

Spanish *lo(s)-le(s)* Clitic Alternations in Psych Verbs: A Multifactorial Corpus-Based Analysis

Viola G. Miglio, Stefan Th. Gries,
Michael J. Harris, Eva M. Wheeler, and Raquel Santana-Paixão

1. Introduction: psych verbs and clitic alternations in Spanish

Psych verbs, i.e. verbs that express specific mental states, changes of state or psychological processes, have consistently been an intriguing subject for traditional grammarians and linguists alike (Cuervo 1895, Bello 1898, Belletti & Rizzi 1988, Melis 1999, Vázquez Rozas 2006, 2012 among many others). They constitute a general class, with cross-linguistically identifiable semantic and syntactic characteristics, comprising verbs of feeling (*like, love, appreciate, fascinate*), perception (*see, feel*), cognition (*know, think, believe*), and choice (*decide, prefer, choose*).¹

In Spanish, verbs such as *gustar* 'to like' exhibit a patterning of syntactic and semantic roles that is different from their English equivalents. The English construction casts its arguments as a nominative subject and an accusative direct object, thus English 'to like' as exemplified in (1) seems a typical transitive construction just like 'I eat apples'.

(1) I_{EXP/SUBJ} like students

However, the semantic roles of its participants are not canonical agent and patient arguments, but rather an *experiencer* (the subject, henceforth EXP) and a *theme* or *stimulus* (the object, henceforth STIM). The Spanish equivalent of (1) in (2) capitalizes on the semantic roles of the arguments and casts the theme/STIM – i.e. the *cause* of the psychological state or process taking place in the mind of the EXP – as the syntactic subject agreeing with the verb and the EXP as a dative (henceforth referred to as OBL or DAT) or at times an accusative (ACC) object, in a so-called *inverse construction*.

(2) Me_{EXP/OBL} gustan los estudiantes_{STIM/SUBJ}
'I like students'

EXPs are entities defined as being in a specific state, which change state, or develop a mental activity, and nominative STIMs are entities that are created or mentally represented by the EXP. Verbs of feeling can also be found with transitive constructions (*amar* and *querer* 'to love,' *odiar* 'to hate', *detestar* 'to detest'), but many verbs appear only in an inverse construction of the *gustar*-type.²

Inverse constructions with psych verbs exist in many languages and generally express the cross-linguistic tendency to encode the experience of feelings (such as like, love, fear, detest) in a fundamentally different way from the encoding of canonical transitive actions (such as hit, break, kill). Inverse constructions are thus typologically associated with experiential predicates (Bauer 1996, 2000, Bossong 1998, Dahl & Fedriani 2012). Feelings, emotions, and states are underspecified as to how to encode the arguments involved, hence the possibility to use either a transitive construction, such as (3) or an inverse construction such as (4) to express the same concept:

(3) Yo_{EXP/SUBJ} adoro [ir al cine]_{STIM}
'I love going to the cinema'

* All authors, University of California, Santa Barbara.

1 These are the *ADESSE* definitions from Vázquez Rozas (2006, 2012). Verhoeven (2007), on the other hand, divides psych verbs into the five subclasses of bodily sensation, emotion, cognition, volition, and perception.

2 Vázquez Rozas (2006) lists various verbs, many of which are very commonly used in Spanish: *agradar* 'please,' *apetecer* 'fancy,' *atañer* 'concern,' *competer* 'be responsible for,' *concernir* 'concern,' *convenir* 'be good for' etc.

- (4) Me_{EXP/DAT} encanta [ir al cine]_{STIM}
'I love going to the cinema'

It is generally accepted that experiential predicates are less transitive – and thus entail a number of semantic properties such as less volition, less control, less affectedness of the object – than agentive ones (Dahl & Fedriani 2012:343), and this is reflected in typologically different case markings for experiencers and prototypical agents (AGTs).

Spanish has a considerable number of such inverse construction verbs, such as *agradar* 'please,' *encantar* 'like a lot,' *fascinar* 'like a lot, fascinate,' or *molestar* 'be bothered by,'³ but the situation is further complicated by the fact that some of these inverse-construction verbs (e.g., *gustar* 'like' or *encantar* 'delight') only take a dative EXP, whereas others take both dative and accusative EXPs. The ones taking both ACC and DAT constitute a larger class of verbs, about which Vázquez Rozas (2006, 4.77) says that they can alternate between the two cases without a clear semantic difference.⁴ In the third person singular specifically, the EXP alternates between *le(s)* and *lo(s)/la(s)*. Thus, with verbs such as *aburrir* 'to bore', *asustar* 'to scare', *molestar* 'to bother', the EXP can be either an oblique, such as an IO pronoun as in (5), or an accusative as in a DO as in (6):

- (5) Le_{EXP/OBL} molestan los estudiantes
To him/her bother the students
'S/he is bothered by the students'

- (6) La_{EXP/ACC} molestan los estudiantes
To her bother the students
'She is bothered by the students'

This is in spite of the fact that these verbs appear in dictionaries with the 'transitive' label, thus raising the expectation that they take accusative DOs only (cf. Vázquez Rosas 2006:103). Vázquez Rozas does, however, offer an analysis highlighting some semantic features of the clause related to transitivity that affect the alternation (such as animacy of subject argument, telicity of the predicate, and affectedness of the experiencer). This alternation between case markings is not unexpected: EXPs have an 'ambivalent status' (Lehmann 1991), since they are similar to patients (PAT) in the lack of control they have over the action or its effects, but similar to AGT in retaining the capacity for independent action (Croft 1993, Bossong 1998, Dahl & Fedriani 2012). In turn, STIMs share some characteristics of AGTs in that they cause the change of state or emotion in EXPs. Thus, while inverse experiential constructions are expected to exhibit some alternations in the way EXPs are encoded, what remains to be established more precisely and based on larger and more diverse data sets are the variables that govern this alternation.

