Defective Agreement in Spanish Dative Clitic-Doubling:
A Cross-Dialectal Study of Morphosyntactic Variation

Gustavo Guajardo

Abstract: In Spanish clitic-doubling constructions, the clitic should agree in number with its coreferential doubled noun phrase. However, the present corpus data analysis reveals that, under certain structural configurations, number agreement is not always realized on the third-person dative clitic. In such cases, plural indirect objects appear doubled with a singular clitic when the indirect object is a lexical noun phrase (80% vs. 20%). In this paper, I use conditional inference trees and generalized mixed-effects logistic regression modeling to study the factors that drive this variation in number agreement. The results show significant effects of the intervening direct object on the dative clitic. When the indirect object is a lexical noun phrase, animacy of the indirect object interacts with the number feature of the direct object but the singular clitic is always more likely whenever the direct object is singular, suggesting a type of attraction effect between the direct object and the clitic. In sharp contrast, when the indirect object is realized as a pronoun, full agreement is almost categorical. Importantly, this type of variable agreement appears to be constrained by the structural context where clitic doubling takes place and I argue that these constraints challenge the idea that the dative clitic can receive a uniform analysis and propose that the categorial status of the clitic depends on the structural configuration in which it appears.

1.   Introduction

Agreement is a pervasive linguistic phenomenon whereby a dependent phrase agrees in certain features with an agreement controller (Corbett 2006). For example, in *John visits his parents on the weekends*, the NP *John* controls agreement on the verb *to visit* so that it is realized as *visit-s* and not *visit.* Naively, one would think that if agreement is obligatory in a language then agreement should be realized whenever its conditions are met. However, there is plenty of evidence that this is not always the case. For example, by studying the effect of interfering elements between the controller and the dependent element researchers have found that agreement mechanisms are indeed subject to interfering or attraction effects with the result that the expected agreement relationship may not be properly established (more on in Section 3). In this paper, I examine agreement between a clitic and its doubled NP in Spanish clitic-doubling constructions with third person dative clitics. Descriptively, I show that agreement in certain clitic-doubling constructions is most often left unrealized and this applies to all Spanish dialects (contra what has been suggested in previous literature). Analytically, I study the linguistic

factors that can predict the realization of agreement between the clitic and its coreferential NP and argue that this phenomenon can shed light on the issue of where clitics originate and the much-debated nature of the category of Spanish clitics[1].

2. The Linguistic Phenomenon under Investigation

In Spanish double-object constructions, the indirect object is typically doubled with a dative clitic which must agree in number with it (1-2).

1. **Le**$_i$ di   un   caramelo   [a   la   niña]$_i$.
   CL.DAT.S   give.1S.PAST   a   sweet   to   the   girl
   'I gave the girl a sweet'

2. **Les**$_i$ di   un   caramelo   [a   las   niñas]$_i$.
   CL.DAT.PL   give.1S.PAST   a   sweet   to   the   girl
   'I gave the girls a sweet'

In (1) the indirect object NP *la niña* "the girl" is singular so the singular clitic *le* is required, whereas in (2) the plural clitic *les* appears because the indirect object *las niñas* "the girls" is plural.

Whether or not doubling occurs depends on a number of factors. Doubling is considered obligatory when the indirect object (IO) is realized as a pronoun (3a), in benefactive constructions (3b), with dative experiencers (3c) and with inalienable possession (3d) (Belloro 2007)

3. a.   Le   di   un   regalo   a   ella.
   CL.DAT.S   give.1S.PAST   a   gift   to   her
   'I gave her a gift'

     b.   Le   hice   los   deberes   a   mi   hermano.
   CL.DAT.S   do.1S.PAST   the   homework   to   my   brother
   'I did my brother's homework'

---

[1] The dataset and Rmarkdown with the code of the statistical analysis will be publicly available once the paper has been published .

<table>
<tr><td>c.</td><td>Le</td><td>molesta</td><td>el</td><td>ruido</td><td>a</td><td>Juan.</td></tr>
<tr><td></td><td>CL.DAT.S</td><td>bother.3S.PRES</td><td>the</td><td>noise</td><td>to</td><td>Juan</td></tr>
<tr><td></td><td colspan="6">'Juan loves ice cream'</td></tr>
</table>

<table>
<tr><td>d.</td><td>Le</td><td>duele</td><td>la</td><td>cabeza</td><td>a</td><td>la</td><td>maestra.</td></tr>
<tr><td></td><td>CL.DAT.S</td><td>hurt.3S.PRES</td><td>the</td><td>head</td><td>to</td><td>the</td><td>teacher</td></tr>
<tr><td></td><td colspan="7">'The teacher has a headache'</td></tr>
</table>

With NPs expressing the notion of goal (4), the clitic is said to be optional in Peninsular Spanish (Demonte 1994) whereas certain American varieties show a stronger preference for the doubled construction (Bentivoglio 1978 for Venezuelan Spanish, Jaeggli 1982 for Argentinean Spanish, Silva-Corvalán 1981 for Chilean Spanish, Company Company 2006 for Mexican Spanish).

<table>
<tr><td>4.</td><td>(Le)</td><td>devolví</td><td>los</td><td>libros</td><td>a</td><td>Federico.</td></tr>
<tr><td></td><td>CL.DAT.S</td><td>return.1S.PAST</td><td>the</td><td>books</td><td>to</td><td>Federico</td></tr>
<tr><td></td><td colspan="6">'I returned the books to Federico'</td></tr>
</table>

A property of this phenomenon that has received little attention is the fact that number agreement between the clitic and its coreferential NP does not always obtain, such that a singular dative clitic may co-occur with a coreferential plural indirect object (5).

<table>
<tr><td>5.</td><td>a.</td><td>No</td><td>**le**</td><td>tiene</td><td>miedo</td><td>a</td><td>**los**</td><td>**gatos**.</td></tr>
<tr><td></td><td></td><td>not</td><td>CL.DAT.S</td><td>have.3S.PRES</td><td>fear</td><td>to</td><td>the</td><td>cats</td></tr>
<tr><td></td><td></td><td colspan="7">'It is not afraid of cats.'</td></tr>
</table>

<table>
<tr><td></td><td>b.</td><td>**Le**</td><td>ofreció</td><td>dinero</td><td>a</td><td>**los**</td><td>**vecinos**.</td></tr>
<tr><td></td><td></td><td>CL.DAT.S</td><td>offer.3S.PAST</td><td>money</td><td>to</td><td>the</td><td>neighbors</td></tr>
<tr><td></td><td></td><td colspan="6">'He offered money to the neighbors.'</td></tr>
</table>

<table>
<tr><td></td><td>c.</td><td>**Le**</td><td>cambió</td><td>la</td><td>mentalidad</td><td>a</td><td>**los**</td><td>**jugadores**.</td></tr>
<tr><td></td><td></td><td>CL.DAT.S</td><td>change.3S.PAST</td><td>the</td><td>mentality</td><td>to</td><td>the</td><td>players</td></tr>
<tr><td></td><td></td><td colspan="7">'He changed the players' mentality.'</td></tr>
</table>

Following Roca (1993), I will call this phenomenon *defective agreement,* that is, the use of a third person singular dative clitic when the apparent "agreeing" element is clearly plural. While it has been acknowledged in the literature that this phenomenon exists (e.g., Kany 1951, DeMello 1992, Franco 1993, Torrego 1995, Belloro 2007), researchers seem to assume that this

type of agreement mismatch is possible in any construction the clitic occurs, that it is a somewhat rare phenomenon and subject to dialectal variation. In this paper, however, I show (i) that the structural context in which defective agreement can occur is highly constrained, (ii) that defective agreement is far more frequent than full agreement when the appropriate structural conditions for it are met and (iii) that this is true across all Spanish varieties.

Defective agreement is only possible in clitic-doubling constructions (6a), but it is not possible when the clitic stands alone without an overt coreferential NP (6b). Defective agreement is also disallowed in left-dislocated constructions (7). With dative experiencers, however, the singular clitic can occur although some speakers report a preference for the plural clitic (8).

6. a.   ¿Qué **le**   diste   a   los   niños?
          what   **CL.DAT.S**   give.2s.past   to   the   children
          'What did you give the children?

    b.   **Les/*Le**   di   un   caramelo.
          CL.DAT.PL/ S   give.1s.past   a   sweet
          'I gave them a sweet'

7. A   los   niños,   **les/ *le**   dieron   un   premio.
    to   the   children   **CL.DAT.PL/S**   give.3pl   a   award
    'They gave the children an award'

8. A   los   chicos **les/ le** gusta el helado.
    to   the   kids   **CL.DAT.PL/S**
    'Children like ice cream.'

These data suggest that defective agreement is highly constrained with respect to the structural contexts in which it is allowed; not every construction with a dative clitic allows for defective agreement. These observations raise two related questions: (i) in those contexts in which defective agreement can occur, which factors favor the choice of clitic? (after all, full agreement is always an option) and (ii) why is defective agreement allowed in some clitic constructions but not in others?

To cover the above facts, the hypothesis offered herein suggests that what we see is a byproduct of the interaction of linear (surface) distance between the clitic and the coreferential NP; the direct object (DO) may intervene as a possible candidate for misanalysis of the reference of the clitic because it is linearly closer to it. Therefore, I expect that defective agreement will be more likely with intervening singular DOs than with plural ones. Moreover, because animacy is a salient feature in Spanish that interacts with other properties within NP-syntax to explain grammaticalized, contextual and dialectal variation (e.g., differential object marking, clitic doubling of DOs), I hypothesized that animacy may also play a role in the alternation of defective agreement; inanimate objects (either direct or indirect) might favor defective agreement to a greater extent than animate objects, not least because animates across the board seem either to require more or greater precision in morphosyntactic marking (e.g., double object marking in Spanish, obligatory plural marking in Japanese and Farsi, etc.). To the extent that defective agreement is driven by the linear order of the objects, it can be expected that animacy of the DO may have a bigger effect on defective agreement than the animacy of the IO.

Since the hypothesis makes reference to the existence of an attraction effect by the DO, a discussion of this phenomenon is in order to understand what I mean by it.

3.   Attraction Effects

The formation of linguistic dependencies between words in a long-distance dependency is subject to what is known as *attraction effect,* the phenomenon whereby the computation of agreement relations can be disrupted in the presence of alternative candidates (Bock and Miller 1991, Nicol et al. 1997). For example, in the by-now-famous sentence *the key to the cabinets are __* , the finite verb agrees with the NP closest to it (i.e., *the cabinets*) even though the

grammatical subject is the singular NP *the key.* Terminologically speaking, the true grammatically agreeing NP is called *the controller* and the interfering NP *the distractor.*

The literature on attraction effects has mainly focused on subject-verb agreement phenomena but more recently researchers have started to look at other types of dependency relations such as antecedent-pronoun agreement to investigate whether this type of dependency is also subject to similar attraction effects. The findings are still somewhat inconclusive as different studies have yielded conflicting results. For example, some studies have found that antecedent-pronoun agreement seems to be as sensitive to attraction effects as subject-verb agreement (Bock et al. 1999, 2004; Patil et al. 2016) whereas others have found that gender agreement in antecedent-reflexive dependencies is not affected by the interference from other competing antecedents (e.g., Nicol and Swinney 1989, Sturt 2003, Dillon et al. 2013).

