

“How many times do I have to tell you?” Clitic tripling in Argentinian Spanish

1 Introduction

1.1 Apparent violations of valency

Every syntactic theory has some mechanism to ensure that required arguments of a predicate appear only once. In English, for example, you cannot utter (1a) if you intend both *John* and *an elephant* to be the object of *saw*, or (1b) if you intend both Peter and Mary to be the subject of ‘being sad’.

1. **I saw John an elephant.*
2. **Peter Mary is sad/are sad.*

In transformational generative grammars in the Chomskyan tradition (e.g., Government and Binding, Minimalism), the valency requirements of a verb are enforced by the combined action of the theta-criterion (every argument needs a thematic role and every thematic role subcategorized for by the verb must be assigned) and the Case filter (every NP (roughly) needs Case). But it has been known since at least the mid-seventies ***Kayne 1975*** that data from “clitic doubling” (3) represent a challenge to syntactic theory precisely because such structures appear to violate the verb’s valency requirements.¹

3. *Yo lo quiero a mi país.* ***Ligatto 1996***
I 3SgM_{CL} I.love A my country
‘I do love my country’

In (3) there is a clitic and a NP (the “double”) that both instantiate the direct object (DO) of the verb. The problem from a GB/Minimalist perspective is determining how case assignment works, i.e., which of the constituents, the clitic or the double, receives case from the verb, and how the other constituent receives case.

The most commonly accepted transformational solution to the CLD puzzle is the “Big DP” analysis ***refs***. Following the DP hypothesis (Abney***), the clitic is considered a determiner which heads a DP. A lexical DP, the “double”, is generated in Spec,DP, and a silent “pro” is generated as a complement to D. When the clitic moves to V (or T), the lexical DP is left “stranded” in the postverbal, “in situ”, position (see Figure 1).

¹ In all the examples in this paper, underlining signals the direct object clitic and its associated constituent. DO clitics are glossed by their agreement features (1, 2, 3 for person, Sg/Pl for singular/plural, M/F for masculine/feminine), accompanied by a “cl” subscript. When necessary, agreement features will appear on other constituents’ glosses, and small capitals will mark focal stress. The gloss “A” will be used for the animacy marker *a* which is obligatory before animate direct objects and is homophonous with the indirect object marker *a* and the preposition *a*. Not naturally occurring (constructed) examples are preceded by the interrobang “‡”. Attested examples are given with their source.

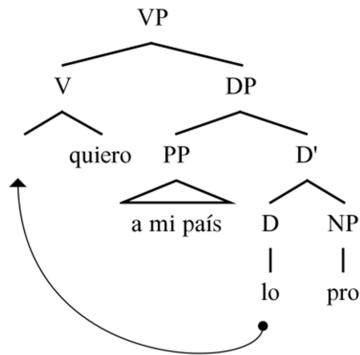


Figure 1: Clitic doubling as a result of clitic movement out of a big DP.

This way, a θ -role can be assigned to the whole Big DP before movement, and Case can be assigned to the clitic in some adequate position, with the lexical part assigned Case by virtue of its relationship to the clitic ***Belletti 2005***. Moreover, this ensures all constituents agree on the relevant agreement features (presumably as an instance of Spec-head agreement ***Torrego 1992***). However, ***Anagnostopoulou 2003*** points out two main problems with the Big DP account.

The first is that the Big DP account depends crucially on the clitics being determiners. But in several languages that have CLD there is no evidence to treat clitics as D heads. For instance, determiners in Greek (or in Spanish) do not cliticize out of their DPs, which is the evidence Uriagereka (***1995***) adduces to treat clitics as determiners in Galician.

- 4. a. *Vimo-lo (a el).*
saw.we-him to him
- b. *Vimo-lo neno.*
saw. we-the child

The second problem is that there is no direct empirical evidence of structures such as Big DPs. As far as I am aware, they are not used in any other constructions but clitic doubling ones (making them *ad hoc* as an explanatory mechanism), and they never surface as such in any language.

To this list, we can add another phenomenon, which I call *Clitic Tripling (CLT)*, which is not easily accounted for via the Big DP hypothesis. CLT is introduced in the next section.

1.2 Clitic tripling constructions.

Comment [BE1]: See also Uriagereka's fn. 6. His account depends crucially on Italian and French not having CLD, and having ne/en cliticization. But of course, Italian does have CLD...

Beyond CLD constructions, there is another class of doubling phenomena that hasn't received attention in the literature, exemplified below:

- (5) (Anuario Barcelona 2003, p. 34)
A esos los voy a extraditar *A TODOS*
 A those 3PIM_{CL} I.go to extradite A all.PL.M
 'I am going to extradite them all'
- (6) (Personal communication, May 10th 2005)
Sí, las fotos las tengo que SUBIR *las nuevas*
 Yes the pictures 3PIF_{CL} I.have that upload the new.PL.F
 'Yes, I have to upload the new pictures'

In these examples, there are three different constituents, all coreferential, that instantiate the direct object of the verbs: a preverbal, left-dislocated constituent; a DOCL; and a postverbal constituent. The construction is attested in Rioplatense Spanish, although it is usually assumed to be impossible. For example, Zagona (2002, p. 222) states that “[t]he CLLD coreferential phrase cannot be overt; it must be an empty category, although this category may require a clitic.” Similarly, Belletti (2005, p. 5) states

An important property which is shared by all cases mentioned above [CLD, CLLD, floating quantifiers, right dislocations, and strong pronoun doubling] is the fact that the two parts in which the original constituent split are such that one contains a lexical noun phrase the other a functional word, either a clitic or a quantifier. Th-theoretic reasons exclude the possibility that the two parts be both constituted by lexical noun phrases.

Some restricted cases of the CLT construction were analyzed by Suñer (2006) as CLLD plus anaphoric epithets.

- (7) (Suñer 2006:129)
A Teresa, la vieron en Rosario a esa borracha empedernida.
 A Teresa 3SgF_{CL} they.saw in Rosario A that.Sg.F drunk.Sg.F inveterate.Sg.F
 'Teresa, they saw that inveterate drunkard in Rosario'

Suñer uses these constructions to argue explicitly against a Big DP account, because CLLD constructions and CLD constructions have different interpretational properties: CLLD (but not CLD) allows non-specific and non-partitive constituents, specific indefinites in CLLD (but not in CLD) may get a generic interpretation, CLLD (but not CLD) can happen with CPs and AdjPs. If both originated within the same DP, one would expect them to allow the same semantic

interpretation and constraints, according to Suñer. López (2009) maintains that a big DP account is viable, if the anaphoric epithet is generated in Spec,D as part of a small clause kind of constituent attached to the constituent that will end up left-dislocated. So, a sentence like

En el trabajo, a su hermano Mara me dijo que no lo aguantan a ese tarugo.

Comment [BE2]: Interestingly, "At work" here cannot modify "told". How is that scope achieved?

In the work A his brother Mara me told that not 3SgM they stand A that screw.anchor
'Mara told me that they can't stand his brother that idiot at work.'

would be, in the "small-clause-in-BigDP" analysis

En el trabajo, Mara me dijo que no lo aguantan [SC ese tarugo su hermano].

In the work Mara me told that not 3SgM they stand that screw.anchor his brother

But as López himself recognizes, some "tweaking" is needed to derive the correct surface form at Spell Out in different cases: insertion of "a", insertion of other morphological forms, changes from a demonstrative to an indefinite article, or word order changes. Moreover, despite López' claims, the small clause above is quite different from a classic small clause like you would find with *consider*:

Considero a tu hermano un tarugo.