An alternation between DAT and ACC forms in clitics does not exclusively affect psych verbs. Originally, the alternation was etymologically inherited from Latin case marking, which in Spanish survives only in the pronominal system. As far as clitics are concerned, specifically in the first and second person singular and plural, ACC and DAT markings are syncretic, thus *me*, *te*, *nos*, *os* could be either ACC or DAT, whereas in the third person the case distinction is preserved (*le(s)* (DAT sg/pl),

3 *Molestar* in Spanish has two main meanings that can be glossed roughly as 'to bother, to bug' (expressing i.e. a more transitive action as in *Hay una mosca que me molesta y no me deja dormir* ('A fly is bugging me and won't let me sleep')) and a more passive psychological state, as in 'to be bothered': *me molestan los ambientes ruidosos*, 'I am bothered by/I dislike noisy environments' – or even 'noisy environments bother me.' The latter is the meaning we are interested in.

4 The complete list from Vázquez Rozas (2006) is: *abrumar* 'overwhelm', *aburrir* 'bore', *admirar* 'admire', *afectar* 'affect', *afligir* 'grieve', *alegrar* 'make happy', *angustiar* 'distress', *apasionar* 'fascinate', *apenar* 'sadden', *asombrar* 'amaze', *asustar* 'scare', *atemorizar* 'frighten', *aterrar* 'terrify', *aterrorizar* 'terrorize', *atormentar* 'torment', *atraer* 'attract', *avergonzar* 'shame', *cansar* 'tire', *complacer* 'please', *consolar* 'console', *convencer* 'convince', *decepcionar* 'disappoint', *deleitar* 'delight', *desanimar* 'dishearten', *descontentar* 'displease', *desconsolar* 'distress', *desesperar* 'exasperate', *disgustar* 'disgust', *distraer* 'amuse', 'distract', *divertir* 'amuse', *emocionar* 'move, touch', *entretener* 'amuse, entertain', *entristecer* 'sadden', *entusiasmar* 'love', *escandalizar* 'scandalize', *espantar* 'scare away', *estorbar* 'bother', *exasperar* 'exasperate', *fascinar* 'fascinate', *fastidiar* 'annoy', *favorecer* 'favor', *halagar* 'flatter', *impresionar* 'impress', *incomodar* 'inconvenience', *inquietar* 'unsettle', *interesar* 'interest', *intranquilizar* 'worry', *intrigar* 'intrigue', *irritar* 'irritate', *maravillar* 'amaze', *molestar* 'bother', *obsesionar* 'obsess', *ofender* 'offend', *pasmar* 'astonish', *perjudicar* 'be bad for', *preocupar* 'worry', *reconfortar* 'comfort', *satisfacer* 'satisfy', *seducir* 'seduce', *sorprender* 'surprise', *tranquilizar* 'calm down.'

lo(s)/la(s) (ACC sg/pl), which is why in our study we limited the analysis to 3rd person clitic pronouns.

Moreover, there also are dialectal differences in the use of clitics in Spanish. The etymological, or case-marking system of the 3rd person is not preserved in cases of phenomena such as *leísmo*, *laísmo*, and *loísmo*, which will be described briefly with regards to the 3rd person singular alternations only. In *leísmo*, *le* refers to both masculine and feminine entities in the DAT (i.e. IOs), but also to the masculine DO (as in *le conocí en la fiesta* 'I met him at the party'). Within the DOs, *le* is most commonly used to refer to masculine, animate entities,⁵ more rarely to things (Fernández Soriano 1993:64). In *laísmo*, the prescriptive feminine DO form *la* is extended to cover also feminine IOs (instead of *le*, as in *la dio la carta* '(s/he) gave her the letter'), and in *loísmo*, the prescriptive masculine DO form *lo* is extended to cover also the masculine IO (also instead of *le*, as in *hoy día lo añaden de todo al vino* 'nowadays they add all sorts of things to wine').⁶

Laísmo and *loísmo*, and – less markedly so – also *leísmo* are typical of peninsular Spanish and not found in Latin America (Roldán 1975:15, Bello 1898:11). Therefore, as Vázquez Rozas warns (2012), the *le-lo* alternation in psych verbs when found in a corpus could be due to the different dialectal provenance of the texts analyzed, rather than to a difference in the semantic features of the pronoun. Since larger and more diverse data were already a general desideratum of future research in this area, such potential dialectal differences point to one particular way in which future work needs to be more diverse, namely by also considering samples of different regional provenances.