In a recent study, Jäger *et al* (2020) replicate the experiment in Dillon *et al* (2013) with a higher number of participants (40 vs. 181). The main conclusion of their study is that they find no difference between attraction effects in subject-verb agreement and antecedent-reflexive dependencies. The caveat to this finding, however, is that the dependent measure used in Dillon et al. is total fixation time. The authors highlight that total fixation time has the potential to be too broad a measure because it subsumes all types of dependent measures from eye-tracking. To explore whether a more fine-grained analysis could shed some light on the status of reflexives in attraction effects they analyze first-pass reading time and first-pass regressions. They find no effect of first-pass reading time in the interaction between type of dependency and attraction effect (i.e., neither subject-verb agreement nor reflexives showed attraction effects). In first-pass regressions, they find an asymmetry between grammatical and ungrammatical sentences. In grammatical sentences reflexives show inhibitory interference (i.e., slowdown in reaction time) whereas in ungrammatical sentences subject-verb agreement shows a clear interference effect but reflexives do not show any interference. Thus, it remains

an empirical question whether, and to what extent, attraction effects are mediated by the type of linguistic dependency or whether all types of agreement relations can, in principle, be subject to attraction effects.

A relevant finding in the literature has been the asymmetry exhibited between singular and plural agreement. It has been found that attraction effects are only present in singular but not plural agreement (Nicol et al. 1997, Wagers et al. 2009). For example, ungrammatical sentences such as *the key to the cabinets are there,* containing a plural distractor and a singular controller, have been shown to facilitate self-paced reading in comprehension as well as elicit higher acceptability ratings than its counterpart, *the keys to the cabinet is there*, where the distractor is singular and the controller plural (Nicol et al. 1997, Wagers et al. 2009). The finding that plural-marked distractors are stronger than singular ones, both in production and comprehension, has been interpreted to suggest that morphological markedness plays a central role in agreement attraction (Santesteban et al. 2017).

Crucially for our purpose, the markedness effect predicts that a singular distractor should not elicit an agreement mismatch when the controller is plural. In other words, a sentence akin to *the keys to the cabinet is there* is predicted not to be likely to be found in corpora (i.e., not produced) whereas a sentence such as *the key to the cabinets are there* should be. Note, however, that defective agreement constitutes an example of the unpredicted pattern: a singular clitic appears with a plural controller in the presence of a singular distractor. This may be an indication that defective agreement may not be a result of an attraction effect after all. I return to this point in Section 6.2.

Before moving on to the present study, a review of the phenomenon of clitic doubling and the theoretical analyses of clitic doubling in Spanish within a generative perspective is in order.

4. Clitic doubling in Spanish

Clitic doubling (CD) is the phenomenon where a phonologically bound element with special placement rules (i.e., the clitic) expresses agreement features of an NP that is an argument to the verb in the same propositional structure (Adger 2003, Anagnostapoulou 2014, Harizanov 2014). This means that the clitic and the co-indexed NP share the same case and *phi*-features. CD is a phenomenon present in many language types as diverse as Romance, Balkan, Bantu and Semitic but each language imposes different constraints on the context in which CD can take place (Anagnostopoulou 2006, 2017). Crosslinguistically, doubling of IOs is more common than doubling of DOs. In Spanish, CD is obligatory with preverbal objects in left-dislocated constructions for both direct (9) and indirect (10) objects.

9. a.   [A   Pedro] **lo**   vi   anoche.
           A   Pedro   CL.ACC.S   see.1S.PAST   last.night
           'I saw Pedro last night'

     b.   [La   mesa] **la**   compré   ayer.
           the   table   CL.ACC.F.S   buy.1S.PAST   yesterday
           'I bought the table yesterday'

10. a.   [A   Ana] **le**   dieron   un   regalo
           a   Ana   CL.DAT.S   give.3PL.PAST   a   gift
           'They gave Ana a gift'

     b.   [A   la   mesa] **le**   cortaron   la   pata.
           A   the   table   CL.DAT.S   cut.3.P.PAST   the   leg
           'They cut one leg off the table'

With postverbal NPs, direct objects cannot generally be doubled in standard Spanish (but this is not true for other varieties, such as Argentinean Spanish, Jaeggli 1982, Suñer 1988). Clitic doubling exceptions seem to be when the DO is the universal quantifier *todo* 'all' (11a), and with strong pronouns (11b) (Ormazabal and Romero 2010).

11. a.   Lo   inventó   todo
           CL.ACC   invent.3S.PAST   everything
           'He made everything up'

b. Hay   que   llamarlo   a   él.
    There.is   that   call.INF-CL.ACC.SG.MASC   to   him
    'We have to call him.'

The analysis of the obligatoriness of clitic doubling in these contexts has been ascribed to the non-argumental status of the left-dislocated NP in (9) and (10) and the strong pronoun in (11b). Regarding the left-dislocated NP, it is generally assumed that it is base generated in this position and therefore the clitic acts as the true argument of the verb; without the clitic, the selectional restrictions of the verb would not be fulfilled (Roca 1992). Similarly, the analysis of the structural position of the strong pronoun has been that it is not inserted in a verb-related argument position. Evidence for this claim comes from a variety of processes that strong pronouns cannot enter into but clitics can. Two of these arguments, for example, are that clitics but not strong pronouns can function as resumptive pronouns, and strong pronouns cannot refer to dislocated elements whereas clitics can (e.g., Rigau 1988, Picallo 1991, Roca 1992).

In contrast to the highly constrained contexts in which doubled DOs are possible, any indirect object can be optionally doubled with the dative clitic (12) (Suñer 1988, Roca 1992, Torrego 1995, Dufter and Stark 2008). The only type of NP that some authors have claimed cannot be doubled is unqualified bare NPs (Suñer 1988), but there is disagreement on this (e.g., Roca 1992). In general, doubling of IOs appears to be the predominant pattern in contemporary Spanish, especially in Latin American Spanish (Becerra Bascuñán 2006, Company Company 2006; but see Hentschel 2013 for a different conclusion).

12. a. Le   pinté   la   pata   a   la   mesa.
        CL.DAT   paint.1S.PAST the   leg   A   the   table
        'I painted the leg of the table.'

     b. Le   corta   el   pelo   a   la   suegra.
        CL.DAT   cut.3S.PRES   the   hair   A   the   mother-in-law
        'S/he cuts her/his mother-in-law's hair.'

c. Le tocó la cabeza a un perro.

CL.DAT touch.3S.PAST the back A a dog

'S/he touched the head of a dog'

In (12a) the IO is a specific definite inanimate noun, (12b) contains an IO that is specific definite and animate and (12c) is an example of a specific indefinite animate IO.

In general, the hierarchy in (13) seems to hold for clitic doubling (Anagnostopoulou 2017). This means that pronouns are most likely to be doubled (and we saw in Spanish this is an obligatory context for CD), followed by IOs and last DOs. Conversely, (13) predicts that if a language allows CD of DOs then it should allow CD of IOs and pronouns.

13. Pronouns > IOs > DOs

4.1 Theoretical analyses of Clitic Doubling

There is a very rich literature on clitic doubling in Spanish and even more generally. In this section, I will focus only on the issues that are relevant for the purposes of this paper: (i) where the clitic originates and (ii) the status of the doubled NPs in cases where clitic doubling is obligatory.

One of the long-standing debates about CD is the origin of the clitic in the syntactic structure. There have been two main approaches within generative grammar: the movement approach (e.g., Kayne 1975, 1989; Rizzi 1986, Sportiche 1989) and the base-generated approach (e.g., Strozer 1976, Rivas 1977, Aoun 1981, Jaeggli 1982, Borer 1984). The movement approach treats clitics as pronominal elements which are generated in the argument position in the VP and then move to a higher functional projection. The base-generated approach argues that Romance clitics are verbal agreement markers and are attached to inflectional heads (e.g., Borer 1984, Saltarelli 1987, Suñer 1988, Roca 1992, Franco 1993).

Proponents of the movement approach argue that the clitic is generated in the argument position of the VP and then moves to an inflectional head (e.g., Kayne 1975, 1989; Sportiche

1989, 1990). This movement then would leave a silent copy of the clitic in the argument position, which helps explain the complementary distribution in some languages between clitics and full NPs like in French. The argument position would still be filled with the silent copy and therefore the lexical NP could not occupy such position. There is quite a great deal of evidence in favor of the movement analysis, all of which involves typical effects of movement (for a thorough overview of these, see Anagnostopoulou 2006). For example, Kayne (1975) shows that the dative clitic cannot reach the main clause over the subject of an embedded clause suggesting a Specified Subject Condition effect, where the clitic could not move over the (intermediate) subject in the embedded clause. Sportiche (1995) shows that, in French, clitics cannot be extracted out of NPs unless the XP being extracted could otherwise appear as the possessor of the NP. This requirement holds of any type of XP more generally, clitics being one of them. If clitics were base generated in the surface position in which they appear, this pattern would be unexpected. Another piece of evidence for the movement approach comes from Spanish clitic climbing, where in the presence of certain multiple verb combinations the clitic can appear as a proclitic of the finite verb or enclitic to the non-finite verb with no apparent change in meaning (14) (e.g., Kayne 1975, Rizzi 1976, 1982; Wurmbrand 2001).

14. a. Quiero   comprar**lo**.
        want.1S.PRES   buy.INF.CL
     'I want to buy it'

    b. **Lo** quiero   comprar
       cl   want.1S.PRES   buy.INF
     'I want to buy it'

However, even though Spanish clitic climbing supports the claim that clitics may originate in the argument position of the verb and move to a higher projection, Spanish also poses a problem for this same approach. In clitic doubling the clitic co-occurs with its co-referential NP so the movement approach becomes more problematic to defend because the clitic could not originate in the argument position of the VP (Anagnostopoulou 2006), it being

filled by the lexical NP. This observation has led to the view that clitics must be generated in the surface position in which they appear.

One of the arguments for the based-generated approach comes from so-called ethical datives (15), which are obligatorily realized as clitics and never alternate with a lexical NP. This suggests that the dative clitic is assigned its own theta-role and this theta-role is never assigned to an NP in argument position.

15. a.
Me	le	secuestraron	a	mi	hijo
CL.DAT	CL.DAT	kidnap.3P.PAST	a	my	son

'They have kidnapped my son'

b.
*Secuestraron	a	mi	hijo	a	mí.
kidnap.3P.PAST	a	my	son	a	me

The base-generated approach and the existence of clitic doubling phenomena lead to the question of what the status of both the clitic and the coreferential NP could be in the syntax. A result of this enterprise was the proposal that clitics were agreement markers (Borer 1984). One of the first proponents that Spanish clitics were agreement markers is Suñer (1988), where she proposes the Matching Principle. According to this principle, the clitic and the NP must agree in *phi*-features. Suñer assumes the clitics are specified with a number of features such as [animacy], [gender], [number] and [person] which have to agree with those of the coreferential NP. Under this analysis, the less constrained nature of CD of IOs over DOs stems from the fact that DOs are assumed to contain a [specific] feature, which IO do not. This means that DOs can only be doubled with specific NPs.

However, this proposal has been challenged by, for example, Franco (1993) who argues that doubling of DOs is possible with non-specific referents and by Fernández Soriano (1989) who shows that IOs cannot be doubled with non-specific NPs. To accommodate these facts, Franco (1993) proposes an alternative theory of agreement where he puts forth *The Agreement Hypothesis* to explain the behavior of Spanish clitics. Accordingly, any theory that postulates

that clitics are agreement markers must be able to explain the apparent optionality of the phenomenon when compared to an undisputed agreement system as subject-verb agreement. In this regard, it is doubling of direct objects that needs an explanation as doubling of indirect objects has become almost categorical (Bentivoglio 1978, Silva-Corbalán 1981). Franco proposes that clitics are agreement heads that project agreement phrases. Much like verbal heads, which show selectional restrictions on its specifiers, so does the agreement head projected by the clitic. He proposes that in those dialects that allow doubling of DO more freely, the agreement head selects for [+definite, + animate] noun phrases. Case assignment and the selectional restriction of the agreement head can then explain the different constraints between IO and DO agreement. Because IOs receive case inherently or lexically, they always occur with a preposition so they do not have to move to the specifier position of any agreement head to check case. In this way, IOs are not subject to the selectional restrictions of the agreement head. However, DOs need to move to the agreement head to receive case. In this process, if the selectional restrictions of the agreement head do not match the features of the DO, then the sentence becomes ungrammatical. Under this analysis, it is the need of the DO to move to the agreement head to receive case that makes doubling of DOs more constrained as this movement must be accompanied by matching the selectional restrictions of the agreement head. To accommodate the apparent optionality of doubling of DOs in this theory, Franco proposes that there is zero object agreement with certain NPs. The support for this proposal comes from crosslinguistic data from languages with object agreement (Paus 1990). In his study, Paus found that in 20 out of 26 languages, third person object agreement is unmarked with a default zero morpheme.

