Coreference and antecedent frequency effects in Spanish: An eye-tracking study[1]

NEREA EGUSQUIZA, KEPA ERDOZIA, ADAM ZAWISZEWSKI

Abstract

In this paper, we explore the nature of the antecedent representations involved during anaphoric pronoun resolution in Spanish. We assume a serial, anaphoric lexical access account derived from the speech production model of Jescheniak and Levelt (1994); Levelt *et al.* (1999) and use word-frequency effects in order to investigate whether antecedent representations vary across languages during coreference resolution by testing Spanish native speakers and comparing our results to those obtained previously in English and German. We report and discuss the results from an eye-tracking while reading experiment in light of the current hypotheses.

Background

Sentence comprehension requires immediate access to the mental lexicon, where conceptual, grammatical and lexical information is stored. We say that speakers understand a sentence, a text or a discourse when they have an accurate mental representation of it. Needless to say, the longer the text or the speech, the more likely will be to find expressions referring to previous entities in the discourse. Anaphoric pronouns belong to this class of referential expressions which depend on the lexical properties of their referent or *antecedent* for interpretation. Evidence from agreement studies in Spanish, Italian, French and German shows that anaphoric pronoun resolution involves retrieving grammatical gender information from its antecedent representation in memory (Garnham *et al.,* 1995; Di Domenico and De Vicenzi, 1995; Cacciari *et al.,* 1997; Irmen and Knoll, 1999; Arnold *et al.,* 2000; Rigalleau *et al.,* 2004; Hammer *et al.,* 2007). Nevertheless, as suggested by Meyer and Bock (1999), antecedent representations could, in principle, differ qualitatively across languages based on the kind of information needed to mediate coreference. It might be the case that in languages like English, for instance, discourse/conceptual information will suffice to interpret anaphoric pronouns (Hankamer and Sag, 1976; Sag and Hankamer, 1984; Cloitrew and Bever, 1988; Lucas *et al.,* 1990). Hence, it is unclear what kind of antecedent representation is retrieved during anaphoric pronoun resolution and whether it is language specific or not.

Here, we adopt a serial, two-stage, activation-based, lexical access model derived from the speech production model of Jescheniak and Levelt (1994); Levelt *et al*. (1999), which suggests that lexical entries store information at two separate levels: grammatical and semantic information at the lemma level and orthographic and phonological information at the lexeme level. Evidence for a separate access to each informational layer during processing comes from a translation latency task where Jescheniak and Levelt (1994) found that homophones (words that only share the lexeme) inherited the lexical frequency of their control pairs –a word frequency effect which suggests that the locus of the effect lies at the lexeme level (see also Laubstein, 1999, Laubstein, 2002; for a dissenting view, see Bonin and Fayol, 2002; Finocchiaro and Caramazza, 2006; Navarrete *et al.,* 2006).

In order to investigate what kind of antecedent representation is retrieved during pronoun processing in Spanish, we use the *word-frequency effect*, which is one of the most robust phenomena in experimental psycholinguistics. It refers to the fact that frequently used words are processed faster and more accurately than infrequent ones. High-frequency words yield faster responses in reading (Rayner and Duffy, 1986), lexical decision (Schilling *et al.,* 1998) and picture naming tasks (Oldfield and Wingfield, 1965; Almeida *et al.,* 2007). Furthermore, data from brain damaged patients suggest that high-frequency words are preserved more often than low-frequency words (Dell, 1990; Colangelo *et al.,* 2004; Knobel *et al.,* 2008) and that speakers experience less tip-of-the-tongue states with high-frequency words (Brown, 2012; Gianico, 2010; Vitevitch and Sommers, 2003). Although all these observations suggest a processing advantage for high-frequency words compared to low-frequency words during lexical access, it remains an open question whether anaphoric pronoun resolution depends on the lexical frequency of the antecedent.

Hypotheses

In order to account for the results found in previous literature, three hypotheses have been proposed: the *full re-access account*, the *lemma re-access hypothesis* and the *saliency account*.

