The impact of relevant environmental sounds on foreign language word learning

Melodie Bellegarda[1,2] ⓘ, Ana Belén García-Gámez[3] ⓘ and Pedro Macizo[1,2] ⓘ

[1]Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain; [2]Department of Experimental Psychology, Faculty of Psychology, University of Granada, Granada, Spain and [3]Faculty of Health Sciences International University of La Rioja, Logroño, Spain

Abstract

Past research suggests that novel word learning is facilitated by multimodal contexts, which enrich semantic representations and strengthen memory traces. We explored whether environmental sounds (e.g., a creaking door) facilitate foreign language (FL) word learning. In all, 36 Spanish-speaking natives learned 60 written Spanish–FL word pairs, each accompanied by one of three sound conditions: a congruent sound matching the word pairs' denotation, a meaningless tone or silence. Participants then completed a semantic priming and lexical decision task where reaction times and accuracy were collected. Performance was similar for congruent sound and tone conditions and, compared to silence, showed lower accuracy in the lexical decision task and a marginal benefit in the semantic task. These findings suggest that environmental sounds can influence learning, with varying effects depending on the task. Results are discussed in terms of current language learning models.

Highlights

- We examined foreign language (FL) word learning via multimodal learning strategies
- Participants learned FL nouns with an environmental sound, a tone, or silence
- Behavioral results showed relevant sounds and tones performed similarly
- Sound negatively impacted performance on the evaluation task requiring form recognition
- Findings highlight task-specific effects of auditory input

1. Introduction

The initial stages of learning a foreign language (FL) are challenging, largely due to the sheer amount of new vocabulary that must be integrated into memory to communicate even the simplest of messages. Research has explored various strategies to improve this demanding process (for reviews, see Plonsky, 2011 and Rasouli & Jafari, 2016), such as using picture–word associations (Bates & Son, 2020) and semantic maps (Hulstijn et al., 1996), looking up definitions (Leach & Samuel, 2007), relying on sentence context (Batterink & Neville, 2011) and using corporal and facial gestures (García-Gámez & Macizo, 2019; Sweller et al., 2020). These studies underscore that the context in which word encoding occurs significantly influences the strength of connections between new words and their meanings. Providing meaningful sensory input during learning by combining different learning modalities (visual, corporal and/or auditory) can create richer cognitive representations of new vocabulary (Ellis, 2019; Li & Deng, 2023). In fact, a multimodal learning environment has been suggested to be optimal because it mirrors the integrated nature of real-world experiences (Shams & Seitz, 2008). This idea aligns with the Dual Coding Theory (Paivio, 1969), which posits that learning is enhanced when verbal and nonverbal sensory inputs are combined and create stronger memory traces. According to this theory, our minds operate with two distinct systems: one for verbal information and another for nonverbal information. By encoding a concept through both systems, it becomes more likely to be remembered because it leaves two separate memory traces. While Paivio primarily emphasized the role of imagery, this theory can be extended to other sensory channels, including auditory inputs and, specifically, environmental sounds.

Environmental sounds, as defined by Vanderveer (1979), are sounds produced by real events that carry meaning by virtue of these causal events, are more complex than laboratory-generated tones, and do not belong to a formal communication system like speech. Sounds such as a door creaking open, pouring rain, or a helicopter flying overhead are encountered regularly in daily life. These sounds are not merely background noise but inherently carry specific, indisputable meanings that trigger mental representations of their corresponding objects or actions (e.g., a door, a downpour, a helicopter). There is evidence that this type of auditory information, like pictures or gestures, can contribute to embodied cognition (Barsalou, 2008; Caramiaux et al., 2011; Grisoni et al., 2016), which posits that cognition is grounded in the body's interactions with the

environment such that sensory input such as sounds could actively shape learning, thinking and remembering. This direct mapping between sound and meaning aligns with the embodied cognition framework, suggesting that environmental sounds can function as a form of nonverbal language that could aid in FL learning by reinforcing the semantic content of new words (Ballas & Howard, 1987).

Research on conceptual priming effects supports the claim that environmental sounds contain semantic information. Environmental sounds have been shown to prime related written words and vice versa (Orgs et al., 2006, 2007, 2008; Van Petten & Rheinfelder, 1995). This effect also extends to spoken words, with Frey et al. (2014) demonstrating evidence for the following prime/target combinations: sound/sound, word/sound and sound/word. In these experiments, the authors found the typical priming effects of enhancing word recognition speed and accuracy when primes and targets were related, regardless of whether the prime was a sound or linguistic stimuli. Moreover, neuroimaging studies including electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) further support the conclusion that there is a common recruitment of cognitive mechanisms for comprehension of linguistic material and environmental sounds as demonstrated by their partially overlapping neural networks, particularly in the left hemisphere regions implicated in semantic processing (Cummings et al., 2006; Dick et al., 2007). Additionally, patients with aphasia, who showed impairments in both linguistic and environmental sound processing, suggest a shared semantic retrieval system located in the left posterior temporal cortex and left inferior frontal circuits (Beauchamp et al., 2004; Saygin et al., 2003).

Despite its potential significance, the influence of environmental sounds on FL vocabulary learning remains underexplored. To our knowledge, to date, the only study that directly investigates the effect of sounds on FL learning comes from Kaplan-Rakowski and Loranc-Paszylk (2019). In their study, Polish university students learning English were presented with low-frequency onomatopoeic nouns (e.g., "lash"), which phonetically imitate or suggest the sound they represent (e.g., the sound of quick blows delivered by a whip). In a classroom environment, participants saw these words presented in four different sound conditions: corresponding sound effect, no audio, spoken pronunciation and a dual presentation of both spoken pronunciation with a corresponding sound effect. Each item was displayed for 15 seconds featuring the target FL word, its Polish translation, a short definition in the FL and an accompanying image representing the word. They found that the sound effect condition improved both immediate and delayed (7-day) free recall in comparison to the no audio condition, which they had included as a control condition. Moreover, no significant differences were found between the no audio and spoken pronunciation conditions nor between the no audio and dual presentation conditions. The authors theorized that the dual auditory input (spoken word and sound effect) during learning might have led to cognitive overload, resulting in lower performance. This interpretation aligns with Cognitive Load Theory (Sweller, 1994), which posits that combining multiple forms of information can exceed a learner's working memory capacity if not carefully coordinated. When overlapping or redundant information is added (a phenomenon known as the redundancy effect), learners must expend additional cognitive resources to reconcile repetitive inputs, thereby hindering rather than enhancing the learning process (Low et al., 2011). Nonetheless, these findings from Kaplan-Rakowski and Loranc-Paszylk highlight that relevant, nonredundant sound effects can function as effective learning cues by providing additional, meaningful context that reinforces word meaning. Thus, by strategically integrating relevant sounds into FL learning and leveraging their inherent semantic qualities, a multimodal learning experience could improve encoding and retrieval while striking a balance that avoids excessive cognitive load and maximizes enriched sensory input.

The Bilingual Interactive Activation developmental model (BIA-d) proposed by Grainger et al. (2010) provides a dynamic framework for understanding how adult learners gradually acquire an FL. Building on the Revised Hierarchical Model (RHM, Kroll & Stewart, 1994) and the Bilingual Interactive Activation model (BIA, Grainger & Dijkstra, 1992; BIA+, Dijkstra & van Heuven, 2002), the BIA-d model describes how learners initially depend on their first language (L1) to interpret second language (L2) words, but, with practice, start forming direct connections between L2 words and their meanings, eventually bypassing L1. Through mechanisms like Hebbian learning (the neural basis for associative learning, i.e., "neurons that fire together wire together"), these connections strengthen over time. Flexible control mechanisms (cognitive systems that regulate interference from the first or additional languages) also develop over time to manage cross-language influence (Beatty-Martínez et al., 2020; Casado et al., 2025).

