



## A Partitioned Chi-square Analysis of Variation in Spanish Clitic Doubling: The Case of *dar* 'give'

Raúl Aranovich

To cite this article: Raúl Aranovich (2016) A Partitioned Chi-square Analysis of Variation in Spanish Clitic Doubling: The Case of *dar* 'give', Journal of Quantitative Linguistics, 23:3, 295-313, DOI: [10.1080/09296174.2016.1169846](https://doi.org/10.1080/09296174.2016.1169846)

To link to this article: <https://doi.org/10.1080/09296174.2016.1169846>



Published online: 09 Jun 2016.



Submit your article to this journal [↗](#)



Article views: 204



View related articles [↗](#)



View Crossmark data [↗](#)

---

## A Partitioned Chi-square Analysis of Variation in Spanish Clitic Doubling: The Case of *dar* ‘give’\*

Raúl Aranovich

University of California Davis, Department of Linguistics, USA

---

### ABSTRACT

In Spanish ditransitive clauses with Dative Clitic Doubling (DCID), the goal may be doubled by a dative clitic pronoun. Applying a chi-square decomposition analysis to examples with the prototypical ditransitive verb *dar* ‘give’, from a corpus of journalistic language characteristic of River Plate Spanish, I show that DCID is disfavoured when the theme is a bare noun, and also that the distinction between bare and non-bare themes is more statistically significant than the distinction between definite and indefinite themes. The relational effect of definiteness on DCID in clauses with the verb *dar* is accounted for by extending Hopper and Thompson’s (1980) Transitivity Hypothesis to clauses with three arguments.

### 1. INTRODUCTION

Current research methods in corpus linguistics make it possible to quantify variation in syntax, leading to more insightful analyses of sentences that were previously considered to be no more than equally possible alternatives of each other. In this paper I will examine Spanish sentences in which the presence of a dative clitic is apparently optional. The issue is illustrated by the examples in (1).

- (1) a. *el Gobierno podría dar una solución política al tema*  
the government could give a solution political to.the issue  
‘The government could give a political solution to the issue.’ [05  
(14)70]  
b. *Paredes le dio un puñetazo a Urquiza.*  
Par edes DAT give a punch to Urquiza.  
‘Paredes gave Urquiza a punch (with the fist).’ [02(51)86]

---

\*Email: [raaranovich@ucdavis.edu](mailto:raaranovich@ucdavis.edu)

These are ditransitive sentences with the verb *dar* ‘give’, extracted from a corpus of journalistic Spanish. In both sentences, the verb is followed by two overt arguments: a theme and a goal/beneficiary (in that order). In (1) b, however, the goal is cross-referenced by an additional pronoun in the dative case, which is proclitic to the verb. This phenomenon is known as Dative Clitic Doubling (DCID). What I will show here is that there are clear preferences for sentences with or without DCID, based on other grammatical features of the clause. In particular, I will show that a strong association between DCID and the definiteness of the theme can be observed, but only after themes are categorized into three different levels of definiteness. What emerges from this analysis is a situation in which DCID is disfavoured by clauses in which the theme is realized as a bare noun. I will also argue that this observation is accounted for by a generalization of Hopper and Thompson’s (1980) Transitivity Hypothesis to cover ditransitives.

The paper is structured as follows. Section 2 presents the data, detailing the structure of the corpus from which the examples are taken. The three levels of definiteness of the theme are characterized, showing attested examples of all three levels in conjunction with dative clitics and also without them. Section 3 summarizes the distribution of the data in a contingency table, which is analysed using statistical methods to deal with categorical data. In particular, I apply the chi-square decomposition methodology developed to analyse  $R \times C$  two-way tables. What I show here is that the null hypothesis of independence between the two variables (DCID and definiteness of the object) is rejected, but also that not all of the levels of object definiteness contribute to the rejection of the null hypothesis with equal strength. Section 4 explores the consequences of the analysis for theories of ditransitives, arguing that Hopper and Thompson’s model for the association between semantic transitivity and the morphosyntactic exponents of transitivity provides an explanatory approach to the distribution of DCID, once their hypothesis is extended to ditransitives. In a nutshell, I claim that doubling of a goal/beneficiary by a dative clitic is a morphosyntactic expression of increased ditransitivity, and that bare objects disfavour DCID because they correspond to decreased semantic ditransitivity. Section 5 concludes the paper with an evaluation of the application of categorical data analysis methods for the development of linguistic theory.

