

## Arbitrary SE and accusative clitics: resolving (mis)matches in the syntax

Monica Alexandrina Irimia & Anna Pineda

(University of Modena and Reggio Emilia & University of Barcelona)

**Intro.** Restrictions on clusters with arbitrary SE clitic ( $SE_{ARB}$ , collapsing passive and impersonal) and 3<sup>rd</sup> person accusative clitics ( $CL_{ACC}$ ), such as in (1)a/b, are pervasive across Romance (Mendikoetxea and Battye [M&B] 1990, MacDonald and Melgares [M&M] 2021, Pescarini 2021, a.o.). We evaluate two main lines of investigation - syntactic and PF - into their complex nature, against less discussed, but problematic patterns, from Catalan (3), American Spanish varieties or other Romance languages. We show that resolution of such clusters *is* a matter of syntax (as opposed to PF), but requires a more in-depth analysis of various parameters, such as (i) internal structure/labelling options in  $CL_{ACC}$ , (ii) licensing needs in CL, or (iii) T-v relation.

### (1) Standard Spanish (and varieties with $*SE_{ARB}-CL_{ACC}$ )

a. \*Se **lo** ve. b. \*Se **los** ve. c. Se **le/les** da eso.  
 $SE_{ARB}$   $CL_{ACC}.M.3SG$  see.3SG  $SE_{ARB}$   $CL_{ACC}.M.3PL$  see.3SG  $SE_{ARB}$   $CL_{IO}.DAT.3SG/PL$  give.3SG it  
Intended: ‘He is seen.’ Intended: ‘They are seen.’ ‘He/they is/are given it.’

**A problem of D match at PF.** M&M (2021) note that grammaticality of (1)a/b is restored when  $SE_{ARB}$  and  $CL_{ACC}$  are *not linearly* adjacent, as in (2)a/b vs (2)c (also Corsican, M&B 1990, Martins and Nunes 2016 for European Portuguese, etc.). Thus, the problem in (1)a/b is not the *impossibility* of licensing structural accusatives, contra Ordóñez and Treviño’s (O&T 2011, 2016) syntactic account. For M&M (2021), it is instead a co-occurrence restriction on clitic clusters. Linear adjacency of [D] heads is banned, but  $CL_{ACC}$  (*lo/los*) and  $SE_{ARB}$  are both specified as [D](efinite):  $SE_{ARB}$ , due to T’s inherently valued D (which blocks the implicit EA’s linking to the discourse, allowing only non-referential readings), and  $CL_{ACC}$ , as proved by doubling contexts (Suñer 1988, Nevins 2007, Ormazabal and Romero [O&R] 2013, M&M 2021, a.o.). Indirect object (IO)  $CL_{DAT}$ , as in (1)c, is underspecified for D, and [ $SE_{ARB}-CL_{DAT}$ ] cluster is licit.

(2) a. En Navidad, se suele poder abrazar { $\sqrt{lo}$ }. (about Santa Claus)  
in Christmas  $SE_{ARB}$  tend.3SG can.INF hug.INF-him.ACC

‘At Christmas, one tends to be able to hug him.’ (M&M 2021: ex. 15a; Honduras Spanish)

b. ..., se suele poder { $\sqrt{lo}$ } abrazar. c. se { $\sqrt{lo}/\sqrt{le}_{DAT=DO.ANIM}$ } suele poder abrazar.

In Spanish (*leísta*) varieties, grammaticality of (1)a/b can also be restored if ACC morphology is replaced by DAT (for animates), as in (2)c, with *no* change in grammatical function; what is spelled out as DAT on the surface is still an accusative direct object (DO) clitic *syntactically*. For M&M (2021), the change from *lo* to *le* in (2)c is due to morphological impoverishment: the [D] feature on  $CL_{ACC}$  is deleted at PF, prior to Vocabulary Insertion. As this is a post-syntactic operation, no interpretive effects ensue (eg., grammatical function stays unchanged as a DO).

