

Focus projection and clitic left dislocation in Spanish

Felix Bildhauer
Freie Universität Berlin
felix.bildhauer@fu-berlin.de

10th July 2008

Abstract

The information-structural status of clitic left dislocated arguments in Spanish has been argued to depend crucially on their thematic role. Earlier HPSG analyses of related phenomena in other languages do not take into account this sort of information. A formalisation will be presented which can handle differences in information-structure arising from different thematic roles of clitic left dislocated phrases.

1 Introduction

Spanish has a left dislocation construction in which the fronted phrase is doubled by a clitic within the core sentence whenever Spanish provides a clitic for the fronted category. The corresponding construction in Italian is discussed in Cinque (1990), where it is termed *clitic left dislocation* (henceforth CLLD). Various authors have pointed out that, from the point of view of information-structure, CLLD is a topic-marking construction (e. g. Zubizarreta 1998; Zagona 2002; Caselles-Suárez 2004). On these approaches, *topic* and *focus* designate disjoint portions of an utterance and are thus mutually exclusive.¹

On the other hand, it has been observed that whether or not a CLLD-ed constituent can be interpreted as part of the focus depends on its thematic role (Contreras 1976; Gutierrez-Bravo 2006, among others). For example, Gutierrez-Bravo (2006) argues that sentence (1a), where the indirect object is CLLD-ed, has unmarked constituent order in the sense that it allows for a sentence focus interpretation (adequate in out-of-the-blue utterances). The examples in (1b)–(1c), each of which constitutes the first sentence of a newspaper article, illustrate the same point. On the other hand, (2) displays no clitic left dislocation, but the subject cannot be interpreted as part of the focused portion of the utterance.²

- (1) a. [A Juan_{IO} le gustan_V los chocolates_S]_F³
 to Juan to.him-CL appeal the chocolates
 ‘Juan likes chocolates.’
 b. [A Carlos Fuentes no le gustan las fronteras]_F
 to Carlos Fuentes not to.him-CL appeal the frontiers
 ‘Carlos Fuentes doesn’t like frontiers.’
 c. [Al dinero_{IO} no le gustan las incertidumbres_S]_F⁴
 to.the money not to.it-CL appeal the insecurities
 ‘The world of finance doesn’t like insecurity.’

(2) Los chocolates_S le gustan_V [a Juan_{IO}]_F.
 [the chocolates to.him-CL appeal to Juan
 ‘JUAN likes chocolates.’

¹For Zagona, a topic is a special part of the ground, which in turn is complementary to focus. Zubizarreta, while assuming a twofold distinction between focus/ground and topic/comment, states explicitly that a topic can never be part of a focus.

²In (2), the indirect object is doubled by a pronominal clitic although it is not CLLD-ed. This phenomenon is pervasive with indirect objects in Spanish and is not directly relevant for the issue at hand. Glosses: S=subject, O=direct object, IO=indirect object, [...] _F=focus.

³*El País*, 09/10/1997. Carlos Fuentes, escritor.

⁴*El País*, 30/06/1997. Inversiones de baja tensión.

The behaviour of the CLLD-ed *experiencer* objects in (1) contrasts with CLLD-ed *theme* objects, which arguably cannot be interpreted as part of the focus. Example (3) illustrates the latter case (square brackets indicate what is assumed here to be largest focus that is compatible with the construction).

- (3) Este partido_o [Boca_s lo está jugando_v desde hace dos meses.]_F
 this match Boca it-CL is playing from make two months
 ‘This match, Boca has been playing it for two months.’

Thus, clitic left dislocation blocks focus projection to the entire structure in some cases, whereas it yields unmarked linear order and the corresponding broad focus reading in other cases. Earlier HPSG accounts of clitic left dislocation, such as Engdahl and Vallduví (1996) for Catalan and Alexopoulou and Kolliakou (2002) for Modern Greek, are based on Vallduví’s (1992) threefold partition into *focus*, *link* and *tail*. A link is a sentence-initial aboutness topic, and link and tail jointly constitute the ground. In these analyses, a CLLD-ed constituent is invariably interpreted as a link. Since links are defined as being part of the ground, these accounts do not in principle allow a wide-focus interpretation of CLLD constructions, and thus do not cover cases like (1) above.

With respect to Spanish, a hard-wired connection between clitic left dislocation and linkhood/topichood seems to lead to an additional problem: it is not clear whether all CLLD-ed phrases in Spanish are links/topics.⁵ As noted by Reinhart (1981), quantified NPs only qualify as aboutness topics if they can be interpreted as referential (e. g. universally quantified NPs and generic or specific indefinites). The sentences in (4) show non-referential dislocated constituents in what seems to be a clitic left dislocation construction. If these examples are indeed instances of CLLD, then CLLD-ed phrases need not be links/topics in all cases.

