

## **Anaphora resolution in L1 English-L2 Spanish in the CEDEL2 corpus: beyond experimental evidence**

Anaphora Resolution (AR) relates to how overt and null pronominal subject anaphors corefer with their antecedents. A particular instance of AR is the Position of Antecedent Strategy (PAS), a purely structural parsing strategy where the syntactic function of the antecedent determines the form of the anaphor: Null pronouns tend to select subject antecedents to mark topic continuity, whereas overt pronouns typically corefer with non-subject antecedents to mark topic shift, (1)

- (1) a. Juan<sub>i</sub> pegó a Pedro<sub>j</sub>. Él<sub>j</sub>/Ø<sub>i</sub> está enfadado.  
b. Lola<sub>i</sub> vio a Eva<sub>j</sub> mientras ella<sub>j</sub>/Ø<sub>i</sub> jugaba en el parque.

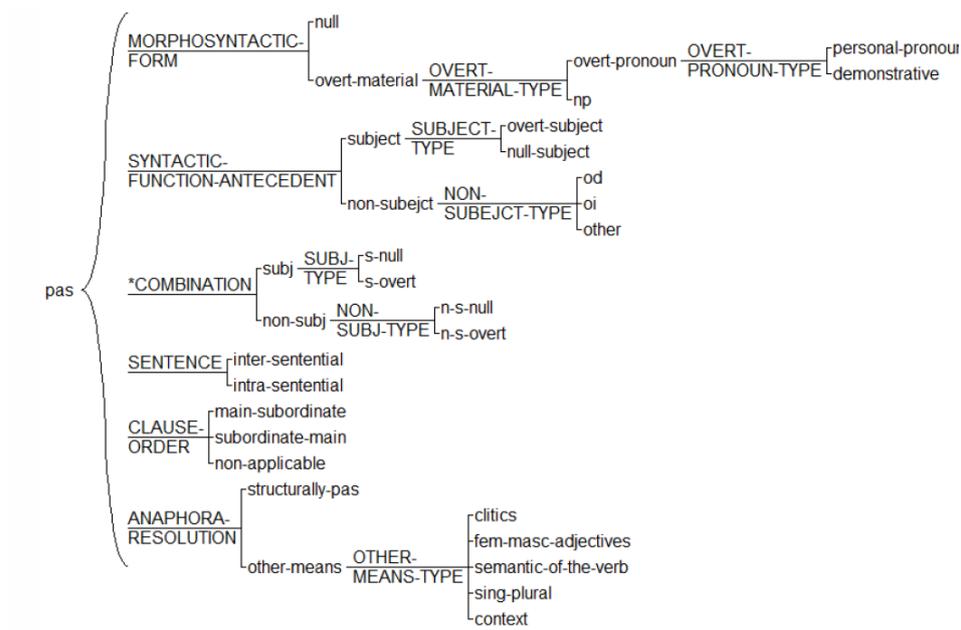
The PAS has been extensively studied experimentally in both native and L2 Spanish (Alonso-Ovalle et al. 2002, Bel & García-Alcaraz 2015, Bel & al. 2016a, 2016b, Filiaci 2010, Filiaci et al. 2014, García-Alcaraz 2015, Jegerski et al. 2011, Keating et al. 2011). In particular, advanced and near-native 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 current theories of bilingual language acquisition like the Interface Hypothesis, IH (Sorace 2011). Similar findings have been also reported for different L2s like Italian (Sorace & Filiaci 2006) and Croatian (Kras, 2008). Importantly, these studies are experimental in nature and have explored PAS with decontextualised and rather artificial scenarios.

In this presentation, PAS is explored in an L2 Spanish corpus (CEDEL2), as production data offers more natural and contextually rich scenarios that reveal crucial factors in AR that have gone undetected in previous experimental studies. CEDEL2 (Corpus Escrito del Español L2) (Author 2) is an 800,000-word corpus of L1 English-L2 Spanish (plus a comparable Spanish native corpus). For this study, a sample of lower- and upper-advanced learners (plus a native control subcorpus) were delicately annotated following (i) a fine-grained tagset (Fig. 1) implemented in the UAM Corpus Tool tagging software (O'Donnell, 2009) and (ii) the standard procedures used in learner corpus research, LCR (Granger, 2009; Myles, 2015).