Formalist analyses have concentrated either on the syntactic properties of these constructions, using native speakers' intuitions of sentences created *ad hoc*.⁷ Functionalist and typological approaches, on the other hand, have tended to concentrate on the semantic properties of the main arguments, often underlining the markedness of experiential predicates in deviating from the more common, canonical transitive pattern of biactant clauses,⁸ and using naturally occurring corpus data. Our approach is aligned with the latter approach and attempts to identify the semantic properties of the arguments and predicates that determine case-marking alternations in the EXP clitic pronouns in the class of psych verbs, whose EXP alternates between ACC and DAT. However, precisely because we used a large Spanish corpus (Mark Davies's *Corpus del Español*, 2002-) for our analysis, we not only paid attention to different regional varieties but also different genres and their potential effect on the distribution of the pronominal alternations, an aspect that to our knowledge has never previously been considered in the analysis of psych verbs, and specifically of the clitic alternations in Spanish. Thus, we made sure to classify the inverse constructions for genre in our tagging system, and to include genre as a predictor in our model.

The generalizations we attain in this study about main predictors and their interactions are relevant not only for the language specific debate on the Spanish clitic alternation, but they also contribute to a wider corpus of linguistic literature on the typology of transitivity in general, and on case-marking of clausal arguments in particular.

More specifically, this study goes beyond previous work in a variety of respects:

- unlike much previous work, this study is not based on acceptability judgments of isolated sentences but on corpus data from naturally-occurring language from Mark Davies's *Corpus del Español* (2002-, henceforth *CDE*);
- we improve on Vázquez Rozas (2006) by including a larger set of different genres and a much larger set of verbs. Specifically, Vázquez Rozas used only oral and theater data for her quantitative analysis as 'most representative of [...] spontaneous language use' and analyzed nine lexemes (2006:107). Our analysis, by contrast, targets all 55 verbs (out of the 61 verbs identified by Vázquez Rozas as having EXPs alternating between accusative and dative that appeared in the *CDE* corpus in the third person singular and plural, masculine and feminine. In addition, we study four different genres.
- we improve upon Harris et al. (2011) by including random factors in our analysis to take author- and verb-specific preferences into consideration;
- we statistically test all of these predictors from different levels of linguistic analysis at the same time in a multifactorial regression model.

5 This form of *leísmo* is even accepted by the Real Academia de la Lengua Española (cf. *Esbozo*, 1973).

6 Example from Fernández Soriano (1993).

7 See Belletti & Rizzi (1988), Sigurðsson (1989), Masullo (1993), Barðdal & Eypósson (2003, 2009), Gutiérrez Bravo (2006), among others.

8 Melis 1999, Bentley 2006, Di Tullio 2004, Vázquez Rozas 2006, 2012, Dahl & Fedriani 2012.

The overall goal is to determine whether speakers' choices of EXP were indeed as unpredictable as suggested in the literature or influenced by, or at least correlated with semantic factors, genre, and dialect and/or any interactions of the above mentioned factors.

2. Data and methods

We retrieved all occurrences of the psych verbs that were singled out in Vázquez Rozas (2006, ex. 4.77, see note 4 above) and occurred with a 3rd person object clitic pronoun (*le(s)/la(s)/lo(s)*) in the 20.5m contemporary part of the *Corpus del Español* (Davies 2002-). As a result, we obtained a total of 1656 instances, which we annotated for the following characteristics:

- EXPERIENCER: accusative vs. oblique;
- ANIMACYSTIMULUS: yes vs. no;
- CLAUSALSTIMULUS: yes vs. no;
- TENSE: imperfect vs. present vs. perfect vs. preterit;
- MOOD: indicative vs. subjunctive
- GENRE: academic vs. literature vs. news vs. oral.

In addition to the above fixed effects, we also annotated each instance for two random effects:

- AUTHOR: for each instance, we noted the name of the author (180 different author names were retrieved);
- VERB: for each instance, we noted the psych verb (55 different verbs from Vázquez Rozas's list were found).

Thus, an example such as (7) was annotated as shown below:

- (7) Lo que ocurre en la actualidad es que a los novelistas
 what happens in the present-day is that to the novelists-EXP/OBL
- les_{RED} preocupa especialmente la estructura y la palabra [...]
 to-them-EXP/OBL worries especially the structure and the word-STIM
- 'What happens these days is that novelists are mostly worried about the structure and the word [...]'

Annotation: EXPERIENCER: oblique; ANIMACYSTIMULUS: no; CLAUSALSTIMULUS: no; TENSE: present; MOOD: indicative; GENRE: news; AUTHOR: Blanca Berasategui; VERB: *preocupar* 'to worry'

This multifactorial data set was then analyzed statistically with a generalized linear mixed-effects model. That is, our dependent variable was the binary variable EXPERIENCER, our fixed-effects predictors were ANIMACYSTIMULUS, CLAUSALSTIMULUS, TENSE, MOOD, and GENRE, and our random effects were AUTHOR and VERB. We used a stepwise backwards model selection process beginning with a model with all main effects, their pairwise interactions for which decent cell frequencies were observed, and the random effects. From this initial model, we successively removed the least significant predictors until all remaining predictors were either significant themselves or participated in significant interactions (using Likelihood-Ratio tests). In a third step and after this fixed-effects structure was determined, we then tested whether the random effects needed to be retained to arrive at the final minimally adequate model.

3. Results

3.1. Overall results

The summary statistics of our final model revealed a very good fit of the model to the data. This final model is highly significantly correlated with the data ($p < 0.001$), yields a high classification accuracy (81.2%), and, more importantly, even higher C - and D_{xy} -values (0.892 and 0.783 respectively, where 'good' C -values are ≥ 0.8 , following Baayen 2008:2004). Three fixed effects survived the model selection process and are significantly associated with EXPERIENCER:

ANIMACYSTIMULUS, GENRE:TENSE, and GENRE:CLAUSALSTIMULUS.

