

Diachronic Null Subject Use across Latin American Spanish: Comparing Corpora¹

It has been widely noted that varieties of Latin American Spanish such as Dominican Spanish (Toribio 2000) increasingly allow both overt and null subject pronouns in the same contexts (opposed to the use of the same term to describe 3rd person referential split in earlier literature), resulting in their labelling as partial null subject languages (PNSLs). Their existence has challenged the NSP as formulated by Rizzi (1982, 1986) and suggested the occurrence of null subject reduction. Given Latin American Spanish's rich history of language contact over the last five centuries, the observed decrease in use of null subjects is likely a result of contact-induced simplification. Following Trudgill's (2011) sociolinguistic typology, this contact would be characterized as short-term adult second-language acquisition which is consistent with the scenario for the African slaves brought over to the Americas by the Spanish from the 16th century onward. Indeed, there are numerous non-standard varieties of Spanish spoken by Afro-Hispanic communities throughout Latin America. These Afro-Hispanic Languages of the Americas (AHLAs) have been described as the result of "conventionalized advanced second languages," much in line with Trudgill's proposal (Sessarego 2013, 2017). The L2-difficulty of the rules surrounding null subject use would have led adult learners to overproduce overt subject pronouns. Their errors would have then become nativized in the speech of the next generation. Similar to the Jespersen cycle for negation, the partial status of null subjects in varieties such as Dominican Spanish could reflect language in change as the variety moves from a null subject language to a non-null subject language á la van Gelderen's (2011) Subject Cycle. What is needed then is a robust diachronic analysis of null subjects.

To achieve this, I constructed a corpus of 57 texts from eight countries over four centuries to evaluate the diachronic rate of null subjects in Latin American Spanish. The corpus is supplemented by transcriptions of previous fieldwork in current AHLA varieties as well as historical texts written in AHLA vernacular. The aim was/is to determine whether the corpus supports the hypothesis that the rate of overt subject pronouns has risen significantly since contact with African L2-speakers. Broadly, we expected to see the highest variation in the form of more overt subject pronouns in varieties with the highest Afro-Hispanic populations. However, despite an attempt to balance the corpus by genre (equal literary and non-literary texts), preliminary data from the Dominican Republic, Bolivia, Panama, and Spain (as a control) has shown much variation, leaving no clear diachronic patterns (Figure 1). In order to see if something similar to but independent of genre could be affecting the pronominal data, Rosemeyer's (2019) method of measuring orality levels in Portuguese plays was adapted for use in this corpus. This ORSCORE was found to have a significant ($p=.001$) relationship with overtness rates (Figure 2²). With a view to account for this obstruction, we set out to compare forms of the private verbs *creer* 'to believe/think' and *pensar* 'to think' which should lessen the influence of orality as its frequency in a text is one of the five criteria used to measure orality. Unfortunately, our corpus is too small to restrict the data to one lexical item, so we are now in the process of sifting through data from the *CORDIAM* and *CORDE/CREA* corpora. We hope to tease apart this orality effect to find underlying diachronic trends.

¹ For consideration as a talk or poster.

² There are two outlying texts with very high ORSCORES (1.77 and 2.35) that might seem to skew the data. However, the best-fit line remains the same when they are excluded.

Figure 1:³ Null vs. Overt Subject Pronouns

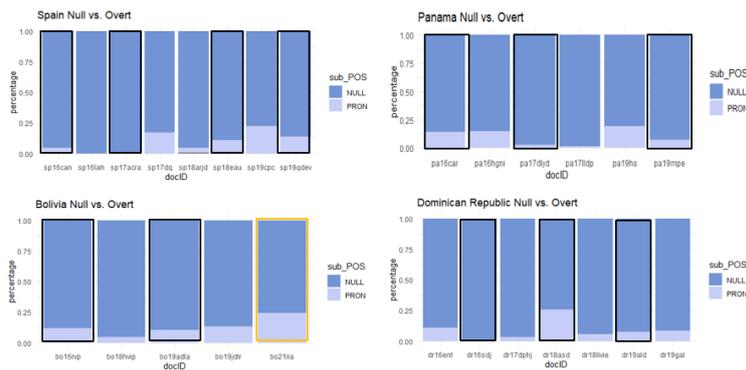
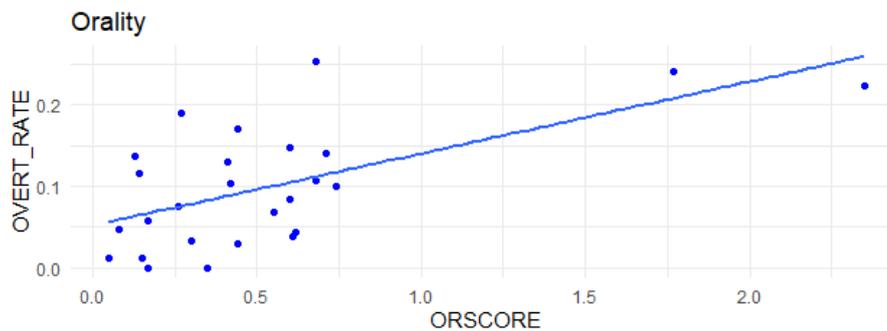


Figure 2: Degree of orality plotted against proportion of overt subject pronouns per text



References:

[CORDIAM] Academia Mexicana de la Lengua, *Corpus Diacrónico y Diatópico del Español de América*, www.cordiam.org

Gelderen, Elly van. 2011. *The Linguistic Cycle*. Oxford: OUP.

REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español*. <<http://www.rae.es>>

REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <<http://www.rae.es>>

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.

Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17: 501–57.

Rosemeyer, Malte. 2019. "Actual and apparent change in Brazilian Portuguese wh-interrogatives." *Language Variation and Change* 31(2): 165–191. CUP.

Sessarego, Sandro. 2013. "Afro-Hispanic Contact Varieties as Conventionalized Advanced Second Languages". *IBERIA* 5 (1): 99-125.

Sessarego, Sandro. 2017. "Revisiting the Null Subject Parameter: New Insights from Afro-Peruvian Spanish". *Isogloss* 3 (1): 43-68.

Toribio, Almeida J. 2000. "Setting parametric limits on dialectal variation in Spanish". *Lingua: International Review of General Linguistics* 110 (5): 315–341.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: OUP.

³ The docID corresponds to the country, century, and title of a text (e.g. dr16ent = Dominican Republic, 16th century, Entremés). The document genre is bolded, and the golden bolded text is a transcript from Afro-Bolivian vernacular.