

Prediction of agreement and phonetic overlap shape sublexical identification

Andrea E. Martin^{1,2*}, Philip J. Monahan^{1,3,4*}, Arthur G. Samuel^{1,5,6}

1 Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián, Spain

2 Department of Psychology, School of Philosophy, Psychology and Language Sciences,
University of Edinburgh, Edinburgh, UK

3 Centre for French and Linguistics, University of Toronto Scarborough, Canada

4 Department of Linguistics, University of Toronto, Canada

5 IKERBASQUE Foundation, Donostia-San Sebastián, Spain

6 Department of Psychology, Stony Brook University, Stony Brook, USA

* Authors contributed equally. Names appear alphabetically.

Contact Information:

Andrea E. Martin
Department of Psychology
School of Philosophy, Psychology and
Language Sciences
University of Edinburgh
7 George Square
EH8 9JZ Edinburgh, UK
andrea.martin@ed.ac.uk
+44 (0) 131 650 34 30

Philip J. Monahan
Centre for French and Linguistics
University of Toronto Scarborough
1265 Military Trail
Toronto, ON M1C 1A4, Canada
philip.monahan@utoronto.ca
+1 (416) 287-7149

Keywords: Prediction, Grammatical Gender Agreement, Phoneme Identification, Speech

Perception, Sentence Processing, Spanish

Short Title: Vowel Identification Shaped by Gender Agreement

Word Count: 5502

Abstract

The mapping between the physical speech signal and our internal representations is rarely straightforward. When faced with uncertainty, higher-order information can be recruited to parse the signal. As such, the lexicon and some aspects of sentential context have been shown to modulate the identification of ambiguous phonetic segments. Here, using a phoneme identification task (i.e., participants judged whether they heard [o] or [a] at the end of an adjective in a noun-adjective sequence), we asked whether grammatical gender cues influence phonetic identification and if this influence is shaped by the phonetic properties of the agreeing elements. In three experiments, we show that phrase-level gender agreement in Spanish affects the identification of ambiguous adjective-final vowels. Moreover, this effect is strongest when the phonetic characteristics of the element triggering agreement and the phonetic form of the agreeing element are identical. Our data are consistent with models wherein listeners generate specific predictions, which are based on the interplay of underlying morphosyntactic knowledge and surface phonetic cues.

Introduction

It has long been known that lexical status affects the identification of ambiguous speech segments (Ganong, 1980). That is, a segment that is acoustically ambiguous between two speech sounds is more likely to be identified as the sound that results in the perception of an existing lexical item (Connine & Clifton, 1987; Fox, 1984; Ganong, 1980; Pitt, 1995; Pitt & Samuel, 1993). It appears then that the integration of higher-order knowledge, in this case, lexical representation, is utilized in the interpretation of lower-level structure. Given the abundant noise and variation that exist in the physical speech signal, the ability of the perceptual system to integrate various knowledge sources would clearly be advantageous (for a review from a speech processing perspective, see Ulsar et al., (2013) and Wendt et al., (2014)). Many models implement some form of top-down information flow as a core tenet, including interactive-activation models (Bowers & Davis, 2004; McClelland & Elman, 1986) and forward-models (Gagnepain, Henson, & Davis, 2012; Halle & Stevens, 1962; Poeppel & Monahan, 2011). Others interpret such “top-down” findings as ancillary task effects only affecting decision-stage nodes (Norris, McQueen, & Cutler, 2000).

In addition to effects that originate at the lexical level, various supra-lexical biases have been identified, arising from lexical-semantic context (Borsky, Shapiro, & Tuller, 2000; Borsky, Tuller, & Shapiro, 1998; Connine, 1987; Miller, Green, & Schermer, 1984), syntactic category constraints (Isenberg, Walker, & Ryder, 1980; van Alphen & McQueen, 2001), and lexical-pragmatic biases of verb selection (Rohde & Ettliger, 2012). Miller et al., (1984) showed that participants are more likely to identify an acoustically ambiguous string between *bath* and *path*, i.e., [ʔæθ], where [ʔ] is acoustically ambiguous between [b] and [p^h], as *bath* in the sentential context *She needs hot water for the ____*, and as *path* in the sentential context *She likes to jog along the ____*. The identification of the same physical stimulus varied as a function of the lexical-semantic context, suggesting a role for higher-order information

in the identification of lower-level segments. Van Alphen and McQueen (2001) created an acoustic continuum between the Dutch tokens *to* [tə] and *the* [də], and placed these tokens in a sentential context that either biased the infinitival *to* [tə] (e.g., ‘I try to/the shoot’) or article *the* [də] (e.g., I try to/the shoes) interpretation. Ambiguous tokens were more commonly recognized as the token that was consistent with the appropriate syntactic frame. Rohde and Ettliger (2012) report that the identification of ambiguous tokens along a *he/she* ([hi]-[fi]) continuum was modulated based on the biological gender of the co-referential subject and object (e.g., Tyler/Sue) and whether the main verb of the sentence was subject- or object-biased, suggesting that pragmatic context and reference resolution can bias phoneme identification. Like previous research, the current study investigates the role of supra-lexical factors in phonetic identification. In contrast previous research, however, we also assess the extent to which lower-level phonetic cues interact with such factors (see below for additional important differences).

We also examine a level of grammatical knowledge that is different than those studied previously. Lexical representations are discrete and presumed to be part of semantic memory (i.e., either a segmental string exists in an individual’s lexicon or not) while sentential context is computed dynamically and therefore more variable (i.e., sentences can be formed in combinatorial and productive ways compared to words). It is unknown whether categorical (i.e., discrete) linguistic knowledge of grammar that is neither lexicalized or contextual/variable in nature, such as grammatical agreement computation, can bias lower-level perceptual identification. An architecture in which both higher level representations, such as discrete lexical entries and variable sentential context, *and* grammatical ‘algorithms’, such as agreement computation, can constrain and shape the processing of phonemes seems highly plausible, but to our knowledge, there is no extant empirical evidence of this type. In addition to asking whether such categorical grammatical information can influence

perceptual identification, we also ask to what extent it interacts with lower-level surface phonetic cues to such information. We test for these effects in a language – Spanish – where adjectives must agree with the grammatical gender and number of their noun.

Psycholinguistic research on agreement has examined many different types of agreement relations (e.g., between nouns and verbs, determiners and nouns, pronouns and antecedents, and nouns and adjectives) in both comprehension and production. Prior research on agreement processing in Spanish has focused on several issues, including the semantic expectancy of an agreeing noun (Wicha, Moreno, & Kutas, 2004), the number of agreement features being violated (Antón-Méndez, Nicol, & Garrett, 2002; Barber & Carreiras, 2005), gender-based lexical priming effects (for a review see Friederici and Jacobsen (1999)), and more recently, effects of retrieval interference on agreement processing (Martin, Nieuwland, & Carreiras, 2012; 2014). From a formal linguistic perspective, it is clear that both in Spanish and more broadly (Comrie, 1999) agreement information can provide a very powerful signal of upcoming sentential and referential relationships that will need to be formed during ongoing production and comprehension. Yet, few studies focus on the implications of that fact for processing of upcoming speech input.

In Spanish, phrasal agreement computations occur routinely during language comprehension and production (Bock, Carreiras, & Meseguer, 2012). While there are some exceptions, the vast majority of adjectives routinely follow nouns in Spanish, resulting in a *noun-adjective* word order. All nouns in Spanish are lexically specified with either masculine or feminine grammatical gender. Determiners and adjectives that modify a given noun also must agree in grammatical gender with the noun. This agreement is categorical and obligatory. Therefore, a noun-adjective sequence is either encoded with masculine or feminine grammatical gender. The surface cues to grammatical gender, however, are less categorical in nature, though there are strong correlations. Teschner and Russell (1984)

observed that 99.89% of Spanish nouns that end in [o] are masculine and 96.6% of nouns that end in [a] are feminine. Thus, if a given Spanish noun ends in [o], it is very likely to have masculine gender; if it ends in [a], it is very likely to have feminine gender.

However, a substantial number of nouns end in sounds other than [o] or [a]. To calculate the relative percentage of masculine nouns that end in [o] and feminine nouns that end in [a], we used www.corpusdelespanol.org, selecting just those citations from the 20th century (20.4 million words; (Davies, 2002)). Looking only at singular-noun token counts, 62% of masculine nouns end in [o], while 55.9% of feminine nouns end in [a]. As such, [o] and [a] are strong but not infallible indicators of nominal grammatical gender in Spanish. For adjectives in Spanish, grammatical gender is often phonetically encoded in the word-final vowel: [o] for masculine agreement and [a] for feminine agreement. The vowel category to grammatical gender mapping is similar for nouns and adjectives. Therefore, we observe the following sequences: masculine: *el cielo bonito* ‘the beautiful sky’, feminine: *la cocina bonita* ‘the beautiful kitchen’.