It is worth pointing out in this context that this final model differs from a previous analysis of this data set (cf. Harris et al. 2011) that did not include the random effects. That previous model, while also yielding a significant correlation between predictors and EXPERIENCER, was characterized by a (i) a much worse classification accuracy (66% correct classifications and $C=0.7$ which is smaller than the usual quality threshold of 0.8) and (ii) a slightly different predictor structure. We will now discuss each significant effect and, where appropriate, relate the results back to those of the previous study.

3.2. The main effect of ANIMACYSTIMULUS

The final model of Harris et al. (2011) also featured ANIMACYSTIMULUS, but only in a significant interaction with CLAUSALSTIMULUS. In our more precise model, ANIMACYSTIMULUS is a significant main effect, whose nature is quite straightforward to describe: when the STIM is animate, then accusatives are more likely, and when it is inanimate, then obliques are more likely. This is represented in Figure 1, which has the predicted probability of oblique EXP on the x-axis, the two levels of ANIMACYSTIMULUS on the y-axis, and the \times represents the relevant average predicted probability of all cases; the heavy line and the error bar represent the standard error and the confidence interval of these averages.

Animacy of stimulus ($p < 0.001$)

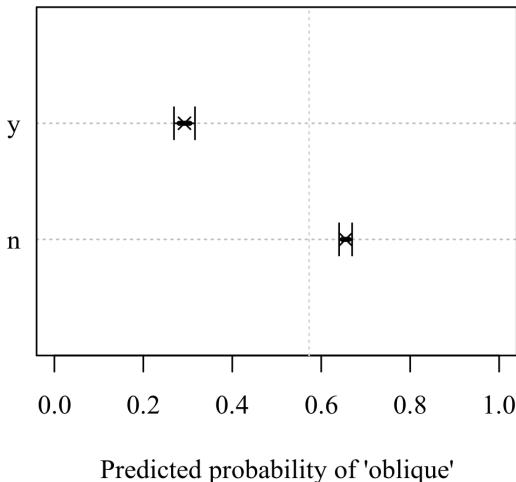


Figure 1: Predicted probabilities of oblique EXPs for ANIMACYSTIMULUS

3.3. The interaction of GENRE:TENSE

In Harris et al.'s (2011) results, GENRE was a significant main effect but in our more precise analysis, GENRE and TENSE interact with each other significantly. In other words, the effect of GENRE is qualified by, or differs across, the different levels of, TENSE. Consider Figure 2, whose two panels show the predicted probabilities of oblique EXP: the left panel shows the genres nested into the tenses; the right panel shows the tenses nested into the genres.

The left panel reveals that the two more atelic tenses imperfect and present behave similarly: literature is characterized by a low probability of oblique EXPs whereas news and oral data are characterized by a high probability of oblique EXPs. A more complex picture is revealed for the more telic tenses perfect and preterit, where literature and oral data prefer ACC and oblique EXPs respectively, but where academic writing prefers ACC with perfect and oblique with preterit; given low cell frequencies the large confidence intervals for the academic genre are a warning not to take the results too literally.

The right panel, on the other hand, reveals that literature and oral data in general have fairly uniform preferences for either accusative or oblique EXPs respectively; also, news data are, on the whole, on the side of preferring oblique EXPs, but the academic writing part of the data behaves extremely heterogeneously.