## 2. EXPLORING DATIVE CLITIC DOUBLING IN A CORPUS OF WRITTEN SPANISH

Dative Clitic Doubling, the phenomenon illustrated by the examples in (1), has a long tradition of research in the field of Spanish linguistics. However, this is often in connection with other phenomena involving clitics, such as the obligatory doubling of some direct objects by accusative clitics (Suñer, 1991, 2000; Torrego, 1995), the alternation between accusative and dative clitics in “*leísta/laísta*” dialects (Bleam, 1999; DeMello, 2002; Romero, 2013), or the contrast between the optional marking of indirect objects by dative clitics and the obligatory marking of experiencers with psychological predicates (Masullo, 1992; Dufter & Stark, 2008). While some investigations regard DCID as an entirely optional phenomenon, independent of any other grammatical factors (Suñer, 1991), others examine the conditions under which ditransitive clauses may favour or disfavour doubling of an indirect object by a dative clitic. Their conclusions, however, are limited to pointing out broad dialectal differences in the preferences for DCID, or the effects of grammatical factors (based on linguistic intuitions). Thus, Parodi (1998) shows that DCID is more frequent in Latin American Spanish than in Iberian Spanish, and Demonte (1995) claims that there is a preference for DCID when the indirect object precedes the direct object. With the availability of large electronic corpora in Spanish, and with the development of powerful statistical methods (and computational implementations) for the analysis of categorical data in linguistics, it is now possible to investigate the systematic association between DCID and other linguistic variables in language use. This is the approach I will follow in this paper.

The literature on dative case and ditransitives is vast and growing. One phenomenon that has received a lot of attention is the alternation between the English double complement construction (2) a and the double object construction (2) b.

- (2) a. Wallace gave a present to Gromit.  
 b. Wallace gave Gromit a present.

While some researchers take the two sentences to be entirely synonymous, deriving one from the other (Dryer, 1986; Larson, 1988; Baker, 1997) others argue that the alternation is conditioned by the intrinsic semantic features of the complements and also by the meaning of the verb (Oehrle, 1976; Jackendoff, 1990; Dowty, 1991; Harley, 2002; Krifka,

2004). Thus, as the contrast between (3) a and (3) b illustrates, the double object construction favours a theme that is indefinite, but a goal that is definite (Ransom, 1979).

- (b) a. They fed the Christian a lamb.  
 b. ?They fed a lion the Christian

This line of inquiry suggests that the definiteness of the theme is a factor that may enter into consideration to choose between alternative structures in ditransitives, providing a lead as to where to start testing for associations between the doubling of a dative clitic in Spanish ditransitives and other linguistic variables.

The source of the data for the analysis is a corpus of about one million Spanish words, taken from the daily edition of the Argentine newspaper *La Nación*. The corpus was created from the online edition of the newspaper, from February 1 to February 13 of 2008. It includes journalistic articles about everyday news, opinion pieces, interviews, and editorials, but also weekly supplements on specific topics (business, travel, arts and culture, sports, etc.), and a weekly magazine. Online commentary about the articles by the readers is also included. The corpus, then, is representative of the variety of South American Spanish spoken in the River Plate area. In order to investigate DCID, I searched the corpus for sentences containing the verb *dar* ‘give’, which is the prototypical ditransitive verb, and one of the most frequent ones in its class. From those, I selected all the sentences that had two overt complements, a theme and a goal/beneficiary, in their canonical order (i.e. with the theme preceding the goal/beneficiary). This resulted in a sample of 349 sentences.

Two variables are used to cross-classify these 349 sentences. One is a dichotomous variable having to do with the optional use of the dative clitic. An example like (1) a would be coded as DCID = no, and (1) b as DCID = yes. The second variable is the definiteness of the object. This variable has three levels: ‘definite’, ‘indefinite’, and ‘bare’. The examples in (1) both have an indefinite object. The classification is based on the type of determiner the object has, in this case the indefinite article *un* ‘a’. In contrast, definite objects have a definite article like *el* ‘the’ (which, like the indefinite article, is conjugated for person and number), as in (4). Bare objects have no determiner, as shown in (5). Notice that in each of these pairs of sentences only the second one has a dative clitic doubling the indirect object.

- (4) a. *este juez dio la orden a Gendarmeria de librar este paso.*  
 this judge gave the order to Border. Police of freeing this pass  
 “This judge gave an order to free this (mountain) pass to the Border Police.” [13(9)138]
- b. *Cuatro jugadas más tarde... le dio la victoria a los Giants.* four plays more late DAT gave the victory to the Giants  
 “Four plays later, he gave the Giants the victory.” [05(51)65]
- (5) a. *Carioca Da Gema ... o logró más que darle emoción al desenlace.*  
 Carioca Da Gema not achieved more than give.DAT excitement to.the ending  
 “Carioca Da Gema didn’t achieve anything other than to give excitement to the ending.” [02(53)69]
- b. *La Casa Rosada aún da impulso a la postulación de Iribarne.* The House Pink still gives impulse to the candidacy of Iribarne  
 “The Pink House is still giving impulse to Iribarne’s candidacy.” [05(10)79]

There is more to the classification of objects according to definiteness, so a more detailed exposition of the criteria I used is in order here. First, definite objects include noun phrases (NPs) with possessive specifiers, as in (6). And indefinites include nouns preceded by “weak” quantifiers, like *cierta* “some” and *muchísimo* “much, a lot”, as in (7).