I consider A your brother a screw.anchor

'I consider your brother an idiot.'

Such small clauses only admit indefinite NPs, whereas CLT (like CLD) is mostly found with definite NPs. Finally, this small clause analysis does not address the interpretive differences between CLLD and CLD. Therefore, this proposal is a new ad hoc, empirically unmotivated, structural adaptation in order to maintain another ad hoc analysis, the Big DP.

Still, a single mechanism accounting for both CLD and CLT is desirable. This paper will provide a common analysis of the syntactic and informational structure of CLT and CLD within the framework of Lexical-Functional Grammar (LFG). Because LFG is non-configurational, it can license doubling and tripling structures without ad hoc machinery; because it is multistratal, it can also model discourse phenomena that interface with syntax and explain CLD and CLT usage.

2. Lexical Functional Grammar Basics

3. Some basics of Spanish syntax in LFG

3.1 Spanish c-structure and functional annotations

We adopt a rather simple structure for Spanish. The sentence is the non-projective, exocentric category S (Bresnan 2001:112). The other categories, at the phrasal level, are endocentric. The following are phrase structure rules relevant to this paper:

1. $S \rightarrow NP VP$
2. $NP \rightarrow (Det)^* (Adj) N'$
3. $N' \rightarrow N (Adj)^* (PP)$
4. $VP \rightarrow V' (NP) (PP)$
5. $V \rightarrow CL V$

Comment [BE3]: Check Payne & Huddleston 2002 for arguments against DP.

Spanish has differential object marking ***Aissen 2003***. All human direct objects (and often animate non-humans) are marked with the preposition *a*, but inanimates are not (generally). I analyze here this “a-personal” as a kind of NP (although this issue is by no means settled in the literature). Likewise, the structure of Spanish NPs is much more complex than what is given here (see Zagona 2002, ch. 2). Here, I use Det as a label encompassing classic determiners, but also quantifiers, cardinals, and demonstratives. Moreover, a distinction is needed between argument and predicative NPs, and between subject and object NPs, but this is probably not at the level of PS rules and is irrelevant for this analysis.

It is important to remember LFG’s principle of economy of expression (Bresnan 2001:91: To be present in c-structure, any syntactic (i. e., non-terminal) phrase node must be required by completeness, coherence, or semantic expressivity. For example, this restricts the use of non-branching nodes required only to provide a full X’ structure in certain flavors of generative syntax. Therefore, the c-structures presented in what follows are really simple. Most of the “grammatical work” is done by lexical entries and by principles of grammatical function assignment, to which we turn next.

3.2 Grammatical function assignment in Spanish

Spanish has relatively free word order. Unlike for a strict configurational language like English, grammatical functions (GFs) are not assigned on the basis of structural positions. Such languages are called *non-configurational*. Generally, non-configurational languages employ morphological methods of GF assignment (“morphology competes with syntax”, ***Bresnan 2001***).

However, Spanish is a language with a relatively impoverished case system. Nominative, accusative, dative, genitive, prepositional, and comitative case are only marked on pronouns. Noun phrases can either be unmarked, or marked with “a-personal”, if they are animate and in an

objective case (accusative/dative). In the example below, “el perro” is the subject and is unmarked, whereas “Juan” is the direct object and is marked with “a”.

(8) *El perro mordió a Juan.*

The dog bit A Juan

‘The dog bit Juan.’

This impoverished morphology would predict a stronger reliance on configurationality, yet Spanish has rather free word order:

(9) *El escritorio tapa la puerta.*

The desk covers the door

SUBJ V OBJ

(10) *Las tierras del señor humedecían su sudor y su llanto día tras día* (Joan Manuel Serrat, “Manuel”, 1969)

The lands of the master wet his sweat and his crying day after day

OBJ V SUBJ

‘His sweat and tears wet the lands of the master day after day.’

In the first example above, “the switch” can be the subject and “the door” the object, or, with a different intonation, the grammatical functions can be reversed. The second example (from a Spanish song) shows that either underlined NP can be the subject or the object of the verb *wet*. (In fact, there is a kind of garden path here, in that when one is hearing the song the first NP is interpreted as subject, until the second NP is encountered.) However, the only function assignment that makes sense semantically is for the preverbal NP to be the object, and the postverbal, conjoined NP to be the subject. Since 1st/2nd person pronouns are case marked and the verb is inflected for subject person and number as well, problems with GF assignment occur mainly when both arguments of a transitive verb are 3rd person and have identical values for number. In that case, in direct object clitics can help to disambiguate, as long as the subject and object differ in gender:

(x) *El escritorio la tapa la puerta.*

‘The desk covers the door.’

(x) *El escritorio lo tapa la puerta.*

‘The desk is covered by the door.’

Therefore, GF assignment in Spanish is a rather constructive process where GF assignment is enforced by coherence and completeness, and several cues are taken into account: subject person and number inflection on the verb, object clitic agreement features, semantics, and prosody. Accordingly, we will annotate the nodes in PS rules with generic GF information, to be solved by the solution algorithm in parsing.

S →	NP	VP
	(↑GF)=↓	↑=↓
VP → V'	(NP)	(PP)
	↑=↓	(↑GF)=↓

Adjectives are treated as introducing a member of a set of adjuncts to a nominal predicate, which is expressed by the annotation $\downarrow \epsilon$ (↑ADJ). Determiners are annotated with a DET function, (↑DET)=↓.

NP → (Det)*	(Adj)	N'
	↓ ϵ (↑ADJ)	↑=↓
N' → N	(Adj)*	(PP)
	↓ ϵ (↑ADJ)	(↑GF)=↓

4. CLD in LFG

4.1 The constituent structure of CLD

Now we can look at an example of clitic doubling: *Las vi las fotos nuevas*. The c-structure generated from the phrase structure rules is as shown in figure X1 (omitting the internal structure of the NP):

Comment [BE4]: But perhaps it is important to show this here, since it becomes relevant for discussing CLT.

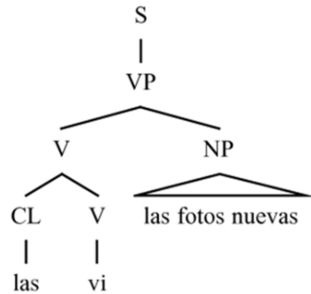
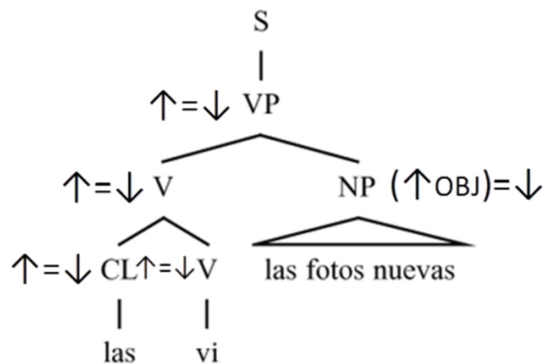


Figure X 1: C-structure of a clitic doubling sentence

In LFG, the NP *las fotos nuevas*, sister to V, is not automatically an object by virtue of its position. But what else could it be? The verb *vi* requires a SUBJ and an OBJ. Its morphology already specifies its subject as 1Sg. Since *las fotos nuevas* is 3PIF, it can only be assigned the OBJ function.



