1), independently of the position that they occupy in relation to the governing verb (cf. (2)-(3)).

3 An OpenCCG Micro-Grammar of Clitics Order Constraints

As stated above, this paper assumes that the reader is familiar with CCG and OpenCCG. In this section, we refer very briefly to specific aspects of the proposed micro-grammar, which addresses only the order constraints governing the distribution of clitics.[5] A grammar declaration within OpenCCG consists of five *.xml* files, which, for example, in the sample tiny-grammar (https://sourceforge.net/projects/openccg/), are named *grammar*, *lexicon*, *morph*, *rules*, and *types*. These files are constructed in accordance with properties defined in the following schemas (*.xsd* files): *grammar*, *lexicon*, *morph*, *rules*, *types*, *tokens*, *parameters*, *hlds*, *dict*, *categories*. The function of these schemas is to guarantee the validity of the *.xml* files, i.e. that the grammar declared is within the class of grammars allowed by OpenCCG.

The micro-grammar of River Plate Spanish clitics proposed here uses the following *.xsd* and *.xml* files: *SpnshGrmr.xsd*, *SpnshLxcn.xsd*, *SpnshMrph.xsd*, *SpnshRls.xsd*, *SpnshTps.xsd*, *SpnshTkns.xsd*, *SpnshPrmtrs.xsd*, *SpnshHLDS.xsd*, *SpnshDct.xsd*, *SpnshCtgrs.xsd,* *grammar.xml*, *SpnshLxcn.xml*, *SpnshMrph.xml*, *SpnshRls.xml*, *SpnshTps.xml*. Note that most file names differ from the original default names given in OpenCCG. The content of the schemas, however, is identical to the one given in OpenCCG. The new names simply serve the purpose of facilitating our own development and maintenance. The content of the *.xml* files is, of course, specific to the micro-grammar of clitics, except for *grammar.xml*, and a few aspects of *SpnshRls.xml* and *SpnshTps.xml*.

3.1 Valence Deletion Rules

Along with the universal rules of functional application and functional composition, the file *SpnshRls.xml* contains the following type-changing rules:

3.1.1 Non-reflexive Dative Deletion

Dat-CLT-Del: $s_{<1>}\$1|{*}clt_{dat,\ non\text{-}rflxv}/{*}vnd_{<2>} \Rightarrow s_{<1>}\$1/{*}vnd_{<2>}$

3.1.2 Non-Reflexive Accusative Deletion

Acc-CLT-Del: $s_{<1>}\$1|{*}clt_{acc,\ non\text{-}rflxv}/{*}vnd_{<2>} \Rightarrow s_{<1>}\$1/{*}vnd_{<2>}$

3.1.3 Nominative Valence Deletion (Pro-Drop)[6]

Nom-NP-Drop: $s_{<1>prsnl}\$1|_{<}np_{nom} \Rightarrow s_{<1>}\$1$

3.2. Sorts and Feature Value Hierarchies

The file *SpnshTps.xml* contains the hierarchies given in Tables A1.1-1.2 at the end of the paper, which, along with the sorts (= types) specified in the sample tiny-grammar of OpenCCG, define syntactic and semantic values that are used in feature structures and, thus, play a crucial role in the unification process. Although the hierarchy of sorts is too basic, it suffices for the purposes of this paper. As for the other hierarchies, they are

provisional and useful only for the current presentation. It is clear that any extension of the observational domain and, therefore, of the data to be accounted for, is likely to require all kinds of refinements. Table A1.1 purports to be a relatively friendly representation of value relationships, including cases of multiple parents. Table A1.2 is the XML specification of three hierarchies just as they appear in the file *SpnshTps.xml*.

3.3 Verbal Roots and Endings

Since OpenCCG does not have a module for morphological processsing, this paper resorts to the "artificial" separation of verbal roots from their endings. Thus, for example, the micro-grammar of clitics contains separate lexical entries for the root *regal* and the ending *ás*, so that a conjugated verb form such as *regalás* is actually a derivation like [*regal* → *ás*], where '→' indicates that the root *regal* takes *ás* as argument to deliver the sign *regal ás*.[7] The artifice obeys our need to show what exactly the morphosyntactic-semantic contribution is of each constituent of a word like *regalás*.

The workings of the micro-grammar are illustrated with a few essential aspects of the verbal root *regal*, defined in Figures A2-10 in detail.[8] There are four primitive entries for *regal*, two reflexive (Figures A2-3) and two non-reflexive (Figures A4-5). Out of these four entries, five other entries are derived, two reflexive (Figures A7 and A9), and three non-reflexive (Figures A6, A8, A10), by deleting non-reflexive valences; see the rules *Dat-CLT-Del* and *Acc-CLT-Del* of the preceding section. Any of these nine roots can combine, for example, with any of the following verbal endings, among many others:

(15a) *o* :- vnd <1> X, E, ind, sg, 1st, vrt1_0, pres, ind-e

(15b) *ás* :- vnd <1> X, E, ind, sg, 2nd, vrt1_0, pres, ind-e

(15c) *a* :- vnd <1> X, E, ind, sg, 3rd, vrt1_0, pres, ind-e

For example, by combining the root of Figure A4 with the ending (15b), the following sign is delivered:

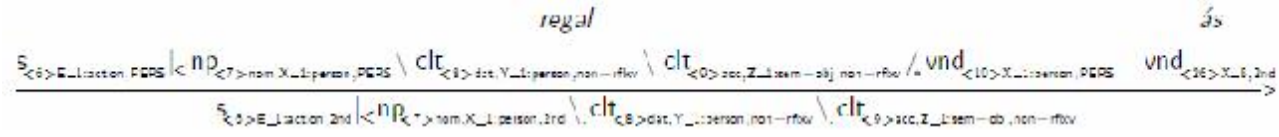

Figure 1. Derivation of *regal ás*.

This sign intervenes in derivations of sentences like *vos me lo regalás*, *me lo regalás vos* and *me lo regalás*, depending on whether the deletion rule *Nom-NP-Drop* is applied or not:

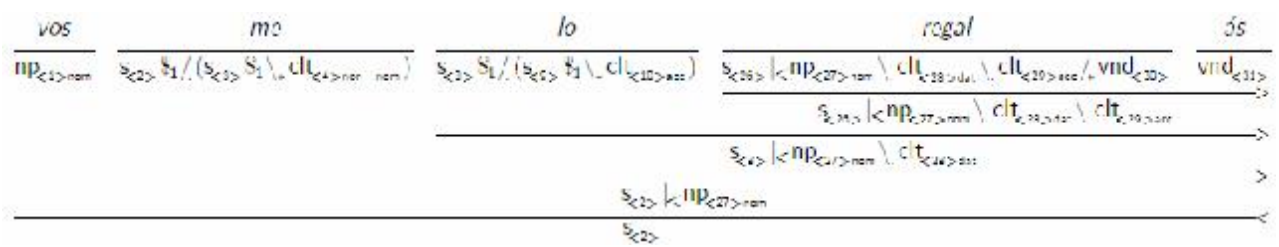

Figure 2. Derivation of *vos me lo regal ás*.

