

Antecedent frequency effects on anaphoric pronoun resolution: Evidence from Spanish.[☆]

Nerea Egusquiza^a, Eduardo Navarrete^b, Adam Zawiszewski^{a,*}

^aDepartment of Linguistics and Basque Studies, University of the Basque Country (UPV/EHU), Spain

^bDipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova

Abstract

High-frequency words are usually understood and produced faster than low-frequency words. Although the effect of word frequency is a reliable phenomenon in many domains of language processing, it remains unclear whether and how frequency affects pronominal anaphoric resolution. We evaluated this issue by means of two self-paced reading experiments. Native speakers of Spanish read sentences containing the anaphoric noun or pronoun at the subject syntactic position (Experiment 1) or at the object syntactic position (Experiment 2) while the antecedent of the anaphor was either a high-frequency or a low-frequency word. Results showed that nominal anaphors were read faster when referring to high-frequency than to low-frequency antecedents, and faster when referring to subjects than to objects. Critically, pronoun reading times were unaffected by the frequency and by the syntactic position of the antecedent. These results are congruent with theories assuming that syntactic information of the words is not frequency sensitive.

Keywords: word frequency, self-paced reading, anaphor resolution

1. Introduction

Word-frequency effects are among the most robust phenomena in language processing: words that are used more frequently are usually processed faster and more accurately. The word frequency effect is a reliable phenomenon observed under several circumstances. For instance, in comparison to low-frequency words, high-frequency words yielded faster responses in reading (e.g. Rayner & Duffy, 1986), lexical decision (e.g. Schilling, Rayner & Chumbley, 1998) and picture naming tasks (e.g. Oldfield & Wingfield, 1965; Almeida, Knobel, Finkbeiner & Caramazza, 2007). Also, linguistic performance of brain damaged

[☆]This research was supported by the Ministry of Science and Innovation (FFI2010-20472), Ministry of Economy and Competitiveness (FFI2012-31360), Department of Education, Universities and Research, Basque Government (IT665-13), Basque Government and University of the Basque Country UPV/EHU to Egusquiza (BFI-2012-219, EHUA13/39) and EC FP7/SSH-2013-1 AThEME (613465) to Zawiszewski. We are also grateful to Mikel Santesteban and Kepa Erdocia for their valuable help during data analysis as well as for their comments on the previous versions of the manuscript. **Please cite as:** Egusquiza, N., Navarrete, E. and Zawiszewski, A. (2016). Antecedent frequency effects on anaphoric pronoun resolution: Evidence from Spanish. *Journal of Psycholinguistic Research*, 45, 71-84. DOI: [10.1007/s10936-014-9325-3](https://doi.org/10.1007/s10936-014-9325-3)

*Corresponding author: adam.zawiszewski@ehu.es

patients is determined by word frequency, as reflected by the fact that high-frequency words are preserved more often in comparison to low-frequency words (e.g. Dell, 1990; Knobel, Finkbeiner & Caramazza, 2008). In addition, speakers tend to experience less tip-of-the-tongue states with high-frequency words than with low-frequency words (e.g. Brown, 2012). Although all these observations clearly suggest an advantage in the processing for high-frequency words compared to low-frequency words; it remains still unclear whether this is also the case during anaphor resolution. Anaphors are deictic words or phrases partially responsible for endowing local coherence to a text by binding a previously mentioned entity (the antecedent) to a referential expression (e.g. a pronoun). In the following example:

1. John saw Mary yesterday. She looked tired that day.

Mary is the object of the first sentence and the antecedent of the pronoun *she* of the second sentence. Anaphoric expressions like (1) have raised much interest in psycholinguistics research (e.g. Branco, McEnery & Mitkov, 2005; Hendrickx, Devi, Branco & Mitkov, 2011). In the last two decades, several studies have examined the word frequency effect in anaphoric comprehension reaching contrasting conclusions (e.g. Simner & Smyth, 1999; Van Gompel & Majid, 2004; Heine, Tamm, Hofmann, Bösel & Jacobs, 2006a; Heine, Tamm, Hofmann, Hutzler & Jacobs, 2006b; Lago, Chow & Phillips, 2011). These studies have explored whether the time involved in the comprehension of an anaphor depends on the lexical frequency of its corresponding antecedent noun. Critically, incongruent evidence has been reached and, up to date, there is still disagreement on the role of antecedent frequency during anaphor resolution (see below). The main aim of the current research was to shed some light on this issue.

The first study that explored the word frequency effect during anaphor resolution comes from Simner & Smyth (1999). Participants in that study read sentences containing a pronoun whose antecedent corresponded to a previously presented picture. The frequency of the noun depicted by the pictures was manipulated; nouns could be high or low-frequency words. Simner and Smyth did not report a frequency effect, that is, pronominal-sentence reading times were unaffected by the frequency of the antecedent noun. These authors interpreted the lack of frequency effects as congruent with the account proposed by Jescheniak & Levelt (1994) about the word frequency effect in speech production. According to Jescheniak and Levelt, lexical access in speech production involves at least two stages of processing. The first stage requires the selection of a semantically and syntactically specified representation (i.e. lemma) corresponding to the word; while the second stage requires the selection of its lexical-phonological representation (i.e. lexeme). Jescheniak and Levelt localize the frequency effect in word production at the lexeme representation. Simner & Smyth (1999) concluded that the lack of frequency effects in pronoun resolution would reflect the fact that, during pronoun reading, participants do access the lemma representation of the antecedent noun (in which grammatical information is stored) but do not access the lexeme representation. If the word frequency effect is localized at the lexeme representation, and assuming that such a representation is not accessed in pronoun resolution during language comprehension, the frequency effect of the antecedent noun would not be transferred to the pronoun. We refer here to the Simner and Smyth account as to the *lemma-access hypothesis*.