Now, the clitic *las*, being a DO clitic, is also assigned the OBJ function. This is an apparent violation of coherence. However, LFG provides a mechanism of structure sharing that allows us to unify the information of both clitic and NP into a single OBJ f-structure, as explained in the next section.

4.2 DOCLs lexical entry

The direct object clitics of Spanish are morphologically marked as accusative. Their lexical entry, therefore, specifies that they contribute information to the OBJ f-structure of the clause they are in. I use here inside-out functional application ****see, e.g., Nordlinger 1997***** and annotate the DOCLs lexical entries with (OBJ ↓). This means that the clitics lexical information (represented by the down-arrow) is the value of an OBJ attribute in the immediately containing f-

structure. The other information contributed lexically is that the object is a pronominal, and that its agreement features are third person, plural number, feminine gender.²

DOCL (3PlFem shown here)
las: CL(OBJ ↓)
 (↓PRED) 'PRO'
 (↓PERS)) 3
 (↓NUM) Pl
 (↓GEND) Fem

4.2.1 Optional PREDs. As discussed above, PREDs are the only f-structure attributes that cannot unify: they have to be singly instantiated in any given f-structure. Therefore, in a CLD sentence it is not possible for both the clitic and the lexical NP in CLD to contribute PRED values to the OBJ f-structure, or this would violate coherence. Andrews' (1990^{***}) solution was to posit that the clitic instantiates its PRED 'PRO' *optionally*, that is, only if required for completeness. (This optionality is indicated below by enclosing the PRED value in parentheses.) If a doubling lexical NP is present, then the clitic does not contribute a PRED.

DOCL (3PlFem shown here)
las: CL(OBJ ↓)
 ((↓PRED) 'PRO')
 (↓PERS) 3
 (↓NUM) Pl
 (↓GEND) Fem

4.2.2 Specificity and animacy effects. CLD is known to be subject to specificity and animacy effects. These restrictions lend support to the proposal that there are two DO clitic sets in Spanish: a doubling set and a non-doubling set. The specificity and animacy restrictions can be encoded in the lexical entry for the doubling clitic, but omitted in the lexical entry for the non-doubling clitic. For example, the specificity effect states that CLD only obtains with specific DOs (or at least indefinite partitive DOs, Estigarribia 2006a). Therefore, it seems natural to posit the following two lexical entries:

Non-doubling DOCL
las: CL(OBJ↓)
 (↓PRED) 'PRO'
 (↓PERS) 3
 (↓NUM) Pl
 (↓GEND) Fem

² In the lexical entries after this section, I will use the metaattribute *agr* to collapse person, number, and gender information.

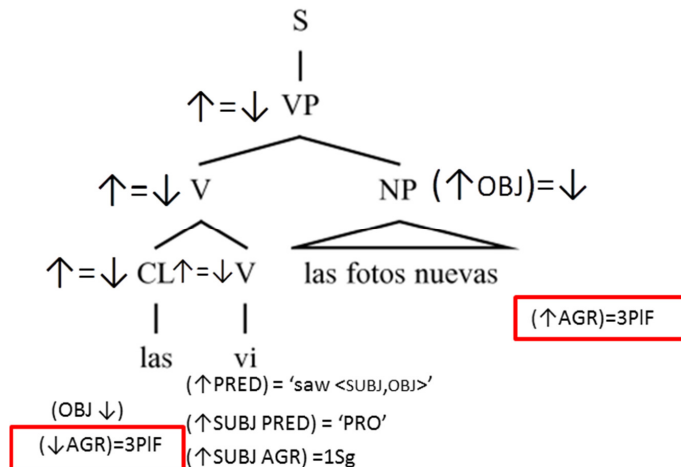
Doubling DOCL
 las: CL(OBJ↓)
 (↓PERS) 3
 (↓NUM) Pl
 (↓GEND) Fem
 ¬(↓SPEC) -

Note that the doubling clitic requires the OBJ not to be instantiated elsewhere as non-specific.

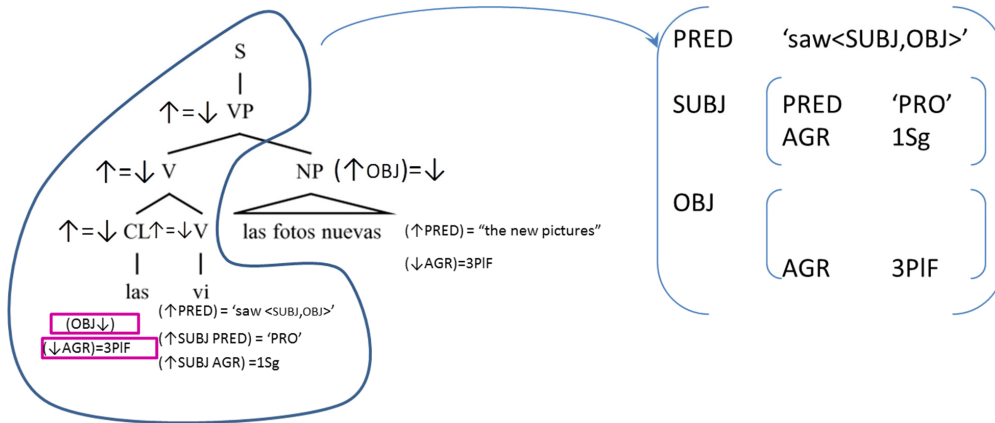
The animacy effect refers to the fact that, in Rioplatense Spanish, only animate DOs can be focused when doubled (Estigarribia 2006a). This could also be modeled by positing two sets of doubling clitics, one specified for discourse function as (TOP ↑), but unspecified for animacy, and a focusing clitic with a (FOC ↑) annotation and a (↓ANIM)=_c + annotation, the latter constraining the object to be specified as animate by other means. This would, however, lead to a proliferation of homophonous clitics. A perhaps more elegant solution is to let constraint competition derive the animacy effect, as done in Estigarribia (2006b) within the framework of OT-LFG.

Comment [BES]: This contradicts Keller's DOUBLEGROUND for Greek (doubled objects must be ground) ["Phonology Competes With Syntax"]

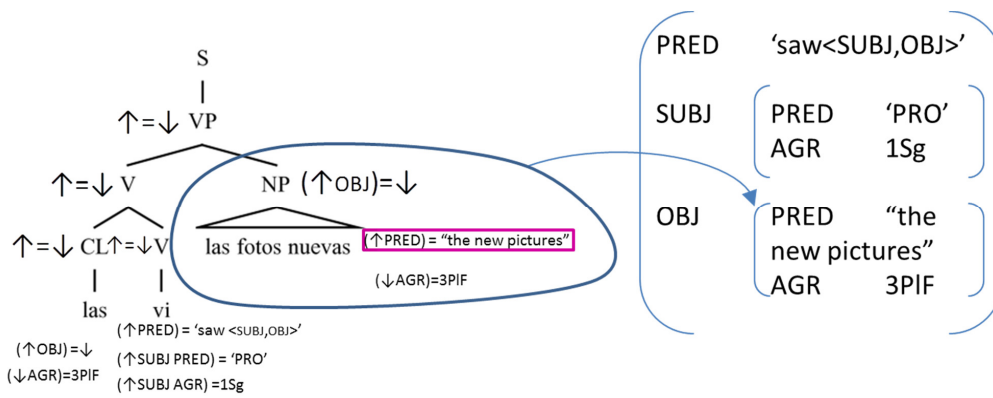
4.2.3 Structure-sharing by unification. The c-structure with functional annotations and lexical annotations is the following:



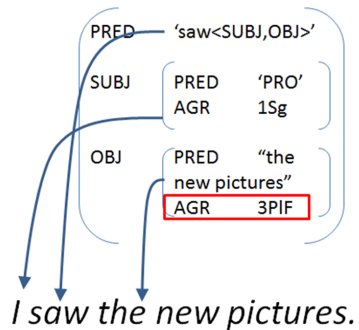
Note that the clitic and the double match in agreement features. This is a necessary condition for their functional information to be unified in a single OBJ f-structure. All of the c-structure nodes annotated ↑=↓ are mapped to the top f-structure, the functional structure of the whole sentence.