Overall, gender agreement production errors are very rare in the natural speech of native speakers (Igoa, García-Albea, & Sánchez-Casas, 1999) or in laboratory settings that attempt to induce speech errors (Antón-Méndez et al., 2002). Montrul et al., (2008) performed a series of experiments testing Spanish speakers’ performance on gender agreement in oral production, as well as written recognition tasks. The primary motivation of the research was to compare L2 Spanish learners and Spanish heritage speakers on their ability to produce and recognize correct gender agreement in the language. Of interest to the current study, however, is that Montrul et al., (2008) included a control group of native monolingual speakers of Spanish and tested these speakers on canonical (e.g., Masculine nouns ending in [o]), non-canonical (e.g., Masculine nouns ending in [e]) and what they termed residual tokens (e.g., Masculine nouns ending in [a]). Across all tasks, these native

Spanish speakers performed at ceiling, with virtually no gender agreement errors. Moreover, Spanish speakers are sensitive to gender agreement violations (Faussart, Jakubowicz, & Costes, 1999), with such violations modulating the left-anterior negativity (LAN) and P600 electrophysiological brain responses (Barber & Carreiras, 2005). Collectively, the literature shows that native speakers of Spanish rarely, if ever, incorrectly select the appropriate gender for nouns, irrespective of whether the noun has a canonical gender ending or not.

Here, we investigate two potential cues that might shift the phonetic identification boundary: 1) underlying grammatical gender cues and 2) surface phonetic cues. In particular, we test whether the preceding noun's grammatical gender can shift phonetic identification, as well as how much of this shift is attributable to how indicative the noun-final vowel is in cuing a lexical item's grammatical gender. Like lexical status, our manipulation is fully predictive: the preceding noun's grammatical gender determines that of the adjective. But unlike lexical status, the preceding noun's grammatical gender cannot be pre-stored in the following adjective's lexical entry.

If identification is in fact shaped by grammatical gender, it is then important to know how listeners coordinate this underlying grammatical information with surface characteristics that robustly indicate this morphosyntactic property, or alternatively, the extent to which phonological priming might be at play in the case of Spanish grammatical gender agreement. Previous research has shown that phonological overlap, defined as the number of overlapping position-dependent phonemes between two words in a pair, can give rise to facilitation in processing (Jakimik, Cole, & Rudnicky, 1985; Radeau, Morais, & Segui, 1995; Slowiaczek & Pisoni, 1986; Slowiaczek, Nusbaum, & Pisoni, 1987). While these effects typically decay with time between presentation (Radeau et al., 1995), the presence of a clear word-final [o] (or [a]) could bias the identification of the ambiguous adjective-final vowel toward [o] (or [a]). Because surface cues ([o] indicates masculine, [a] indicates feminine) are not fully

predictive of gender, there are situations in which the underlying and surface cues are consistent, inconsistent, or neither (when the surface cues are absent), where we take *consistent* to mean that masculine nouns end in [o], and feminine nouns end in [a]. When nouns do not end in either of these two vowels, there is no strong predictor of the adjective's gender. These three cases allow us to test the relative weighting of underlying versus surface cues to grammatical gender agreement.

In three methodologically parallel experiments, we tested (1) nouns with a final vowel that robustly cued its grammatical gender, i.e., masculine nouns with final [o] and feminine nouns with final [a]; (2) nouns ending in segments (e.g., [e]) that were not strongly associated with a grammatical gender; and (3) nouns that ended in a vowel that is strongly associated with the 'opposite' grammatical gender of that noun. The relevant grammatical gender marking vowels ([o] and [a]) are adjacent in the Spanish vowel space, making it possible to construct a continuum between the two categories, and thus, a continuum between a masculine and feminine adjective.

Experiment 1

Experiment 1 tested transparent nouns' influence on identification of adjective-final vowels, i.e., [o] for masculine nouns (e.g., *cielo bonito* 'beautiful sky') and [a] for feminine nouns (e.g., *cocina bonita* 'beautiful kitchen'). Because the nouns follow the typical pattern, with [o] indicating masculine and [a] indicating feminine, any effects we find may be due to the underlying grammatical gender of the noun and/or the surface properties, including phonological priming (Radeau et al., 1995; Slowiaczek et al., 1987; Slowiaczek & Pisoni, 1986). In this case, the underlying and surface cues are consistent with one another.

Participants

Twenty-four (19 females; mean age = 21.9 years) native speakers of Spanish participated in Experiment 1. All participants reported normal hearing, provided written informed consent and were remunerated for their participation.

Materials

Eight Spanish adjectives that alternate in their final vowel between [o] and [a] (e.g., *bonito/bonita* ‘beautiful’) to agree with the grammatical gender of the modified noun were paired with eight masculine and eight feminine nouns. Each adjective was paired with one masculine noun and one feminine noun. The nouns were matched across grammatical gender for log frequency; on a paired t-test, the two sets of nouns were not reliably different. No determiners/articles, e.g., *el/la* ‘the (masc./fem.)’, were included in the experiment as they reliably signal the grammatical gender of the noun.

To ensure that there were no reliable differences attributable to the semantic fit between the noun and adjective in the masculine and feminine pairs, we had ten native speakers of Castilian Spanish naïve to the purposes of the experiments rate how well each noun and adjective fit together on a five-point Likert scale (1 = not very good semantic fit, e.g., non-sensical; 5 = very good semantic fit, e.g., makes sense). On the whole, the pairs were rated as having a good semantic fit (3.975/5), and the mean ratings were nearly identical across the masculine and feminine pairs (Masculine: 3.96/5; Feminine: 3.99/5), with no reliable statistical difference between them ($p = 0.92$). See Table 1 for a list of stimuli used in Experiment 1.

INSERT TABLE 1 ABOUT HERE

Stimulus Creation

We recorded a female native speaker of Castilian Spanish, who was naïve as to the purpose of the experiment, reading all noun-adjective pairs in a neutral sentential context, (e.g., *Repetiré cielo bonito otra vez más* ‘I will repeat beautiful sky again’) to ensure natural prosody. The entire recording time was approximately 1 hour and 15 minutes, but was broken up into three 15 minute sessions with at least 15 minutes between sessions. Recordings were made in a sound-attenuated booth and digitally sampled at 44.1 kHz on a computer located outside the booth with Adobe Audition (Version 4, CS5.5; Adobe Systems Inc., San Jose, California).

The following synthesis and re-synthesis procedures were carried out using Praat (Boersma & Weenink, n.d.). Experimental items, i.e., the nouns and adjectives, were spliced from the continuous recording and resampled to 11.025 kHz with 50 sample precision for the purpose of re-synthesis. A representative natural token of [o] produced in adjective final position was selected as the stimulus from which the seven-step vowel continuum was constructed. Synthesis was accomplished using a linear predictive coding (LPC) analysis/reanalysis method. First, the source (glottal spectrum) and filter (extracted using LPC coefficients) of the vowel token were separated. A formant object was created using the Burg method as implemented in Praat. Specifically, the vowel token was resampled to twice the maximum formant (in this case 11 kHz given that the maximum formant to be found was at 5500 Hz) and a Pre-emphasis was applied from 50 Hz. Subsequently, in 25 ms windows, LPC coefficients were computed (Burg, 1978; Press, Teukolsky, Vetterling, & Flannery, 1992) using ten poles. The first (F1) and second formants (F2) of the extracted filter were manipulated in a FormantGrid object to create the seven-step continuum and later recombined with the source (Zölzer, 2002). The fundamental pitch (f0) and higher formants (F3, F4, etc.) were not manipulated. The seven vowel tokens, ranging from the endpoint [o] (center frequency F1: 523 Hz, F2: 1200 Hz) to the endpoint [a] (center frequency F1: 777 Hz,

F2: 1772 Hz) in equal first (F1) and second formant (F2) steps (F1=43 Hz/step; F2=95/Hz step), were re-spliced onto the adjectives, such that each adjective had seven variants: one token for each step along the continuum. The endpoints of the F1 and F2 synthesized [o]-[a] adjective continua were determined by selecting the closest natural tokens of [o]- and [a]-final (produced in the context of the adjectives) to the mean F1 and F2 values of the noun final-vowel tokens. This continuum was used in Experiment 1 for all experimental items (note: the same continuum was also used in Experiments 2 and 3). The formant center F1 and F2 frequencies are provided in Table 2.

INSERT TABLE 2 ABOUT HERE

We chose adjectives in which the final consonant was always the voiceless stop [t] to avoid misleading formant transition cues between the release of the consonant and the adjective-final vowel when we re-spliced the seven vowel versions. Stimulus intensity was normalized such that the average, root-mean-square, intensity was 70 dB SPL, and the output volume via headphones remained constant across participants.

To ensure that the resulting vowel continuum was, in fact, perceived as systematically shifting from [o] to [a], we conducted a pretest with participants who did not take part in Experiment 1. Native speakers of Castilian Spanish (n=10) listened to one presentation of each of the seven tokens (one of each step of the continuum) of the eight adjectives from Experiment 1, for a total of 56 trials. The order of presentation was pseudo-randomized and participants were asked to identify whether they heard [o] or [a]. The identification curve (see Figure 1) shows the typical sigmoidal function for speech continua (Liberman, Fry, Abramson, & Eimas, 1962).

