New factors in anaphora resolution: topic continuity in L1 English-L2 Spanish (CEDEL2 corpus)

Fernando Martín-Villena, Universidad de Granada, fmartinvillena@ugr.es

Cristóbal Lozano, Universidad de Granada, cristoballozano@ugr.es

L1 English-L2 Spanish learners (even at very advanced stages) show deficits with anaphora resolution (AR) at the syntax-discourse interface. In particular, when compared to Spanish natives in topic-continuity contexts (TC), learners significantly produce infelicitous forms such as:

  (i) redundant overt pronouns from the outset (Montrul & Rodríguez-Louro 2016; Rothman 2009; Lozano 2009), as in (1), though it is not entirely clear whether this could be due to L1 influence (since their L1 English is a non-null-subject language), or to a more general overspecification phenomenon in L2 acquisition (Hendriks 2013; Ryan 2015).
  (ii) full Noun Phrases (NPs) (Lozano 2016), as in (2), whereas in native Spanish a null pronoun is expected to mark topic-continuity (3).

Unlike previous research, this paper focuses on the multiple factors constraining AR exclusively in TC contexts. The anaphoric forms produced were finely annotated with UAM Corpus Tool (Fig. 1) based on: 1) their pragmatic (in)felicity; 2) the syntactic patterns in which they occur (coordination vs. non-coordination); and 3) the chains created between the anaphoric form and their antecedent(s), a factor that has not been fully explored in previous studies. We tagged and analysed the written compositions of L1 English-L2 Spanish at 3 proficiency levels (beginner, intermediate, advanced/near-natives; N=57) and compared them against a Spanish native control subcorpus (N=20) from CEDEL2 (http://cedel2.learnercorpora.com). Additionally, some examples from near-natives were analysed qualitatively in order to explore other factors conditioning the overproduction of overt forms in comparison with natives.
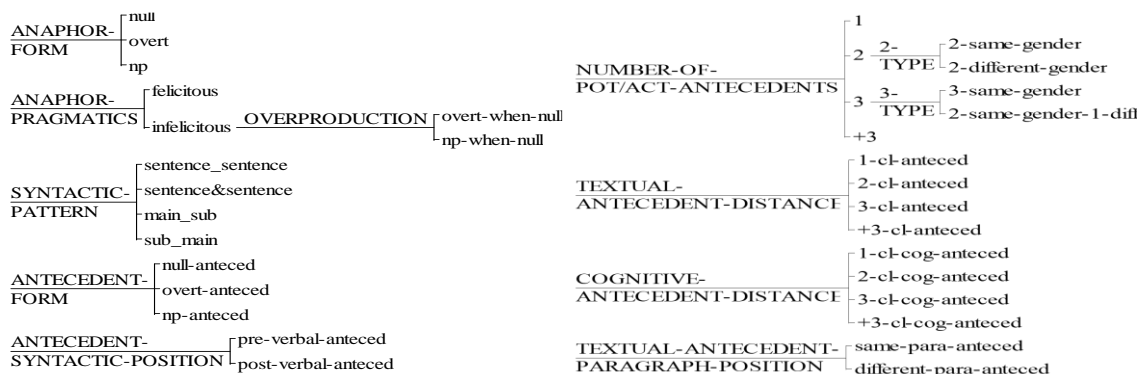
**Figure 1. Tagset (3rd person singular anaphoric forms tagged)**

As in previous research, learners initially overspecify: they redundantly use overt forms (pronouns, NPs) in contexts where null pronouns are pragmatically expected (Fig. 2). While their production of pragmatically adequate forms tends towards the native norm as proficiency increases (Fig. 3), native-like behaviour cannot be eventually attained at very advanced levels, thus supporting the Interface Hypothesis (Sorace & Filiaci 2006; Sorace 2011), which claims that at the syntax-discourse interface learners at the end state show residual deficits such as overuse of overt pronouns in null-subject L2s.