- (4) a. [A nadie]_{io} le gusta que lo tilden de ladrón⁶
 to nobody to.him-CL appeals that him-CL call of thief
 ‘Nobody likes being called a thief.’
 b. [A mucha gente]_{io} le gustan las promesas de los partidos⁷
 to many people to.them-CL appeal the promises of the parties
 Many people like the promises of the political parties.

In what follows, I will propose an HPSG approach that can handle the differences in focus projection arising from different thematic roles of the CLLD-ed phrase. Instead of Vallduví’s (1992) three-way categorisation of information-structural primitives, an orthogonal two-dimensional distinction between topic/comment and focus/ground is assumed. The topic/comment and focus/ground partitions of a sentence are allowed to overlap in ways excluded under Vallduví’s approach. In particular, nothing prevents topics from being embedded within foci, such that out-of-the-blue utterances like those in (1) may still be analysed as containing an aboutness topic.⁸

2 Clitic left dislocation

As noted by Cinque (1990), Balari (1998) and others, clitic left dislocation differs from other long distance dependencies in that it fails to correlate with phenomena typically observed with ordinary extraction, such as sensitivity to islands and obligatory subject inversion in Spanish. Balari argues that clitic left dislocation constructions are weak unbounded dependencies: only indices are shared between the fronted phrase and the clitic, while binding theory is relied on for ruling out ungrammatical dislocations. However, CLLD-ed phrases in Spanish show case agreement with the corresponding clitic, and since CASE is not represented on indices, it is hard to see how such an approach can rule out case mismatches. Alexopoulou and Kolliakou (2002) propose an account of *clitic left dislocation* in Modern Greek. At the heart of their proposal lies a set-valued CLITIC feature, which is an additional non-local feature and serves to optionally

⁵This probably also holds for Catalan, since data parallel to (4) can easily be found.

⁶*Revista Hoy*, 29/12/1997. El técnico que no ríe.

⁷*Diario de Jerez Digital*, 27/01/2004. Hablando en el desierto.

⁸The idea that topic and focus may be embedded within each other is not new. Chafe (1976) suggested that all-new sentences can be construed as conveying information about a particular entity, and thus contain an aboutness topic (a subject, in his terminology). More recently, Frey (2004) has argued that the focused part of a sentence may in principle contain an aboutness topic. Conversely, Krifka (2007) proposes that contrastive topics (as discussed in Büring 1997) are contrastive precisely because they contain a focus which introduces alternatives. See also Steedman (2000). In this paper I will not be concerned with topic instantiation. The HPSG formalisation presented below leaves topic instantiation in CLLD constructions underspecified.

collect information about cliticised arguments. As with other non-local features, the CLITICS set is passed up to dominating nodes. In analogy to SLASH dependencies, a phrasal type *cldd-phrase* finally licenses the combination of a left dislocated constituent with a head daughter that has an appropriate object in its CLITIC set. Alexopoulou and Kolliakou argue that these objects cannot be of type *local*, as is commonly assumed for SLASH dependencies. The reason is that *local* objects contain semantic information specifying (in the case of nouns) the subtype of *nominal object*. Assuming that object clitics are specified as *pronominal*, this may conflict with the specification of the dislocated phrase, which may or may not be pronominal. In order to overcome this difficulty, Alexopoulou and Kolliakou propose that a dislocated phrase and the corresponding clitic share only HEAD values. To ensure sharing of agreement information, they are forced to modify the commonly assumed HPSG feature geometry such that agreement features are represented in the HEAD path (instead of INDEX). However, such a move will complicate an account of symmetric coordination, and it does not really seem to be necessary either: if *cldd-phrase* does not require token identity of the non-head daughter’s LOCAL value with some element in the head daughter’s CLITIC set, the problem does not arise. Instead, sharing of only HEAD and INDEX values can be specified in *cldd-phrase*, as illustrated in Figure 1.⁹ Note that the head daughter must be saturated for its complements, but may still subcategorise for a subject, thus allowing CLLD-phrases to intervene between the subject and the VP.