But the OBJ lacks a PRED and therefore does not satisfy completeness. Hence, the PRED value is contributed by the double.



The reading of the CLD sentence (in a way that is truth-conditionally equivalent) is therefore:



Note that this elegant structure-sharing solution also solves a binding puzzle. Cecchetto (1999) notes that [in-situ] CLD constructions should be a Principle C violation, but says that Principle C possibly does not apply because the CL and the NP share the same theta-role. This is straightforward in the LFG formalism, because binding conditions are stated in terms of f-structure functions, and both clitic and double belong to a single function OBJ. In a nutshell, the structure of CLD poses no particular problem for LFG, since this theory was designed to accommodate non-configurational phenomena.

However straightforward a structural account is, the important question of why CLD structures are used remains, and it is to that question I turn now.

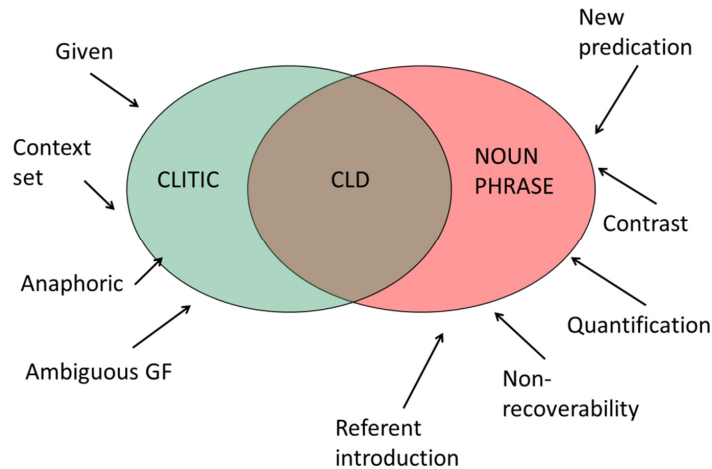
5. The Discourse forces model of CLD

Following Estigarribia (2006a,b), I propose that CLD results as the interaction of two partially independent sets of discourse conditions, or “discourse forces”.³ On the one hand, factors that favor expression of a referent as a clitic are the referent’s givenness at a particular point in discourse, its salience, its anaphoricity (in particular anaphoric constructions, e.g., CLLD), the possibility of ambiguous assignment of a GF to the referent (in which case the clitics, being morphologically marked for case, disambiguate that assignment), or the fact that the referent will be quantified (with the clitics providing a context set for quantification).

On the other hand, factors that favor expression of a referent as a lexical NP are non-recoverability, contrast, newness of the referent, or a shift in perspective or new predication over a given referent (see Estigarribia 2006a for extensive discussion and examples). Crucially, factors from these two sets can be operative at the same time. Under those conditions, CLD is used to satisfy both sets of discourse forces simultaneously:

Comment [BE6]: Look at Stirling 2008 on Double Reference in Kala Lagaw Ya.

³ This term was suggested to me by Peter Gordon.



For example, below is a CLD sentence used when recoverability is compromised because several referents can be picked up by a clitic:

(11) b. <Lig-115>

Es mucho más fácil montar un jardín que un hospital o un consultorio

Is much more easy mount a garden than a hospital or a consultancy

'It is much easier to start a kindergarten than a hospital or a private practice – you

el jardín lo podés hacer poner con nada casi

the garden 3SgM you.can make put with nothing almost

can set up a kindergarten with almost no investment'

In the example above, *jardín* 'kindergarten', *hospital* 'hospital', and *consultorio* 'private practice' are all given and salient. Furthermore, they are all 3SgM. So when the DO clitic is used to refer to one of these, agreement features cannot disambiguate. Hence, a lexical NP is used concurrently, giving rise to a doubling pattern.

Below is an example where new content is predicated of a given referent:

(12) (e-mail exchange)

Speaker A: *Queremos vender la camioneta.*

We want sell the van

‘We want to sell the van.’

Speaker B: *Sí, espero que la vendan esa chatarra de una vez.*

Yes I hope that 3SgF youPl.sell that junk once.and.for.all.

Here, *camioneta* ‘van’ is given, salient, and recoverable, so it can be referred back to with the DO clitic *la* alone. But the speaker (myself, incidentally) wanted to say something else about the van under discussion, namely that it was a piece of junk, hence the need for a “doubling” lexical NP.

5.1 An unexpected prediction

The discourse forces model is successful in accounting for a wide range of uses of CLD (Estigarribia 2006a). Of interest, the discourse forces model does not constrain how many of the factors in each set can be operative at any given point in discourse. Therefore, it predicts that there should be cases where more than two constituents instantiate a direct object. Such cases, where two lexical NPs are coreferential with a DOCL, do exist, and are the *Clitic Tripling (CLT)* cases that are introduced in this paper.

(13) *Las fotos las vi las nuevas.*

The pictures 3PIF I.saw the new-Pl

‘I saw the new pictures.’

(14) *A esos los voy a extraditar a todos.*

A those 3PIM I.go to extradite A all

‘Those people, I am gonna extradite them all.’

Suñer (2006) already recognized a restricted class of tripling constructions, CLLD + anaphoric epithets, but she explicitly excluded postverbal inanimate constituents, which (13) above shows to be empirically incorrect. Moreover, the construction is much more general than what Suñer posited. The postverbal constituent in fact need not be an anaphoric epithet. It can be a headless anaphoric NP functioning as an adjectival modifier (either a Det + Adj or a Det + Rel Clause sequence, 15–17), a Det + Possessive pronoun (18), a quantifier (19), a deictic NP (20), or a full headed NP with an anaphoric interpretation (21,22)

(15) *Las fotos las tengo que subir las nuevas.*

'I have to upload the new pictures.'

(16) *?A mi vecina la atropellaron a la vieja.*

'My neighbor was run over, the old woman.'

≠ 'My old neighbor was run over.'

(17) *?Las fotos las tengo que subir las que saqué ayer.*

'I have to upload the pictures that I took yesterday.'

(18) *?El auto lo tengo que lavar el mío.*

'I have to wash my car.'

(19) *A esos los voy a extraditar a todos.*

'Those people, I am going to extradite them all.'

(20) *?La moto la tengo que lavar esa porquería.*

'I have to wash the motorcycle, that piece of crap.'

'I have to wash that crappy motorcycle.'

(21) *?A mi vecina la atropellaron a la pobre mujer.*

'My neighbor was run over, the poor woman.'

'My poor neighbor was run over.'