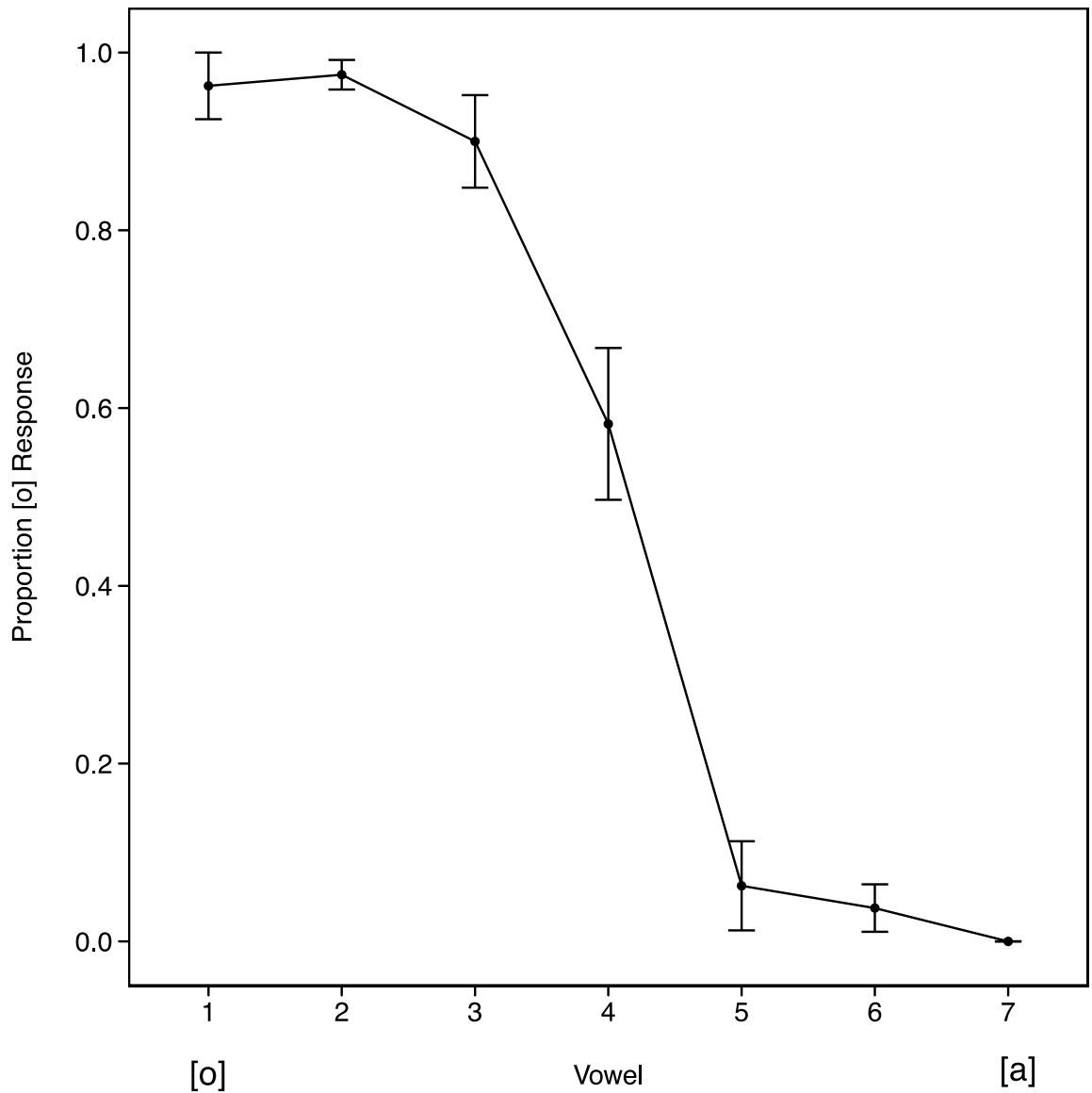


Figure 1

Identification curve for the results from the pretest. Step 1 corresponds to [o], while Step 7 corresponds to the vowel [a]. Error bars represent one standard error of the mean.

Apparatus and Procedure

Participants were seated in front of a computer monitor and listened to stimuli over Beyer Dynamic DT770pro headphones at a comfortable volume in sound-attenuated cabins. Stimuli were delivered using DMDX (K. I. Forster & Forster, 2003). Each trial began with a fixation point, i.e., “+”, to indicate the start of a trial. The fixation point remained on the screen for

780 ms and was immediately followed by the presentation of the auditory stimulus. On half the trials, an adjective alone (No Context) was played and on the other half of trials, the adjective was preceded either by a grammatically masculine (Masculine) or feminine (Feminine) noun. In a two-alternative forced choice task (2AFC), participants were asked to indicate as quickly and accurately as possible whether they heard [o] or [a] at the end of the adjective. The presentation of the stimuli was randomized and the inter-trial interval (ITI) pseudo-randomly varied between 390 ms and 835 ms. Participants first completed a short practice session to become familiarized with the task. All participants received all possible stimuli in a fully within-subjects design (224 total trials; 16 nouns-with-adjectives \times 7 vowel steps + 8 adjectives-only \times 7 vowel steps \times 2 repetitions). Response button assignments were counter-balanced across participants.

Results

Trials with response times greater or less than 2.5 standard deviations of each participant's overall mean reaction time were eliminated from the analysis (2.92% of all items; Ratcliff (1993)). For phonetically ambiguous tokens, we predicted more [o] responses following a masculine noun and fewer [o] responses following a feminine noun. To test this hypothesis, we submitted our results to a mixed-effects model (Baayen, Davidson, & Bates, 2008) with a logistic link function (Jaeger, 2008) using the `glmer()` function in the `lme4` library (Bates, Maechler, & Bolker, 2011) for the R statistical package (R Development Core Team, 2012). The fixed effects structure of the model contained the factors Context (three levels: Masculine, Feminine, No Context), Vowel (seven steps: 1-7; centered: Step 1 = -3, Step 4 = 0, Step 7 = 3) and their interaction. The random effects structure contained random by-subject

and by-item slopes for vowel step.¹ The reported models accounted for significantly more variance than the null model, which included only the random effects structure ($p < 0.001$). Post-hoc pairwise comparisons were performed using generalized linear hypotheses testing using Tukey contrasts as implemented with the `glht()` function in the `multcomp` library (Hothorn, Bretz, & Westfall, 2008). The following number of data points were included in the model for each condition: Masculine = 1279, Feminine = 1300 and No Context = 2623. Mixed effects models are particularly suited for handling unbalanced designs (Baayen et al., 2008).

¹ Note that the maximal model, i.e., including Vowel Step and Context as by-subject and by-item random slopes, as well as including Context alone as a random slope, as advocated by Barr et al., (Barr, Levy, Scheepers, & Tily, 2013), caused a failure of the model to converge. See Barr et al., (Barr et al., 2013, p. 276) for discussion of the limitations of maximal random effects structure in this class of models.