Other authors have taken a more nuanced approach and argued for a mixed analysis of Spanish clitics by proposing that the accusative third person clitic is a determiner and the dative clitic is an agreement marker (e.g., Roca 1992, Torrego 1995, Ormazabal and Romero 2010).

Although these authors agree that the accusative and the dative clitics must be different syntactic elements and they all agree on the status of the accusative clitic as heading an NP, they differ in the status they assign the dative clitic. Roca (1992) does not commit himself to calling it an agreement morpheme but prefers to leave it as an undefined category, whereas Ormazabal and Romero (2010) argue that the dative clitic is an agreement morpheme. Similarly to Franco (1993), Ormazabal and Romero argue that Spanish has a multiple agreement system like Quechua or Basque with agreement markers for DO, IO and subject agreement. But, similar to the third person singular present tense in Spanish, which lacks an overt person marker, they argue that the third person singular DO marker agreement is also lacking because the third person accusative clitics are determiners. They use the restrictions on DO doubling as an argument for their different status from IO clitics, which, as we saw, do not present such constraints. Besides belonging to different syntactic categories, under these analyses the accusative and the dative clitics are said to originate in a different structural position. The accusative clitic is proposed to be the head of an NP in the argument position of the verb and the dative clitic as a head of a (higher) functional projection. This analysis then means that the accusative clitic must move in the course of the derivation whereas the dative clitic need not.

Vázquez Rozas and García Salido (2012) take a diachronic perspective to the study of clitics in Spanish. They argue that clitics are agreement object markers whose evolution as object markers is due to the frequency with which they occurred with certain NPs. More specifically, their claim is that clitics appeared most frequently with highly accessible NPs and these are the NPs that have developed object agreement. Thus, their conclusion is that the object agreement patterns of contemporary Spanish is a case of fixation and morphologization of highly frequent patterns. According to their proposal, the difference between DO and IO in their ability to be doubled stems from the fact that IO tend to be highly animate and human (i.e., highly accessible) whereas DO tend to be the opposite. Within this approach the difference in

behavior between the two clitics may be explained by a difference in the degree of grammaticalization that each clitic has undergone or finds itself in. Thus, the dative clitic would be more advanced in this process, expressing a higher degree of semantic bleaching (Enrique Arias 2003) whereas the accusative clitic is in an earlier stage of this process.

What these proposals highlight is that the status of the dative clitic as an agreement object marker is much less controversial than that of the accusative clitic: doubling of IOs appears to be more categorical and there are no semantic restrictions on the type of IO that can be doubled (with the caveat mentioned above). Most proposals then have the burden to find a mechanism that constrains doubling of the DO. Another observation that becomes clear from the literature is the assumption that clitic phenomena can be characterized in very general terms independent of the particular syntactic structures in which clitics appear. To put matters differently, the assumption has been that if clitic X appears in construction Y and the same clitic appears in construction Z, then the analysis of the clitic must be general enough, and even independent of the construction Y or Z. This is not a trivial issue for a morpheme whose behavior has been so difficult to account for so it needs to be empirically motivated and shown whether this is indeed the case. I return to this question in Section 6.3.

This paper contributes to the debate about the nature of clitics in, at least, two ways. On the one hand, the phenomenon under study constitutes further support to the treatment of accusative and dative clitics as distinct elements (be it because they are syntactically different or on a different stage of grammaticalization) in that a phenomenon similar to defective agreement but with DOs in CD is, to my knowledge, unattested in any Spanish variety (note that the phenomenon in Andean Spanish where a default third person accusative clitic is used is different from defective agreement because it applies across the board and there is no alternation of gender or number; there is simply lack of agreement altogether, (see Escobar 1978, Escobar 2000 for a description and discussion of the phenomenon and more recently

Zdrojewski and Sanchez 2014 for an analysis and comparison to other dialects)). Therefore, data such as (16), where a singular accusative clitic appears doubled with a plural lexical DO, are not grammatical in the dialects that allow CD of DOs.

16. a   *Lo   vi   a   los   chicos anoche.
           CL.ACC.MASC see.1S.PAST   A   the   guys   last.night
           'I saw the guys last night'

    b.   *La   invité   a   mis   amigas a   casa   ayer
           CL.ACC.FEM   invite.1S.PAST A   my   friends to   house   yesterday
           'I invited my friends over yesterday'

In both (16a) and (16b) the accusative clitic is singular but the coreferential NP is plural and therefore the sentence is ill-formed in the syntax. Needless to say, these data raise the question why defective agreement is only possible with IOs so I return to this issue in Section 6.

The second contribution into the nature of clitics is the fact that defective agreement is sensitive to the structural context in which the clitic appears. In other words, I will show that the contexts in which defective agreement can and cannot occur are those same contexts in which clitic doubling of the DO is obligatory and I will propose an explanation for this overlap based on the analyses we just discussed about the nature of the doubled NPs and pronouns in clitic doubling.

5. Methodology

Corpus data was extracted and annotated and then used to fit two complementary machine learning methods. First a conditional inference tree was fit to the data and these results were used to inform the choice of variable predictors in a mixed-effects logistic regression model. In this section, I describe the procedure for data extraction, annotation and choice of final model for each method.

5.1 Data

The data were extracted from the Web/Dialect version of Corpus del español (Davies 2016). The corpus contains nearly 2 billion words from 21 Spanish-speaking countries, including the U.S. The data is tagged for morphosyntactic, lexical and semantic information and comes from websites, including blogs and forums. The source of each sentence, whether it is a website or a blog is included in the results of the search as is the website's address of each sentence. Data extraction was done manually alternating the position (*pre-* or *postverbal*) and the number of the clitic (*sg* vs. *pl*), the gender, number and definiteness features of the DO (*masc* vs. *fem*; *sg* vs. *pl*; *bare* vs. *definite* vs. *indefinite*) and the gender of the IO. Only sentences with definite IOs and with a token frequency of at least 2 were extracted.

After manual removal of duplicates, a total of relevant 2376 sentences were left for analysis which were coded for the following variables: SOURCE, COUNTRY, ADDRESS, FINITE VERB, NON-FINITE VERB, PERSON, NUMBER OF VERB, ANIMACY OF DO, ANIMACY OF IO, POSITION OF CLITIC, NUMBER OF DO, GENDER OF DO, GENDER OF IO, DEFINITENESS OF DO, TYPE OF IO. A brief explanation and the levels of each variable are presented in table 1.

Variable Name	Levels	Variable Name	Levels
SOURCE	B(log), G(eneral)	ANIMACY_IO	animate, inanimate
COUNTRY	All 21 countries	POSITION OF CLITIC	pre, post
ADDRESS	website address	NUMBER_DO	sg, pl
FINITE_V	any verb	GENDER_DO	fem, masc
NON-FINITE_V	any verb	GENDER_IO	fem, masc
PERSON_V	1st, 2nd, 3rd	DEFINITENESS_DO	def, indef, bare
NUMBER_V	sg, pl	TYPE_IO	pron, NP
ANIMACY_DO	animate, inanimate		

Table 1. Coded variables and their levels

4.2. Statistical Analysis

Statistical analysis was performed in *R* (R Core Team 2016). Although the final dataset contained 2,376 sentences, the distribution of each clitic was very disproportionate with singular *le* occurring 78% of the time and plural *les* only 22%. As is well-known in the machine

learning literature imbalanced datasets are extremely problematic for classification algorithms because they are biased towards the majority class, irrespective of which machine learning algorithm is used (Provost 2000, Chawla *et al.* 2004, He and Garcia 2009, Wallace *et al.* 2012, Lopez *et al.* 2013, Raeder *et al.* 2013, Krawczyk 2016). With this issue in mind, the first step in the analysis was to remedy the data imbalance by using random undersampling in the ROSE package in *R* (Lunardon et al. 2014). This resampling method simply keeps all the data points from the minority class (i.e., *les*) and randomly removes cases of the majority class. The desired proportion of majority/minority class can be set at different values. The machine learning literature suggests that a fully balanced proportion of majority and minority class (i.e., 50/50) is not always the best solution for the class imbalance problem (Provost 2000, Weiss and Provost 2006) so it is suggested that different proportions be explored to find the best value for the particular dataset at hand. Thus, I randomly constructed four datasets with different proportions of minority/majority class: 30/70, 40/60, 50/50, 60/40 with the *ovun.sample* function in the ROSE package. Then each undersampled dataset was used to fit a conditional inference tree with the *party* package (Hothorn et al. 2006) and balanced accuracy and the Kappa statistic for each tree was calculated and used to determine the best distribution.[2] The highest balanced accuracy and Kappa values were obtained with a 40/60 proportion ($Bal.Acc_{30/70}$ = 0.59, Kappa = 0.22; $Bal.Acc_{40/60}$ = 0.70, Kappa = 0.39; $Bal.Acc_{50/50}$ = 0.65 , Kappa = 0.30 ; $Bal.Acc_{60/40}$ = 0.68, Kappa = 0.38). This final dataset contained 1,333 sentences (60% *le* vs. 40% *les)*[3].

The balanced dataset was used to fit the conditional inference tree and a mixed effects logistic regression model with ADDRESS and VERB as random intercepts. The variable ADDRESS

---

[2] The conditional inference trees were built without the variables ADDRESS and VERB because these contained too many levels and were considered random effects.
[3] As undersampling is a random procedure, a seed must be set to ensure the same dataset is obtained every time the algorithm is run. All the undersampled datasets were generated with seed = 1.

was used as a proxy for speaker as this information is not available in the corpus. ANIMACY_DO was only used in the conditional inference tree but was removed from the logistic regression analysis because most DOs were inanimate. Only ten out of 2,376 sentences contained an animate DO, with just one sentence containing the plural clitic.

6. Results

In this section, I will first present the descriptive statistics to give the reader a sense of the characteristics of the data and will then present the results of the two models. The descriptive statistics section is based on the original dataset of 2,376 sentences.