The *full reaccess account* states that pronoun comprehension is similar to lexical access for non-anaphoric words: readers will retrieve the same information from the antecedent as when they first accessed to it. Hence, it predicts that the lexical frequency of the antecedent should be transferred to the anaphoric pronoun. This effect has been observed in language production tasks. Navarrete *et al*. (2006), for instance, reported faster naming latencies for sentences containing a pronoun with a high-frequency antecedent noun than with a low-frequency antecedent noun. Similarly, participants tested by Finocchiaro and Caramazza (2006) elicited faster pronominal clitic production latencies when the replaced noun was highly frequent.

The *lemma-reaccess account*, however, assumes that pronoun processing differs from non-anaphoric lexical access in that there is only partial re-access to the antecedent representation in memory; more specifically, to its lemma, where grammatical and semantic information is stored (Simner and Smyth, 1998). Based on evidence from speech production showing that the word frequency effect lies at the lexeme (Jescheniak and Levelt, 1994), the lemma reaccess account predicts that pronouns referring to low-frequency nouns should not be harder or easier to process than those linked to high-frequency nouns (Simner and Smyth, 1999). In a series of experiments carried out by Simner and Smyth (1999), participants read English sentences containing a pronoun

whose antecedent had been previously presented in picture form. The lexical frequency of the depicted noun was either high or low. Simner and Smyth found that reading times for sentences containing a pronoun were unaffected by the lexical frequency of the antecedent. Further evidence supporting this account comes from Lago (2014), where she manipulated the lexical frequency of English common and proper noun antecedents in two eye-tracking experiments and concluded that low-frequency antecedents had no additional cost compared to high-frequency antecedents during pronoun resolution.

The *saliency account* challenges any null antecedent frequency effect. Van Gompel and Majid (2004), for instance, ran an eye-tracking study in which participants read sentence pairs in English containing a pronoun referring to a high or low-frequency antecedent. Pronouns with low-frequency antecedents elicited shorter reading times at the post-anaphoric region in first-fixation, first-pass and total reading times measures. The authors explained their reverse word-frequency effect in terms of saliency (Pynte and Colonna, 2000): infrequent words tend to have rather unusual —salient— features and their processing is thus more costly compared to that of high-frequency words, as evidenced by the word-frequency effect in nouns. The investment of additional attentional resources on infrequent words increases their probability of being better encoded in long-term memory (Garnham *et al.*, 1996; Malmberg and Nelson, 2003; Foraker and McElree, 2007) and, consequently, low-frequency words are recognized more quickly than high-frequency words (O'Brien and Myers, 1985; Glanzer and Adams, 1990; Shiffrin and Steyvers, 1997; Garnham, 2001; Malmberg and Nelson, 2003; Nicol and Swinney, 2003). The *saliency account* predicts that word saliency determines faster processing when pronouns refer to low-frequency antecedent nouns. Further evidence supporting this account comes from the EEG study in Heine *et al.* (2006b), where pronouns referring to high-frequency antecedents elicited a larger P300 component than those referring to low-frequency antecedents in German, suggesting a higher processing cost for the former (see also Heine *et al.*, 2006a).

The three accounts described above make different predictions. The *full reaccess account* predicts the same effects in nouns as in pronouns, i.e. faster reaction times with pronouns referring to high-frequency words than with pronouns referring to low-frequency words. The *lemma-reaccess account* predicts no frequency effects, whereas the *saliency account* predicts a reverse frequency effect, with faster processing for pronouns referring to low-frequency words.

Exploratory research

Previously, in Egusquiza *et al.* (2016), we reported two self-paced reading tasks showing that pronoun resolution in Spanish was affected by the syntactic role of the antecedent noun, but unaffected by its lexical frequency, in support of the lemma-reaccess account. The reliability of our results was evidenced by the fact that the frequency manipulation was strong enough to replicate the word-frequency effect in the antecedent regions as well as in the control conditions. Nevertheless, the self-paced reading method might not have been sensitive enough to detect any antecedent frequency effects during pronoun processing. Therefore, in the present study, we opted for eye-tracking, a methodology with a higher temporal resolution and which makes reading more natural than self-paced.

