

Stormy words: Bilingual emotional lexical access under virtual weather conditions

Francisco Rocabado  and Jon Andoni Duñabeitia 

Nebrija Research Center in Cognition, Universidad Nebrija, Spain

Research Article

Cite this article: Rocabado, F. and Duñabeitia, J.A. (2026). Stormy words: Bilingual emotional lexical access under virtual weather conditions. *Bilingualism: Language and Cognition* 1–9. <https://doi.org/10.1017/S1366728926101382>

Received: 7 October 2025

Revised: 10 February 2026

Accepted: 17 March 2026

Keywords:

Virtual Reality; Disfluency; Second Language Processing; Bilingualism; Emotions; Bilingual Lexical Access

Corresponding author:

Francisco Rocabado;

Email: jrocabado@nebrija.es

  This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

Language, emotion, and environment jointly shape how words are processed in real life. This study tested how valence and simulated weather influence bilingual lexical access in virtual reality (VR). Forty Spanish–English bilinguals completed a language-decision task with negative high-arousal and neutral low-arousal words under sunny and rainy conditions. Accuracy was high, with no reliable effects. Reaction times were faster for negative than for neutral words and slower under rain than sun, with no significant language effect. A Weather by Trial Order interaction reflected a practice-related speeding under sun under sunny weather. Valence and weather exerted additive influences, and weather did not modulate language or valence effects. These findings suggest that realistic perceptual load imposes general costs without altering emotional or language-related processing. The study underscores VR's potential to integrate ecological validity into psycholinguistic paradigms, revealing how intrinsic and extrinsic factors jointly constrain bilingual emotional word processing.

Highlights

- VR shows additive effects of Valence and Weather on speeded language decisions, with no reliable Language effect
- Negative high-arousal words speed lexical decisions in L1 and L2
- Rainy VR induces an overall slowing without modulating valence or language effects
- The weather effect becomes more pronounced over time, practice-related speeding emerges in sun but is constrained under rain
- Supports VR as a tool for ecologically realistic research in bilingual psycholinguistics

1. Introduction

Language and emotion are deeply interconnected. Language conveys affective content while also shaping how emotions are perceived, experienced, and regulated (Ferré et al., 2025). Psycholinguistic research has consistently shown that the degree to which a word is experienced in terms of positive or negative polarity (i.e., the word's valence) directly influences its recognition (Kousta et al., 2009). Across lexical decision and emotional categorization tasks, positive words typically elicit faster responses than neutral and negative words (Crossfield & Damian, 2021; Kauschke et al., 2019; Scott et al., 2014). Evidence regarding the processing of negative words is mixed. Early accounts of 'automatic vigilance' proposed that negative stimuli capture attention and slow responses, a view supported by large-scale analyses showing slower recognition of negative relative to positive words (Kuperman et al., 2014). However, other studies controlling for lexical confounds suggest that both negative and positive words can be recognized faster than neutral words, consistent with the idea that motivationally relevant stimuli enjoy a processing advantage (Kousta et al., 2009; see Hinojosa et al., 2019 for a review). Critically, these effects are not only driven by valence but also by arousal. Interestingly, high-arousal negative words have been shown to facilitate recognition relative to neutral ones, whereas low-arousal negative words tend to slow down responses (Citron et al., 2014; Hofmann et al., 2009; Larsen et al., 2008). This variability highlights that the impact of negativity on word recognition is not uniform and may depend on task demands, stimulus properties, and individual differences. Importantly, large-scale evidence confirms robust valence effects while also showing that their size varies across individuals (Haro et al., 2024). Complementarily, computational work suggests that positive words reach stable semantic representations earlier and more reliably during development than negative words, which may underpin their processing advantage (Martínez-Huertas et al., 2025). Together, these strands position valence and its interaction with arousal as a central determinant of lexical access.

In bilingual contexts, emotional effects are further modulated by language. The emotional resonance of words is typically attenuated in a second language (L2) compared to the native language (L1). This phenomenon, often termed the foreign-language effect, has been extensively reviewed (see Aguilar et al., 2024 for an extensive review on the topic) and demonstrated in experimental studies across fields (Costa et al., 2014; Geipel et al., 2015; Hayakawa et al., 2017).

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Emotional words in L2 often evoke weaker neural and physiological responses than in L1. For instance, Chen et al. (2015) showed that emotional words elicited reduced ERP and fMRI responses in L2 compared to L1, indicating attenuated neural engagement. Similarly, Iacozza et al. (2017) reported reduced pupil dilation when participants read emotional sentences in L2 relative to L1. Such asymmetry is often attributed to differences in acquisition context and frequency of emotional use (Caldwell-Harris, 2014; Opitz & Degner, 2012). Typically, L1 is learned in emotionally rich, embodied contexts, while L2 often lacks such grounding experiences. Yet, when proficiency in both languages is accounted for, lexical decision tasks do not show reliable language-specific differences in emotional word processing (Ponari et al., 2015). This tentatively suggests that as L2 proficiency increases, its disembodiment decreases and responses can parallel those observed in L1.

Beyond internal linguistic and affective variables, environmental context can also shape lexical processing. Outside the laboratory, recognition unfolds within perceptual variability and concurrent sensory demands. Research on processing fluency shows that perceptual disfluency, such as degraded or visually challenging input, slows recognition and can alter processing strategies (Alter & Oppenheimer, 2009). In this regard, weather is one of the most pervasive factors in everyday perception. It has been associated with fluctuations in mood and expressed sentiment (Baylis et al., 2018; Jiang et al., 2022), even though some experimental effects are often subtle and depend on the manipulation (Behnke et al., 2021), underscoring the need for paradigms that are both controlled and ecologically grounded. Virtual reality (VR) affords precisely this combination (Rocabado et al., 2025a).

