

**Not All Clitics are Made the Same:
Defective Agreement in Spanish Dative Clitic-Doubling.
Gustavo Guajardo
UiT, The Arctic University of Tromsø**

INTRODUCTION: In Spanish clitic-doubling constructions, the dative clitic must agree in number with its co-referential NP (1a). However, under certain circumstances, agreement does not always obtain and the singular clitic is found with a co-referential plural NP (1b). I call this *defective agreement* (DA). Although this may appear to be a phonological process of s-weakening, defective agreement is constrained syntactically: it is not possible with preverbal indirect objects (IOs) (2a) or in non-doubling contexts (2b). Therefore, DA does not appear to be a phonological phenomenon.

1. a. Les di un regalo a mis padres.
b. [Le]_{SG} di un regalo [a mis padres]_{PL}
'I gave my parents a present'

2. a. A mis padres, *le/ les di un regalo.
'To my parents, I gave (them) a present'
b. Le di un regalo a mis padres pero no *le/ les dije dónde lo compré.
'I gave my parents a present but I didn't tell them where I bought it'

RESEARCH QUESTIONS: (1) What linguistic factors favour DA? (2) What can DA tell us about the syntax of clitics? **HYPOTHESIS:** defective agreement will be more likely with singular direct objects (DOs) because the DO is linearly closer to the clitic. **METHOD:** Data from 21 Spanish-speaking countries was extracted from Corpus del Español (Davies 2002) Web Dialects version. Data extraction was done manually alternating the clitic position (*pre-* or *postverbal*), the number of the clitic (*sg* vs. *pl*) and the gender, number and definiteness of the DO (*masc* vs. *fem*; *sg* vs. *pl*; *bare* vs. *definite* vs. *indefinite*). Only sentences with plural definite IOs and with a minimum token frequency of 2 were extracted. The final dataset comprises 2376 sentences and was coded for the following variables: COUNTRY, ADDRESS, FINITE-VERB, NON-FINITE VERB, PERSON, NUMBERVERB, ANIMACYDO, ANIMACYIO, POSITIONCLITIC, NUMBERDO, GENDERDO, GENDERIO, DEFINITENESSDO, TYPEIO. **STATISTICAL ANALYSIS:** The data had class imbalance with 78% of the data comprised of the singular clitic and 22% of the plural clitic. To remedy this problematic class imbalance (e.g., Provost 2000, Chawla et al 2004), random undersampling was performed to create a more balanced dataset with 60/40 proportion of singular and plural clitics, respectively. This proportion was determined by fitting a conditional inference tree with different proportions and calculating the Balanced Accuracy and Kappa for each tree (Provost 2000, Weiss and Provost 2006). 60/40 was the proportion with the highest values for both measures (1330 sentences). This balanced dataset was used to fit a mixed-effects logistic regression model with ADDRESS and VERB as random intercepts. **RESULTS:** Figure 1 shows the proportion of DA by country. The data shows that the singular clitic is the most frequent form in every single country in plural dative clitic-doubling. A conditional inference tree fit with COUNTRY as a predictor variable and DA as the dependent variable found no group differences across countries. The mixed-effects model (Model-1) ($C\text{-index} = 0.87$, *Balanced Accuracy* = 0.80) contains PERSON, GENDERIO and TYPEIO as single terms and the interactions

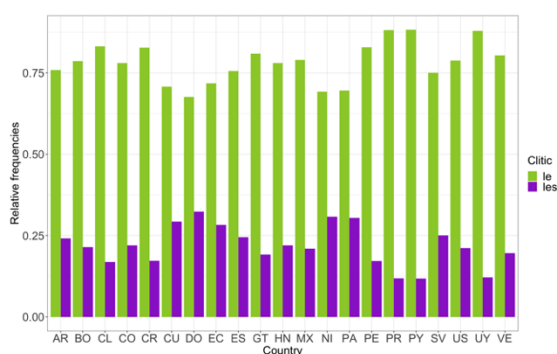


Figure 1. Proportion of DA by country.