However, the conclusions reached by Simner & Smyth (1999) have been challenged by the study of Van Gompel & Majid (2004), in which a pronominal frequency effect was

reported. Specifically, Van Gompel and Majid observed that pronouns with low-frequency antecedent nouns were read faster in comparison to pronouns which high-frequency antecedent nouns. This result was interpreted in accordance with other studies showing that infrequent words require more attentional resources than more frequent words. For instance, [Malmberg & Nelson \(2003\)](#) suggest that “*low-frequency words are more difficult to process because they tend to have rather unusual features or configurations; resources initially devoted to the processing of relatively common features may be reallocated to unusual features that have not yet been fully decoded.*” (p. 38). The investment of more attentional resources on the processing of unusual features results in a relatively high probability of infrequent words being encoded in long-term memory. According to this, low-frequency words would become more salient (i.e. prominent) than high-frequency words, and when the antecedent has to be accessed during pronominal resolution, low-frequency words would be more quickly recognized than high-frequency words (e.g. [Glanzer & Adams, 1985, 1990](#); [Shiffrin & Steyvers, 1997](#); [Malmberg & Nelson, 2003](#)). Further evidence supporting this account comes from the Event-Related Potentials study conducted by [Heine et al. \(2006a\)](#). In that study, pronouns referring to high-frequency antecedents elicited a larger P300 component in comparison to those referring to low-frequency antecedents, suggesting a higher processing cost for the former than for the latter. Under the name of *saliency-hypothesis*, in this article we refer to the prediction that word saliency determines faster processing when pronouns are linked to low-frequency antecedent nouns compared to when these same pronouns refer to high-frequency antecedent nouns.

In a more recent study, [Lago et al. \(2011, but see Lago, 2014\)](#) observed a word frequency effect with faster reading times for pronouns with a high-frequency antecedent noun than with a low-frequency antecedent noun. These results are similar to those observed in pronominal utterances during language production tasks. For instance, [Navarrete, Basagni, Alario & Costa \(2006\)](#) reported faster naming latencies for pronominal sentences containing a pronoun with a high-frequency antecedent noun than with a low-frequency antecedent noun (see also [Finocchiaro & Caramazza, 2006](#)). The results by Lago and colleagues would suggest that, during pronoun resolution, the antecedent is retrieved up to a stage that is sensitive to lexical frequency, similarly to what has been argued in speech production studies (e.g. [Navarrete et al., 2006](#)). We will call *the full-access hypothesis* the prediction that the frequency of the antecedent should be transferred to the anaphoric pronoun.

1.1. Motivation of the current research

In light of these studies, it remains unclear how the frequency of the antecedent noun may affect anaphoric processing during language comprehension. Indeed, the three hypotheses described above make different predictions. The *lemma-access hypothesis* predicts no frequency effects; *the salience hypothesis* predicts a reversed frequency effect, with faster processing for pronouns referring to low-frequency words. Finally, *the full-access hypothesis* predicts the same effects in nouns than in pronouns, that is, faster reaction times with pronouns referring to high-frequency words. The main aim of the present study was to examine the influence of antecedent frequency during anaphoric pronoun resolution. There were two specific objectives. First, previous studies addressing this issue have been performed in non pro-drop languages, such as English and German, where the use of an overt subject is obligatory. It has been argued that pronoun processing may be language-specific ([Meyer & Bock, 1999](#)) and different for pro-drop and non pro-drop

languages (Fernández Soriano, 1989). These differences in pronoun processing between pro-drop and non pro-drop languages may be due to the fact that while the users of the non pro-drop languages must rely on the lexical characteristics of the (pronominal) anaphors, the speakers of pro-drop languages can (although do not have to) use null subjects (*pro*'s) in similar contexts. Consequently, it becomes relevant to explore anaphoric frequency effects in pro-drop languages, such as Spanish, which was used in the two experiments reported here. Spanish is a Romance pro-drop language with two grammatical genders (i.e. masculine and feminine) in which the form of gender-marked and number-marked pronouns depends on the grammatical gender and number of the referent noun.

As a second objective, we aimed to investigate whether the syntactic position of the antecedent modulates the frequency effect during anaphor resolution. This is relevant because some studies have reported faster reading times for anaphors when the antecedent was a subject than when it was an object (Purkiss, 1978, cited in Sanford & Garrod (1981)). Specifically, Purkiss reported faster reading times for sentences referring to topic antecedents (i.e. subjects) than for sentences referring to non-topic antecedents (i.e. objects), which suggests that syntactic position have a privileged status with respect to ease of accessing. Critically, all the studies that have explored so far the frequency effect during anaphor resolution have been tested only in those cases where the antecedent was a grammatical object. In order to explore the difference in anaphor resolution as a function of word frequency and the syntactic position, we conducted two experiments with the same materials, but manipulating syntactic context. Since it is a well known that subjects are more likely to prime subjects than objects are likely to prime subjects (e.g. Meyerhoff, 2009), in Experiment 1, both the antecedent and the anaphor occupied the subject syntactic position; while in Experiment 2, both occupied the object syntactic position. In Experiment 1, both the antecedent and the anaphor occupied the subject syntactic position; while in Experiment 2, both occupied the object syntactic position. Critically, the distance between the antecedent and the pronoun, in terms of the number of intervening words was kept constant across the two experiments, in order to allow for a direct comparisons between the two experiments.