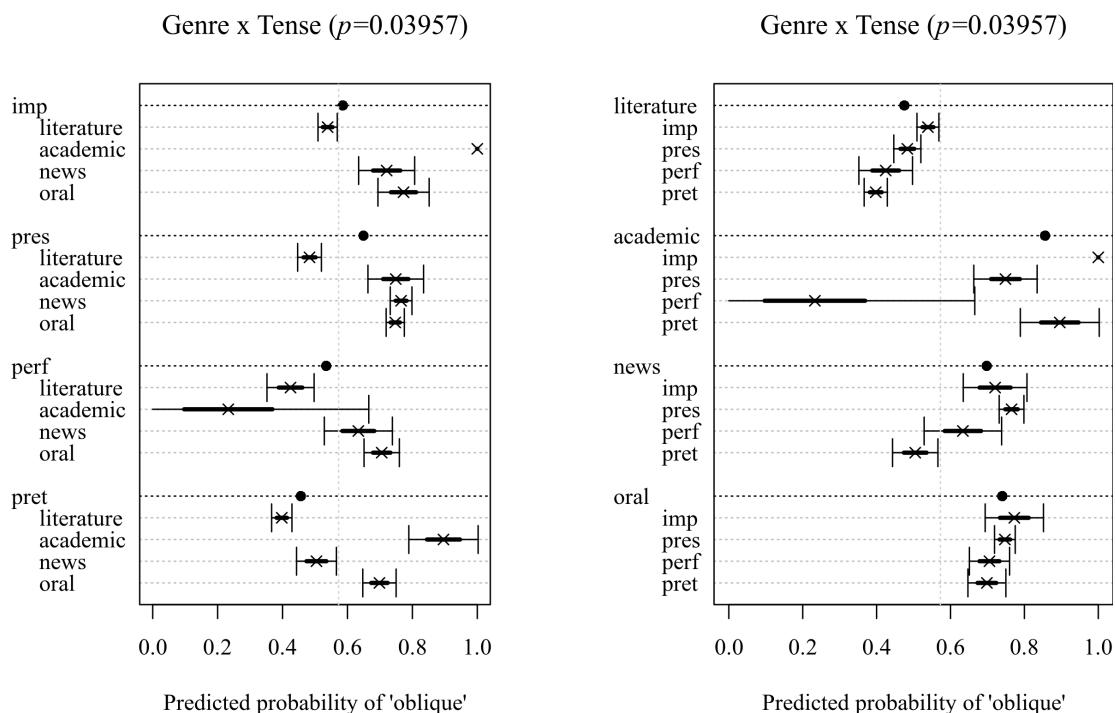


Figure 2: Predicted probabilities of oblique EXPs for GENRE:TENSE

3.4. The interaction of GENRE:CLAUSALSTIMULUS

The final fixed effect is that of GENRE:CLAUSALSTIMULUS, which shows that main effect of Genre is not just qualified by Tense, but also by whether the STIM is clausal or not. Figure 3 reveals several trends. First, oblique EXPs are generally preferred with the main exception of non-clausal STIMs in literature. Second, in all genres but academic writing (which suffers from the same huge variability as before) clausal STIMs prefer oblique EXPs whereas non-clausal ones prefer accusative EXPs. Third, this differential is strong in literature and in news, and still present, but less so in oral data.

3.5. Academic writing as part of the effect of GENRE

The academic register needs a separate comment, as it never patterns with the other genres in any of the analyses we carried out. The anomalous behaviour of the academic genre will be mainly due to two factors. First, the academic genre is represented by a much smaller number of data points than all other genres: our data contained only 56 instances within academic writing while the next most frequent genre – news – already comprised 252 instances, which explains the large confidence intervals.

Second, it is also interesting to see that, even if we disregard the large confidence intervals, not even the average predicted probabilities of the oblique EXPs in the academic data pattern as any other genre. At least part of this discrepancy may be due to the data representing that register in the *Corpus del Español*: academic writing in this corpus is *entirely* made up of the engaging, popular-science encyclopedia MS Encarta and does not include academic writing in the form of university essays, scientific papers etc. A certain idiosyncratic behavior of this part of the data is therefore to be expected; a more detailed analysis would require a comparable corpus, but one in which academic data have been sampled more widely.

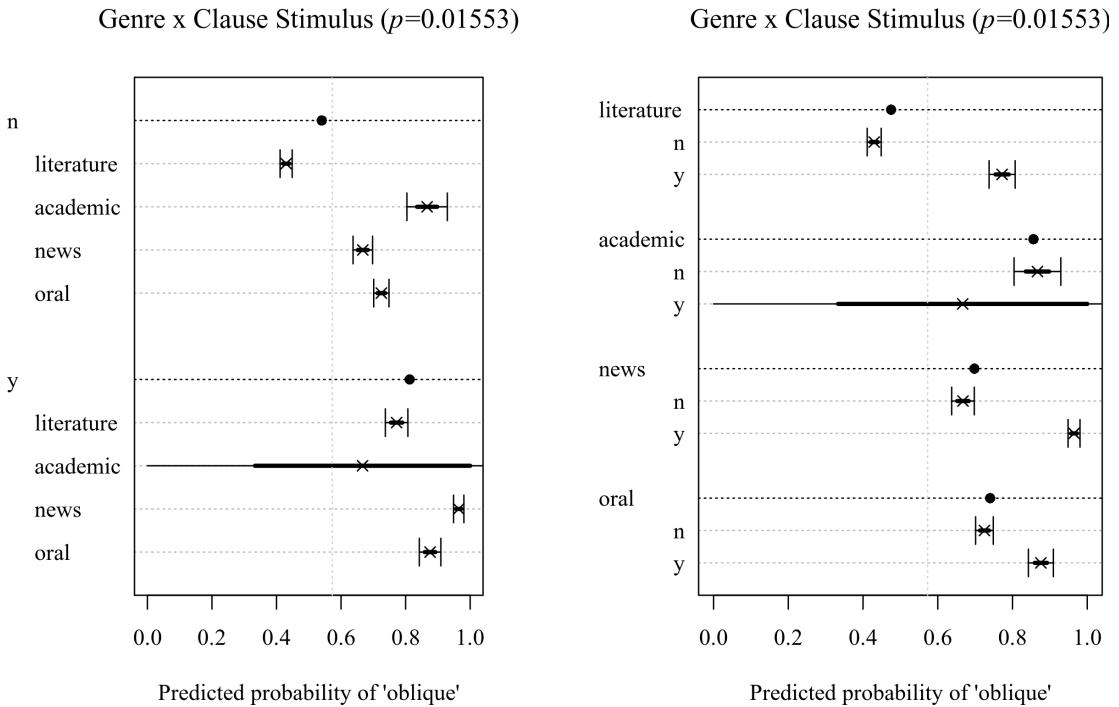


Figure 3: Predicted probabilities of oblique EXPs for GENRE:CLAUSALSTIMULUS

3.6. Further exploration: the random adjustments for authors

As we mentioned above, the minimally adequate model also contained two random effects, to which we turn now; first, the authors. A visual inspection of the random effect of AUTHOR revealed stark differences between authors in terms of their EXP choices. To explore this variation in more detail, we classified the authors into eight regions of origin: North America, Central America, South America, Rio de la Plata, Caribbean, Spain, Europe, and unknown. Then, we checked to what degree the eight different geographical regions differed in terms of the predicted EXP choices. A Kruskal test shows that much of the author variability boils down to regional differences ($\chi^2=3.44$, $df=7$, $p<0.001$), which is represented in Figure 4.