(22) *?A los soldados los mataron a los que no se rindieron.*

'They killed the soldiers that did not surrender.'

That in this CLT construction the postverbal constituent is not right-dislocated is shown by two facts (see the right-dislocation tests in, e.g., [Cecchetto 1999?2000?](#)). First, prima facie no pause is needed before it. Second, it is possible to have certain adverbs which are commonly taken to be VP-internal follow it:

(23) *?Las fotos las tengo que subir las nuevas cuidadosamente.*

(24) *?A mi vecina la atropellaron a la pobre mujer descaradamente.*

Therefore, the postverbal constituent can be construed as a sister of V, and is not necessarily right-dislocated (although it can be as an instance of a different construction). This supports a

Comment [BE7]: Some of the glosses are problematic, to say the least. I have decided to go with "neutral" truth-conditionally equivalent versions for the early examples, until I can discuss the complexities. But clearly, in some cases it doesn't work.

Comment [BE8]: But see the tests in Alexiadou & Anagnostopoulou 2000 that don't seem to work.

Comment [BE9]: Are these VP-internal adverbs in Spanish?

view of CLT as CLLD+CLD. It is a case of CLLD because it shares the properties of that construction (as demonstrated by Suárez 2006 for CLLD+ep): the clitic is obligatory (25), the left-dislocated constituent enters in an unbounded dependency (26), and CLT can appear in embedded contexts (27).⁴

(25) **Las fotos* tengo que subir *las nuevas*.⁵

The pictures I.have that upload the new-PIF

(26) ?*Las fotos* Juan me dijo que *las* tengo que subir *las nuevas*.

The pictures Juan me told that 3PIF I.have that upload the new-PIF

(27) ?*Juan me dijo que las fotos las* tengo que subir *las nuevas*.

Juan me told that the pictures 3PIF I.have that upload the new-PIF

The question whether CLT is a genuine case of CLD is more complex. I will discuss later whether CLT also shows the specificity and animacy effects found in CLD.

6. The constituent structure of CLT.

In this section, I show that the c-structure of CLT is licensed by exactly the same mechanisms that license CLD.

6.1 Split NPs.

Interestingly, LFG analyses exist of a phenomenon similar to CLT: Split NPs.

(28) Jiwarli. (Dixon 1977, Austin 2001, Mushin & Baker 2008)

Karla *wantha-nma-rni* *jarnpa juma*.

Fire *give* *light* *small*

'Give me a small fire light.'

This construction is a focusing construction, where the most important information is given first, followed by less important information. But at the f-structure level, "karla jarnpa juma" is the direct object of the verb, even though it appears in different positions for information-structural reasons.

⁴ I am not using tests referring to island sensitivity or weak crossover here because it is clear to me that judgments on those constructions are more variable and less reliable than it is usually claimed in the literature.

⁵ This example is OK with a different prosody that marks the left-edge constituent as a hanging topic.

(33) ?*Juan las fotos las tiene que subir las nuevas.*

Juan the pictures 3PIF he.has that upload the new-PIF

‘Juan has to upload the new pictures.’

Another important difference is that the Chicheña discontinuous constituents surface with exactly the same morphology as their continuous counterparts. However, in Spanish CLT the two lexical nouns both must have determiners (as in the German Split NP):

(34) a. ?*El auto lo lavé el rojo.* [CLT]

The car 3SgM I.washed the red-SgM

b. ?*Lo lavé el auto rojo.* [CLD]

3SgM I.washed the car red-SgM

The CLT version has two masculine singular definite articles, whereas the CLD version has only one. This is because each constituent has to be licensed in c-structure with all its obligatory determiners. For this reason, despite all the similarities, I will not refer to CLT as a discontinuous constituency phenomenon, but rather as a doubling phenomenon.

Comment [BE10]: Comment on how this bears on the BigDP issue.

6.2 Word order

Even though Spanish is a relatively free word order language, CLT has a rather strict word order: most of the examples I have cited in the paper are not reversible:

(35) **Las nuevas las tengo que subir las fotos.*

The new-PIF 3PIF I.have that upload the pictures

‘I have to upload the new pictures.’

(36) **A la vieja la atropellaron a mi vecina.*

A the old-SgF 3SgF they.ran.over A my neighbor-SgF

‘My neighbor was run over, the old woman.’

≠ ‘My old neighbor was run over.’

(37) **Las que saqué ayer las tengo que subir las fotos.*

‘I have to upload the pictures that I took yesterday.’

(38) **El mío lo tengo que lavar el auto.*

‘I have to wash my car.’

(39) **Esa porquería la tengo que lavar la moto.*

‘I have to wash the motorcycle, that piece of crap.’

- 'I have to wash that crappy motorcycle.'
- (40) **A la pobre mujer la atropellaron a mi vecina.*
'My neighbor was run over, the poor woman.'
'My poor neighbor was run over.'
- (41) **A los que no se rindieron los mataron a los soldados.*
'They killed the soldiers that did not surrender.'

One possible exception is CLT with quantifiers, where it seems acceptable *prima facie* to have a preverbal focused (but not a non-focused) quantifier:

- (42) **A todos los voy a extraditar a esos.*
?A TODOS los voy a extraditar a esos.
'Those people, I am going to extradite them all.'

In fact, some of the "reversed" versions seem *prima facie* more acceptable if the preverbal constituent is focused:

- (43) ?*LAS NUEVAS las tengo que subir las fotos.*
'I have to upload the new pictures.'
- (44) **A LA VIEJA la atropellaron a mi vecina.*
'My neighbor was run over, the old woman.'
≠ 'My old neighbor was run over.'
- (45) ?*LAS QUE SAQUÉ AYER las tengo que subir las fotos.*
'I have to upload the pictures that I took yesterday.'
- (46) ?*EL MÍO lo tengo que lavar el auto.*
'I have to wash my car.'
- (47) **ESA PORQUERÍA la tengo que lavar la moto.*
'I have to wash the motorcycle, that piece of crap.'
'I have to wash that crappy motorcycle.'
- (48) **A LA POBRE MUJER la atropellaron a mi vecina.*
'My neighbor was run over, the poor woman.'
'My poor neighbor was run over.'
- (49) ?*A LOS QUE NO SE RINDIERON los mataron a los soldados.*
'They killed the soldiers that did not surrender.'

My intuition is that the focused versions are better if the interpretation of the preverbal constituent is that of an intersective modifier. The word order constraints, then, would not be syntactic, but semantic in nature. Caution, however, is required. Not only are these constructed

examples with author judgments attached to them, but also the required prosody to make them grammatical is not clear. Therefore, I will consider CLT to be generally non-reversible.

7. The functional Structure of CLT

7.1 Structure-sharing in CLT

In LFG, you can license constituents in all possible positions and share their information in a single f-structure provided all DO constituents have the same values for person, number, gender, animacy. Structure-sharing, then, rules out as ungrammatical sentences where there is any feature mismatch:⁶

<i>*Las fotos</i>	<i>las</i>	<i>vi</i>	<i>la nueva</i>
The pictures.F-Pl	3PIF	I.saw	the new-SgF
<i>*Las fotos</i>	<i>las</i>	<i>vi</i>	<i>los nuevos</i>
The pictures.F-Pl	3PIF	I.saw	the new-PIM
<i>*Las fotos</i>	<i>las</i>	<i>vi</i>	<i>las locas lindas</i>
The pictures.F-Pl	3PIF	I.saw	the crazy-PIF pretty-PIF

The first two examples have agreement mismatches; the last example is an animacy mismatch: the expression *loco lindo* means “a nice person”, hence, it is animate and cannot be coreferential with the inanimate *las fotos* (“the pictures”).