Within VR, initial studies suggest that naturalistic environmental load (e.g., simulated rain) negatively impacts L1 single word processing and sentence reading (Rocabado et al., 2024). Importantly for the current study, in an affective L1 word judgment task, rainy VR conditions have been shown to slow responses without altering perceived valence, while intrinsic affective evaluations were kept intact (Rocabado & Duñabeitia, 2025). This dissociation indicates that the impact of simulated rain is better interpreted as perceptual interference rather than mood induction, implying that environmental disfluency taxes early visual processing while leaving affective representations largely unchanged. In another study, Rocabado et al. (2025b) employed a bilingual language-decision task with orthotactically unmarked Spanish and English words distributed across two visually degraded conditions, simulated rain and laboratory degradations (namely, a noise mask). Rainy weather induced a general slowdown across both languages, whereas visual noise-induced masking selectively reduced the L1 advantage. These findings point to a dissociation where naturalistic environmental disfluencies, such as simulated rain, tend to impose broad additive costs on processing, whereas laboratory perceptual degradations, such as masking with artificially superimposed visual noise, can more selectively disrupt bilingual lexical dynamics. Such dissociation motivates a critical next step: testing whether realistic weather manipulations modulate emotional word processing during timed lexical access, where early perceptual constraints and decisional pressure are paramount.

Despite these advances, the interaction between valence and environmental disfluency during timed bilingual word recognition in real-life-like settings remains largely unexplored, especially for negative items which are, presumably, the most sensitive to interference and thus the most likely to reveal boundary conditions. Prior VR studies have either examined bilingual language decisions with valence uncontrolled items (Rocabado et al., 2025a) or assessed valence ratings through reflective appraisal tasks without

testing lexical access efficiency of retrieving linguistic information during speeded online processing tasks (Rocabado & Duñabeitia, 2025). Consequently, it is unclear whether the tendency for negative words to elicit longer response times and less consistent performance due to attentional capture and associated costs (Hinojosa et al., 2019; Kousta et al., 2009; Kuperman et al., 2014) is amplified, attenuated, or preserved under perceptually challenging and realistic conditions like simulated rain. This issue is especially relevant in bilingual settings. Given that emotional processing in the L2 is often described as attenuated, one might argue that the negative-over-neutral contrast could be smaller in L2 than in L1. Nevertheless, in tightly controlled lexical decision tasks, highly proficient L2 speakers show the same effects for valenced over neutral words as native speakers do, with no reliable Language by Valence interaction. Late L2 speakers are slower overall and more frequency-sensitive, but the valence advantage persists in both languages (Ponari et al., 2015; Wang et al., 2023).

The present study embeds a Spanish (L1) and English (L2) language-decision task in a VR realistic environment under sunny versus rainy conditions while manipulating valence (negative versus neutral). This task was selected for its heightened sensitivity to both lexical and sublexical processing demands, as it requires the successful integration of phonotactic, orthotactic, and semantic cues. Unlike traditional lexical decision tasks that can be administered in blocked language contexts, the language-decision task requires explicit identification of the language of each item, thereby maximizing cross-linguistic competition and providing a stringent test of bilingual lexical access under perceptual load (van Hell & Tanner, 2012). This design reflects situated bilingual cognition within an immersive framework, simulating how languages compete for attention in dynamic real-world settings rather than being processed in isolated blocks. We focus on the contrast between impactful highly negative words (i.e., words with low valence and high arousal) and neutral items (i.e., words with neutral valence and low arousal), since such negative stimuli are theoretically and empirically the most sensitive to interference, offering maximal leverage for detecting contextual modulation. This contrast is particularly relevant for bilingual processing, because negative high-arousal words are precisely those for which emotional resonance tends to be most attenuated in L2 compared to L1. If L2 emotional grounding is weaker, one might expect the facilitation associated with high-arousal negativity to be reduced, absent, or delayed in the L2 relative to the L1. Equally, if motivational salience, meaning its capacity to capture attention due to biological or behavioral relevance, overrides language-based attenuation, facilitation could emerge in both languages, paralleling the patterns observed in prior studies (Ponari et al., 2015). By focusing on this critical region of the affective space: negative high-arousal versus neutral low-arousal words, and targeting lexical-semantic and decisional processes, our design provides a stringent test of whether L2 emotional disembodiment persists under realistic perceptual load, or whether arousal-driven vigilance effects generalize across languages even when both languages are processed with comparable efficiency, acting as domain-general constraints on the time course of word recognition.

2. Method

2.1. Participants

A total of 40 students from Nebrija University participated in the present experiment. Among the participants, 32 self-identified as female (Mage = 22.22, SD = 4.53) and 8 as male (Mage = 22, SD = 4.41). All participants had normal or corrected-to-normal

vision and hearing and were native speakers of Spanish (L1) with English as their second language (L2). To enroll in the experimental session, participants were required to self-report a B2 level of English proficiency according to the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001). This was further confirmed using LexTALE (Lemhöfer & Broersma, 2012), a vocabulary knowledge test designed to estimate lexical proficiency. Scores around 60–70 correspond to a B2 level, and participants achieved an average score of 70.96 (SD = 7.66). Written informed consent was obtained prior to participation, and the study was approved by the Research Ethics Committee at Nebrija University (approval code UNNE-2022-0017).

2.2. Materials

A total of 160 Spanish words and their English partial cognate translations were selected from Multilex (Spanish: Martínez et al., 2024a; English: Brysbaert et al., 2024). Relying on the Spanish word properties available we chose 80 negative word pairs and 80 neutral word pairs. To minimize trivial form-based cues, items were orthographically non-identical cognates (Comesaña et al., 2015). This partial overlap was confirmed using the length-corrected orthographic Levenshtein Distance (see Duñabeitia et al., 2013, 2020), a metric that ranges from 0 (totally different translation equivalents) to 1 (completely overlapping cognates). The corrected Levenshtein Distance averaged 0.66, with values ranging from 0.34 (e.g., *diablo* – *devil*) to 0.80 (e.g., *secta* – *sect*). Word pairs were additionally matched on frequency, length, and orthographic neighborhood density (OLD20, Keuleers, 2013; Yarkoni et al., 2008; see Table 1). Crucially, the negative and neutral sets differed not only in valence but also in arousal. Normative ratings from Martínez et al. (2024b) confirmed that negative words were both highly negative and high in arousal, whereas neutral words were mid-valence and low in arousal. This ensured that the manipulation contrasted very negative, high-arousal items with neutral, low-arousal ones, directly targeting the region of the affective space most likely to reveal vigilance-driven differences in lexical access.