Critically, the BIA-d model indicates that early in FL acquisition, learners coactivate multiple representations: L1 form, L2 form and the corresponding semantic concept. When semantically relevant multimodal input (e.g., environmental sounds) is introduced at this stage of learning, it could serve as an additional cue that directs attention to meaning. By reinforcing the conceptual node, relevant sounds could act as a catalyst for faster integration of L2 words into the semantic network in novel learners, shifting from L1 dependency to direct L2-concept links. Based on the BIA-d model, these sounds might act similarly to the process of Hebbian learning, where repeated coactivation of sound and meaning strengthens the connection between the L2 word and its corresponding semantic features, aligning with the idea that coactivation fosters more robust associative networks.

Moreover, the BIA-d model provides a framework for understanding how environmental sounds may facilitate tasks emphasizing semantic retrieval, although their role in developing orthographic representations is less clear. In the early stages of FL acquisition, balancing form-level encoding (i.e., orthography) and semantic-level integration (i.e., meanings) can be particularly challenging, and certain cues may overshadow others. If learners rely too heavily on sound-meaning associations, they risk neglecting the new FL word form. Nonetheless, once strong semantic links take hold, words typically become more resistant to forgetting and are accessed with greater ease, as shown in bilingualism research (Martin et al., 2009; Ning et al., 2020; Thierry & Wu, 2007).

These theoretical insights directly shaped the structure of our study. Drawing on the BIA-d model, we designed a learning paradigm that introduces environmental sounds during early FL exposure and assesses the effects on both semantic integration and form-level encoding. Accordingly, the evaluation phase includes one task targeting semantic retrieval and another targeting lexical form recognition.

While our approach is grounded in existing theoretical models, the empirical research on this topic remains limited. Despite growing interest in multimodal strategies for FL learning, only one prior study has examined the effects of relevant environmental sounds on FL vocabulary acquisition – and that study focused exclusively on onomatopoeic words (Kaplan-Rakowski & Loranc-Paszylk, 2019). As such, it remains unclear whether semantically rich, everyday sounds can enhance early FL vocabulary learning. Addressing this gap, the present study investigates whether environmental sounds

can support the formation of semantic and lexical connections in adult FL learners.

1.1. The current study

Our study focuses on whether environmental sounds originating from everyday real events can facilitate adult FL vocabulary learning when these sounds correspond to the meaning of target words. Unlike Kaplan-Rakowski and Loranc-Paszylk (2019), who focused on onomatopoeic words and their associated sound effects (words typically acquired at more advanced stages), we selected high-frequency vocabulary commonly encountered during the early stages of language learning and paired it with corresponding environmental sounds. In our study, participants were visually presented with pairs of Spanish and FL words, accompanied by three sound conditions: congruent sound, neutral sound and silence. For the congruent sound condition, the environmental sound (e.g., the sound of thunder) semantically matched the known word (e.g., "thunder"). The neutral condition consisted of monochromatic single-frequency tones that provided no relevant conceptual information to the task. This neutral, nonsemantic sound condition, absent in Kaplan-Rakowski and Loranc-Paszylk (2019), was introduced to compare two conditions with concurrent auditory and visual stimuli, helping us disentangle the effects of semantic information from those due to the mere presence of auditory input (Cummings et al., 2006). Inclusion of this condition was important to account for the inherent alerting effect of auditory stimuli (Heikkilä et al., 2015). The silence condition served as a baseline, capturing performance due to pure visual learning without any auditory input.

We deliberately chose not to include semantically incongruent sound stimuli. Although such a condition could maximize the opportunities to identify differences in learning, our primary interest lies in understanding what conditions favor language learning, not in examining learning interference effects. Therefore, we selected the condition that we hypothesized would most support language acquisition by enhancing semantic integration. In addition, incongruent learning situations are less representative of typical second language learning environments, where learners rarely encounter mismatched auditory and visual information during the acquisition of new words (e.g., both naturalistic and formal learning contexts).

To assess the impact of environmental sounds on word learning, we used two well-established evaluation tasks: a semantic priming task and a lexical decision task. In the semantic priming task, participants categorized FL words previously learned preceded by semantically related or unrelated Spanish cues. We expected that FL words learned with congruent sounds would be processed faster when preceded by related cues, compared to the silence and neutral conditions (Collins & Loftus, 1975; McNamara, 2005). These results may reflect more of the semantically mediated processing of words learned under the congruent sound condition. In the lexical decision task, participants discriminated between learned words and new FL words that were not part of the previous learning material. We hypothesized that words learned in the congruent sound condition would be recognized more readily than words learned in the other conditions, reflecting enhanced lexical retrieval and better organization of the L2 mental lexicon. Although semantic processing is not required in this task, previous studies have found that semantically mediated learning methods not only promote the establishment of semantic connections but also aid in the consolidation of lexical links (García-Gámez & Macizo, 2019; García-Gámez et al., 2022).

With regard to the other two sound conditions, we predicted that participants would perform better in both evaluation tasks for words learned in the neutral condition compared to the silence condition. The tone sound of the neutral condition, despite lacking semantic information, is expected to create an alerting effect that will enhance attention and engagement during learning. Auditory stimulation, even when nonspecific, has been shown to increase arousal and thus could serve as an attentional scaffold during the learning (Banbury et al., 2001; Näätänen, 1990).

Moreover, there is substantial evidence that language experience (LEX) plays a crucial role in language acquisition (Bonnet & Siemund, 2018; Kaiser et al., 2015). Language experience, including bilingualism and multilingualism, has been shown to enhance cognitive flexibility and metalinguistic awareness, with multilingual individuals often outperforming monolinguals specifically in language-learning tasks (Cenoz, 2013; Hirosh & Degani, 2018; Kaushanskaya & Marian, 2009). Therefore, it seems relevant to take into account LEX when studying FL acquisition. We addressed this factor by examining whether higher language experience scores, as assessed by a modified version of the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007), were associated with better learning outcomes.

We also aim to account for variability due to participants' socioeconomic status (SES). SES is closely linked to factors such as parental education, income and access to resources, all of which collectively influence cognitive development and learning opportunities, including FL acquisition. While the negative impact of lower SES on brain development and language learning is well documented in children (for a review, see Abo Hamza et al., 2024), research indicates that cognitive disparities between high- and low-SES children can persist into adulthood (Dong, 2024). To control for this potential confounding factor, we assessed SES using the MacArthur Scale of Subjective Social Status (Adler et al., 2000).

Considering these intervening variables, we included LEX and SES in an exploratory analysis, as they may influence participant's ability to learn a new language. Prior research suggests that both variables may positively correlate with language learning, with higher LEX and higher SES being associated with better learning performance.

2. Methods

2.1. Participants

Thirty-six (19 females) Spanish-speaking natives (M_{age} = 22.58, SD_{age} = 3.58) from the University of Granada participated in the study in exchange for course credit. All reported not having any visual, hearing, neurological or language impairments. Informed consent was obtained from the participants before they began the experiment. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1964 and its later amendments.

G*Power version 3.1.9.6 software (Faul et al., 2009) was used to calculate the sample size needed to capture the effects evaluated in this experiment. It was calculated that for a repeated measures analysis of variance (ANOVA) with one within-subject factor (sound condition with 3 levels), a total sample size of N = 36 was needed to achieve 95% statistical power with α = .05 and a medium effect size of 0.275 (η^2_p = .07).