Finally, the possibility that the discrepant results from the revised studies might be due to the differences in the experimental design cannot be discarded. For instance Simner & Smyth (1999) used pictures as antecedents whereas the other studies (Van Gompel & Majid, 2004; Heine *et al.*, 2006a) tested participants while reading written sentences. Thus, it is important to provide more empirical evidence on how frequency influences anaphor resolution while keeping experimental settings and materials identical.

1.2. Experimental overview

Two self-paced reading experiments are reported here with the aim to investigate whether or not word frequency modulates the way anaphoric pronouns are processed during language comprehension. In Experiment 1, we tested how the frequency of the antecedent noun located at the subject syntactic position modulates pronominal anaphor resolution; while in Experiment 2, we tested to what extent anaphor resolution is influenced by the frequency of the antecedent noun located at the object syntactic position. Three alternative possibilities were considered: (1) lexical frequency of the antecedent will not affect the way anaphoric pronouns are processed, thus supporting *the lemma-access hypothesis*; (2) pronouns referring to low-frequency antecedents will be processed faster than those referring to high-frequency antecedents. If so, *the saliency hypothesis* would be confirmed;

and (3), pronouns referring to high-frequency antecedents will be processed faster than those related to low-frequency ones, as proposed by the *full-access hypothesis*. In order to test the reliability of the frequency effect at the same region of the sentence in which the anaphoric pronoun was located, we added a noun phrase control condition in which the noun instead of the pronoun was presented (see below). The present research allowed us to investigate another relevant issue: the influence of syntactic class on anaphor resolution. To this respect, we hypothesized that, given the highly prominent status of the subject, anaphors referring to subject antecedents will be processed faster than those referring to the object antecedents (Arnold, 1998; Falk, 2006). Moreover, we explored also whether the expected frequency effect during anaphor resolution depends on its syntactic position. Three possibilities emerge here. One possibility is that the frequency effect in anaphor resolution is more pronounced when the antecedent is a subject than when the antecedent is an object, due to the fact that subjects have privileged status as compared to objects. However, one could also make the opposite prediction, that is, having objects a less privileged status, readers would need a full access to the antecedent just when the antecedent is an object in order to resolve the anaphor. If this is the case, we may also predict a larger frequency effect for an object than for a subject antecedent. Finally, another plausible scenario can take place, where a similar frequency effect will be observed both in the subject and in the object antecedent contexts. This would suggest that lexical frequency of the antecedent plays a similar role during anaphor resolution, independently on the syntactic status of the co-referring element (subject or object).

2. Experiment 1: Antecedent noun in a subject position

As mentioned above, previous studies exploring word frequency effects during anaphor resolution have localized the antecedent noun in the object position. In this experiment, we explored frequency effects during pronoun reading when the antecedent noun was in subject position. Participants were presented with two sentences. Antecedents could be high-frequency or low-frequency Spanish noun phrases (e.g. *hombre* or *bedel*, ‘man’ or ‘porter’ in English, respectively).

2.1. Method

2.1.1. Participants

Thirty-two monolingual native speakers of Spanish (eighteen women), aged 18-46 years ($SD = 23.7$), were recruited from the University of the Basque Country (UPV/EHU) (Vitoria-Gasteiz campus). Participants in this and in the subsequent experiment performed only one experiment.

2.1.2. Materials

A total of 80 words were selected as experimental items. According to the B-Pal Spanish standard database (Davis & Perea, 2005), half of them were high-frequency words (mean 67.62 occurrences per million) and the other half were low-frequency words (mean 1.72 occurrences per million). A total of 160 experimental sentence pairs were constructed. The first sentence of each pair contained a singular masculine subject and a singular feminine object. The second sentence started either with the repeated masculine noun or with the anaphoric pronoun “Él” (he), which co-referred unambiguously with the

subject of the preceding sentence. Experimental sentences were distributed over 4 conditions: (1) Repeated noun phrase, High-frequency antecedent; (2) Repeated noun phrase, Low-frequency antecedent; (3) Pronoun, High-frequency antecedent; and (4) Pronoun, Low-frequency antecedent (see Table 1).

Table 1: Sample of the materials used in the Experiment 1.

