The processing of object relative clauses in Spanish: Some evidence in favor of usage-based approaches

Florencia Reali

Universidad de los Andes, Bogotá, Colombia

Universidad de los Andes

Universidad de los Andes

Introduction

Object relative clauses

Sentences with object relative (OR) clauses such as *The politician that the secretary called resigned*, are difficult to process.

This well-established phenomenon has been explained by different theories relying on structural or functional variables including working memory limitations, discourse/pragmatics and experience (Roland et al., 2012; Reali & Christiansen, 2007).

However, when the embedded subject is a pronoun as in *The politician that we called resigned*, OR clauses become much easier to process.

Such facilitation has been explained in terms of discourse expectations (Roland et al., 2012) or, in line with usage-based approaches, as a consequence of exposure to frequent pronominal OR constructions that may become automated into processing units that get easier to access (Reali & Christiansen, 2007).

Usage based approaches and nested structure

Usage-based models of acquisition within the constructivist tradition (Bybee & Hopper, 2001; Goldberg, 2006) propose a cognitive model of linguistic knowledge in which speakers learn language as a series of larger chunks ("established units") from which they may derive representations of varying degrees of cohesion or schematicity.

From this perspective, language use statistics gain a central place in linguistic theory: Corpus data becomes a model of the way an 'average' speaker derives linguistic knowledge from usage. Along these lines, nested clauses formed by frequent chunks might lead to stronger representations that have become fluent through language use and repetition, and therefore are quite relatively easy to access. Previous results in English support this hypothesis.

Here, we provide further evidence in favor of usage-based approaches by combining corpus analysis and two off-line rating tasks designed to test complexity of relative clause structures in Spanish.

Flexibility of surface constituents in Spanish OR Clauses

1. Spanish allows word order flexibility within relative clauses.
Word order in OR clauses (verb-first/verb-final) is flexible.
Example: "la persona que [los abogados conocen/ conocen los abogados] acaba de llegar" (t.: The person that the lawyers know just arrived)

2. In Spanish the embedded subject can be dropped.
Omission of the embedded subject is particularly natural in pronominal clauses since subject information is encoded in verb conjugation when the subject is a personal pronoun. Example:
a. El sapo que nosotros perseguimos. (Overt pronoun)
b. El sapo que perseguimos. (null pronoun)
[The frog that (we) chased]

Questions of study:

¿What is the relative frequency of OR clauses in Spanish?
¿Are there significant differences in the relative frequency of surface structure variations in OR clauses in Spanish?

References:
Bybee, J. & Hopper, P. (2001). Frequency and the emergence of linguistic structure. Amsterdam & Philadelphia: John Benjamins.
Roland D., Mauner, G., O'Meara, C. & Yun H. (2012). Discourse expectations and relative clause processing. *J Mem Lang*, 66:479-508.
Reali, F. & Christiansen, M.H. (2007). Word-chunk frequencies affect the processing of pronominal object-relative clauses. *Q J Exp Psychol*, 60:161-170.
Marin, M. (1991). Corpus lingüístico de referencia de la lengua española, *Boletín de la Academia Argentina de Letras*, 56:129-155.

Corpus Analysis

Previous corpus analyses have shown that pronominal ORs are extremely frequent in English (Roland et al., 2012; Reali & Christiansen, 2007).

Goals:
1. Investigate whether the distributional regularities of nested structure found in English reproduce in Spanish.
2. Explore patterns of occurrence of flexible surface structure in Spanish.

Materials and methods:
We used data from *Corpus de Referencia de la Lengua Española Contemporánea: Corpus Oral Peninsular* – free-access corpus of spoken Spanish. The corpus contains 1.100.000 words recorded and transcribed from public radio sources. We analyzed a total of 110 different monologues and conversations distributed over six different labeled topics (total over 200.000 words). We searched for lexical items used as subordinating connectives in Spanish, including the relativizer "quien(es)" [who] and the relative pronoun "que"[that]. All sentences containing those lexical items were individually analyzed and only sentences in which the words were used to connect a relative clause and a head noun phrase were considered.