Demographic information was collected before starting the experiment through two questionnaires. The participants indicated

SES through the MacArthur's Scale of Subjective Social Status (Adler et al., 2000), where they placed themselves in terms of income, education and occupation on a scale ranging from 1 to 10. The participants subjectively rated their SES on a scale of 1–10 as $M = 5.75$, $SD = 1.46$ and range 4–9. LEX was assessed via a modified version of the LEAP-Q (Marian et al., 2007). Participants were asked to list the languages they knew in order of acquisition, in order of fluency and in order of percentage of time exposed to each one. They were also asked to list the percentage of cases they would choose to read a text and speak to someone in each of their languages. Finally, they were asked to list the cultures they identified with. Descriptive data of this demographic information can be seen in Table 1. From this data, we selected the number of languages spoken to represent the linguistic diversity of the participants, which we expected to affect vocabulary learning (Hirosh & Degani, 2018). The questionnaire indicated that the average number of languages spoken (language experience, LEX) was $M = 2.97$, $SD = 1.18$ and range 1–7.

2.2. Design and materials

In this study, 60 Spanish (L1)–FL word pairs were presented visually and simultaneously with or without a sound (silence, congruent sound or neutral tone). In this within-subject design, 20 words were learned under each of the sound conditions, giving a total of 60 new

Table 1. Language use by order of acquisition as well as cultural identities self-reported by participants. Standard deviations are shown in parentheses.

Language use			
Language	% Daily exposure	% Reading	% Speaking
L1	79.31 (17.52)	81.39 (23.44)	79.35 (25.05)
L2	16.65 (13.52)	15.85 (19.40)	15.46 (19.26)
L3	5.83 (8.67)	4.80 (8.20)	6.74 (12.69)
L4	2.67 (2.35)	1.44 (2.13)	0.89 (1.24)
L5	1.83 (2.36)	1.67 (2.89)	17.50 (28.17)
L6	5.00 (0.00)	2.50 (0.00)	2.50 (0.00)
L7	0.50 (0.00)	0.00 (0.00)	0.00 (0.00)

Cultural identity		
Cultural group	% Identifying	Strength of identification
Spanish (National)	91.67	6.70 (2.19)
Spanish (Regional or Local)[a]	86.11	7.46 (0.39)
Mediterranean	52.78	4.66 (1.90)
European[b]	36.11	4.65 (2.15)
Anglophone[c]	27.78	5.39 (0.52)
Latin American[d]	13.89	7.40 (0.00)

Note. Language use measures reported mean daily exposure (i.e., the percentage of time participants reported exposure to each language) and preferences for reading and speaking (i.e., the percentage of time participants reported choosing to read or speak in each language). Participants reported experience with 1–7 languages. There were 2 monolinguals, 11 bilinguals, 14 trilinguals and 9 multilinguals with experience with 4 or more languages. Cultural identity measures the percentage of participants identifying with each cultural group, along with the mean strength of identification (scale 1–10). Participants could indicate multiple cultural identities.
[a]Includes Andalusian, Canary Islander, Catalan, Galician, Manchegan, Granadan and. Sevillian.
[b]Includes German, French, Italian, Romanian and European.
[c]Includes Canadian, English and US American.
[d]Includes Venezuelan, Colombian, Mestiza, Norteña and Latino.

words. The assignment of words to each of the sound conditions was counterbalanced between participants. After the learning task, participants performed two evaluation tasks: a semantic priming task and a lexical decision task. These tasks are commonly used in novel word-learning experiments and evidence the establishment of semantic and lexical links between new FL words, L1 words and the conceptual system (Grainger et al., 2010; Lindsay & Gaskell, 2013; Mestres-Missé et al., 2007). While the semantic priming task primarily engages semantic retrieval and conceptual activation, the lexical task engages recognition processes and familiarity with the word form, and thus lexical connections (Coltheart et al., 2001; Perry et al., 2007; Stanovich, 1991).

2.2.1. Materials
Sixty Spanish nouns were selected, with half representing natural entities (e.g., "gato," which is "cat" in English) and half representing man-made entities (e.g., "alarma," which is "alarm" in English). All selected nouns were familiar words with a mean lexical frequency of 18.99 per million words ($SD = 43.96$) (SUBTLEX-ES corpus, Cuetos et al., 2011). Sixty FL words were selected from the Vimmi corpus, an artificial language based on Italian phonotactics (Macedonia et al., 2011; Macedonia & Knösche, 2011), which had previously been used for multimodal learning experiments in our laboratory (García-Gámez & Macizo, 2019, 2020, 2023; García-Gámez et al., 2024). Using an artificial language had the advantage of removing any bias or experience the participants might have toward/with the language. The selected FL words had a low number of Spanish orthographic neighbors ($M = 0.20$, $SD = 0.94$). The FL words and Spanish nouns were paired together at random (although word pairings that started with the same phonemes were avoided), creating 30 Spanish–FL natural pairs and 30 Spanish–FL man-made pairs. The pairs were grouped into three sets of 20 words. Each set contained 10 natural and 10 man-made pairs. The three sets were equated in terms of Spanish lexical–semantic variables (lexical frequency, number of syllables, orthographic length, orthographic neighbors, shared graphemes with FL pair, familiarity, imageability and concreteness) and FL lexical variables (orthographic length, orthographic neighbors and number of syllables). Orthographic neighbors, familiarity, imageability and concreteness were consulted in the EsPal database (Duchon et al, 2013). The set of materials used can be consulted in Supplementary Materials, Table 1. Statistical description of the material is detailed as well in Supplementary Materials, Table 2.

Sixty environmental sounds congruent with the 60 selected Spanish nouns were chosen from open access online databases such as Soundbible (https://www.soundbible.com), Zapsplat (https://www.zapsplat.com) and Freesound (https://www.freesound.org). The sounds came from several different semantic classes: alarms and alerts ($n = 7$, e.g., alarm clock ringing), household sounds ($n = 9$, e.g., someone pulling a curtain), sports ($n = 1$, e.g., someone bowling), tools ($n = 10$, e.g., someone hammering), transport ($n = 3$, e.g., a car starting), animal vocalizations ($n = 19$, e.g., horse whinnying), environmental phenomenon ($n = 8$, e.g., wind whistling), and insect sounds ($n = 3$, e.g., bee buzzing). To ensure our selected sounds were representative of the Spanish nouns and easily identifiable, a total of 23 participants, who did not take part in the main experiment, participated in a normative study (see García-Gámez & Macizo, 2019, for a similar method). In each trial, participants heard an auditory stimulus and saw a Spanish word on the screen. They were instructed to rate the degree of match between the meaning of the word and the sound on a 7-point Likert scale (1 = "high mismatch," 7 = "high match"). To reduce fatigue

and repetition – given that the original 60 stimuli would otherwise be presented only in congruent form – each sound stimulus was presented twice: once with a congruent word and once with an incongruent word. Congruent and incongruent sound conditions were randomly presented. In the incongruent condition, the mismatched word was always drawn from the same semantic category (e.g., oinking paired with "marrano" [hog in English] in the congruent condition and with "caballo" [horse in English] in the incongruent condition). In total, participants rated 120 stimulus pairs during the normative study. The sound–word pairs were rated significantly higher in the congruent condition ($M = 6.24, SD = 1.57$) than in the incongruent condition ($M = 1.62, SD = 1.42$), $t(22) = 33.05, p < .001$. These results confirm that participants perceived a strong match between the selected sounds and their corresponding words.

For the neutral sound condition, 20 monochromatic unique sinusoidal tones within the range of 525–1000 Hz were created using Praat software (Boersma & Weenink, 2020). Both the tones and the environmental sounds were trimmed to 4 seconds and digitized at 48 kHz with a 16-bit sampling rate and the acoustic pressure level was normalized to the same RMS level of 70 dB SPL (Dittinger et al., 2016).