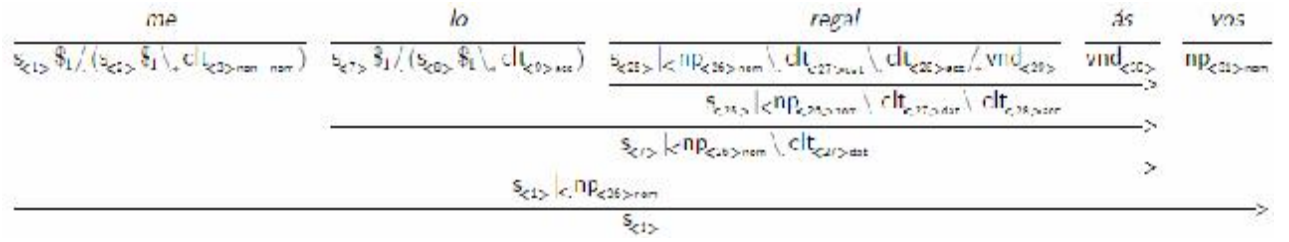

Figure 3. Derivation of *me lo regal ás vos*.

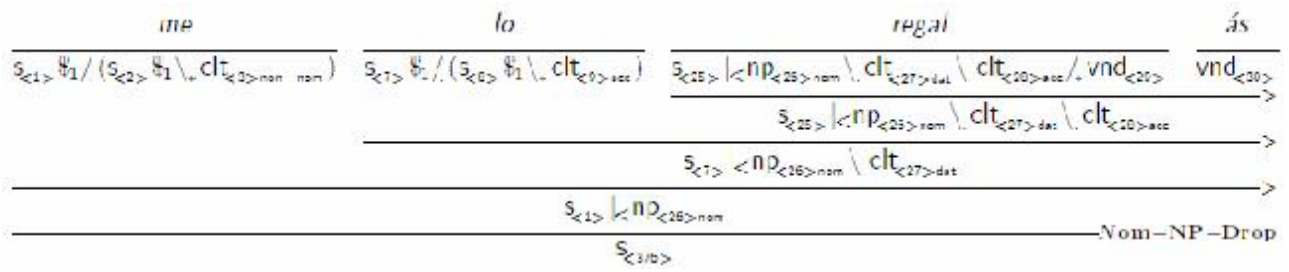

Figure 4. Derivation of *me lo regal ás*.

3.4 Clitics

3.4.1 Clitics as Functors

To account for the morpho-syntactic distribution observed in (5)-(14), clitics are here conceived as functor signs that seek for signs that in turn seek for clitics to deliver signs with the appropriate argument saturation. The micro-grammar contains three different classes of sign that seek for clitics, namely, verbs, verbal endings and clitics themselves, e.g., Figures A10-11. As functors, then, clitics can combine both with verbs and other clitics.

In contrast with *s*, *np*, and *vnd*, we use the atomic category *clt* to refer to clitics. Thus, there are complex categories such as, for example, *s|np\clt*, where both *np* and *clt* are arguments. We say that an expression belonging to such a category is a functor that requires to combine backwards with an expression of category *clt* to deliver an expression of category *s|np*. Then, clitics belong to categories of the following form:

(16) s$1/(s$1\clt) : *lf*

According to (16), clitics seek forward for argument signs of type *s$1\clt* (i.e. signs that backward seek for clitics), and deliver signs of type *s$1*, where *$1* is a variable (possibly null) ranging over one or more arguments.

Each clitic, in turn, defines its specific morpho-syntactic and semantic properties in the feature structures of *s* and *clt*, and also in *lf* of (16). Thus, for example, the clitic *lo* looks like (17) in simplified form:[9]

(17) *lo* :- s[cltc3mrk]$1/.(s[un-hdd, non-imp-e]$1\clt[acc, msc, X, sg, 3rd, non-rflxv]) : … (Cf. Figure A11.)

The simplified version of *le* is as in (18):

(18) *le* :- s[cltc3mrk]$1/.(s[un-hdd, non-imp-e]$1\clt[dat, X, sg, 3rd, non-rflxv]) : …

The clitics *me*, *te* and *se* are as in (19-21) to capture the reflexive and non-reflexive interpretations of which they are capable of participating: (Cf. Figures A12-13 for more complete representations of *me*.)

(19a) *me* :- s[cltc1, 1st]$1/.(s[cltc3mrk, non-imp-e, 1st]$1*clt[non-nom, X, sg, 1st, rflxv]) : …

(19b) *me* :- s[cltc1, non-1st]$1/.(s[cltc3mrk, non-imp-e, non-1st]$1*clt[non-nom, X, sg, 1st, non-rflxv]) : …

(20a) *te* :- s[cltc2, 2nd]$1/.(s[cltc13, non-imp-e, 2nd]$1*clt[non-nom, X, sg, 2nd, rflxv]) : …

(20b) *te* :- s[cltc2, non-2nd]$1/.(s[cltc13, non-imp-e, non-2nd]$1*clt[non-nom, X, sg, 2nd, non-rflxv]) : …

(21) *se* :- s[cltc3unk, 3rd]$1/.(s[cltc213, non-imp-e, 3rd]$1*clt[non-nom, X, sg, 3rd, rflxv]) : …

The constraints controlling the order patterns among clitics (cf. schema (1) and sequences (4)-(10)), and between clitics and their governing verbs (cf. schemas (2)-(3) and sentences (11)-(14)), are captured as described in §§3.4.1.1-2 immediately.

3.4.1.1 Order among Clitics

Clitics forward combine with signs, verbs or other clitics, which satisfy the value associated with the attribute 'lft_hd' in the feature structure of the second *s* in (16);[10] i.e. clitics, as functors, require of argument signs to have a given value for the attribute 'lft_hd', whose possible values are those defined in the column CLITIC HEAD of Table A1.1. Thus, for example, both entries of the clitic *me* require of its argument to have the value 'cltc3mrk', while the clitics *lo* and *le* require of the argument to have the value 'un-hdd'. The sign resulting from the combination, the first *s* in (16), takes the inherent value of the functor clitic, 'cltc3mrk' for *lo* and *le*, and 'cltc1' for *me* in its two entries.

Before combining with other signs, personal verb forms are *not* specified with respect to 'lft_hd'; then, in principle, any clitic can take them as arguments, for the order constraints governing clitics in relation to 'lft_hd' are all compatible with the absence of a specification in verbs, i.e. feature unification is possible. Thus, both *lo regalo* and *me regalo* are correctly parsed by the micro-grammar:

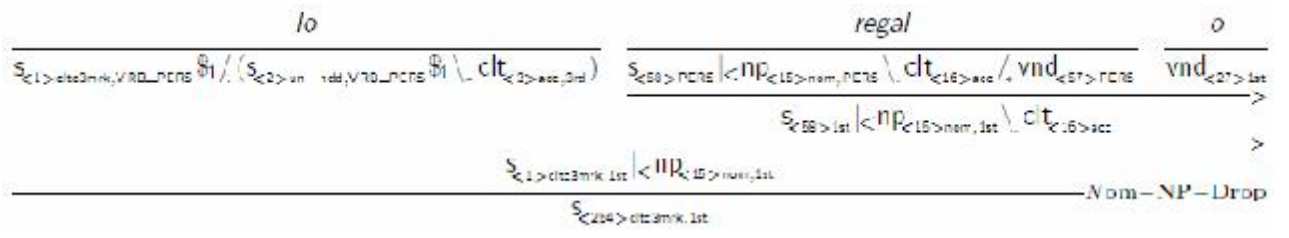

Figure 5

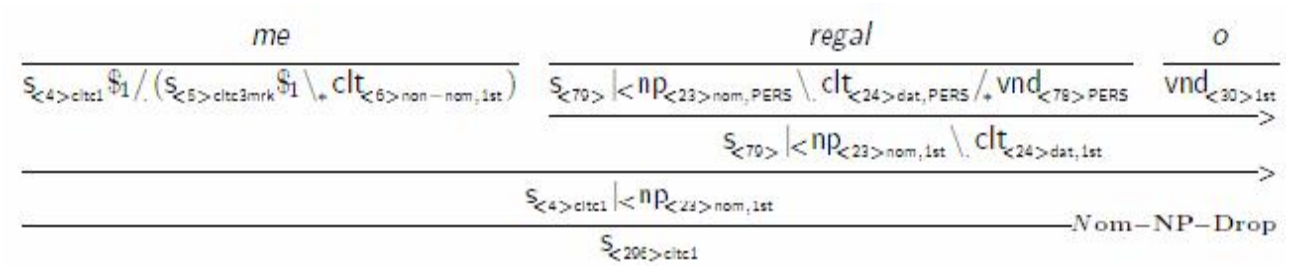

Figure 6

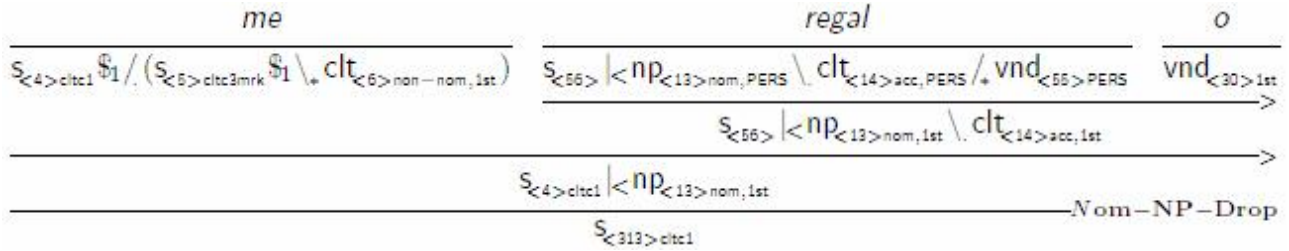

Figure 7

Note now that the resulting signs are indeed specified with a value for the attribute 'lft_hd', namely, 'cltc3mrk' (Figure 5) and 'cltc1' (Figures 6-7), respectively.[11] This specification turns out to play a crucial role in future combinations with clitics. For example, although the signs of Figures 8-9 are accepted by the micro-grammar, those in (22) and (23) are not:

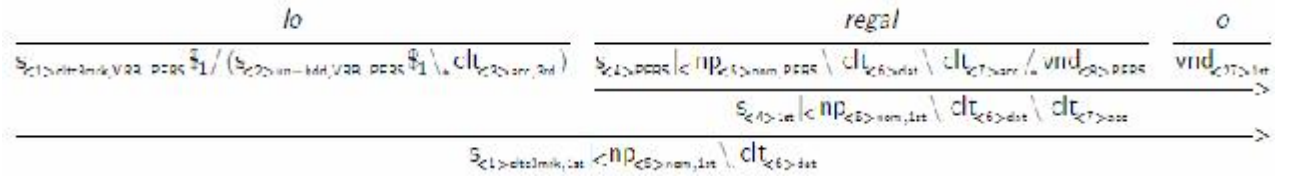

Figure 8

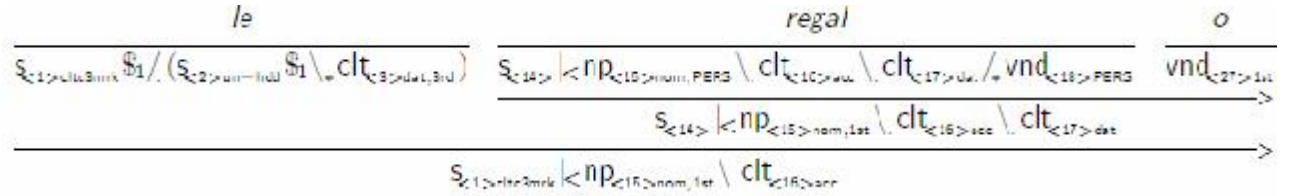

Figure 9

(22) le →× [lo → [regal → o]] (23) lo →× [le → [regal → o]]

The derivations of **le lo regal o* and **lo le regal o* are impossible because the signs *lo regal o* and *le regal o* are *s[cltc3mrk]* (cf. last line of Figures 8-9), whereas the signs *lo* and *le* require an *s[un-hdd]* (cf. (17) and (18)), which causes feature unification to fail. There are, however, derivations like Figures 10-11 for *me lo regalo* and *me le regalo* because the signs for *me* (cf. (19a-b) above) require an *s[cltc3mrk]*, which is precisely the specification of the signs *lo regal o* and *le regal o* in Figures 8-9:

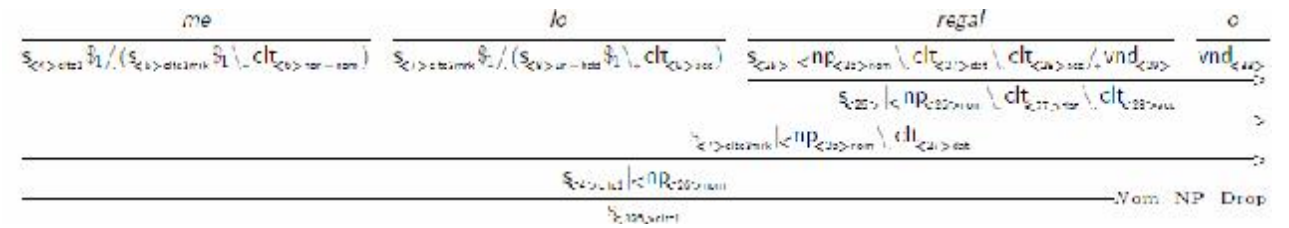

Figure 10

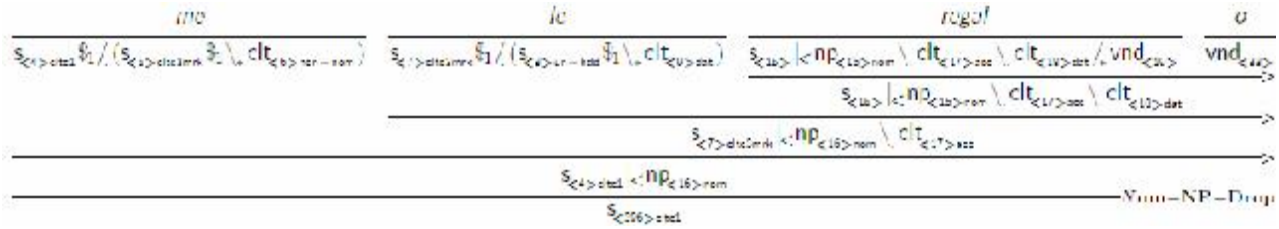

Figure 11

Similarly, the micro-grammar accepts the derivations of Figures 12-13 but not derivations like (24)-(25):