2.3 Virtual reality setting and apparatus

The experimental setting was designed and created using Vizard 7.5 (WorldViz, 2024), a virtual reality (VR) software platform, and presented to participants through a head-mounted display (HMD). The VR task was programmed and designed using Python 3 and

Vizard 7 (WorldViz). The experimental environments and stimuli were displayed using the HTC VIVE Pro HMD, which offers a rendering resolution of 2880 × 1600 pixels (1440 × 1600 pixels per eye), a 90 Hz refresh rate, and a 110° field of view. The participant's viewpoint remained fixed throughout the task, irrespective of their real-world movements.

Modifications to the 3D model and the main environment were implemented using Vizard Inspector (WorldViz, 2024). This tool was utilized to remove unnecessary 3D objects and incorporate a white canvas for the presentation of experimental stimuli. Additionally, ambient sounds were included to enhance the realism of the VR experience, with rain sounds for the rainy condition and sounds of a fountain and pigeons for the sunny condition. To ensure auditory consistency, all ambient sounds were normalized and presented at a constant volume across conditions. These sounds were selected to be strictly congruent with the respective virtual contexts, providing a realistic level of environmental noise.

The background sky was animated to reflect the respective weather conditions (see [Supplementary Materials](#) for a video demonstration). Furthermore, the rainy condition featured darker modified building textures to reflect a cloudy, wet setting. Crucially, these environmental changes did not extend to the central white canvas used for stimulus presentation, which maintained uniform brightness and contrast across both weather conditions. This ensured that any visual interference was strictly a result of the dynamic falling raindrops between the participant and the target, rather than changes in the legibility or illumination of the display itself.

The items were presented in a 3D open street residential neighborhood, chosen as the main scenario for its high realism, familiarity, and the openness of the simulated environment. This immersive setting allowed participants to experience simulated weather in a more naturalistic context (see [Figure 1](#) and [Supplementary Material](#)).

2.4. Task and procedure

Participants were seated on a rotating chair and provided with a HMD to immerse them in the 3D virtual environment, allowing a full 360° stationary perspective. Once the headset was placed and calibrated, participants were given two controllers, representing virtual hands, for interacting within the environment. The instructions for the language-decision task were displayed on a floating canvas in the virtual setting. These instructions outlined the two phases of the experiment: a practice phase followed by the main experimental phase. Participants categorized the linguistic stimuli as either Spanish or English by pressing buttons on the left or right controller, respectively, and were encouraged to respond as quickly as possible based on their first impressions.

The target words were displayed centrally on a simulated white canvas in black Courier New monospaced font to ensure readability. Each trial began with a central fixation mark presented for 500 ms to orient the participants' gaze toward the stimulus area and minimize potential distractions from environmental elements. This was followed by the target word, which remained visible for a maximum of 3000 ms or until a response was provided. An inter-stimulus interval blank screen was presented for 500 ms before the next trial.

The order of presentation of the weather conditions (i.e., rainy or sunny) was counterbalanced across participants. In each weather condition, participants completed 160 unique randomized trials: 40 negative and 40 neutral words in Spanish, and 40 negative and 40 neutral words in English. After a 10–15 minute unrelated filler task, participants completed a second language-decision session

Table 1. Descriptive statistics of characteristics of the materials

Word properties	Spanish		English	
	Negative	Neutral	Negative	Neutral
Word frequency	3.83 (0.66)	3.89 (0.57)	3.87 (0.64)	3.91 (0.57)
Word length	6.40 (0.76)	6.40 (0.72)	6.29 (1.07)	6.29 (1.14)
Corrected LD	0.67 (0.15)	0.65 (0.15)	0.67 (0.15)	0.65 (0.15)
Neighborhood density	2.30 (0.41)	2.35 (0.49)	2.30 (0.50)	2.28 (0.53)
Valence	2.17 (0.71)	5.01 (0.29)	2.14 (1.04)	7.04 (1.07)
Arousal	6.85 (0.75)	2.57 (0.60)	5.1 (0.34)	1.94 (1.13)

Note: Values reported are means with standard deviation in parentheses for word frequency (Zipf scale), word length (number of letters), and the length-corrected LD metric for interlingual orthographic overlap distance score (Duñabeitia et al., 2013). Neighborhood density is indexed using OLD20, defined as the mean orthographic distance to the 20 closest lexical neighbors (lower values indicate denser neighborhoods). Finally, valence and arousal properties are presented for Spanish words.



Figure 1. Example of the participant's perspective in the main scenario under sunny (left) and rainy (right) weather conditions.

under the remaining weather condition, again with 160 trials. Critically, translation pairs were split across weather conditions, so that translation equivalents were never presented in the same weather condition. The two 160-word lists were counterbalanced across weather conditions, and all words within each list were presented in a random order.

2.5. Data processing

All data handling was conducted in RStudio (RStudio Team, 2022). Reaction-time (RT) analyses were restricted to correct trials. Trials with incorrect responses were removed prior to trimming. For each participant, RT outliers exceeding ± 2.5 standard deviations from that participant's mean were excluded; this procedure led to the removal of 3.77% of the data. Accuracy was computed on the full set of valid trials prior to RT trimming. The final analysis datasets were inspected to verify normality of residuals for RT models and to ensure that exclusion rules were applied uniformly across experimental cells.

2.6. Statistical analysis

All inferential statistics were performed in jamovi (The jamovi project, 2022) using the GAMLj module (Gallucci, 2019), which interfaces R's mixed-effects engines. Accuracy was analyzed with a binomial generalized linear mixed-effects model (GLMM) and RTs with linear mixed-effects models (LMMs). Fixed factors were Language (Spanish/L1 versus English/L2), Valence (negative high-arousal versus neutral low-arousal), and Weather (sunny versus rainy). In addition, Trial Order was included as a covariate (*z*-scored), following psycholinguistic modeling conventions to control for systematic variance related to time on task (Baayen et al., 2008).

All predictors were coded using deviation (sum-to-zero) contrasts so that main effects reflect average differences across the other factor levels. Random intercepts for participants and items were included in all models. We began with a maximal random-effects structure (Barr et al., 2013; Brysbaert & Debeer, 2025), specifying by-participant slopes for all within-subject predictors (Language, Valence, Weather, and Trial Order), and by-item slopes for Weather and Order. Random slopes were modeled as uncorrelated to mitigate convergence issues (Matuschek et al., 2017).