In cases where the postverbal constituent is a headless NP (either Det + Adj or Det + Relative Clause), a structure-sharing analysis is rather straightforward. I model headless NPs by a slight modification to Hahn & Wechsler’s ***2007*** treatment of Russian long-form adjectives. Spanish adjectives introduce their lexical information in an f-structure that is an element of the adjunct set of the f-structure that immediately dominates them. This is captured by the first functional annotation in the following lexical entry:

nuevas: Adj ↓ ∈ (↑ ADJ)

(↓PRED) = 'new '

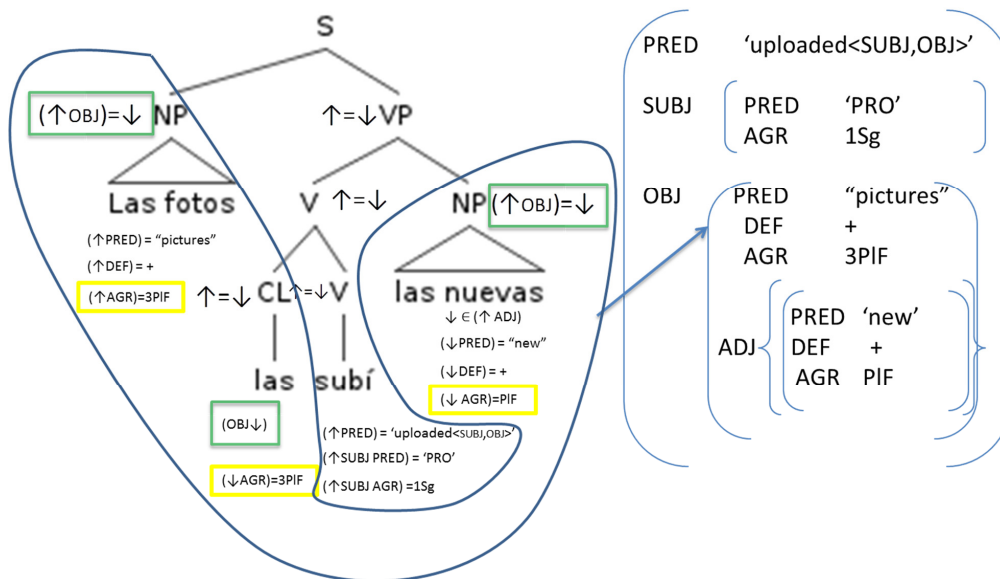
(↓*agr*) = SgF

(↓DEF) = +

⁶ However, definiteness and/or specificity might be violated, under conditions that are at present unclear (see Estigarribia 2006a): *?Las fotos las vi unas que tenían en tu computadora*

C-structurally speaking, the adjective heads a noun phrase, a phenomenon known as hybridization in the Cambridge Grammar of the English Language ****cit, Payne 2007****, and usually associated with gerunds and nominalizations in English.

So, an example like *Las fotos las subí las nuevas* (again simplifying the internal structure of NPs) receives the following analysis:



In this sentence, grammatical function assignment is enforced by coherence and completeness. The c-structure node annotations (↑OBJ)=↓ are assigned so as to make the sentence coherent. Note that, by coherence, the two NP nodes can only be assigned SUBJ or OBJ. Because the verb is inflected for 1Sg subject, and the NPs are 3PI, they cannot be assigned (↑SUBJ)=↓ or there would be a clash of features. Therefore, they must be assigned (↑OBJ)=↓. Note that if the subject was 3PIF, completeness would still require that at least one lexical NP was assigned (↑OBJ)=↓.

The three constituents marked (↑OBJ)=↓ or (OBJ)↓ share their information in a single f-structure, which is made possible by the fact that they all agree in *agr* features. If one of them was not compatible with 3PIF, then the OBJ f-structure would have two different values for the (meta)attribute *agr*, violating uniqueness.

Importantly, in the absence of prosody and discourse context, many similar c-structures are in fact ambiguous between a CLT, a CLD, or a CLLD. In the examples below all constituents (including the subject) are compatible with 3PIF agreement:

[Las fotos]OBJ las subieron [las nuevas]OBJ [CLT]

'They uploaded the new pictures.'

[Las fotos]OBJ las subieron [las nuevas]SUBJ [CLLD]

'The new ones uploaded the pictures.'

[Las fotos]OBJ las subieron [las nuevas secretarias]SUBJ [CLLD]

'The new secretaries uploaded the pictures.'

[Las secretarias]SUBJ las subieron [las nuevas]OBJ [CLD]

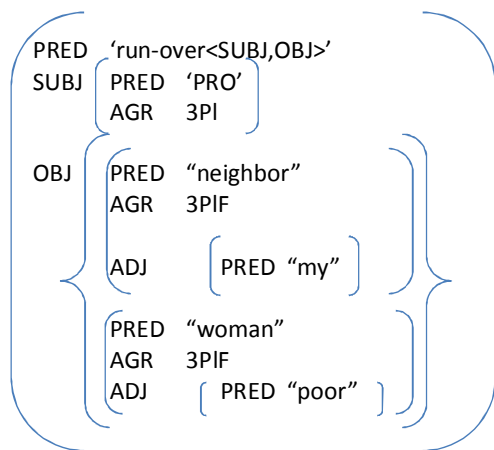
'The secretaries uploaded the new ones.'

In these structure-sharing cases, it is clear that the CLT structure behaves very much like a Split-NP, and "reconstructs" to a single GF with adjectival modification at the level of f-structure. But in other cases, the construction seems to work more like appositions:

?A mi vecina la atropellaron a la pobre mujer.

Comment [BE11]: Here I need to read the notes from Meyer 1992, "Apposition in Contemporary English".

Following ***Sadler & Nordlinger 2006***, these cases have the f-structure representation below:



This difference in f-structure representation correlates with a difference in interpretation. In the “Split-NP” CLT case the postverbal constituent identifies a subset of the referents picked up by the preverbal constituent. In the “apposition” CLT case, the postverbal constituent does not delimit the reference of the preverbal one, but rather introduces an additional semantic predication that obtains of that referent.

In order to derive this f-structure, one has to allow (as it is done by Sadler & Nordlinger 2006) for argument GFs to be set-valued, just like non-argument GFs (UDF, ADJ) are. Hence, this extension is not ad hoc, but is independently needed to treat appositional and coordination phenomena in a variety of languages. What is needed is to allow NP c-structure nodes to be annotated $\downarrow \in (\uparrow GF)$. Thus, the following is the annotated c-structure for *a mi vecina la atropellaron a la pobre mujer*:

Comment [BE12]: Restricting overgeneration and violations of coherence here is CRUCIAL.

‘Juan told Carmen that he/she had to upload the new pictures.’

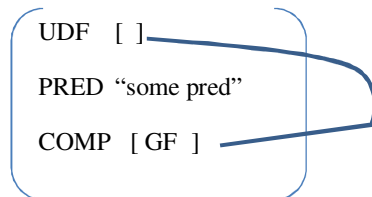
In this case, simple structure sharing will not work, since the preverbal DO and the postverbal one are in different clauses. Note that this is not a filler-gap construction: a CLT sentence minus the CLLDed constituent has all its GFs realized:

Las tenía que subir las nuevas.