6.1 Descriptive Statistics

We will start by looking at the distribution of *defective agreement* across all Spanish-speaking countries. As can be observed in Fig 1. the use of singular *le* with overt indirect plural objects is present in all varieties of Spanish. Perhaps the most surprising finding that stands out from these data is the fact that in no variety of Spanish is the plural clitic used more often than the singular form in this context. The country with the lowest proportion of singular *le* is the Dominican Republic at 0.67 whereas Paraguay contains the highest proportion at 0.88. On average, the relative frequency of the singular clitic is 0.78.
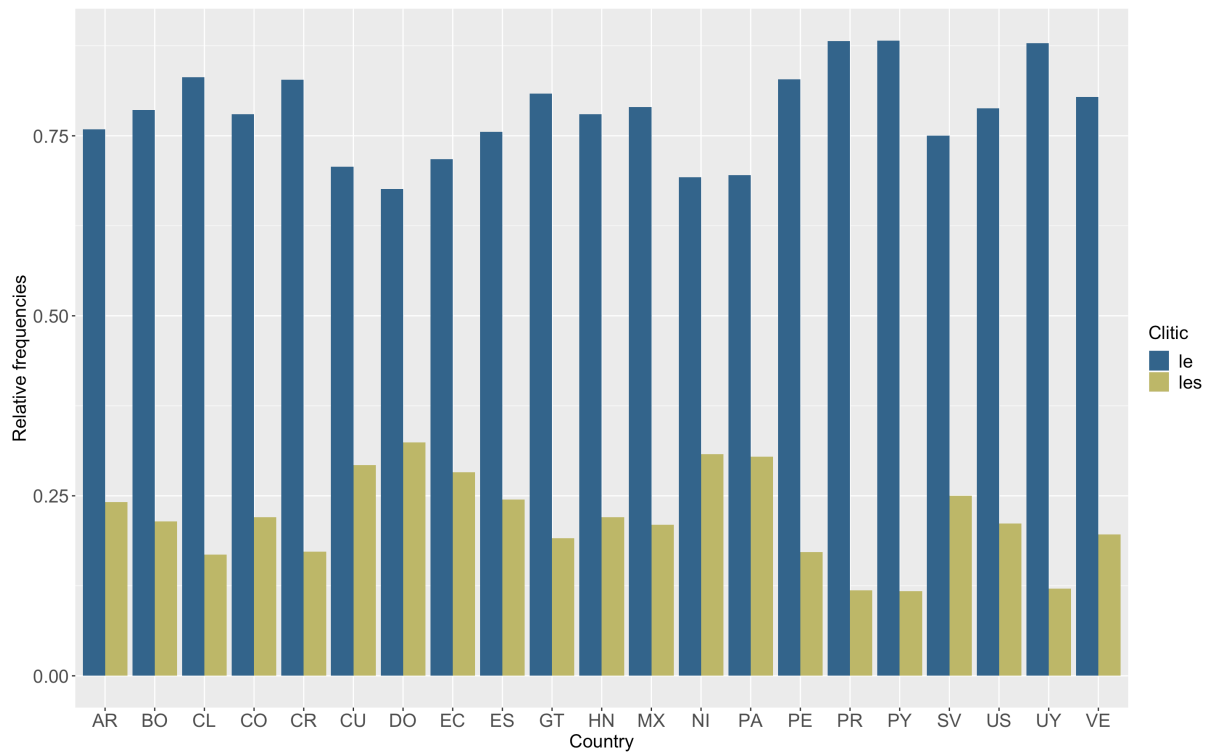
Figure 1. Relative frequency of *le* and *les* in clitic doubling with plural indirect objects by country.

In Figure 2, I present the distribution of *le* and *les* across all of the remaining variables. The first two plots show the distribution of the clitic with respect to animacy of the direct and indirect object, respectively. The singular clitic seems to appear more with animate direct objects than with inanimate (0.90 vs. 0.77). In contrast, there appears to be a relatively large difference between animate and inanimate indirect objects such that most inanimate indirect objects appear with the singular clitic (0.90 with singular vs. 0.10 with plural). Animate indirect objects show the same direction in preferring the singular clitic but the difference is not as big as with inanimates (0.70 with singular vs. 0.30 with plural). Regarding the definiteness feature of the direct object, we see that the plural clitic has the lowest frequency with indefinites and bare nouns with a relative frequency of 0.20. As for the gender of the direct object, there seems to be a small difference across the two genders where feminine direct objects show a slightly higher frequency of the plural clitic relative to masculine nouns (0.25 vs. 0.19). The exact

opposite seems to be true for indirect objects as the plural clitic shows a higher relative frequency with masculine nouns.

Regarding the number feature of the direct object, we can see that the singular clitic appears more often with singular direct objects than with plural ones (0.82 vs. 0.60). With respect to the position of the clitic, the singular clitic occurs only slightly more often when the clitic appears postverbally (0.80 vs. 0.77). Likewise, there does not seem to be a big difference in the type of source (i.e., blogs vs. general websites), but there is a slightly higher frequency of the plural clitic in the general category (0.20 vs. 0.24). The next variable is the syntactic type of the indirect object. There are two noteworthy observations that are clear from this plot: (i) the syntactic type of the indirect object (*full NP* vs. *pronoun*) is the variable with the largest difference between the two values of the variable and, most importantly, (ii) this is the only case in which one of the levels of a variable favors the plural clitic. As is clear from the plot, when the indirect object is a pronoun, the relative frequency of the singular clitic is extremely low and this is the only time in which the plural clitic is the preferred form (0.90 vs. 0.10). The opposite is true with full NPs where, in line with the rest of the data, the preferred form appears to be the singular clitic. Although the difference between the singular and plural clitic is also very large with NPs, the difference between the two forms of the clitic is slightly smaller than we saw with pronouns (0.80 vs. 0.20). The last two plots show the distribution of the clitic with respect to verbal features. There appears to be no difference regarding the number feature of the verb with the singular clitic being the most frequent form with both singular and plural verbs. In contrast, there is a clear difference between second person on the one hand and first and third person on the other, in that second person seems to favor use of the singular clitic much more strongly relative to the other two persons (1st: 0.77, 2nd: 0.90, 3rd: 0.77).
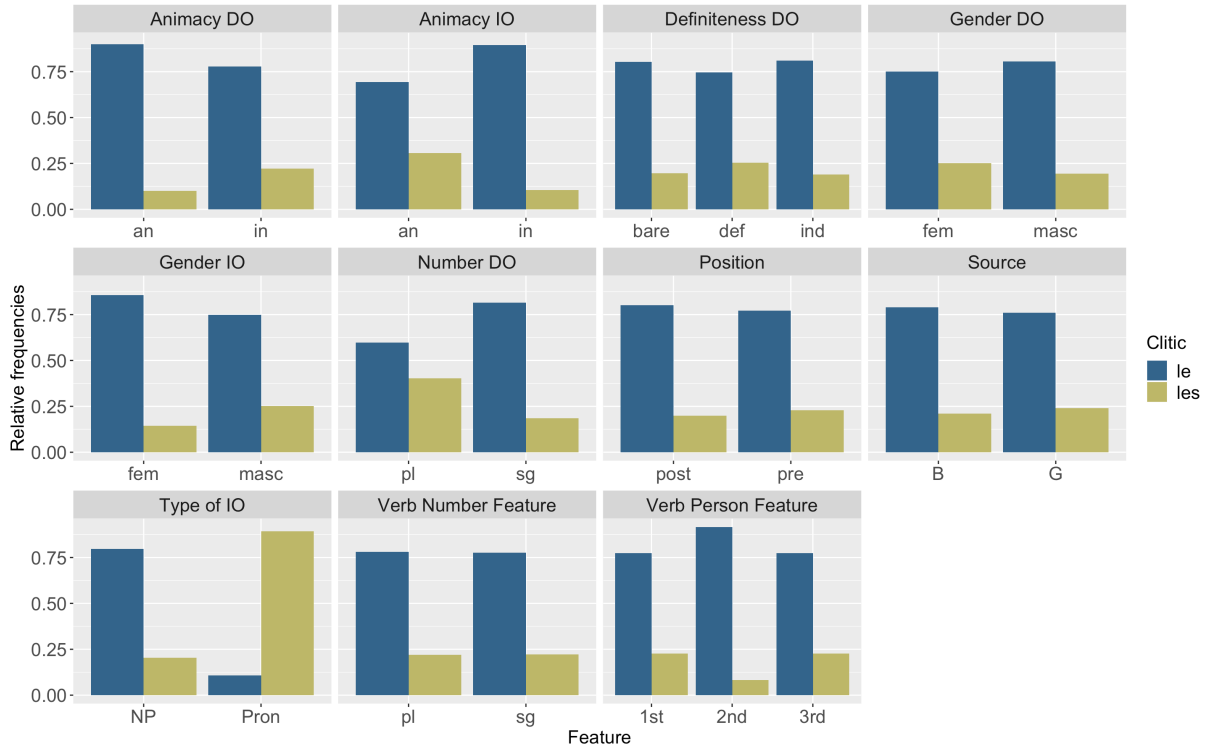
Figure 2. Relative frequency of *le* and *les* by variable.

In short, the descriptive results show that with the exception of pronouns, all other variables appear with the singular clitic much more often than with the plural clitic but there are some small differences across some levels of each variable. In the next section, I will study these differences in detail with the use of two modeling techniques: conditional inference tree and mixed-effects logistic regression to arrive at converging evidence about the factors that seem to drive the variation in number agreement of the dative clitic.

6.2   Modeling results

In this section, I will present the results of the two models that were used to analyze the data. The results will show that the two models identify the same key variables and interactions rendering the findings robust and reliable.

5.3   Conditional Inference Tree Model

The first step in the analysis was to construct a conditional inference tree without the random

variables ADDRESS and VERB. In a second step, a conditional inference tree was built with just

the variable COUNTRY to determine whether the different countries could be grouped somehow

to reduce the number of levels of this variable[4] and also to determine whether this phenomenon

is subject to significant dialectal variation.

The result of the first tree is shown in Figure 3 and the tree with only COUNTRY as variable

is shown in Figure 4. From the first tree (*C-index* = 0.77, *Balanced Accuracy* = 0.70, *Kappa* =

0.39, *F1* = 0.66) we can observe that the most important variable is animacy of the indirect

object, which interacts with two other predictors. On the one hand, if the IO is inanimate, then

the number features of the direct object come into play, such that the singular clitic is favored

with singular direct objects whereas the plural clitic is favored with plural direct objects. On

the other hand, if the IO is animate then, as we saw in the previous section, there is a clear split

between NPs and pronouns where pronouns almost categorically favor the plural clitic. If the

IO is realized as a full NP then other predictors come into play. The first of these predictors is

again the number features of the direct object and, as with inanimate IOs, plural DOs favor the

plural clitic. However, with singular DOs there appears to be a distinction between definite and

indefinite DOs on the one hand, and bare DOs on the other. Both definite and indefinite DOs

show a split depending on the number features of the verb. Interestingly, plural verbs favor the

singular clitic whereas singular verbs do not seem to favor either form. Bare DOs seem to be

sensitive to the gender of the DO; masculine bare DOs show a very strong preference for the

singular clitic where there is no preference for either clitic with feminine bare DOs.

With regards to the second tree, as is clear from Fig. 4, no significant differences or

groupings were found across the Spanish varieties so the algorithm simply produced the relative

---

[4] Both trees were built with random seed 123.

frequency of each level. This means that COUNTRY *per se* is not a good predictor for the alternation between the two forms of the clitic, suggesting that this phenomenon is not subject to (significant) dialectal variation (contra what has been claimed in the literature (e.g., Franco 1993, Torrego 1995))

