Is there a clear division of labour in Spanish? Corpus approaches to test the Position of Antecedent Strategy

Teresa Quesada & Cristóbal Lozano (Universidad de Granada)

The Position of Antecedent Strategy (PAS) (Carminati 2002) is a purely structural parsing strategy where the syntactic function of the antecedent determines the form of the anaphor. Carminati proposed that, in Italian, null pronouns tend to select subject antecedents, whereas overt pronouns typically corefer with non-subject antecedents. The PAS has been extensively studied experimentally in native and L2 Italian and other null-subject languages like Spanish (e.g., Alonso-Ovalle et al. 2002, Bel & García-Alcaraz 2015, Bel & al. 2016, Filiaci et al. 2014, Jegerski et al. 2011, Sorace & Filiaci, 2006), as in *Juan$_i$ pegó a Pedro$_j$. Él$_j$/Ø$_i$ está enfadado*.

If we focus on Spanish data, advanced and near-native learners of Spanish typically show certain deficits when processing PAS, arguably as a result of their limitations when integrating simultaneously syntactic information (overt/null alternation) with discourse information (topic/focus) at the syntax-discourse interface, as predicted by the Interface Hypothesis, IH (Sorace 2011). Importantly, these studies are experimental in nature and have explored PAS in decontextualised and unnatural scenarios.

This study explores the PAS in a corpus, as production data offers natural and contextually richer scenarios. A sample of intermediate and near-native L1 English – L2 Spanish learners plus a control group of Spanish natives was selected from CEDEL2 (Corpus Escrito del Español L2) (Lozano & Mendikoetxea 2013). This sample was manually annotated with a tagset (Fig. 1) in the UAM Corpus Tool tagging software following a fine-grained tagset. Some of those results reveal the following:

1. Near-native learners behave similarly to Spanish natives in terms of the PAS, as both advanced groups produce mainly a null subject pronoun (and not an overt pronoun) to refer to a subject antecedent, except for the intermediate group that shows variability. Therefore, corpus data confirm previous experimental findings (Fig. 2).

2. However, regarding overt anaphoric forms, LCR methodology reveals that the anaphoric choices for non-subject antecedents are more complex than previously assumed. Importantly, it is not only overt forms (e.g., *él*) but also NPs (e.g., *Pedro/el hombre*) that show a strong bias towards antecedents in non-subject position in natives (Fig. 3), a fact that has gone undetected in previous experimental work. The division of labour in Spanish is between null pronouns vs. overt material (=overt pronouns & NPs). Advanced learners do not show a clear bias with NP forms, as was the case for overt pronominal forms.

3. In short, intermediates show variability and even near-natives fall short of attaining complete native-like knowledge at the syntax-discourse interface.

Corpus methods therefore reveal that there are additional factors that have gone undetected in previous experimental studies. Results also suggest that full native-like competence is not attainable in very advanced levels, which supports the IH predictions.

Carminati, M. N. (2002). The processing of Italian subject pronouns. PhD thesis: University of Massachusetts at Amherst.

Alonso-Ovalle, L., Fernández-Solera, S., Frazier, L., & Clifton, C. (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. Journal of Italian Linguistics, 14(2), 151–169.

Bel, A., & García-Alcaraz, E. (2015). Subject pronouns in the L2 Spanish of Moroccan Arabic speakers. In T. Judy & S. Perpiñán (Eds.), The Acquisition of Spanish in Understudied Language Pairings (pp. 201–232). Amsterdam: John Benjamins.

Bel, A., Sagarra, N., Comínguez, J. P., & García-Alcaraz, E. (2016). Transfer and proficiency effects in L2 processing of subject anaphora. Lingua, 184, 134–159.

Filiaci, F., Sorace, A., & Carreiras, M. (2014). Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. Language, Cognition and Neuroscience, 29(7), 825–843.