The present study

The aim of the present study is to examine the nature of the antecedent representations involved during anaphoric pronoun resolution in Spanish by manipulating the word frequency of the antecedent noun in an eye-tracking experiment. Two factors were crossed: lexical frequency of the antecedent (high vs. low) and anaphor type (pronoun vs. repeated noun). *Repeated noun* refers to the noun phrase control condition where the antecedent was repeated in order to pre-empt a null frequency effect across conditions. Based on Rayner *et al*. (1995), frequency effects should persist at repeated noun phrases. We considered three hypotheses in our study: (i) faster reding times for pronouns referring to high-frequency antecedents (*full reaccess account*); (ii) no antecedent frequency effects for pronouns (*lemma-reaccess account*); (iii) faster processing for pronouns referring to low-frequency words (*saliency account*).

The eye-tracking experiment

The underlying initial assumption in eye-tracking studies is that eye movements are strongly linked to cognitive processing (Just and Carpenter, 1980; cf. Anderson *et al*., 2004). Reading times are interpreted as an index of language processing effort; with longer reading times (and more fixations) revealing processing difficulty compared to faster reading times (and less fixations), much as in self-paced reading tasks. Eye-movement data, however, offer a more detailed map of what happens during online sentence comprehension. Regressions (i.e. re-reading earlier parts of a text), for instance, cannot be captured in self-paced reading tasks and are interpreted as a sign of language processing difficulty at the discourse/semantic integration levels. Despite its higher temporal resolution compared to self-paced reading, eye-tracking while reading shows higher data loss rates due to word skipping and blinking (Rayner *et al*., 2011), among other factors. The method provides highly accurate information about where readers look, for how long they gaze and how many eye fixations they make in a region.

Method

Participants

A set of twenty-two native speakers of Spanish (twelve women), ages 18-24 (SD = 1.96), who had not participated in Egusquiza *et al*. (2016), were recruited from the University of the Basque Country (Vitoria-Gasteiz campus). All participants had normal or corrected-to-normal vision. Two of them were excluded from further analysis because their data accuracy was below 70%.

Materials and Design

The materials were adapted from those used in Experiment 2 in Egusquiza *et al*. (2016). Two factors were crossed: frequency of the antecedent noun (high vs. Low) and anaphor type (repeated noun vs. pronoun). 160 sentence pairs were distributed in four lists (10 items per condition per list) so that each participant read only one version of the same item. 80 filler sentences were randomly intermixed with the experimental sentence pairs of each list and 40 yes/

no comprehension questions (half of them about the experimental sentences) were included. Both filler and experimental sentences shared a similar structure and length.

Our materials differed from Experiment 2 (Egusquiza *et al*., 2016) in two aspects: (i) a three-word spillover region was included after the post-anaphoric one to capture any wrap-up, sentence-final effect; (ii) an additional discourse entity was inserted in the prepositional phrase following the verb in order to improve the naturalness of the materials in the repeated noun conditions. It mismatched in gender with the first element in the prepositional phrase (the retrieval target.) The subject was feminine and singular across conditions, omitted in the second sentence of each pair and it always mismatched the retrieval target in gender. Table 1 contains a sample of our materials.

Table 1

Sample of the materials used in the eye-tracking experiment

		EYE-TRACKING EXPERIMENT
Repeated noun phrase	HF	01. La senadora criticó **[a un ministro]** y a una parlamentaria durante el discurso de ayer. The senator$_{fem}$ criticized [a minister$_{masc}$] and a member of parliament$_{fem}$ during yesterday's speech. Posteriormente arremetió **[contra el ministro]** en los medios nacionales y europeos. Later, (she) attacked [the minister] on the national and European media. *Meaning: The senator criticized a minister and a member of parliament during yesterday's speech. Later, (she) attacked the minister on the national and European media.*
	LF	02. La senadora criticó **[a un banquero]** y a una alcaldesa durante el discurso de ayer. The senator$_{fem}$ criticized [a banker$_{masc}$] and a mayoress during yesterday's speech. Posteriormente arremetió **[contra el banquero]** en los medios nacionales y europeos. Later, (she) attacked [the banker] on the national and European media. *Meaning: The senator criticized a banker and a mayoress during yesterday's speech. Later, (she) attacked the banker on the national and European media.*
Pronoun	HF	03. La senadora criticó **[a un ministro]** y a una parlamentaria durante el discurso de ayer. The senator$_{fem}$ criticized a minister$_{masc}$ and a member of parliament$_{fem}$ during yesterday's speech. Posteriormente arremetió **[contra él]** en los medios nacionales y europeos. Later, (she) attacked [him] on the national and European media. *Meaning: The senator criticized a minister and a member of parliament during yesterday's speech. Later, (she) attacked him on the national and European media.*
	LF	04. La senadora criticó **[a un banquero]** y a una alcaldesa durante el discurso de ayer. The senator$_{fem}$ criticized [a banker$_{masc}$] and a mayoress during yesterday's speech. Posteriormente arremetió **[contra él]** en los medios nacionales y europeos. Later, (she) attacked [him] on the national and European media. *Meaning: The senator criticized a banker and a mayoress during yesterday's speech. Later, (she) attacked him on the national and European media.*