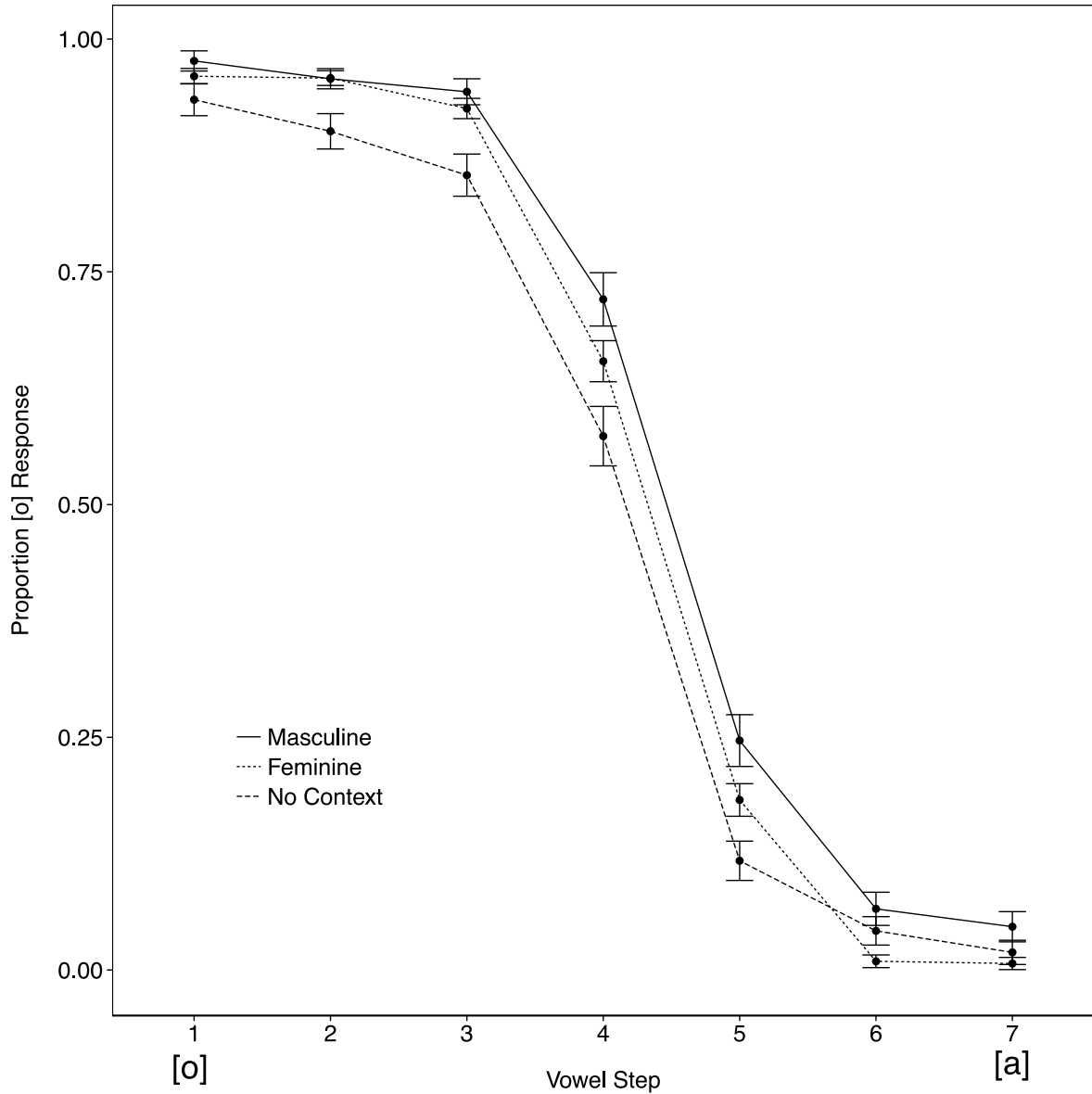


Figure 2 Identification curves as a function of proportion /o/ responses for Experiment 1.

Responses are to the adjective final vowel when preceded by nouns of distinct grammatical genders (masculine: solid line; feminine: dotted line; no context: dashed line). Overall, more [o] responses were made when the nominal context was masculine while more [a] responses were made when the nominal context was feminine. Step 1 on the continuum was most often responded to as [o] and Step 7 as [a]. Error bars represent the standard error of the mean.

In Experiment 1, we tested whether the grammatical gender of a preceding noun influences the identification of an ambiguous adjective-final vowel, using nouns that were phonologically transparent, i.e., masculine nouns ended in [o] and feminine nouns ended in [a]. Figure 2 shows the identification functions. As expected from the pilot testing of the continuum, there was a main effect of Vowel ($\beta = -3.29$, $SE = 0.26$, $z = -12.46$, $p < 0.001$). The central question of Experiment 1 is whether the identification function is shifted as a function of the gender of the preceding noun.

In fact, masculine noun contexts elicited more /o/ responses than feminine contexts ($\beta = 1.26$, $SE = 0.17$, $z = 7.50$, $p < 0.001$), and both were reliably different from the No Context condition (Masculine/No Context: $\beta = 0.57$, $SE = 0.16$, $z = 3.64$, $p < 0.01$; Feminine/No Context: $\beta = -0.69$, $SE = 0.14$, $z = -4.92$, $p < 0.001$). Thus, the preceding noun shifted the identification of adjective-final ambiguous vowels in the direction of grammatical gender agreement. Consistent with previous reports, the largest shift occurred in the middle of the continuum (Pitt & Samuel, 1993), producing an interaction between Vowel and Masculine/No context ($\beta = 0.48$, $SE = 0.14$, $z = 3.45$, $p < 0.001$) and Vowel and Feminine/No context ($\beta = 0.70$, $SE = 0.13$, $z = 5.53$, $p < 0.001$).

Discussion

Alternating vowel-final adjectives end in [o] in Spanish when preceded by a masculine noun and in [a] when preceded by a feminine noun. Our results demonstrate that listeners are sensitive to this morphosyntactic pattern, producing a reliable shift in the identification of ambiguous adjective-final vowels in the expected directions. Such vowels were more reliably identified as /o/ in the context of a grammatically masculine noun than when there was no context, or when the context was a grammatically feminine noun. Similarly, we found a reliable identification shift in the direction of [a] when the nominal context was

grammatically feminine, compared to when it was preceded by either a masculine noun or with no noun context. Thus, when presented with ambiguous acoustic information, listeners appear to use the relatively abstract grammatical gender feature of the preceding noun to interpret the ambiguous signal.

A potential limitation of Experiment 1 is that all of the context nouns ended in either [o] or [a], with [o] always indicating a masculine noun and [a] always indicating a noun that was feminine. Conceivably, the results might not be due to the grammatical property of the noun, but instead from a sort of phonological priming (Radeau et al., 1995; Slowiaczek et al., 1987; Slowiaczek & Pisoni, 1986). That is, the noun-final vowel could influence the identification of the adjective-final vowel, independent of any morphosyntactic properties. Therefore, in the second experiment, we used grammatically masculine and feminine nouns that do not phonologically overlap with the adjective-final vowels of interest.

Experiment 2

In Experiment 2, we isolated the contribution of a noun's inherent grammatical gender from any overlapping phonetic cues shared with the adjective by testing grammatically masculine and feminine nouns that do not overlap with the adjective-final vowels of interest. The adjective tokens (and continua) from Experiment 1 were used, but nouns that did not end in [o] or [a] now preceded the adjective. If identification of the ambiguous vowels is shifted under these conditions then the shift can be directly attributed to the underlying grammatical gender.

Participants

Thirty-one native speakers of Castilian Spanish (13 female; mean age=22.5 yrs.) who did not take part in Experiment 1 participated in Experiment 2. All reported normal hearing, provided written informed consent and were remunerated for their participation.

Materials

The same adjective tokens (and continua) from Experiment 1 were used again in Experiment 2. We selected 16 new nouns (8 masculine, 8 feminine) that did not end in [o] or [a]. Each adjective was again paired with two nouns, one masculine and one feminine. For the items in Experiment 2, the same ten native speakers of Castilian Spanish who took part in the pre-experiment ratings for Experiment 1 judged how well the noun and adjective fit together on a five-point Likert scale (1 = not very good semantic fit, e.g., non-sensical; 5 = very good semantic fit, e.g., makes sense). Participants rated the new pairs as having a good semantic fit (3.95/5), with mean ratings nearly identical across the masculine and feminine pairs (Masculine: 3.98/5; Feminine: 3.92/5); there was no reliable difference between them ($p = 0.86$). Additionally, the nouns were matched across grammatical gender for log frequency ($p = 0.11$) and length ($p = 0.33$). The recording of the nouns was done during the same session as the recording of nouns and adjectives for Experiment 1. Stimulus construction procedures were identical to those in Experiment 1. See Table 3 for a list of stimuli used in Experiment 2.

INSERT TABLE 3 ABOUT HERE

Apparatus and Procedure

The procedure was identical to Experiment 1. Response button assignments were counter-balanced across participants. This was a fully within-subjects design – all participants were presented with all items.

Results

Trials with response times greater or less than 2.5 standard deviations of each participant's overall mean reaction time were eliminated from the analysis (3.03% of all items). The linear mixed effects model structure and post-hoc comparisons were identical to those in Experiment 1. The following number of data points were included in the model for each condition: Masculine = 1665, Feminine = 1672 and No Context = 3361. As expected, the main effect of Vowel ($\beta = -2.67$, $SE = 0.16$, $z = -16.93$, $p < 0.001$) was significant; it did not interact with either Masculine/No Context or Feminine/No Context (all $ps > 0.5$).