(1) Una persona famosa es **Beyonce Knowles**. **Beyonce** esta muy bonita. **Ella** canta fantastico. **Ella** baila muy bien. [CEDEL2 L1Eng-L2Spa Beginner]

'A famous person is **Beyonce Knowles**. **Beyonce** is very beautiful. **She** sings really well. **She** dances very well.'

(2) **John Lennon** es una persona famosa. **John** nace en Liverpool, England en mil novecientos cuarenta. **John** era moreno, alto [CEDEL2 L1Eng-L2Spa Beginner]

'**John Lennon** is a famous person. **John** was born in Liverpool, England, in 1940. **John** was dark-haired, tall'.
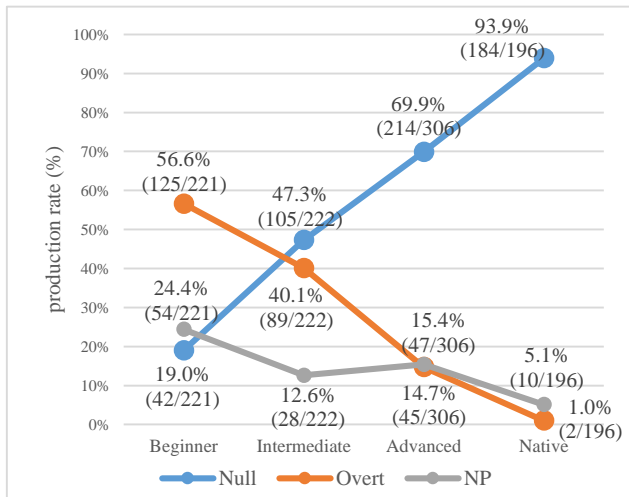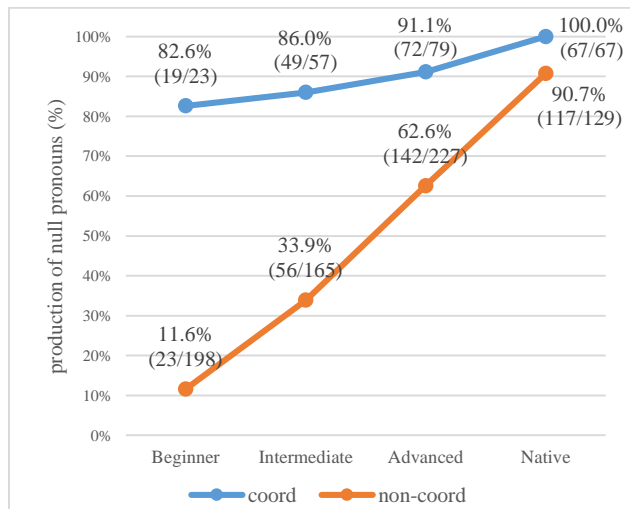
Figure 2. Referential forms in TC
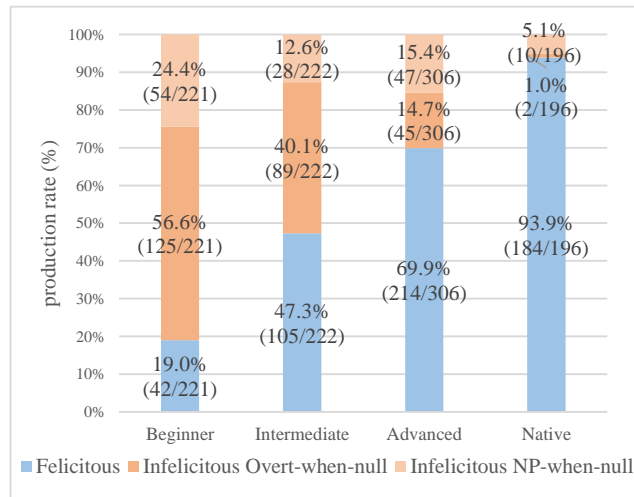
Figure 3. Pragmatic (in)felicity

Additionally, the L1 was shown to be a modulating factor (Fig. 4): learners use felicitous null pronouns in syntactic coordination from early stages (see 4 and 5), which results in native-like behaviour (simply) because this is possible in their L1 English. By contrast, in non-coordinate contexts there is a gradual increase of null pronouns, which eventually approaches native levels in near-natives. This indicates that learners acquire the pragmatic rules of AR at the syntax-discourse interface as proficiency increases.