In short, the conditional inference tree revealed the presence of significant interactions. On the one hand, the animacy features of the IO interact with the number features of the DO, and on the other hand, there are several nested interactions among the features of the DO for IOs realized as NPs. But as the goodness-of-fit measures above show, there is still room for improvement in the model as this conditional inference tree only reaches a C-index of 0.77. This may suggest that the random variables that were not included in the model, namely VERB and ADDRESS may help explain some of the variation that is still left unexplained. With this observation in mind and the results of the conditional inference tree, a generalized mixed-effects model was fitted with VERB and ADDRESS as random intercepts. Recall that ADDRESS is a proxy for speaker/ author as this information is not available in the version of the corpus used in this study.
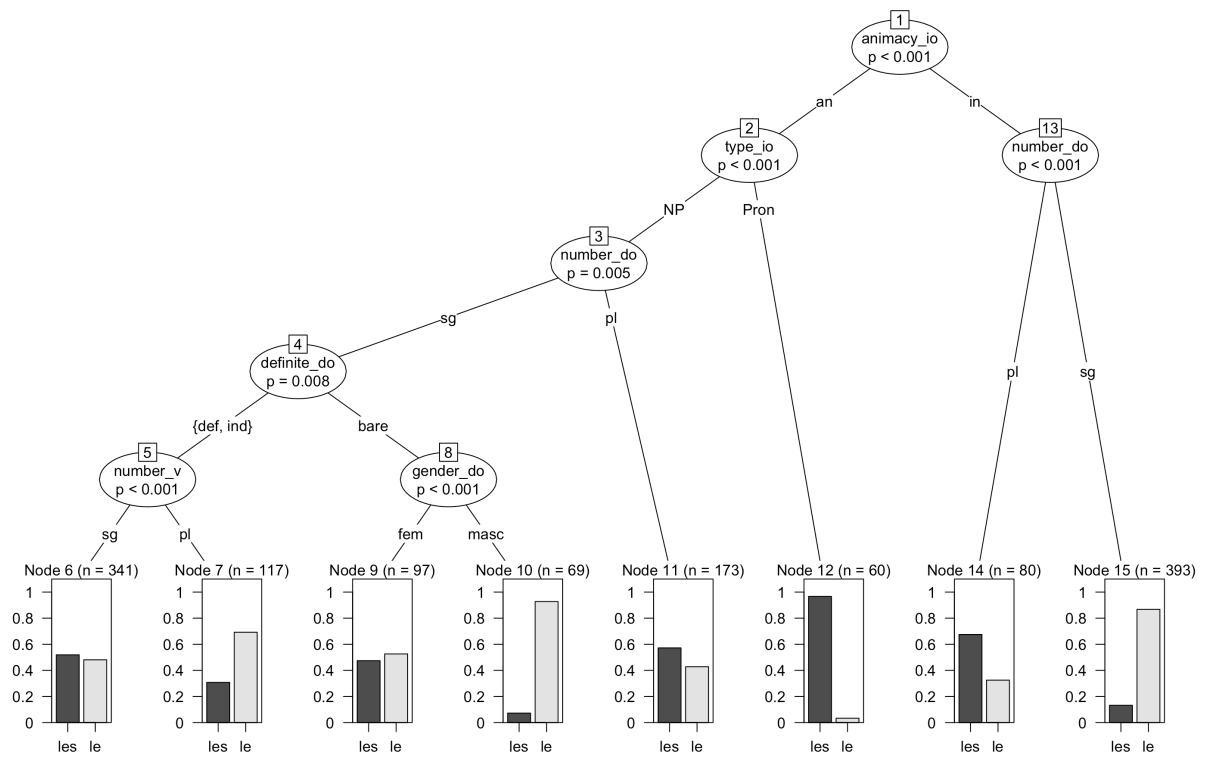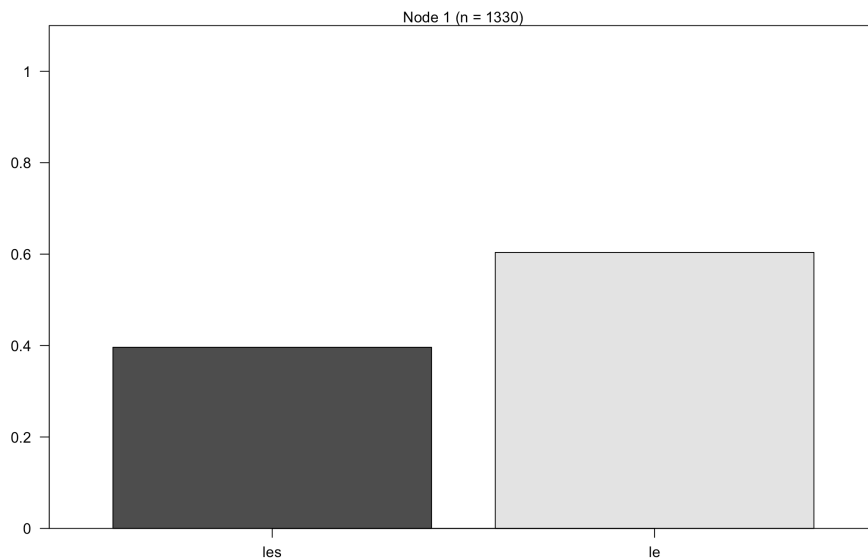
Figure 3. Conditional inference tree

Figure 4. Conditional inference tree with only COUNTRY as predictor.

5.4.   Generalized Mixed-Effects Regression Model

A full model was first fitted with all fixed-effects variables, with the interactions NUMBER_DO*ANIMACY_IO, DEFINITENESS_DO*GENDER_DO and DEFINITE_DO*NUMBER_V, and the two random effects. The model with all fixed-effects available was fitted in order to ensure that the variables that were not chosen by the conditional inference tree were in fact not important. It may be the case that once the random effects are included in the model, some of the "unimportant" variables in the conditional inference tree were found to help explain some of the variance left or that some of the variables are weaker predictors than those chosen in the conditional inference tree model, yet still make a significant contribution to the model. The predictor variables that were tested for significant effects in this step (i.e., those that were not chosen by the conditional inference tree) were: GENDER_IO, PERSON_V, POSITION and SOURCE. Log-likelihood ratio tests of nested models with and without each of the variables under consideration showed that indeed some of the variables contributed significantly to the model once the random effects were incorporated. The two significant variables identified were PERSON_V ($p = 0.002$) and GENDER_IO ($p = 0.003$). Neither POSITION ($p = 0.47$) nor SOURCE ($p = 0.27$) appeared to be statistically significant so they were removed from the model and were no longer considered.

The final model (*C-index* = 0.87, *Balanced Accuracy* = 0.80, *Kappa* = 0.59, *F1* = 0.74) includes the predictors PERSON_V, GENDER_IO and TYPE_IO as single terms and the three interaction terms NUMBER_DO*ANIMACY_IO, DEFINITENESS_DO*GENDER_DO and DEFINITE_DO*NUMBER_V. The first thing to note is that all goodness-of-fit measures have improved relative to the conditional inference tree, suggesting that the incorporation of the random effects and the single terms that do not appear in the conditional inference tree contribute significantly to explaining some of the variance in the data.

The coefficient estimates of the model are shown in Table 2. Note that the reference level for the dependent variable is the plural clitic *les*. The presentation of the results of the regression analysis will be based on effects displays from the *ggeffects* package (Lüdecke 2018) to make interpretation easier. Figures 5 shows the effects of the single terms and Figure 6 the effects of interaction terms.

Let us start by looking at the variance in the random effects shown in Table 1. We can observe that there are 1178 different web addresses and 178 different verbs. Although there are fewer verb types than there are addresses, there is much more variance among the verbs than among the web addresses (0.47 vs. 0.26). This suggests that individual verbs appear to behave slightly differently from one another. On the other hand, the lower variance in ADDRESS may indicate that inter-speaker variation is not very high, suggesting that this is a relatively stable phenomenon across speakers.

With respect to the fixed effects, let us first focus on the single terms in Figure 5. Plot A shows the effect of type of indirect object, which refers to whether the IO is realized as a full NP or a pronoun. As was discussed above and was clear from the descriptive results as well, the distinction between NPs and pronouns is nearly categorical such that the predicted probability of finding the singular clitic when the IO is a pronoun is close to 0 (0.03). On the other hand, when the IO is a full NP, then there is a 0.66 probability of finding the singular clitic. With respect to the person features of the verb, we can observe that the singular clitic is most likely to appear with second person (0.90). First and third person both have a predicted probability of around 0.60. The difference between masculine and feminine IOs is not very large but feminine IOs show a slight but significant preference for the singular clitic (feminine 0.71, masculine 0.58).

Moving on to the interaction terms, Plot A in Figure 6 shows the interaction NUMBER_DO*ANIMACY_IO. There is an interesting difference in how these two variables

interact in the choice of clitic. Overall, the singular clitic is much more likely to appear with singular DOs and this effect is stronger with inanimate IOs (predicted inanimate: 0.85 vs. animate: 0.57). However, if the DO is plural, the singular clitic is nearly twice as likely to appear with animate IOs than with inanimate IOs (predicted inanimate: 0.21 vs. animate: 0.40). Consequently, the singular clitic is most favored when the IO is inanimate and the DO is singular and it is least favored when the IO is inanimate and the DO is plural. Plot B shows the interaction DEFINITENESS_DO*GENDER_DO. There appears to be a very small difference between definite NPs across the two genders, such that the predicted probability of finding the singular clitic with a definite NP is around 0.60 for both genders. There is, however, a larger difference across the genders of the DO with bare NPs. This means that there is a higher probability of finding a singular clitic with bare masculine DOs than with bare feminine nouns (predicted masculine 0.74 vs. feminine 0.57). Interestingly, the relationship between definiteness and gender of the DO reverses with indefinite DOs. In this context, the singular clitic is much more likely to appear when the DO is indefinite and feminine than when it is indefinite and masculine (predicted feminine 0.74 vs. masculine 0.45).

Model Estimates									
	Estimate	**SE**	**z-value**	**p-value**			**Estimate**	**SE**	**z-value**	**p-value**
(Intercept)	0.62	0.33	1.88	0.06		Definiteness DO: Bare	-0.98	0.32	-3.04	< 0.01 **
Person: 1st	-0.22	0.23	-0.928	0.35		Definiteness DO: Indef	0.12	0.93	0.13	0.90
Person: 2nd	1.78	0.60	2.95	< 0.01 **	Number Verb: SG	-0.96	0.23	-4.12	< 0.001 ***
Type IO: Pron	-4.31	0.78	-5.529	< 0.001 ***	Number DO: SG*Animacy IO: Inanimate	2.32	0.40	5.80	< 0.001 ***
Gender IO: Masc	-0.54	0.19	-2.90	< 0.01 **	Gender DO: Masc*Definiteness DO: Bare	0.60	0.33	1.83	0.07
Number DO: SG	0.69	0.22	3.15	< 0.01 **	Gender DO: Masc*Definiteness DO: Indef	-1.42	0.70	-2.10	< 0.05 *
Animacy IO: Inanimate	-0.90	0.35	-2.61	< 0.01 **	Definiteness DO: Bare*Number Verb: SG	1.39	0.33	4.16	< 0.001 ***
Gender DO: Masc	0.17	0.21	0.82	0.41	Definiteness DO: Indef*Number Verb: SG	0.91	0.75	1.21	0.23

	Random Effects			
Groups	**Name**	**Variance**	**Std.Dev**	**N**
Address	(Intercept)	0.26	0.51	1178
Verb	(Intercept)	0.47	0.68	178

