

Prediction during sentence comprehension

Any leftovers from a discarded prediction? Evidence from eye-movements during sentence comprehension

Arella E. Gussow*,
Basque Center on Cognition, Brain and Language (BCBL)
Donostia-San Sebastián, Spain

Efthymia C. Kapnoula,
Basque Center on Cognition, Brain and Language (BCBL)
Donostia-San Sebastián, Spain

and

Nicola Molinaro
Basque Center on Cognition, Brain and Language (BCBL)
Donostia-San Sebastián, Spain
Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Running Head: PREDICTION DURING SENTENCE COMPREHENSION

*Corresponding Author:

Arella E. Gussow

Currently at Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson St., Madison, WI 53706, USA

Email: gussow@wisc.edu

Abstract

We investigated how listeners use gender-marked adjectives to adjust lexical predictions during sentence comprehension. Participants listened to sentence fragments in Spanish (e.g., “*The witch flew to the village on her...*”) that created expectation for a specific noun (*broomstick_{fem}*), and were completed by an adjective and a noun. The adjective either agreed (*new_{fem}*), disagreed (*new_{masc}*), or was neutral (*big_{fem/masc}*) with respect to the expected noun’s gender. Using the visual-world paradigm, we monitored looks toward images of the expected noun versus an alternative of the opposite gender (*helicopter_{masc}*). While listening to the initial fragment, participants looked more towards the expected noun. Once the adjective was heard, looks shifted toward the noun that matched the adjective’s gender. Finally, upon hearing the noun, looks were affected by both previous context and adjective gender. We conclude that predictions are updated online based on gender cues, but sentence context still affects integration of the expected noun.

Keywords: Spoken sentence processing, prediction, visual-world paradigm, anticipatory eye-movements

1. Introduction

The predictive view of language processing (e.g., DeLong, Urbach, & Kutas, 2005; Farmer, Brown, & Tanenhaus, 2013; Federmeier, 2007; Kuperberg & Jaeger, 2016; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2004) posits that anticipation of future words or structures in incoming linguistic input is a prominent mechanism in sentence comprehension. Under this account, the language processor is not limited to integrating incoming sensory information (e.g., auditory, visual) to construct higher-level meaning in a bottom-up manner; but instead uses higher-level representations to form hypotheses about what is most likely to appear next. For example, upon hearing “*The witch flew to the village on her...*”, people are likely to expect the word “*broomstick*” as a completion. These hypotheses in turn pre-activate predicted items at the lower sensory levels – a top-down mechanism, preparing the system for the input to come (DeLong et al., 2005; Dikker & Pylkkanen, 2011; Dikker & Pylkkänen, 2013). The interplay between directions of information flow is assumed to make comprehension more efficient, as shown in facilitation of performance when stimuli are supported by context (Hess, Foss, & Carroll, 1995; Jordan & Thomas, 2002; James L. McClelland & O’Regan, 1981).

In the present study, we examined how comprehenders use spoken sentence context in combination with gender-marked adjectives to create and adjust, respectively, lexical predictions in real time. Specifically, we asked whether participants’ lexical expectations were affected by prenominal adjectives marked by a grammatical gender that either supported, contradicted, or was neutral with respect to a context-predicted noun. We also asked whether these predictive

cues would affect target integration¹ once it is encountered. Finding such evidence would support an account of prediction whereby the language processor pre-activates expected words and their features, and further adjusts these predictions in real time to facilitate online comprehension.

Investigating prediction using the visual world paradigm

Predictive language comprehension relies on the incremental nature of the unfolding signal, as the limited information already processed is used not only to integrate what has been revealed, but also to predict future material in a continuous manner (Kamide, Altmann, & Haywood, 2003a) – updating predictions with the receipt of each new piece of information, and pre-activating features of expected items before they are encountered (DeLong et al., 2005; Federmeier & Kutas, 1999). One method that allows monitoring incremental processing is the visual-world paradigm (originally by Cooper, 1974). In this paradigm, participants' eye-movements are tracked as they view a display of images and listen to spoken language stimuli referring to objects in the display. Eye-movements towards the different objects can be time-locked to critical points in the auditory stimuli, enabling researchers to follow how the linguistic input affects participants' visual mapping in real time (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Importantly, it is possible to track looks towards displayed objects even before they are mentioned in a spoken sentence. These anticipatory looks as a function of linguistic input can provide an online window into the factors leading to prediction of a certain referent (for a review of the paradigm and its experimental uses, see Salverda & Tanenhaus, 2017; Tanenhaus & Trueswell, 2006)

¹Although we refer to processing during target presentation as “integration”, it should be noted that we cannot draw a clear distinction between recognition and integration of the word, as these processes highly overlap (Molinaro, Conrad, Barber, & Carreiras, 2010).

Altmann and Kamide (Altmann & Kamide, 1999) used this technique to investigate prediction as a function of the semantic constraints a verb poses on its objects. Participants were presented with a visual scene made up of images such as a boy, a cake, a ball, a toy train, and a toy car; while an auditory sentence unfolded in one of two versions: “*the boy will move the cake*” or “*the boy will eat the cake*”. Results showed that in the first condition (“*move*”), looks towards the cake image increased when the word “*cake*” was heard, as expected from the observation that linguistic stimuli can direct eye-movements to corresponding referents (Tanenhaus et al., 1995). However, in the second condition (“*eat*”), eye-movements to the cake (the only edible object in the display) began earlier— at the verb “*eat*”. This was taken as evidence of the verb’s semantic properties restricting options for upcoming references, hence supporting a predictive mechanism in the mapping of linguistic stimuli onto a visual display.

The visual-world paradigm has further been used to examine prediction based on various properties such as tense (Altmann & Kamide, 2007), case marking (Kamide, Altmann, & Haywood, 2003b; Kamide, Scheepers, & Altmann, 2003), transitivity (Arai & Keller, 2013), argument status (Boland, 2005), prepositions (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002), discourse context (Kaiser & Trueswell, 2004), and prosody (Weber, Grice, & Crocker, 2006). Moreover, evidence for integration of information from different domains such as syntax, semantics, and world knowledge (Kamide, Altmann, et al., 2003a; Kamide, Scheepers, et al., 2003) indicates that cues from different levels of representation can interact with each other to guide eye-movements toward the most supported prediction.

Prediction and Grammatical Gender

Grammatical gender information has also been shown to affect linguistic prediction. For example, Lew-Williams and Fernald (2007) report that native Spanish speakers make use of gender-marked articles to predict subsequent noun referents. Participants viewed pictures of two objects whose names in Spanish are either of the same grammatical gender (e.g., *pelota* [ball_{fem}], *galleta* [cookie_{fem}]) or different (e.g., *pelota* [ball_{fem}], *zapato*, [shoe_{masc}]); while listening to a sentence that referred to one of the objects (e.g., “*Encuentra la pelota*” [Find the_{fem} ball_{fem}]). Results showed that participants were faster to look at the target object in different-gender trials, where a gender-marked article in the sentence was informative about which noun would follow, suggesting that grammatical gender information helps predict upcoming nouns (for similar findings in French, see Dahan, Swingley, Tanenhaus, & Magnuson, 2000).

Researchers have also used grammatical gender properties to examine lexical prediction in the absence of visual referents, employing electrophysiological (EEG) methods (Molinaro, Giannelli, Caffarra, & Martin, 2017; Van Berkum et al., 2005; Wicha, Bates, et al., 2003; Wicha et al., 2004). For example, Van Berkum et al. (2005) capitalised on this feature in Dutch, in which adjectives carry a gender-marked suffix that matches the gender of the noun being modified. Participants listened to discourse contexts that were constraining towards a specific noun, e.g. “*The burglar had no trouble locating the secret family safe. Of course, it was situated behind a...*”. The contexts were followed by a prenominal adjective in one of two versions: 1. the adjective agreed in gender with the expected noun (of either neuter- or common-gender) and was followed by it, “*big_{neuter} but unobtrusive painting_{neuter}*”; or 2. it disagreed and was followed by an alternative compatible noun, “*big_{common} but unobtrusive bookcase_{common}*”. A differential ERP (Event-Related Potential) effect was found at the adjective when it mismatched the expected noun in gender as compared to one that agreed with the expected noun. This difference

disappeared when sentences were not supported by a predictive discourse, suggesting that comprehenders were able to use context information to predict specific words and their grammatical gender.

Results from Van Berkum et al. (2005) therefore support the notion that grammatical gender is included in the prediction of a noun, and is pre-activated as one of its features (Molinaro et al., 2017; Wicha, Bates, et al., 2003; Wicha et al., 2004; Wicha, Moreno, & Kutas, 2003). Crucially, the effect was found before any noun was encountered (a point at which either gender marker was compatible without causing overt agreement violation). This suggests that during spoken language comprehension, listeners compute syntactic agreement relations taking into account both received input and predicted information.

Adjusting Predictions

The common manipulation in ERP prediction studies, comparing brain responses to expected versus unexpected continuations (DeLong et al., 2005; Federmeier & Kutas, 1999; Federmeier, McLennan, Ochoa, & Kutas, 2002; Wicha, Moreno, et al., 2003), points to an important property of prediction; it is not always fulfilled as anticipated. There is no restricted set of messages to select from during conversation, but on the contrary – we can spontaneously create new, original sentences never heard before. Considering this open-ended nature of human language, it is reasonable that at times the system will be proved wrong.

Prediction errors, however, are not necessarily just a “surprise” for the system, but may be further used to adjust predictions (Clark, 2013; Friston & Stephan, 2007; Winkler & Czigler, 1998; Winkler, Karmos, & Näätänen, 1996). For example, a grammatical gender marker that is incompatible with a context-predicted word can signal that a change in pre-activation is needed –

so that a gender-matching option gains preference over the previous prediction. In fact, Van Berkum et al. (2005) suggest that the differential ERPs they found may be interpreted beyond processing consequences of covert agreement violations. Instead, the effect may be reflecting an *adjustment* of the prediction to keep in line with the incoming stimuli. Clearly, a system that can keep up with the input will be more advantageous for online comprehension.

Open Questions and Further Issues

Despite the large number of prediction studies in the literature, there are still a number of issues that remain unresolved. First, while prediction effects have been widely reported on Spanish articles (Molinaro et al., 2017; Wicha, Bates, et al., 2003; Wicha et al., 2004; Wicha, Moreno, et al., 2003; and see DeLong et al., 2005, for similar findings in English), the strong association between nouns and articles limits conclusions from these findings. This is because Spanish nouns are rarely used in isolation, but instead appear together with their articles even when there is no apparent need for a determiner (e.g., *'me gusta el arte'* [I like the art], meaning 'I like art'). As a result, it is possible that each article-noun pair is stored as a single unit in the mental lexicon, such that pre-activation of a noun entails pre-activation of its matching article form, without tapping into grammatical gender. Adjectives, on the other hand, vary widely and are completely optional, therefore generally characterised by low co-occurrence patterns with the nouns they modify. Creating constraining sentences in which gender is marked on a prenominal adjective would, therefore, not only extend research on predictive gender cues in Spanish to constituents other than articles, but also minimise the chance of tight article-noun co-occurrence patterns driving observed effects.