This is best seen as an instance of anaphoric binding, where the CLLDed constituent is integrated into the relevant clause semantically. The extended coherence condition ***Fassi Fehri 1984; Bresnan 2001*** on f-structures supplements the coherence condition in these cases: any discourse function in an f-structure that is already complete has to be identified with (i.e., anaphorically linked to) an integrated GF. Therefore, following ***Dalrymple***, we annotate the left-dislocated constituent as follows:

$$((\text{COMP}^* \text{GF} \uparrow) \text{UDF})_{\sigma} = \uparrow_{\sigma}$$

The subscript σ indicates here that we are talking about the s(ematic)-structures corresponding to the relevant c-structure nodes, not about the f-structures. The above equation is read as follows: beginning with the mother of the node that bears the annotation (designated by the first up-arrow), you can “go up” one GF (SUBJ, OBJ, or other), and then an arbitrary number of COMPs (possibly zero), to find yourself in an f-structure that contains a UDF **whose semantic projection** is equated with the semantic projection of the node we started at. That is, if we instantiate the Kleene star as 1, this is equivalent to:



This anaphoric binding mechanism is independently needed to account for reflexive pronoun binding ***Dalrymple, Maxwell, & Zaenen 1995***, and other types of bound anaphors ***Bresnan 2001***. Again, we see that the mechanisms of structure-sharing LFG has already built-in suffice to account for CLT without added machinery.

In the next section, I will show how to model the information-structural properties of CLT in a LFG framework.

8. The informational structure of CLT

Comment [BE13]: WHAT SEEMS TO BE GOING ON HERE IS THAT THERE IS A FAMILY OF CLT CONSTRUCTIONS (CLT WOULD BE IN TURN A SUBTYPE OF THE CLD AND CLLD CONSTRUCTIONS) WHERE EACH NEEDS A SLIGHTLY DIFFERENT ANALYSIS.

So far, we have seen that LFG can easily model both CLD and CLT constructions in a principled way, without recourse to ad hoc machinery. Rather, because LFG was partly designed to account for the structure of non-configurational languages, mechanisms for structure-sharing are built-in, and they are enough to model a variety of discontinuous constituency and doubling phenomena. This section looks at the information-structure properties of CLT. I assume the treatment of information structure sketched above, in section 2, that follows **Butt & King 2000**. I distinguish 4 discourse functions represented at i-structure.

Discourse function	Attribute name	definition
<i>topic</i>	TOPIC	Discourse old, relevant in current context
<i>focus</i>	FOCUS	Discourse new, prominent
<i>background information</i>	BACK.INF	Discourse old, indicates relationship between new and old information
<i>Completive information</i>	COM.INF	Discourse new, not prominent

8.1 Discourse functions in CLT

It is often assumed that CLD encodes topicality of the DO referent, but this is not necessarily the case, as Estigarribia (2006a) demonstrates. Likewise, Focused CLT is also possible with animate DOs. The examples below show that either the LD or the post-verbal constituent can be focused:

- *¿Echaron a tu tía del trabajo?*
 They.fired A your aunt of.the job
 ‘Was your aunt fired?’

- *No, a mi mamá la echaron a la pobre.*
 No A my mom 3SgF they.fired A the poor.Sg
 ‘No, they fired my poor mom.’

In this example *a mi mamá* is a contrastive focus, and *a la pobre* is completive information.

- *¿Las querés a tus hijas?*
 3PIF you.love A your daughters
 ‘Do you love your daughters?’
- *Sí, a mis hijas las quiero a las que me tratan bien.*

Yes a my daughters 3PIF I love a the.PIF that 1Sg treat well
'Yes, my daughters I love those that treat me well.'

In this example *a mis hijas* is a *topic* (or perhaps *background information*), and *a las que me tratan bien* is the *focus*. These examples show the need for a very flexible mapping between c-structure and i-structure, one in which any constituent can bear any discourse function.

Some cases of CLD involve v-focusing or secondary predicate focusing. Estigarribia (2006a) shows that, when either the verb or a secondary predication constituent are focused, CLD is obligatory.

(13) a. <PerInv-15>

*¿No la SABÉS a ésa? / *¿No SABÉS a ésa?*

no CL you.know A that

'You don't [know]FOC that one [i.e., that story]?'

b. <Lig-45>

*Yo las tenía GUARDADAS las cartas / * Yo tenía GUARDADAS las cartas*

I CL had stored the letters

'I had the letters [stored]FOC'.

In CLD, therefore, the direct object can be a topic or background constituent (both old information) in sentences where some other constituent is focused (subject focusing, V-focusing, secondary predication focusing). Moreover, in Rioplatense Spanish, if the DO is animate it can also be focused. As for CLT, the attested examples are cases of V-focusing or quantifier focusing:

Las fotos [las tengo que subir]_{FOC} las nuevas.

A esos los voy a extraditar [a todos.]_{FOC}

Other examples can be constructed where the preverbal constituent is focused:

A MI MAMA la echaron a la pobre.

Apparently, the postverbal constituent can be focused in some cases despite not being a quantifier:

**A mi mama la echaron A LA POBRE.*

A mis hijas las quiero A LAS QUE ME TRATAN BIEN.

The constraint seems to be pragmatic/semantic rather than syntactic, namely that the postverbal constituent can be focused if it represents a **restriction** on the set identified by the preverbal constituent.

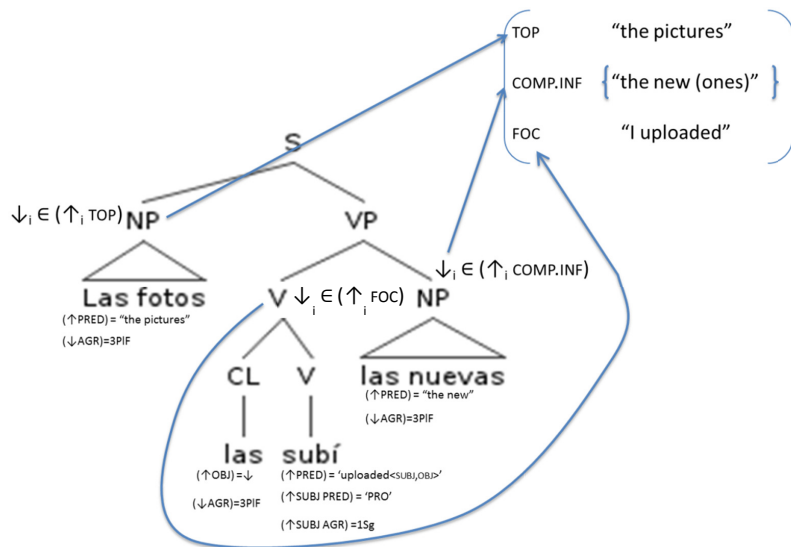
8.2 Mapping to i-structure

The mapping to i-structure I am assuming is from c-structure (for discussion of other possibilities see ***King & Zaenen 2004, Mycock 2006***). Each maximal node that dominates only one syntactically independent morpheme contributing a PRED value to f-structure is annotated with some discourse function. This allows NPs to be annotated with a discourse function, and also cliticized and inflected V nodes. This reflects the property that affixes and clitics share of not being able to carry a DF by virtue of their position in the sentence (or prosody). Clitics however, can be **lexically** specified with a DF. Since Rioplatense Spanish DOCLs can corefer with both topic and focused constituents, I hypothesize they are lexically unmarked for DFs.