EXPERIMENT 1	
Repeated noun phrase	01. [Un ministro] criticó a la reina durante el discurso de ayer. [A minister] criticized the queen during the speech of yesterday.
	HF [El ministro] censuró la monarquía. [The minister] disapproved the monarchy. <i>Meaning: A minister criticized the queen during yesterday’s speech.</i> <i>The minister disapproved the monarchy.</i>
	02. [Un senador] criticó a la reina durante el discurso de ayer. [A senator] criticized the queen during the speech of yesterday.
	LF [El senador] censuró la monarquía. [The senator] disapproved the monarchy. <i>Meaning: A senator criticized the queen during yesterday’s speech.</i> <i>The minister disapproved the monarchy.</i>
Pronoun	03. [Un ministro] criticó a la reina durante el discurso de ayer. [A minister] criticized the queen during the speech of yesterday.
	HF [Él] censuró la monarquía. [He] disapproved the monarchy. <i>Meaning: A minister criticized the queen during yesterday’s speech.</i> <i>He disapproved the monarchy.</i>
	04. [Un senador] criticó a la reina durante el discurso de ayer. [A senator] criticized the queen during the speech of yesterday.
	LF [Él] censuró la monarquía. [He] disapproved the monarchy. <i>Meaning: A senator criticized the queen during yesterday’s speech.</i> <i>He disapproved the monarchy.</i>

2.2. Design

Four counterbalanced lists containing forty experimental sentences (each list containing 10 items per condition) were created in such manner that each participant saw just one version of the same item. In addition, eighty filler sentences and forty yes/no comprehension questions (twenty about the filler sentences and the other twenty about the experimental ones) were included. Filler sentences were randomly intermixed with the experimental sentence pairs of each list.

2.3. Procedure

Participants were tested individually. Self-paced moving window paradigm was used to perform the experiment. Each list contained forty stimuli sentence pairs randomly intermixed with eighty filler sentences. Participants sat in front of a 19-inch screen and a

keyboard in a quiet, lit up booth and read the instructions on the screen. The task consisted of a practice session (3 sentences) followed by the set of experimental sentences (120 sentences, that is, 40 experimental sentences and 80 filler sentences). Participants were told to read carefully at his or her normal rate. They were first presented with an array of preview dashes: each dash corresponded to a word in the current sentence pair. Every time the reader pressed the “space bar”, a constituent of the sentence appeared, replacing the corresponding dashes. Sentences were unmasked one constituent at a time, keeping the past and future parts of the sentence hidden. Space bar automatically served readers an incoming sentence pair and allowed them to proceed with the task. Two optional short breaks every forty sentences were included to prevent subjects from fatigue. The entire experimental session lasted 15 minutes and was controlled by Linger software (Rohde, 2001).

2.4. Analysis

Reaction times below 100 ms or above 2500 ms and those above 3 standard deviations from the participant’s mean were excluded from the analyses (2 %). Two within-subject factors, Frequency (High Frequency vs. Low Frequency) and Anaphor (Repeated noun phrase vs. Pronoun), and their interaction, were included in the analysis. Separate analyses were carried out treating participants and items as random factors, yielding F_1 and F_2 statistics respectively. Analyses were performed on the Anaphor region as well as on the Antecedent region, where the noun was located. Finally, given the fact that the effects sometimes emerge on words following the critical region (Rayner & Pollatsek, 1989; Van Gompel & Majid, 2004), analyses were also performed in the Post-anaphoric region.

2.5. Results

With respect to the reading times, in the Anaphor region, the main effect of Frequency was significant in the participants analysis and marginally significant in the item analysis ($F_1(1, 31) = 5.27, p < 0.03$; $F_2(1, 39) = 3.21, p = 0.08$). The main effect of Anaphor was significant ($F_1(1, 31) = 69.56, p < 0.001$; $F_2(1, 39) = 95.01, p < 0.001$). The interaction between these two factors was significant in the subject analysis and marginally significant in the item analysis ($F_1(1, 31) = 5.77, p < 0.02$; $F_2(1, 39) = 2.83, p = 0.1$). Paired t-tests revealed that the interaction was due to the fact that Frequency effect was observed in the Repeated noun phrase anaphoric condition ($t_1(31) = -2.87, p < 0.01$; $t_2(39) = -2.26, p < 0.02$), but not in the Pronoun anaphoric condition pronouns ($ts < 1$) (see Table 2).

In the Antecedent region, the main effect of Frequency was significant ($F_1(1, 31) = 14.67, p < 0.01$; $F_2(1, 39) = 13.08, p < 0.01$), but the effect of Anaphor was not significant ($F_s < 1$). There was no interaction between these two factors ($ts < 1$). In the Post-anaphor region, the main effect of Frequency yielded no significance ($F_1(1, 31) = 1.96, p > 0.17$; $F_2 < 1$), but the main effect of Anaphor was significant ($F_1(1, 31) = 15.85, p < 0.001$; $F_2(1, 39) = 27.65, p < 0.001$). The interaction between both factors was not significant ($F_s < 1$).

2.6. Discussion

Results in the Antecedent region reported a word frequency effect, with high-frequency nouns eliciting faster reaction times than low-frequency nouns, replicating previous studies (e.g. Forster & Chambers, 1973; Rayner & Duffy, 1986; Besner & McCann, 1987;

Table 2: Mean reading latencies (RT) and standard deviations (SDE) in ms according to Frequency and Anaphor type at the Anaphor region in Experiments 1 and 2.