Second, certain Spanish adjectives are of identical form regardless of which gender they assume for noun agreement (e.g., ‘*gran escoba*’ [big_{fem} broomstick_{fem}]; ‘*gran helicóptero*’ [big_{masc} helicopter_{masc}]). This particular feature of Spanish has been neglected in the study of language prediction. However, these adjectives, henceforth termed “neutral”, can provide an additional comparison for understanding how information about morphosyntactic gender affects prediction computation; using neutral adjectives we can better assess the effect of a supportive cue (gender-matching adjective) versus that of a contradicting cue (gender-mismatching adjective) on lexical prediction, with an uninformative cue (neutral adjective) serving as the baseline.

Lastly, most studies examine prediction while focusing on a specific point in time. A more dynamic approach, however, would look at how prediction unfolds over time (e.g. the time-course of a sentence), zooming in at different time points of interest. This would allow us to a) closely follow how expectations are adjusted as new information is received, and b) examine within the same paradigm how prediction ultimately affects integration.

All of these issues are addressed in the present study.

The Present Study

In the present study, we used the visual-world eye-tracking paradigm to investigate lexical prediction in spoken sentences. In doing so, we focused on the role of grammatical gender, specifically as marked on adjectives. In Spanish, every noun is assigned a feminine or masculine grammatical gender, stored with it in the mental lexicon (Harris, 1991). Furthermore, articles, pronouns, and adjectives must agree in gender with the nouns they are associated with (e.g., ‘*la nueva escoba*’ [the_{fem} new_{fem} broomstick_{fem}], ‘*el nuevo helicóptero*’ [the_{masc} new_{masc}

helicopter_{masc}]). The grammatical gender system has been shown to facilitate comprehension in various languages (for a review, see Friederici, 1999) and has also been proposed as a syntactic cue that can aid predictive processing (Wicha et al., 2005). A number of ERP studies have taken advantage of this feature to show expectation of the grammatical gender of articles (e.g., Wicha, Bates, et al., 2003; Wicha et al., 2004) or adjective inflections (Van Berkum et al., 2005) based on a context-expected noun. Furthermore, behavioral studies combining semantic context and grammatical gender manipulations have found evidence for an interaction between these sources of information, reflected in reaction times (Bentrovato et al., 2003; Wicha et al., 2005).

The goal of the present study was to examine how lexical predictions are shaped by sentence context and grammatical gender cues. Specifically, we asked whether a prenominal adjective that mismatches a context-expected noun in gender can trigger rapid adjustment of the prediction, in time before the noun is reached. Additionally, we looked at whether the updated prediction overrides the initial expectation (i.e., whether it is still affected by the sentence bias or not). We were also interested in examining the influence of both sentence context and grammatical gender cues on noun integration, as a clue to the role of prediction in spoken language comprehension.

Using the visual-world paradigm, we tracked participants' eye movements towards an array of four images, as they listened to a sentence fragment (e.g., *'La bruja voló a la aldea en su...'* [The witch flew to the village on her...]) that builds up expectation for one of the displayed noun images ("expected"; e.g., *'escoba'* [broomstick_{fem}]). Another noun image was also semantically compatible with the sentence, yet less expected and of the opposite grammatical gender ("alternative"; e.g., *'helicóptero'* [helicopter_{masc}]). The remaining two images were of semantically unrelated nouns, one of which matched the expected noun's gender ("gender filler");

e.g., ‘*silla*’ [chair_{fem}]), and the other was of the opposite gender (“distractor”; e.g., ‘*micrófono*’ [microphone_{masc}]). Critically, an adjective was presented at the end of the sentence context just before the noun, and it either agreed in gender with the expected noun (“congruent”; e.g., ‘*nueva*’ [new_{fem}]), disagreed (“incongruent”; e.g., ‘*nuevo*’, [new_{masc}]), or was neutral (“neutral”; e.g., ‘*gran*’ [big_{fem/masc}]). Finally, the sentence was completed by the noun that was always semantically acceptable and compatible with the adjective’s gender – that is, either the initially expected or the alternative noun.

We hypothesised that as the sentence unfolded and became more restrictive regarding possible completions, the proportion of looks towards the image of the context-expected noun (e.g., broomstick_{fem}) would rise – reflecting the incremental formation of a prediction. Assuming morphosyntactic gender is considered throughout this process, when an adjective of the expected gender was presented (e.g., new_{fem}) we expected to see a strengthening of this bias, reflected as more looks towards the expected noun (e.g., broomstick_{fem}). In the incongruent condition, upon encountering an adjective marked by a gender that disagrees with the expected noun (e.g., new_{masc}), we expected to see a shift of fixations towards the less expected, but gender matching, alternative image (e.g., helicopter_{masc}). This would suggest that grammatical gender is anticipated as part of noun expectation, and immediately takes part in guiding online prediction. As the neutral adjective (e.g., big_{fem/masc}) carries no gender indication that can strengthen or weaken the prediction, in this condition we expected to see a simple continuation of the trend of more looks towards the expected noun, following the semantic context.

Finally, we were also interested in effects of previous cues on target integration, that is, effects seen post noun onset. We hypothesised that noun integration would be affected by both context and grammatical gender. We expected that whichever target noun was heard in the

stimuli would receive an overwhelming proportion of the looks; however, target activation should still be affected by participants' predictions based on previous information. First, the effect of the constraining context should be apparent such that when the target is the expected noun (e.g., broomstick_{fem}), it should receive more looks than when the target is the alternative noun (e.g., helicopter_{masc}). Moreover, we expected a larger advantage in looks to the target when preceded by a gender-marked adjective compared to a neutral adjective, due to the unequivocal grammatical cue directing looks to the gender-matching noun. Overall, a larger advantage for a target supported by more cues (i.e., context and grammatical gender) would demonstrate the effective combination of predictive cues and their continuing role in noun integration.

2. Methods

Participants

Thirty-six participants were recruited from the participant pool of the Basque Center on Cognition, Brain, and Language, San-Sebastián, Spain. Data from three participants were excluded from analyses; two of them due to inappropriate language profile, and one due to eye-tracking problems. The remaining thirty-three participants (27 females, age range 19-40, mean: 24, SD: 4.3) were native Spanish speakers, had normal or corrected-to-normal vision and no reported hearing deficits. All provided written informed consent and were compensated with 12 EUR.

Design and Materials

Seventy-eight experimental sentence contexts were constructed, each designed to create expectation for a specific noun (henceforth “expected”) that was preceded by a critical adjective.

Although adjectives in Spanish typically occur post-nominally, we designed the sentences and chose the adjectives in a way that allows for natural pre-nominal placement. The grammatical gender of the adjectives was manipulated to create three conditions:

1. **Congruent:** the adjective agreed with the expected noun and was followed by it.

La bruja voló a la aldea en su nueva escoba.

[The witch flew to the village on her new_{fem} broomstick_{fem}]

2. **Incongruent:** the adjective disagreed with the expected noun and was followed by an alternative noun (henceforth “alternative”) that was less expected but semantically and grammatically compatible with the sentence.

La bruja voló a la aldea en su nuevo helicóptero.

[The witch flew to the village on her new_{masc} helicopter_{masc}]

3. **Neutral:** an adjective that can assume either gender (henceforth “neutral”) without changing form; followed by either the expected or the alternative noun.

La bruja voló a la aldea en su gran escoba/helicóptero.

[The witch flew to the village on her big_{fem/masc} broomstick_{fem} / helicopter_{masc}]

Three versions of each context were therefore created, resulting in a total of 234 experimental sentences, displayed in Appendix. These were divided into three lists such that each participant was presented with every sentence context, but only in one of its conditions (congruent, incongruent, or neutral). Within each list, an equal number of sentences occurred in each condition. For each sentence, the same adjective was used for the congruent and incongruent conditions, only varying the inflectional marker in the last position (-o for masculine, -a for feminine). The neutral condition required using a different adjective, as this

class is unique and does not have an identifying marker that could be applied as in the gender-marked adjectives. Half of the sentences in the neutral condition were followed by the expected noun, and the other half were followed by the alternative noun. Out of all expected nouns 46% were feminine and 54% were masculine.

Cloze probabilities were assessed by native Spanish speakers (N=20) who did not take part in the experiment. These participants were instructed to read all sentence contexts (truncated before the adjective) and complete each of them with the first word that came to mind. The mean cloze probability for expected nouns was .64 (SD=.28). Synonyms of the expected nouns accounted for .27 of the remaining non-target responses. The mean cloze probability for alternative nouns was .01 (SD=.04); for unrelated items it was 0.

Mean log frequency of the expected and alternative nouns (extracted from the EsPal database: Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013) did not differ significantly (expected: $M = 1.14$, $SD = 0.58$, alternative: $M = 1.05$, $SD = 0.52$; $t(132.34) = 0.96$, $p = .34$). There was no difference in frequency between the gender-marked ($M = 2.09$, $SD = .05$) and neutral ($M = 2.16$, $SD = .72$) adjectives ($t(122.28) = -0.63$, $p = .52$). Mean sentence duration was 7,542 ms ($SD = 1,327$), mean adjective duration² was 643 ms ($SD = 141$), mean duration of the time-window between adjective offset and noun onset was 280 ms ($SD = 98$), and mean noun duration³ was 766 ms ($SD = 203$).

A number of experimental items were excluded from analyses. Four items could not be used due to technical difficulties in pre-processing; and one item due to a coding error. We also excluded four items that were found to have inappropriate cloze-probabilities (both the expected

² Adjective duration did not differ between gender-marked and neutral items ($M = 637 \pm 87$ and $M = 653 \pm 207$ respectively; $t(90.01) = -.59$, $p = .55$).

³ Noun duration did not differ between expected and alternative items ($M = 755 \pm 202$ and $M = 778 \pm 205$ respectively; $t(133.98) = -.65$, $p = .52$).

and alternative items had similarly low clozes). One more item was excluded after finding that the wrong image was clicked by over 17% of the participants. Error rate for other items was otherwise close to zero, so we suspect that naming the expected and alternative images for this item was ambiguous due to visual similarity. All descriptive information reported here (frequencies, durations, cloze) pertains to the 68 items remaining in our analyses.