Due to persistent convergence and identifiability warnings (e.g., degenerate Hessian, negative eigenvalues), we adopted a principled

stepwise reduction strategy. The final RT model retained random intercepts for participants and items, by-participant slopes for Language and Order, and fixed effects for Language, Valence, Weather, and Trial Order, along with the critical interaction between Weather and Order. No other interactions improved model fit or reached significance (all *ps* > .19), so the final model was additive.

For accuracy, the final GLMM included fixed effects of Language, Valence, Weather, and Trial Order, but no interaction terms, as none reached significance or improved model fit (all *ps* > .10). The random-effects structure included intercepts for participants and items, and a by-participant random slope for Language.

Likelihood-ratio χ^2 tests are reported for GLMMs, *F*-tests for LMMs, and associated *p*-values. Odds ratios and mean differences in milliseconds are provided where relevant. Estimated marginal means are plotted with 95% confidence intervals in all figures.

3. Results

3.1. Accuracy

Accuracy was generally high across conditions. The final binomial GLMM included fixed effects of Language (Spanish versus English), Valence (negative versus neutral), Weather (sunny versus rainy), and centered Trial Order, with random intercepts for participants and items.

There were no significant effects of Language, $\chi^2(1) = 0.21$, *p* = .645, or Weather, $\chi^2(1) = 1.93$, *p* = .165. The effect of Trial Order was also not significant, $\chi^2(1) = 0.83$, *p* = .362. A marginal effect of Valence was observed, $\chi^2(1) = 3.49$, *p* = .062, with post-hoc contrasts indicating that responses to negative words were 1.42 times more likely to be correct than responses to neutral words (see Figure 2). In sum, accuracy did not vary systematically with language or weather and showed only a weak tendency toward better performance for negative relative to neutral stimuli.

3.2. Reaction times

RTs were analyzed using a linear mixed-effects model with fixed effects of Language, Valence, Weather, and Trial Order (centered), as well as the Weather by Trial Order interaction. The final model included random intercepts for participants and items, and by-participant random slopes for Language and Trial Order.

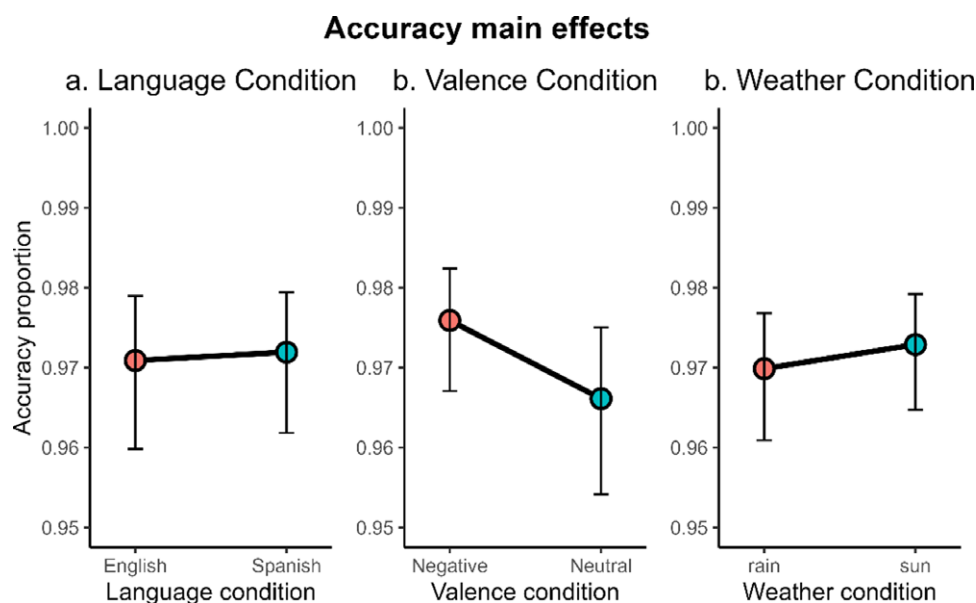


Figure 2. Estimated marginal means of accuracy proportions for the three main factors: (left) language condition, (middle) valence condition, and (right) weather condition. Each panel presents the main effects of each factor on accuracy. Vertical bars represent ± 1 standard error of the mean.

A robust main effect of Valence emerged, $F(1, 312.7) = 10.34$, $p = .001$, with faster responses to negative words compared to neutral ones, yielding an average facilitation of 28 ms. The main effect of Weather was also significant, $F(1, 11,895.6) = 4.39$, $p = .036$, with overall slower responses in rainy compared to sunny conditions (mean difference = 9 ms). In contrast, the main effect of Language did not reach significance, $F(1, 90.5) = 2.33$, $p = .13$ (see Figure 3).

Crucially, we observed a significant Weather by Trial Order interaction, $F(1, 12,024) = 7.22$, $p = .007$ (Figure 4). This indicates that the performance gap between weather conditions changed over time, driven primarily by an improvement under the sunny weather condition rather than a progressive decline under the rainy weather

condition. Simple effects analyses revealed that in early trials (-1 SD), the difference between conditions was minimal (2.65 ms, $SE = 6.09$, $t = 0.435$, $p = .664$). At the task midpoint, responses under rainy weather were significantly slower than under sunny weather conditions by 8.98 ms ($SE = 4.29$, $t = -2.10$, $p = .036$). By the final phase of the task ($+1$ SD), this difference reached 20.61 ms ($SE = 6.09$, $t = -3.38$, $p < .001$), indicating that the sunny weather condition became progressively faster while the rain condition remained relatively stable across trials.