	Experiment 1 (Subject antecedent)	Experiment 2 (Object antecedent)
Repeated noun	RT (SDE)	RT (SDE)
HF antecedent	534 (22)	687 (27)
LF antecedent	569 (21)	744 (40)
Effect	-35	-57
Pronoun	RT (SDE)	RT (SDE)
HF antecedent	463 (14)	499 (16)
LF antecedent	459 (16)	500 (19)
Effect	4	-1

Schilling *et al.*, 1998). In the Anaphor Region, high-frequency nouns were read faster than low-frequency nouns, suggesting the high reliability of the effect (for an overview see Ellis, 2002). Critical for our purposes, there was no frequency effect in the Anaphor region in the Pronoun condition. Before drawing conclusions from these results, in the next experiment, the frequency effect was explored when the antecedent’s syntactic position is an object. This is important because the syntactic prominence of the subject might have obscured any possible frequency effect. If so, by creating a context where the antecedent’s position is not prominent, this would help avoiding this confounds.

3. Experiment 2: Antecedent noun in an object position

The same materials and procedure as in Experiment 1 were used here with the only difference that sentences were modified in order to locate the antecedent noun and the anaphor at the object syntactic position. In order to allow the comparison between the two experiments, the distance in the total number of words between the antecedent and the anaphoric pronoun in Experiment 2 was the same as in Experiment 1.

3.1. Method, Participants and Materials

Thirty-two native speakers of Spanish (twenty-four women) from the same population as in Experiment 1, aged 18-34 years ($SD = 22.6$), took part in the experiment. The same materials as in Experiment 1 were used here with the difference that the sentences were modified in order to place the antecedent at the object position (see Table 3). The same design and procedure as in Experiment 1 were used.

3.2. Results

Following the same criteria as in Experiment 1, 2% of the data points were discarded from the analysis.

In the Anaphor region, the main effect of Frequency was significant in the participant analysis and marginally significant in the item analysis ($F_1(1, 31) = 4.99, p < 0.04$; F_2

Table 3: Sample of the materials used in the Experiment 2.

EXPERIMENT 2	
Repeated noun phrase	HF 01. La reina criticó [a un ministro] durante el discurso de ayer. The queen criticized [a minister] during the speech of yesterday. Posteriormente arremetió [contra el ministro] en el parlamento. Later attacked [against the minister] at the parliament. <i>Meaning: The queen criticized a minister during yesterday’s speech. Later on, she attacked the minister at the Parliament.</i>
	LF 02. La reina criticó [a un senador] durante el discurso de ayer. The queen criticized [a senator] during the speech of yesterday. Posteriormente arremetió [contra el senador] en el parlamento. Later attacked [against the senator] at the parliament. <i>Meaning: The queen criticized a senator during yesterday’s speech. Later on, she attacked the senator at the Parliament.</i>
Pronoun	HF 03. La reina criticó [a un ministro] durante el discurso de ayer. The queen criticized [a minister] during the speech of yesterday. Posteriormente arremetió [contra él] en el parlamento. Later attacked [against him] at the parliament. <i>Meaning: The queen criticized a minister during yesterday’s speech. Later on, she attacked him at the Parliament.</i>
	LF 04. La reina criticó [a un senador] durante el discurso de ayer. The queen criticized [a senator] during the speech of yesterday. Posteriormente arremetió [contra él] en el parlamento. Later attacked [against him] at the parliament. <i>Meaning: The queen criticized a senator during yesterday’s speech. Later on, she attacked him at the Parliament.</i>

(1, 39) = 3.33, $p = 0.076$), as well as the main effect of Anaphor ($F_1(1,31) = 83.44$, $p < 0.001$; $F_2(1,39) = 209.5$, $p < 0.001$). The interaction between these two factors was significant ($F_1(1, 31) = 5.21$, $p < 0.03$; $F_2(1, 39) = 3.90$, $p = 0.055$). Paired t-tests revealed that the interaction was due to the frequency effect observed in the Repeated noun phrase condition ($t_1(31) = -2.52$, $p < 0.02$; $t_2(39) = -2.09$, $p < 0.05$), but not in the Pronoun anaphoric condition ($ts < 1$) (see Table 2).

In the Antecedent region, the main effect of Frequency was marginally significant in both analyses ($F_1(1, 31) = 3.7$, $p = 0.06$; $F_2(1, 39) = 2.91$, $p = 0.09$). No other effects were found at this region ($F_s < 1$). In the Post-anaphor region, the only significant effect was the main effect of Anaphor ($F_1(1, 31) = 5.28$, $p < 0.03$; $F_2(1, 39) = 4.33$, $p < 0.05$). No other effects yielded statistical significance at this region ($F_s < 1$).

3.3. Discussion

Results of this experiment replicated those of the Experiment 1. First, a word frequency effect was found in the Antecedent region, with faster reaction times elicited by high-frequency nouns than by low-frequency nouns. At the Anaphora position, the frequency effect was found in the Repeated noun condition, with high-frequency nouns being read

faster than low-frequency nouns; but not in the Pronoun condition, replicating Experiment 1.