Authors from Spain or the rest of Europe are characterized by high probabilities of oblique EXPs, which may follow from the accepted *leísmo* of standard peninsular Spanish, i.e. the use of *le* for animate/definite/masculine DOs (Bello 1898, Roldán 1975, Vázquez Rozas 2012). For much of the Americas, we find a north-south cline: as one goes from North America to South America, the use of ACC appears to grow, and the Caribbean as well as the Rio de la Plata regions are similar in particular to Central America. Obviously, regional varieties⁹ differ considerably in terms of their preferences, which is nicely revealed in the present multifactorial methodological context.

⁹ An anonymous reviewer asked how the dialectally-related metadata was obtained, since these metadata are not easily retrievable from the *CdE*. *CdE* does code for author, so for each construction analyzed, we noted the name of the author and title of the work (180 different author names were retrieved from the *CdE* metadata); then a manual, biographical search for each author (or author+title of the work) was carried out to establish where the author was born and grew up. This information provided us with informed judgements as to the authors' most likely variety of Spanish.

Author region ($p < 0.001$)

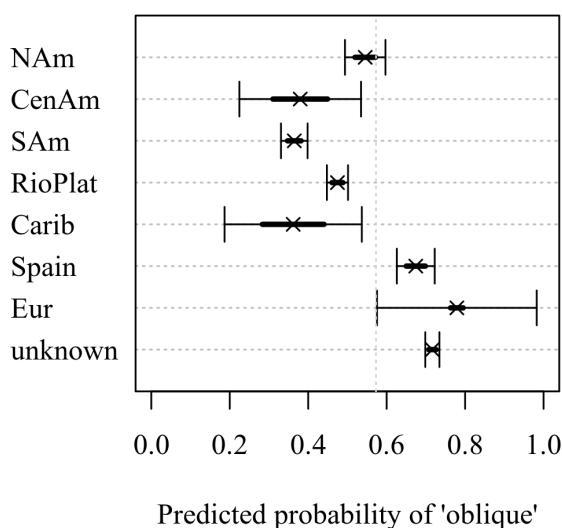


Figure 4: Predicted probabilities of oblique EXPs for authors' different geographical origins

4. Discussion and concluding remarks

4.1. Linguistic implications

As has become obvious, the assumption that there is no systematic patterning in the alternations is not borne out all. On the contrary, even the small number of factors we included in our analysis makes for a very good classification accuracy of ACC vs. DAT/OBL experiencer choice. In addition, several of the hypotheses suggested in the literature were generally borne out.

4.1.1. The effect of Animacy

We find a strong effect of ANIMACY in the analysis of the psych verbs studied here such that animate and inanimate STIMs prefer accusative and oblique EXPs respectively. Although not referring specifically to psych verbs, other authors had mentioned the role of animacy, from traditional grammarians such as Bello and Cuervo (Bello 1898, Cuervo 1895) to Lapesa observing that dative marking on the objects of some Latin verbs assigning lexical case had extended to other verb classes when the pronominal referents were [+human], and that this hybrid and ultimately just referential system had been finally adopted by peninsular Spanish (1993), to modern linguists such as Klein-Andreu (1993), and Fernández Soriano (1993).

The fact that ANIMACY is a significant predictor is also compatible with previous studies such as Vázquez Rozas (2006, 2012), who emphasized the importance of semantic factors related to transitivity in the *le-lo* alternation in Spanish psych verbs, following Hopper and Thompson's model (1980). Our psych verbs are typically stative (see also Islas Canales 2004:248, 254), i.e. very low in transitivity, but if the STIM causing the emotion in the EXP is animate, it becomes more similar to an AGT and thus raises the level of transitivity of the clause. In turn, the EXP will be construed more along the lines of a PAT – after all it is affected, and typically has no control over the emotion experienced. In this case, because of the increased transitivity of the clause, the EXP is more likely to be expressed by the ACC clitic *lo(s)*.

4.1.2. The effect of Genre:Tense

The interaction GENRE:TENSE reveals the expected cline of oblique preferences for three of the four genres: in literature, news, and oral texts we find that the occurrence of typically atelic tenses such as present and imperfect correlates with a high likelihood of oblique EXPs. This is especially true in the oral data, and seems to reflect the fact that interlocutors do attribute sensations and emotions to people

they talk about, even if verbs of emotions reflect a mental, and thus very personal, internal activity. This result, however, contradicts statements in the literature that EXPs are typically 1st and 2nd person, because after all 'speakers do not claim to feel what another individual is feeling' (Mithun 1991:522).

Our findings show that in inverse constructions, where transitivity is reduced, 3rd person EXPs are found despite infringing the 'animacy hierarchy,' a semantic scale that predicts that first persons will be located at the maximum pole of agentivity, followed by second and third person; cf., e.g., Comrie (1989). This finding can still be explained because agentivity is not the only axis along which Spanish psych verbs can be organized – another important dimension is *empathy* (Vázquez Rozas 2012:851): speakers use these constructions to report about the feelings and emotions of 3rd person EXPs, which makes them particularly common in the oral genre.

It should be highlighted how the data in the 'literature' genre exhibit extremely low frequencies of OBL EXPs and are therefore clearly distinct from the results in the 'oral' section of the corpus. These results warrant a first methodological conclusion, namely a warning against using literary texts for linguistic analysis as if they were sources of spontaneous speech: the literary register here behaves quite differently from all the others.

