Using corpus methods to investigate parsing strategies: The position of antecedent strategy in late bilinguals (L1 English – L2 Spanish learners)

Teresa Quesada & Cristóbal Lozano
University of Granada

A well-studied phenomenon in the psycholinguistic literature is Anaphora Resolution (AR). It relates to how anaphoric forms corefer with their antecedents. PAS is a structural parsing strategy whereby a null pronominal subject (Ø) biases towards an antecedent in a structurally higher (subject) position, which marks topic continuity. By contrast, an overt pronoun (él/ella) biases towards a non-subject antecedent, which marks topic shift.

The PAS has been extensively studied in the psycholinguistic literature with Spanish natives, L2 Spanish late bilinguals and Heritage Speakers of Spanish (Alonso-Ovalle et al. 2002, Bel & García-Alcaraz 2015, Bel et al. 2016a, 2016b, Filiaci et al 2014, Jegerski et al. 2011, Keating et al. 2016). Advanced and near-native learners of L2 Spanish typically show deficits when processing PAS, arguably as a result of their limitations when integrating simultaneously syntactic information (overt/null alternation) with discursive information (topic/focus) at the syntax-discourse interface during online processing, as predicted by the Interface Hypothesis (Sorace 2011). However, most of these studies are experimental and have investigated PAS in decontextualized scenarios.

Production data offers contextually richer scenarios and reveal new insights into AR processing. We therefore present evidence from naturalistic production data from the CEDEL2 corpus (http://cedel2.learnercorpora.com), an 800,000-word corpus of L1 English-L2 Spanish (Lozano & Mendikoetxea 2013). Two proficiency-level samples of late bilinguals (plus a Spanish native control subcorpus) were linguistically annotated following a linguistically-informed tagset implemented in UAM Corpus Tool (O’Donnell, 2009).

The corpus results (figure 1) confirm previous experimental work: late bilinguals show native-like behaviour in terms of the PAS, as all groups produce a null subject pronoun to refer to a subject antecedent. But corpus data reveal that preferences for non-subject antecedent are more complex than previously assumed because overt pronouns do not show a clear bias and NPs are also produced in these contexts (which was previously overlooked). Additionally, results show cross-
linguistic influence from specific properties of their L1. Finally, the IH is confirmed because late bilinguals show deficits at the syntax-discourse interface.

Figure 1

Subject antecedent

Subject and coordinated

Subject and not coordinated