Ninety filler sentences without an adjective preceding the noun were added to the stimuli lists to diversify the sentence structure and make the manipulation less salient. Seventy-five of the fillers followed the same pattern as the experimental stimuli by providing a constraining context, with half of them ending in the expected noun and half in an alternative noun. The remaining fifteen fillers were not constraining, serving to further conceal the experimental manipulation.

In sum, over the course of the experiment (across experimental and filler trials), the target noun was always semantically and grammatically compatible with the sentence, and matched the sentence-based expectation half of the time. The same set of trials was presented twice to examine how prediction biases may be affected by previous exposure, however we here report data only from the first presentation (see Results).

Visual Displays

Eighty-four displays were created, each comprised of four images depicting the expected noun (e.g., *broomstick*_{fem}), the alternative noun (e.g., *helicopter*_{masc}), and two unrelated nouns of opposite genders (“gender filler”, e.g., *chair*_{fem}; “distractor”, e.g., *microphone*_{masc}). The two unrelated items (e.g., *chair*_{fem} and *microphone*_{masc} served as expected and alternative nouns of a different sentence context (i.e., in that case, *broomstick*_{fem} and *helicopter*_{masc} acted as the unrelated items). That is, each display was used twice for each participant; once with one noun

pair being the expected and alternative items, and once with the other noun pair being the expected and alternative items. Figure 1 presents an example of a visual display. Location of images varied across trials.

(Figure 1 about here)

Images

For our visual stimuli we used 329 images taken from the Spanish version of the MultiPic database (Duñabeitia et al., 2017), five images from the MACLab database (Kapnoula & McMurray, 2016; McMurray, Samelson, Lee, & Tomblin, 2010), and two from an electronic database of object and action pictures (Akinina et al., 2015). All images were sized 240 x 240 pixels; Microsoft Paint was used for light editing when needed.

Recordings

Stimuli were recorded by a female native Spanish speaker in a sound-proof booth, sampling at 44,100 Hz. Sentence recordings were spliced such that the same base context was combined with the different condition-dependent adjectives and nouns, ensuring identical stimuli across all three conditions until the critical point at which conditions diverged. Nouns and adjectives were taken from two separate sentence recordings; nouns were spliced in isolation to control for coarticulation, adjectives were spliced together with the word that preceded them to preserve naturalness. Splicing was always at the nearest zero-crossing, and 50 ms of silence were added at the beginning and end of each complete sentence. Editing and noise removal were performed using Praat and Audacity, and critical time points (adjective onset, adjective offset, noun onset) were marked for later analyses.

Procedure

Participants were seated in front of a 19'' monitor with a resolution of 1024×768 pixels for visual stimuli presentation; auditory recordings were transmitted through high quality headphones. Monocular eye movements were recorded at a sampling rate of 1000 Hz using the SR Research Desktop Mount EyeLink 1000 plus eye-tracker. Calibration and validation were carried out before beginning the experiment using a standard nine-point procedure.

After calibration, written instructions for the task were presented on the screen until participants pressed a key to begin the experiment. Each trial began with the presentation of a display of four images with a red circle in the center. After 500 ms the red circle turned green, as a cue for participants to click on it. This allowed a brief view of the images before the onset of the auditory stimulus, thereby minimizing eye-movement due to visual search. Once clicked, the circle and the cursor disappeared from the screen and the spoken sentence recording began. When the recording ended, the cursor reappeared on the screen and participants were required to click on the image that depicted the final noun in the auditory sentence, thereby terminating the trial and proceeding to the next one. Order of trials was randomised for each participant. Every 12 trials participants could take a break, and once ready to proceed they performed a drift correct. After completing all trials once, a repetition block began (we report data only from the first presentation, see Results). Procedures were implemented in SR research Experiment Builder. The whole session lasted approximately one hour and twenty minutes.

Data Analysis

The real-time record of gaze in screen coordinates was automatically parsed into saccades and fixations using the default psychophysical parameters, and adjacent saccades and fixations were combined into a single “look”. Boundaries of the ports containing visual stimuli were extended by 100 pixels to account for noise and head-drift (see McMurray, Tanenhaus, & Aslin,

2002). Eye movements were recorded from the onset of the trial until the participant's response. This resulted in a variable trial offset time, depending on the individual response time. Following prior studies (Allopenna, Magnuson, & Tanenhaus, 1998; McMurray et al., 2002) we set a fixed trial duration of 10,000 ms. If a trial ended before this point, we extended the last eye movement; trials longer than 10,000 ms were truncated. This approach assumes that any fixations made in the very late portions of a trial reflect the word the participant settled on and can, thus, be interpreted as an estimate of the final state of the system. Given the time needed for initiating eye-movement, every time-point was shifted forward by 200 ms to account for this delay (Altmann & Kamide, 2004; Matin, Shao, & Boff, 1993). All analyses and figures in the current paper already include the delay time.

For statistical analyses, we focused on four time-windows of interest as illustrated in Figure 2. The first time-window (*Context End*) captured the last 400 ms of the preceding sentence context that was equal across conditions, chosen to reflect the anticipation of a certain image type based on the constraining sentence. The second time-window (*Adjective*) covered adjective duration (mean length: 643 ms, SD=141), the point at which stimuli diverge for each of the three experimental conditions. The third time-window (*Adjective offset - Noun onset*) was limited between adjective offset and noun onset (mean length: 280 ms, SD=98), the critical locus for successful adjustment before reaching the predicted item. The fourth time-window (*Noun*) covered noun duration (mean length: 766 ms, SD=203), where the target word was finally presented.

(Figure 2 about here)

3. Results

Although the same stimuli were presented twice for each participant, visual inspection of results from the second presentation revealed that participants clearly remembered the experimental items, fixating more on the target noun that was yet to be mentioned (either the expected or the alternative) throughout the trial. We therefore report results only from the first presentation that was of interest for testing our hypotheses.

Behavioral Data

Accuracy in picking the correct image was 99.64% (SD=.06%), indicating that participants were indeed attentive to the stimuli and task. Trials that were answered incorrectly were excluded from further analyses. Mean reaction time was 823 ms (SD=808), calculated from noun offset until participants clicked on one of the images.

Eye-tracking Data

The proportion of trials on which participants fixated each image in the display was calculated at 4-ms increments. Figure 3 presents a time-course graph plotting the proportion of trials with fixations on the expected (e.g., ‘*escoba*’ [broomstick_{fem}]) and alternative (e.g., ‘*helicóptero*’ [helicopter_{masc}]) noun images, shown separately for each of the three auditory stimuli conditions (congruent, incongruent, neutral). The graph is time-locked to adjective offset as computed per individual trial.

(Figure 3 about here)

Within each pre-determined time-window (Context End, Adjective, Adjective offset - Noun onset, Noun), fixation proportions were computed on a trial-by-trial basis considering for that specific trial where the window began (e.g., adjective onset) and where it ended (e.g., adjective offset). Table 1 summarises proportions of fixations on each image per condition for

the three time-windows prior to noun onset; Table 2 displays proportions from the fourth time-window of noun duration. We focused our statistical analyses only on looks to the expected and alternative noun images; data from the two unrelated images of opposite genders (“gender filler”, “distractor”) are displayed in Table 1 and Table 2 to assure the low and almost identical proportions of fixations they received, indicating that they were not considered at any point during the trial.

(Table 1 about here)

(Table 2 about here)

To compare proportions of looks towards the expected and alternative noun images without violating assumptions of independence between samples, for each of the first three time-windows prior to noun revelation we calculated the ratio between the proportion of fixations to the expected image and the sum of the proportions of fixations towards the expected and the alternative images together:

$$\text{expected-to-alternative ratio} = \frac{\text{looks to expected}}{\text{looks to expected} + \text{looks to alternative}}$$

A constant of .00000001 was added to each observation (the proportion of looks to an image per subject per item within each time window) before computing ratios, to avoid undefined expressions in cases where both the expected and the alternative did not receive any looks. A ratio greater than .5 indicates that out of all looks towards the expected and the alternative images taken together, the expected one attracted more than half of those looks (see Dahan & Tanenhaus, 2005; Huettig & McQueen, 2007; McQueen & Huettig, 2012). Figure 4 displays the ratios as a function of condition within each of the first three time windows.

Ratios were logit-transformed for statistical analyses, such that equal probabilities of fixating the expected or alternative image (a ratio of .5) corresponded to 0 in the transformed

values. Ratios were then evaluated as a function of condition within each of the first three time-windows. We estimated linear mixed-effects models using the lme4 (Bates, Maechler, & Dai, 2009) package in R (R Core Team, 2016). For each mixed-effects model, the maximal random-effects structure permitting model convergence was chosen (Barr, Levy, Scheepers, & Tily, 2013), including by-subject and by-item random intercepts and random slopes. Model simplification was required for convergence in time-window 1 (Context End) where we removed covariances among by-subject and by-item random effects; and in time-window 2 (Adjective Duration) where we removed covariance among by-subject random effects.

Next, the lmerTest package using the Satterwaite approximation for degrees of freedom was employed to obtain F and p -values for the omnibus test of condition. Pairwise comparisons are reported after adjusting for multiple comparisons using the holm-bonferroni method.

Additionally, supportive analyses tested whether the individual ratios for each condition within each time-window were significantly different from chance. To do so, logit-transformed ratios were evaluated in linear mixed-models that included the fixed intercept, with by-subject and by-item random intercepts. The fixed intercept in these models represents the mean of the logit-transformed ratios for that condition in that time window, and a statistically significant intercept indicates that the logit-transformed ratios are different than 0 – that is, different than chance.

(Figure 4 about here)

Time-window 1: Context End

There was no effect of condition on the expected-to-alternative ratio ($F(2, 49.26) = .2, p > .1$), reflecting that until this point stimuli were equal across conditions. The expected-to-alternative ratio was above chance in the congruent ($B = 2.81, SE = .79, t(34.08) = 3.58, p < .01$), incongruent ($B = 3.07, SE = .75, t(33.98) = 4.07, p < .001$), and neutral ($B = 2.58, SE = .8, t(35.93) = 3.21, p < .01$) conditions, indicating that the expected image was fixated more than the alternative image in all three conditions.

Time-window 2: Adjective duration

There was a significant effect of condition on the expected-to-alternative ratio ($F(2, 97.996) = 7.57, p < .001$). The ratio was significantly higher in the congruent as compared to the incongruent condition ($B = 2.66, SE = .69, z = 3.85, p < .001$). No other comparisons were significant ($ps > .1$).