- (6) a. *Piccoloti dio su apoyo a los cortes.*  
 Piccoloti gave his support to the cuts  
 “Piccoloti gave his support to the cuts.” [03(13)22]
- b. *le di mi auto a un asistente muy joven...*  
 DAT gave my car to an assistant very young.  
 “I gave a very young assistant my car.” [03(115)117]
- (7) a. *Los sondeos dan hoy cierta ventaja al ex obispo Fernando Lugo*  
 the surveys give today some advantage to the former bishop Fernando Lugo  
 “Today, the surveys give some advantage to former bishop Fernando Lugo.” [05(11)78]
- b. *Esta movida...puede darle muchísimo oxígeno a Yahoo!*  
 this move may give.DAT much oxygen to Yahoo!  
 “This move may give Yahoo! a lot of oxygen.” [02(15)107]

The distinction between definites and indefinites can be instrumentalized, then, as a contrast between the kinds of NPs that can or cannot appear in an existential construction, as in (8) a. Notice, however, that this criterion would classify bare nouns as indefinite (8) b. But bare nouns can be distinguished from indefinites based on their form alone.

- (8) a. *hay un/ cierto/ mucho/ \*el/ \*mi prejuicio contra la lectura.*  
 there.is a/ certain/ much/ the/ my prejudice against the reading  
 “There is some/much/the/my prejudice against reading.”  
 b. *hay prejuicio contra la lectura.*  
 there.is prejudice against the reading  
 “There is prejudice against reading.”

I will return later to consider the semantic differences between bare nouns and indefinites in more detail. For the present purpose of making explicit the criteria I used in classifying objects according to definiteness, it is important to point out that bare nouns include singular and plural nouns too (9) a, and also that when an NP does not have a determiner it is classified as an indefinite if it is modified by an adjective or other phrases (9) b-c.

- (9) a. *Buscamos darle soluciones a la población.*  
 seek.1.pl give.inf.dat solutions to the population  
 “We are trying to give solutions to the population.” [04(9)102]  
 b. *Hay orden en el gobierno de darle trato preferencial a*  
 there.is order in the government of give.INF. DAT treatment preferen-  
 tial to *Santa Fe*. *Santa Fe*  
 “There is an order from the government to give preferential treat-  
 ment to Santa Fe.” [01(13)68]  
 c. *Los negocios y restaurantes dan vida nocturna al lugar.*  
 the shops and restaurants give life nocturnal to the place  
 “Shops and restaurants give night life to the place.” [02(36)50]

To summarize, then, the three levels of the definiteness variable are identified by the following criteria:

- Definite: Proper nouns; NPs with definite determiners (e.g. *el, las* “the.M.SG, the.F.PL”), demonstrative determiners (*este, esta* “this.M, this.F”, *ese, esa* “that.M.SG, that.F.SG”), possessive pronouns (e.g. *mi*

- “my”, *su* “his/her/its/their”), and strong quantifiers (e.g. *todo* “all”, *cada* “each”); free relatives headed by *el que*, *aquel que* “he who”.
- Indefinite: NPs with indefinite determiners (e.g. *un*, *unas* “a.MSG, a.F.PL”); NPs without determiners, but modified by adjectives; weak quantifiers (e.g. *nadie* “nobody”, *uno* “one”); NPs introduced by weak quantifiers (e.g. *mucho* “a lot of”, *cierto* “some”, *más* “more”, *tanto* “many, much”, *varios* “several”, *algún* “some”, *demasiado* “too much, too many”) or cardinal determiners (e.g. *dos* “two”); free relatives headed by *quien* “whomever”.
  - Bare: Common nouns, plural or singular, without any determiners or modifiers.

### 3. A CHI-SQUARE ANALYSIS

Once the sentences in my sample are classified according to the two variables discussed previously (doubling of the indirect object by a clitic and definiteness of the object), the frequencies can be summarized in a  $3 \times 2$  contingency table (Table 1).

The null hypothesis is that DCID is independent of the definiteness of the object. The alternative hypothesis is that the two variables are associated, that is, that the difference between the expected and observed frequencies in the table cannot be attributed to sampling error if the expected frequencies are a function of the marginal totals for rows and columns independently of each other. The expected frequencies under independence are presented below (Table 2):

One way to test the null hypothesis of independence is with a chi-square statistic. This can be easily calculated with the help of the *loglin* function in the statistics software package R (accessed through the *loglm* interface).

Table 1. Frequencies of *dar* clauses.

ODEF	LE		
	n	y	
B	114	57	171
I	51	79	130
D	25	23	48
	190	159	349



Table 2. Expected frequencies for Table 1.