NUMBERDO*ANIMACYIO,

Table 1. Coefficient estimates of Model-1.

Model Estimates					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
	(Intercept)	0.47	0.68	178

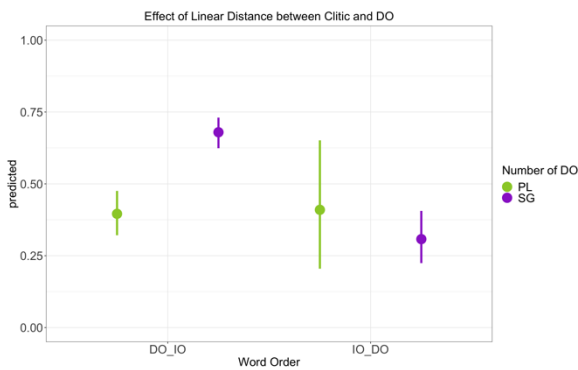


Figure 3. Predicted probabilities of DA from interaction Word order*Number of DO.

plural DOs ($p = 0.91$). **Discussion:** The results of Model-2 support the hypothesis that linear order has an effect on DA as when the DO is closest to the clitic, the probability of DA reaches its highest value. Given the prevalence of DA over full agreement in the data, DA seems to be the result of grammaticalization caused by an attraction effect of the DO. This process has all the characteristics of grammaticalization processes: the clitic has undergone semantic bleaching (loss of plural marking) and phonetic reduction (loss of the segment /s/) (Hopper and Traugott 1993) and the cline *pronoun > clitic > agreement marker* is a well-attested process in the evolution cycle of pronominal forms (Hopper and Traugott 1993). In addition, we found a new context where DA is disallowed, namely when the doubled IO is realized as a strong pronoun. The three contexts where DA is not allowed share the fact that the dative clitic is the true argument of the verb because it is a non-doubling context as in (2b) or the lexical NP is right or left-dislocated (i.e., strong pronouns are right dislocated and preverbal IOs are left-dislocated) (Roca 1992). This suggests that DA is not allowed when the clitic is in an argument position. Based on these observations, I propose that the dative clitic behaves both as a pronominal clitic and as an agreement marker and DA is sensitive to this distinction: DA is only possible when the clitic is an agreement marker.

Conclusion: This study provides evidence that defective agreement of the dative clitic is a pervasive phenomenon across all Spanish varieties; it is the norm, not the exception. The most important variables for DA are the type of IO, animacy of the IO, the number of the DO and linear distance between the DO and the clitic. I have also proposed that the structural constraints on DA may provide indirect evidence for the categorial status of the clitic as DA is only possible when the clitic is in a non-argumental position and it is banned elsewhere. The categorial status of the clitic, then, depends on the structural configuration in which it appears, suggesting that analyses where the dative clitic is analysed as either an agreement morpheme or a pronominal clitic *across the board* cannot be maintained. The results herein also highlight the potential of morphosyntactic variation to inform linguistic theory.

DEFINITENESSDO*GENDERDO and DEFINITE DO*NUMBERV.

Table 1 shows the coefficient estimates of the model (reference level: plural clitic). The model shows that TYPEIO has the biggest effect such that pronouns almost categorically reject DA. The second largest effect is

the interaction NUMBERDO*ANIMACYIO such that inanimate objects with singular DOs provide the most likely context for DA. To investigate whether linear order of the DO was significant for DA a second mixed-effects model (Model-2) was fit with the interaction WORD_ORDER (Clitic IO DO vs. Clitic IO DO) and NUMBERDO as a predictor. The interaction is highly significant ($\beta = -1.62, p < 0.01$) and the results are shown in Figure 3 as predicted probabilities of DA. We can see that singular DOs with word order DO_IO increase the predicted probability of DA to 0.70 while, in the opposite word order, it drops to 0.30. Word order is not a significant predictor with