3.4. Cross-experiment analysis

We carried out an additional analysis comparing Experiments 1 and 2 in the Anaphor region, in order to further explore whether syntactic position of the antecedent plays a role in anaphor resolution. We addressed two critical issues (see above Experimental overview). First, we explored whether the syntactic position of the antecedent affects anaphor resolution by analyzing the factor Experiment (Experiment 1 vs. Experiment 2). The main effect of Experiment was significant ($F_1(1, 62) = 10.40, p < 0.01$; $F_2(1, 78) = 402.86, p < 0.001$), with faster reading times in Experiment 1 than in Experiment 2 (507 ms and 608 ms, respectively). In other words, reading anaphora was faster with subject than with object syntactic positions. The interaction between the factor Experiment and Anaphor type (Repeated noun vs. Pronoun) was significant ($F_1(1, 62) = 22.49, p < 0.001$; $F_2(1, 78) = 17.15, p < 0.001$). Further paired t-tests revealed that the interaction was due to the fact that the Experiment effect was larger on the Repeated noun (164 ms; $t_1(62) = -4.12, p < 0.001$; $t_2(78) = -13.58, p < 0.001$) than in the Pronoun anaphora (39 ms; $t_1(62) = -1.15, p > 0.25$; $t_2(78) = -18.99, p < 0.001$).

However, it is known that besides lexical frequency the context of use of lexical items plays a significant role during language comprehension (Gahl & Garnsey, 2006; Brown & Rivas, 2012). In order to make sure that the reported effects were due to the lexical frequency of the nouns used in the experiments rather than to the frequency they occur in a given syntactic context (subject vs. object), we performed a comparison based on GOOGLE where we contrasted the occurrence of the nouns in subject and object positions with the verbs used in both experiments (i.e. *un ministro criticó* ‘a minister criticized’ and *criticó a un ministro* ‘criticized a minister’). Two variables were used to perform statistical analyses: Frequency (high / low) and Position (subject / object). Besides the expected Frequency effect ($F(1, 39) = 12.41, p = 0.001$), no Position effect ($F(1, 39) = 0.804, p = 0.375$) or Frequency x Position interaction ($F(1, 39) = 0.950, p = 0.336$) were found, suggesting that the reported findings must be due to the lexical frequency of the nouns rather than to the specific position these nouns appear within the sentences.

A second empirical question we aimed to address was whether the Frequency effect during anaphor resolution depends on the syntactic position of the antecedent. In order to test this, the interaction between Anaphor type and Frequency was analyzed. The statistical analysis showed a significant interaction between these two factors ($F_1(1, 62) = 7.16, p < 0.01$; $F_2(1, 78) = 7.343, p < 0.01$). The interaction was due to the fact that high-frequency nouns were read faster than low-frequency nouns (612 ms vs. 659 ms, $t_1(63) = -3.63, p < 0.01$; $t_2(79) = -2.25, p < 0.03$); while in the Pronoun condition it was irrelevant whether the antecedent was a high or a low-frequency word (488 ms vs. 483 ms, $t_1(63) = 0.39, p > .69$; $t_2(79) = -0.742, p > 0.45$). Figure 1 plots a summary of the results.

Finally, in order to investigate whether the faster reading of repeated high frequency vs. low frequency nouns is due to the repeated mention rather than due to the frequency of the noun itself, we performed additional statistical tests with full nouns considering Frequency (high / low) and Mention (antecedent / anaphor) as within-subject factors and Experiment (Exp. 1 / Exp. 2) as between-subject factor. The analyses revealed a significant effect of Frequency ($F_1(1, 62) = 17.07, p < 0.001$; $F_2(1, 78) = 13.09, p = 0.001$) and

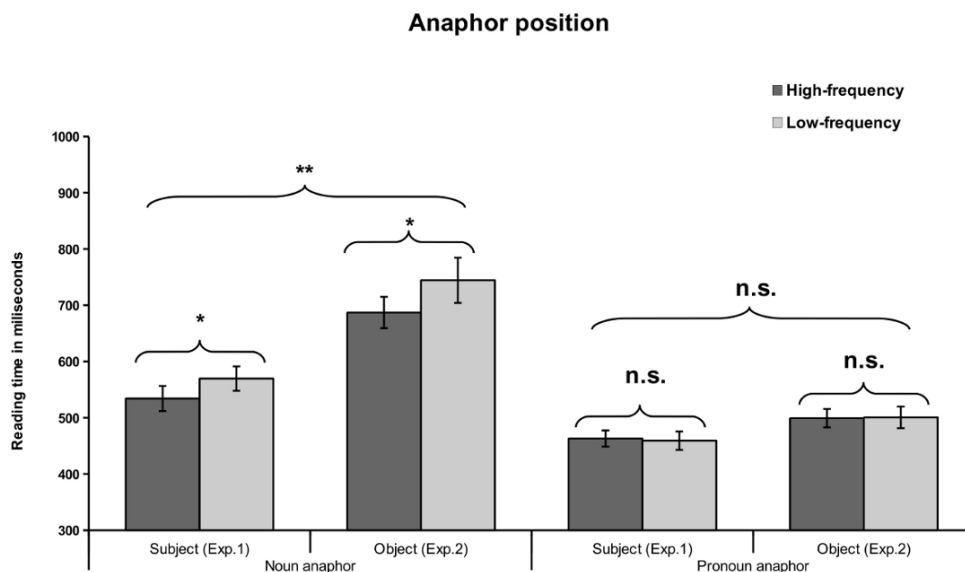


Figure 1: Mean reading times at the Anaphor position in the Experiments 1 and 2. ** $p < 0.01$; * $p < 0.05$

Mention ($F_1(1, 62) = 21.93, p < 0.001$; $F_2(1, 78) = 17.43, p < 0.001$), that is, faster reading for high frequency (649 ms, SDE = 20.85) than low frequency items (704 ms, SDE = 25.85) and faster reading in anaphoric contexts (560 ms, SDE = 14.59) than in the antecedent contexts (712 ms, SDE = 25.53).