The expected-to-alternative ratio was significantly above chance in the congruent condition ($B = 3.43, SE = .8, t(38.6) = 4.3, p < .001$), and in the neutral condition ($B = 2.14, SE = .93, t(44.27) = 2.3, p < .05$), indicating that the expected image was fixated significantly more than the alternative image. The ratio was not significantly different than chance in the incongruent condition ($B = .8, SE = .72, t(33.35) = 1.105, p > .1$), indicating similar probabilities of looking to the expected or alternative noun images. Considering that the critical information about an adjective's gender inflection in Spanish is at the coda, arriving just before adjective offset at the end of this time window, these equal probabilities during the adjective in the incongruent condition are unsurprising.

Time-window 3: Adjective offset - Noun onset

Condition had a significant effect on the expected-to-alternative ratio ($F(2, 38.703) = 49.42, p < .001$). Pairwise comparisons showed that the ratio was significantly higher in the congruent condition as compared to both the incongruent ($B = 10.06, SE = 1.01, z = 9.95, p < .001$) and the neutral conditions ($B = 4.21, SE = .88, z = 4.8, p < .001$). Moreover, the ratio was significantly higher in the neutral condition as opposed to the incongruent condition ($B = 5.85, SE = .97, z = 6.05, p < .001$).

The expected-to-alternative ratio was significantly above chance in both the congruent ($B = 6.15, SE = .94, t(40.23) = 6.55, p < .001$), and the neutral ($B = 1.93, SE = .9, t(40.38) = 2.145, p < .05$) conditions, indicating that the expected image was fixated significantly more than the alternative image. In the incongruent condition, the expected-to-alternative ratio was significantly below chance ($B = -3.9, SE = .77, t(40.15) = -5.08, p < .001$) indicating that the alternative image was fixated more than the expected one.

Time-window 4: Noun duration

Finally, we turn to evaluate the effects of prediction on word integration. During the fourth time-window participants heard the final word (either the expected or the alternative noun), and were instructed to click on the corresponding image. Hence for analyzing looks in this time-window we refer to the noun revealed auditorily as the “target” while the other is the “competitor” – such that in the congruent condition the expected noun was the target and the alternative was the competitor; in the incongruent condition the alternative noun was the target and the expected was the competitor; and in the neutral condition both cases occurred at equal

proportions. Similar to the previous analyses, we computed a ratio for the proportion of fixations, however this time based on the target:

$$\text{target-to-competitor ratio} = \frac{\text{looks to target}}{\text{looks to target} + \text{looks to competitor}}$$

The target-to-competitor ratio was examined in light of two cues: *Expectancy* – whether the target was expected based on sentence context (expected noun) or not (alternative noun); and *gender-Markedness* – whether the prenominal adjective was gender-marked (congruent and incongruent conditions), or unmarked (neutral condition). Figure 5 shows that the target-to-competitor ratios were all higher than .5, as expected considering that participants heard the target and were required to click on it (expected – gender-marked: $M=0.89$, $SD=0.07$; expected – gender-unmarked: $M=0.85$, $SD=0.1$; alternative – gender-marked: $M=0.86$, $SD=0.09$; alternative – gender-unmarked: $M=0.78$, $SD=0.08$).

(Figure 5 about here)

We again used logit-transformed ratios to estimate a linear mixed-effects model accounting for both by-subject and by-item variance. The analysis revealed an effect of *Expectancy* ($B = 2.29$, $SE = 1.03$, $t(47.84) = 2.22$, $p < .05$), such that ratios were higher when the target was the expected image as opposed to the alternative image. We also found an effect of *gender-Markedness* ($B = 3.43$, $SE = .74$, $t(34.43) = 4.59$, $p < .001$), such that ratios were higher when the target was preceded by a gender-marked adjective as compared to a neutral adjective. The *Expectancy* \times *gender-Markedness* interaction was not significant ($B = -.44$, $SE = 1.0001$, $t(49.11) = -.44$, $p > .1$). Figures 6 and 7 present timelines of the raw proportion of looks to the target and competitor images, respectively, as a function of *Expectancy* and *gender-Markedness* throughout noun duration.

(Figure 6 about here)

(Figure 7 about here)

An additional analysis was run on fixations post-noun offset, to see whether the expected noun effect lasted even beyond noun presentation. We examined looks to non-target images in trials where the target was the alternative noun, comparing fixations on the expected noun image ($M = .05$, $SD = .04$) versus the unrelated fillers (gender filler, $M = .01$, $SD = .01$; distractor, $M = .01$, $SD = .02$; a composite score of the fillers was used). This time-window began 500 ms post-noun offset and ended either when participants clicked on the image or at 1000 ms post-noun offset, whichever occurred first (note that trials where reaction time was under 500 ms were excluded). Logit-transformed expected-to-filler ratios were significantly above chance, ($B = .58$, $SE = .28$, $t(55.98) = 2.1$, $p < .05$) demonstrating that the initially expected noun continued to pose more of a competition than the fillers, well beyond noun presentation.

4. Discussion

The current study investigated how comprehenders create and update lexical predictions based on sentence context and grammatical gender cues; and how these predictions ultimately affect word integration. To examine this, participants listened to sentence contexts (e.g., *The witch flew to the village on her...*) that created expectation for a certain noun (e.g., *broomstick_{fem}*). Contexts were continued by a modifying adjective that either agreed (e.g., *new_{fem}*), disagreed (e.g., *new_{masc}*), or was neutral (e.g., *big_{fem/masc}*) with respect to the expected noun's gender. We then tracked participants' eye movements towards an array of images

depicting the expected noun, an alternative noun of the opposite gender (e.g., *helicopter*_{masc}), and two unrelated nouns of the two genders (e.g., *chair*_{fem}, *microphone*_{masc}), as they listened to the spoken sentences and reached the critical adjective point.

Results show more looks towards the expected noun (e.g., *broomstick*_{fem}) compared to the alternative one (e.g., *helicopter*_{masc}) at the end of the sentence context, while the two unrelated noun images were virtually neglected – suggesting that participants indeed used the context to create expectations about the final noun. As we hypothesised, once the gender-marked adjective was revealed, looks towards the expected and alternative noun images changed abruptly according to experimental condition. When the adjective's morphosyntactic gender matched that of the expected noun (congruent condition), the proportion of looks towards the expected noun image rose drastically, while looks towards the alternative one diminished. When the adjective mismatched the expected noun in gender (incongruent condition), the pattern of looks reversed such that the alternative noun – of a gender that agreed with the adjective – was fixated more than the initially expected noun. A neutral adjective continued the prediction bias trend, but did not provide the boost seen with the congruent adjective. As these effects were apparent before any noun was revealed auditorily, the sudden change in fixations cannot be ascribed simply to adapting responses once the sentence completion is revealed, but specifically to adjusting the earlier underlying prediction.

Finally, upon hearing the target noun, participants looked more at the target image—whether of the expected or the alternative noun, in accordance with experimental condition. Interestingly, even at this point the initial context bias continued to affect the pattern of fixations; expected nouns attracted more looks when they were revealed as the target than did alternative nouns in the opposite case. Moreover, adjectives that were gender-marked boosted the matching

target activation compared to neutral adjectives. Taken together, these results suggest that word integration was directly affected by prediction based on both sentence context and grammatical gender information.

Prediction and Grammatical Gender

The fact that a morphosyntactic gender feature had such a clear impact on anticipatory eye-movements strongly supports the idea that grammatical gender information plays a role in the prediction of a noun (Molinaro et al., 2017; Van Berkum et al., 2005; Wicha et al., 2004). We interpret the change in the pattern of looks in the incongruent condition as resulting from a mismatch between the revealed adjective's gender and the expected (yet unrevealed) noun's gender. This covert grammatical violation signals a prediction error, initiating a prediction adjustment to support a noun compatible with the new input. Our use of adjectives instead of the widely studied articles (Wicha, Bates, et al., 2003; Wicha et al., 2004; Wicha, Moreno, et al., 2003) is crucial for interpreting this point. As described in the introduction, articles in Spanish are strongly associated with their nouns due to high co-occurrence patterns in everyday language. Thus, it may be argued that the effects found on articles already reflect integration of the noun phrase, distinct from prediction effects rooted in morphosyntactic properties. Under this account, we would have expected to see a relatively small effect on adjectives that are less associated with the nouns they modify due to lower co-occurrence in their distribution in natural language. However, the robust results we obtained suggest that the effects are not driven solely by an associative or distributional relationship, but indeed reflect the prediction of a word with its features.

To clarify, we are not claiming that prediction of grammatical gender is necessarily inherent in every linguistic prediction. If a lexical representation is defined by the combination of

its properties at the phonological, syntactic, and conceptual levels of representation (Jackendoff, 2002; Van Berkum et al., 2005), pre-activation of a word may focus on a subset of those to a larger extent than others, depending on the relevant constraints in the input (Van Berkum et al., 2005). In our sentences, for example, there were no syntactic cues until reaching the adjective, so that semantic information was most relevant at first. However, once the adjective was heard, the gender of the predicted word became extremely relevant for prediction. Assuming a tight association between features of a lexical unit, and considering the rich contexts in which words usually appear, spreading pre-activation between features of the predicted word is only expected from an efficient processor (Van Berkum et al., 2005). Once a syntactic cue became available, it turned into an inherent part of the prediction and its evaluation.

In addition, finding an effect driven by grammatical gender verifies that anticipatory looks were mediated specifically by the *lexical* representation of images. The feature that yielded a robust difference in the pattern of looks had to be accessed through the mental lexicon where it is stored together with the noun (Harris, 1991), and could not be extracted based on the visual image alone. This further supports the linking between eye-movements and properties of normal linguistic processing (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan & Tanenhaus, 2005; Eberhard et al., 1995; Tanenhaus, Magnuson, Dahan, & Chambers, 2000), on which visual-world language experiments base their inferences.

Congruency

Our congruency manipulation (i.e. using adjectives of congruent, incongruent, or neutral gender with respect to the predicted noun) confirms the prominent role of morphosyntactic gender in predictive processes. Firstly, the congruent adjective that matched the expected noun in

gender caused a boost in looks towards the corresponding image. While this finding can be seen as similar to those of Lew-Williams and Fernald (2007) described earlier, showing the advantage of gender-marked articles in predicting subsequent nouns, it differs in two main aspects (aside from the use of adjectives as opposed to articles). For one, the sentences used in Lew-Williams and Fernald (2007) consisted of simple recurring questions or instructions such as ‘find the ball’, so that morphosyntactic gender was the only predictive cue available. In contrast, our sentences varied in length and carried a much richer semantic content, yet participants still singled out the gender cue and used it efficiently. Secondly, our sentences created a prediction even before the gender marker was reached. This allows us to show the clear contribution of the morphosyntactic gender cue to an already established prediction, which was strengthened by the new information.