2.2.2. Learning phase

The three sets of Spanish–FL pairs were associated with a sound condition (silence, congruent sound or neutral tone), and the assignment of word sets to sound conditions was counterbalanced across the participants. Participants learned the pairs of words in a blocked manner (the items within sets were not mixed), with one round of learning consisting of exposure to all three sets of words and thus, exposure to all 60 word pairs (see Figure 1). There were 20 rounds of learning in all, meaning 20 exposures to each
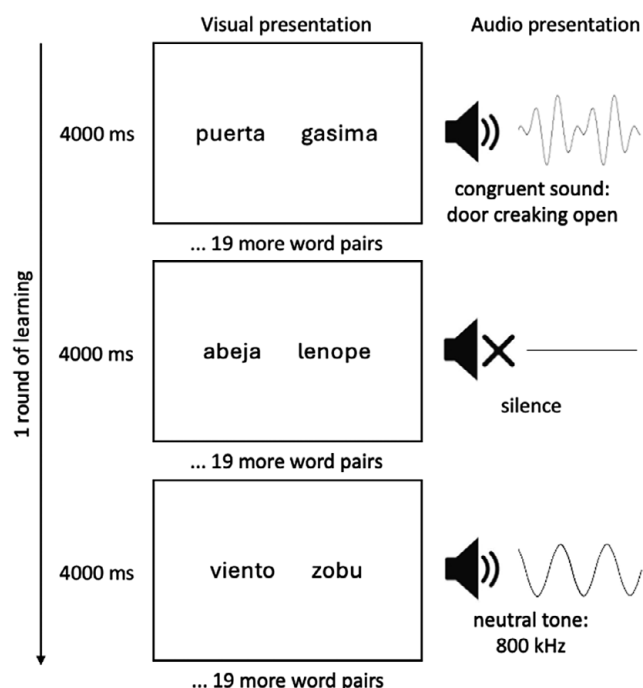
individual item, and 1200 trials in all. The presentation of the sound conditions was counterbalanced between participants. Participants were allowed a brief pause every two rounds of learning (about every 10 minutes). A trial consisted of the presentation of a centrally appearing fixation point (+) for 900 ms, followed by the central visual presentation of the word pair (L1 word–FL translation) simultaneously with the auditory presentation of the sound condition (silence, congruent sound or neutral tone) for 4000 ms and a final wait period of 700 ms (trial events based on García-Gámez & Macizo, 2020). Participants were instructed to pay close attention to the new words and focus on learning their meaning, without taking any written notes. Within each set, the word pairs were presented in a random order. The side assignment (left/right) for L1 and FL words during stimulus presentation was counterbalanced across participants.

2.2.3. Evaluation phase

Semantic priming task. A pilot study was first carried out with 48 Spanish-speaking university participants who did not participate in the main experiment in order to select the cues that would be used for the semantic priming task. In an online questionnaire, participants read 60 Spanish nouns (see Supplementary Materials, Table 1) and were instructed to type the first word that came to mind for each (e.g., for the item "cortina" ["curtain" in English], responses included, among others, "ventana" ["window" in English], "tela" ["cloth" in English] and "persiana" ["blinds" in English]). With the results from this free association task, the strength of the degree of association for each response was calculated by dividing the production frequency of an item by the total sample size. The associated words with the highest strength were selected as cue words for the semantic priming task and had a mean association frequency of 28% ($SD = 18\%$). We believe this association strength is high enough to elicit priming effects (see Anaki & Henik, 2003, Experiment 1, with 10% association frequency for weak associates and 42% association frequency for strong associates). There was no significant difference in associative strength for cues associated to the nouns from the learning task in the man-made group and those in the natural group, $t(58) = 1.48, p = .15$. We report the lexical–semantic variables of the cue words (associative strength, lexical frequency, number of syllables, orthographic length, orthographic neighbors, shared graphemes cues – FL items, familiarity, imageability and concreteness) in Supplementary Materials, Table 3.

In the semantic priming task, participants judged whether each of the 60 FL words from the learning task represented a natural item or a man-made item by pressing either a left or right key. Response keys were M and Z on the keyboard. The assignment of the M/Z key to natural/man-made items was counterbalanced across participants. Each FL word was shown after a cue (in Spanish), which was either semantically related (30 cues) or unrelated (30 cues) to the word. Two lists of words were created and counterbalanced among participants. In the first list, 30 FL words were preceded by related cues, while the other 30 were preceded by cues that were pseudo-randomly assigned, making them unrelated. In the second list, the relationships between cues and FL words were shuffled: related pairs from the first list were now unrelated, and unrelated pairs became semantically related. From the total 60 learned words, 30 FL words represented natural items (15 with related cues and 15 with unrelated cues) and 30 FL words represented man-made items (15 with related cues and 15 with unrelated cues).

A trial consisted of a fixation point presented in the middle of the screen (300 ms), followed by a cue word (semantically related or unrelated) in the same position (500 ms). After that, the FL item

Figure 1. One round of word learning in the learning phase, in which participants learned 60 Spanish-foreign language (FL) word pairs in a blocked design. Each block was associated with a different sound condition (congruent sound, silence and neutral tone), which was presented simultaneously with the word pairs. In this example, the Spanish stimuli are: "puerta" ("door" in English), "abeja" ("bee" in English) and "viento" ("wind" in English).

appeared with the choice alternatives "man-made" and "natural" displayed at the bottom right and left of the screen, corresponding to the side of the associated keypress. The FL word remained on the screen until the participants' response, and the trial concluded with a 450 ms blank screen. The stimuli appeared in random order for each participant.

Lexical decision task. Sixty new FL words were selected from the Vimmi corpus, which, like the 60 FL words presented in the learning task contained a low number of Spanish orthographic neighbors ($M = 1.17$, $SD = 3.87$). Statistical description of the material is detailed in Supplementary Materials, Table 4 for orthographical length, orthographic neighbors and number of syllables. In this task, participants were instructed to indicate with M and Z keys whether the item presented was a learned FL word or a new, never-seen-before FL word. The assignment of the keys to the yes/no responses was counterbalanced across participants. A trial consisted of a fixation point presented centrally (300 ms), followed by the item presented in the same location with the options "learned" or "not learned" displayed at the bottom right and left of the screen in accordance with the side of the associated keypress. As in the semantic priming task, the target item remained on screen until a response was given, and the trial then concluded with a blank screen (450 ms). In this way, a total of 120 trials were performed in this task, with 60 of them being part of the previously learned words and the remaining 60 words corresponding to never-seen words. The stimuli appeared in random order regardless of the condition for each participant.

2.3. Procedure

Before arriving at the laboratory, participants completed the two online questionnaires (LEAP-Q and SES), which they accessed through a link on the participant-recruiting platform. Once in the laboratory, the participants completed the learning and evaluation tasks in a single experimental session lasting about 120 minutes. During the session, they were situated in a quiet room to minimize distractions and wore a Trust Mauro USB headset to hear the binaural stimuli during the learning task (100 minutes). They then completed the semantic priming task (10 minutes) and the lexical decision task (10 minutes). The order of the evaluation tasks was counterbalanced across participants. All stimuli in this experiment were presented using OpenSesame version 3.2.8 software (Mathôt et al., 2012) on a 19-inch Captiva E1903 monitor.

2.4. Data analysis

Data was analyzed with linear mixed-effects models (for continuous data) and general linear mixed-effects models (for binomial data) using the lme4 package (Bates, Maechler, et al., 2015a) for R version 3.6 (R Core Team, 2019). The statistical models were constructed by fitting random intercepts and/or random slopes by participants and by items, following the recommendation for using parsimonious random structures to account for variability in the data, without overfitting (Bates, Kliegl, et al., 2015b). For the reaction time analyses, only correct trials were included. We eliminated univariate outliers that deviated more than ±2.5 SD from each participant's mean and log-transformed the values. Likelihood ratio tests were conducted to assess the significance of fixed effects and interactions, and estimated marginal means were computed using the emmeans package (Lenth, 2023). Tukey's correction (Tukey, 1953) was used when performing post hoc comparisons to adjust for multiple comparisons.

