Null subjects in Portuguese and the typology of conditioning on variation
Gregory R. Guy – New York University

Variable subject pronoun expression (SPE) – the alternation between null and overt subject pronouns – is conditioned by a variety of factors in 'null subject languages' such as Spanish and Portuguese (Otheguy & Zentella 2010, Duarte 2020). These factors include all the types of conditioning proposed by Tamminga, MacKenzie and Embick (2016). They distinguish *social* or *s*-conditioning, *internal linguistic* or *i*-conditioning, such as morphological structure, and *psychophysiological* or *p*-conditioning, such as priming. The i-conditions can participate in the rules of grammar, and along with social conditions, can vary between languages and dialects, while p-conditions are postulated to derive from the physical, neurological, and cognitive properties of the human body and language faculty, and hence should be cross-linguistically constant or similar. The constraints on SPE provide a test of this typology: do the three types adequately characterize the conditioning of SPE, and are the p-conditions constant across dialects and speech communities?

This paper examines the distinction between s-, i- and p- conditioning by comparing constraint effects across four dialects of Portuguese: Lisbon, from mainland Portugal, Madeira island in the Atlantic, and São Paulo and Rio de Janeiro in Brazil. The Brazilian Portuguese (BP) dialects differ dramatically from European Portuguese (EP) in SPE rates – twice as high in Brazil as in Portugal (see Table 1), and all four are distinguished by phonological and other features.

Following the typology of Tamminga et al., we consider priming as a p-condition. As predicted it has a consistent effect across dialects – nulls favor subsequent nulls, overt pronouns favor overts. Clause type – main vs. subordinate – is an i-condition, and is also consistent across dialects: subordinate clauses favor SPE. Social conditioning is modest – no significant effect of education or age, a gender difference in all communities (women use more overt forms than men), but there is an interaction between age and gender in Lisbon. The biggest social distinction in the data is dialectal: the great difference in rates between Brazil and Portugal.

Questions arise with respect to two other constraints. Reference continuity – is the subject referent the same or different as the previous clause – is a discourse-level phenomenon, which is not addressed by Tamminga et al. It is one of the strongest conditions on SPE with a systematic effect in our communities and many other studies, such that same referents favor nulls, and switched referents favor overt pronouns. It thus behaves like a p-condition, possibly implying a general cognitive system for tracking potential referents in discourse.

Person/number might appear to be a straightforward i-condition, but EP makes the T/V distinction between familiar and formal 2nd person forms, requiring a social calculation over pronoun choice in direct address. The person/number constraint has also been explained in functional terms, which is not addressed in the Tamminga et al. typology. Distinctive verbal inflections (like 1st plural -*mos* in Portuguese) make overt pronouns redundant, while inflections with greater ambiguity may require more overt pronouns to disambiguate. The 3rd singular verbal inflection in Portuguese is used with reference to 2nd sg *você*, 3rd sg *ele/ela*, and 1st plural *a gente*. As Table 2 shows, the four dialects are relatively inconsistent for this constraint. Socially motivated differences between communities are evident for 2nd person forms and for the new pronoun *a gente*. The distinctive 1st pl forms disfavor SPE, per the functional hypothesis, but the equally distinctive 1st sg strongly favors SPE. We conclude that the Tamminga et al. typology is partially sustained by these findings but leaves some important questions unaddressed. Social factors may impact i- and p-conditions in unexpected ways.

Speech community	Number of Speakers	Social dimensions	Tokens	% overt pronouns
São Paulo (SP2010 corpus, USP)	44	2 ed., 3 age, 2 gender	15895	66.2
Rio de Janeiro (Comparaport corpus, UFRJ)	36	3 ed., 3 age, 2 gender	9776	64.6
Lisbon (Comparaport, UFRJ)	36	3 ed., 3 age, 2 gender	9746	32.0
Funchal (Comparaport, UFRJ)	18	3 ed., 3 age, 2 gender	4602	32.3
Total	101		40,019	

Table 1. The corpus

Person/ number	SP	RJ	LIS	FUN
1sg	.48	.52	.54	.53
2sg	---	---	.45	.21
você	.71	.59	.27	.24
3sg	.43	.40	.48	.62
a gente	.70	.73	.92	.88
1pl	.26	.27	.41	.34
vocês	---	---	.76	.42
3pl	.35	.32	.35	.33

Table 2. Person/number conditioning of SPE in four dialects of Portuguese.
Relative weights.

References
Duarte, M.E.L. 2020. Sociolinguistics as a powerful tool to follow the course of a parametric change. *PWPL* 25(2):3
Otheguy, R. and A.C. Zentella 2012. *Spanish in New York*. Oxford University Press
Tamminga, M., L. MacKenzie & D. Embick, 2016. The dynamics of variation in individuals. *Linguistic Variation* 16: 300-336.