Taken together, these results show that emotional valence consistently facilitates lexical access, and that virtual weather imposes a perceptual cost that becomes increasingly apparent as the task progresses. This pattern emerges because performance improves

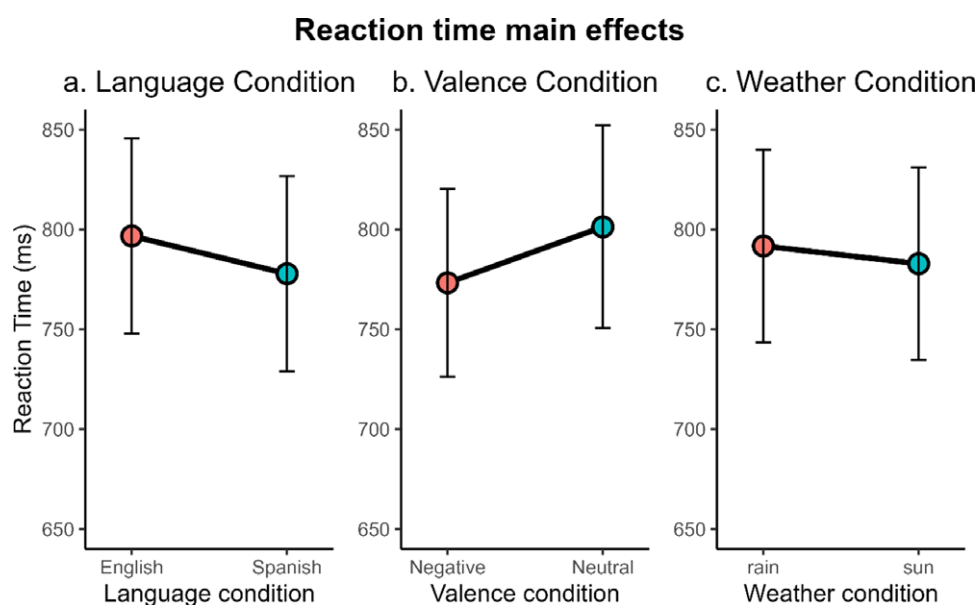


Figure 3. Estimated marginal means of reaction times (in milliseconds) for the three main model factors: (left) language condition, (middle) valence condition, and (right) weather condition. Each panel plot illustrates the main effects of each factor on reaction time. Vertical bars represent ± 1 standard error of the mean.

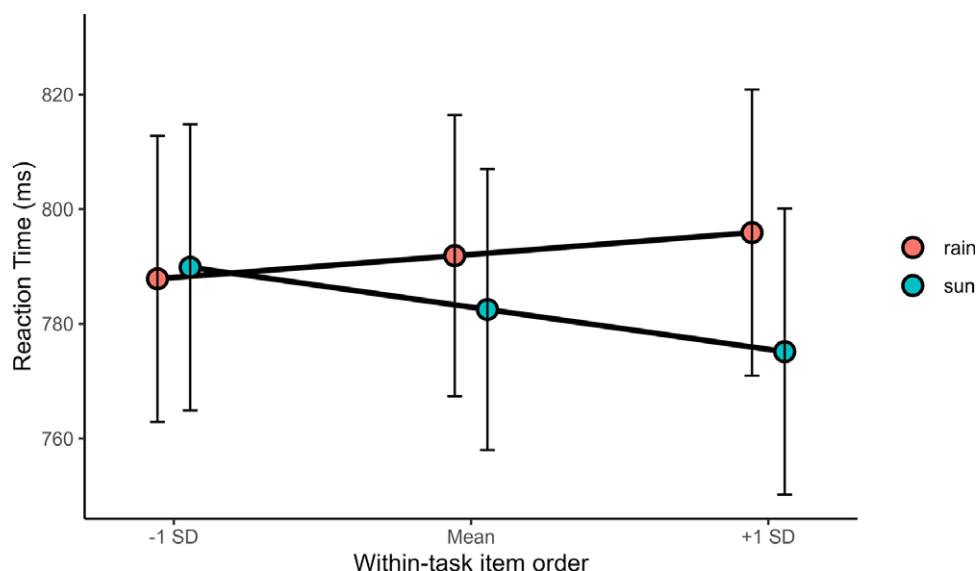


Figure 4. Estimated marginal means of reaction times (in milliseconds) as a function of weather condition across item order, shown at -1 SD, the mean, and $+1$ SD of the order distribution. The panel illustrates the significant weather by item order interaction, with the sun-related acceleration increasing over the course of the task. Vertical bars represent ± 1 standard error of the mean.

over time in sunny weather conditions, while response speed remains relatively stable under rain. Importantly, these effects were additive since no two-way or three-way interactions among Language, Valence, and Weather reached significance, with all p -values above .15. These findings support the interpretation that emotional and environmental factors independently constrain word recognition during dynamic VR-based processing.

4. Discussion

The present study examined how bilingual word recognition is modulated by emotional valence and environmental context under realistic perceptual conditions. Using a VR-based language-decision task, we tested whether simulated weather (sunny versus rainy conditions) interacts with word valence (negative versus neutral) and language (L1 Spanish versus L2 English). Moreover, exploratory models including trial-level order did not reveal any modulating effects on these main patterns, except for a significant interaction between Weather and Trial Order in the RT data, which indicated that rainy conditions led to increasingly slower responses as the task progressed. This suggests that the perceptual cost imposed by rain may accumulate under sustained load but does not alter the fundamental additive architecture of the effects. In addition, given that accuracy was high across conditions and no interactions involving valence, weather, or language emerged in the accuracy data, further reinforcing the additive pattern of effects and supporting the interpretability of the main RT results without confounding speed–accuracy trade-offs. Taken together, the findings indicate that language, valence, and environmental context each exert reliable influences on word recognition, but their effects were additive rather than interactive.

While previous studies have emphasized the robustness of L1 advantages in bilingual lexical processing, our findings revealed no significant effect of Language in the RT data. Although responses tended to be slower to English than to Spanish items, this trend did not reach statistical significance. This contrasts with the classic L2 cost typically observed in bilingual lexical access, even when the two

languages are closely matched in their properties and when proficiency is relatively high (Duyck et al., 2008; Gollan et al., 2008). Similar developmental evidence has shown that slower responses in the non-dominant language emerge even in balanced bilingual children, and that this asymmetry gradually diminishes as reading skills consolidate (Duñabeitia et al., 2016). In the present context, the lack of a reliable Language effect provides an important baseline for interpreting the emotional and environmental effects, as it suggests that processing demands across languages may converge under immersive and ecologically valid conditions. Although our design was sensitive enough to detect other small main effects and interactions, the absence of a robust Language effect in this case may reflect the relatively high proficiency and frequent usage of both languages in our bilingual sample. This aligns with previous accounts suggesting that language dominance effects may attenuate in immersive or ecologically valid contexts where both languages are highly active (van Hell & Tanner, 2012).

