

## **Supplementary Material**

The Supplementary Material contains the details of the Method that complement the essential information provided in the main text. The complete list of sentences is available in Foucart et al. (Foucart, Ruiz-Tada, & Costa, 2015).

### **MATERIALS AND DESIGN**

We used the same 52 highly constrained sentences used in Foucart et al. (2014). Expected and unexpected nouns were matched in frequency (expected: 50.9 per million, SD: 58.25; unexpected: 43.18 SD: 55.43  $p=0.36$ ) and length (mean 6 letters for both expected and unexpected, range of 4 to 9 letters for expected, 3 to 11 for unexpected), using the program *Buscar Palabras* (Davis & Perea, 2005). Out of the 52 expected nouns, 28 were feminine nouns and 24 were masculine nouns (and vice-versa for the unexpected nouns).

The mean cloze probability of expected sentences was assessed by native ( $N = 12$ ) and non-native speakers ( $N = 12$ ) of Spanish who did not take part in the experiment. They were asked to complete sentence frames with the first word that came to their mind. Sentence contexts were provided up to the word before the article; hence, participants were free to complete the sentence frame with a noun of either gender. Mean cloze probability for expected and unexpected words was .82 (SD: .18) and .02 (SD: .04), respectively, for monolinguals and .81 (SD: .19) and .04 (SD: .09) for bilinguals with no significant difference between the groups ( $p= .42$  and  $p= .12$ ). The 104 sentences were divided into two lists containing an equal number of expected and unexpected trials. Lists were balanced for number of cognates, noun gender, word frequency and length.

Each sentence appeared only once in each list either in the expected or unexpected condition.

#### Audio recording

Each sentence was recorded in its expected and unexpected version by a female native speaker of Spanish (resulting in 104 sentences, 52 in each condition). Recordings were then edited using the audio editing program Audacity (Version 2.0.2 <http://audacity.sourceforge.net>). We ensured sentences had an identical sentence context in both conditions by keeping one version of the recording (either expected or unexpected) up to the word preceding the article and then we spliced the rest of the sentence from the other condition. To reduce any bias due to cues (e.g., prosody, coarticulation) that may potentially convey information regarding the upcoming word, sentence contexts from the expected and unexpected conditions were selected equally (i.e., 52 from each condition). In each experimental sentence, the noun was completely muted for 500 ms after article offset and then the sentence resumed normally until the end (the 500 ms silence window covered the totality of the word with the exception of two words in each condition for which the silence was extended up to 550 ms). The mean duration of the (muted) nouns was similar in the expected (345 ms, SD: 85 ms) and unexpected conditions (350 ms, SD: 96 ms). The average duration of the sentence context before article onset was 2835 ms (SD: 707 ms) and that of the article was 124 ms (SD: 38 ms). The article onset was found by audio and visual inspection of the spectrogram. The total duration of a sentence was 3802 ms (SD: 770 ms). To avoid participants to guess the manipulation, we included silences (200 and 300 ms long) in the filler sentences at different positions in the sentence but never on the nouns that would then be presented as “old” words in the recognition task. Finally, to account for the interferences created by the silences, participants were told the sentences had been

extracted from a phone call; hence, all sentences were equalized to reproduce a phone-like sound quality.

### *Listening phase.*

#### Procedure:

Participants were sitting comfortably in front of a computer screen in a softly lit sound proof room. Instructions were given visually as well as verbally. The sentences were presented through two loudspeakers placed on either side of the screen using the software E-prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002a, 2002b). Participants were instructed to minimize blinking, eye movements and to stay still during the experiment. After a two sentence practice, participants were presented with one of the two lists, randomised for each participant. Each list contained 26 sentences in the “Expected” condition, 26 in the “Unexpected” condition, and 52 filler sentences. Each sentence was preceded by a white fixation cross displayed on the screen for 500 ms; the cross turned red during sentence presentation. The listening phase lasted for about 15-20 min with a break half-way through the session.

#### EEG recording and data analysis

Electrophysiological data was recorded from 31 tin electrodes attached to an elastic cap (Electrocap International) distributed according to the standard International 10-20 system (Jasper, 1958) and referenced to the nose. EEG activity was amplified with a band-pass between 0.1 Hz and 200 Hz (BrainAmps DC amplifier, Brain Products GmbH, Munich, Germany), continuously digitised at a sampling rate of 500 Hz, and re-filtered offline at 30 Hz low pass. The electrode impedance was set below 5 k $\Omega$ . Epochs ranged from -100 ms to 1000 ms after the onset of the critical article. Artifacts were automatically rejected using the procedure implemented in Brain Analyzer 2.0

(differences in values 200  $\mu\text{v}$  in 200 ms intervals, and amplitudes of +/- 100  $\mu\text{v}$ ), resulting in 8.7% and 7.1% of rejected trials in the expected and unexpected condition, respectively. Baseline correction was performed in reference to pre-stimulus activity (-100; 0 ms) and individual averages were digitally re-referenced to the average of the mastoid electrodes. The ERP data were quantified by calculating the mean voltage amplitudes. Based on previous studies, we expected a negativity peaking around 400 ms (Foucart, Martin, Moreno, & Costa, 2014; Wicha, Moreno, & Kutas, 2003), potentially preceded by an early differential (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

#### *Lexical recognition task*

To balance the number of yes-no answers, an extra 52 words were selected from the sentences heard in the listening phase (not analysed). All the words were recorded individually by the same speaker and presented auditorily using E-prime. Participant's answer triggered a fixation cross preceding the following word.