Semantic priming task. A generalized linear mixed-effects model was fitted for the binomial accuracy data (i.e., "correct" and "incorrect" responses), and a linear mixed-effects model was fitted for the reaction times. The fixed-effects predictors were: Sound (congruent, silence, neutral) and Cue (related, unrelated). We included the following interaction as well: Sound x Cue.

Lexical decision task. A generalized linear mixed-effects model was fitted for the binomial accuracy data (i.e., "correct" and "incorrect" responses). In order to take into account the different types of incorrect answers (misses, or responding "new" to an old stimulus, and false alarms, or responding "old" to a new stimulus), we performed a z transformation of the hit rates and the false alarm rates to obtain d' values for each participant (MacMillan & Creelman, 2005), and a linear mixed-effects model was fitted. Another linear mixed-effects model was fitted for the reaction times as well. The fixed-effects predictor was Sound (congruent, silence, neutral).

Questionnaires. As an exploratory analysis, we extended our primary model analyses by incorporating two additional predictors: SES and LEX. Based on previous literature, we thought that these variables could have an effect on learning in general. They were included as fixed-effects predictors, and the values were scaled and centered. For each task and each dependent variable (accuracy, reaction times, d'), we ran two additional models, one with SES and one with LEX. To account for multiple comparisons in these exploratory analyses, we applied the Benjamini–Hochberg procedure to control the false discovery rate across all exploratory models (Benjamini & Hochberg, 1995).

3. Results

3.1. Semantic priming task

Accuracy. Participants showed a general accuracy rate of $M = 75.79\%$, $SD = 13.68\%$, range 45% – 100%. The analysis with our model glmer ($ACC \sim sound*cue + (cue|subject) + (1|item)$) showed a main effect of Cue (χ^2) (1) = 8.67, $p < .003$ with participants having more accuracy on trials containing related cues ($M = 81.02\%$) than unrelated cues ($M = 70.56\%$). There was also a marginally significant main effect of Sound, Chi-squared test (χ^2) (2) = 5.96, $p < .051$. *Post-hoc t-*tests indicated that participants performed marginally more accurately on trials for items that had been learned with a congruent sound ($M = 77.64\%$) in comparison with those that had been learned with silence ($M = 73.06\%$) (logit estimate: 0.30, $SE = 0.13$, $p = .063$), see Figure 2. There were no significant differences in accuracy for trials where items had been learned with tone ($M = 76.67\%$) in comparison with those from the congruent (logit estimate: 0.08, $SE = 0.14$, $p = .83$) or silent conditions (logit estimate: -0.22, $SE = 0.13$, $p = .22$). No interaction was found, $p > .49$.

Reaction times. The percentage of data excluded as outliers was 3.67%. Results for our model $lmer(logRT \sim sound*cue + (1|subject) + (cue|item))$ showed no significant main effects or interactions, all ps > .51.

3.2. Lexical decision task

Accuracy. Participants showed a general accuracy rate of $M = 85.65\%$, $SD = 8.26\%$, range 68.33%–98.33%. The analysis described above with our model $glmer(ACC \sim sound + (1|subject) + (1|item))$ revealed a main effect of Sound (χ^2) (2) = 10.49, $p < .005$.

*Post-hoc t-*tests indicated that participants were more accurate on trials involving items learned in silence ($M = 83.75\%$) compared to those learned with a congruent sound ($M = 77.78\%$) (logit estimate: -0.45, $SE = 0.14$, $p = .006$). Similarly, accuracy was higher
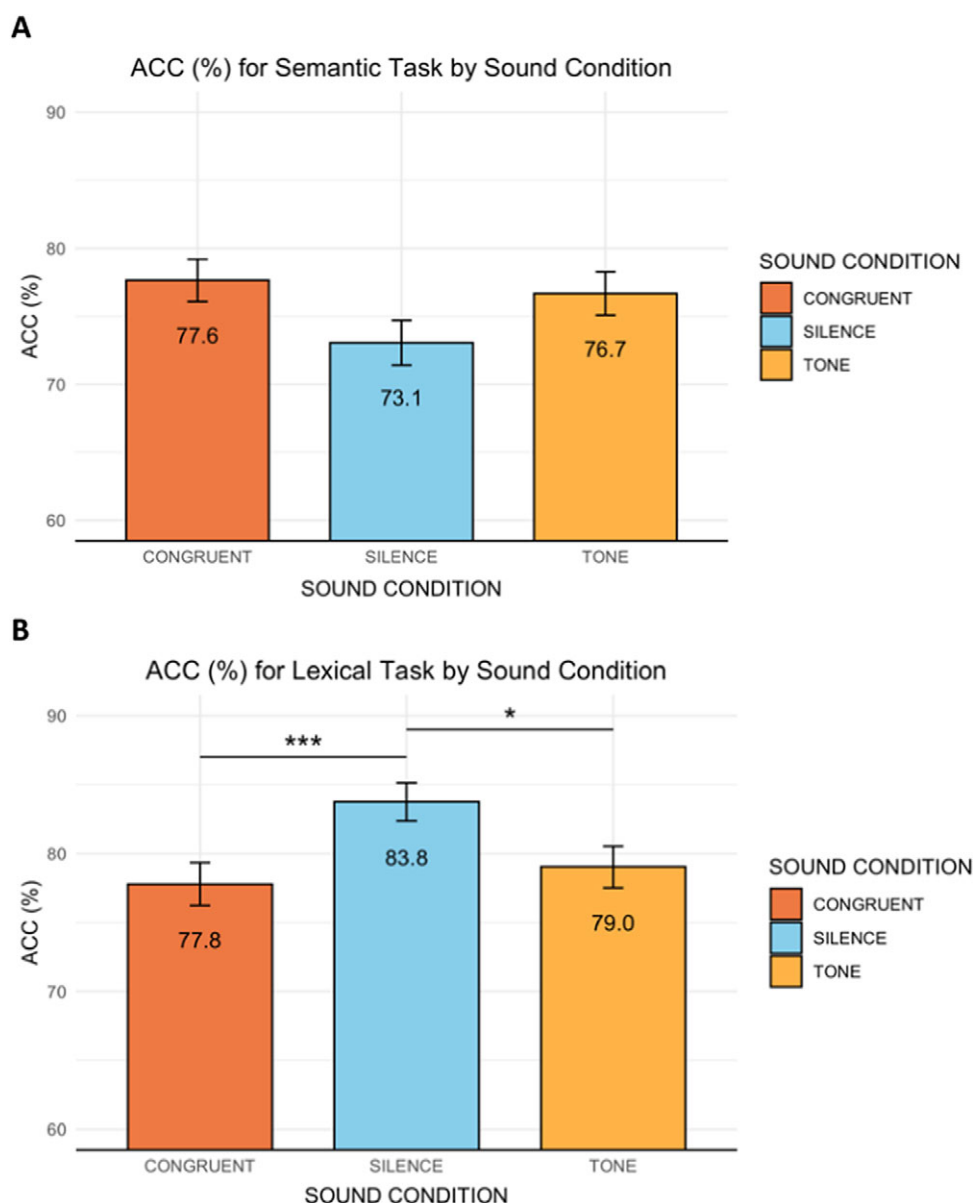
A

B

Figure 2. Recall percentage (ACC, in percentage) in (A) Semantic Task and (B) Lexical Task for the different sound conditions (congruent sound, silence, neutral tone). Standard error is shown with vertical lines, and average recall is indicated for each condition.

for items learned in silence compared to those learned with a tone ($M = 79.03\%$) (logit estimate: 0.36, $SE = 0.14$, $p = .04$), see Figure 2. There were no significant differences found for the tone and congruent sound conditions (logit estimate: -0.09, $SE = 0.14$, $p = .80$).