4.1.3. *The effect of Genre: Clausal Stimulus*

The interaction `GENRE:CLAUSALSTIMULUS` reinforces the observation that inanimate STIMs prefer oblique EXPs: clausal STIM are prototypically inanimate, which in turn relates back to transitivity (cf. above). This preference is very similar in oral and news and it is even stronger in literature. Much of the lesser 'obliqueness' of EXPs in the 'literature' genre is due to the non-clausal STIMs, which are rather the actors/participants in the stories told.

4.2. *Methodological implications*

Apart from the above results, there are also some methodological implications. As we already mentioned, the results are a clear warning to not neglect genre/register variation and, thus, to also not assume that one particular genre (such as literature) can represent another one (such as spoken conversation). This fact cannot be highlighted enough: there is still much corpus work where either `GENRE` is not entered into models as a predictor or where `GENRE` is entered as a predictor but not in interactions with all other predictors: to uncover genre effects or just to be certain that once genre choices were useful/revealing, one has to add interactions such as `GENRE:TENSE` or `GENRE:CLAUSALSTIMULUS` in this study.

An additional important lesson to be learnt from the comparison of this analysis to the previous one (Harris et al. 2011) is the huge gain in accuracy resulting from the inclusion of random effects in linguistic analysis. Not only do such models deal better with imbalanced data and dependent data points, the exploration of random effects can also give rise to new hypotheses (cf. the results on regional distinctions).

4.3. *Conclusions*

We found that clitic choice in psych verbs supporting ACC-OBL clitic alternation can be predicted well from semantic factors related to transitivity and others. Of the semantic factors analyzed, `ANIMACY` seems to be the main one governing the alternations of Spanish 3rd person clitics: The animacy of the STIM increases its agentivity and in turn demotes the EXP to an argument closer to a PAT. By invoking these semantic roles, many other related characteristics found in the literature as having a cross-linguistic effect on ACC-OBL case marking choice in general can also be explained, such as loss of control/volition over the event, a higher affectedness of the EXP, and telicity. Manipulating these factors through clitic choice (*le(s)-lo(s)* or *las(s)*), speakers essentially vary the degree of transitivity of the clause, making even tendentially stative verbs such as psych verbs more or less transitive (in the sense of Hopper and Thompson 1980).

As expected, we also find that dialectal differences correlate with high or low occurrences of OBL EXPs, depending on whether authors are from Spain or Latin America respectively. This has been underlined by many, but what is unusual and worthy of further exploration is the decrease in frequency of OBL EXPs along the axis North → Central America/Caribbean → South America. To the best of

our knowledge this type of dialectal distinction in the usage of 3rd person clitics has not been detected before in Latin American Spanish.

As for the importance of genre, the prediction of a cline topped by the oral section of the corpus with highest frequencies of OBLs was borne out in the order: literature, news, oral. This mirrors findings for 1st and 2nd person EXPs (Mithun 1991, Vázquez Rozas 2006), but unlike Mithun's view and since we analyzed only 3rd person clitics, we found that speakers do not only speak about their own feelings using inverse constructions, but that they use these constructions also when they attribute sensations and emotions to *other* people they talk about.

Our study's conclusions about main predictors and their interactions (the importance of animacy and agentivity, dialectal differences in pronominal use along lines of areal distribution different from those suggested in previous literature, the contribution of different genres to the distribution of the pronouns) are clearly relevant not only for the specific study of the Spanish clitic alternation, but also for a broader typological literature on transitivity in general, and on case-marking of clausal arguments in particular.