(3) Chloe$_i$ se enamora de Chris$_j$, **Ø$_i$** le introduce en la empresa familiar y **Ø$_i$** acaba casándose con él$_j$. [CEDEL2 L1Spa Native]

'**Chloe** falls in love with Chris, (**she**) introduces him in the family business and (**she**) ends up marrying him'.

(4) Ella$_i$ ha aparecido en muchas películas y **Ø$_i$** ha registrado cuatro álbumes[CEDEL2 L1Eng-L2Spa Beginner]

'**She** has appeared in many films and (**she**) has recorded four albums'

(5) Porque **ella$_i$** gasta tanto dinero, **ella$_i$** está en deuda y **Ø$_i$** no puede pagar sus cuentas de tarjeta de crédito [CEDEL2 L1Eng-L2Spa Intermediate]

'Because **she** spends a lot of money, **she** is in debt and (**she**) cannot pay her credit card bills'.

Figure 4. Null pronouns x syntactic pattern

Finally, in topic chains, null antecedents are typically followed by null anaphors in native Spanish in line with previous research (Blackwell & Quesada 2012; Cameron & Flores-Ferrán 2004; Geeslin & Linford 2012), though our learners are sensitive to this with increasing proficiency.

In short, our corpus approach reveals facts about AR in TC contexts that have previously gone unnoticed in the L2 acquisition literature as well as some additional qualitative findings.

References

Blackwell, S. E., & Quesada, M. L. (2012). Third-person subjects in native speakers' and L2 learners' narratives: Testing (and revising) the Givenness Hierarchy for Spanish. In K. L. Geeslin & M. Díaz-Campos (eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium* (pp. 142–164). Somerville, MA: Cascadilla Press. **Cameron, R., & Flores-Ferrán, N.** (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context*, 1(1), 41–65. **Geeslin, K. L., & Linford, B.** (2012). *Cross-sectional study of the effects of discourse cohesiveness and perseveration on subject expression. Paper presented at the Sixth International Workshop on Spanish Sociolinguistics*. Tucson, Arizona. **Hendriks, H.** (2003). Using nouns for reference maintenance: A seeming contradiction in L2 discourse. In A. Giacalone Ramat (Ed.), *Typology and Second Language Acquisition* (pp. 291–326). Berlin: Mouton de Gruyter. **Lozano, C. (2009).** Selective deficits at the syntax-discourse interface: Evidence from the CEDEL2 corpus. In: N. Snape, Y-k. I. Leung & M. Sharwood-Smith (eds). *Representational Deficits in Second Language Acquisition* (pp. 127–166). Amsterdam: John Benjamins. **Lozano, C.** (2016). Pragmatic principles in anaphora resolution at the syntax-discourse interface: advanced English learners of Spanish in the CEDEL2 corpus. In: M. Alonso Ramos (ed.), *Spanish Learner Corpus Research: Current Trends and Future Perspectives* (pp. 236–265). Amsterdam: John Benjamins. **Montrul, S., & Rodríguez-Louro, C.** (2006). Beyond the syntax of the Null Subject Parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. In: V. Torrens & L. Escobar (eds.), *The Acquisition of Syntax in Romance Languages* [Language Acquisition and Language Disorders 41] (pp. 401–418). Amsterdam: John Benjamins. **Rothman, J.** (2009). Pragmatic deficits with syntactic consequences? L2 pronominal subjects and the syntax-pragmatics interface. *Journal of Pragmatics*, 41, 951–973. **Ryan, J.** (2015). Overexplicit Referent Tracking in L2 English: Strategy, Avoidance, or Myth? *Language Learning*, 65(4), 824–859. **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.