D prime. D′ calculations showed an average of $M = 3.84$, $SD = 3.10$, range 1.02–13.66. Our model $lmer(dprime \sim sound + (1|subject)$ showed no significant main effect, $p > .45$.

Reaction times. The percentage of data excluded as outliers was 3.62%. Results for our model $lmer(logRT \sim sound + (sound|subject) + (1|item)$ showed no significant main effect, $p > .87$.

3.3. Questionnaires

In the lexical decision task, results showed a main effect of LEX, Chi-squared test (χ^2) (1) = 3.97, $p < .046$, with a positive relationship between the number of languages learned and d', where the model

predicted a higher d' value for more languages spoken. However, this effect was no longer significant when we adjusted the p-value for multiple comparisons, $p > .46$. No other main effects for LEX were found, all ps > .48, and corrected ps > .82. No main effects for SES were found for either task, all ps > .65, and corrected ps > .82.

4. Discussion

In the present study, we investigated the impact of environmental sounds on FL vocabulary acquisition, examining how semantically meaningful sounds can enhance the learning process. We developed a learning task with three sound conditions to explore how sounds that provide semantically rich information (environmental sounds congruent with the linguistic stimuli) can help integrate information differently from sounds that do not provide any semantic information (neutral tone condition), or learning without the auditory modality (silent condition). By including both a

semantic priming and a lexical decision task to evaluate learning, we were able to capture how environmental sounds influence different levels of processing – semantic/conceptual learning and form-based lexical recognition. Our findings reveal that the effect of sound on vocabulary learning varied based on the type of assessment task used, highlighting the complex interplay between auditory input and cognitive processing in language acquisition. Moreover, our results suggested that learning with sound, independent of its relation to the words' meaning, appeared to engage attentional mechanisms in a comparable way. Our results also shed light on some evidence for language experience affecting new vocabulary learning. The implications of these findings are discussed below in relation to previous literature.

To the best of our knowledge, this is the first study to employ an in-lab paradigm testing the influence of semantically informative environmental sounds on FL vocabulary learning. Previous literature has established the general utility of environmental sounds for enhancing learning (Heikkilä et al., 2015) and, specifically, for FL acquisition (Kaplan-Rakowski & Loranc-Paszylk, 2019). However, our results indicate that while environmental sounds can positively influence learning outcomes, the extent of benefit is task-dependent. For example, learning with congruent sounds yielded a marginal accuracy improvement in the semantic priming task compared to silent learning, but it resulted in a decrease in learning in the lexical decision task. This contrast suggests that environmental sounds may enhance semantic processing but could introduce cognitive load detrimental to tasks requiring form-based recognition.

The contrasting results in our study may be explained by the different cognitive demands and processing requirements of our two evaluation tasks. The semantic priming task required participants to engage in deeper lexical–semantic processing by activating not only the word form but also its associated meaning and related concepts. Such a task is commonly used to assess the automatic activation of semantic links, relying on faster retrieval of related words compared to unrelated ones (Kumar, 2021; McNamara, 2005). In this task, congruent sounds provided a marginal benefit, likely because they facilitated the establishment of meaningful connections between the new words and their corresponding sounds. While the extra sensory information during encoding was disruptive for the lexical decision task, this cognitive load could be considered germane (Paas et al., 2004) for the semantic priming task, as it supported the processing of semantic information during later retrieval. This aligns with research showing that environmental sounds can enhance conceptual understanding by providing contextual cues that strengthen semantic links between words and their real-world referents (Heikkilä et al., 2015; Kaplan-Rakowski & Loranc-Paszylk, 2019). Although only marginal improvements were observed in the semantic priming task, the presence of congruent sounds may suggest a trend toward deeper semantic integration. In contrast, these sounds were less useful for tasks requiring simple form-based recognition, such as the lexical decision task.

In the lexical decision task, participants were asked to determine whether a written item had been previously encountered. Our findings indicate that participants who learned words with auditory input (congruent sound or neutral tone) performed worse than those in the silence condition. This suggests that the introduction of sounds may have imposed extraneous cognitive load (Sweller et al., 2011), which could interfere with the ability to focus on orthographic forms during recognition-based tasks (Paas et al., 2004). Research in dual-processing tasks, such as listening while reading, has similarly shown that simultaneous engagement of auditory and

visual modalities can disrupt performance by increasing cognitive demands in situations requiring simple form recognition (Baddeley, 1992; Zhang et al., 2018).

Grainger's BIA-d model (Grainger et al., 2010) provides a useful theoretical framework for understanding why sound might facilitate semantic processing but interfere with tasks that require orthographic recognition. According to the model, in the initial stages of FL acquisition, an FL word is coactivated with its L1 word form and the corresponding semantic representation. The environmental sounds presented during encoding could serve to enhance this semantic access and in the process minimize the importance of the L2 word form. This is consistent with findings from studies on bilingualism, which have shown that the establishment of durable semantic connections is more pronounced in relatively fluent bilinguals, making these connections more resistant to forgetting (Martin et al., 2009; Ning et al., 2020; Thierry & Wu, 2007). From an attentional perspective, introducing *any* auditory cue – semantic or otherwise – can heighten alertness and direct processing away from orthographic form, favoring conceptual connections. In our study, the sounds provided during learning may have functioned similarly to how bilinguals rely on multiple cues to establish robust lexical–semantic links, enhancing retention in semantic tasks but leading to interference in tasks requiring the recall of orthographic form.

Regarding the neutral condition, performance was comparable for words learned with both types of sounds, whether semantically related (congruent sound) or unrelated (neutral tone). This suggests similar engagement of attentional mechanisms. However, the gap in performance compared to the silent condition was consistently larger for congruent sounds than for tones. In the semantic priming task, both congruent sound and tone conditions outperformed silence, although only the effect for congruent sound approached significance. This marginal trend may reflect a modest benefit of congruent sounds for semantic processing, whereas neutral tones did not show a similar pattern. Conversely, in the lexical decision task, both sound conditions performed significantly worse than silence, but the impact was again more pronounced for congruent sounds, reflecting subtle differences in processing demands between tasks. This contrast highlights the potentially disruptive role of auditory input (semantic or not) in tasks requiring orthographic form recognition. These findings also hint at an underlying attentional mechanism: even a nonsemantic sound could heighten alertness (Banbury et al., 2001; Näätänen, 1990), during encoding, functioning as an "attentional scaffold." Overall, this interplay among attention, semantic access and cognitive load provides a nuanced view of how auditory cues can both support and impede FL learning.

It was also within the scope of this study to explore how language experience can affect FL vocabulary learning. A multilingual background has been shown to provide various benefits for language acquisition, such as a larger phonological network and familiarity with language-learning strategies (Hirosh & Degani, 2018). In the lexical task, we observed a positive relationship between language experience and sensitivity to stimuli, suggesting that participants who spoke more languages were better at recognizing previously learned words and rejecting unfamiliar words. However, this effect did not survive multiple comparisons and therefore should be interpreted with caution. While not statistically robust, the pattern may reflect underlying individual differences related to lexical access and managing ongoing lexical competition (Kroll & Bialystok, 2013) or familiarity with language-learning strategies among multilingual participants (Kaushanskaya & Marian, 2009;

Kaushanskaya, 2012). Curiously, this relationship between language experience and discrimination ability was only observed in the lexical decision task and not in the semantic priming task. As explained in the previous paragraphs, the processing requirements in these tasks differ, with the lexical decision task relying heavily on recognition memory and the retrieval of previously learned vocabulary. Participants with a more diverse language experience may benefit from their broader lexical network, enabling quicker access to familiar words and better discrimination of novel ones. Future research with more targeted designs is needed to clarify whether language experience reliably contributes to improved vocabulary discrimination in similar tasks.