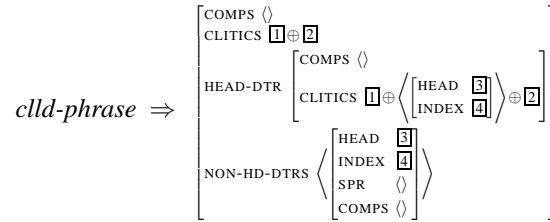


Figure 1: Constraint on phrasal type *cldd-phrase*

As noted above, the information-structural partitioning I am assuming here divides an utterance into *focus/ground* and *topic/comment*. Unlike Engdahl and Vallduvf’s (1996) and Alexopoulou and Kolliakou’s (2002) approaches, the focused portion of a sentence may include the non-head daughter in a *cldd-phrase* in some cases. Moreover, the non-head daughter need not always be interpreted as a topic. Focus projection will be modelled by means of an interface constraint between linearisation and information-structure, which will take into account information associated with the thematic role of the fronted constituent.

3 Thematic roles

Within the current HPSG feature geometry, information about thematic roles is contained in the semantic contribution of the head that assigns these roles to its dependents. Role attributes proposed in the HPSG literature range from specific features for every semantic relation (Pollard and Sag 1994) to very generic attributes (Flickinger et. al. 2003), with most approaches falling somewhere in between (e. g. Davis 2001). The current grammar architecture does not provide a means to retrieve this kind of information from the dependents when these are realised syntactically. While it seems clear that information derived from thematic roles is needed in order to appropriately constrain focus projection in clitic left dislocation constructions, it is not desirable to directly associate discourse function with thematic roles. The reason is that the relationship between linearisation, thematic role and discourse function may be affected by specific constructions (such as passive, see Contreras 1976), and possibly also by extensions of a head’s argument structure. Moreover, as pointed out by Müller (1999) (who discusses a suggestion by Uszkoreit 1986), representing thematic roles on the dependents that carry them is problematic because a dependent may be assigned different roles by different verbs in a complex predicate. To avoid these complications, I suggest to model the connection between semantics and linearisation by means of a mediating boolean-valued feature UPV (*unmarked preverbal*), located under LOCAL. A head may then specify which of its dependents, if any, can be realised as a non-head daughter in a broad-focus CLLD construction. These specifications need not be stipulated for every single verb, but can be expressed as constraints on lexical

⁹Here, the CLITICS feature takes a list as its value.

types. To illustrate, the statements in (5a)–(5b) constrain the UPV value of transitive verbs and intransitive psych verbs of the “gustar”-class, respectively. Linking of arguments to thematic roles is included here for expository purposes, and nothing hinges on the rather specific role attributes.

$$\begin{aligned}
 (5) \quad a. \text{ strict-tr-v-lxm} &\Rightarrow \left[\begin{array}{l} \text{ARG-ST} \left\langle \left[\begin{array}{c} \text{UPV} \quad + \\ \text{INDEX} \quad \boxed{1} \end{array} \right], \left[\begin{array}{c} \text{UPV} \quad - \\ \text{INDEX} \quad \boxed{2} \end{array} \right] \right\rangle \\ \text{RELS} \left\langle \left[\begin{array}{c} \text{AGENT} \quad \boxed{1} \\ \text{THEME} \quad \boxed{2} \end{array} \right] \right\rangle \end{array} \right] \\
 b. \text{ io-unerg-itr-v-lxm} &\Rightarrow \left[\begin{array}{l} \text{ARG-ST} \left\langle \left[\begin{array}{c} \text{UPV} \quad - \\ \text{INDEX} \quad \boxed{1} \end{array} \right], \left[\begin{array}{c} \text{UPV} \quad + \\ \text{INDEX} \quad \boxed{2} \end{array} \right] \right\rangle \\ \text{RELS} \left\langle \left[\begin{array}{c} \text{THEME} \quad \boxed{1} \\ \text{EXPERIENCER} \quad \boxed{2} \end{array} \right] \right\rangle \end{array} \right]
 \end{aligned}$$

As for (5a), the assumption is somewhat simplified since it presupposes that all transitive verbs assign a THEME and an AGENT role. In fact, verbs like *temer* ‘to fear’ are transitive, but the subject is arguably an EXPERIENCER rather than an AGENT. However, the constraint in (5a) could be further refined such that it applies only to the relevant subset of transitive verbs.