4. General discussion

Two self-paced reading experiments were performed in order to investigate how word frequency and syntactic position of the antecedent influence anaphoric pronoun processing in Spanish. In Experiment 1, we tested native speakers of Spanish while reading items containing either a high-frequency or a low-frequency antecedent in the first clause, and a repeated noun phrase or an anaphoric pronoun in the second. Both the antecedent and the anaphor were placed at the subject position. Critical for our purposes here, anaphoric pronoun resolution was unaffected by the word frequency status of the antecedent. Critically, this null effect was observed in the context of other reliable phenomena. First, the results showed a frequency effect at the antecedent position in which infrequent antecedents took longer to read than frequent antecedents. This result is indeed consistent with previous studies (Forster & Chambers, 1973; Rayner & Duffy, 1986; Besner & McCann, 1987; Schilling *et al.*, 1998). Second, in the Anaphor region nouns were read slower in comparison to pronouns, in congruency with the repeated-name penalty effect reported in previous studies (e.g. Gordon, Grosz & Gilliom, 1993; Kennison & Gordon, 1997), which refers to a processing delay for repeated nouns when compared to pronoun processing. This effect has been reported for syntactic subjects that co-refer with the subjects, but it does not occur for direct objects that refer to the object of the preceding sentence (Kennison & Gordon,

1997). And third, in the Anaphor region, noun phrases showed a frequency effect, with high-frequent words being read faster than low-frequent words (e.g. Ellis, 2002). In Experiment 2, we observed a similar pattern of results to the one detected in the Experiment 1. More precisely, at the Anaphor region, there was not effect of word frequency when the anaphor was a pronoun. However, when the Anaphor region was occupied by a noun phrase, the frequency effect emerged, with high-frequency nouns being read faster than low-frequency nouns. These last results challenge Gordon *et al.* (1993) and Kennison & Gordon (1997) findings. These authors found no repeated-name penalty effects in the conditions where the antecedent was not a topic or a subject of the sentence (i.e. when it was an object of the sentence). Our results indicate not only that the effect is present when the antecedent is the syntactic subject of a sentence (as in Experiment 1), but also when it is the object (as in Experiment 2). Furthermore, the frequency effect was significantly larger when the noun occupied an object (57 ms) than a subject syntactic position (35 ms) (see however Kennison & Gordon, 1997). Importantly, the distance between the anaphoric expression and its antecedent was identical in both experiments and, consequently, the degree of accessibility remained the same. Therefore, the effect must be attributed to structural differences between the antecedent positions (subject vs. object) rather than to other factors (e.g. working memory load), although the possibility that stronger priming might have occurred when the nouns had the same syntactic role (subject) in comparison to the contexts where the antecedent was a direct object but the anaphor was a prepositional object cannot be discarded. Thus, in light of these data, syntactically prominent arguments such as subjects are easier to refer to than less prominent arguments (i.e. objects). These findings are in line with other experimental results showing for instance that subject-relative clauses are easier to process than object-relative clauses (e.g. Traxler & Morris, 2002, but see also Carreiras, Duñabeitia, Vergara, de la Cruz-Pavía & Laka (2010) for an opposite finding). On the other hand, the fact that participants were faster in processing anaphors referring to subjects compared to those referring to objects may be related to the order of mention effect, that is, the advantage in reaccessing first-mentioned characters within a clause. It does not depend on linguistic factors and occurs even if the first participant is not the initial word in the sentence (Gernsbacher & Hargreaves, 1988). According to Gernsbacher (1990), during sentence processing, comprehenders devote more attention to first participants, because initial elements are considered the foundations of discourse understanding.

5. Conclusion

The main aim of the current research was to test for frequency effects on pronoun anaphoric comprehension. In the two experiments reported here, no such effect was obtained. In relation to the three hypothesis described above, these results support the lemma-reaccess hypothesis proposed by Simner & Smyth (1999), which postulates that during pronoun resolution there is not access to the lemma level (i.e. syntactic information) of the antecedent noun. According to such an account, the word frequency effect would be located at the lexeme level, which would not be accessed during pronoun resolution in sentence comprehension. Finally, the current research extended such conclusion to a pro-drop language (such as Spanish) and to the syntactic context where the anaphor occupies subject position.