The discourse entities were taken from Es-Pal Spanish Lexical database (Duchon *et al.*, 2013), and pairs in the conjoined prepositional phrase were matched for relative frequency within each item to keep the prepositional-phrase saliency level constant. The Es-Pal Spanish Lexical database showed that the frequency type of the high-frequency antecedents (mean: 93.06 per million words, range: 18.25-912.26) was much higher than that of the low-frequency antecedents (mean: 3.10 per million words, range: 0.08-5.83).

Norming study. We ran a norming study where sixteen native speakers of Spanish judged the acceptability of our materials on a 7-point Likert scale, where 7 meant perfectly acceptable and 1 was completely unacceptable. Each participant read ten experimental sentences intermixed with thirty fillers. None of them took part in the following eye-tracking experiment. Target sentences with a rate higher than 3.5 were considered for the follow-up experiment. None of them had to be excluded. The mean score value for the four lists was 5.58 (range: 5.15-6.09).

Procedure

The experiment involved an eye-tracking reading task. Participants were tested individually in a dimed-light, sound-proof booth, and eye-movements were recorded using a remote Tobii X120 Eye Tracker interfaced with a PC. The sampling rate for recordings was 120 Hz. The maximum gaze angle was 36 degrees and viewing was binocular. Participants were seated 67 cm from a 19-inch LCD screen. The forty experimental sentence pairs were randomly intermixed with the eighty fillers and displayed on a single line in 13.5 pt. fixed width Arial white font on a black screen. Each participant was randomly assigned one of the four lists of stimuli. The experiment was implemented using the *Tobii Studio 3.2* software. A calibration check was run at the beginning of each individual session. Previous to the task, participants were instructed to read carefully at their normal rate and completed a short practice session with 5 sentence pairs and their corresponding yes/no comprehension questions. A fixation cross on the left edge of the computer screen was displayed for 800 milliseconds before each item. The position of the fixation cross coincided with the beginning of every sentence to prevent participants from random looking and to make sure that the starting point of every sentence was the same for all conditions and items. Forty yes/no comprehension questions were included to ensure participants were attending to the stimuli. The entire experimental session lasted approximately 45 minutes and six optional breaks were included along the task. All participants had at least one break.

Analysis

Only subjects whose data accuracy was 70% or above were taken into consideration for further analysis. Two subjects did not meet this criterion and thus, were excluded. Skips of a region in any particular measure were treated as missing data points. Following Van Gompel and Majid (2004), three regions of analysis were delimited: the antecedent region, the anaphor region and the post-anaphor region. The antecedent region was the first prepositional phrase in the conjoined object position (e.g. *a un ministro* 'to a minister'). The anaphor region was the prepositional phrase immediately after the verb in the second sentence (e.g. *contra el ministro/ contra él* 'against the minister/against him'). The post-anaphor region consisted of the three-word prepositional phrase after the anaphor region. We report three eye-tracking measures: first-fixation

