Implicit Causality in Romanian Pronoun Production and Comprehension

Ruxandra Ionescu, Fabian Istrate, Barbara Hemforth

Implicit Causality in Romanian Pronoun Production and Comprehension

Ruxandra Ionescu, Fabian Istrate & Barbara Hemforth
Laboratoire de Linguistique Formelle,
Université de Paris-CNRS

1 Introduction

Romanian, as other Romance languages (not including French), is a so-called pro-drop language, meaning that the subject can be realized as a null subject, a pronominal subject or a lexical subject . The choice between these alternatives depends on a variety of factors. Istrate et al. (2023) for example show based on corpus studies that null subjects are preferred for discourse persons (1st and 2nd person) and somewhat less so for 3rd person. Studies conducted on Romanian with experimental items containing a transitive verb with two human arguments (subject and object), followed by an embedded temporal clause have revealed a preference for null subjects over pronominal ones (Istrate at al. 2022). Moreover, negative polarity has been found to be a factor that also plays a role in the equation with more pronominal subjects in negated sentences. The polarity effect could be explained as a contrast effect (Dobrovie-Sorin & Giurgea 2013).

Similar to other Romance languages (see for example Carminati 2002 for Italian, Mayol 2010 for Catalan, de la Fuente & Hemforth 2013 for Spanish, Fernandes et al. 2018 for European Potuguese), null and pronominal subjects show a division of labor effet in Romanian. Istrate et al. (2022) showed that generally null subjects have a preference for a subject antecedent while pronominal subjects have a preference for non-subject antecedents. This preference has been explained by the *Position of Antecedent Hypothesis (PAH)* (Carminati 2002) suggesting a preference based on the syntactic position of the antecedent. However, it has been shown that other factors can also play a role such as, for example, the informational status of the antecedent with a preference for a topical antecedent for a null subject (Runner & Ibarra 2016).

Based on Runner & Ibarra (2016), an alternative to the *PAH* could be that the saliency of the antecedent plays a role for the choice of subject type. Since the subject is the default topic in most of the languages for which the *PAH* has been suggested, a syntactic position effect and an information structure effect would be hard to distinguish. Following (Ariel 1994, Givón 1983), the weakest referential expression would go with the most salient entecedent (in production as well as in comprehension). In English, for example, a pronominal subject is supposed to be the weakest referential expression that refers back to the most salient antecedent (typically the subject). The weakest pronoun in Romanian would be the null pronoun. In our research project we want to find out whether saliency induced by implicit causality biases contributes to the differential choice between the null vs pronominal subjects in production as well as the choice of antecedent in comprehension. Following Ferstl et al. (2010), we will also look at possible gender biases in Romanian in regards to antecedent choice. Ferst et al. found a small preference for male causers in their implicit causality corpus.

Language comprehension relies to a large extent on the understanding of connected sequences of events or states which are sometimes linguistically expressed by certain verbs (e.g. Ferstl et al. 2010, van den Hoven & Ferstl 2018, Hartshorne 2013). It is essential to determine the causal relationships between the events and the states presented in a text in order to properly understand it. We will focus here on a class of interpersonal verbs such as (1a,b) that increase the prediction of an upcoming cause to be attributed to one of its arguments.

(1a) Mary fascinated Peter because …
(1b) Mary criticized Peter because …

Depending on whether the cause of the event is assigned to the subject or to the object of a simple active sentence, two categories of verbs emerge: subject-bias and object-bias verbs. *Facinate* in (1a) is a subject-biased verb that increases the expectation of a cause attributed to Mary (why is she fascinating?), *criticize* in (1b) increases the expectation of a cause with the object antecedent as a causer (what has Peter done to be criticized). In order to investigate the null/pronominal subject alternation as well as possible gender biases in the production of pronouns in Romanian, we will use implicit causality as a factor to see whether or not saliency affects the production and/or understanding of null and pronominal subjects.