References

- Almeida, J., Knobel, M., Finkbeiner, M. & Caramazza, A. (2007). The locus of the frequency effect: When recognizing is not enough. *Psychonomic Bulletin and Review*, **14**, 1177–1182.
- Arnold, J. (1998). *Reference form and discourse patterns*. Ph.D. thesis, Stanford University.
- Besner, D. & McCann, R.S. (1987). Word frequency effects and pattern distortion in visual word identification and production: An examination of four classes of models. In M. Coltheart, ed., *Attention and performance XII: The psychology of reading*, 201–219, Erlbaum, Hillsdale, NJ.
- Branco, A., McEnery, T. & Mitkov, R., eds. (2005). *Anaphora processing. Linguistic, cognitive and computational modelling. Current Issues in Linguistic Theory series*. John Benjamins. Amsterdam.
- Brown, A.S. (2012). *The tip of the tongue state*. New York: Psychology Press.
- Brown, E. & Rivas, J. (2012). Grammatical relation probability: How usage patterns shape analogy. *Language Variation and Change*, **24**, 317–341.
- Carreiras, M., Duñabeitia, J., Vergara, M., de la Cruz-Pavía, I. & Laka, I. (2010). Subject relative clauses are not universally easier to process: Evidence from Basque. *Cognition*, **115**, 79–92.
- Davis, C.J. & Perea, M. (2005). Buscapalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, **37**, 665–671.
- Dell, G. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, **5**, 313–349.
- Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, **24**, 143–188.
- Falk, Y.N. (2006). *Subjects and Universal Grammar*. Cambridge: Cambridge University Press.
- Fernández Soriano, O. (1989). Strong pronouns in null subject languages and the avoid pronoun principle. *MIT Working Papers in Linguistics*, **11**, 228–240.
- Finocchiaro, C. & Caramazza, A. (2006). The production of pronominal clitics: Implications for theories of lexical access. *Language and Cognitive Processes*, **21**, 141–180.
- Forster, K.I.F. & Chambers, S.M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, **12**, 627–635.
- Gahl, S. & Garnsey, S.M. (2006). Knowledge of grammar includes knowledge of syntactic probabilities. *Language*, **82**, 405–410.
- Gernsbacher, M.A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- Gernsbacher, M.A. & Hargreaves, D. (1988). Accessing sentence participants: the advantage of first mention. *Journal of Memory and Language*, **27**, 699–717.
- Glanzer, M. & Adams, J.K. (1985). The mirror effect in recognition memory. *Memory and Cognition*, **13**, 8–20.
- Glanzer, M. & Adams, J.K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16**, 5–16.

- Gordon, P.C., Grosz, B.J. & Gilliom, L.A. (1993). Pronouns, names and the centering of attention in discourse. *Cognitive Science*, **17**, 311–348.
- Heine, A., Tamm, S., Hofmann, M., Bösel, R.M. & Jacobs, A.M. (2006a). Event-related theta activity reflects memory processes in pronoun resolution. *Cognitive Neuroscience and Neuropsychology*, **17**, 1835–1839.
- Heine, A., Tamm, S., Hofmann, M., Hutzler, F. & Jacobs, A.M. (2006b). Does the frequency of the antecedent noun affect the resolution of pronominal anaphors? an ERP study. *Neuroscience Letters*, **400**, 7–12.
- Hendrickx, I., Devi, S., Branco, A. & Mitkov, R. (2011). Anaphora processing and applications. 8th discourse anaphora and anaphor resolution colloquium, daarc 2011. Revised Selected Papers, Springer, Berlin-Heidelberg, Faro, Portugal.
- Jescheniak, J.D. & Levelt, W.J.M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **20**, 824–843.
- Kennison, S.M. & Gordon, P.C. (1997). Comprehending referential expressions during reading: Evidence from eye tracking. *Discourse Processes*, **24**, 229–252.
- Knobel, M., Finkbeiner, M. & Caramazza, A. (2008). The many places of frequency: Evidence for a novel locus of the lexical frequency effect in word production. *Cognitive Neuropsychology*, **25**, 256–286.
- Lago, S. (2014). *Memory and Prediction in Cross-Linguistic Sentence Comprehension*. Ph.D. thesis, University of Maryland.
- Lago, S., Chow, W.Y. & Phillips, C. (2011). Word frequency affects pronouns and antecedents identically: Distributional evidence. CUNY Conference on Human Sentence Processing, Stanford University.
- Malmberg, K.J. & Nelson, T.O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory and Cognition*, **31**, 35–43.
- Meyer, A.S. & Bock, J.K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language*, **41**, 281–301.
- Meyerhoff, M. (2009). Replication, transfer and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change*. **21**, 1–21, **21**, 1–21.
- Navarrete, E., Basagni, B., Alario, F.X. & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology*, **59**, 1681–1690.
- Oldfield, R. & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology: General*, **118**, 43–71.
- Purkiss, E. (1978). The effect of foregrounding on pronominal reference, unpublished undergraduate thesis, Glasgow University, Glasgow, Scotland.
- Rayner, K. & Duffy, S.A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, **14**, 191–201.
- Rayner, K. & Pollatsek, A. (1989). *The psychology of reading*. Prentice Hall, Englewood Cliffs, NJ.
- Rohde, D. (2001). Linger. <http://tedlab.mit.edu/~dr/Linger/>.
- Sanford, A.J. & Garrod, S.C. (1981). *Understanding Written Language*. Chichester: John Wiley & Sons.
- Schilling, H.H., Rayner, K. & Chumbley, J.I. (1998). Comparing naming, lexical decision,

- and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition*, **26**, 1270–1281.
- Shiffrin, R.M. & Steyvers, M. (1997). A model for recognition memory: Rem - retrieving effectively from memory. *Psychonomic Bulletin and Review*, **4**, 145–166.
- Simner, J. & Smyth, R. (1999). Phonological activation in anaphoric lexical access. *Brain and Language*, **68**, 40–45.
- Traxler, M. & Morris, R., R.K. and Seely (2002). Processing subject and object relative clauses. evidence from eye movements. *Journal of Memory and Language*, **47**, 69–90.
- Van Gompel, R.G.P. & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, **90**, 255–264.