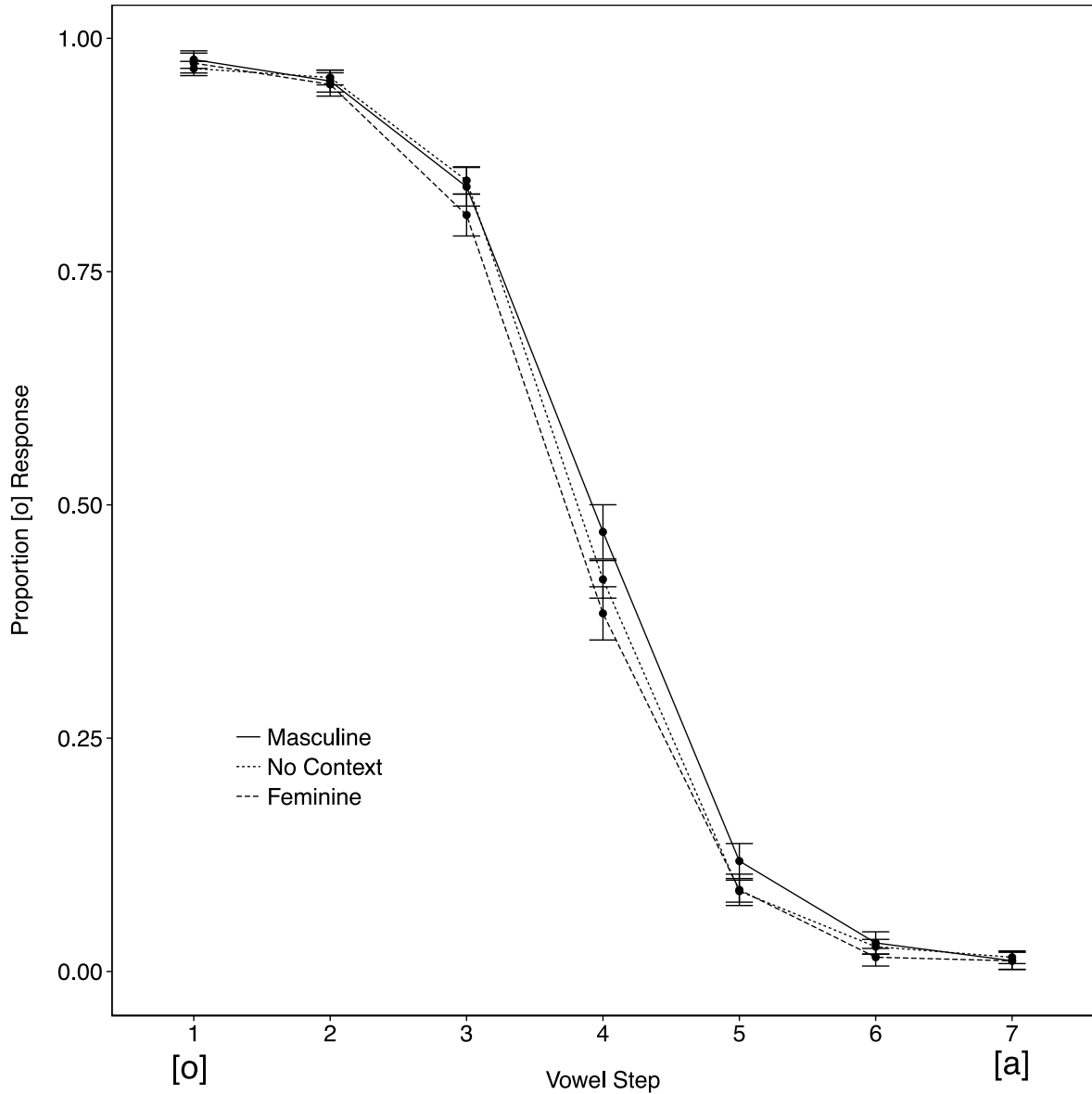


Figure 3 Identification curves as a function of proportion /o/ responses for Experiment 2.

Responses are to the adjective final vowel when preceded by nouns of distinct grammatical genders (masculine: solid line; feminine: dotted line; no context: dashed line). Overall, more [o] responses were made when the nominal context was masculine while more [a] responses were made when the nominal context was feminine. Step 1 on the continuum was most often responded to as [o] and Step 7 as [a]. Error bars represent the standard error of the mean.

In Experiment 2, the noun contained no phonologically transparent cue to its grammatical gender. As Figure 3 shows, this clearly produced smaller shifts. Nonetheless, we found reliably more /o/ responses when the adjective was preceded by a masculine noun compared to a feminine noun ($\beta = 0.51$, $SE = 0.15$, $z = 3.44$, $p < 0.01$), while the difference between the feminine and No Context conditions was marginal ($\beta = -0.29$, $SE = 0.13$, $z = -2.22$, $p = 0.07$). There was no difference between the No Context condition and the Masculine condition ($p > 0.1$).

Discussion

In Experiment 1, we found a reliable difference in the identification of ambiguous final vowels in adjectives, consistent with the grammatical gender of the preceding noun. A limitation in that experiment was that the final vowel of the noun in each item overlapped with the phonetic realization of the grammatical agreement of the adjective. In Experiment 2, we eliminated the surface cues between the noun and adjective in a pair, thus providing a purer test of the effect of underlying grammatical gender. Even in the absence of phonetic overlap between the final vowels of the noun and adjective, we observed a bias in identification of ambiguous vowel tokens. This suggests that the abstract property of grammatical gender can shift the identification of an ambiguous stimulus in a subsequent word.

The weaker effect found in Experiment 2 compared to Experiment 1 suggests that in addition to this grammatical effect, there is also an effect of the consistency of the surface features of the noun and adjective. Specifically, identification shifts were larger when the two factors were consistent with each other. Spanish allows an even stronger test of whether grammatical cues alone can influence the identification of phonetic segments. There is a small set of nouns that violate the tendency toward grammatically masculine nouns ending in

[o] and grammatically feminine nouns ending in [a] (Harris, 1991): There are some masculine nouns that end in [a] and a very small number of feminine nouns that end in [o]. The results of Experiment 2 show that in the absence of surface [o] or [a] on a noun, abstract gender can affect vowel identification on the following adjective. The exception words that violate the usual pattern allow us to pit surface cues against abstract grammar. In Experiment 3, we tested whether the abstract gender effect is strong enough to produce a detectable effect when pitted against surface cues.

Experiment 3

To produce an extreme case of morphological and phonological divergence, a case where cues are inconsistent and thus potentially working against each other, Experiment 3 tested pairs in which the final vowel of the masculine nouns was [a] and the final vowel of the feminine nouns was [o], the opposite pattern from the norm (e.g., *foto favorita* ‘favorite photograph’).

Participants

Twenty-four native speakers of Castilian Spanish (12 female; mean age=22.6 yrs.) who did not participate in Experiments 1 or 2 took part in this experiment. All reported normal hearing, provided written informed consent and were remunerated for their participation.

Materials

The same adjective tokens (and continua) from Experiment 1 and Experiment 2 were used again in Experiment 3 with the inclusion of 2 additional new adjectives. We selected 10 new

nouns (5 masculine, 5 feminine), whose final vowel is the opposite of the canonical ending. There is a possible concern as to whether the noun-final [o] and [a] differ acoustically as a function of the grammatical gender of the noun. That is, perhaps the masculine-final [o] is more *prototypical* than the feminine final [o], as tested here in Experiment 3. As such, to determine if the noun-final vowels in Experiment 3 were acoustically distinct from those in Experiment 1, the F1 and F2 for each noun-final vowel was measured at its steady-state portion and these formant frequencies were submitted to a linear regression model with the given formant (F1 or F2) as the dependent variable and Experiment (Experiment 1 vs Experiment 3) as the predictor. Experiment was not a reliable predictor for either F1 ($\beta = 38.5$, $SE = 74.21$, $t = 0.52$, $p = 0.61$) or F2 ($\beta = -205.6$, $SE = 121.2$, $t = -1.69$, $p = 0.10$) suggesting that the primary acoustic characteristics of the noun-final vowels were not substantially distinct (see Table 4 for a comparison of the mean and standard deviation for F1 and F2 of the noun-final vowels used in Experiments 1 and 3).

INSERT TABLE 4 ABOUT HERE

Fewer nouns were utilized in Experiment 3 because items of this type are relatively rare in the language (although a number of them are frequent, e.g., *la mano* ‘hand’, *el tema* ‘theme’, *la moto*, ‘motorcycle’, *la foto* ‘photograph’).² The same speaker of Castilian Spanish who recorded the items in Experiments 1 and 2 recorded the nouns and adjectives for Experiment 3. The stimuli were recorded and processed identically to those in the preceding experiments, including use of the same [o]-[a] continuum in the previous two experiments. Because fewer

² Two of the items in Experiment 3 are back-formations of longer forms, i.e., *la moto* ‘motorcycle’ from *la motocicleta* and *la foto* ‘photograph’ from *la fotografía*. Given that these back-formation nominals are the more frequent forms, we see no reason why an item being a back-formation should influence our findings.

nouns were available, each noun was paired with two adjectives. As such, participants in Experiment 3 heard each adjective more often than the participants in Experiments 1 and 2. The nouns were controlled for log frequency ($p = 0.58$) and length ($p = 0.14$). See Table 5 for a list of stimuli used in Experiment 3. Five native speakers assessed the naturalness of the noun-adjective pairs used in Experiment 3; all pairs were judged to be equally natural within and between conditions.

INSERT TABLE 5 ABOUT HERE

Apparatus and Procedure

As in Experiments 1 and 2, participants listened to adjectives that were either presented alone or preceded by a noun. Response button assignments were counterbalanced across participants. The experiment was fully within-subjects – all participants were presented with all items.

Results

The data were analyzed using the same linear mixed effects model and post hoc comparisons as in Experiments 1 and 2. Trials with response times greater or less than 2.5 standard deviations of each participant's overall mean reaction time were eliminated from the analysis (3.26% of all items). The following number of data points were included in the model for each condition: Masculine = 1596, Feminine = 1621 and No Context = 3250. As in the previous experiments, we observed a main effect of Vowel ($\beta = -2.67$, $SE = 0.18$, $z = -14.47$, $p < 0.001$).