4 Interface constraints

Instantiation of the FOCUS value in CLLD constructions can now be made sensitive to the UPV value of the dislocated constituent. In addition to the HEAD and INDEX values, the UPV value must be shared between the relevant object in the head verb’s CLITICS list and the dislocated phrase. This can be achieved by slightly modifying the constraint on *clld-phrase*, as shown in Figure 2.

$$\text{clld-phrase} \Rightarrow \left[\begin{array}{l} \text{COMPS} \langle \rangle \\ \text{CLITICS} \boxed{1} \oplus \boxed{2} \\ \text{HEAD-DTR} \left[\begin{array}{l} \text{COMPS} \langle \rangle \\ \text{CLITICS} \boxed{1} \oplus \left[\begin{array}{c} \text{HEAD} \quad \boxed{3} \\ \text{INDEX} \quad \boxed{4} \\ \text{UPV} \quad \boxed{5} \end{array} \right] \oplus \boxed{2} \end{array} \right] \\ \text{NON-HD-DTRS} \left\langle \left[\begin{array}{c} \text{HEAD} \quad \boxed{3} \\ \text{INDEX} \quad \boxed{4} \\ \text{UPV} \quad \boxed{5} \\ \text{SPR} \quad \langle \rangle \\ \text{COMPS} \langle \rangle \end{array} \right] \right\rangle \end{array} \right]$$

Figure 2: Revised constraint on phrasal type *clld-phrase*

The interface constraint in Figure 3 may now refer to the UPV feature: if the dislocated constituent is [UPV −], then its semantic contribution cannot be part of the focus. In this case, the FOCUS value of the entire *clld-phrase* must be identical to that of the head daughter. On the other hand, if the fronted phrase is [UPV +], the constraint in Figure 3 does not apply, thus allowing for a reading where the *clld-phrase* as a whole contributes to focus.

$$\left[\begin{array}{l} \text{clld-phr} \\ \text{NON-HD-DTRS} \left\langle \left[\text{UPV} \quad - \right] \right\rangle \end{array} \right] \Rightarrow \left[\begin{array}{l} \text{FOCUS} \boxed{1} \\ \text{HEAD-DTR} | \text{FOCUS} \boxed{2} \end{array} \right]$$

Figure 3: Syntax/information-structure interface constraint on *clld-phrase*

FOCUS is a list-valued feature here, and a phrase’s FOCUS value may become instantiated in one of two ways: either all the daughters’ FOCUS values are collected (see De Kuthy 2002), or the phrase’s FOCUS list contains as its single element the phrase’s RELS value. The latter case corresponds to focus projection, where the semantics of the entire phrase contributes to focus. Focus projection is assumed here to be generally possible unless some constraint blocks it. The interface constraint in Figure 3 is one such constraint. It will block focus projection whenever a CLLD-ed constituent is not the one which, according to its thematic role, may appear preverbally in unmarked constituent order.

Analyses of sentences (1a) and (3) above are given in Figures 4 and 5 on the next page, respectively. In each case, they describe an interpretation with a maximally large constituent in focus. Both sentences have other readings, not illustrated here, in which the focused part is smaller. The important point is that the dislocated constituent cannot be interpreted as focused in (3), while it may or may not be part of the focused portion in (1a).