So, for the following discourse fragment:

- *Tenés que subir las fotos.*

- *Sí, las fotos las subí las nuevas.*



Comment [BE14]: Find the ACTUAL context. This one is reconstructed from memory.

The NPs *las fotos* and *las nuevas* dominate only one syntactically independent word that contributes a PRED value ("fotos", "nuevas"), and therefore get an information structure

annotation. The V node annotated with the FOC DF dominates two words, but only one of them, the verb, is syntactically independent. The resulting i-structure is shown. The topic (old information that is relevant in the current context) is “las fotos”. The focus in this case is (*las*) *subí*, which is counter to the interlocutor’s expectation. Finally, “las nuevas” is completive information, new information in discourse that is nevertheless not prominent (because what is prominent is the content that is counter to expectation). The completive information status of such constituents is clear in the following (constructed) examples:

6. *Mataron a los soldados que no se rindieron*
7. *¿Qué les pasó a los soldados?*
 - a) *A los soldados, los mataron a los que no se rindieron*
 - b) *A los soldados, a los que no se rindieron los mataron*
 - c) *#Mataron a los soldados que no se rindieron*

8.3 Information structure and grammatical function mismatches

In fact, LFG’s multistratal architecture is perfectly suited to account for cases where information structure and grammatical function are mismatched. These cases are well known in the LFG literature ***e.g., King 1997, King & Zaenen 2004***. For instance, for English VP focus, the focused constituent is not one single f-structure unit:

Mary [**ate the cake**]_{FOC}.

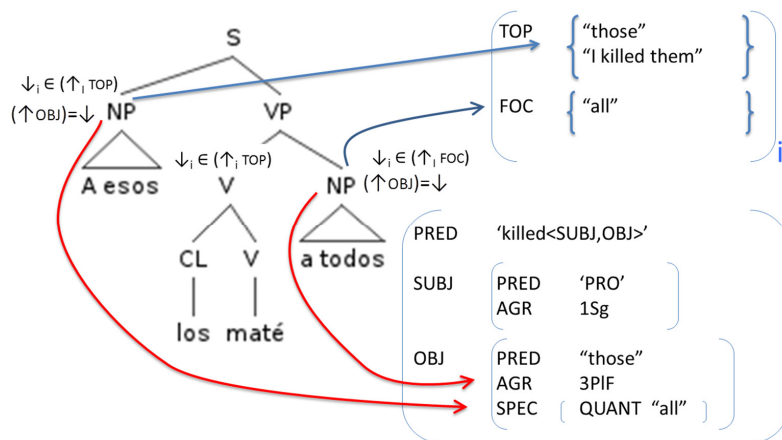
[PRED 'eat<SUBJ,OBJ>
 TNS past
 SUBJ [PRED 'Mary']
OBJ [PRED 'cake']]

[FOCUS { eat cake }]

The converse type of mismatch—where only part of an f-structure unit is focused—also occurs in CLT:

*[[A esos]_{TOP}]_{OBJ} los maté[[a todos]_{FOC}]_{OBJ}
*[[A mamá]_{FOC}]_{OBJ} la echaron [[a la pobre]_{TOP}]_{OBJ}**

In each of the above cases, the two lexical NPs that instantiates the OBJ of the verb carry different information structure functions. In the first case, the left-edge constituent is topical, and the postverbal constituent is focused, whereas in the second case the arrangement is reversed. Since the grammatical function assignment (mapping to f-structure) and the discourse function assignment (mapping to i-structure) are independent, such mismatches pose no problem and are straightforwardly accounted for. For example, for *[[A esos]_{TOP}]_{OBJ} los maté[[a todos]_{FOC}]_{OBJ}:*



9. Further questions

9.1 Restrictions on discourse function mappings

An important question left unresolved in the previous sections is whether the animacy effect in operative in CLT—that is, whether focus is restricted to animate objects as Estigarribia (2006a) showed for CLLD. Estigarribia (2006b) proposed an OT model that derives the animacy effect as a by-product of the interaction of case marking / grammatical function assignment, and differential object marking. Interestingly, CLT apparently violates the animacy restriction:

?Las fotos las tengo que subir [las nuevas]foc, las viejas ya las subí.
 The pictures 3PIF I.have that upload the new-PIF the old-PIF already 3PIF
 I.uploaded
 'I have to upload the new pictures; the old ones I uploaded already.'

Although this observation needs confirmation and further analysis, I would like to propose that an OT analysis would be viable here: even though focused in-situ DOs cannot be doubled by a clitic, this constraint is outranked by an anaphoric constraint that requires all CLLD constituents to be doubled. This latter constraint is independently needed, since clitics are obligatory in simple CLLD ***refs***.

9.2 Double postverbal tripling

I have described CLT as CLLD+CLD, that is, one preverbal, left-dislocated constituent, one clitic, and one postverbal constituent. However, attested examples show that two constituents can appear in postverbal position, without any CLLD.

(From personal emails)

- *GAP no hace más los buzos tradicionales, cambiaron el estilo, así que veo que le compro a Florencia.*

‘Gap does not manufacture their traditional sweatshirts anymore, they changed the style, so I have to think about what I will bring Florencia.’

- *En internet los venden esos buzos gap los tradicionales, no te pongas en un gasto grande.*

They sell them, on the Internet, those Gap sweatshirts, the traditional ones. Don’t spend too much.’

My intuition is that no pause is required here for either of the postverbal DOs, which is Cecchetto’s 1999 diagnostic for right-dislocation, suggesting the postverbal DOs are VP-internal. A question for further study is then whether there are syntactic or pragmatic restrictions on the position and the number of constituents that corefer with a DO clitic.

10 Conclusion

The discourse forces model (Estigarribia 2006a) was put forth to answer a little researched question in the literature on CLD: why do speakers produce such seemingly redundant structures, when the language has “simpler” constructions without doubling. The usual monofactorial answers (clitics encode topicality, or presuppositionality, or specificity) found wanting, Estigarribia (2006a) proposed that different instantiations of CLD respond to different discourse pressures. Indeed, there is a set of discourse-related factors that favor cliticization of an argument, and another set that favors lexical NP expression. CLD is used when both a clitic and a lexical NP are needed to satisfy these simultaneous pressures.

An unforeseen consequence of the discourse forces model was that, given the variety of factors that influence the form of referential expressions, it is expected that sometimes more than two constituents might be needed. This implicit prediction is true: Rioplatense Spanish has a set of clitic tripling constructions that express a direct object via two lexical NPs coreferential with a direct object clitic. This paper showed how the structural, functional, and discourse properties of both CLD and CLT can be accounted for in the LFG formalism. LFG was explicitly designed to deal with nonconfigurational languages. Hence, it can accommodate a host of doubling, tripling, and discontinuous constituency structures without ad-hoc machinery. Moreover, it represents grammatical information via several projections: c-structure for constituency and linear order; f-structure for grammatical function information; i-structure for discourse function and information structure. These projections are related by formally well-defined mappings. This multistratal architecture allows for a flexible yet formal treatment of the interfaces.

Comment [BE15]: Beef up. Take up the differences between BigDp and this analysis, and other general issues about grammar design, the syntax of modification/apposition, and the interface with info structure...