2 The Experiment

Our experiment focuses on testing at the same time the production of referential expressions (lexical vs pronominal vs null subject) in Romanian as well as the preference for an antecedent using a paradigm similar to Kehler & Rohde (2019). In order to increase the salience of an antecedent, we choose implicit causality verbs that increase the next mention probability of the subject as in (3a, b) (subject-biased verbs) or the object (object-biased verbs) as in (4a, b). One of the antecedents was always a feminine first name, the other a masculine first name. We created two conditions for each sentence switching subject and object to test for possible gender effects as they have been found in Ferstl et al. (2010). The experiment consisted of a sentence-completion task conducted on IbexFarm (created by Alex Drummond 2010, installed and maintained on a lab server at University Paris Cité by Achilles Falaise) with 48 experimental items, 2 conditions each, no distractors) where participants were asked to continue sentences following the pattern in (2).

(2) Female/Male first name + IC subject/object-bias verb + Male/Female first name + "because"

The participants would be asked to continue the sentences using a plausible continuation of their choice (freely choosing the subject of the causal clause). According to literature, participants continue with either a pronominal, a lexical or a null subject in more than 85% of the cases. (Kehler, A. & Rohde, H. 2019).

(3a) Maria îl dezamăgește pe Victor pentru că…
Mary CL.3SG.M.ACC dissapoint.PRS DOM Victor because…
'Mary disappoints Victor because…'

(3b) Victor o dezamăgește pe Maria pentru că…
Victor CL.3SG.F.ACC dissapoint.PRS DOM Mary because…
'Victor disappoints Mary because…'

(4a) Alexandra îl adoră pe Albert pentru că…
Alexandra CL.3SG.M.ACC adore.PRS DOM Albert because…
'Alexandra adores Albert because…'

(4b) Albert o adoră pe Alexandra pentru că…
Albert CL.3SG.F.ACC adore.PRS DOM Alexandra because…

'Albert adores Alexandra because…'

3 Participants

31 native Romanian speakers participated in our experiment. We only present data from the first 18 participants here whose continuations we have annotated so far. All of the participants were raised in Romania. They were students recruited at the University of Bucharest. The age range of the participants was between 19 and 32 with a mean age of 27 years. Given that the participants are enrolled in an institution of higher education, their level of instruction is very similar (a minimum of 12 years of instruction). Thus, the participants had no difficulty in reading, understanding and continuing the sentences. Participation was voluntary, participants were not paid. The experiment was run on a version of Ibex farm installed on our local server. Participants' data were immediately anonymized. At no moment identifying information was stored.

4 Hypotheses

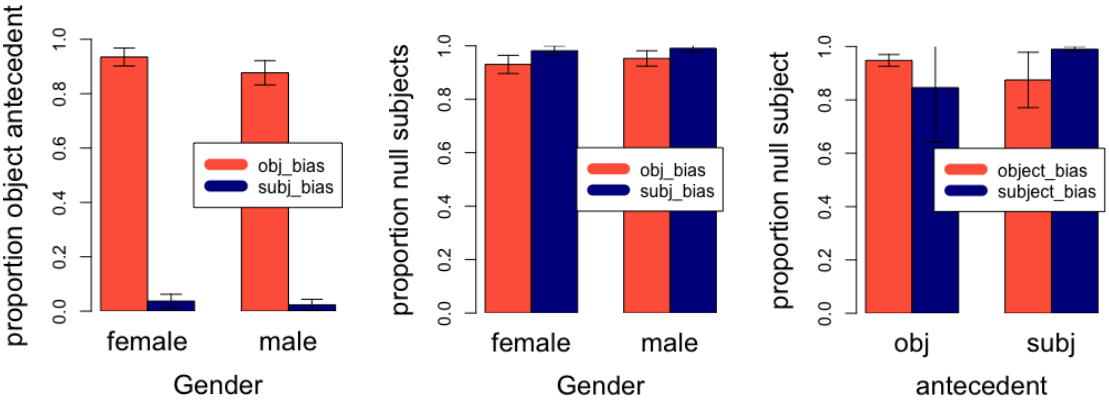
The experiment follows two hypotheses that are, to some extent, interconnected. As a pro-drop language, Romanian can have null subjects to refer back to an antecedent in the previous clause or sentence. According to the literature (e. G. Carminati 2002), null subjects have a strong tendency to go with a subject antecedent (see Istrate et al. 2022 for Romanian). If this hypothesis is correct, we should expect that participants produce more null subjects when referring back to a subject antecedent as well as more continuations referring back to the subject antecedent when a null subject is produced. Pronouns should be used more when participants refer back to the object. If, however, null pronouns prefer salient antecedents that are likely to be mentioned next, verb biases may change the picture: Null subjects should be more frequent when the continuation is in line with the verb bias.