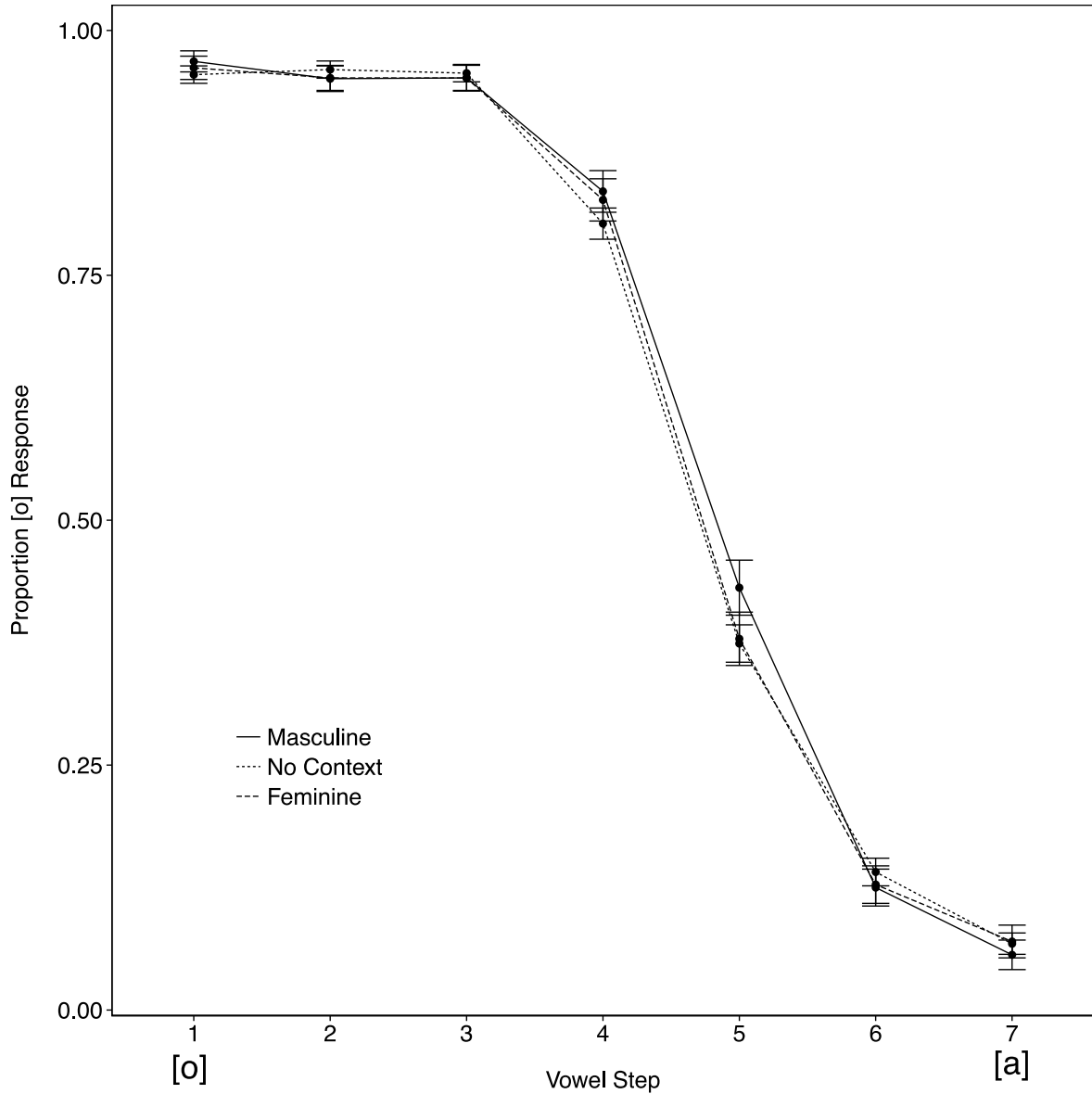


Figure 4 Identification curves as a function of proportion /o/ responses for Experiment 3.

Responses are to the adjective final vowel when preceded by nouns of distinct grammatical genders (masculine: solid line; feminine: dotted line; no context: dashed line). Overall, more [o] responses were made when the nominal context was masculine while more [a] responses were made when the nominal context was feminine. Step 1 on the continuum was most often responded to as [o] and Step 7 as [a]. Error bars represent the standard error of the mean.

We observed a significant difference between masculine and feminine contexts ($\beta = 0.55$, $SE = 0.23$, $z = 2.44$, $p < 0.05$) and masculine noun contexts elicited reliably more /o/ responses compared to the No Context condition ($\beta = 0.77$, $SE = 0.20$, $z = 3.90$, $p < 0.001$), with the interaction between Vowel and Masculine/No Context ($\beta = -0.45$, $SE = 0.15$, $z = -2.94$, $p < 0.01$) indicating that this difference was primarily in the most ambiguous range of the continuum. There was no difference between the feminine and No Context conditions ($p > 0.4$), and there was no interaction between Vowel and Feminine/No Context ($p > 0.1$).

INSERT FIGURE 2 ABOUT HERE

Discussion

The normative situation in Spanish is that masculine nouns end in [o], feminine nouns end in [a], and adjectives that modify them have final vowels that match those in the nouns. In Experiment 3, we selected stimuli that could actually mislead listeners regarding gender – the nouns came from the relatively rare set of items that have the opposite surface mapping of final vowels and gender. In particular, if phonetic overlap between the final vowels in the noun and adjective were primarily responsible for the perceptual bias shift in Experiment 1 (i.e., more [o]-responses when preceded by a masculine noun because the masculine nouns ended in [o]), then we might have anticipated the opposite pattern in Experiment 3 (i.e., more [a]-responses following masculine nouns). That the overall shift occurred in the same direction as in Experiments 1 and 2 suggests a strong role for the noun's grammatical gender in determining the observed shifts in bias. Despite the misleading cues, there was still a measurable (though clearly diminished) effect of underlying grammatical gender, with a significant shift in the predicted direction for masculine nouns relative to both the Feminine

and No Context cases. Note that, as discussed previously, it is very unlikely that the participants retrieved the wrong gender for the nominal items used in Experiment 3 (see Montrul et al., 2008).

Across-Experiment Comparison

Examining the relative effects of bias across experiments, the largest difference in proportion [o] responses between the Masculine and Feminine contexts was found in Experiment 1, while the smallest difference in proportion [o] responses was found in Experiment 3 (see Figure 5). This is borne out in the Cohen's d estimates of effect size for the Masculine/Feminine contrast across experiments (for the ambiguous regions of the continuum only): Experiment 1: 0.52; Experiment 2: 0.24; Experiment 3: 0.11. This pattern is consistent with the prediction that phonetic cues, in addition to grammatical properties of the noun, influence phonetic identification. To determine the relative contribution of the phonetic and morphosyntactic cues, we submitted the results of all three experiments to a linear mixed effects model with a logistic link function, as above. The model contained fixed effects of Context (Masculine, Feminine and No Context (default contrast)), Experiment (One (default contrast), Two, Three) and Vowel Step (1-7; centered), interactions between all three fixed effects and random by-subject and by-item intercepts.

INSERT TABLE 6 ABOUT HERE

To determine the full pairwise comparisons for Experiment and Context, we submitted the output of the logistic mixed-effects model to simultaneous tests for General Linear Hypotheses using Tukey Contrasts, as implemented in the `glht()` function (Hothorn et al.,

2008) in the R statistical environment. We find reliable differences for each of the three Context contrasts (Masculine/Feminine: $\beta = 1.25$, $SE = 0.17$, $z = 7.48$, $p < 0.001$; Masculine/No Context: $\beta = 0.56$, $SE = 0.16$, $z = 3.60$, $p < 0.01$; Feminine/No Context: $\beta = -0.69$, $SE = 0.14$, $z = -4.93$, $p < 0.001$). Moreover, we also observe differences in the proportion of [o] responses for Experiments 1 and 2 ($\beta = -1.14$, $SE = 0.40$, $z = -2.85$, $p < 0.05$) and Experiments 2 and 3 ($\beta = -2.05$, $SE = 0.40$, $z = -5.12$, $p < 0.001$) and a marginal difference between Experiments 1 and 3 ($\beta = 0.91$, $SE = 0.43$, $z = 2.14$, $p = 0.08$). Please refer to Table 6 for the full statistical output of the model and Figure 5 for a comparison of the proportion [o]-responses aggregated over the ambiguous regions of the continuum (Steps 3-6) by Context and Experiment.