Table 2. Coefficient estimates, standard errors and *p*-values for the mixed-effects logistic regression model.
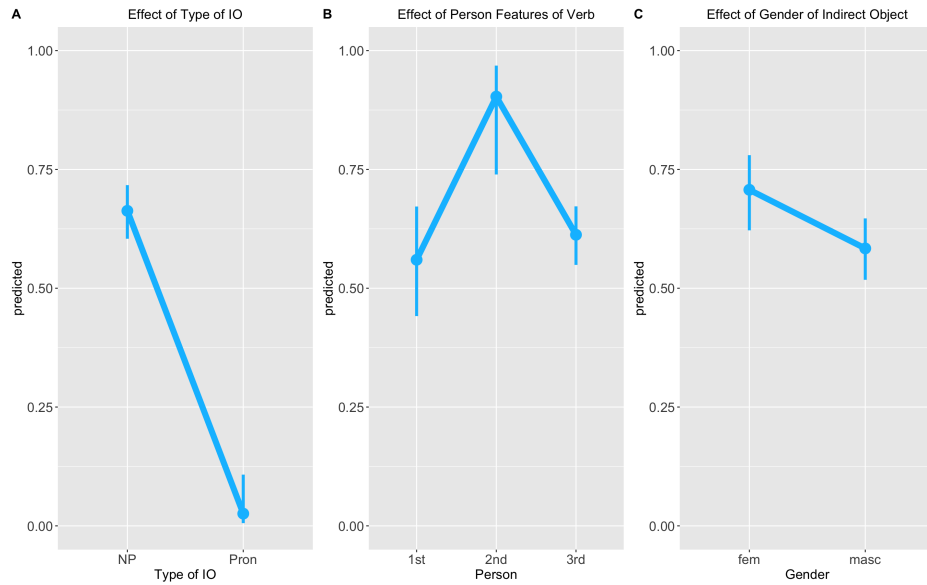
Figure 5. Effects displays showing the predicted probability of defective agreement for the single terms in the generalized mixed-effects regression model.
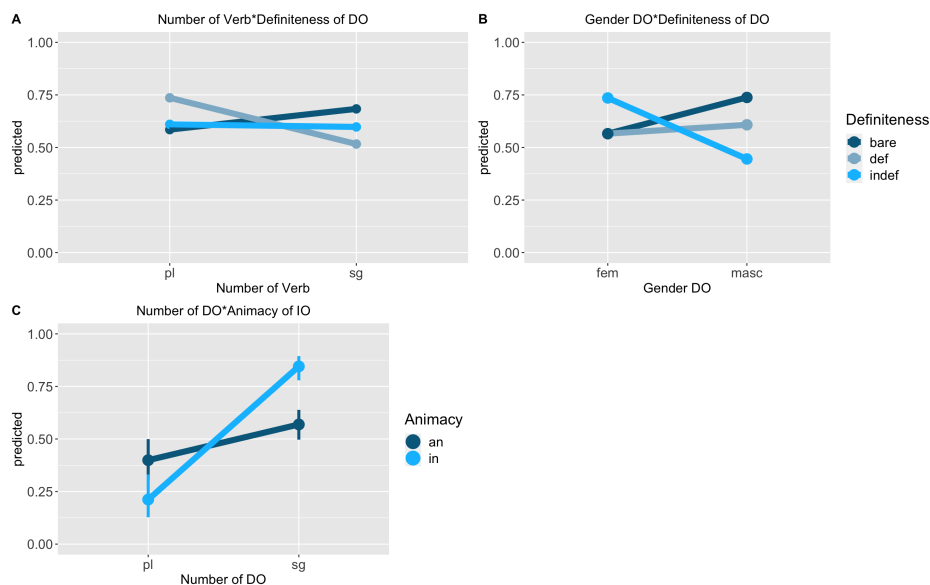

Figure 6. Effects displays showing the predicted probability of defective agreement for the interaction terms in the mixed-effects logistic regression model.

6. Discussion

In this section, I will first revise the hypotheses of the study and then focus the discussion on the two main results that we have learned from the models, namely the interaction between the

number of the direct object and animacy of the indirect object and the nearly categorical split between pronouns and lexical NPs. I will argue that these results provide evidence for an analysis of clitics as object agreement markers as opposed to pronominal elements.

6.1 Hypotheses

In Section 1, I laid out the main hypotheses, namely that (i) defective agreement would be more likely with singular DOs because DOs are linearly closer to the clitic and that (ii) defective agreement would be triggered by animacy features of both objects assuming (i) was correct. As the reader must already have noticed, the first hypothesis was borne out in that singular DOs do favor defective agreement. Notice that although this feature interacts with the animacy of the IO, whenever the DO is singular, defective agreement is the preferred option. This brings us to our second part of the hypothesis about the animacy feature of both objects. As it turns out, it is unlikely that animacy of the DO may play a role because most DOs tend to be inanimate and this forced us to remove the variable from further analysis. However, the animacy features of the IO do appear to play a significant role but only in conjunction with singular DOs. Inanimate IOs that appear with plural DOs do not favor defective agreement, which suggests that the number features of the DO outrank the animacy features of the IO. Therefore, the second part of the hypothesis was only partially borne out.

6.2. Attraction effects of the Direct Object

The two models show that the number of the direct object is a very strong predictor for defective agreement. Plot C in Figure 6 shows that, although there is a clear interaction between number of the DO and animacy of the IO, for both values of the animacy feature of the IO a singular DO results in the highest predicted probability of defective agreement. This result can be interpreted as the presence of an attraction effect by the DO in the computation of the

dependency between the clitic and its coreferential NP, the DO being linearly closer to the clitic than the IO. I will argue that this type of attraction effect allows us to probe into the syntactic structure of clitics but first I will highlight some important peculiarities about the attraction effect instantiated by the DO.

The type of attraction effect responsible for defective agreement appears to challenge the general claim that attraction effects are more likely with singular agreement both in production and comprehension (Bock and Miller 1991, Bock and Cutting 1992, Bock and Eberhard 1993, Eberhard 1997, Nicol et al. 1997, Wagers et al 2009). This claim would predict that we should find sentences where a plural clitic is found with a singular correfering IO due to the presence of an intervening plural DO such as in *les di [los caramelos] al niño* "I gave the sweets to the boy", where the DO *los caramelos* intervenes between the clitic and the singular IO. However, this type of agreement mismatch is unattested in Spanish. One possible reason that it is the unpredicted pattern that is found with the dative clitic may be that this is simply not an example of an attraction affect. In other words, is it true that the linear distance between the clitic and the DO predicts the realization of defective agreement? Luckily, Spanish allows us to answer this question by switching the order of the objects as in (17), where the clitic is linearly closer to the IO than to the DO (i.e., IO_DO word order).

17. [...] les ahorra a los trabajadores el viaje [...]
   CL.DAT save.3s.pres to the.PL workers the trip
   'It saves workers the trip'

Thus, we can directly test whether there is an intervention effect between the DO and the clitic by comparing the realization of defective agreement in the two word orders, namely DO_IO and IO_DO. The prediction is that if defective agreement is an intervention effect phenomenon, then the order of objects should matter such that defective agreement should be significantly more likely with singular DOs in the DO_IO word order. To test this prediction, I extracted the more marked order CLITIC IO DO from the same corpus and with the same procedure used

with the main model (i.e., correcting for class imbalance), I built a mixed-effects generalized logistic regression model with the interaction WORD ORDER*NUMBER DO as the predictor variable and CLITIC as the dependent variable. The random factors were VERB and ADDRESS. In addition, DEFINITENESS DO and POSITION were added to the model in order to control for these features as there were very few bare DOs and postverbal clitics in the marked word order.

The results of the model (*C-index*= 0.82, *Kappa* = 0.41, *F1*= 0.61) are shown in Table 3 and the effects display in Figure 7[5]. The interaction between word order and the number feature of the DO is highly significant ($p < 0.01$), such that when the DO is singular and it appears adjacent to the clitic (i.e., DO_IO) the expected probability of defective agreement is 0.70. This contrasts with the marked word order (i.e., IO_DO) where there is a significant drop to 0.30 in the probability of defective agreement. When the DO is plural, word order is not a significant predictor of defective agreement ($p = 0.91$, predicted probability = 0.40 for both word orders). These results are compatible with, and provide further support for, the interpretation of defective agreement as an attraction effect exerted by the DO on the clitic. This may be the first reported case of an attraction effect triggered by a singular attractor.

Model Estimates					
	Estimate	SE	z-value	p-value	
(Intercept)	-0.70	0.19	-3.70	< 0.001	***
Word Order: IO_DO	0.06	0.52	0.11	0.91	
Number DO: SG	1.17	0.17	6.80	< 0.001	***
Word Order: IO DO* Number DO: SG	-1.62	0.56	-2.91	< 0.01	**

Random Effects				
Groups	Name	Variance	SD	N
Address	(Intercept)	0.19	0.44	1528
Verb	(Intercept)	0.62	0.79	213

Table 3. Coefficients, standard errors and *p*-values for the mixed-effects logistic regression model with WORD ORDER*NUMBER_DO as predictor variable.

---

[5] For ease of exposition I have not included the controlling variables as they are not relevant for the analysis. A complete table with coefficient estimates can be found in the Appendix.
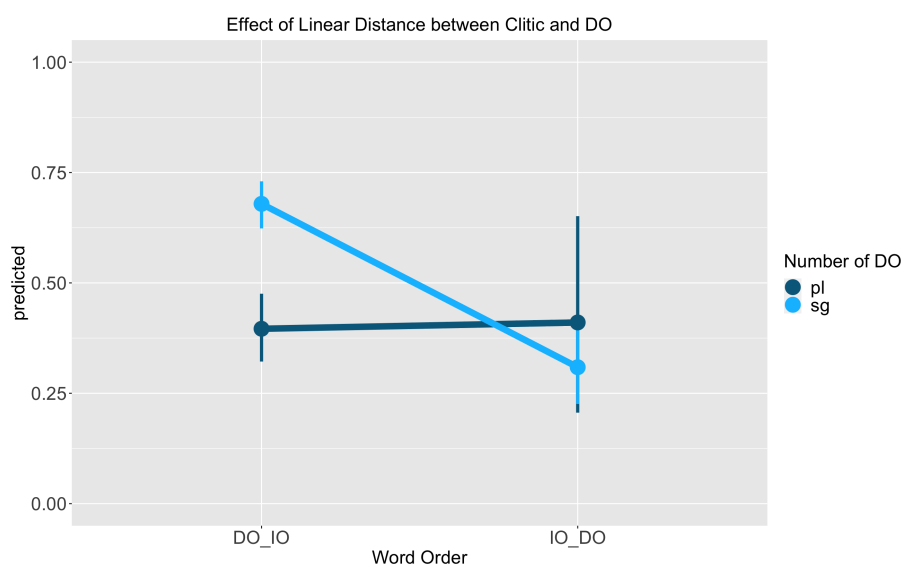
Figure 7. Predicted probability of defective agreement as a function of the linear distance between the clitic and the direct or indirect object.

6.3. Defective Agreement and the Syntax of Clitics

Having established that defective agreement is an attraction effect phenomenon, I will now discuss how this finding can shed light on the syntax of clitics and of agreement more generally. In particular, defective agreement can be interpreted in two ways: (i) the clitic is based generated in its surface position or (ii) agreement must be post-syntactic.

Given the findings just presented, it is clear that surface word order is relevant in determining whether defective agreement might occur. I think that the fact that defective agreement is a surface-word-order phenomenon can be used to inform the long-standing debate of whether clitics move to their surface position or are based generated there. A movement approach to clitics is not compatible with a language with defective agreement unless other assumptions or conditions are added to the analysis, namely where agreement takes place. Assuming the syntactic structure in (18) (López 2009) and that agreement takes place in syntax proper pre-movement, the clitic originates in *v* and then moves to a higher functional projection.

Note that according to (18) the agreement dependency between the clitic and the IO must be established via *upward* Agree where the goal c-commands the probe (Wurmbrand 2011, Zeijlstra 2012). In this syntactic structure, it is not clear how attraction between the clitic and the DO would take place unless it were assumed that the clitic could probe both upward and downward. Although there are cases where both upward and downward Agree are possible within the same language, each option occurs under different structural configurations and the choice between one or the other is not optional. For example, in Standard Arabic subjects can appear in pre- or post-verbal position and each position displays a different type of Agree; upward Agree with preverbal subjects and downward Agree with postverbal ones (Bjorkman and Zeijlstra 2019). Thus, if a probe in a specific configuration probes upward, it is not theoretically principled that it may optionally probe down in the same exact configuration. Therefore, the structure in (18) and the conditions on upward Agree predict that there should be no attraction effects because the conditions for it are never met in the syntactic representation before movement. In other words, in order for Agree to be computed between the clitic and the DO, the DO would have to c-command the clitic (Zeijsltra 2012). As is clear from (18), the DO is in the wrong position for Agree to take place between these two elements. Thus, in the syntactic structure in (18) the conditions for attraction effects do not seem to be met nor are they predicted to be possible.