Regarding the valence effect observed (namely, faster responses to negative relative to neutral words), our results converge with studies showing that both negative and positive words can enjoy a processing advantage over neutral items (Citron et al., 2014; Hofmann et al., 2009). This facilitation is consistent with the approach–withdrawal model (Robinson et al., 2004), by which negative valence combined with high-arousal signals potential threats and triggers rapid avoidance-oriented responses. Although the classic automatic vigilance account predicts a generic slowdown for negative words (Algom et al., 2004; Estes & Adelman, 2008), subsequent work has clarified that negativity is not categorical, and not all negative words produce the same degree of interference. In fact, interactions with arousal show that low-arousal negative words are most likely to produce slowdowns, whereas high-arousal negative words often produce less interference and sometimes even facilitation (Larsen et al., 2008; Haro et al., 2024). Our findings fit this pattern, suggesting that the use of highly arousing negative items may have shifted processing from interference toward vigilance-driven facilitation.

Importantly, the facilitation observed for negative words was parallel across L1 and L2, with no evidence for a Language by

Valence interaction. At first glance, this absence of attenuation in L2 may seem unexpected given frequent reports of weaker emotional resonance in the second language (Caldwell-Harris, 2014; Pavlenko, 2012). However, our results converge with previous well-controlled lexical decision studies showing that, when proficiency and stimulus properties are carefully matched, valence effects can be equivalent across L1 and L2. In this line, Ponari et al. (2015) found that both native speakers and highly proficient L2 speakers exhibited comparable facilitation for emotional relative to neutral words in lexical decision (see Sutton et al., 2007 for an early Stroop-task evidence). Although previous reviews emphasize that findings are heterogeneous (see Aguilar et al., 2024), our results converge with this view, suggesting that arousal-driven vigilance effects generalize across L1 and L2 when motivational salience is high.

Simulated weather also reliably modulated word processing and language-decision times. Responses under rainy conditions were slower than under sunny conditions, an effect that was consistent across both languages and valence categories. This pattern supports the perceptual load account suggested by Rocabado et al. (2025a), stating that environmental disfluency imposed by simulated rain taxes early perceptual processing, yielding a general slowdown, but does not selectively amplify or attenuate the effects of language or emotion. These results are in line with previous VR studies in which rain slowed lexical access and reading but did not alter intrinsic affective evaluations of words (Rocabado & Duñabeitia, 2025). By contrast, laboratory degradations such as visual masking can disrupt bilingual dynamics more selectively (e.g., reducing the general L1 advantage). Notably, however, the effect of Weather was not constant across time. The observed interaction between Weather and Trial Order indicated that performance improved progressively under sunny conditions, while response speed under rainy conditions remained stable across the task. This shows that the classic practice effect, in which participants respond faster as they become familiar with the task, was only present in the sunny weather condition. By contrast, the difficulty introduced by the rainy environment appears to have prevented this improvement. Although a central fixation mark was used to orient attention before each trial, the falling droplets introduced persistent visual disturbance that contributed to perceptual disfluency without necessarily diverting gaze. Prior research has shown that dynamic but task-irrelevant visual input can impair performance by introducing noise that disrupts perceptual clarity, reduces processing fluency, and slows target detection (Li & Shimomura, 2025; Oppenheimer, 2008). Perceptual environments with stable and predictable features, like the sunny condition, tend to support practice-related gains by allowing participants to allocate attention more efficiently to the task demands (Oppenheimer, 2008). In this regard, the sunny weather may have facilitated faster adaptation, whereas the rain constrained performance by maintaining a constant level of interference. Importantly, the Weather effect was not accompanied by any interaction with Valence or Language, further supporting that the influence of simulated rain was additive and did not modulate emotional or linguistic processing.

At first sight, our results may seem at odds with some preceding VR studies that report mood- or situation-congruent effects in emotional processing (Wang et al., 2023). In those studies, exposure to negative versus positive VR contexts biased subsequent word selection and amplified physiological responses such as galvanic skin conductance (Wang et al., 2023). However, a key distinction needs to be noted at the level of the task timing. Affect-labeling and appraisal tasks require reflective judgment and therefore allow emotional context to influence decision criteria. In contrast, lexical

or language decision demands rapid categorization, tapping earlier stages of lexical access. Our findings indicate that in such speeded tasks, realistic weather manipulations function primarily as early perceptual load rather than as affective inductions, producing additive slowdowns without interacting with emotional content. This dissociation highlights the importance of task characteristics in determining when environmental factors shape emotion–language coupling.

Furthermore, our results confirm that the disruptive influence of simulated rain is not uniform, but accumulates over time, as evidenced by the significant Weather by Trial Order interaction. This dynamic pattern supports the interpretation of environmental disfluency as a cumulative perceptual constraint rather than a transient mood induction, consistent with prior findings by Rocabado et al. (2025a). At the same time, the absence of interactions with language or valence indicates that the interference induced by simulated rain operates independently of lexical–emotional content. This points to a domain-general interference mechanism – likely due to reduced visual clarity or increased attentional demands – rather than to affective congruence or semantic integration.

Finally, it is important to acknowledge certain methodological constraints regarding environmental manipulation. While ambient sounds were matched for volume to ensure consistency, they were not controlled for other acoustic dimensions such as pitch or frequency spectrum. This limitation is relevant, as distinct sound profiles across weather conditions could introduce uncontrolled variance in auditory salience. Although these sounds were selected for their contextual congruence and ecological plausibility, future research should explicitly control or manipulate spectral features to examine whether auditory complexity interacts with visual load or lexical processing. Moreover, while word lists were matched for frequency, length, and orthographic neighborhood density, other uncontrolled lexical and sublexical factors could have exerted their influence in the observed effects. These dimensions represent residual design limitations that future research should address.