With regard to SES, we found no evidence that this variable influenced learning outcomes in our study. This may be due to the relatively homogeneous nature of our participants, who were primarily university students and self-identified as largely middle class. The limited variance in SES may not have allowed for it to exert a detectable influence on language-learning performance.

4.1. Broader implications and educational relevance

The present findings carry implications for FL instruction and curriculum design. In modern classrooms, multimodal input is increasingly used, not only due to documented cognitive benefits but also because of greater access to technology and the technological disposition of today's learners. Such input can also accommodate diverse learning profiles, making it especially valuable in inclusive educational settings. Our results suggest that while environmental sounds may not always yield strong semantic gains, they can still enhance learning by drawing attention to the material and supporting more efficient encoding. From a pedagogical standpoint, strategically incorporating relevant auditory input during early vocabulary instruction may improve retention, particularly in meaning-based tasks. However, educators should be cautious when applying this approach in tasks that require precise form recognition (e.g., spelling or written recall), where additional auditory input may impose extraneous cognitive load. These findings underscore the importance of aligning the type and timing of auditory input with the specific demands of the learning task.

4.2. Potential limitations and future directions

A potential limitation is that the positive effect observed in the congruent sound condition in the semantic priming task was only marginal, suggesting that further research is needed to fully understand the role of relevant sounds in fostering semantic connections and supporting vocabulary acquisition. One possibility is that the task may not have been challenging enough to capture the benefits of deeper processing. Additionally, the lack of significant differences between the congruent and neutral tone conditions suggests that both types of auditory input may have enhanced attention during learning (Murphy et al., 2017). To better understand the mechanisms underlying these effects, it would be necessary to use neuroimaging techniques in order to disentangle the impact of the two types of sound conditions on lexical and semantic access.

Future studies could also consider introducing continuous evaluations during the learning phase to better understand how lexical–semantic representations evolve over time, from initial exposure to more advanced stages of acquisition. In this study, the assessment of novel word learning was conducted only after the learning phase had concluded, and therefore did not capture the dynamic process by which participants' representations of new items evolve over time.

Monitoring participants throughout the learning phase (see García-Gámez & Macizo, 2023) would yield more nuanced data on the learning trajectory and could also reveal potential variations in how emerging representations are influenced by different sound conditions.

Similarly, considering the role of offline consolidation in vocabulary integration could provide insights about the mechanisms that underpin effective language learning. Evidence suggests that offline consolidation, which includes sleep, plays a crucial role in transforming new word representations from episodic into stable lexical–semantic forms by facilitating the integration of new vocabulary into long-term memory systems (Bakker et al., 2015). Future studies should implement multiday paradigms to explore how consolidation contributes to the long-term retention and integration of new vocabulary.

5. Conclusions

Taken together, our results provide a basis for evaluating the use of environmental sounds in FL learning. Our findings suggest that while environmental sounds may offer benefits in tasks that rely on deeper semantic processing, such as semantic priming, they can also impose extraneous cognitive load that hinders performance in tasks focused on form recognition, such as lexical decision. These results highlight the importance of task specificity when evaluating the impact of auditory input on learning and call attention to the nuanced interaction between cognitive load and sensory input in language processing. Future research should consider incorporating neuroimaging techniques to further investigate the distinct impacts of relevant versus nonrelevant sounds on learning outcomes to better understand their roles in vocabulary acquisition.

References

Abo Hamza, E., **Tindle, R.**, **Pawlak, S.**, **Bedewy, D.**, & **Moustafa, A. A.** (2024). The impact of poverty and socioeconomic status on brain, behaviour, and development: A unified framework. *Reviews in the Neurosciences*, **35**(6), 597–617. https://doi.org/10.1515/revneuro-2023-0163.

Adler, N. E., **Epel, E. S.**, **Castellazzo, G.**, & **Ickovics, J. R.** (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, **19**(6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586.

Anaki, D., & **Henik, A.** (2003). Is there a "strength effect" in automatic semantic priming? *Memory & Cognition*, **31**(2), 262–272. https://doi.org/10.3758/BF03194385.

Baddeley, A. (1992). Working memory. Science, **255**(5044), 556–559. https://doi.org/10.1126/science.1736359

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia*, **79**, 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020.

Ballas, J. A., & Howard, J. H. (1987). Interpreting the language of environmental sounds. *Environment and Behavior*, **19**(1), 91–114. https://doi.org/10.1177/0013916587191005.

Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors*, **43**(1), 12–29. https://doi.org/10.1518/001872001775992462.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, **59**, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Bates, J., & Son, J. B. (2020). English vocabulary learning with simplified pictures. *TESL-EJ: The Electronic Journal for English as a Second Language*, **24**(3) 1–20. http://www.tesl-ej.org/wordpress/issues/volume24/ej95/ej95a12

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015a). *lme4: Linear mixed-effects models using Eigen and S4.* R package version 1.1–10. http://CRAN.R-project.org/package=lme4.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015b). *Parsimonious mixed models. Preprint, arXiv:1506.04967*.

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience*, **23**(11), 3181–3196. https://doi.org/10.1162/jocn_a_00013.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, **41**(5), 809–823. https://doi.org/10.1016/s0896-6273(04)00070-4.

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **46**(6), 1022. https://doi.org/10.1037/xlm0000770.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(1), 289–300. http://www.jstor.org/stable/2346101

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* [computer program]. Version 6.1.33. http://www.praat.org/

Bonnet, A., & Siemund, P. (2018). Multilingualism and foreign language education. A synthesis of linguistic and educational findings. In A. Bonnet & P. Siemund (Eds.), *Foreign language education in multilingual classrooms* (Vol. 7, pp. 1–29). John Benjamins Publishing Company.

Caramiaux, B., Susini, P., Bianco, T., Bevilacqua, F., Houix, O., Schnell, N., & Misdariis, N. (2011). Gestural embodiment of environmental sounds: An experimental study. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 144–148. https://doi.org/10.5281/zenodo.1177979.

Casado, A., Walther, J., Wolna, A., Szewczyk, J., Sorace, A., & Wodniecka, Z. (2025). Investigation of long- and short-term adaptations of the bilingual language system to different language environments: Evidence from the ERPs. *Journal of Neurolinguistics*, **74**, 101242. https://doi.org/10.1016/j.jneuroling.2024.101242.

Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching*, **46**(1), 71–86. https://doi.org/10.1017/S0261444811000218.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**(6), 407. https://doi.org/10.1037/0033-295X.82.6.407.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**(1), 204–256. https://doi.org/10.1037/0033-295x.108.1.204.

Cuetos, F., González-Nosti, M., Barbón, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica*, **32**, 133–143.

Cummings, A., Ceponiene, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, **1115**(1), 92–107. https://doi.org/10.1016/j.brainres.2006.07.050.

Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Bentrovato, S., D'Amico, S., Wilson, S., Bates, E., & Pizzamiglio, L. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: Evidence from functional magnetic resonance imaging and lesion data. *Journal of Cognitive Neuroscience*, **19**, 799–816. https://doi.org/10.1162/jocn.2007.19.5.799.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, **5**(3), 175–197. https://doi.org/10.1017/S1366728902003012.

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: From faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, **28**(10), 1584–1602. https://doi.org/10.1162/jocn_a_00997.

Dong, L. (2024). Examining the relationship between socioeconomic status, self-regulated learning strategies, and writing proficiency in English as a second language learning context. *Journal of Educational Psychology*, **116**(5), 686–702. https://doi.org/10.1037/edu0000834.

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

Ellis, N. C. (2019). Essentials of a theory of language cognition. *The Modern Language Journal*, **103**(S1), 39–60. https://doi.org/10.1111/modl.12532.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, **41**, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Frey, A., Aramaki, M., & Besson, M. (2014). Conceptual priming for realistic auditory scenes and for auditory words. *Brain and Cognition*, **84**, 141–152. https://doi.org/10.1016/j.bandc.2013.11.013.