In the incongruent condition, participants received morphosyntactic gender information that contradicted their initial expectation, therefore leading to a prediction error. Following this cue, the pattern of looks shifted towards the alternative noun – that was semantically compatible with the sentence, and of the gender that matched the new incoming input. This adjustment following a prediction error is even more interesting considering the fillers on screen, one of which matched the adjective in gender. Despite this filler not being semantically compatible with the sentence context, we may have expected some automatic activation based on gender association alone. However, fillers maintained very low, negligible fixation proportions, firstly assuring that the effect is not merely local gender priming. Moreover, this suggests that the semantic context constrained the options for sentence completion, and only within those options (the two semantically compatible nouns), the gender cue could exert its influence. The mismatching gender didn’t lead to a complete rejection of the prediction, and didn’t activate just

any syntactically compatible noun. It was specifically the error – morphosyntactic gender mismatch – that was used for targeted adjustment of relevant features.

Our third condition using neutral adjectives is unique in the literature on prediction and grammatical gender. The data show that the trend of looks to the expected image continues when hearing the neutral adjective, but clearly lacks the boost in looks seen with the congruent adjective. While this simple trend continuation cannot answer as to what might be happening during the neutral adjective, a number of considerations should be taken in to account. Since neutral adjectives lack any gender marker that could support or contradict an initial prediction, comprehenders may turn to evaluate their semantic aspects as combined with compatible nouns. This would provide an additional semantic-based cue, in line with cues provided by the sentence context as a whole. It is also possible that co-occurrence probabilities of adjectives with certain nouns cause a more automatic effect of neutral adjectives on predictions. In that case, the adjectives may elicit automatic activation of a noun that is more associated with them, which could either support or contradict the initial prediction. It would be of interest for future studies to use corpus analyses to evaluate predictions based on semantic relationships or co-occurrence patterns of adjectives and nouns.

Time Course

Our time-window analysis confirms that participants immediately used gender information carried by the adjective to adjust predictions before the final noun was revealed. While we observed a small change in fixation patterns beginning already during the adjective, our third time-window between adjective offset and noun onset clearly shows that looks were directed even more towards the image that matched the morphosyntactic gender cue. This effect was obtained despite the time constraint that our design posed on the processor, for the adjective

carrying gender information was placed immediately before the target noun with no intervening constituents. Moreover, in the case of the masculine and feminine adjectives in Spanish, gender information is encoded only on the last phoneme of the word, further constraining the window available for adjustment.

This immediate change is in line with previous experiments showing very early prediction effects of gender information on adjectives, emerging within the first 50 ms after the critical inflection onset (Van Berkum et al., 2005). It is important to point out again, however, that the neutral adjectives in Spanish do not have a single point of inflection that categorises them as one gender or the other, but are a unit that as a whole form is ambiguous regarding gender. This may suggest a different time-course of their effect as opposed to gender-marked adjectives, especially within the adjective duration time-window. Future studies may provide more information by examining the earliest point of disambiguation in the input at which the adjective is recognizable, which would presumably be when the gender “neutrality” is available for extraction.

Noun Integration

Lastly, our data from the fourth time-window show that the influence of predictions persisted even once the target input was received. Looks at this point were unsurprisingly directed toward the target image – following the auditory reference and as preparation for clicking on it. However, the focus of interest was whether and how the advantage of the target over the competitor image would be affected by preceding information (sentence context and adjective gender). Our results revealed higher target-to-competitor fixation ratios when the target matched participants’ initial expectation, indicating that it was more easily integrated compared to a target that was not supported by sentence context. Indeed, facilitative effects of supportive

contexts on word integration are well established in the literature (Hess et al., 1995; Jordan & Thomas, 2002; James L. McClelland & O'Regan, 1981; Van Berkum et al., 2005); however, here we show that context effects persisted even after the presentation of additional predictive input (i.e., the adjective). This is especially interesting in the cases where the gender-marked adjective contradicted the context-based expectation. It seems that despite the clear prediction update triggered by the mismatching adjective, the initial context bias was not completely overridden in participants' updated predictions. While in the time-window between adjective offset and noun onset this could be interpreted as an attempt to reconcile the contradicting cues toward a final prediction (Lowder & Ferreira, 2016), the main effect of Expectancy during noun presentation suggests a lingering influence of the unrealised prediction – the context-expected noun.

Lingering effects have previously been reported in repair disfluencies, where an initially uttered word continues to influence participants' offline structural interpretation, despite the immediate correction to an alternative word (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Corley, 2010; Ferreira, Lau, & Bailey, 2004; Hartman & Hasher, 1991; Lau & Ferreira, 2005). This has been attributed to a trace left by the originally uttered word that is not eliminated from the comprehender's sentence representation; and seen as a consequence of incremental processing (Ferreira & Bailey, 2004; Ferreira et al., 2004; Lau & Ferreira, 2005; Lowder & Ferreira, 2016). Our data, however, demonstrate that even a predicted word can carry such a persisting effect: the initial prediction remained significant despite the receipt of a contradicting gender cue, despite being falsified when the target appeared, and, crucially – despite never even being uttered in the input. We see such persistence as support for the prominence of predictions

during sentence processing; they are not merely a side outcome of constraining contexts, but a significant component with lasting implications.

The effect of adjective gender-markedness in the noun time-window shows that target integration was facilitated by the presence of gender-marked adjectives compared to neutral adjectives. Gender information provided participants with an additional predictive cue allowing stronger pre-activation of the matching noun, whereas neutral adjectives provided no such advantage. We again emphasize that looks to the gender-matching, but context-unrelated, filler images in our displays were not affected by gender-marked adjectives, and instead remained negligible. That is, it was not the case that hearing a grammatical gender marker immediately directed participants' attention towards nouns that matched this gender; but instead the sentence context and gender-marked adjective were used *together* to converge on the most supported prediction. This finding firstly adds to the accounts of semantic and grammatical predictive cues working together for efficient noun integration (Bentrovato, Devescovi, D'Amico, Wicha, & Bates, 2003; Wicha et al., 2005), and further emphasises the lasting effects of top-down predictions as they combine with the actual bottom-up input received.

However, it also highlights a certain asymmetry between the semantic and grammatical cues. The gender cue did not cause activation of a semantically-incompatible yet gender-matching distractor; but the semantic expectancy effect did persist during noun integration even after receiving a gender-mismatch cue. This hints at a certain advantage for semantic over grammatical information, at least in the current experiment. One reason for this may be the temporal precedence of the semantic context. Being established earlier could have given the semantic context an advantage, in that a mismatching gender cue would have to compete with – and override – the existing prediction. Relatedly, the semantic constraint was built up over a

longer time period and contained more linguistic material; the leading sentence context consisted of several words taking a number of seconds to unfold, while the gender cue was only a single phoneme.

This quantitative aspect also lends itself to a more qualitative question about whether certain cues are less credible than others during prediction, perhaps because they are more prone to error. For example, it only takes a small pronunciation or perception error at one phoneme for the gender of an adjective to be the opposite of what was intended. On the other hand, changing the meaning of a semantic context – usually comprised of several words – would typically require much more than that. Moreover, a grammatical gender error may be frequent in certain situations, causing it to be processed differently – for example, coming from a non-native speaker (Caffarra & Martin, 2018). There seem to be plausible situation-dependent reasons for the system to favor certain cues over others. Clearly, the weight given to different cues and the interaction between them warrants further investigation.

Conclusion

In conclusion, our study shows that during spoken language comprehension people use sentential context to anticipate specific words, and rapidly adjust these expectations as new information (e.g. gender markers) arrives. Moreover, the effects of prior context and gender cues are apparent even after the target word is revealed, confirming the prominent role that prediction plays in word integration.

Acknowledgments

This work was partially supported by the Spanish Ministry of Economy and Competitiveness (MINECO), the Agencia Estatal de Investigación (AEI) and the Fondo Europeo de Desarrollo Regional (FEDER) (grant PSI2015-65694-P, “Severo Ochoa” programme SEV-2015-490 for

Centres of Excellence in R&D), and by the Basque government (grant PI_2016_1_0014). Further support derived from the AThEME project funded by the European Commission 7th Framework Programme, the ERC- 2011-ADG-295362 from the European Research Council. This project was also supported by the Spanish Ministry of Economy and Competitiveness (MINECO) through the convocatoria 2016 Subprograma Estatal Ayudas para contratos para la Formación Posdoctoral 2016, Programa Estatal de Promoción del Talento y su Empleabilidad del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016, reference FJCI-2016-28019. We thank Jordi Martorell and the Proactive group for helpful discussion throughout this work.

Disclosure of Interest

The authors report no conflict of interest.

References

- Akinina, Y., Malyutina, S., Ivanova, M., Iskra, E., Mannova, E., & Dragoy, O. (2015). Russian normative data for 375 action pictures and verbs. *Behavior Research Methods*, *47*(3), 691–707.
<https://doi.org/10.3758/s13428-014-0492-9>
- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. <https://doi.org/10.1006/jmla.1997.2558>
- Altmann, G. T. M., & Kamide, J. (2004). Now You See It, Now You Don't: Mediating the Mapping between Language and the Visual World. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347–386). New York, NY, US: Psychology Press.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. <https://doi.org/10.1016/j.jml.2006.12.004>
- Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560.
<https://doi.org/10.1080/01690965.2012.658072>
- Bates, D., Maechler, M., & Dai, B. (2009). lme4: Linear mixed-effects models using Eigen and Eigen. 2009. R package version 0.999375-31. URL: [Http://CRAN.R-project. Org/Package= Lme4](http://CRAN.R-project.org/Package=lme4).
- Bentrovato, S., Devescovi, A., D'Amico, S., Wicha, N., & Bates, E. (2003). The Effect of Grammatical Gender and Semantic Context on Lexical Access in Italian Using a Timed Word-Naming Paradigm. *Journal of Psycholinguistic Research*, *32*(4), 417–430.
<https://doi.org/10.1023/A:1024899513147>

- Boland, J. E. (2005). Visual arguments. *Cognition*, 95(3), 237–274.
<https://doi.org/10.1016/j.cognition.2004.01.008>
- Caffarra, S., & Martin, C. D. (2018). Not all errors are the same: ERP sensitivity to error typicality in foreign accented speech perception. *Cortex*. <https://doi.org/10.1016/j.cortex.2018.03.007>
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing Referential Domains during Real-Time Language Comprehension. *Journal of Memory and Language*, 47(1), 30–49. <https://doi.org/10.1006/jmla.2001.2832>
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, 42(4), 368–407.
<https://doi.org/10.1006/cogp.2001.0752>
- Clark. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, 6(1), 84–107. [https://doi.org/10.1016/0010-0285\(74\)90005-X](https://doi.org/10.1016/0010-0285(74)90005-X)
- Corley, M. (2010). Making predictions from speech with repairs: Evidence from eye movements. *Language and Cognitive Processes*, 25(5), 706–727. <https://doi.org/10.1080/01690960903512489>
- Dahan, D., Swingle, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic Gender and Spoken-Word Recognition in French. *Journal of Memory and Language*, 42(4), 465–480.
<https://doi.org/10.1006/jmla.1999.2688>
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12(3), 453–459.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121.
<https://doi.org/10.1038/nn1504>