**Catalan.** We turn now to Catalan  $SE_{ARB}$ , which is challenging for both these accounts - it allows the 3<sup>rd</sup> p. *el(s)*  $CL_{ACC=DO}$ , as in (3)a (M&B 1990, Bartra 2002, GIEC 2016, a.o.). However, the ‘neuter’ 3<sup>rd</sup> p. *ho*  $CL_{ACC=DO}$ , in (3)b, is ungrammatical;  $CL_{DAT=IO}$ , as expected, goes through, - (3)c.

### (3) Catalan

a. Se’**ls** veu. b. \*S’**ho** veu. c. Se **li** dona això.  
 $SE_{ARB}-CL_{ACC}.M.3SG/3PL$  see.3SG  $SE_{ARB}-CL_{ACC}.3N$  see.3SG  $SE_{ARB}$   $CL_{DAT}.IO.3SG$  give.3SG this  
‘He/they is/are seen.’ Intended: ‘It is seen.’ ‘He is given this (thing).’

Given that the clitics in (3)a/b are clearly accusative (Catalan does not have subject clitics; dative ((3)c), locative or partitive clitics have distinct morphology), such examples are counterevidence to O&T. A PF-based account (M&M 2021) is not intuitive either: *el* DO clitic in (3)a is not spelled out with dative morphology ((3)a vs (2)c/(3)c); there are important *interpretive* differences between *el* in (3)a and *ho* in (3)b, as *ho* cannot be used for animates - see especially GIEC (2016: §18.6.2); as opposed to  $CL_{DAT}$ , both *el* and *ho* give evidence of the

presence of a [D] feature: *ho* doubles only the universal quantifier *tot* ((4)d), known to interact with definiteness; *el* is restricted to pronouns, being ungrammatical with indefinites.

**Feature clashes in the syntax.** We agree with M&M (2021) in that accusative Case can be licensed in  $SE_{ARB}$  (in the languages under discussion); we differ in that we illustrate cases where ungrammaticality in  $SE_{ARB}$  clitic clusters is not a matter of PF. Subsequently, we investigate a different path: the clash is caused when categories with *mismatching* (values of) features are found on the same (extended) head in the syntax, such as in the cluster  $\{[+D]_{CL_{ACC}} [-D]_{SE_{ARB}}\}$ . We build on Otero (1986, p.91) and subsequent work which formalize the functional role of  $SE_{ARB}$  as “to ‘absorb’ the plus value of [+D(ef)] in a finite Infl”, thus blocking linking to the discourse. If a clitic with [+D] ends up on the same head as  $SE_{ARB}$ , a clash should arise. Of course, this would predict that the cluster  $\{[+D]_{CL_{ACC}} [-D]_{SE_{ARB}}\}$  should never be grammatical; we see however that (Romance) languages have ways to work around this restriction, besides non-adjacency: Catalan  $[CL_{ACC}-SE_{ARB}]$  clusters contrast in grammaticality, although both *el* and *ho* give indication of the presence of [+D]. The question then becomes: what type of mechanism is responsible for rendering the [+D] feature opaque in some environments?

**Clashes in labels.** We propose that it is the *label* that counts; even if the category might contain [+D], the higher label might ‘protect’ [+D], thus rendering it unproblematic for [-D] in  $SE_{ARB}$ . Under this hypothesis, certain types of  $CL_{ACC}$  (animate, individualized, etc.; *el* form in (3)a), or  $le_{DAT=DO.ANIM}$  in (2)c) contain additional features above D, and it is these features that provide the label, as in (4)a. This avoids a clash with [-D] label of  $SE_{ARB}$ . In turn, *ho* (or  $CL_{ACC}$  in (1)a, b) lacks additional labelling structure; being default [+D] ((4)c), cannot escape the clash. We take advantage of research on: i) whether features related to animacy, individuation, etc. are simply an index on D or head their own (labelling) projection (e.g., K head; López 2012 for differential object marking, a.o.); ii) enriched DP configurations, with a difference between so-called ‘low D’ (above which various projections related to individuation, animacy, referentiality, gender<sub>γ</sub> etc. can be merged) and ‘high D’ (Bernstein, Ordóñez & Roca 2019, 2020). Higher INDIV, ANIM,  $\gamma$  ((4)c), act as categories that can ‘protect’ D in cluster formation. This also correctly predicts the existence of dialects where the restriction in (1)a holds only with  $CL_{ACC}.MASC$  (default [+D]), but not with  $CL_{ACC}.FEM$  ( $*se_{ARB} lo_{CL_{ACC}.M} ve$  vs  $\checkmark se_{ARB} la_{CL_{ACC}.F} ve$ ).