5. Conclusion

In this study, we examined how language (L1 Spanish versus L2 English), emotional valence (negative high-arousal versus neutral low-arousal), and realistic weather (sun versus rain) shape speeded lexical access in an additive manner in VR. The findings demonstrate that while rainy weather imposes a general processing cost, both languages are processed with comparable efficiency under these immersive conditions, and the relative pattern of valence effects remains consistent across languages. Crucially, the disruptive influence of rain was shown to increase over time, as revealed by the significant Weather by Trial Order interaction, indicating that perceptual load accumulates under sustained exposure rather than operating as a fixed or transient effect. These findings help delineate the boundary conditions of the foreign-language effect in emotion: while immersive VR contexts can bias affective judgments in reflective tasks, early lexical decision appears dominated by perceptual load and shows parallel valence effects in L1 and L2. The study also highlights that realistic environmental manipulations can systematically affect cognitive processing without necessarily altering emotional or linguistic dynamics, supporting the interpretation of simulated rain as a domain-general source of perceptual interference rather than an affective induction. More broadly, the study illustrates the utility of VR for bridging laboratory control with ecological realism, providing a framework for testing how environmental and emotional factors converge in situated bilingual cognition.

Supplementary materials. Video samples of the virtual reality task conducted under both sunny and rainy weather conditions, along with a list of experimental materials, can be found at the following link: <https://doi.org/10.17605/OSF.IO/2WK8S>.

Data availability statement. The data and analysis code that support the findings will be available at Open Science Framework at the following link: <https://doi.org/10.17605/OSF.IO/2WK8S>.

Acknowledgements. This research was partially funded by PID2024-161331NB-I00 (MCIN/AEI/10.13039/501100011033). We thank Nebrija University participants for their contribution.

Competing interests. The authors declare no competing interests.

Ethics statement. 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 1975, as revised in 2008.

AI tools disclosure. The authors declare no use of artificial intelligence (AI) tools in the preparation of the present document.

References

- Aguilar, M., Ferré, P., & Hinojosa, J. A. (2024). Chapter one – The landscape of emotional language processing in bilinguals: A review. In E. K. D. Federmeier (Ed.), *Psychology of learning and motivation* (Vol. 80, pp. 1–32). Academic Press. <https://doi.org/10.1016/bs.plm.2024.03.002>.
- Algom, D., Chajut, E., & Lev, S. (2004). A Rational Look at the Emotional Stroop Phenomenon: A Generic Slowdown, Not a Stroop Effect. *Journal of Experimental Psychology: General*, 133(3), 323–338. <https://doi.org/10.1037/0096-3445.133.3.323>.
- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc.*, 13(3), 219–235. <https://doi.org/10.1177/1088868309341564>.
- 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, Special Issue: Emerging Data Analysis*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>.
- 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. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Baylis, P., Obradovich, N., Kryvasheyeu, Y., Chen, H., Coviello, L., Moro, E., Cebrian, M., & Fowler, J. H. (2018). Weather impacts expressed sentiment. *PLoS One*, 13(4), e0195750. <https://doi.org/10.1371/journal.pone.0195750>.
- Behnke, M., Overbye, H., Pietruch, M., & Kaczmarek, L. D. (2021). How seasons, weather, and part of day influence baseline affective valence in laboratory research participants? *PLoS One*, 16(8), e0256430. <https://doi.org/10.1371/journal.pone.0256430>.
- Brysbaert, M., & Debeer, D. (2025). How to run linear mixed effects analysis for pairwise comparisons? A tutorial and a proposal for the calculation of standardized effect sizes. *Journal of Cognition*, 8(1). <https://doi.org/10.5334/joc.409>.
- Brysbaert, M., Martínez, G., & Reviriego, P. (2024). Moving beyond word frequency based on tally counting: AI-generated familiarity estimates of words and phrases are an interesting additional index of language knowledge. *Behavior Research Methods*, 57(1), 28. <https://doi.org/10.3758/s13428-024-02561-7>.
- Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01055>.
- Chen, P., Lin, J., Chen, B., Lu, C., & Guo, T. (2015). Processing emotional words in two languages with one brain: ERP and fMRI evidence from Chinese–English bilinguals. *Cortex*, 71, 34–48. <https://doi.org/10.1016/j.cortex.2015.06.002>.
- Citron, F. M. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*, 56, 79–89. <https://doi.org/10.1016/j.neuropsychologia.2014.01.002>.
- Comesaña, M., Ferré, P., Romero, J., Guasch, M., Soares, A. P., & García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: The role of stimuli list composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 614–635. <https://doi.org/10.1037/xlm0000065>.
- Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). Piensa twice: On the foreign language effect in decision making. *Cognition*, 130(2), 236–254. <https://doi.org/10.1016/j.cognition.2013.11.010>.
- Council of Europe, Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Crossfield, E., & Damian, M. F. (2021). The role of valence in word processing: Evidence from lexical decision and emotional Stroop tasks. *Acta Psychologica*, 218, 103359. <https://doi.org/10.1016/j.actpsy.2021.103359>.
- Duñabeitia, J. A., Borragán, M., de Bruin, A., & Casaponsa, A. (2020). Changes in the sensitivity to language-specific orthographic patterns with age. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.01691>.
- Duñabeitia, J. A., Dimitropoulou, M., Morris, J., & Diependaele, K. (2013). The role of form in morphological priming: Evidence from bilinguals. *Language and Cognitive Processes*, 28(7), 967–987. <https://doi.org/10.1080/01690965.2012.713972>.
- Duñabeitia, J. A., Ivaz, L., & Casaponsa, A. (2016). Developmental changes associated with cross-language similarity in bilingual children. *Journal of Cognitive Psychology*. (world). <https://doi.org/10.1080/20445911.2015.1086773>.
- Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word processing: Evidence from the semantic Simon paradigm. *Psychonomic Bulletin & Review*, 15(5), 961–966. <https://doi.org/10.3758/PBR.15.5.961>.
- Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion* (Washington, D.C.), 8(4), 441–444; discussion 445–457. <https://doi.org/10.1037/1528-3542.8.4.441>.
- Ferré, P., Fraga, I., & Hinojosa, J. A. (2025). The interplay between language and emotion: Introduction to the special issue. *Cognition and Emotion*, 1–13. <https://doi.org/10.1080/02699931.2025.2549966>.
- Gallucci, M. (2019). *GAMLj: General analyses for the linear model in jamovi*.
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015). The foreign language effect on moral judgment: The role of emotions and norms. *PLoS One*, 10(7), e0131529. <https://doi.org/10.1371/journal.pone.0131529>.
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787–814. <https://doi.org/10.1016/j.jml.2007.07.001>.
- Haro, J., Hinojosa, J. A., & Ferré, P. (2024). The role of individual differences in emotional word recognition: Insights from a large-scale lexical decision study. *Behavior Research Methods*, 56(8), 8501–8520. <https://doi.org/10.3758/s13428-024-02488-z>.
- Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science*, 28(10), 1387–1397. <https://doi.org/10.1177/0956797617720944>.
- Hinojosa, J. A., Moreno, E. M., & Ferré, P. (2019). Affective neurolinguistics: Towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience*, 35(7), 813–839. <https://doi.org/10.1080/23273798.2019.1620957>.
- Hofmann, M. J., Kuchinke, L., Tamm, S., Vo, M. L. H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience*, 9(4), 389–397. <https://doi.org/10.3758/9.4.389>.
- Iacozza, S., Costa, A., & Duñabeitia, J. A. (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and