- Dikker, S., & Pylkkänen, L. (2011). Before the N400: Effects of lexical–semantic violations in visual cortex. *Brain and Language*, *118*(1–2), 23–28. <https://doi.org/10.1016/j.bandl.2011.02.006>
- Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language*, *127*(1), 55–64. <https://doi.org/10.1016/j.bandl.2012.08.004>
- Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*(4), 1246–1258. <https://doi.org/10.3758/s13428-013-0326-1>
- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2017). MultiPic: A standardized set of 750 drawings with norms for six European languages. *The Quarterly Journal of Experimental Psychology*, 1–24. <https://doi.org/10.1080/17470218.2017.1310261>
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436. <https://doi.org/10.1007/BF02143160>
- Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences*, *36*(03), 211–212. <https://doi.org/10.1017/S0140525X12002312>
- Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. <https://doi.org/10.1111/j.1469-8986.2007.00531.x>
- Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. <https://doi.org/10.1006/jmla.1999.2660>
- Federmeier, K. D., McLennan, D. B., Ochoa, E. D., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146.

- Ferreira, F., & Bailey, K. G. D. (2004). Disfluencies and human language comprehension. *Trends in Cognitive Sciences*, 8(5), 231–237. <https://doi.org/10.1016/j.tics.2004.03.011>
- Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension, and Tree Adjoining Grammars. *Cognitive Science*, 28(5), 721–749.
<https://doi.org/10.1016/j.cogsci.2003.10.006>
- Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3), 417–458.
<https://doi.org/10.1007/s11229-007-9237-y>
- Harris, J. W. (1991). The Exponence of Gender in Spanish. *Linguistic Inquiry*, 22(1), 27–62.
- Hartman, M., & Hasher, L. (1991). Aging and Suppression: Memory for Previously Relevant Information. *Psychology and Aging*, 6(4), 587–594.
- Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, 124(1), 62–82.
<https://doi.org/10.1037/0096-3445.124.1.62>
- Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57(4), 460–482. <https://doi.org/10.1016/j.jml.2007.02.001>
- Jackendoff, R. (2002). *Foundations of knowledge*. Oxford: Oxford University Press.
- Jordan, T. R., & Thomas, S. M. (2002). In search of perceptual influences of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 34–45. <https://doi.org/10.1037/0278-7393.28.1.34>
- Kaiser, E., & Trueswell, J. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94(2), 113–147. <https://doi.org/10.1016/j.cognition.2004.01.002>
- Kamide, Y., Altmann, G. T. ., & Haywood, S. L. (2003a). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156. [https://doi.org/10.1016/S0749-596X\(03\)00023-8](https://doi.org/10.1016/S0749-596X(03)00023-8)

- Kamide, Y., Altmann, G. T. ., & Haywood, S. L. (2003b). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156. [https://doi.org/10.1016/S0749-596X\(03\)00023-8](https://doi.org/10.1016/S0749-596X(03)00023-8)
- Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of Syntactic and Semantic Information in Predictive Processing: Cross-Linguistic Evidence from German and English. *Journal of Psycholinguistic Research*, 32(1), 37–55. <https://doi.org/10.1023/A:1021933015362>
- Kapnoula, E. C., & McMurray, B. (2016). Training alters the resolution of lexical interference: Evidence for plasticity of competition and inhibition. *Journal of Experimental Psychology: General*, 145(1), 8–30. <https://doi.org/10.1037/xge0000123>
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59. <https://doi.org/10.1080/23273798.2015.1102299>
- Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, 20(5), 633–666. <https://doi.org/10.1080/01690960444000142>
- Lew-Williams, C., & Fernald, A. (2007). Young Children Learning Spanish Make Rapid Use of Grammatical Gender in Spoken Word Recognition. *Psychological Science*, 18(3), 193–198. <https://doi.org/10.1111/j.1467-9280.2007.01871.x>
- Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the visual-world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(9), 1400–1416. <https://doi.org/10.1037/xlm0000256>
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53(4), 372–380. <https://doi.org/10.3758/BF03206780>

- McClelland, J. L., & O'Regan, J. K. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 7(3), 634–644. <https://doi.org/10.1037/0096-1523.7.3.634>
- McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition : Implications for SLI. *Cognitive Psychology*, 60(1), 1–39. <https://doi.org/10.1016/j.cogpsych.2009.06.003>
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42. [https://doi.org/10.1016/S0010-0277\(02\)00157-9](https://doi.org/10.1016/S0010-0277(02)00157-9)
- McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, 131(1), 509–517. <https://doi.org/10.1121/1.3664087>
- Molinaro, N., Conrad, M., Barber, H. A., & Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience*, 1(1), 1–7. <https://doi.org/10.1080/17588920903373952>
- Molinaro, N., Giannelli, F., Caffarra, S., & Martin, C. (2017). Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals. *Cognition*, 164, 61–73. <https://doi.org/10.1016/j.cognition.2017.03.012>
- R Core Team. (2016). R: A Language and Environment for Statistical Computing. *Vienna, Austria*. Retrieved from <http://www.R-project.org>
- Salverda, A. P., & Tanenhaus, M. K. (2017). 5 The Visual World Paradigm. In A. M. B. de Groot & P. Hagoort (Eds.), *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide* (p. 89). John Wiley & Sons.
- Tanenhaus, M K, Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634.

- Tanenhaus, Michael K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye Movements and Lexical Access in Spoken-Language Comprehension: Evaluating a Linking Hypothesis between Fixations and Linguistic Processing. *Journal of Psycholinguistic Research*, 29(6), 557–580.
<https://doi.org/10.1023/A:1026464108329>
- Tanenhaus, Michael K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. *Handbook of Psycholinguistics*, 863–900.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467.
<https://doi.org/10.1037/0278-7393.31.3.443>
- Weber, A., Grice, M., & Crocker, M. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, 99(2), B63–B72.
<https://doi.org/10.1016/j.cognition.2005.07.001>
- Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346(3), 165–168.
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting Gender: An Event Related Brain Potential Study on the Role of Grammatical Gender in Comprehending a Line Drawing Within a Written Sentence in Spanish. *Cortex*, 39(3), 483–508. [https://doi.org/10.1016/S0010-9452\(08\)70260-0](https://doi.org/10.1016/S0010-9452(08)70260-0)
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16(7), 1272–1288.
<https://doi.org/10.1162/0898929041920487>
- Wicha, N. Y. Y., Orozco-Figueroa, A., Reyes, I., Hernandez, A., Gavaldón de Barreto, L., & Bates, E. A. (2005). When zebras become painted donkeys: Grammatical gender and semantic priming

interact during picture integration in a spoken Spanish sentence. *Language and Cognitive Processes*, 20(4), 553–587. <https://doi.org/10.1080/01690960444000241>

Winkler, I., & Czigler, I. (1998). Mismatch negativity: deviance detection or the maintenance of the standard? *NeuroReport*, 9(17), 3809–3813. <https://doi.org/10.1097/00001756-199812010-00008>

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, 742(1–2), 239–252. [https://doi.org/10.1016/S0006-8993\(96\)01008-6](https://doi.org/10.1016/S0006-8993(96)01008-6)

Table 1

Means (and standard deviations in parentheses) for the proportion of fixations on each image as a function of condition within each time window.

Time Window	1. Context end			2. Adjective Duration			3. Adjective Offset-Noun Onset		
	Congruent	Incongruent	Neutral	Congruent	Incongruent	Neutral	Congruent	Incongruent	Neutral
Expected	.39(.17)	.41 (.14)	.39 (.16)	.43 (.16)	.39 (.15)	.40 (.17)	.54 (.18)	.30 (.11)	.45 (.15)
Alternative	.26 (.09)	.26 (.10)	.26 (.10)	.28 (.09)	.33 (.10)	.32 (.11)	.21 (.09)	.49 (.13)	.34 (.11)
Gender filler	.09 (.06)	.08 (.05)	.09 (.06)	.08 (.06)	.08 (.05)	.08 (.06)	.07 (.05)	.08 (.05)	.06 (.06)
Distractor	.10 (.06)	.09 (.06)	.10 (.06)	.08 (.05)	.08 (.06)	.08 (.06)	.08 (.07)	.07 (.04)	.06 (.06)

Table 2

Means (and standard deviations in parentheses) for the proportion of looks to each image during the final noun duration, as a function of condition. Data from the neutral condition is divided into “expected” and “alternative” since the neutral adjective was followed by either the expected or the alternative noun at equal proportions.

Condition	Congruent	Incongruent	Neutral	
			Expected	Alternative
Expected	.77 (.13)	.11 (.06)	.74 (.15)	.18 (.07)
Alternative	.07 (.05)	.76 (.12)	.11 (.08)	.68 (.10)
Gender filler	.04 (.03)	.04 (.03)	.03 (.04)	.05 (.04)
Distractor	.04 (.03)	.03 (.03)	.03 (.04)	.04 (.03)



Fig. 1.

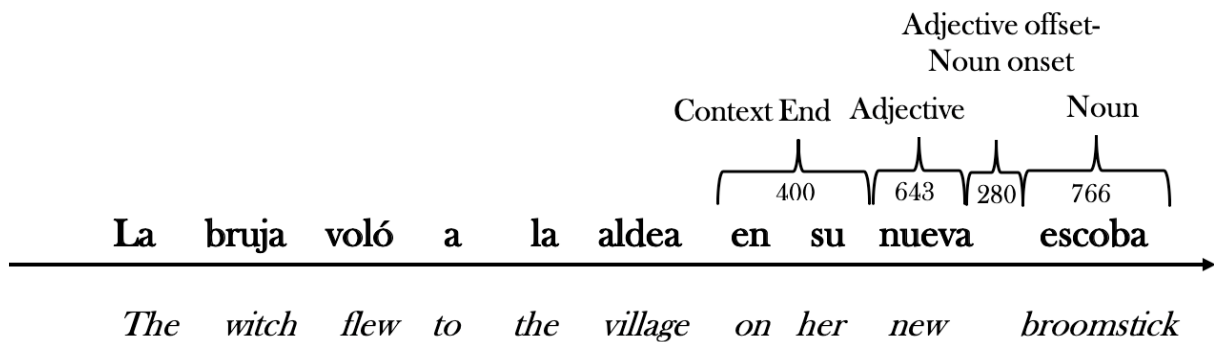


Fig. 2.