References

- Barðdal, Jóhanna & Thórhallur Eythórsson. 2003. The Change that Never Happened: The Story of Oblique Subjects. *Journal of Linguistics* 39(3). 439-472.
- Barðdal, Jóhanna & Thórhallur Eythórsson. 2009. The origin of the oblique-subject construction: an Indo-European comparison. In Vit Bubenik, John Hewson, & Sarah Rose (eds). *Grammatical Change in Indo-European Languages*. Amsterdam & Philadelphia: John Benjamins. 179-93.
- Bauer, Brigitte L. M. 2000. *Archaic Syntax in Indo-European. The Spread of Transitivity in Latin and French*. Berlin & New York: Mouton de Gruyter.
- Belletti, Adriana & Luigi Rizzi. 1988. Psych-Verbs and θ -Theory. *Natural Language & Linguistic Theory* 6(3). 291-352.
- Bello, Andrés. 1898. *Gramática de la lengua castellana destinada al uso de los americanos*. Paris: A. Roger and F. Chernoviz Publishers.
- Bentley, Delia. 2006. *Split intransitivity in Italian*. Berlin & New York: Mouton de Gruyter.
- Bossong, Georg. 1998. Le marquage de l'experient dans les langues de l'Europe. In Jack Feuillet (ed.). *Actance et Valence dans les Langues de l'Europe*. Berlin: Mouton de Gruyter. 259-294.
- Brewer, William B. 1970. Extent of Verbal Influence and Choice between *le* and *lo* in Alphonsine Prose. *Hispanic Review* 38(2). 133-146.
- Cuervo, Rufino José. 1895. Los casos enclíticos y proclíticos del pronombre de tercera persona en castellano. *Romania* 24. 95-113, 219-263.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. Chicago: University Press.
- Croft, William. 1993. Case marking and the semantics of mental verbs. In James Pustejovsky (ed.), *Semantics and the Lexicon*. 55-72. Boston & London: Kluwer.
- Dahl, Eystein & Chiara Fedriani. 2012. The argument structure of experience: experiential constructions in Early Vedic, Homeric Greek, and Early Latin. *Transactions of the Philological Society*. 110(3). 342-362.
- Davies, Mark. (2002-) *Corpus del Español: 100 million words, 1200s-1900s*. Available online at <http://www.corpusdelespanol.org>.
- Di Tullio, Ángela. 2004. Los verbos psicológicos y la estatividad: realizaciones del español. *Cuadernos de Lingüística del Instituto Universitario Ortega y Gasset*, 11. 23-43.
- Esbozo de una nueva gramática de la lengua española*. 1973. Madrid: Real Academia de la Lengua Española.
- Fernández Soriano, Olga (ed.). 1993. *Los pronombres átonos*. Madrid: Taurus Universitaria/ Santillana.
- García, Erica C. 1975. *The Role of Theory in Linguistic Analysis: The Spanish Pronoun System*. Amsterdam: North Holland Publishing Company.
- Gutiérrez-Bravo, Rodrigo. 2006. A Reinterpretation of Quirky Subjects and Related Phenomena in Spanish. In C. Nishida & Jean-Pierre Y. Montreuil (eds.). *New Perspectives on Romance Linguistics: Vol. I: Morphology, Syntax, Semantics, and Pragmatics. Selected papers from the 35th Linguistic Symposium on Romance Languages (LSRL)*. 127-142. Amsterdam & Philadelphia: John Benjamins.
- Harris, Michael J., Raquel Santana Paixão, Eva Wheeler, & Viola G. Miglio. 2011. Accusative-oblique alternations in Spanish clitics. Paper presented at the American Association of Corpus Linguistics Conference (AACL 2011). Atlanta, GA. October 8th, 2011.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language*, 56(2). 251-299.
- Islas Canales, Martha R. 2004. *Verbos de emoción y aktionsart*. Ph.D. Dissertation, Mexico City: El Colegio de México. Available at <<http://consulta.telecom.cide.edu/conacytfiles/TCONACYT200572.pdf>>.

- Klein-Andreu, Flora. 1993. Distintos sistemas de empleo de «le», «la», «lo». Perspectiva sincrónica, diacrónica, y sociolingüística. In Fernández Soriano, Olga (ed.). 1993. *Los pronombres átonos*. 337-354. Madrid: Taurus Universitaria/Santillana.
- Lehmann, Christian. 1991. Predicate classes and PARTICIPATION. In Hansjakob Seiler & Waldfried Premper (eds.), *Partizipation: Das Sprachliche Erfassen von Sachverhalten*. 183-329. Tübingen: Gunter Narr.
- Lapesa, Rafael. 1993. Sobre los orígenes y evolución del leísmo, laísmo y loísmo. In Olga Fernández Soriano(ed.), *Los pronombres átonos*, 313-36. Madrid: Taurus Universitaria/Santillana.
- Masullo, Pascual J. 1993. Two types of quirky subjects: Spanish versus Icelandic. *Proceedings of Nels* 23. 303-317. Amherst: University of Massachusetts Press.
- Melis, Chantal. 1999. Variación sintáctica con los verbos de emoción. *Español Actual*, 71. 49-62.
- Mithun, Marianne. 1991. Active/agentive case marking and its motivations. *Language* 67(3). 510-546.
- Roldán, Mercedes. 1975. The Great Spanish *le-lo* Controversy. *Linguistics* 147. 15-30.
- Sigurðsson, Halldór A. 1989. *Verbal syntax and case in Icelandic*. Ph.D. Dissertation, Lund University.
- Vázquez Rozas, Victoria. 2006. Gustar-type verbs. In J. Clancy Clements & Jiyoung Yoon (eds.), *Functional Approaches to Spanish Syntax. Lexical semantics, discourse and transitivity*, 80-114. Hampshire/New York, Palgrave MacMillan.
- Vázquez Rozas, Victoria. 2012. Construyendo emociones: sintaxis, frecuencia y función comunicativa . In T. Jiménez, B. López Meirama, V. Vázquez Rozas, & A. Veiga (eds.). *Cum corde et in nova grammatica. Estudios ofrecidos a Guillermo Rojo*, 841-854. Santiago de Compostela: USC Editor.
- Verhoeven, Elisabeth. 2007. *Experiential Constructions in Yucatec Maya: A Typologically Based Analysis of a Functional Domain in a Mayan Language*. Amsterdam & Philadelphia: John Benjamins.

Selected Proceedings of the 16th Hispanic Linguistics Symposium

edited by Jennifer Cabrelli Amaro,
Gillian Lord, Ana de Prada Pérez,
and Jessi Elana Aaron

Cascadilla Proceedings Project Somerville, MA 2013

Copyright information

Selected Proceedings of the 16th Hispanic Linguistics Symposium
© 2013 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-459-1 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Miglio, Viola G., Stefan Th. Gries, Michael J. Harris, Eva M. Wheeler, and Raquel Santana-Paixão. 2013. Spanish *lo(s)-le(s)* Clitic Alternations in Psych Verbs: A Multifactorial Corpus-Based Analysis. In *Selected Proceedings of the 16th Hispanic Linguistics Symposium*, ed. Jennifer Cabrelli Amaro et al., 268-278. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2939.