- foreign language. *PLoS One*, 12(10), e0186027. <https://doi.org/10.1371/journal.pone.0186027>.
- Jiang, J., Murrugara-Llerena, N., Bos, M. W., Liu, Y., Shah, N., Neves, L., & Barbieri, F. (2022). Sunshine with a chance of smiles: How does weather impact sentiment on social media? *Proceedings of the International AAAI Conference on Web and Social Media*, 16, 393–404. <https://doi.org/10.1609/icwsm.v16i1.19301>.
- Kauschke, C., Bahn, D., Vesker, M., & Schwarzer, G. (2019). The role of emotional valence for the processing of facial and verbal stimuli—Positivity or negativity bias? *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01654>.
- Keuleers, E. (2013). *vwr: Useful functions for visual word recognition research*. <https://CRAN.R-project.org/package=vwr>
- Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112(3), 473–481. <https://doi.org/10.1016/j.cognition.2009.06.007>.
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, 143(3), 1065–1081. <https://doi.org/10.1037/a0035669>.
- Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, 8(4), 445–452. <https://doi.org/10.1037/1528-3542.8.4.445>.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>.
- Li, Y., & Shimomura, Y. (2025). The effect of dynamic stimuli on attention under different perceptual loads. *Journal of Physiological Anthropology*, 44(1), 18. <https://doi.org/10.1186/s40101-025-00398-3>.
- Martínez, G., Conde, J., Reviriego, P., & Brysbaert, M. (2024a). AI-generated estimates of familiarity, concreteness, valence, and arousal for over 100,000 Spanish words. *Quarterly Journal of Experimental Psychology*, 17470218241306694. <https://doi.org/10.1177/17470218241306694>.
- Martínez, G., Molero, J. D., González, S., Conde, J., Brysbaert, M., & Reviriego, P. (2024b). Using large language models to estimate features of multi-word expressions: Concreteness, valence, arousal. *Behavior Research Methods*, 57(1), 5. <https://doi.org/10.3758/s13428-024-02515-z>.
- Martínez-Huertas, J. Á., Jorge-Botana, G., Martínez-Mingo, A., Iglesias, D., & Olmos, R. (2025). Are valence and arousal related to the development of amodal representations of words? A computational study. *Cognition and Emotion*, 39(7), 1465–14. <https://doi.org/10.1080/02699931.2023.2283882>.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>.
- Opitz, B., & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, 50(8), 1961–1967. <https://doi.org/10.1016/j.neuropsychologia.2012.04.021>.
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, 12(6), 237–241. <https://doi.org/10.1016/j.tics.2008.02.014>.
- Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, 47(6), 405–428. <https://doi.org/10.1080/00207594.2012.743665>.
- Ponari, M., Rodríguez-Cuadrado, S., Vinson, D., Fox, N., Costa, A., & Vigliocco, G. (2015). Processing advantage for emotional words in bilingual speakers. *Emotion*, 15(5), 644–652. <https://doi.org/10.1037/emo0000061>.
- Robinson, M. D., Storbeck, J., Meier, B. P., & Kirkeby, B. S. (2004). Watch out! That could be dangerous: Valence-arousal interactions in evaluative processing. *Personality & Social Psychology Bulletin*, 30(11), 1472–1484. <https://doi.org/10.1177/0146167204266647>.
- Rocabado, F., & Duñabeitia, J. A. (2025). Clouded judgments? The role of virtual weather in word valence evaluations. *Cognition and Emotion*, 39(7), 1446–1456. <https://doi.org/10.1080/02699931.2024.2354488>.
- Rocabado, F., Muntini, L., González Alonso, J., & Dunabeitia, J. A. (2024). Weathering words: A virtual reality study of environmental influence on Reading dynamics. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1433781>.
- Rocabado, F., Muntini, L., Jubran, O. F., Lachmann, T., & Duñabeitia, J. A. (2025a). Transforming language research from classic desktops to virtual environments. *Scientific Reports*, 15(1), 23118. <https://doi.org/10.1038/s41598-025-08319-1>.
- Rocabado, F., Schmitz, G., & Duñabeitia, J. A. (2025b). You can stand under my umbrella: Cognitive load in second-language Reading. *Behavioral Sciences*, 15(8), 1051. <https://doi.org/10.3390/bs15081051>.
- RStudio Team. (2022). *RStudio: Integrated Development Environment for R* [Software]. RStudio, PBC. <http://www.rstudio.com/>
- Scott, G. G., O'Donnell, P. J., & Sereno, S. C. (2014). Emotion words and categories: Evidence from lexical decision. *Cognitive Processing*, 15(2), 209–215. <https://doi.org/10.1007/s10339-013-0589-6>.
- Sutton, T. M., Altarriba, J., Gianico, J. L., & Basnight-Brown, D. M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish–English bilingual speakers. *Cognition and Emotion*. (world). <https://doi.org/10.1080/02699930601054133>
- The Jamovi Project. (2022). *Jamovi (Version 2.3)* [Software]. <https://www.jamovi.org/>
- van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, 62(s2), 148–171. <https://doi.org/10.1111/j.1467-9922.2012.00710.x>.
- Wang, X., Zhang, S., & Zhang, X. (2023). How do word valence and classes influence lexical processing? Evidence from virtual reality emotional contexts. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1032384>.
- Worldviz. (2024). *Vizard 7* [Software]. WorldViz LLC. <https://www.worldviz.com/>.
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15(5), 971–979. <https://doi.org/10.3758/PBR.15.5.971>.