ODEF	LE		
	n	y	
B	93.09	77.91	171
I	70.77	59.23	130
D	26.14	21.86	48
	190	159	349

The results of testing the hypothesis of independence on Table 1 are displayed in Table 3.

The *loglin* function returns a likelihood ratio statistic, with an approximate chi-square distribution. With 2 degrees of freedom (df), it is clear that the value of the Likelihood Ratio statistics is more than high enough to reject the null hypothesis.

A chi-square statistic like the one in Table 3 may let us infer that there is interaction between the variables, but not much more. It does not indicate the strength of the association, or the direction of the association (negative or positive). Fortunately, there are methods of categorical data analysis that can offer answers to these questions (Goodman, 1968; Fienberg, 2007 [1980]). Moreover, when one of the variables is polychotomous, it is possible to discern the contribution of separate sub-tables to the overall chi-square statistic, to arrive at a more fine-grained understanding of the structure of the data. The first step in this exploratory process is to obtain residuals. These are the differences, for each individual cell, between the observed and expected values (under the hypothesis of independence). Residuals (adjusted residuals, to be more precise) can be obtained with the R function *resid*, applied to the output of *loglm*. For Table 1, the results are displayed in Table 4.

The table of residuals shows that large departures from the expected values happen in the rows labelled Bare and Indefinite, and that the values in

Table 3. Chi-square analysis of Table 1.

	$X^2$	df	$P(> X^2)$
Likelihood Ratio	22.77292	2	1.134811e-05

Table 4. Residuals of Table 1.

ODEF	LE	
	n	y
B	2.0923616	2.4884146
I	-2.4750473	2.4432641
D	-0.2230325	0.2399838

the Definite row are very close to the expected values (thus contributing little to the breakup of the hypothesis of independence). Moreover, the sign of the residuals shows that the indefinite and definite rows move in the same direction, contributing to an increase in the frequencies of the “yes” column. The Bare row, on the other hand, contributes to a decrease in the frequencies of the “yes” column.

Based on the analysis of residuals, I propose to test the hypothesis that the contrast between the classification of an object as either indefinite or definite is statistically independent of the absence or the presence of a dative clitic doubling the indirect object. In other words, I want to know if the association between object definiteness and DCID can be reduced to the opposition between bare nouns and non-bare nouns (definite and indefinite). One way to do this is to partition Table 1 into two  $2 \times 2$  sub-tables. The first sub-table, Table 5(a), is made of the two bottom rows and the two columns of Table 1. The second sub-table, Table 5(b), is made of the remaining (top) row, and another row obtaining from adding over the frequencies in each of the columns of Table 5(a).

For each sub-table it is possible to calculate a chi-square statistic with  $df=1$ , to examine their contribution to the overall chi-square value of Table 1. In the case of the Likelihood Ratio statistic, the chi-square values of the sub-tables add up to the overall chi-square value, as shown in Table 6.<sup>1</sup>

From the partitioning of chi-square analysis it is clear that only the opposition between bare and non-bare objects is significant to establish an association between object definiteness and DCID. A cursory examination of Table 1 may make it seem that there is also a difference between sentences

<sup>1</sup>The logline function also returns a Pearson statistic, which has an approximate chi-square distribution as well. When considering the contributions of each sub-table to the overall chi-square, however, the results are only approximate for the Pearson chi-square statistic. This is the reason why I only consider the Likelihood Ratio statistic in the analysis.

Table 5a and 5b. Partition of Table 1 ([DI]B).

ODEF	LE		
	n	y	
I	51	79	130
D	25	23	48
	76	102	178

  

ODEF	LE		
	n	y	
B	114	57	171
D+I	76	102	178
	190	159	349

Table 6. Decomposition of chi-square for Tables 5a-b.

Component due to:	LRX <sup>2</sup>	df	<i>p</i> (> X <sup>2</sup> )
Indefinite vs Definite	2.35027	1	0.1252614
Bare vs. non-bare	20.42265	1	6.209056e-06
Total (overall)	22.77292	2	

with definite objects and sentences with indefinite objects in their tendency towards favouring the doubling of indirect objects by a clitic. After all, among the sentences with definite objects, the ones without dative clitics slightly outnumber those with dative clitics. But a detailed statistical analysis of residuals and partial chi-square values shows that whatever departures from expected values happen in that sub-table can also be attributed to sampling error under independence.

Another way to look at the same problem is to consider what would happen to the null hypothesis of independence between object definiteness and DCID if the Bare row were to be collapsed with the Indefinite row. The reason for asking the question in this way is that, as I have shown before, bare nouns pattern with indefinites in their ability to appear in the existential construction. To test this hypothesis, I partition Table 1 again into two  $2 \times 2$  sub-tables. The first sub-table is made of the top two rows and the two columns (Table 7 (a)), and the second sub table of the bottom row of Table 1 and a new row that results from adding over the columns of Table 7(a). The results are summarized in Table 8.