Jegerski, J., VanPatten, B., & Keating, G. D. (2011). Cross-linguistic variation and the acquisition of pronominal reference in L2 Spanish. Second Language Research, 27(4), 481–507.

Lozano, C., & Mendikoetxea, A. (2013). Learner corpora and second language acquisition: the design and collection of CEDEL2. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), Automatic Treatment and Analysis of Learner Corpus Data. (pp. 65–100). Amsterdam: John Benjamins.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. Linguistic Approaches to Bilingualism, 1(1), 1–33.
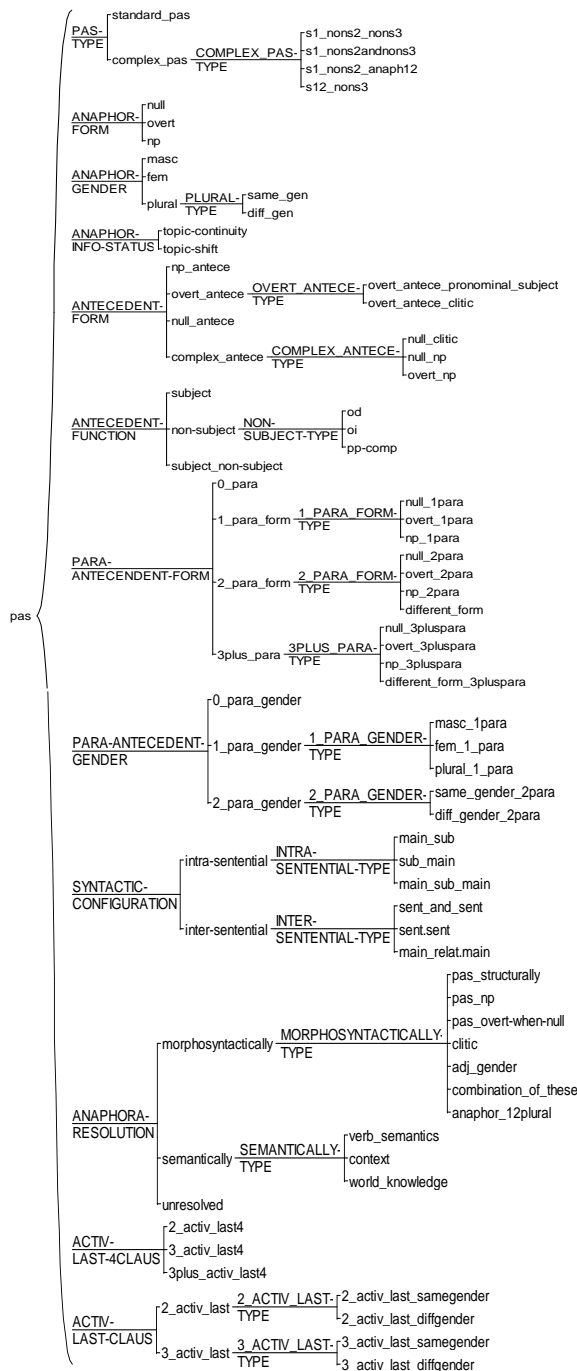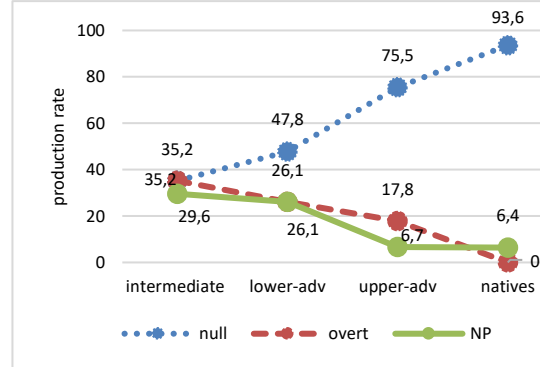
Figure 1: Tagset

Figure 2: Subject-antecedent results

Figure 3: Division of labour of anaphoric forms (natives)