duration, fixation duration and total fixation duration. *First-fixation duration* is the duration of the reader's first fixation in a region. *Fixation duration* corresponds to the sum of all fixations on a critical region before the reader leaves it for the first time (only first-pass reading is taken into account; no regressions). *Total fixation duration* is the sum of all fixation durations in a region (possible regressions included). During the recordings, Tobii Studio 3.2 software's IV-T *Fixation Filter* (Olsen, 2012) was set on: adjacent fixations were automatically merged provided the fixation was shorter than 75 ms (see Salojärvi *et al.*, 2005) and the maximum angle between both fixations was 0.5 degrees - to filter out micro-saccades (Komogortsev *et al.*, 2010), which usually have an amplitude of 0.5 or less (Yarbus, 1967). Fixations shorter than 120 ms or longer than 890 ms were deleted (for similar approaches, see Drieghe *et al.*, 2010; Folk and Morris, 2003; Johnson *et al.*, 2007; Lowder *et al.*, 2013; Rayner *et al.*, 2010; Van Gompel and Majid, 2004). The remaining data points that exceeded a threshold of 2.5 standard deviations from the mean by region and condition were excluded from further analysis (Ratcliff, 1993). This process affected 1.6% of the data in first-fixation duration, 2.99% in fixation duration and 1.84% in total fixation duration. Two ANOVAs were conducted for each eye-tracking measure and region, one with participants (F1) and one with items (F2) as the random factors, with Frequency (High Frequency / Low Frequency) and Anaphor type (Noun Phrase / Pronoun) as within-participants and within-items factors. Only frequency effects and its interaction with anaphor type will be reported. In order to directly compare frequency effects, paired T-test analyses were carried out whenever appropriate.

Results

Antecedent region. Low-frequency antecedents took longer to read than high-frequency antecedents in all measures but total fixation duration (first fixation: $F_1 (1, 19) = 8.11$, $p = .01$; $F_2 (1, 39) = 7.01$; $p = .01$; fixation duration: $F_1 (1, 19) = 6.03$; $p = .02$; $F_2 (1, 39) = 3.99$; $p = .05$; total fixation duration: $F_1 (1, 19) = .27$; $p = .61$; $F_2 (1, 39) = .49$; $p = .48$).

	Antecedent	Anaphor	Post-anaphor
First-Fixation times			
Repeated NP-HF	237 (9)	269 (16)	248 (13)
Repeated NP-LF	275 (15)	303 (20)	255 (10)
Pronoun-HF	247 (13)	252 (14)	284 (17)
Pronoun-LF	279 (14)	242 (11)	280 (17)
Fixation times			
Repeated NP-HF	241 (9)	255 (14)	248 (13)
Repeated NP-LF	269 (14)	286 (17)	270 (10)
Pronoun-HF	245 (9)	254 (16)	268 (16)
Pronoun-LF	274 (13)	241 (11)	276 (12)
Total Fixation times			
Repeated NP-HF	507 (51)	449 (37)	589 (48)
Repeated NP-LF	559 (50)	513 (44)	620 (51)
Pronoun-HF	550 (56)	384 (34)	817 (65)
Pronoun-LF	529 (39)	334 (27)	725 (49)

Anaphor region. In first-fixation duration, no frequency effect (F 1 (1, 19) = .7; p = .41; F 2 (1, 39) = .01; p =.91) or interaction with anaphor type (F 1 (1, 19) = 2.53; p = .13; F 2 (1, 39) = 3.89; p = .56) was found. However, in fixation duration, there was an interaction between both factors (F 1 (1, 19) = 4.94; p =.04; F 2 (1, 39) = 11.65; p = .002). Paired T-tests showed that repeated low-frequency nouns took longer to read than repeated high-frequency nouns ((t 1 (19) = –2.31; p = .03; t 2 (39) = –2.89; p = .006), but pronoun conditions did not elicit any significant difference ((t 1 (19) = 1; p =.32; t 2 (39) = 1.54; p = .13) In total fixation duration, the interaction between frequency and type of anaphor turned out to be statistically significant (F 1 (1, 19) = 7.33; p = .01; F 2 (1, 39) = 6.58; p = .01). Pronouns referring to low-frequency antecedents were read faster than those referring to high-frequency ones in the analysis by subjects ((t 1 (19) = 2.47; p = .02; 334.49 ms. vs. 384.80 ms.); but not by items (t 2 (39) = 1.34; p = .18; 343.44 ms. vs. 394.79 ms.) Regarding the control condition, the opposite pattern was observed: repeated high-frequency nouns were read faster than repeated low-frequency nouns in the analysis by items (t 2 (39) = –2.31; p = .02; 449 ms vs. 522.48 ms.), but not by subjects (Repeated HF NP vs. Repeated LF NP: (t 1 (19) = –1.67; p = .11; 449.39 ms. vs. 513.27 ms.)