Table 7a and 7b. Partition of Table 1 ([BI]D).

ODEF	LE		
	n	y	
B	114	57	171
I	51	79	130
	165	136	301

  

ODEF	LE		
	n	y	
B+I	165	136	301
D	25	23	48
	190	159	349

Table 8. *Decomposition of chi-square for Tables 7a-b.*

Component due to:	LRX <sup>2</sup>	df	<i>p</i> (> X <sup>2</sup> )
Indefinite vs Bare	22.64842	1	1.945205e-06
Definite vs. non-definite	0.12450	1	0.7242021
Total (overall)	22.77292	2	

Clearly, the association between object definiteness and DCID would be obscured if the distinction between bare nouns and other indefinites is not taken into account.

The conclusion of a statistical analysis of Table 1 based on partitioning of the chi-square statistic is that the null hypothesis of independence between object definiteness and DCID is rejected, but that this is mainly because of the opposition between bare objects and other objects (non-bare indefinites and definites). Moreover, it becomes clear that dative clitics are favoured by clauses containing a non-bare noun object, but disfavoured in the case of sentences with bare noun objects. In the next section I will develop an explanatory account of these generalizations based on an extension of Hopper and Thompson's (1980) Transitivity Hypothesis.

#### 4. TRANSITIVITY, DITRANSITIVES, AND DCLD

Transitive sentences typically consist of a verb with a subject and an object. Morphosyntactic means of coding subject and object, such as word order,

verb agreement, or case inflexions, are employed to clearly distinguish one argument from the other. An example comes from the verbal agreement system of some Finno-Ugric and Uralic languages (Aranovich, 2007). In these languages there is a complex system of agreement that employs portman-teau morphemes to cross-reference the subject and the object. In Hungarian, for instance, the suffix *-lak* indicates that the subject is first person and the object second, while the suffix *-om* indicates that the subject is first person and the object third person.

- (10) a. *Lat-lak*.  
 see-1.sg.S/2.O  
 “I see you (sg/pl.fam)”  
 b. *Lat-om a hazat*.  
 see-1.sg.S/3.O the house  
 “I see the house.”

The suffix *-om* is used if the object is a definite NP. A different suffix, *-ok*, is used to express agreement with a first person singular subject when the object is an indefinite, as in (11). Interestingly enough, this is the same suffix that occurs on intransitive verbs.

- (11) *Lat-ok egy hazat*.  
 see-1sg a house  
 ‘I see a house.’

The Hungarian definite conjugation, then, can be treated as a morphosyntactic expression of transitivity, since it encodes features of object and subject. The indefinite conjugation, on the other hand, is used to encode features of the subject alone in clauses with an (indefinite) object, and it is also used in intransitive clauses. Therefore, it cannot be treated as an expression of transitivity. What is interesting here is that the semantic features of the object are associated with the choice of a transitive or an intransitive verbal conjugation: definite objects require the transitive encoding, indefinite objects the intransitive encoding.

It is facts like this one that give support to Hopper and Thompson’s (1980) Transitivity Hypothesis. Hopper and Thompson make a distinction between morphosyntactic transitivity, of the sort that is encoded in agreement patterns like those of the Hungarian verb, for instance, and semantic transitivity. From a semantic point of view, a clause is transitive if it has

two arguments that are clearly seen as separate entities participating in an event. Hopper and Thompson list several features that contribute to increase or decrease the transitivity of a clause at the semantic level. Among them, the degree of individuation of the object is important: a clause has increased transitivity if the participants in the event are seen as distinct individuals (both from each other and from the event itself). Arguments that can be named, for instance, are more individuated than arguments referred to by common nouns, and arguments in the singular are more individuated than plural arguments. The Transitivity Hypothesis is that there is a correlation between increased transitivity at the semantic level with a tendency to use transitive means to encode the arguments at the morphosyntactic level. Thus, in the Hungarian case discussed above, an intransitive conjugation is chosen when the object is indefinite because indefinite arguments are less identifiable than definite ones, and therefore a clause with an indefinite object is semantically less transitive than a clause with a definite object.

The TH is originally designed to account for a diverse range of facts concerning sentences with two arguments, but it can certainly be extended to handle ditransitive sentences as well. Kittilä (2007) does so to account for alternations in the realization of the goal. In Czech, for instance, the goal is marked dative when it is highly affected (12) b, but it occurs as the complement of a preposition when it is less affected (12) a.

- (12) a. *ani pes by od něho kůrku chleba nevzal.*  
 not.even dog.NOM would from him.GEN crust.ACC bread not.-  
 took  
 “Not even a dog would take a crust of bread from him.”
- b. *ani pes by mu kůrku chleba nevzal.*  
 not.even dog.NOM would him.DAT crust.ACC bread not. took  
 “Not even a dog would take away his crust of bread.”