18.
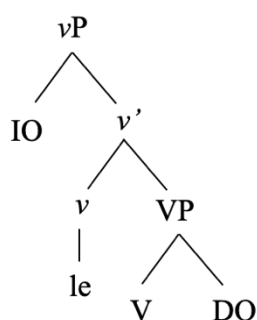
As I mentioned above, there are two possible solutions for the problem just described. Perhaps the most obvious solution is to say that the clitic is base-generated in its surface position (and therefore (18) is not the correct syntactic representation for clitic doubling) and so the attraction effect falls out from the linear order of constituents and no other assumptions need to be made; the clitic is linearly closer to the DO than to the IO in the canonical word order where the DO precedes the IO. Another, albeit more elaborate solution, would be to say that Agree is post-syntactic (Bobaljik 2008), that is, agreement would take place after/ during linearization of hierarchical structure à la Distributed Morphology (Halle and Marantz 1993). Both accounts would predict the existence of attraction effects but the proposal that Agree may take place post-syntactically might be more attractive given the body of evidence showing that clitics do move (e.g., clitic climbing, extraction effects, intervention effects, etc). Therefore, putting together the evidence that clitics move and the existence of defective agreement in the same language we can conclude that agreement between the clitic and the doubled NP must be computed post-syntactically.

6.4. Constraints on Defective Agreement: What do they Tell us?

One question that follows from these results is why defective agreement is possible in certain structural contexts but not in others. Recall that defective agreement is not possible with preverbal IOs and when the IO is a strong pronoun. There is an interesting asymmetry between DO clitic doubling and defective agreement that may shed light on this issue: those contexts in which clitic doubling of the accusative clitic *is possible* (i.e., with strong pronouns and preverbal DOs) are the same contexts in which defective agreement *is not possible*. Is there a unifying explanation for why the structures that allow accusative clitic doubling coincide with those that disallow defective agreement in the dative clitic? I will attempt to provide one such

explanation in what follows and argue that this can be accounted for by the different categorial status of the clitic across structures.

In Section 3, I presented the theoretical analyses of DO clitic doubling and of the status of the doubled nominal phrase, be it a strong pronoun or a preverbal full NP. In a nutshell, we saw that the doubled element was considered to be in a non-argumental position in both cases (right-dislocated for strong pronouns vs. left-dislocated for preverbal NPs). Therefore, in both of these instances the clitic is fulfilling the selectional requirements of the verb to have an object of the right kind (i.e., direct or indirect). The fact that in these two contexts defective agreement cannot take place suggests that this phenomenon is sensitive to the structural position the clitic is in. In other words, defective agreement appears not to be possible when the dative clitic is the true argument of the verb. Recall also that defective agreement is banned when there is no doubled NP, which is another instance of the clitic as the true argument of the verb. The generalization that seems to emerge, then, is that defective agreement and doubling of the accusative clitic are morphosyntactic processes that are sensitive to the syntactic role of the clitic in the sentence; when the clitic is an argument of the verb an accusative clitic is obligatory and full agreement of the dative clitic must take place. Regardless of the analysis of what occupies the argument position of the verb, whether it is filled with *pro* or with a silent copy of the clitic, this generalization can be restated in the following way:

i.   *Defective agreement cannot occur if the NP coreferential with the dative clitic is not phonologically overt in an argument position.*

This licensing requirement for the argument position to be filled with an overt NP can be interpreted in terms of recoverability (Chomsky 1964), which states that only recoverable deletions are permitted in the grammar. In particular, the number feature may be left unspecified on the dative clitic as long as this information can be recovered somewhere else in the syntactic structure, in this case from the coreferential NP in the argument position of the verb.

Evidence that the NP in the contexts in which defective agreement is possible is in an argument position comes from the fact that, although the clitic-doubled alternative is always preferred, the non-doubled version is grammatical, albeit more formal as shown in (19), where the clitic *le* is optional though highly preferred.

19. (Le)   donaron   leche   a   las   escuelas.
     CL.DAT   donate.3P.PAST   milk   to   the   schools
     'They donated milk to the schools'

Another piece of evidence that defective agreement is sensitive to the argumental status of the clitic comes from raising predicates. With this type of predicate, the dative clitic is obligatory (Fontana 1993) suggesting that the clitic and not the NP in IO position is an argument of the verb. As expected, defective agreement with raising predicates is highly dispreferred. In the corpus, the raising predicate *parecer* 'to seem' appears only 17% of the time with defective agreement, a very large difference when compared with the nearly 80% of cases with non-raising predicates.

The fact that clitic doubling of the direct object and defective agreement in the indirect object are processes that are sensitive to the structural configuration of the sentence would seem to indicate that clitics may have more than one syntactic role depending on the type of construction in which they appear. My proposal is that in Spanish clitics can be pronominal clitics or agreement markers but this is dependent on the structure in which they are found. In cases of true clitic doubling, that is, when the clitic is not in an argument position, then the clitic is an agreement marker. However, in cases where the clitic is not truly doubling an argument but it is the argument itself then the clitic is pronominal. Thus, analyses where clitics are analyzed as either object markers or agreement morphemes *across the board*, regardless of the specific structure in which they appear, cannot be maintained; not all instances of the dative clitic are of the same syntactic type and this is why a seemingly optional phenomenon such as

defective agreement turns out to be constrained by the structural configuration in which the clitic appears.

The proposal that defective agreement is a reflex of the status of the clitic as an agreement-marker is supported by well-attested processes of grammaticalization. On the one hand, in diachrony pronouns develop into clitics and clitics develop into agreement morphemes as in the grammaticalization path in (20) (e.g., Lehmann 1985, 2002; Hopper and Traugott 1993). Moreover, this process is accompanied by phonological attrition (i.e., the gradual loss of phonological material) and semantic bleaching (i.e., loss of semantic material) (Hopper and Traugott 1993, Fuß 2005). Phonological attrition is characterized by the loss of segments or of marked phonological features and loss of prosodic independence. In our case, the loss of /s/ signaling plural marking follows the expected grammaticalization path where the clitic undergoes both phonological attrition (loss of a phoneme) and semantic bleaching (loss of plural marking). Therefore, it would appear that *le* has grammaticalized as an agreement marker of the dative NP lacking both gender and number features.

20. Independent pronoun > weak pronoun > clitic pronoun > affixal agreement marker > fused agreement marker > Ø

It must be acknowledged that an attraction effect account cannot capture the extent to which defective agreement is found in contemporary Spanish. After all, nearly 80% of the data shows defective agreement, so defective agreement is the norm not the exception. From a diachronic perspective, it seems logical to assume, based on the results of the models, that it was the attraction effect of the DO that triggered the grammaticalization of the dative clitic as an unmarked agreement marker. The linguistic change now seems to be quite advanced and, except for the cases in which defective agreement is banned, the variation we find in the data

may simply be a byproduct of language standardization (e.g., Rothman 2007, Kato et al. 2009). I leave the study of the diachronic development of defective agreement for future research.

6.5.   Predictions and Future Research

The claim that the syntactic properties of the clitic may be dependent on the structural context in which it appears makes clear predictions that can be empirically tested. For example, there is evidence that dependencies established by antecedent-pronoun relations are processed differently from dependencies established in agreement-type relations such as subject-verb agreement. In light of this, in those contexts in which I have argued that the clitic is an argument of the verb, we should expect that it will elicit responses akin to pronominal dependencies whereas in contexts where the clitic is not an argument of the verb (i.e., in cases where defective agreement is possible) we should expect responses suggesting the formation of agreement relations similar to those elicited by subject-verb agreement. The first part of this prediction is borne out in Santesteban et al. (2017). In their study, they look at attraction effects between a left-dislocated antecedent NP and a following clitic in sentences such as (15) where they manipulated the number of both the object and the clitic (in bold).

15.   El   cartero   afirmó   que   **el**   **paquete**   para   el
       The   mailman   state.3s.past that   the   package   for   the

       vecino   **lo**   entregó   a   tiempo.
       neighbor   it   deliver.3s.past to   time.
       'The mailman stated that he delivered the package on time'

Their goal was to investigate what type of dependency would be established between the clitic and its antecedent: (a) agreement or (b) pronominal dependency. I will focus on the results of the two online tasks they conducted, namely a self-paced reading task and an ERP task but participants were also asked for acceptability ratings of the stimuli.

Unlike subject-verb agreement, which has been shown to be sensitive to attraction effects in self-paced reading tasks, the antecedent-clitic relation in their study is not sensitive to

attraction effects in this task. This is more in line with other studies that found that the relation between a reflexive pronoun and its antecedent is not sensitive to attraction effects (Dillon et al. 2013; Parker and Phillips 2017 but see Jäger et al. 2020 for a different result). Likewise, they observe that subject-verb agreement resolution in attraction effects has been shown to elicit a late positive P600 component in ERP tasks (Osterhout et al. 1996, Nevins et al. 2007, Frenck-Mestre et al. 2008) and the absence of a frontal negativity. However, their results show that the antecedent-clitic dependency mainly affected early frontal *negative* ERP components. Altogether, they interpret these results to suggest that the dependency established by the antecedent-clitic relation and subject-verb agreement represent different types of linguistic dependencies so their results favor an analysis of clitics as pronouns. They do note, however, that further research needs to be conducted in different structural contexts to arrive at a better understanding of the nature of Spanish clitics. Importantly, these results are predicted by the analysis herein where clitics with left-dislocated antecedents are claimed to occupy an argumental position and therefore should behave like true pronouns. Whether the second part of the prediction is true, namely that the clitic is not pronominal in the contexts where defective agreement is allowed, such that the relationship between it and its agreeing NP should mirror that of subject-agreement dependencies, is an empirical question that has yet to be pursued.

7. Conclusion

In this paper we looked at defective agreement of the third person dative clitic in Spanish clitic doubling. I have shown that defective agreement is a pervasive phenomenon in all Spanish varieties such that when the structural conditions for it are met, defective agreement is the norm, not the exception, across all Spanish dialects. Using conditional inference trees and mixed-effects logistic regression models we saw that linguistic features of both the direct and indirect objects play a role in the realization of defective agreement. The most important of these are animacy of the indirect object and number of the direct object: inanimate IOs coupled with

singular DOs provide the context in which defective agreement is most likely. In addition, I have argued that the existence of defective agreement can only be predicted under two competing scenarios: (i) the clitic is base-generated or (ii) agreement takes place in the post-syntactic component. Moreover, I have proposed that the structural constraints on defective agreement may provide indirect evidence for the categorial status of the clitic as defective agreement is only possible when the clitic is in a non-argumental position and it is banned elsewhere. Consequently, this appears to imply that the categorial status of the clitic is dependent upon the structural configuration in which it appears, suggesting that an analysis where the dative clitic is considered an agreement morpheme across the board cannot be maintained.

References

Adger, D. (2003) *Core Syntax*. Oxford: Oxford University Press

Alpaydin, E. (2020). *Introduction to machine learning*. Cambridge, MA: MIT press.

Anagnostopoulou, E. (2006). Clitic doubling. In Martin Everaert and Henk van Riemsdijk
(eds.), *The Blackwell companion to syntax 1*. 519–581. Oxford: Blackwell Publishers.
DOI: https://doi.org/10.1002/9780470996591.ch14

Anagnostopoulou, E. (2017). Clitic Doubling. In Martin Everaert and Henk van Riemsdijk
(eds.). *The Blackwell Companion to Syntax*, 2nd edition. Oxford: Blackwell
Publishers.

Aoun, J. (1981), The Formal Nature of Anaphoric Relations, MIT PhD dissertation.

Belloro, V. A. (2007). Spanish clitic doubling: A study of the syntax-pragmatics interface.
Buffalo, NY: State University of New York dissertation.

Becerra Bascuñán, S. (2006). Estudio diacrónico y sincrónico del objeto indirecto en el
español peninsular y de América. *Etudes Romanes 57*, Copenhagen: Museum
Tusculanum.

Bentivoglio, P. (1978). Formación de clíticos: análisis sobre el habla culta de Caracas.
In H. López Morales (ed), *Corrientes actuales en la dialectología del Caribe Hispánico*.
Puerto Rico, Editorial Universitaria: 13-29.

Bobaljik, J. D. (2008). Where's phi? Agreement as a post-syntactic operation. In Harbour, D.,
Adger, D. and Béjar, S., (eds.), *Phi-Theory: Phi features across interfaces and
modules*, 295–328. Oxford: Oxford University Press.