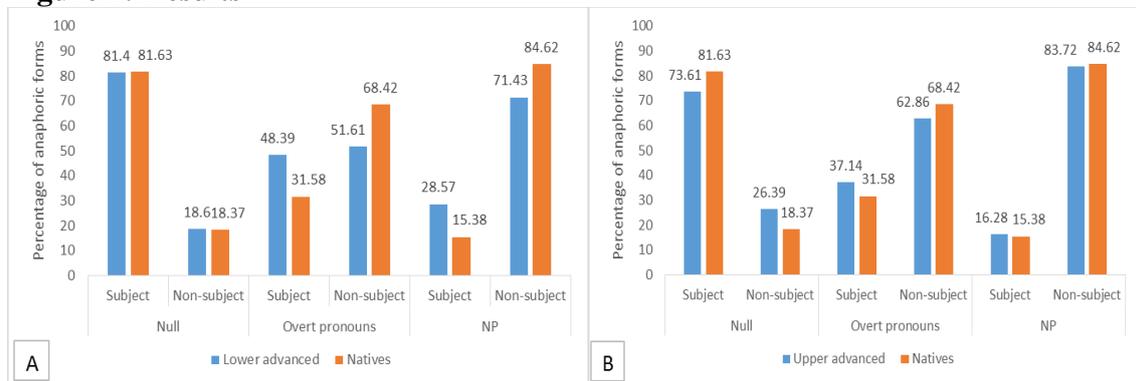
The corpus results (Fig. 2) reveal that both lower- and upper-advanced learners behave similarly to Spanish natives in terms of the PAS, as all groups produce a null subject pronoun to refer to a subject antecedent. Regarding overt anaphoric forms, LCR methodology reveals that the syntax of topic shift is more complex than previously assumed. Overt personal pronouns are more neutral and do not show any biases towards a subject or non-subject antecedent, as some authors have previously reported. Importantly, it is NPs that show a strong bias towards antecedents in non-subject position, both in natives and learners' grammars, a fact that has gone undetected in previous experimental work. The corpus data show that there are other additional factors in AR that are crucial for a better understanding of the PAS.

In short, LCR methods reveal that there are additional factors that have not been previously explored in experimental studies. Additionally, results suggest that native-like competence is attainable in very advanced levels of competence, which runs against the IH predictions.

**Figure 1: Tagset**



**Figure 2: Results**



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., García-Alcaraz, E., & Rosado, E. (2016a). Reference comprehension and production in bilingual Spanish: The view from null subject languages. In A. A. de la Fuente, E. Valenzuela, & C. Martínez Sanz (Eds.), *Language Acquisition Beyond Parameters* (pp. 37–70). Amsterdam: John Benjamins.

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

Filiaci, F. (2010). Null and overt subject biases in Spanish and Italian: A cross-linguistic comparison. In C. Borgonovo, M. Español-Echevarría, & P. Prévost (Eds.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium* (pp. 171–182). Somerville, MA: Cascadia Press.

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.

García-Alcaraz, E. (2015). *Comprensión y producción de los pronombres nulos y explícitos de tercera persona en posición de sujeto en la adquisición temprana del español L2*. UPF, Barcelona.

Granger, S. (2009). The contribution of learner corpora to SLA and FLT. In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 13–32). Amsterdam: John Benjamins.

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.

- Keating, G. D., VanPatten, B., & Jegerski, J. (2011). Who was walking on the beach? Anaphora resolution in Spanish heritage speakers and adult second language learners. *Studies in Second Language Acquisition*, 33(2), 193–221.
- Kras, T. (2008). Anaphora resolution in near-native Italian grammars: Evidence from native speakers of Croatian. In L. Roberts, F. Myles, & A. David (Eds.), *EUROSLA Yearbook 8* (pp. 107–134). Amsterdam: John Benjamins.
- Myles, F. (2015). Second language acquisition theory and learner corpus research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 309–332). Cambridge: Cambridge University Press.
- O'Donnell, M. (2009). The UAM CorpusTool: Software for corpus annotation and exploration. In C. M. Bretones & et al (Eds.), *Applied Linguistics Now: Understanding Language and Mind/La Lingüística Aplicada actual: Comprendiendo el Lenguaje y la Mente* (pp. 1433–1447). Almería: Universidad de Almería.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1–33.
- Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), 339–368.