- (4) a. Animate  $CL_{ACC}$ :  $[_{ANIM} Anim/Indiv [D D [_{ACC} ACC...]]]$ ; d. **Ho** vull tot.  
 b. Gender  $CL_{ACC}$ :  $[\gamma Gender [D D [_{ACC} ACC...]]]$   $CL_{ACC}.3N$  want.1SG all  
 c.  $CL_{ACC}$  (*ho*, (1)a, b):  $[D +D [_{ACC} ACC...]]$ ; ‘I want it all/I want everything.’

What is at stake is precisely the internal composition of clitics in the syntax, even if additional structure in Catalan  $CL_{ACC}$  *el* is *not* signalled at PF via *dative* morphology, as it happens in (2)c.

**Licensing needs.** Additional evidence that the restriction in (1)a/b is not simply a PF one comes from contrasts such as in (5), holding in the same dialects with the restriction in (1)a/b. The only difference between (5)a and (5)b resides in the nature of  $me_{CL.DAT}$  – ethical dative in (5)b vs. goal dative in (5)a. For (5)a,  $CL_{DATIO=GOAL}$  is possible by itself with  $SE_{ARB}$ , as expected (*se me da* ‘it is given to me’), and  $le_{DAT=DO.ANIM}$  is possible by itself too (as we saw in (2)c), as they do not count as [+D]. If all that matters is avoidance of linear adjacency of [D], PF accounts predict (5)a to be grammatical. In turn, syntactic accounts have shown that both  $CL_{DATIO=GOAL}$  and  $CL_{DAT=DO.ANIM}$  have licensing needs in the syntax (O&R 2007, a.o.), eg., due to the presence of a labelling [+PERSON] above D. As there is only one [PERSON] licenser but two [PERSON] features to license, (5)a ends up being ungrammatical, independently of the fact that  $CL_{DATIO=GOAL}$  and  $CL_{DAT=DO.ANIM}$  are not ‘seen’ as [+D].

- (5) a. \*Al hombre, **se me le** envía. b. A mi hijo, **se me le** suspende (mucho = much).  
 DOM man  $SE_{ARB}$   $CL_{DATGOAL}$   $DAT_{DO.ANIM}$  sends. DOM my son  $SE_{ARB}$   $CL_{DATETH}$   $DAT_{DO.ANIM}$  suspends  
 Intended: ‘The man, he is sent to me.’ ‘My son, he is suspended on me (quite a lot).’

**To conclude**, mapping internal structure/licensing needs reduces apparently disparate  $CL_{ACC}-SE_{ARB}$  patterns to a unified mechanism in syntax, while also deriving variation; eg., Italian  $CL_{ACC}$

is only possible in a higher position than  $SE_{ARB}$  (M&B 1990, D'Alessandro 2007; ✓*lo si vede* vs \**si lo vede*), indicating a higher licenser for the labelling category in  $CL_{ACC}$ . Lastly, we evaluate conditions of [-D] on  $SE_{ARB}$  disrupting the T- $\nu$  relation (Saab 2014), leading to ungrammaticality of *referential accusatives* (beyond clitics), as in eg. Romanian (Dobrovie-Sorin 1998).