

Pronominal Ambiguity Resolution in Spanish Child Dialogue: A Corpus Based Developmental Language Acquisition Approach.

Martha Robinson
Department of Linguistics
University of Edinburgh
martha.robinson94@gmail.com

Abstract

This investigation is a CHILDES (MacWhinney, 2000) corpus based study of how 3-4 and 5-6 year old monolingual Spanish speaking children learn to disambiguate null and overt pronominal reference from the input they are exposed to whilst engaging in adult-child dyadic communicative acts, as they are involved in the active development of pragmatic awareness. It was found, that although there was no significant difference between both groups in terms of tokens belonging to either pronominalization strategy in the input the children received, there was, however, a difference in the types of lexical verbs and modes of child-adult interaction at each developmental step.

1 Introduction

In the literature, it is often mentioned that pronouns have no fixed reference and their interpretation is highly context dependent (Kempson et al., 2001). In fact, more recent psycholinguistic experimental work has shown that syntactic, pragmatic and discursive factors figure prominently in their interpretation (Stevenson et al., 1994; Kehler, 2002; Kehler et al., 2008; Kehler and Rohde, 2015; inter alia). From an early age, children are able to produce personal pronominal forms correctly and their acquisition has been said to be closely embedded in early communicative experiences (Salazar Orvig et al., 2010). Nevertheless, comparison of adults and older children has shown that there are considerable differences in the mastery of their appropriate use among children and that many so called failures are in fact, developmental (Song and Fisher, 2007).

2 Null and Overt Pronominals in Spanish.

From a typological perspective, Spanish has been described as belonging to the group of languages licencing a partial *pro*-drop strategy (and related to Italian, Portuguese, Catalan, among others). This means, that subject pronominal expressions tend to be phonetically null and, dropped readily. In the literature, this has been related rich verbal morphology. However, Spanish also presents an overt pronominal counterpart as shown in (a) with the third person stressed form *Él* (he).

- a. Felix_i le pegó a Max_j y luego Ø_i/ÉL_j le pegó a Pedro.
Felix hit Max and then Ø_i/HE_j hit Pedro.

Here, the null form in the second conjunct co-refers naturally with the higher subject *Felix*. In contrast, the overt stressed form *él* (he) shows a marked natural preference to attach to the lower object *Max*. The alternation between overt and null pronouns has been at the centre of a great deal of debate in the linguistics literature for decades, especially in these inter-sentential instances of anaphoric co-reference (RAE, 1999). In fact, a prevalent view is that they occur in complementary distribution and display a division of labour strategy, a position that has also been widely adopted in more recent experimental psycholinguistics literature (as in Carminati, 2002 for Italian and Alonso-Ovalle, et al., 2002 for Spanish, for example). Although (a) appears to be a clear example of a strict division of labour strategy, it is also the case, that these types of anaphoric relations are often affected by verb semantics as we see in (b) and (c) with IMPLICIT CAUSALITY verbs such as *asustar* (frighten) and *temer* (fear):

- b. María_i asusta a Ana_j porque Ø_i/ELLA_j es; antipática.
María frightens Ana because Ø_i/SHE_j is horrible.
c. María_i teme a Ana_j porque Ø_j/ELLA_j es antipática.
María fears Ana because Ø_j/SHE_j is horrible.

In contrast to (a) above with a verb like *pegar* (hit) where null and overt forms appear to enter into an either/or relation, here both are able to enter into parallel coreference relations. In (b) both forms co-refer with *María* and in (c) with *Ana*. And here, the fact that a null form can actually co-refer with a lower object and the overt form with the higher subject NP, is particularly unexpected and undermines the strict division of labour perspective often proposed in the literature. Other factors disputing this perspective are issues of dialectal variation and formal syncretism in certain verbal paradigms (especially between 1st and 3rd person singular verbal inflections), in spoken varieties of Spanish where the appearance of overt pronouns is favoured in order to disambiguate reference to the subject. And finally, more general pragmatic and discursive principles also figure prominently in Spanish (in line with Kehler, 2002; Kehler et al., 2008; Kehler and Rohde, 2015, inter alia, on pronominal co-reference in English).

Experimental work with children both in English and Spanish has shown that they display a marked preference for lower attachment coreferential interpretations (Hartshorne, et al., 2014, for English and Ruigendijk, et al., 2011 for Spanish). However, in a study involving 5-6 year old children and adults conducted by Kehler, Hayes and Barner, (2011) involving *Transfer of Possession* and *Transfer of Location* verbs in English, a main effect of verb type and age was found. Although children are already highly adept at knowing how discourse works from an early age (Clark, 2005), the anaphoric value of pronominal expressions is first acquired by being involved in dialogue, before it is extended to monological uses (Salazar Orvig et al., 2010). Therefore, since the fundamental skills underlying the communicative process develop gradually between the ages of 3 and 5, until these become the highly sophisticated conversational acts of adults, the main question here, is how children learning Spanish achieve this task from the input they are exposed to, especially since they have to acquire two pronominalization strategies. We contend here that it is not only the input children are exposed to, but also how children conduct themselves in communicative exchanges that aid the development of anaphoric interpretations.