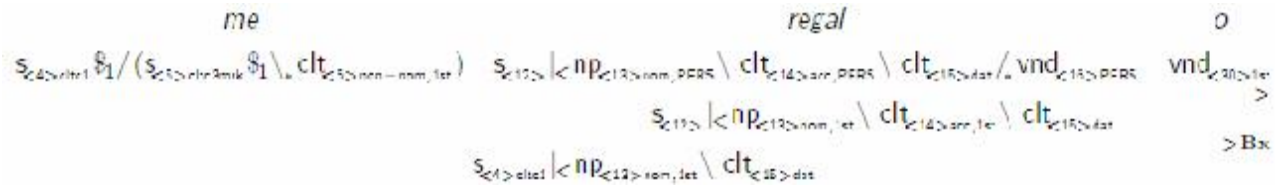

Figure 12

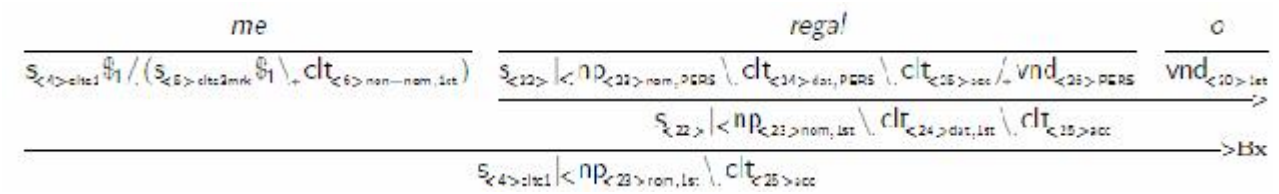

Figure 13

(24) le →× [me → [regal → o]] (25) lo →× [me → [regal → o]]

Derivations for **le me regal o* and **lo me regal o* are not possible because the sign *me regal o* is *s[cltc1]* (cf. last lines of Figures 12-13), whereas the signs *le* and *lo* require an *s[un-hdd]* (cf. (18) and (17) above); again, feature unification fails. Although (24) and (25) are excluded, the following derivations are possible:

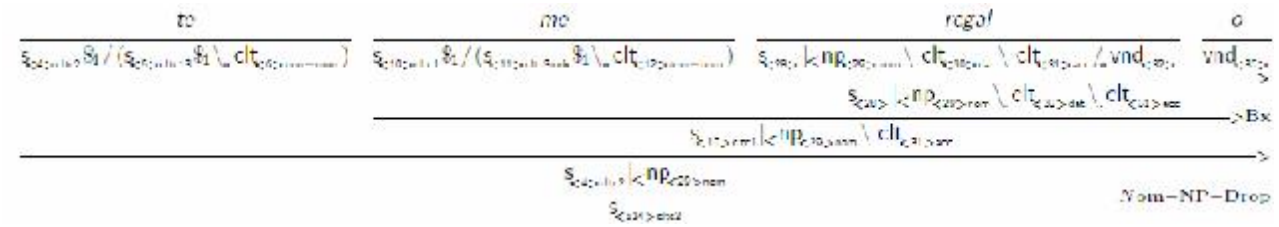

Figure 14

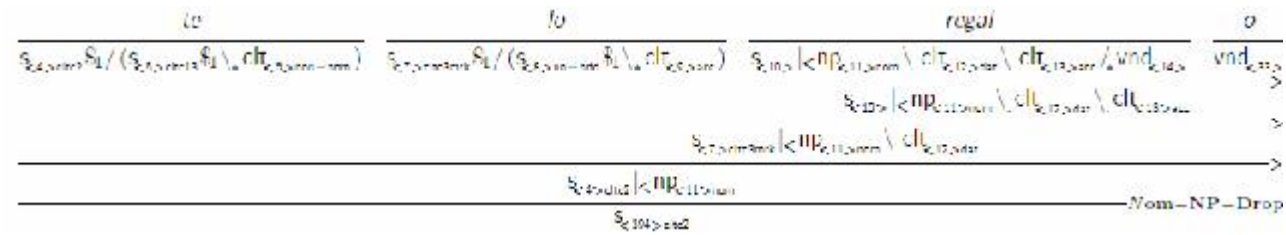

Figure 15

These two derivations illustrate the scope of value ‘cltc13’ as a constraint imposed by *te* on its argument sign. According to the hierarchy of values in column CLITIC HEAD of Table A1.1, ‘cltc1’ and ‘cltc3mrk’ are daughters of ‘cltc13’ and, thus, can unify with ‘cltc13’. The

sign for *me regal o* in Figure 14 has the value 'cltc1' for the attribute 'lft_hd'; the sign for *lo regal o* in Figure 15 has the value 'cltc3mrk' for 'lft_hd'. However, there are no derivations like (26)-(27):

(26) me $\rightarrow\times$ [te $\rightarrow$ [regal $\rightarrow$ o]] (27) lo $\rightarrow\times$ [te $\rightarrow$ [regal $\rightarrow$ o]]

Being specified as *s[cltc2]*, the sign *te regal o* in (26) and (27) cannot be taken as argument by the signs *me* and *lo* because they require of the argument sign to have the values 'cltc3mrk' and 'un-hdd', respectively; in fact, none of these values unifies with 'cltc2'.

In all the derivations presented so far, the sign resulting from the combination of a clitic with a verb (whether the verb contains a clitic or not) is specified with the inherent value of the functor clitic. See, for example, *s[cltc3mrk]* in the last line of the derivation in Figure 8, where the value 'cltc3mrk' is contributed by the clitic *lo*.

A type like (16) for clitics, and the lexical entries (17) and (19), along with the rule of Functional Composition, allow for the derivation of signs like the following:

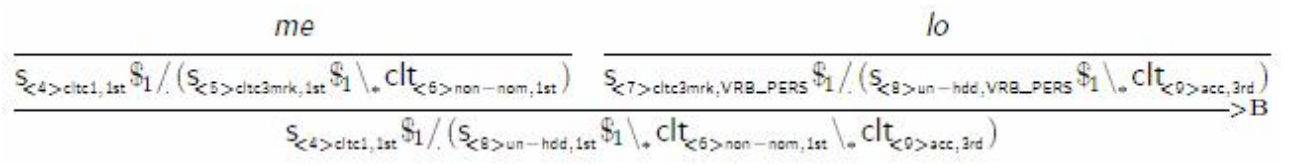

Figure 16

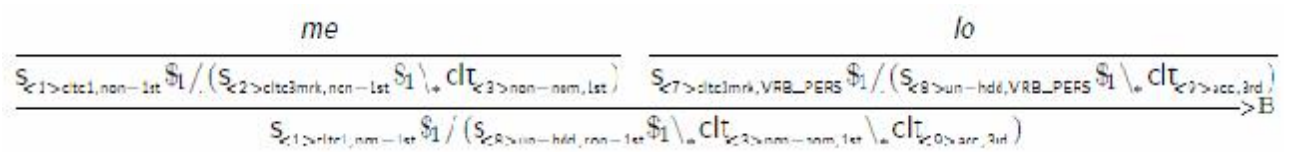

Figure 17

Because of the intervention of functional composition, the sign resulting from the combination of a clitic with another clitic is specified with values associated with the attribute 'lft_hd' in two "places": the first and second *s* of category (16). The value specified in the first *s* is contributed by the functor clitic. The value specified in the second *s* is contributed by the argument clitic. This is how the idea is captured that a clitic *package* "inherits" the order constraints of the argument clitic, and, at the same time, is specified with the inherent properties of the functor clitic, so that requirements of other functor clitics can subsequently be met. Thus, for example, the sign for *me lo* in Figures 16-17 is specified in the second *s* with the constraint that it is a functor sign that can forward combine with an argument sign that is specified with the value 'un-hdd' for the attribute 'lft_hd' (as if it were simply *lo*). Furthermore, the sign *me lo* is specified with the value 'cltc1' for the attribute 'lft_hd' in the first *s* (as if it were simply *me*), whereby it is defined

how it can eventually be taken as argument by other (clitic) functors: *se* and *te*, for all practical purposes:

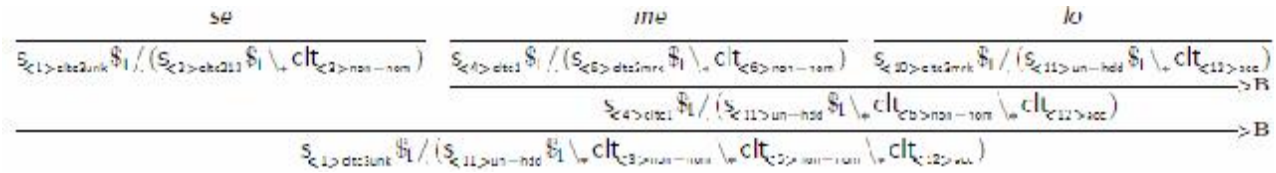

Figure 18

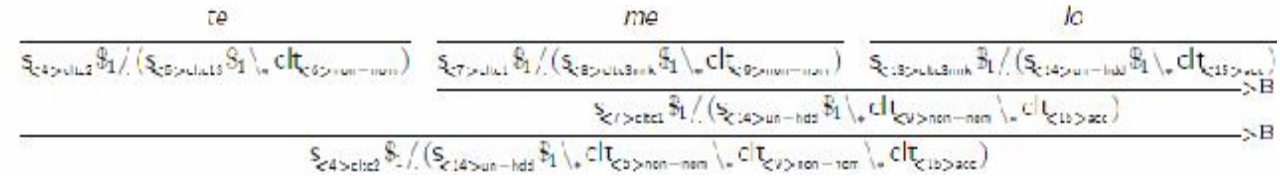

Figure 19

In Figure 18, the clitic *se* requires its argument to be 'cltc213', which is indeed satisfied because the sign *me lo* is specified as 'cltc1'. According to the hierarchy in the column CLITIC HEAD of Table A1.1, the value 'cltc1' is a daughter of 'cltc213' and, then, the unification between these two features is possible. Similarly, in Figure 19, the clitic *te* requires its argument to be 'cltc13', which again is satisfied as long as the sign *me lo*, as stated above, is specified as 'cltc1', making the unification with 'cltc13' possible, for 'cltc1' is a daughter of 'cltc13' in the hierarchy of values in the column CLITIC HEAD of Table A1.1.

Similarly, the clitic packages in (28) can be derived, while, at the same time, derivations like (29) are correctly excluded:

(28a) [te → lo] (28b) [se → [te → lo]] (28c) [se → [te → lo]]

(28d) [se → [te → [me → lo]]] (29a) lo →× te (29b) te →× [se → lo]

(29c) me →× [te → lo] (29d) te →× [se → [me → lo]]

In general, the mechanism proposed here accepts all the clitic packages that are well-formed in accordance with schema (1), namely, those in (5)-(7), and it excludes all the ill-formed sequences that violate the schema, some of which are illustred in (8)-(10). To put it in a nutshell, the micro-grammar accounts for the order constraints among clitics captured by Perlmutter (1971)'s schema (1).

3.4.1.2 Order between Clitics and the Governing Verb

According to the schemas (2) and (3), clitics occur on the left side of personal verb forms, except for imperatives, and on the right side of infinitives, gerunds and imperatives. See the grammaticalness contrasts illustrated by the examples in (11)-(14). We now define the mechanism responsible for this distribution of clitics in relation to their governing verb.

The relevant feature is 'non-imp-e' in the second *s* of (16), as illustrated by the entries for clitics (17)-(21). This possible value of attribute 'vrb_end' is interpreted in the context of the

hierarchy of the column VERB ENDING of Table A1.1. According to this hierarchy, the feature 'non-imp-e' ranges over all personal verb forms, except the imperative; i.e. any verbal sign with a feature capable of unifying with 'non-imp-e', namely, any value dominated by 'non-imp-e', can be taken by a clitic from the left.[12] For example, the derivations in Figures 20-21, as well as the (simplified) derivations in (30), but not those in (31), are accepted by the micro-grammar:

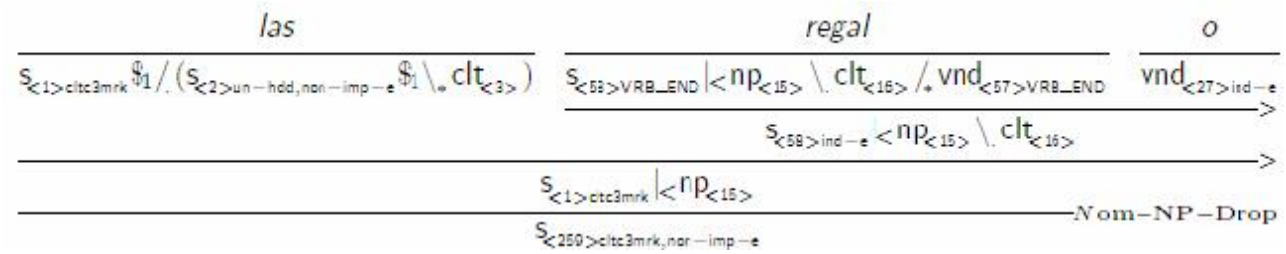

Figure 20

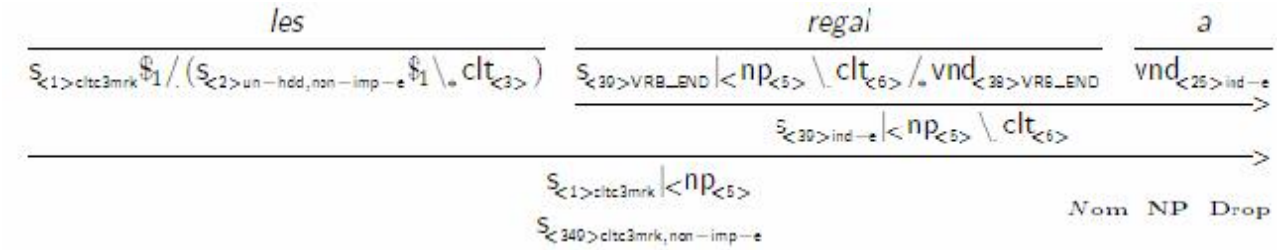

Figure 21

(30a) … [se → [las → [regal → e]]]	(30b) … [me → [los → [regal → ó]]]	(30c) … [te → [lo → [regal → aría]]]
(31a) las →× [regal → á]	(31b) les →× [regal → ar]	(31c) me →× la →× [regal → á]
(31d) te →× los →× [regal → ar]	(31e) se →× [regal → ando]	(31f) te →× la →× [regal → ando]

The derivations in (31) are not accepted by the micro-grammar because in all the cases the constraint 'non-imp-e' imposed by the functor sign on the argument sign is violated. The signs for *regal á*, *regal ar*, and *regal ando* are the following:

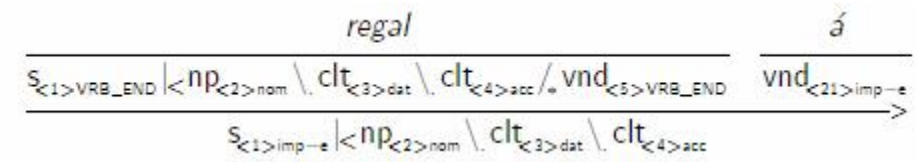

Figure 22

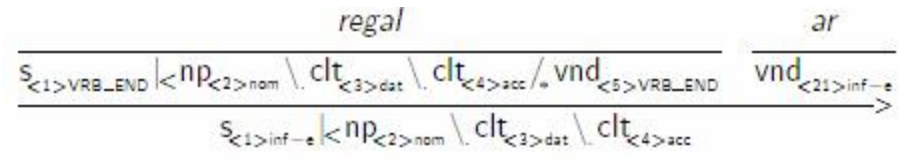

Figure 23

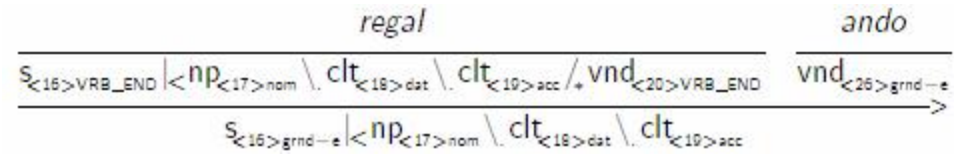

Figure 24

When the functor sign “tries” to combine with these argument signs, feature unification fails, for neither ‘imp-e’ nor ‘inf-e’ nor ‘grnd-e’ of the signs in Figures 22-24, respectively, unify with the value ‘non-imp-e’ required by the clitic. Cf. the hierarchy of values of column VERB ENDING of Table A1.1.

3.4.2 Clitics as Arguments

How do we now account for the distribution captured by schema (3), i.e. the occurrence of clitics on the right side of infinitival, gerundive, and imperative verb forms? The micro-grammar also defines infinitive, gerund and imperative endings as in Figures 25-27: (For a more complete specification of *ar*, see Figure A14.)

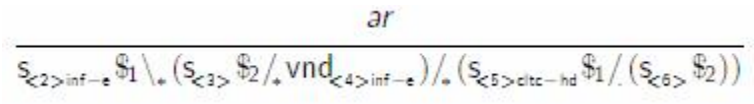

Figure 25

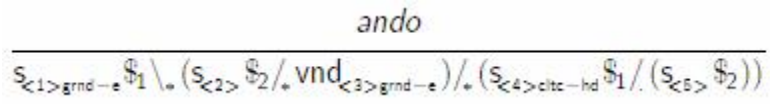

Figure 26

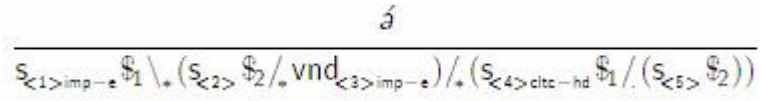

Figure 27

According to this categorial assignment, infinitive, gerund and imperative endings are also functors seeking forward for clitics (or clitic packages) to deliver functors that seek backward for verbal roots to deliver, finally, infinitival, gerundive and imperative verb forms.[13] Thus, the consumption of the outermost argument delivers the signs in Figures 28-29 and (32), but not (33):

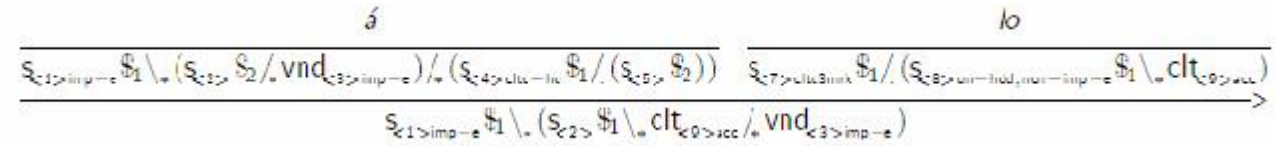

Figure 28

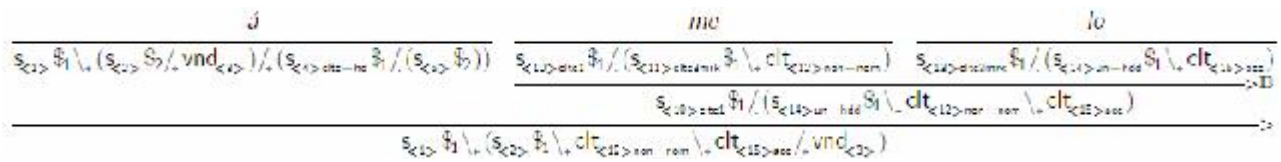

Figure 29

(32a) [ar → me]	(32b) [ar → [me → las]]	(32c) [á → [te → les]]	(32d) [ando → [te → [me → los]]]
(33a) ar →× las →× me	(33b) á →× les →× te	(33c) ando →× me →× [te → los]	

In (32b-d), clitic packages are derived first and then they are taken by infinitive, gerund and imperative endings. There are no derivations like (33) because the micro-grammar

does not accept derivations for the sequences *las te*, *les te*, and *me te los*, for these violate the order constraints among clitics captured by schema (1).

The signs in Figures 28-29 seek backwards for verbal roots with a specific argument configuration, and deliver imperative verbal forms with the appropriate argument consumption. Thus, for example, the sign *á lo* in Figure 28 seeks backwards for a verbal root to consume its accusative valence and deliver the sign *regal á lo* in Figure 30.

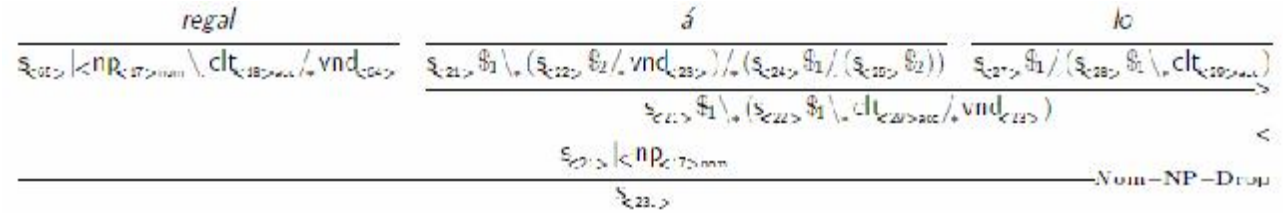

Figure 30

The sign *á me lo* in Figure 29 seeks backwards for verbal roots to deliver verbal signs like *regal á me lo*, where two of its clitic valences have been consumed, namely, the dative and accusative valences:

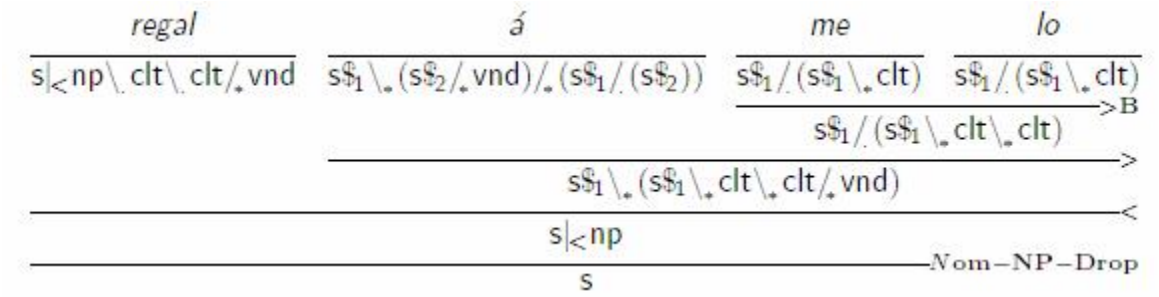

Figure 31

Similarly, the infinitive ending allows for the derivation of signs like the following:

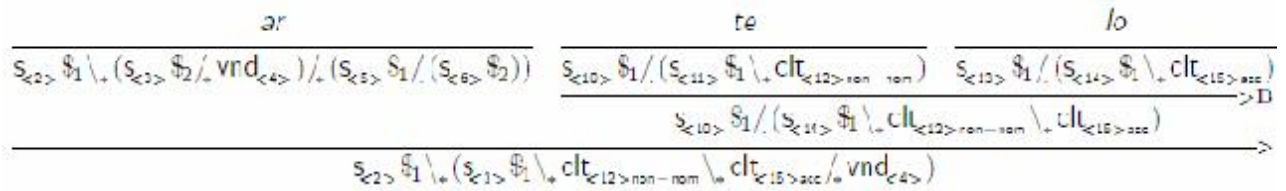

Figure 32

By combining *ar te lo* with a verbal root like *regal*, the sign *regal ar te lo* is obtained with a derivation like the following:

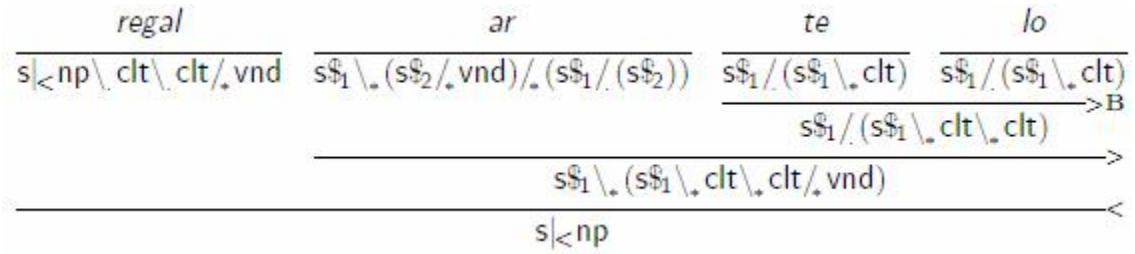

Figure 33

Again, the functor sign *ar te lo* consumes the dative and accusative clitic valences of its argument sign *regal*, but in this case the nominative valence is left available for subsequent consumption by a control verb. The reason is that there is no intervention of *Nom-NP-Drop*. In fact, this rule is sensitive only to signs that are specified as 'prsnl'. Note that the sign *regal ar te lo* is 'inf-e' and, therefore, 'non-prsnl'; however, the signs *regal á lo* and *regal á me lo* are 'imp-e' and, therefore, 'prsnl'. Cf. the hierarchy of values of the column VERB ENDING in Table A1.1.

To sum up, clitics occur on the left side of non-imperative personal verb forms (cf. schema (2)) because they are functors seeking forward for signs specified as 'non-imp-e' (cf. hierarchy of values of column VERB ENDING in Table A1.1). However, clitics occur on the right side of infinitives, gerunds and imperatives (cf. schema (3)) because they can be taken by the infinitive, gerund and imperative endings so that, later, the resulting signs can combine backwards with verbal roots.[14] One positive consequence of this treatment is that there is a single set of constraints controlling the order among clitics (captured by schema (1)), independently of the relative order between clitics and the governing verb (captured by schemas (2) and (3)).[15]

4 Conclusions

The paper has shown a few essential aspects of the design of a micro-grammar of River Plate Spanish clitics within the framework of OpenCCG (Bozşahin *et al* 2007). The resulting specification accounts for the order constraints among clitics and between clitics and their governing verbs.

The distinctive contribution of the paper is the conception of clitics as functor signs that have the capacity to seek forward either (i) for verbal signs to deliver signs with the relevant argument consumption, or (ii) for other clitics with which to construct clitic packages that can later be sought for by the infinitive, gerund or imperative endings to deliver verbal signs so that the relevant argument consumption is carried out, except that now more than one argument can be involved at the same time. This approach captures adequately and elegantly all the order constraints (1)-(3).

The proposed mechanism consists of a single lexical specification for clitics, which is not only independent of the meaning of the clitics but also of the position they occupy in relation to the governing verb (on the left or right side). This solution is better than other categorial treatments like, for example, Baschung *et al* (1987) for French, which declare different order constraints (in fact, mirror image constraints), depending on whether the clitics occur on the left or right side of the governing verbs.

References

Baschung, Karine, G.G. Bès, A. Corluy and T. Guillotin, 1987. "Auxiliaries and Clitics in French UCG Grammar", *The Third Conference of the European Chapter of the Association for Computational Linguistics*, Copenhagen.

Bastida, Salvador, 1974. "Restricciones de orden en las secuencias de clíticos del castellano: dos requisitos", *Revista de Filología Española* (sn): 58-69.

Bozşahin, Cem, G-J. Kruijff and M. White, 2007. *Specifying Grammars for OpenCCG: A Rough Guide*. Available from http://groups.inf.ed.ac.uk/ccg/index.html.

Castel, Víctor M., 1994. *Paquets de clitiques de l'espagnol. Calcul de la corrélation entre rôle sémantiques et proprietés morphologiques, et spécification dans une grammaire catégorielle d'unification*. PhD thesis, Université Blaise Pascal, Clermont-Ferrand II, France. Available from http://www.cricyt.edu.ar/institutos/incihusa/ul/WebHelpPhDThesis/Victor_M_Castel_PhD_Thesis.htm.

_____, 1996. "Paquetes de clíticos del español: restricciones de orden y expresión de roles semánticos desde una perspectiva categorial", *Signo & Seña* 5, 241-278.

_____, 1997. "Gramática categorial de unificación de los clíticos del español", *Signo & Seña* 7, 359-418.

_____, 2007a. "An *OpenCCG* Implementation of Perlmutter's Generalization on Spanish Clitics Word Order", *Infosur*, 3-14. Available from http://www.infosurrevista.com.ar.

_____, 2007b. "Hacia una gramática categorial combinatoria de los clíticos del español rioplatense", ponencia presentada en el *VIII Congreso Argentino de Hispanistas*.

_____, En prensa. "Una micro-gramática *OpenCCG* de los clíticos del español rioplatense", *Filología* ¿?.

García, Erica, 1975. *The Role of Theory in Linguistic Analysis: the Spanish Pronoun System*, Amsterdam, North-Holland Publishing Company.

Grimshaw, Jane, 2001. "Optimal Clitic Positions and the Lexicon in Romance Clitic Systems", en Geraldine Legendre, J. Grimshaw, and S. Vikner (eds.). *Optimality-Theoretic Syntax*. Cambridge: MIT Press.