Word order patterns un OR clauses
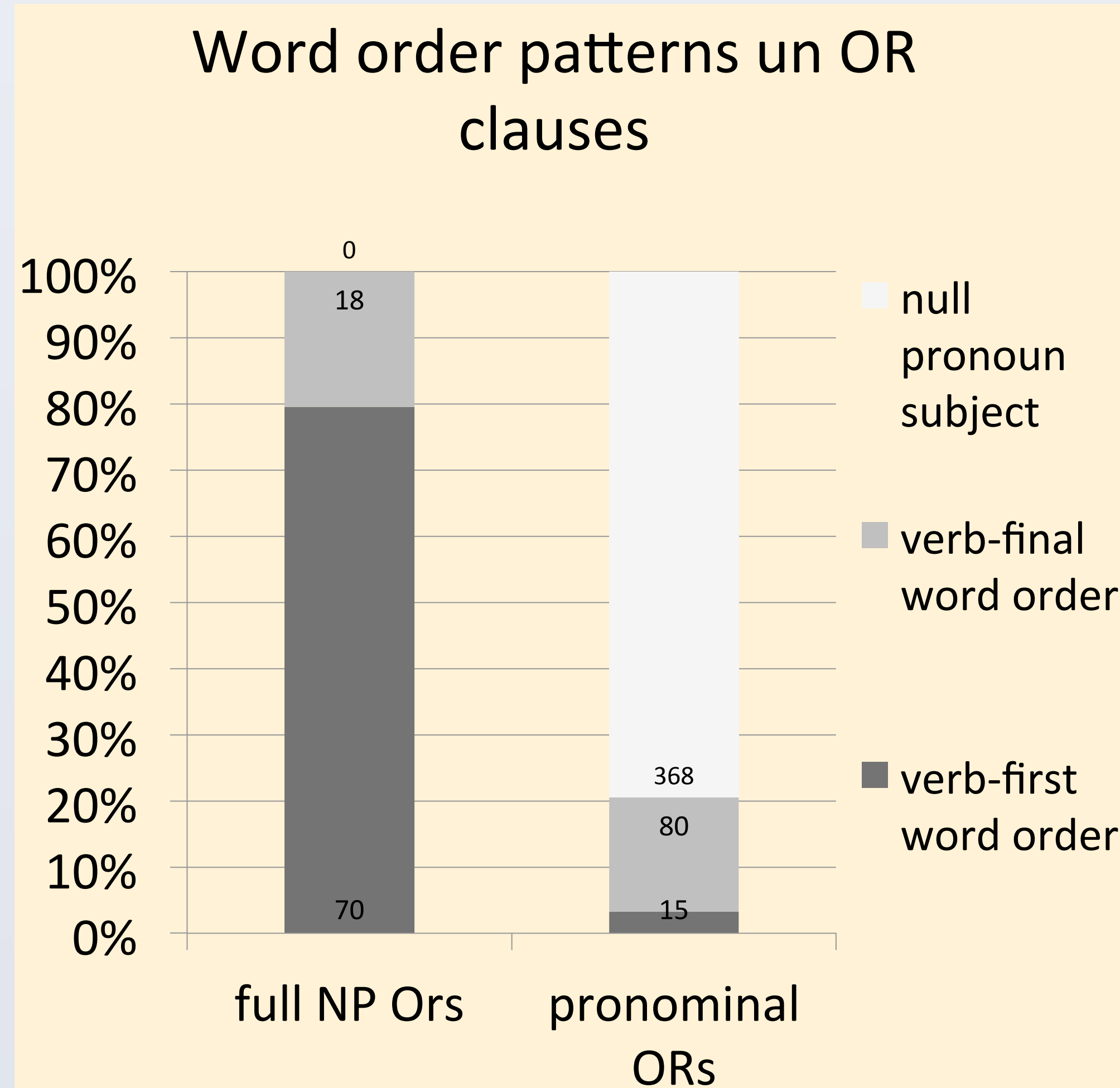
Results:

We found 394 transitive subject relative clauses and 564 object relative clauses. There was a significant asymmetry between pronominal object and subject relative clauses: 12% of SR clauses were pronominal, while 85% of OR clauses were pronominal (85%). This is qualitatively similar to the pattern distribution found in English (Roland et al, 2012; Reali & Christiansen 2007).

The corpus analysis revealed that pronouns are omitted in the majority of pronominal ORs (77%), and that, when present, pronouns occur in verb-final clauses 85% of the times. However, when the embedded subject is a full noun-phrase the pattern reverses: 70% of ORs are verb-first (χ^2 >100; p< .0001).

Experimental hypothesis: Frequent chunks should facilitate the processing of OR clauses in Spanish. Complexity ratings should mirror the distributional patterns found.

Experiments

Two off-line complexity rating experiment

Materials
Experiment 1
Twelve sentences containing OR clauses in Spanish. Embedded subjects were full noun phrases and word order was varied across conditions:

a. La araña que [persiguieron los sapos] se comió la mosca. (V-NP/High-frequency condition)
b. La araña que [los sapos persiguieron] se comió la mosca. (V-NP/Low-frequency condition)
(Translation: The spider that the frogs chased ate the fly.)

Experiment 2
Twelve sentences containing OR clauses in Spanish. Embedded subjects were pronouns. Word order/overt vs. null pronoun was varied across conditions:

a. La araña que [perseguimos] se comió la mosca. (NP-V, null/High-frequency condition)
b. La araña que [perseguimos nosotros] se comió la mosca. (NP-V, overt/Medium-frequency condition)
c. La araña que [nosotros persiguimos] se comió la mosca. (V-NP, overt/Low-frequency condition)
(Translation: The spider that we chased ate the fly.)