Besides affectedness, Kittilä considers the effects of aspect (i.e. completeness of the event), the animacy features of the goal, and purposefulness on the transfer event on the marking of the goal. His conclusion is that dative case-marking marks goals that have high transitivity features.

To account for variation in DCID in Spanish, however, I need to take a slightly different approach. First, I do not think the variation in DCID reflects a change in the grammatical function of the goal. In this sense, the

Spanish examples are different from the examples in Kittilä's paper. The main issue for me is what conditions the overt expression of dative case by means of the clitic pronoun cross-referencing the goal.<sup>2</sup> Second, I am interested in the effects that features of the theme have on the marking of the goal, not the features of the goal itself. In particular, I am investigating the effects of definiteness on DCID, a feature that Kittilä's study does not mention (he notices an individuation effect, but linked to animacy, rather than definiteness).

Indirect (or secondary) objects, then, are coded by their own specific morphosyntactic resources: dative case marking, cross-reference by verbal agreement affixes (in some cases a separate set of affixes from those used for object agreement), and word order (with respect to the direct or primary object). According to the TH (in its more general form), these structural features correlate with semantic ditransitivity, which I understand as a clear differentiation between the object (or theme) and indirect object (goal/beneficiary) of the event, or between the object and the event itself. The more clearly differentiated the two internal arguments of the verb are, the more likely the clause is to sport the morphosyntactic features that mark ditransitivity. Again, definiteness (as one of the ingredients of argument individuation) has a big role to play.

Having extended the Transitivity Hypothesis to cover ditransitive clauses in this way, it is possible to account for the association between object definiteness and DCID in Spanish. Cross-referencing the indirect object by a dative clitic is clearly a way to encode ditransitivity. Clauses with a dative clitic, then, are higher than clauses without the clitic in the morphosyntactic dimension of ditransitivity. On the semantic side, then, clauses whose objects are bare nouns have a lower index of ditransitivity, since a bare noun is much less individuated than an indefinite or a definite noun. The distribution of DCID, then, can be accounted for by the following generalization:

- (13) The more highly individuated the direct object is, the more likely the indirect object is to be overtly marked as a dative complement.

---

<sup>2</sup>The hypothesis that variation in the marking of the goal corresponds to changes in its grammatical function is defended in Demonte (1995) and Cuervo (2003). In this approach, the prepositional phrase that is not doubled by a dative clitic is an oblique. DCID, then, is not optional, but an obligatory way to mark an indirect object. I will not explore this approach here.

The generalization in (13) follows from the extension of the Transitivity Hypothesis to ditransitives, considered from the point of view of the relative degree of individuation (specificity) of the patient with respect to the goal. And here is where the novelty and the explanatory power of the analysis can be found, since it appeals to the general notion of a correlation between semantic and morphosyntactic transitivity features, but extending it to a new domain. There are, however, a couple of further issues that require clarification. The first one is the status of the opposition between definite and indefinite objects with respect to the distribution of dative clitics and the Transitivity Hypothesis. If definite objects make the clause more ditransitive than indefinite objects, then one would expect to find more dative clitics when the object is definite. But the data show this to be the opposite. However, the partitioning of chi-square carried on in the previous section shows that the sub-hypothesis of independence between DCID and object definiteness (excluding bare objects) cannot be rejected. And in any case, it is not uncommon for particular languages to set up cut-off points when hierarchies of semantic features like these are in correlation with morphosyntactic features (Silverstein, 1976; Aissen, 1999).

The second issue is that, unlike the Hungarian agreement examples, the decision as to whether or not to cross-reference the indirect object with a clitic pronoun depends not on the semantic features of the argument that is being cross-referenced, but on the features of another argument (the object, in this case). This is not an uncommon occurrence, however. For instance, in the Paleo-Siberian language Chukchi (Kozinsky et al., 1988), subjects are marked ergative (the case that normally codes transitive subjects in this type of languages) when the object is definite (14) a, but when the object incorporates to the verb, as in (14) b, the case marking of the subject changes to absolutive (the case used to mark intransitive subjects).

- (14) a. *ənan qaa-t qərir-nin-et.*  
 he.ERG deer-ABS.PL look.for-3sg.S/3sg.O.AOR-PL  
 “He was looking for the deer.”  
 b. *əton qaa-rer-gʔe.*  
 he.ABS deer-look.for-3sg.AOR  
 “He was looking for deer.”

Incorporated objects, like the one in (11) b, are often less individuated than their canonical (i.e. non-incorporated) counterparts. Incorporation, then, can be seen as another morphosyntactic way to encode the object in a tran-



sitive clause with reduced transitivity at the semantic level. This reduction in transitivity, triggered by the semantic features of the object, has then an effect on the marking of the subject: absolutive instead of ergative.