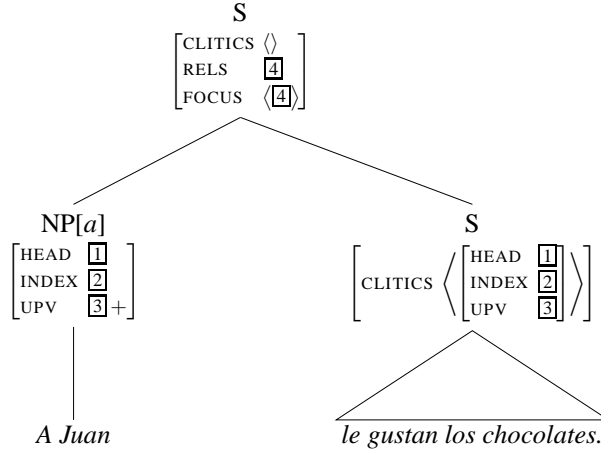


Figure 4: Broad focus reading with a CLLD-ed *experiencer* object

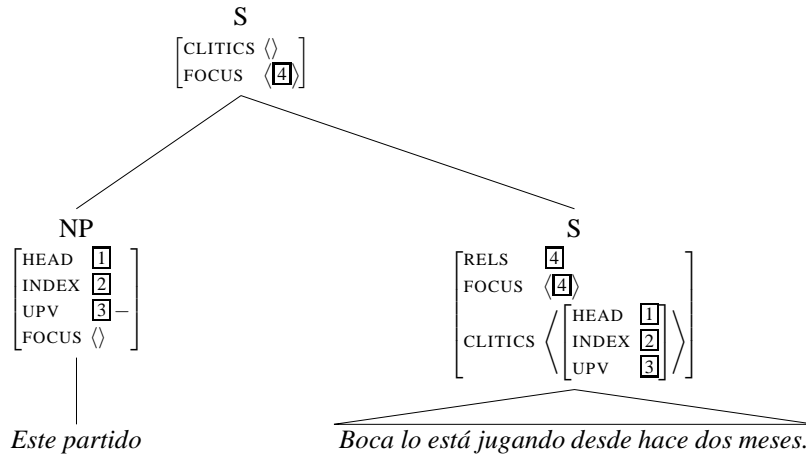


Figure 5: Restricted focus projection with a CLLD-ed *theme* object

5 Summary

The formalisation proposed here makes available on a verb's dependents just the right amount of information that is necessary in order to constrain focus projection in Spanish clitic left dislocation constructions. Following Contreras (1976), Gutierrez-Bravo (2006) and others in assuming that the crucial factor is the thematic role of the dislocated constituent, I showed how the connection between thematic roles and unmarked constituent order can be established at a point where information about thematic roles is retrievable without complications, that is, in the lexicon. Constraining focus projection indirectly by using a mediating feature (UPV) seems to be more promising than stating a direct connection between thematic roles and unmarked order, since changes in a verb's argument structure may affect unmarked linearisation of the arguments, while their thematic roles need not change. Focus instantiation has only been sketched in the present proposal, and it is clear that prosodic factors as well as linearisation constraints in the postverbal field have to be taken into account in order to restrict it further.

- Alexopoulou, Theodora and Kolliakou, Dimitra. 2002. On Linkhood, Topicalization and Clitic Left Dislocation. *Journal of Linguistics* 38, 193–245.
- Balari, Sergio. 1998. Pronouns, Variables, and Extraction in HPSG. In Sergio Balari and Luca Dini (eds.), *Romance in HPSG*, pages 151–217, Stanford: CSLI.
- Büring, Daniel. 1997. *The Meaning of Topic and Focus. The 59th Bridge Accent*. London and New York: Routledge.
- Casielles-Suárez, Eugenia. 2004. *The Syntax-Information Structure Interface. Evidence from Spanish and English*. London: Routledge.
- Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. In Charles N. Li (ed.), *Subject and Topic*, pages 25–55, New York: Academic Press.
- Cinque, Guglielmo. 1990. *Types of A'-dependencies*. Cambridge, Mass.: MIT Press.
- Contreras, Heles. 1976. *A theory of word order with special reference to Spanish*. Amsterdam: North-Holland.
- Davis, Anthony R. 2001. *Linking by Types in the Hierarchical Lexicon*. Stanford: CSLI Publications.
- De Kuthy, Kordula. 2002. *Discontinuous NPs in German*. Stanford: CSLI Publications.
- Engdahl, Elisabeth and Vallduví, Enric. 1996. Information Packaging in HPSG. In Claire Grover and Enric Vallduví (eds.), *Studies in HPSG*, Edinburgh working papers in cognitive science, No. 12, pages 1–31, Edinburgh: Centre for Cognitive Science, University of Edinburgh.
- Flickinger, Dan, Bender, Emily M. and Oepen, Stephan. 2003. MRS in the LinGO Grammar Matrix. A Practical User's Guide, <http://faculty.washington.edu/ebender/papers/userguide.pdf>.
- Frey, Werner. 2004. A Medial Topic Position for German. *Linguistische Berichte* 198, 153–190.
- Gutierrez-Bravo, Rodrigo. 2006. *Structural Markedness and Syntactic Structure*. New York and London: Routledge.
- Krifka, Manfred. 2007. Basic Notions of Information Structure. In Caroline Féry, Gisbert Fanselow and Manfred Krifka (eds.), *Interdisciplinary Studies on Information Structure*, Working Papers of the SFB 632, No. 6, pages 13–56, Potsdam: Universitätsverlag.
- Müller, Stefan. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Linguistische Arbeiten, No. 394, Tübingen: Max Niemeyer Verlag, <http://www.cl.uni-bremen.de/~stefan/Pub/hpsg.html>.
- Pollard, Carl J. and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, London: University of Chicago Press.
- Reinhart, Tanya. 1981. Pragmatics and Linguistics. An Analysis of Sentence Topics. *Philosophica* 27, 53–94.
- Steedman, Mark. 2000. Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry* 31, 649 – 689.
- Uszkoreit, Hans. 1986. Constraints on Order. *Linguistics* 24, 883–906.
- Vallduví, Enric. 1992. *The informational component*. New York: Garland.
- Zagona, Karen. 2002. *The syntax of Spanish*. Cambridge: Cambridge University Press.
- Zubizarreta, Maria Luisa. 1998. *Prosody, Focus and Word Order*. Cambridge, Massachusetts: MIT Press.