Participants. Thirty (Experiment 1) and thirty (Experiment 2) native Spanish speakers, undergraduate students from Univ. de los Andes.
*Procedure:* Participants rated complexity ratings of sentences on a scale from 1 to 7.

Results
Experiment 1 showed that V-NP/High-frequency condition OR clauses were rated easier (t(1,29)=3,31; p=.002).

Experiment 2 showed that – consistent with statistics - pronominal ORs were rated easier in the (NP-V, null/High-frequency condition), followed by NP-V, overt/Medium-frequency condition, while Os in the V-NP, overt/Low-frequency condition were rated the most difficult (F_1(2,58)=10.6; p< .000).

While the results are consistent with the predictions of usage-based accounts, they are are hard to explain in terms of structural, working-memory or discourse factors.

Complexity ratings in ORs with full NP subjects

Complexity Ratings Pronominal ORs
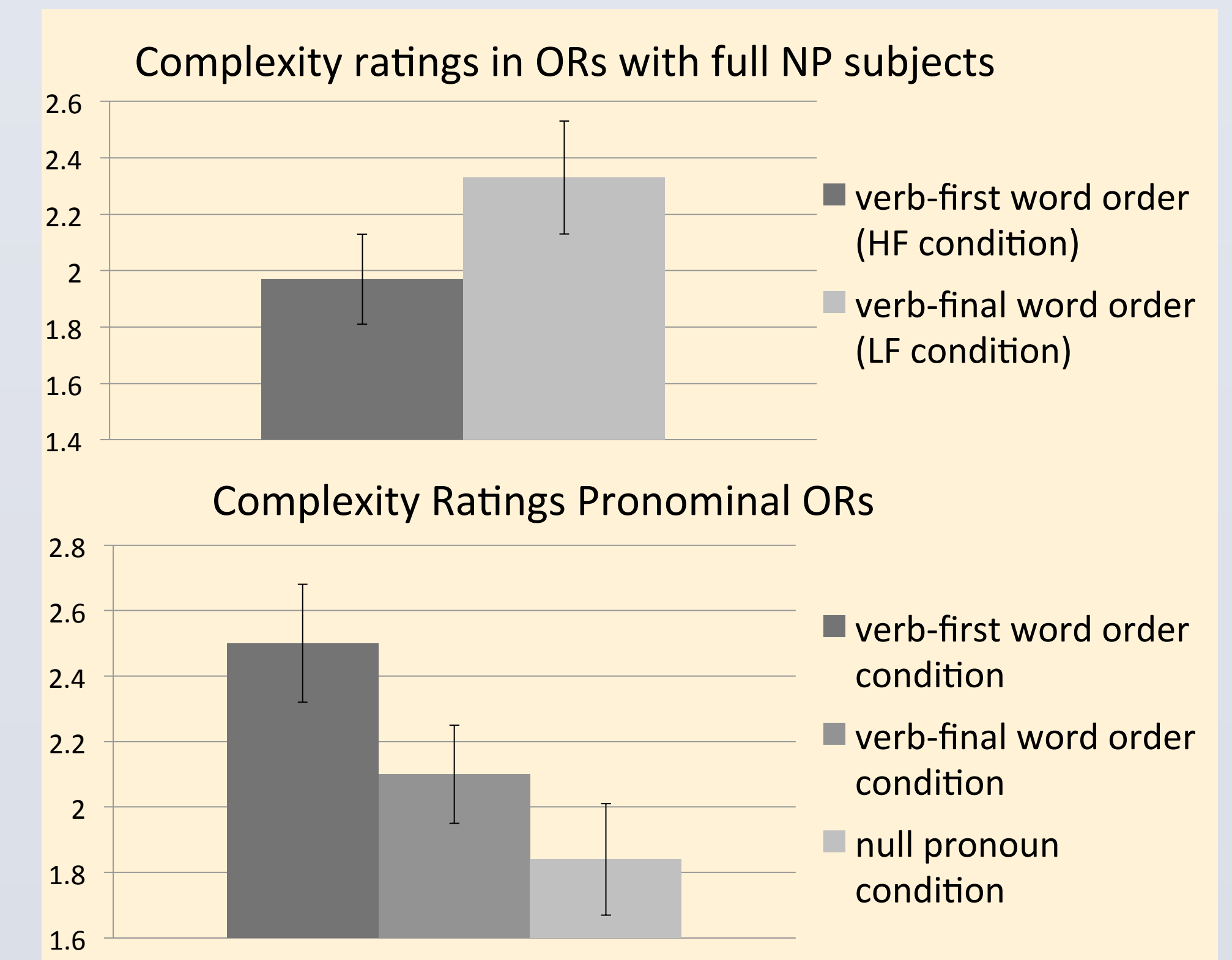
Conclusion
We conclude in favor of usage-based accounts, arguing that frequent pronominal OR constructions may become easier to access as a consequence of use.