Post-anaphor region. No frequency effects or interaction with anaphor type occurred in this region in any measure (first-fixation duration: (F 1 (1, 19) = 0; p = .93; F 2 (1, 39) = .09; p = .76); fixation duration: (F 1 (1, 19) = .31; p = .58; F 2 (1, 39) = .77; p = .38) and total fixation duration: (F 1 (1, 19) = 2.24; p = .15; F 2 (1, 39) = 2.94; p = .09).

Discussion

The aim of this eye-tracking-while-reading study was to determine how the lexical frequency of the antecedent affects anaphoric pronoun resolution. Similarly to the eye-tracking results obtained by Van Gompel and Majid (2004) and Lago (2014), we found a lexical frequency effect at the antecedent region in first-fixation and fixation duration; however, we do not observe any in total fixation duration. This difference suggests that high and low-frequency antecedents in our materials elicited a similar amount of regressions.

In the anaphor region, there was an interaction between frequency and anaphor type in fixation duration and total fixation duration, such as repeated noun phrases referring to high-frequency antecedents were read faster than those linked to low-frequency nouns. Nevertheless, this frequency effect was fully significant in fixation duration only.

Results elicited in total fixation duration should be interpreted with caution since they are partially significant: pronouns linked to low-frequency antecedents were read faster than the same pronouns referring to high-frequency nouns, but this saliency effect was only significant in the analysis by subjects[2]. The fact that there was no effect in fixation duration for pronouns compared to repeated noun phrases suggests that the saliency effect in total fixation duration was caused by readers looking back to the anaphor region. Similarly to Van Gompel and Majid (2004) and Lago (2014), the pronoun was a short word at the edge of the region, so we expected participants' eyes to move onto the next without having fully processed it. However, there were no spill-over effects of antecedent lexical frequency at the post-anaphor region (cf. Van Gompel and Majid, 2004; Lago, 2014), which led us to think that the saliency effect detected in the previous region could be a Type I error.

---

[2] Despite the fact that we ran a norming study to evaluate the grammatical acceptability of our materials, a plausibility effect might be responsible for the lack of significance in the analysis by items. It is likely that high-frequency words fitted better in the context than low-frequency words.

Conclusion

The aim of this study was to investigate the nature of the antecedent representations involved during anaphoric pronoun resolution in Spanish after Meyer and Bock (1999) suggested that pronoun processing could, in principle, differ across languages based on the kind of information needed to mediate coreference. Whereas we replicated the word-frequency effect in the case of repeated nouns in the control condition, no reliable effect was found for pronouns: infrequent antecedents did not cause any processing cost during pronoun comprehension. We interpret this finding as evidence in favor of the lemma-reaccess account by Simner and Smyth (1998, 1999), which postulates that anaphoric pronoun resolution involves lemma retrieval (i.e. retrieval of the syntactic and semantic properties of the antecedent) whereas lexical (re)access to nouns additionally requires orthographic or phonological information. Simner and Smyth (1999) based their proposal on the serial, two-stage, activation-based speech production model proposed by Jescheniak and Levelt (1994), which locates the word-frequency effect at the lexeme level. The fact that this effect was elicited at the antecedent region as well as with repeated nouns in the control condition suggests thus that the lexeme must have been retrieved during sentence comprehension. Therefore, we argue that the absence of antecedent frequency effects during anaphoric pronoun resolution is a reliable null effect indeed rather than evidence for a difference between the production and comprehension systems. The current research extends Simner and Smyth and Lago's conclusion in English to Spanish. Likewise, it contradicts the results reported in English and German by Van Gompel and Majid (2004) and Heine *et al.* (2006a,b), respectively, in support of the saliency account: i.e. the idea that since infrequent words require more attention and effort during language processing, they are better encoded in memory and, therefore, reaccessed faster than frequent antecedents during anaphoric pronoun resolution. Van Gompel and Majid (2004) suggests that their results can be accommodated within the lemma re-access account by positing two processing stages: one where the memory access mechanism finds the antecedent and retrieves the infrequent word faster, and a second stage where the processor re-accesses the lemma information of the antecedent only. The question which follows then is if discourse information would be enough to target and retrieve the antecedent during stage one of pronoun resolution when the grammatical properties of the antecedent are not yet available and there is not enough context.