Bock, K. and Miller, C. (1991). Broken agreement. *Cognitive Psychology*. 23: 35–43.
doi: 10.1016/0010-0285(91)90003-7

Bjorkman, B. and Zeijlstra, H. (2019). Checking up on (Phi-) Agree. *Linguistic Inquiry* 50:
527-569.

Bock, K. and Cutting, J. C. (1992). Regulating mental energy: performance units in language production. *Journal of Memory and Language*. 31, 99–127. doi: 10.1016/0749-596X(92)90007-K

Bock, K. and Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, *8*(1): 57-99.

Bock, K., Nicol, J., and Cutting, J. C. (1999). The ties that bind: creating number agreement in speech. *Journal of Memory and Language*. 40: 330–346. doi: 10.1006/jmla.1998.2616

Bock, J. K., Eberhard, K. M., and Cutting, J. C. (2004). Producing number agreement: how pronouns equal verbs. *Journal of Memory and Language*. 51: 251–278. doi: 10.1016/j.jml.2004.04.005

Borer, H. (1984). *Parametric Syntax*. Dordrecht: Foris Publications.

Chawla, N.V., Japkowicz, N. and Kotcz, A. (2004). Special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter*, *6*(1), pp.1-6.

Chomsky, N. (1964). Current issues in linguistic theory. *The structure of language: readings in the philosophy of language*, in Fodor, J.A. and Katz, J.J., (eds), 50-118. Princeton, New Jersey: Englewood Cliffs.

Company Company C. (2006). El objeto indirecto. In: Company Company C (ed.) *Sintaxis histórica de la lengua española*. Mexico City: Universidad Nacional Autónoma de México, Fondo de Cultura Económica, 477–572.

Corbett, G. (2006). Agreement. Cambridge: Cambridge University Press.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press on Demand.

De Mello, G. (1992). Se los for se lo in the spoken cultured Spanish of eleven cities. *Hispanic Journal*, 13 (1):165-179.

Demonte, V. (1994). La ditransitividad en español. *Gramática del español*. México, El

    Colegio de México: 431-470.

Dillon, B., Mishler, A., Slogett, S., and Phillips, C. (2013). Contrasting intrusion profiles for

    agreement and anaphora: experimental and modeling evidence. Journal of Memory

    and Language. 69: 85–103. doi: 10.1016/j.jml.2013. 04.003

Dufter, A. and Stark, E. (2008): Double indirect object marking in Spanish and Italian, in

    Elena Seoane y María José López-Couso (eds.), *Theoretical and Empirical Issues in*

    Grammaticalization, Amsterdam, John Benjamins, 111-129.

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of*

    Memory and Language, 36, 147–164.

Escobar, A. (1978). *Variaciones Sociolingüísticas del Castellano en el Perú*. Instituto de

    Estudios Peruano, Lima.

Escobar A. (2000). *Contacto Social y Lingüístico*. Pontífica Universidad Católica del Perú,

    Lima.

Fernández Soriano, O. (1989). *Rección y ligamiento en el español: Aspectos del*

parámetro de sujeto nulo. PhD Dissertation, Universidad Autónoma de Madrid, Spain.

Fontana, J. M. (1993). *Phrase structure and the syntax of clitics in the history of Spanish*.

    PhD Dissertation, University of Pennsylvania, Philadelphia.

 Franco, J. (1993). *On Object Agreement in Spanish*. PhD Dissertation, USC, Los Angeles.

Frenck-Mestre, C., McLaughlin, J., Osterhout, L., and Foucart, A. (2008). The effect of

    phonological realization of inflectional morphology on verbal agreement in French:

    evidence from ERPs. *Acta Psychologica*. 128, 528–536. doi: 10.1016/j.actpsy.

    2007.12.007

Fuß, Eric (2005). *The rise of agreement: A formal approach to the syntax and*

    grammaticalization of verbal inflection. Amsterdam: John Benjamins.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language.* Oxford University Press on Demand.

Gries, S. T. and Stefanowitsch, A. (eds.). (2007). *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis* (Vol. 172). Berlin: Walter de Gruyter.

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Hale, K. and Jay Keyser, S., (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.

Harizanov, B. (2014). Clitic doubling at the syntax-morphology interface. A-movement and morphological merger in Bulgarian. In *Natural Language and Linguistic Theory, 32:1033–1088*

He, H. and Garcia, E.A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, 9:1263-1284.

Hentschel, B. (2013). La duplicación pronominal del objeto indirecto en español. *Romanische Forschungen*, *125*(3): 313-330.

Hopper, P. J. and Traugott, E.C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A. and Van Der Laan, M.J., 2005. Survival ensembles. *Biostatistics*, *7*(3), pp.355-373.

Jaeggli, O. (1982). *Topics in Romance syntax*. Dordrecht: Foris Publications.

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, *111*.

Kany,C. (1951). *American Spanish Syntax*. Chicago: Chicago University Press.

Kato, M. A., S. M. L. Cyrino and V. R. Corrêa (2009). Brazilian Portuguese and the recovery of lost clitics through schooling. In A. Pires and J.Rothman (eds) *Minimalist inquiries*

into child and adult language acquisition: Case Studies across Portuguese, 245-272. Berlin/New York: Mouton DeGruyter.

Kayne, R. S. (1975). *French syntax: The transformational cycle* (Vol. 30). Cambridge, MA: MIT press.

Kayne. R. (1990). Null subjects and clitic climbing. In O. Jaeggli and K. Sar (eds) *The null subject parameter.* Dordrecht: Kluwer

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4): 221-232.

Lakoff, G. (1990). Cognitive versus generative linguistics: How commitments influence results. *Language and Communication*, *1*(1).

Lehmann, C. (1985). Grammaticalization: Synchronic variation and linguistic change. *Lingua e Stile*, 20: 303–310.

Lehmann, C. (2002) *Thoughts on Grammaticalization*. 2nd, revised edition. Erfurt: Seminar für Sprachwissenschaft der Universität.

López, L. (2009). *A derivational syntax for information structure*. Oxford: Oxford University Press.

López, V., Fernández, A., García, S., Palade, V. and Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information sciences*, *250*, pp.113-141.

Lunardon, N., Menardi, G. and Torelli, N. (2014). ROSE: a Package for Binary Imbalanced Learning. *R Journal*, 6(1): 82-92.

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, 3(26), 772. doi: 10.21105/joss.00772.

Mitchell, T. M. (2017). *Machine Learning* (chapter 14 draft for second edition). New York: McGraw-Hill Education.

Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature number and feature-type in processing Hindi verb agreement violations. *Brain Research*. 1164, 81–94. doi: 10.1016/j.brainres.2007.05.058

Nicol, J. and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistics*. Res. 18: 5–19. doi: 10.1007/BF01069043

Nicol, J. L., Forster, K. I., and Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal Memory and Language*. 36: 569–587. doi: 10.1006/jmla.1996.2497

Ormazabal, J. and Romero, J. (2010). Object clitics and agreement. Ms., University of the Basque Country/HiTT and University de Extremadura/HiTT.

Osterhout, L., McKinnon, R., Bersick, M., and Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: is the syntactic positive shift a member of the P300 family? Journal of Cognitive Neuroscience. 8, 507–526. doi: 10.1162/jocn.1996.8.6.507

Parker, D. and Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language*. 94: 272–290. doi: 10.1016/j.jml.2017.01.002

Patil, U., Vasishth, S., and Lewis, R. L. (2016). Retrieval interference in syntactic processing: the case of reflexive binding in English. *Frontiers in Psychology*. 7:329. doi: 10.3389/fpsyg.2016.00329

Paus, C. (1990). *"Subject and Direct Object in Transitive Verb Agreement: A typological survey"*. Ms, University of Southern California.

Picallo, M. C. (1991). Nominals and Nominalizations in Catalan, *Probus* 3: 279- 316.

Provost, F. (2000). July. Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI'2000 workshop on imbalanced data sets* (Vol. 68, No. 2000, pp. 1-3). AAAI Press.

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Raeder, T., Forman, G. and Chawla, N.V. (2012). Learning from imbalanced data: Evaluation matters. In *Data mining: Foundations and intelligent paradigms* (pp. 315-331). Springer, Berlin, Heidelberg.

Rigau, G. (1988). Strong Pronouns. *Linguistic Inquiry* 19.3: 503-511.

Rivas, A. (1977). *A Theory of Clitics*, PhD Dissertation, MIT.

Rizzi, L. (1976). Ristrutturazione. *Rivista de Grammatica Generativa* 1: 1-54.

Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris Publications.

Rizzi, Luigi. (1986). On chain formation. In *The Syntax of Pronominal Clitics*, in Borer H., vol. 19 of *Syntax and Semantics*, 65–95. New York: Academic Press.

Roca, F. (1992). On the Licensing of Pronominal Clitics: the properties of object clitics in Spanish and Catalan. Unpublished manuscript, Universitat Autònoma de Barcelona.

Rothman, J. (2007) Heritage speaker competence differences, language change and input type: inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism*, 11 (4): 359-389.

Saltarelli, Mario (1987). The Acquisition of Agreement. In Bahner, W. et al, (eds). *Proceedings of the Fourteenth International Congress of Linguistics*, 810-815. Berlin: Akademie-Verlag

Santesteban M., Zawiszewski A., Erdocia K. and Laka I. (2017). On the Nature of Clitics and Their Sensitivity to Number Attraction Effects. *Frontiers in Psychology*. 8:1470. doi: 10.3389/fpsyg.2017.01470

Silva-Corvalán, C. (1981). The Diffusion of Object-Verb Agreement in Spanish. *Papers in Romance (3)*: 163-176.

Sportiche, D. (1989) "Le Mouvement Syntaxique: Constraintes et Parametres", *Langages* 95, 35-80.

Sportiche, D. (1990) "Movement, Agreement and Case", Ms. UCLA.

Sportiche, D. (1995). Clitic Constructions. Phrase Structure and the Lexicon. In Zaring, I. and

Rooryck, J. (eds). Dordercht: Kluwer Academic Publishers: 213-276.

Strozer, J. (1976). *Clitics in Spanish*, PhD Dissertation, UCLA.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*. 48, 542–562. doi: 10.1016/S0749-596X(02) 00536-3

Suñer, M. (1988). The Role of Agreement in Clitic-Doubled Constructions. *Natural Language and Linguistic Theory* 6: 391-434.

Torrego, E. (1995). On the nature of clitic doubling. In H. Campos and P. Kempchinsky (eds) *Evolution and Revolution in Linguistic Theory,* 399-418. Washington DC: Georgetown University Press.

Rozas, V. V., and Salido, M. G. (2012). A discourse-based analysis of object clitic doubling in Spanish. *Grammaticalization and Language Change: New Reflections*, *130*, 271.

Wagers, M. W., Lau, E. F., and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *Journal of Memory and Language*. 61, 206–237. doi: 10.1016/j.jml.2009.04.002

Wallace, B. C. and Dahabreh, I. J. (2012). Class probability estimates are unreliable for imbalanced data (and how to fix them). In *Institute of Electrical and Electronics Engineers (IEEE) 12th International Conference on Data Mining (International Conference on Data Mining)* 695–704. IEEE Computer Society, Washington, DC.

Weiss, G. M. and Provost, F. (2003). Learning when training data are costly: The effect of class distribution on tree induction. *Journal of artificial intelligence research*, *19*, 315-354.

Wurmbrand, S. (2001*). Infinitives: restructuring and clause structure*. New York: Mouton de Gruyter

Zdrojewski, P., & Sánchez, L. (2014). Variation in accusative clitic doubling across three Spanish dialects. *Lingua*, *151*, 162-176.

Zeijlstra, Hedde. 2012. There is only one way to agree. *The Linguistic Review* 29(3). 491–539