García-Gámez, A. B., & Macizo, P. (2019). Learning nouns and verbs in a foreign language: The role of gestures. *Applied PsychoLinguistics*, **40**(2), 473–507. https://doi.org/10.1017/S0142716418000656.

García-Gámez, A. B., & Macizo, P. (2020). The way in which foreign words are learned determines their use in isolation and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **46**(2), 364–379. https://doi.org/10.1037/xlm0000721.

García-Gámez, A. B., & Macizo, P. (2022). Lexical and semantic training to acquire words in a foreign language: An electrophysiological study. *Bilingualism: Language and Cognition*, **25**(5), 768–785. https://doi.org/10.1017/S1366728921000456

García-Gámez, A. B., & Macizo, P. (2023). Gestures as scaffolding to learn vocabulary in a foreign language. *Brain Sciences*, **13**(12), 1712. https://doi.org/10.3390/brainsci13121712.

García-Gámez, A. B., Cervilla, Ó., Casado, A., & Macizo, P. (2024). Seeing or acting? The effect of performing gestures on foreign language vocabulary learning. *Language Teaching Research*, **28**(3), 1055–1086. https://doi.org/10.1177/13621688211024364.

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. J. Harris (Ed.), *Advances in psychology* (Vol. 83, pp. 207–220). North-Holland. https://doi.org/10.1016/S0166-4115(08)61496-X.

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (Vol. 52, pp. 267–283). John Benjamins Publishing.

Grisoni, L., Dreyer, F. R., & Pulvermüller, F. (2016). Somatotopic semantic priming and prediction in the motor system. *Cerebral Cortex*, **26**(5), 2353–2366. https://doi.org/10.1093/cercor/bhw026.

Heikkilä, J., Alho, K., Hyvönen, H., & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Experimental Psychology*, **62**(2), 123–130. https://doi.org/10.1027/1618-3169/a000279.

Hirosh, Z., & Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin Review*, **25**, 892–916. https://doi.org/10.3758/s13423-017-1315-7.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, **80**(3), 327–339. https://doi.org/10.1111/j.1540-4781.1996.tb01614.x.

Kaiser, A., Eppenberger, L. S., Smieskova, R., Borgwardt, S., Kuenzli, E., Radue, E. W., Nitsch, C., & Bendfeldt, K. (2015). Age of second language acquisition in multilinguals has an impact on gray matter volume in language-associated brain areas. *Frontiers in Psychology*, **6**, 638. https://doi.org/10.3389/fpsyg.2015.00638.

Kaplan-Rakowski, R., & Loranc-Paszylk, B. (2019). The impact of verbal and nonverbal auditory resources on explicit foreign language vocabulary learning. *System*, **85**, 102114. https://doi.org/10.1016/j.system.2019.102114.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, **16**, 705–710. https://doi.org/10.3758/PBR.16.4.705.

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition*, **15**(3), 470–489. https://doi.org/10.1017/S1366728911000472

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, **25**(5). https://doi.org/10.1080/20445911.2013.799170.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, **33**(2), 149–174. https://doi.org/10.1006/jmla.1994.1008.

Kumar A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic bulletin & review*, **28**(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

Leach, L., & Samuel, A. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Journal of Cognitive Psychology*, **55**(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001.

Lenth, R. (2023). emmeans: Estimated marginal means, aka Least-squares means. R package version 1.8.5. https://CRAN.R-project.org/package=emmeans.

Li, J., & Deng, S. W. (2023). Facilitation and interference effects of the multisensory context on learning: A systematic review and meta-analysis. *Psychological Research*, **87**(5), 1334–1352. https://doi.org/10.1007/s00426-022-01733-4.

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, **39**(2), 608–622. https://doi.org/10.1037/a0029243.

Low, R., Jin, P., & Sweller, J. (2011). Cognitive load theory, attentional processes and optimized learning outcomes in a digital environment. In C. Roda (Ed.), *Human attention in digital environments* (pp. 93–113). Cambridge University. https://doi.org/10.1017/CBO9780511974519.004.

Macedonia, M., & Knösche, T. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain and Education*, **5**, 196–211. https://doi.org/10.1111/j.1751-28×.2011.01129.x.

Macedonia, M., Müller, K., & Friederici, A. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, **32**, 982–998. https://doi.org/10.1002/hbm.21084.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates Publishers.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, **44**(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, **50**(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067).

Martin, C. D., Dering, B., Thomas, E. M., & Thierry, G. (2009). Brain potentials reveal semantic priming in both the 'active' and the 'non-attended' language of early bilinguals. *NeuroImage*, **47**(1), 326–333. https://doi.org/10.1016/j.neuroimage.2009.04.025.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press. https://doi.org/10.4324/9780203338001.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, **17**(8), 1858–1866. https://doi.org/10.1093/cercor/bhl094.

Murphy, S., Spence, C., & Dalton, P. (2017). Auditory perceptual load: A review. *Hearing Research*, **352**, 40–48. https://doi.org/10.1016/j.heares.2017.02.005.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, **13**(02), 201–233. https://doi.org/10.1017/s0140525x00078407.

Ning, S., Hayakawa, S., Bartolotti, J., & Marian, V. (2020). On language and thought: Bilingual experience influences semantic associations. *Journal of Neurolinguistics*, **56**, 100932. https://doi.org/10.1016/j.jneuroling.2020.100932.

Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition*, **62**, 267–272. https://doi.org/10.1016/j.bandc.2006.05.003.

Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2007). Is conceptual priming for environmental sounds obligatory? *International Journal of Psychophysiology*, **65**, 162–166. https://doi.org/10.1016/j.ijpsycho.2007.03.003.

Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2008). N400-effects to task irrelevant environmental sounds: Further evidence for obligatory conceptual processing. *Neuroscience Letters*, **436**, 133–137. https://doi.org/10.1016/j.neulet.2008.03.005.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, **32**(1), 1–8. https://doi.org/10.1023/B:TRUC.0000021806.17516.d0.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, **61**, 179–211. http://doi.org/10.1037/h0027272.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, **114**(2), 273. https://doi.org/10.1037/0033-295X.114.2.273.

Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, **61**(4), 993–1038. https://doi.org/10.1111/j.1467-9922.2011.00663.x.

R Core Team (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Rasouli, F., & Jafari, K. (2016). A deeper understanding of L2 vocabulary learning and teaching: A review study. *International Journal of Language and Linguistics*, **4**(1), 40–46. https://doi.org/10.11648/j.ijll.20160401.16.

Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds: Evidence from aphasia. *Brain: A Journal of Neurology*, **126**(4), 928–945. https://doi.org/10.1093/brain/awg082.

Shams, L., & Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, **12**(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006.

Stanovich, K. E. (1991). Word recognition: Changing perspectives. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. **2**, pp. 418–452). Lawrence Erlbaum Associates, Inc.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, **4**(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Sweller, N., Shinooka-Phelan, A., & Austin, E. (2020). The effects of observing and producing gestures on Japanese word learning. *Acta Psychologica*, **207**, 103079. https://doi.org/10.1016/j.actpsy.2020.103079.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, **104**(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104.

Tukey, J. W. (1953). The problem of multiple comparisons. *Technometrics*, **5**(2), 99–114. https://doi.org/10.1080/00401706.1953.10405634.

Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, **33**(4), 485–508. https://doi.org/10.1016/0028-3932(94)00133-A.

Vanderveer, N. J. (1979). *Ecological acoustics: Human perception of environmental sounds* (Order No. 8004002). [Doctoral Dissertation, Cornell University]. ProQuest Dissertations & Theses Global.

Zhang, H., Miller, K., Cleveland, R., & Cortina, K. (2018). How listening to music affects reading: Evidence from eye tracking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **44**(11), 1778–1791. https://doi.org/10.1037/xlm0000544.