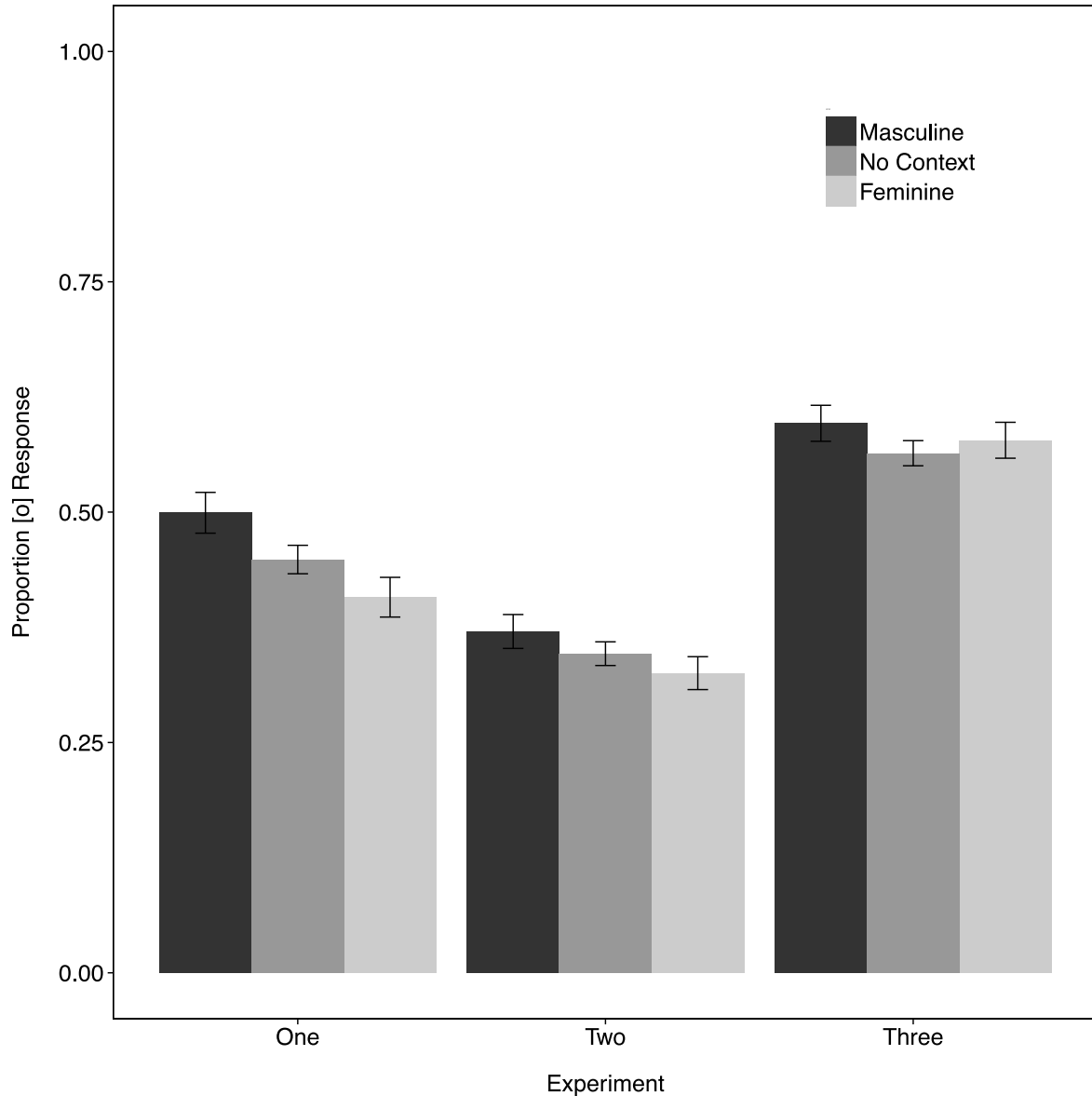


Figure 5 Mean proportion [o] responses aggregated over the ambiguous steps of the continuum (Steps 3, 4, 5,6) by experiment and condition. Error bars represent the standard error of the mean.

General Discussion

This is the first work to investigate whether a categorical grammatical cue – gender agreement – influences phonetic identification and whether canonical surface cues modulate the strength of this effect. When presented with ambiguous acoustic information, listeners

used underlying grammatical cues from the preceding noun to interpret the ambiguous signal. In all three experiments, we found evidence for this, with the effect systematically reduced as surface information changed from consistent, to neutral, to inconsistent with the abstract morphosyntactic form. The decrease in the shift across experiments points to a contribution from surface phonetic information, but the across-experiment pattern demonstrates that surface information alone cannot account for our effects. In particular, when such surface cues were neutral (Experiment 2), there were still significant shifts due to grammatical gender. There was even a small residual shift when the surface cues should work against the effect (Experiment 3). Speculatively, the apparent asymmetry in Experiment 3 (when only underlying cues are available, only the masculine condition shifts significantly) could be related to differences in frequency between [o] and [a] being indicative of masculine and feminine nouns in Spanish, or the number of masculine nouns ending in [a] compared to the number of feminine nouns ending in [o].

Our findings extend the class of previously observed effects by using a productive, non-probabilistic morphosyntactic manipulation and by providing evidence for underlying and surface cues acting additively. Our results demonstrate that abstract grammatical information from one word is carried forward and affects the phonetic processing of the subsequent word, and that there is an important interplay between abstract linguistic representations and bottom-up phonetic cues.

One observation for which we do not have a good understanding is that the between-experiment variability in the proportion [o]-responses across experiments. In Experiment 1, for example, the proportion [o]-responses at Step 4 was 68%, versus 42% in Experiment 2 and 82% in Experiment 3. Moreover, a comparison of the identification plots suggests that the entire perceptual boundary is shifted at least one step toward the [a]-end of the continuum,

i.e., overall, more [o]-responses, in Experiment 3 compared to Experiments 1 and 2. The locus of this shift requires further investigation.

It is not obvious how models built specifically for lexical and phonemic processing (McClelland & Elman, 1986; Norris, 1994) can account for such sublexical effects from a source that is clearly supra-lexical. Both manifestations of the adjective are extant lexical items (or variants of a single lexical item) in Spanish, and the particular form required is dependent upon the grammatical characteristics of a preceding lexical item that shares some syntactic relationship, in this case, a head-complement relationship within a nominal phrase. As a consequence, the locus of these effects must be accounted for within an architecture that permits the use of supra-lexical information to inform sublexical identification, and to do so forward in time. Several models of sentence comprehension assume that information can be projected ahead in the parse, such that syntactic structure is projected based on phrase structure rules (Frazier & Clifton, 1996) or verb information projects word categories and their structure (Gorrell, 1995).

We are not aware, however, of a computationally implemented model where the incoming input's lexical features (aside from word category in the case of category-ambiguous words) and/or identity are subject to the current state of the system, or to the cumulative information of the previous inputs. It might be possible to adopt the approach taken by Townsend and Bever (2001) in extending the Halle and Stevens (1962) analysis-by-synthesis framework to a sentence processing architecture. Although the levels of representation relevant to our design are not explicitly mentioned in their schematized model, the analysis-by-synthesis principle entails two important computational requirements: (1) Information from the previous parse/cycle is carried forward in the form of representational hypotheses about what the next cycle's input is likely to be, and crucially, (2) based on these hypotheses, the grammar constrains the ultimate representational state of the current input

and the postulation of hypotheses for the next cycle. This approach may also be compatible with surprisal-based models of sentence comprehension (Hale, 2003). Although this class of model has focused on syntactic ambiguity resolution, prediction or expectation of grammatical agreement relationships between words could be said to contribute to one adjective form's conditional probability relative to another form. More generally, the results of the current experiments are compatible with the notion that the observed identification shifts arise from such real-time hypothesis generation during online language comprehension.

References

- Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, 5(1), 1–25. <http://doi.org/10.1111/1467-9612.00045>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <http://doi.org/10.1016/j.jml.2007.12.005>
- Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17(1), 137–153. <http://doi.org/10.1162/0898929052880101>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <http://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D. M., Maechler, M., & Bolker, B. (2011). Lme4: Linear mixed-effects classes using S4 classes. Retrieved from <http://cran.r-project.org/package=lme4>
- Bock, K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language*, 66(1), 17–37. <http://doi.org/10.1016/j.jml.2011.07.008>
- Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer [Computer Program]. Retrieved from <http://www.praat.org>.
- Borsky, S., Shapiro, L. P., & Tuller, B. (2000). The temporal unfolding of local acoustic information and sentence context. *Journal of Psycholinguistic Research*, 29(2), 155–168.
- Borsky, S., Tuller, B., & Shapiro, L. P. (1998). “How to milk a coat.” The effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, 103, 2670–2676.
- Bowers, J. S., & Davis, C. J. (2004). Is speech perception modular or interactive? *Trends in Cognitive Sciences*, 8(1), 3–5. <http://doi.org/10.1016/j.tics.2003.10.015>
- Burg, J. P. (1978). Modern spectrum analysis. *Childers, DG (Ed.)*, 34–41.
- Comrie, B. (1999). Grammatical gender systems: A linguist's assessment. *Journal of Psycholinguistic Research*, 28(5), 457–466. <http://doi.org/10.1023/A:1023212225540>
- Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, 26(5), 527–538.
- Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2), 291–299.
- Davies, M. (2002). *Corpus del Español: 100 Million Words, 1200s-1900s*. Retrieved from <http://www.corpusdelespanol.org/>
- Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista Di Linguistica* 11(1): 75-101.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116–124.
- Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10(4), 526–540.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: The MIT Press.
- Friederici, A. D., & Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research*, 28(5), 467–484. <http://doi.org/10.1023/A:1023264209610>
- Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken

- words in auditory cortex. *Current Biology*, 22(7), 615–621.
<http://doi.org/10.1016/j.cub.2012.02.015>
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125.
- Gorrell, P. (1995). *Syntax and Parsing*. Cambridge, UK: Cambridge Univ Press.
- Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2), 101–123. <http://doi.org/10.1016/j.bandl.2014.10.006>
- Halle, M., & Stevens, K. N. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory*, 8(2), 155–159.
- Harris, J. W. (1991). The exponence of gender in Spanish. *Linguistic Inquiry*, 22(1), 27–62.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <http://doi.org/10.1002/bimj.200810425>
- Igoa, J. M., García-Albea, J. E., & Sánchez-Casas, R. (1999). Gender-number dissociations in sentence production in Spanish. *Rivista Di Linguistica*, 11(1), 163–196.
- Isenberg, D., Walker, E. C. T., & Ryder, J. M. (1980). A top-down effect on the identification of function words. *The Journal of the Acoustical Society of America*, 68, S48.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. <http://doi.org/10.1016/j.jml.2007.11.007>
- Jakimik, J., Cole, R. A., & Rudnicky, A. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language*, 24, 165–178.
- Lieberman, A. M., Fry, D. B., Abramson, A. S., & Eimas, P. D. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5(4), 171–189. <http://doi.org/10.1177/002383096200500401>
- Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage*, 59(2), 1859–1869. <http://doi.org/10.1016/j.neuroimage.2011.08.057>
- Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain and Language*, 135, 42–51. <http://doi.org/10.1016/j.bandl.2014.05.001>
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86.
- Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Attention, Perception, & Psychophysics*, 36(4), 329–337.
- Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The Effects of Age and Context of Acquisition. *Language Learning*, 58(3), 503–553. <http://doi.org/10.1111/j.1467-9922.2008.00449.x>
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23(3), 299–325. <http://doi.org/10.1017/S0140525X00003241>
- Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 1037–1052.
- Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, 19(4), 699–725.
- Poeppl, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, 26(7), 935–951.

- <http://doi.org/10.1080/01690965.2010.493301>
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing*. Cambridge University Press, Cambridge, MA.
- R Development Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1297–1311.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.
- Rohde, H., & Ettliger, M. (2012). Integration of pragmatic and phonetic cues in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 967–983. <http://doi.org/10.1037/a0026786>
- Slowiaczek, L. M., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition*, *14*(3), 230–237.
- Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*(1), 64–75.
- Teschner, R. V., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, *1*, 115–132.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: The MIT Press.
- Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., & Kollmeier, B. (2013). Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test. *The Journal of the Acoustical Society of America*, *134*(4), 3039–18. <http://doi.org/10.1121/1.4818760>
- van Alphen, P. M., & McQueen, J. M. (2001). The time-limited influence of sentential context on function word identification. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1057–1071.
- Wendt, D., Brand, T., & Kollmeier, B. (2014). An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities. *PLoS ONE*, *9*(6), e100186–14. <http://doi.org/10.1371/journal.pone.0100186>
- 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. <http://doi.org/10.1162/0898929041920487>
- Zölzer, U. (2002). Source-filter processing. In D. Arfib, F. Keiler, & U. Zölzer (Eds.), *DAFX: Digital Audio Effects*. Chichester, UK: John Wiley & Sons, Ltd.

Acknowledgements

We thank Larraitz Lopez for assistance with data collection and Oihana Vadillo for lending her voice for the recordings. AEM was supported by a Juan de la Cierva Fellowship [JCI-2011-10228] from the Spanish Ministry of Science and Innovation, and a Future Research Leaders grant from the Economic and Social Research Council of the United Kingdom [ES/K009095/1]. PJM was supported by a Marie Curie Fellowship from the European Research Council [FP7-People-2010-IIF; Project No. 275751]. This work was also supported by grants, PSI2010-17781 and PSI2014-53277 from the Spanish Ministerio de Economía y Competitividad. Additional support was provided by Ministerio de Ciencia E Innovacion Grant #PSI2014-53277 and by Ayuda Centro de Excelencia Severo Ochoa SEV-2015-0490 to the BCBL.

Table 1

List of nouns, adjectives and their pairings used in Experiment 1. The grammatical gender of the noun is provided in parentheses.

Nouns	Adjectives
salto ‘jump’ (M)	alto/a ‘high/tall’
planta ‘plant’ (F)	
cielo ‘sky’ (M)	bonito/a ‘beautiful’
cocina ‘kitchen’ (F)	
ruido ‘noise’ (M)	distinto/a ‘different’
balda ‘shelf’ (F)	
dinero ‘money’ (M)	fortuito/a ‘chance/fortuitous
salida ‘exit’ (F)	
pulso ‘pulse’ (M)	lento/a ‘slow’
entrega ‘delivery’ (F)	
hombro ‘shoulder’ (M)	molesto/a ‘irritating’
pantalla ‘screen/monitor’ (F)	
estudio ‘study’ (M)	astuto/a ‘shrewd’
guerra ‘war’ (F)	
fallo ‘error’ (M)	tonto/a ‘stupid’
lucha ‘fight’ (F)	

Table 2

Center formant frequencies for the first (F1) and second (F2) formants (in Hz) for the seven-step continuum used in Experiments 1, 2 and 3. The continuum was synthesized from a natural token of [o] using the LPC re-synthesis method described in the Materials section. F1 was increased in equal-sized increments of 43 Hz, while F2 in equal-sized increments of 95 Hz.

Token	1 = [o]	2	3	4	5	6	7 = [a]
F1	523	564	606	649	692	735	777
F2	1200	1295	1391	1486	1581	1676	1772

Table 3

List of nouns, adjectives and their pairings used in Experiment 2. The grammatical gender of the noun is provided in parentheses.

Nouns	Adjectives
monitor 'screen' (M)	alto/a 'high/tall'
madre 'mother' (F)	
jardín 'garden' (M)	bonito/a 'beautiful'
luz 'light' (F)	
cinturón 'belt' (M)	distinto/a 'different'
ración 'portion/share' (F)	
examen 'exam' (M)	fortuito/a 'chance/fortuitous'
suerte 'luck' (F)	
camión 'truck' (M)	lento/a 'slow'
muerte 'death' (F)	
alcohol 'alcohol' (M)	molesto/a 'irritating'
verdad 'truth' (F)	
personal 'staff/personnel' (M)	astuto/a 'shrewd'
nación 'nation' (F)	
placer 'pleasure' (M)	tonto/a 'stupid'
juventud 'youth' (F)	

Table 4

List of nouns, adjectives and their pairings used in Experiment 3. The grammatical gender of the noun is provided in parentheses.

Nouns	Adjectives
cometa 'comet' (M)	alto/a 'high/tall'
líbido 'libido' (F)	
enigma 'mystery' (M)	bonito/a 'beautiful'
radio 'radio' (F)	
sistema 'system' (M)	distinto/a 'different'
foto 'photograph' (F)	
sistema 'system' (M)	fortuito/a 'chance/fortuitous'
líbido 'libido' (F)	
cometa 'comet' (M)	lento/a 'slow'
moto 'motorcycle' (F)	
enigma 'mystery' (M)	molesto/a 'irritating'
radio 'radio' (F)	
poema 'poem' (M)	astuto/a 'shrewd'
mano 'hand' (F)	
poema 'poem' (M)	tonto/a 'stupid'
moto 'motorcycle' (F)	
tema 'theme' (M)	favorito/a 'favorite'
foto 'photograph' (F)	
tema 'theme' (M)	resuelto/a 'decisive'
mano 'hand' (F)	

Table 5

The mean and one standard deviation for F1 and F2 of the noun-final vowels utilized in Experiments 1 and 3.

Experiment		F1 (Hz)	SD	F2 (Hz)	SD
1	o	502	51.2	1530	399.9
3		456	104.8	1250	253.8
1	a	751	70.7	1758	155.9
3		874	86.5	1627	123.6

Table 6

Output of linear mixed effects model with logistic link function across all three Experiments.

Consult the text for the specifics of the model specification. Note: β = Estimate; SE =

Standard Error; M = Masculine Context; F = Feminine Context; N = No Context; 1 =

Experiment 1; 2 = Experiment 2; 3 = Experiment 3. Indicators of p -values: * = < 0.05, ** <

0.01, *** < 0.001.

	β	SE	z -score	p (> z)	
(intercept)	0.9961	0.3515	2.834	< 0.01	**
Feminine	-0.6871	0.1394	-4.928	< 0.001	***
Masculine	0.5651	0.1571	3.597	< 0.001	***
Experiment 2	0.9129	0.4266	2.14	< 0.05	*
Experiment 3	-1.1352	0.3988	-2.847	< 0.01	**
F/N x 1/3	-3.2101	0.212	-15.14	< 0.001	***
M/N x 1/3	0.9066	0.22	4.121	< 0.001	***
F/N x 1/2	0.2133	0.2519	0.847	0.40	
M/N x 1/2	0.4019	0.1893	2.123	< 0.05	*
F/N x Vowel	-0.348	0.2009	-1.733	0.08	
M/N x Vowel	0.705	0.1267	5.565	< 0.001	***
1/2 x Vowel	0.4704	0.1392	3.38	< 0.001	***
1/3 x Vowel	0.5077	0.2765	1.836	0.07	
F/N x 1/2 x Vowel	0.5392	0.2625	2.054	< 0.05	*
M/N x 1/2 x Vowel	-0.8857	0.1857	-4.77	< 0.001	***
F/N x 1/3 x Vowel	-0.9277	0.2074	-4.473	< 0.001	***
M/N x 1/3 x Vowel	-0.6882	0.175	-3.932	< 0.001	***