To sum up our hypotheses:

- Based on the implicit causality biases of the verbs, continuations should refer to the antecedent foregrounded by the verb.
- The choice of the referential expression (null vs pronominal subject) will be influenced by verb bias.
- Following results from Ferstl et al (2010), we may also find a gender effect with a preference to choose male antecedents as the causer of an event.

4.1 Results

The continuations were annotated by two native Romanian speakers with respect to the intended antecedent of the continuation as well as with respect to the referential expression used for the subject of the because clause. All data were analyzed using logistic regressions (glmer function in the lme4 package, Bates et al. 2015, p-value were estimated using lmer test, Kutzenova et al. 2017). We first analyzed the effect of implicit causality and gender on antecedent choice. Gender of the subject of the root clause as well as verb bias were added as mean centered fixed factors and participants and items as random factors. Random slopes could not be added due to convergence failure. This is true for all models presented here. As shown in Figure 1, participants' continuations were highly consistent with the verb bias. They systematically chose a continuation consistent with a subject antecedent after subject biased verbs and with the object antecedent after object biased verbs (Est. = -6.90, std. error = .6176, z=-11.179, $p < .001$). There is also a small numeric effect of gender with slightly more subject choices when the subject antecedent is male (Est. = -.5938, std. error = .3586, z=-1.656, p = .0977).



Participants chose null or pronominal subjects in more than 95% of the cases so we will not report on other referential expressions. They overwhelmingly chose null subjects in the *because* clause independent of subject bias (Est. = 4.8020, std. error = .6262, z=7.669, p < .001). Null subjects were chosen more frequently for sentences with subject biased verbs where verb bias and the general preference of null subjects for subject antecedents align (Est. = 1.6934, std. error = .5866, z=2.887, p < .01). We finally looked at the frequencies of null and pronominal subjects depending on the antecedent choice made by the participants. Logistic regressions included antecedent choice and verb bias as fixed factors and participants and items as random factors. Figure 3 shows that participants chose null subjects more often in cases where the verb bias and the antecedent choice aligned, i.e. when they produced a continuation consistent with an object antecedent in sentences with object biased verbs or a continuation consistent with a subject antecedent in sentences with subject biased verbs (Est. = 4.3433, std. error = 1.2946, z=3.355, p < .001).

5 Discussion

All in all, the results of our experiment confirm our hypothesis that null subjects can easily be produced for non-subject antecedents when they are salient enough. To our knowledge, this is the clearest evidence against the *Position of Antecedent Hypothesis* that stipulates a general preference for subject antecedents. It might be argued that the results we found, could be due to the artificial experimental situation. However, the continuations produced by the participants with object antecedents in the case of object biased verbs (5a,b) are judged as highly natural by both native speakers of Romanian who annotated them.

(5a) Laura îl felicită pe Ionuț pentru că a luat permisul.
Laura CL.3SG.M.ACC congratulate.PRS DOM Ionut because AUX take.PST license.
'Laura congratulates Ionuț because he got his driving licence.'

(5b) Ionuț o felicită pe Laura pentru că a câștigat un premiu.
Ionut CL.3SG.F.ACC congratulate.PRS DOM Laura because AUX win.PST a prize.
'Ionuț congratulates Laura because she won a prize.'

To make sure that the intuition of our annotators are shared by the larger community, acceptability studies are planned for the near future. It will also be interesting to see how the preference of a null subject for a salient antecedent translates to other null subject languages like Italian or Spanish.

Participants produced pronominal subjects more often when the antecedent was less salient. Other cases of pronominal subjects included continuations that contained a contrast (Dobrovie-Sorin & Giurgea, 2013).

(6a) Maria îl dezamăgește pe Victor pentru că și el a făcut
Mary CL.3SG.M.ACC dissapoint.PRS DOM Victor because also he AUX make.PST
același lucru.
same thing.
'Mary disappoints Victor because he also did the same thing.'