Given our results, further research on pronoun processing and lexical access in general needs to be carried out in order to provide data from a larger pool of participants and typologically different languages before drawing any strong conclusions on the nature of the antecedent representations retrieved across languages.

Summary

This article provides empirical evidence in favor of a lexical access model where nouns and pronouns retrieve qualitatively different antecedent representations from memory, in support of a theory of anaphoric lexical access in comprehension (Simner and Smyth, 1998, 1999). We show that coreference resolution in Spanish only involves lemma retrieval —i.e. retrieval of the syntactic and semantic properties of the antecedent—, whereas lexical access to nouns and repeated nouns additionally targets orthographic-phonological information.

References

Almeida, J., Knobel, M., Finkbeiner, M., and Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review*, 14(6):1177-1182.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4):1036-1060.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., and Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eye-tracking. *Cognition*, 76(1):B13-B26.

Bonin, P. and Fayol, M. (2002). Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology*, 14(3):289-313.

Brown, A. S. (2012). *The tip of the tongue state*. Taylor & Francis.

Cacciari, C., Carreiras, M., and Cionini, C. B. (1997). When words have two genders: Anaphor resolution for italian functionally ambiguous words. *Journal of Memory and Language*, 37(4):517-532.

Cloitrew, M. and Bever, T. G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes*, 3(4):293-322.

Colangelo, A., Holden, J. G., Buchanan, L., and Van Orden, G. C. (2004). Speculation about behavior, brain damage, and self-organization: The other way to herd a cat. *Brain and Language*, 90(1):151-159.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5(4):313-349.

Di Domenico, E. and De Vicenzi, M. (1995). Gender and number in the retrieval of pronoun antecedents: differences in use and representation. *Actes du deuxième colloque Langues et grammaire'*, Paris 8 juin 1995, pages 95-109.

Drieghe, D., Pollatsek, A., Juhasz, B. J., and Rayner, K. (2010). Parafoveal processing during reading is reduced across a morphological boundary. *Cognition*,116(1):136-142.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., and Carreiras, M. (2013). Espal: One-stop shopping for spanish word properties. *Behavior Research Methods*, 45(4):1246-1258.

Egusquiza, N., Navarrete, E., and Zawiszewski, A. (2016). Antecedent frequency effects on anaphoric pronoun resolution: Evidence from spanish. *Journal of Psycholinguistic Research*, 45(1):71-84.

Finocchiaro, C. and Caramazza, A. (2006). The production of pronominal clitics: Implications for theories of lexical access. *Language and Cognitive Processes*, 21(1-3):141-180.

Folk, J. R. and Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory & Cognition*, 31(1):87-99.

Foraker, S. and McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, 56(3):357-383.

Garnham, A. (2001). Mental *models and the interpretation of anaphora*. Psychology Press.

Garnham, A., Oakhill, J., Ehrlich, M.-F., and Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. J*ournal of Memory and Language*, 34(1):41.

Garnham, A., Traxler, M., Oakhill, J., and Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, 35(4):517-543.

Gianico, J. L. (2010). *Word concreteness and word frequency as moderators of the tip-of-the-tongue effect*. State University of New York at Albany.

Glanzer, M. and Adams, J. K. (1990). The mirror effect in recognition memory: data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1):5.

Hammer, A., Goebel, R., Schwarzbach, J., Münte, T. F., and Jansma, B. M. (2007). When sex meets syntactic gender on a neural basis during pronoun processing. *Brain Research*, 1146:185-198.

Hankamer, J. and Sag, I. (1976). Deep and surface anaphora. *Linguistic inquiry*,7(3):391-428.

Heine, A., Tamm, S., Hofmann, M., Bösel, R. M., and Jacobs, A. M. (2006a). Event-related theta activity reflects memory processes in pronoun resolution. *Neuroreport*, 17(18):1835-1839.

Heine, A., Tamm, S., Hofmann, M., Hutzler, F., and Jacobs, A. M. (2006b). Does the frequency of the antecedent noun affect the resolution of pronominal anaphors?: An ERP study. *Neuroscience Letters*, 400(1):7-12.