Even though noun incorporation is not an active grammatical process in Spanish, the analogy between the effects of a bare object on DCID and object incorporation in Chukchi is suggestive, since many of the examples of *dar* “give” followed by a bare noun can be paraphrased as a single verb, cognate with the bare noun. Thus, *dar impulso* “give impulse” can also be expressed as *impulsar* “to push (something) forward”. In these cases, *dar* is functioning like a light verb, that is, a verb that is there to bear tense, agreement, and other functional information, but whose descriptive content comes from its complement noun. In light verb complexes, then, the complement noun is not a true semantic argument of the verb, but rather a semantic constituent of the predicate (Grimshaw & Mester, 1988; Samek-Lodovici, 2003).<sup>3</sup>

## 5. CONCLUSION

In this paper I have shown that there is an association between the presence or absence of a dative clitic and the definiteness of the direct object. Evidence for this claim came from a statistical analysis of the actual frequency of DCID in a corpus of River Plate Spanish. The association, however, is mostly due to the opposition between bare objects and non-bare objects, which include definite objects and other indefinites. This conclusion was arrived at through a partitioning of chi-square to test sub-hypothesis of independence. I proposed an account based on an extension of Hopper and Thompson’s (1980) Transitivity Hypothesis (to cover the relationship between direct and indirect objects in ditransitives), which rested on two insights: that clauses with bare objects have a lower coefficient of semantic (di)transitivity than clauses with other types of objects, and that dative clitics are the structural manifestation of a clause with a higher ditransitivity coefficient at the morphosyntactic level, compared to clauses without dative clitics. The direction of the association between DCID and object definiteness is then predicted by the Transitivity Hypothesis, since clauses without dative clitics have a significantly higher proportion of bare objects than

---

<sup>3</sup>Alternatively, the bare object can be seen as an argument that is semantically incorporated (Van Geenhoven, 1995, 1998), without undergoing incorporation in the syntax.

clauses with dative clitics, and clauses with dative clitics have a significantly higher proportion of non-bare objects than clauses with dative clitics.

The relationship between DCID and object definiteness that I have uncovered is not absolute, but a well-defined tendency. There are still many ditransitive sentences with dative clitics and bare objects, and also many ditransitive sentences without dative clitics and definite or indefinite objects. The advantage that corpus research methods has over other ways of obtaining data (i.e. introspective grammaticality judgments) is that it makes it possible to quantify variation in actual usage; coupled with well-developed methods for statistical data analysis, this kind of research results in objective claims about the sources of variation. Recent work on language usage has concluded that the variable syntactic phenomena discovered in corpus research are the mirror image of more rigid grammatical facts observed across languages. Bresnan et al. (2001), for instance, show that the effects of a hierarchy of person marking (1st/2nd > 3rd) are variable in English passives: if a patient outranks the agent in the person hierarchy, then there is a tendency for that clause to be passivised. This tendency of the English grammatical system is found in Lummi (a Salish language) as a categorical process.<sup>4</sup>

My investigation into the grammatical factors that correlated with variation in DCID allows me to show how well the Transitivity Hypothesis fits the hypothesis that the tendencies found in language usage are gradient counterparts of rigid constraints identified across languages. Hopper and Thompson's approach to the relationship between form and content is not deterministic, since it leaves room for a lot of variation across languages. It does not specify that a certain semantic feature (say, specificity) will be universally coded by a certain morphosyntactic feature (case marking, or agreement, or word order), but that there will be a tendency across languages for that semantic feature to be associated with a morphosyntactic feature of equal sign with respect to transitivity. In sum, by combining corpus research with a solid statistical analysis of the data, the same principles that account for variation across languages in Hopper and Thompson's approach can be shown to work for variation within a language as well.

---

<sup>4</sup>More recently, Bresnan and Hay (2008) apply similar concepts to an analysis of the dative alternation in English, showing significant differences in the effect of animacy between New Zealand and the United States.

## ACKNOWLEDGEMENTS

I have been fortunate to be able to present some of my ideas about the quantitative analysis of clitic doubling in Spanish to a number of audiences. For their helpful comments, I thank Luís Paris, Monique Lamers, Chiyo Nishida, John Beavers, Stefan Gries, Viola Miglio, and Eduardo Raposo. All mistakes are my own.

## DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author.

## REFERENCES

- Aissen, J. (1999). Markedness and subject choice in optimality theory. *Natural Language & Linguistic Theory*, 17, 673–711.
- Aranovich, R. (2007). Optimizing verbal agreement in Mordvin. *Studia Linguistica*, 61, 185–211.
- Baker, M. C. (1997). *Thematic roles and syntactic structure*. In L. Haegeman (Ed), *Elements of Grammar* (pp. 73–137). Dordrecht: Kluwer.
- Bleam, T. M. (1999). *Leista Spanish and the Syntax of Clitic Doubling*. PhD Dissertation, University of Delaware.
- Bresnan, J., & d Hay, J. (2008). Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua*, 118, 245–259.
- Bresnan, J., Shipra, D., & Manning, Ch. (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In M. Butt & T. Holloway King (Eds.), *Proceedings of the LFG 01 conference*.
- Cuervo, C. (2003). Structural asymmetries but same word order: The dative alternation in Spanish. In A. M. Di Sciullo (Ed), *Asymmetries in Grammar* (pp. 117–144). Amsterdam: Benjamins.
- DeMello, G. (2002). Leísmo in contemporary Spanish American educated speech. *Linguistics*, 40, 261–283.
- Demonte, V. (1995). Dative alternation in Spanish. *Probus*, 7, 5–30.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619.
- Dryer, M. S. (1986). Primary objects, secondary objects, and antidative. *Language*, 62, 808–845.
- Dufter, A., & Stark, E. (2008). Double indirect object marking in Spanish and Italian. In E. Seoane & M. J. López-Couso (Eds), *Theoretical and Empirical Issues in Grammaticalization* (pp. 111–129). Amsterdam: John Benjamins.
- Fienberg, S. E. (2007 [1980]). *The Analysis of Cross-Classified Categorical Data*. New York: Springer.
- Goodman, L. A. (1968). The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. *Journal of the American Statistical Association*, 63, 1091–1131.
- Grimshaw, J., & Mester, A. (1988). Light verbs and theta-marking. *Linguistic Inquiry*, 19, 205–232.

- Harley, H. (2002). Possession and the double object construction. In P. Pica & J. Rooryck (Eds), *Linguistic Variation Yearbook 2* (pp. 31–70). Amsterdam: John Benjamins.
- Hopper, P., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 56, 251–299.
- Jackendoff, R. (1990). On Larson's treatment of the double object construction. *Linguistic Inquiry*, 21, 427–456.
- Kittilä, S. (2007). On the encoding of transitivity-related features on the indirect object. *Functions of Language*, 14, 149–164.
- Kozinsky, I. Š., Nedjalkov, V. P., & Polinskaja, M. S. (1988). Antipassive in Chukchee: Oblique object, object incorporation, zero object. In M. Shibatani (Ed), *Passive and Voice* (pp. 651–706). Amsterdam: John Benjamins.
- Krifka, M. (2004). Semantic and pragmatic conditions for the dative alternation. *Korean Journal of English Language and Linguistics*, 4, 1–32.
- Larson, R. K. (1988). On the double object construction. *Linguistic Inquiry*, 19, 335–391.
- Masullo, P. J. (1992). *Incorporation and Case Theory in Spanish: A Crosslinguistic Perspective*. PhD Dissertation, University of Washington.
- Oehrle, R. (1976). *The Grammatical Status of the English Dative Alternation*. Doctoral Dissertation, Cambridge: MIT.
- Parodi, T. (1998). Aspects of clitic doubling and clitic clusters in Spanish. In R. Fabri, A. Ortman & T. Parodi (Eds), *Models of Inflection* (pp. 85–102). Tübingen: Max Niemeyer Verlag.
- Ransom, E. N. (1979). Definiteness and animacy constraints on passive and doubleobject constructions in English. *Glossa*, 13, 215–240.
- Romero, J. (2013). Accusative datives in Spanish. In B. Fernández & R. Etxepare (Eds), *Variation in Datives: A Microcomparative Perspective* (pp. 283–300). Oxford: Oxford University Press.
- Samek-Lodovici, V. (2003). The internal structure of arguments and its role in complex predicate formation. *Natural Language and Linguistic Theory*, 21, 835–881.
- Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed), *Grammatical categories in Australian languages* (pp. 112–171). Canberra: Australian Institute of Aboriginal Studies.
- Suñer, M. (1991). Two properties of clitics in clitic-doubled constructions. In C.-T. James Huang & R. May (Eds.), *Logical Structure and Linguistic Structure: Cross-Linguistic Perspectives* (pp. 233–251). Dordrecht: Kluwer.
- Suñer, M. (2000). Object-shift: Comparing a Romance language to Germanic. *Probus*, 12, 261–289.
- Torrego, E. (1995). On the nature of clitic doubling. In H. Campos & P. Kempchinsky (Eds), *Evolution and revolution in linguistic Theory* (pp. 399–418). Washington DC: Georgetown University Press.
- Van Geenhoven, V. (1995). Semantic incorporation: a uniform semantics of West Greenlandic noun incorporation and West Germanic bare plural configurations. In A. Dainora, R. Hemphill, B. Luka, B. Need & S. Pargman (Eds), *Papers from the 31st Regional Meeting of the Chicago Linguistic Society* (pp. 171–186). Chicago, IL: Chicago Linguistic Society.
- Van Geenhoven, V. (1998). *Semantic Incorporation and Indefinite Descriptions: Semantic and Syntactic Aspects of Noun Incorporation in West Greenlandic*. Stanford: CSLI Publications.