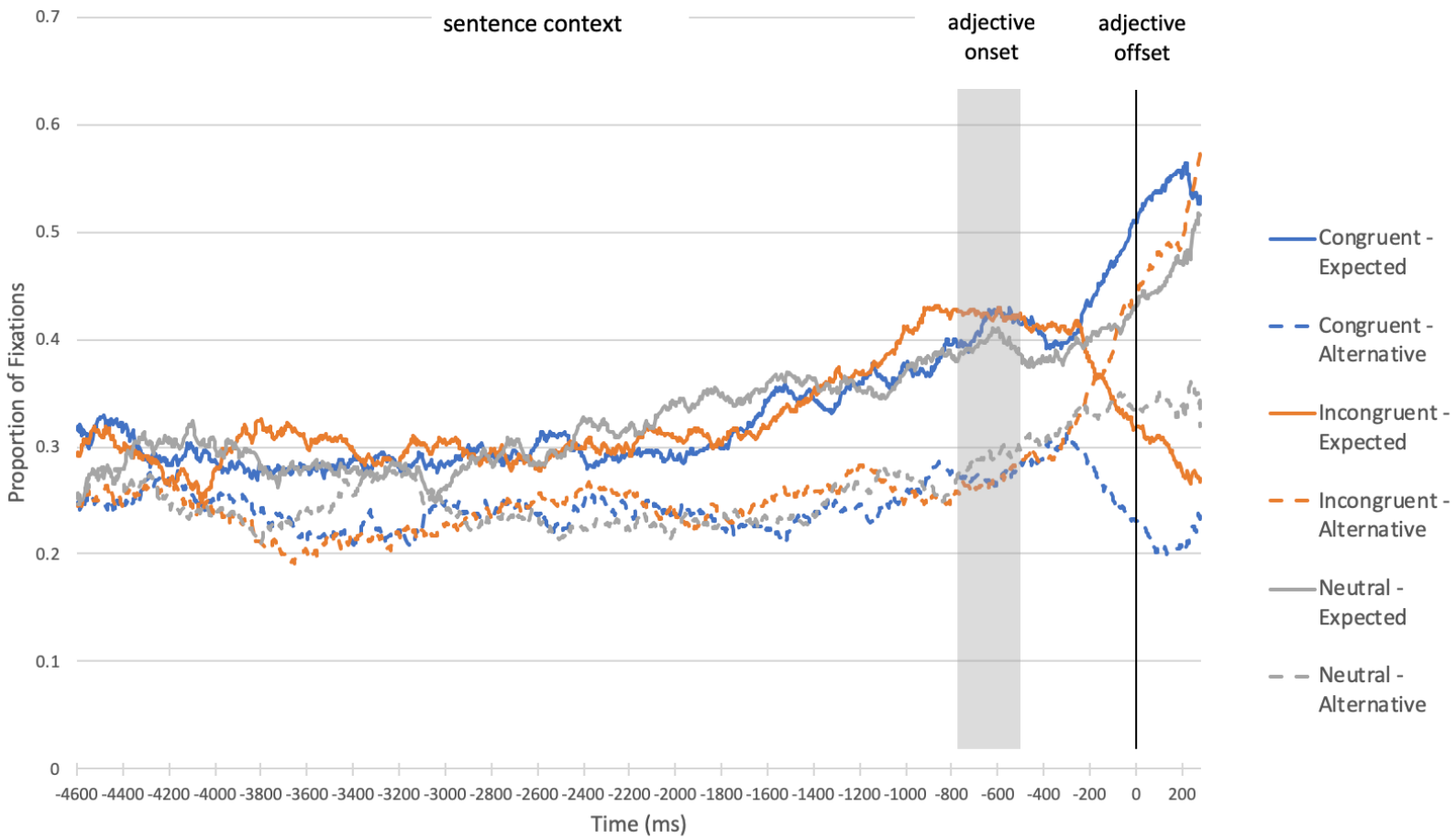


Fig. 3.

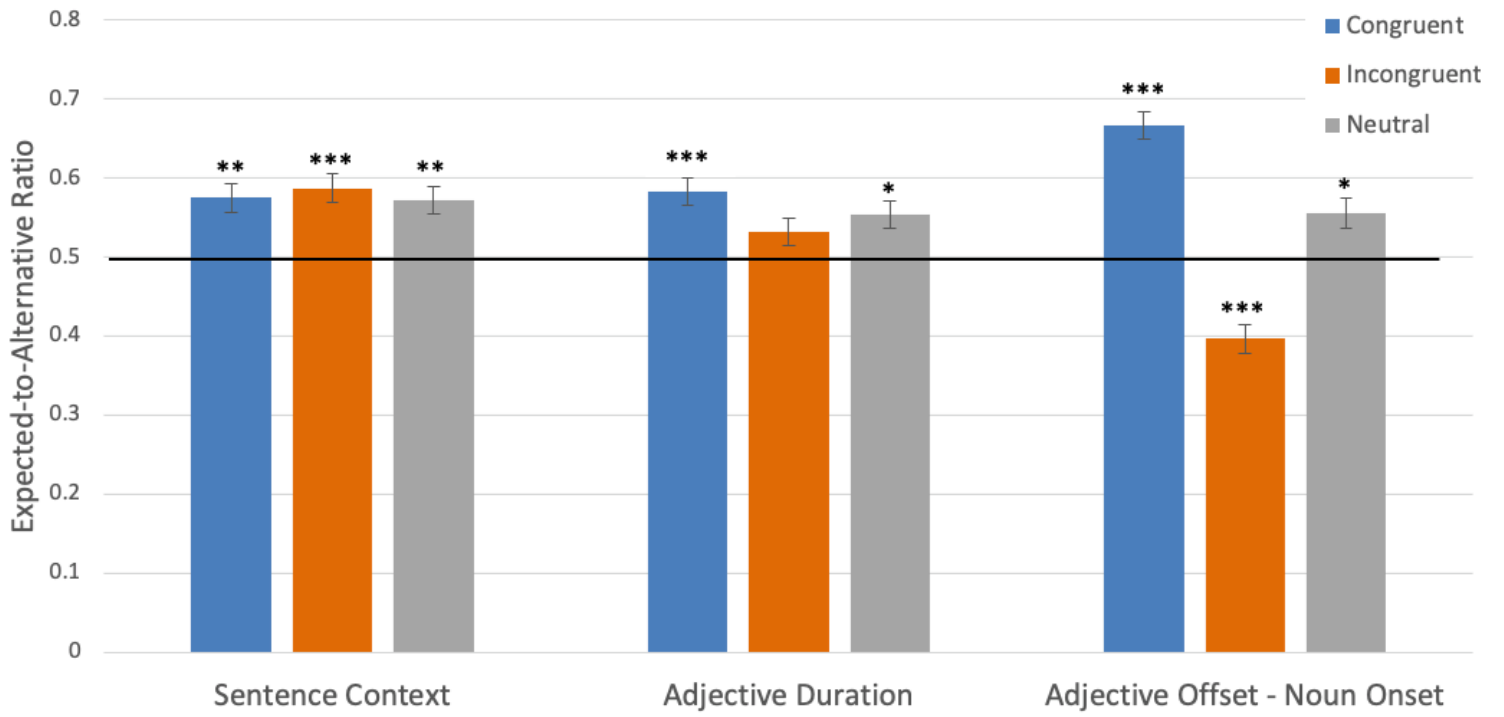


Fig. 4.

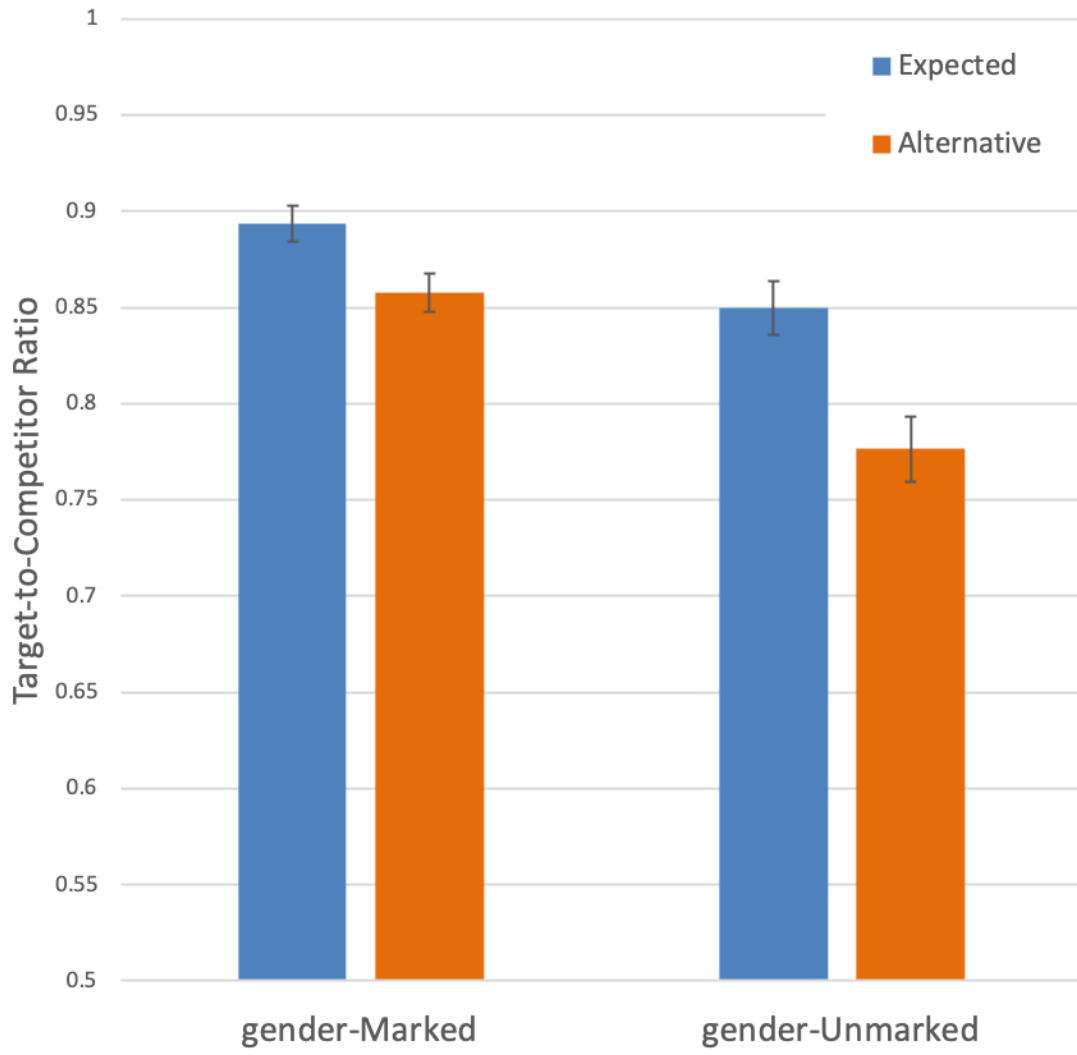


Fig. 5.