Haddock, N., E. Klein and G. Morril, eds., 1987. "Categorial Grammar, Unification Grammar, and Parsing", *Working Papers in Cognitive Science*, Vol. 1, Centre for Cognitive Science, University of Edimburgh.

Jaeggli, Osvaldo, 1982. *Topics in Romance Syntax*, Dordrecht, Foris.

Luján, Marta, 1979. *Clitic Promotion and Mood in Spanish Verbal Complements*, Bloominton, Indiana University Linguistic Club.

Nuessel, Frank, Ed., 1980. *Contemporary Studies in Romance Languages*. Proceedings of the Eighth Annual Symposium on Romance Languages, Indiana, Indiana University Linguistic Club.

Perlmutter, David, 1971. *Deep and Surface Structure Constraints in Syntax*, New York, Holt, Rinehart and Winston.

Steedman, Mark, 2000. *The Syntactic Process*, Massachussets, MIT Press.

Suñer, Margarita, 1980. "Clitic Promotion in Spanish Revisited". In Frank Nuessel (1980: 300-330).

Zeevat, Henk, E. Klein and J. Calder, 1987. "Unification Categorial Grammar". In Haddock *et al*., 1987.

Appendix

SORTS	*SEMANTIC PERSON*	*CASE*	*MORPHOL. PERSON*	*CLITIC HEAD*	*VERB ENDING*	*MORPHOL. MOOD*
sem-obj phys-obj animate-being person performer addressee male female thing situation change action state	sem-pers-vals participant interactant performer sg-prfmr pl-prfmr addresse sg-addrss pl-addrss outsider sg-outsdr pl-outsdr	case-vals nom non-nom acc dat	pers-vals 1st 2nd 3rd non-1st 2nd 3rd non-2nd 1st 3rd non-3rd 1st 2nd	lft-hd-vls hdd cltc-hd cltc13 cltc1 cltc3mrk cltc213 cltc1 cltc2 cltc3mrk cltc3unk non-cltc-hd un-hdd	vrb-end-vals vrb-end prsnl imp-e non-imp-e ind-e sbjn-e non-prsnl non-prtp-e imp-e inf-e grnd-e prtp-e	m-mood-vals ind sbjn imp inf grnd prtp

Table A1.1. Hierarchies of feature values.

Values of the attribute *lft_hd*	Values of the attribute *vrb_end*	Values of the attribute *pers*
<type name="lft-hd-vls"/> <type name="hdd" parents="lft-hd-vls"/> <type name="cltc-hd" parents="hdd"/> <type name="cltc13" parents="cltc-hd"/> <type name="cltc213" parents="cltc-hd"/> <type name="cltc2" parents="cltc213"/> <type name="cltc1" parents="cltc13 cltc213"/> <type name="cltc3mrk" parents="cltc13 cltc213"/> <type name="cltc3unk" parents="cltc-hd"/> <type name="non-cltc-hd" parents="hdd"/> <type name="un-hdd" parents="lft-hd-vls"/>	<type name="vrb-end" parents="vrb-end-vals"/> <type name="prsnl" parents="vrb-end"/> <type name="imp-e" parents="prsnl non-prtp-e"/> <type name="non-imp-e" parents="prsnl"/> <type name="ind-e" parents="non-imp-e"/> <type name="sbjn-e" parents="non-imp-e"/> <type name="non-prsnl" parents="vrb-end"/> <type name="non-prtp-e" parents="non-prsnl"/> <type name="inf-e" parents="non-prtp-e"/> <type name="grnd-e" parents="non-prtp-e"/> <type name="prtp-e" parents="non-prsnl"/>	<type name="pers-vals"/> <type name="1st" parents="pers-vals non-2nd non-3rd"/> <type name="2nd" parents="pers-vals non-1st non-3rd"/> <type name="3rd" parents="pers-vals non-1st non-2nd"/> <type name="non-1st" parents="pers-vals"/> <type name="non-2nd" parents="pers-vals"/> <type name="non-3rd" parents="pers-vals"/>

Table A1.2. Hierarchies of feature values in the file *SpnshTps.xml*.

Notes

[*]Aspects of this paper were presented at the *VIII Congreso Argentino de Hispanistas* (Castel 2007b), and the *IV Encuentro de Gramática Generativa*, held in May and July 2007, respectively, in Mendoza, Argentina. The discussion in §§3.4.1.1-2 is also addressed in Castel (2007a), though the treatments differ radically in various respects; in particular, no provision is made in Castel (2007a) for the occurrence of clitics on the right side of imperatives, infinitives and gerunds. Moreover, Castel (en prensa) is an OpenCCG micro-grammar of River Plate Spanish clitics that integrates the treatment of order constraints with semantic role assignments

[1]OpenCCG (Bozşahin *et al* 2007) "is an open source natural language processing library written in Java, which provides parsing and realization services based on Mark Steedman's Combinatory Categorial Grammar formalism" (Steedman 2000).

[2]To facilitate the presentation, we take advantage of TeXnicCenter, which is an integrated development environment for developing LaTex documents on Microsoft Windows. The resulting visualization greatly simplifies the discussion of abstract and complex structures.

[3]Note that there are various phenomena that are left out of the scope of the current proposal, some of which are the following: (a) the distribution of semantic roles; (b) clitic doubling; (c) the co-occurrence of clitics with full noun phrases in non-reflexive configurations; and (d) clitic climbing in infinitival and gerundive constructions. See Castel (en prensa), for an account of the distribution of semantic roles and a treatment of clitic climbing with Equi-Subj-NP-Deletion verbs.

[4]To test the micro-grammar, the reader needs first to download the OpenCCG environment from https://sourceforge.net/projects/openccg/.

[5]See Castel (en prensa), for a micro-grammar that accounts also for the correlations between clitics and semantic roles.

[6]Except for minor details, the discussion of which we leave for another place, this rule is quite similar to the one proposed in Bozşahin *et al* (2007: 25).

[7]Similarly, clitics occurring on the right side of imperatives, infinitives and gerunds are separated, so that instead of, say, *regalármelas* and *regalales*, the micro-grammar outputs *regal ar me las* and *regal á les*.

[8]Figure numbers preceded by *A* refer to figures in the Appendix to Castel (en prensa). If the reader does not have this paper, (s)he can run the micro-grammar and thus check all sorts of details of complete derivations.

[9]Henceforth, some simplified representations in the text will have a reference to more complete representations in figures of the Appendix to Castel (en prensa). The reader is reminded that (s)he can always run the micro-grammar to output complete derivations.

[10]The expression 'lft_hd' abbreviates *left head*.

[11]Note also that, in Figures 6 and 7, the clitic *me* consumes the valences *clt[dat]* and *clt[acc]*, respectively.

[12]Of course, as long as the order constraints among clitics are also met, which are based on the values of the hierarchy CLITIC HEAD of Table A1.1.

[13]The micro-grammar also assigns these endings to a category that is an argument that can be taken by (functor) verbs. Cf. Figure 22-24 above, and Figures A31-32.

[14]Note that the requirement 'non-imp-e' imposed by clitics on their arguments is "neutralized" when clitics are taken by the imperative, infinitive and gerund endings. In fact, these endings deliver functor signs that seek backwards for verbal roots.

[15]The treatment of French clitics by Baschung *et al* (1987), for example, assumes two different sets of order constraints among clitics depending on whether they occur on the left or right side of verbs.