(6b) Victor o invidiază pe Maria pentru că ea are note mai bune
Victor CL.3SG.F.ACC envy.PRS DOM Mary because she have.PRS grades better
decât el.
than he.
'Victor envies Mary because she has better grades than him.'

The small possible gender effect we found in our data encourages us to look at potential gender effects in more detail. Following Kaiser et al. (2022), we are currently running a study looking at the role of family names for antecedent choice with the hypothesis that family names are dispreferred for female antecedents. The study will be finished and annotated to be included in the final version of this paper.

References

Ariel, Mira. 1994. Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics* 30.1: 3-42. https://doi.org/10.1017/s0022226700016170

Bates, Douglas, Mächler, Martin, Bolker Ben, Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01. https://doi.org/10.18637/jss.v067.i01

Carminati, Maria Nella. 2002. The processing of Italian subject pronouns. University of Massachusetts Amherst.

De la Fuente, Israel & Barbara Hemforth (2013). Effects of clefting and left-dislocation on subject and object pronoun resolution in Spanish. In Jennifer Cabrelli Amaro, Gillian Lord, Ana de Prado Pérez, Jessi Elana Aaron, *Selected proceedings of the 16th Hispanic Linguistic Symposium*, 27-45.

Dobrovie-Sorin, Carmen, and Giurgea, Ion eds. 2013. *A reference grammar of Romanian: Volume 1: The noun phrase*. Vol. 207. John Benjamins Publishing, 2013.https://doi.org/10.1075/la.207

Fernandes EG, Luegi P, Correa Soares E, de la Fuente I, Hemforth B. Adaptation in pronoun resolution: Evidence from Brazilian and European Portuguese. J Exp Psychol Learn Mem Cogn. 2018 Dec;44(12):1986-2008. doi: 10.1037/xlm0000569. Epub 2018 Apr 26. PMID: 29698041.

Ferstl, Evelyn C., et al. 2010. "Implicit Causality Bias in English: A Corpus of 300 Verbs." *Behavior Research Methods*, vol. 43, no. 1, pp. 124–135, https://doi.org/10.3758/s13428-010-0023-2. Accessed 12 Nov. 2021.

Givón, Talmy. 1983. Topic continuity in discourse: The functional domain of switch reference. *Switch reference and universal grammar* 51: 82. https://doi.org/10.1075/tsl.2.06giv

Hartshorne, Joshua K. 2013. "What Is Implicit Causality?" *Language, Cognition and Neuroscience*, vol. 29, no. 7, pp. 804–824, https://doi.org/10.1080/01690965.2013.796396. Accessed 10 Aug. 2020.

Istrate, Fabian, Abeillé, Anne, & Hemforth, Barbara. 2022. Romanian Subject Alternation: The Position of Antecedent Hypothesis. *Linguistic Evidence Proceedings. 61-64. https://le-2022.sciencesconf.org/data/pages/LinguisticEvidenceBooklet.pdf*

Istrate, Fabian, Abeillé, Anne, Hemforth, Barbara. 2023. *53 Symposium on Romance Languages< Juin 2023.*

Kaiser, Elsie, Deborah Ho, Haley Hsu, Claire Post, & Madeline Rouse. 2022. Referring to someone only using their last name. *Proceedings of Linguistic Evidence 2022*, 69-72. https://le-2022.sciencesconf.org/data/pages/LinguisticEvidenceBooklet.pdf

Kehler, Andrew & Rohde, Hannah. 2019. Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics, Special Issue on Prominence in Pragmatics,* 154, 63-78.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Mayol, Laia. 2010. Refining salience and the position of antecedent hypothesis: a study of Catalan pronouns. University of Pennsylvania Working Papers in *Linguistics* 16.1: 15.

Runner, Jeffrey T., & Alyssa Ibarra. 2016. Information structure effects on null and overt subject comprehension in Spanish. *Empirical perspectives on anaphora resolution*: 87-112. https://doi.org/10.1515/9783110464108-006

van den Hoven, Emiel, and Evelyn C. Ferstl. 2018. “The Roles of Implicit Causality and Discourse Context in Pronoun Resolution.” *Frontiers in Communication*, vol. 3, https://doi.org/10.3389/fcomm.2018.00053. Accessed 16 Jan. 2023.