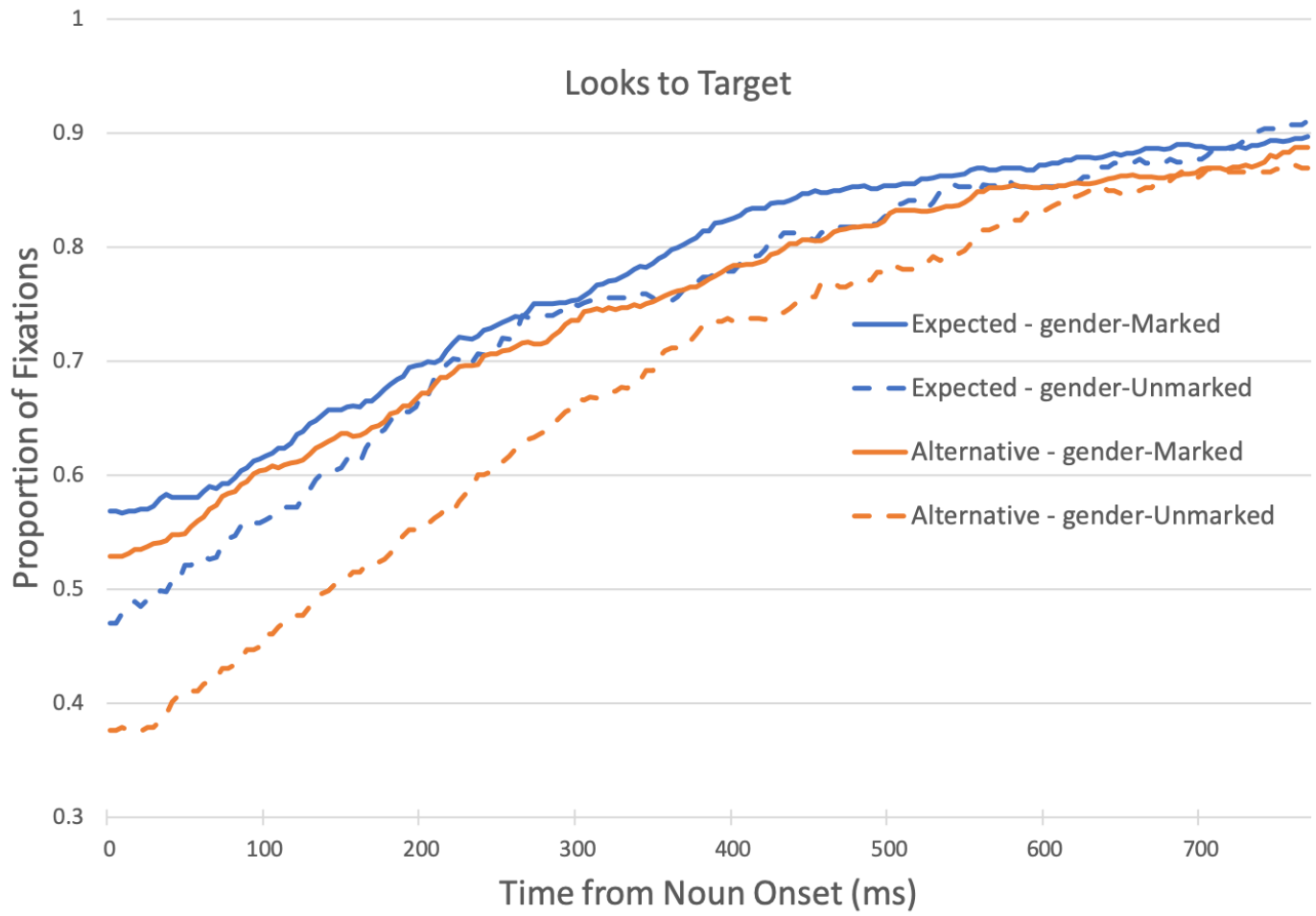


Fig. 6.

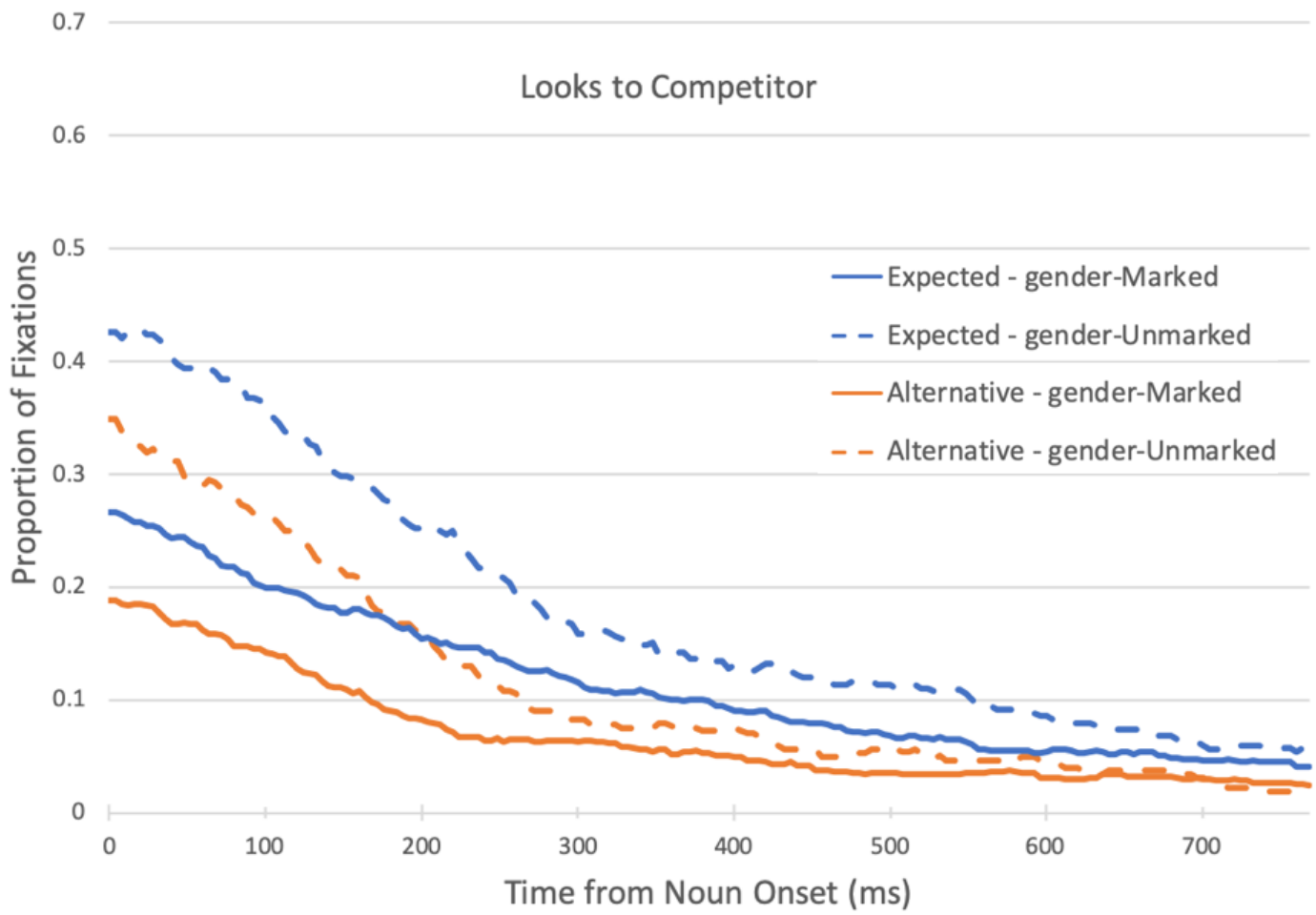


Fig. 7.

Figure Captions List

Fig. 1. Example of a visual display.

Fig. 2. The four time-windows used for data analyses, illustrated on an example sentence. The numbers in the brackets show the mean duration of each time-window in milliseconds (Context Ends was a fixed duration of 400 ms). The relative sizes of spaces in Fig. 2 are for illustration only and do not correspond to the actual relative lengths in the auditory recordings.

Fig. 3. Proportion of trials with fixations on the Expected and Alternative noun images in each condition. Computations were time-locked to adjective offset (marked by the vertical line at zero) and blocked by noun onset at the upper limit. Given the time needed for initiating eye-movement, every time-point was shifted forward by 200 ms. The graph presents data beginning 4600 ms before adjective offset to illustrate changes in proportions during the unfolding context; the grey rectangle illustrates average adjective onset at 643 ± 141 ms prior to adjective offset.

Fig. 4. Bar plots of the Expected-to-Alternative ratio as a function of condition within the first three time-windows. Asterisks indicate a significant difference against the value of .5 marked by the black horizontal line. Bars above this line indicate that the Expected noun received more fixations than the Alternative noun, bars below this line indicate that the Alternative noun received more fixations than the Expected noun. Ratios were logit-transformed for statistical analyses, but presented here in their raw form for ease of visualization. Error bars indicate ± 1 standard error (Morey, 2008). * = $p < .05$, ** = $p < .01$ *** = $p < .001$.

Fig. 5. Bar plots of the Target-to-Competitor ratios as a function of Expectancy and gender-Markedness, in the fourth time-window covering noun duration. Ratios were logit-transformed for statistical analyses, but presented here in their raw form for ease of visualization. Error bars indicate ± 1 standard error (Morey, 2008).

Fig. 6. Proportion of trials with fixations on the Target image, time-locked to noun onset and shown for average noun duration. The target was either the Expected or the Alternative noun, and followed either a gender-Marked or gender-Unmarked adjective. Given the time needed for initiating eye-movement, every time-point was shifted forward by 200 ms.

Fig. 7. Proportion of trials with fixations on the Competitor image, time-locked to noun onset and shown for average noun duration. The competitor was either the Expected or the Alternative noun, and followed either a gender-Marked or gender-Unmarked adjective. Given the time needed for initiating eye-movement, every time-point was shifted forward by 200 ms.