3 Method

We took a sample of adult-child (investigator or parent) dyadic interactions from three Iberian Spanish monolingual child dialogue corpora and created a sub-corpus (27,277 total number of words) based on 100 utterances per child (approx. 200 adult-child turns) on which we calculated their Mean Length Utterance (Brown, 1973) or MLU_w (ie. MLU measured in words, as discussed in Ezeizabarrena and García Fernández, 2018), totalling 3972 turns. The dialogues involved traditional interactive story telling sessions (we excluded monological narratives) as well as communicative acts recounting children's daily routines at school and at home. The children were then matched for socio-economic status, MLU_w and strict monolingual Spanish linguistic background (ie. excluding familial bilinguals and monolingual and multilingual children cohabiting in areas with linguistic minority languages). Samples were collected at two developmental steps 3-4 (N=10) and 5-6 (N=8) years old, identified as Group A and Group B respectively. We excluded adult-adult exchanges at this point of the investigation.

4 Results

In the first instance, we found an overwhelming marked preference for the null form of the pronominal in adult-child interactions in both groups with only a few instances of the overt form and this was fairly consistent among both age groups (Group A: null 87.6% vs overt 12.4% and Group B: null 89.25% vs overt 10.75%). This is consistent with the fact that in a language like Spanish, the null form is considered to be the default and the overt, the exception (albeit the overt appears in certain highly predictive contexts). For the overt pronominal we found that adult-child ratio showed no significant difference either (Group A: adult 48.95% vs child 51.05% and Group B: adult 48.54% vs child 49.5%) and this means that children are exposed to and have already learned of the availability of the overt form from a very young age. However, the difference can be seen in the types of lexical verbs and adult-child interactional strategies utilised whilst engaging in communicative exchanges which differ at both developmental steps, an issue that nevertheless, merits further investigation.

References

- Roger Brown. 1973. *A first Language: The early stages*. Cambridge, Harvard University.
- Maria Nella Carminati. 2002. The processing of Italian subject pronouns. PhD Thesis, University of Massachusetts, Amherst.
- Eve Clark. 2005. *Pragmatics and Language Acquisition*. The Handbook of Pragmatics. Laurence R. Horn and Gregory Ward (eds.), Blackwell.
- María José Ezeizabarrena and Iñaki García Hernández. 2017. Length of Utterance, in morphemes or in words?: MLU3-w, a reliable measure of language development in Early Basque. *Frontiers in Psychology*, 8, 2265: <https://doi.org/10.3389/fpsyg.2017.02265>
- Theres Gruter, Hannah Rohde and Amy J. Schafer. 2017. Coreference and discourse coherence in L2: The Roles of grammatical aspect and referential form. *Linguistic approaches to Bilingualism*, 7(2), 199-229.
- Joshua K. Hartshorne and Jesse Snedeker. (unknown date). *Integration of discourse and semantic structure in children's resolution of ambiguous pronouns*. https://d2dg4e62b1gc8m.cloudfront.net/pub_pdfs/HartshorneSnedeker_BUCLD09_abstract.pdf
- Joshua K. Hartshorne, Rebecca Nappa, and Jesse Snedeker. 2014. Development of the first-mention bias. *Journal of Child Language*, 42(2), 423-446.
- Andrew Kehler. (2002). *Coherence, reference and the theory of grammar*. Stanford, CA: CLSI Publications.
- Andrew Kehler, Laura Kertz, Hannah Rohde, Jeffrey L. Elman. 2008. Coherence and coreference revisited. *Journal of Semantics*, 25, 1-44.
- Andrew Kehler and Hannah Rohde. 2015. Pronominal reference and pragmatic enrichment: A Bayesian account. In Proceedings of the 37th Annual Conference of the Cognitive Science society, pp. 1063.
- Andrew Kehler, Emily Hayes and David Barner. 2011. *Pragmatically driven biases in children's pronoun interpretation*. Poster presented at the 24th CUNY Conference on Human Sentence Processing, March 24-26, 2011, Stanford.
- Ruth Kempson, Wilfried Meyer-Viol, Dov Gabbay. 2001. *Dynamic Syntax: The flow of language understanding*. Wiley-Blackwell.
- Evan Kidd, Andrew Stewart and Ludovica Serratrice. 2011. Children do not overcome lexical biases where adults do: the role of referential scene in garden-path recovery. *Journal of Child Language*, 38 (222-234).
- Brian MacWhinney. 2000. The CHILDES Project: Tools for analysing talk, 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- RAE: *Gramática Descriptiva del Español*. 1999. Violeta Demonte and Ignacio Bosque (eds.), Real Academia Española, colección Nebrija y Bello. Espasa Calpe.
- Esther Ruigendijk, Sergio Baauw, Shalom Zuckerman, Nada Vasic, Joke de Lange, Sergei Avrutin. 2011. A Cross-linguistic Study on the interpretation of Pronouns by Children and Agrammatic Speakers: Evidence from Dutch, Spanish and Italian. In Edward Gibson and Neal J. Pearlmutter (eds.), *The Processing and Acquisition of Reference*. MIT Press, Cambridge Massachusetts.
- Anne Salazar Orvig, Haydée Marcos, Aliyah Morgenstern, Rouba Hassan, Jocelyne Leber-Marin, Jacques Parès. 2011. Dialogical beginnings of anaphora: The use of third person pronouns before the age of 3. *Journal of Pragmatics* 42, 1842-1865.
- Hyun-joo Song and Cynthia Fisher. 2007. Discourse prominence effects on 2.5-year old children's interpretation of pronouns. *Lingua*, 117(11), 1959-1987.
- Stevenson, R., Crawley, R. & Kleinman, D. (1994): Thematic roles, focusing and the representation of events. *Language and Cognitive Processes*, 9, 519-548.