Irmen, L. and Knoll, J. (1999). On the use of the grammatical gender of anaphoric pronouns in german. a comparison between finns and germans. *Sprache & Kognition*, 18(3-4):123-135.

Jescheniak, J. D. and Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4):824.

Johnson, R. L., Perea, M., and Rayner, K. (2007). Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1):209.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4):329.

Knobel, M., Finkbeiner, M., and Caramazza, A. (2008). The many places of frequency: Evidence for a novel locus of the lexical frequency effect in word production. *Cognitive Neuropsychology*, 25(2):256-286.

Komogortsev, O. V., Jayarathna, S., Koh, D. H., and Gowda, S. M. (2010). Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms. In *Proceedings of the 2010 Symposium on eye-tracking research & applications*, pages 65-68. ACM.

Lago, M. S. (2014). *Memory and Prediction in Cross-Linguistic Sentence Comprehension*. PhD thesis, University of Maryland-College Park.

Laubstein, A. S. (1999). Lemmas and lexemes: The evidence from blends. *Brain and Language*, 68(1):135-143.

Laubstein, A. S. (2002). Experimental evidence for serial models of lexical access: A judgment task. *Brain and Language*, 81(1-3):424-431.

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1):1-38.

Lowder, M. W., Choi, W., and Gordon, P. C. (2013). Word recognition during reading: The interaction between lexical repetition and frequency. *Memory & Cognition*, 41(5):738-751.

Lucas, M. M., Tanenhaus, M. K., and Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory & Cognition*, 18(6):611-631.

Malmberg, K. J. and Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, 31(1):35-43.

Meyer, A. S. and Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from dutch. J*ournal of Memory and Language*, 41(2):281-301.

Navarrete, E., Basagni, B., Alario, F.-X., and Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology*, 59(10):1681-1690.

Nicol, J. L. and Swinney, D. (2003). The psycholinguistics of anaphora. *Anaphora: A reference guide*, pages 72-104.

O'Brien, E. J. and Myers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(1):12.

Oldfield, R. C. and Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4):273-281.

Olsen, A. (2012). The tobii i-vt fixation filter. *Tobii Technology*.

Pynte, J. and Colonna, S. (2000). Decoupling syntactic parsing from visual inspection: The case of relative clause attachment in french. *Reading as a Perceptual Process*, pages 529-547.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3):510.

Rayner, K., Castelhano, M. S., and Yang, J. (2010). Preview benefit during eye fixations in reading for older and younger readers. *Psychology and Aging*, 25(3):714.

Rayner, K. and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3):191-201.

Rayner, K., Raney, G. E., and Pollatsek, A. (1995). *Eye movements and discourse processing*. Lawrence Erlbaum Associates, Inc.

Rayner, K., Slattery, T. J., Drieghe, D., and Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2):514.

Rigalleau, F., Caplan, D., and Baudiffier, V. (2004). New arguments in favour of an automatic gender pronominal process. *The Quarterly Journal of Experimental Psychology Section A*, 57(5):893-933.

Sag, I. A. and Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and Philosophy*, 7(3):325-345.

Salojärvi, J., Puolamäki, K., Simola, J., Kovanen, L., Kojo, I., and Kaski, S. (2005). Inferring relevance from eye movements: Feature extraction. In *Proceedings of the NIPS 2005 Workshop on Machine Learning for Implicit Feedback and User Modeling*. Publications in Computer and Information Science.

Schilling, H. E., Rayner, K., and Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, 26(6):1270-1281.

Shiffrin, R. M. and Steyvers, M. (1997). A model for recognition memory: Rem—retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4(2):145-166.

Simner, J. and Smyth, R. (1998). Anaphoric vs. stimulus-based lexical access. In *CUNY Conference on Human Sentence Processing*, Rutgers University.

Simner, J. and Smyth, R. (1999). Phonological activation in anaphoric lexical access (ala). *Brain and Language*, 68(1):40-45.

Van Gompel, R. P. and Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, 90(3):255-264.

Vitevitch, M. S. and Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. Memory & *Cognition*, 31(4):491-504.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision*, pages 171-211. Springer.