



Article

You Can Stand Under My Umbrella: Cognitive Load in Second-Language Reading

Francisco Rocabado ^{1,*} , Gianna Schmitz ² and Jon Andoni Duñabeitia ¹

¹ Centro de Investigación Nebrija en Cognición (CINC), Department of Education, Universidad Nebrija, 28015 Madrid, Spain; jdunabeitia@nebrija.es

² Facultat de Filologia y Comunicació, Universitat de Barcelona, 08007 Barcelona, Spain; gianna.schmitz@icloud.com

* Correspondence: jrocabado@nebrija.es; Tel.: +34-677724174

Abstract

Second-language (L2) written processing has often been linked to cognitive disfluency, resembling fluency disruptions caused by perceptual challenges, such as visual degradation. This study used Virtual Reality to investigate whether cognitive disfluency in L2 mirrors perceptual disfluency by simulating adverse weather conditions (sunny vs. rainy) and applying visual masking. Spanish–English bilinguals completed a language decision task, identifying orthotactically unmarked words as either Spanish (L1) or English (L2) while experiencing these perceptual manipulations. Results showed that visual masking significantly increased reaction times, particularly for L1 words, suggesting that masking can diminish the native language advantage. Spanish words under masking elicited slower responses than unmasked ones, whereas L2 word recognition remained comparatively stable. Additionally, rainy weather conditions consistently slowed responses across both languages, indicating a general effect of environmental disfluency. A significant interaction between language and masking emerged, highlighting distinct cognitive effects for different disfluency types. These findings suggest that cognitive disfluency in L2 does not equate to perceptual disfluency; each affects processing differently. The use of Virtual Reality enabled the controlled manipulation of realistic environmental variables, offering valuable insights into how perceptual and linguistic challenges jointly influence bilingual language processing.

Keywords: cognitive disfluency; second-language processing; bilingualism; virtual reality; disfluency



Academic Editor: James H. Smith-Spark

Received: 5 June 2025

Revised: 18 July 2025

Accepted: 23 July 2025

Published: 3 August 2025

Citation: Rocabado, F., Schmitz, G., & Duñabeitia, J. A. (2025). You Can Stand Under My Umbrella: Cognitive Load in Second-Language Reading. *Behavioral Sciences*, 15(8), 1051. <https://doi.org/10.3390/bs15081051>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Language comprehension does not occur in a vacuum. In our daily lives, we read and understand language while simultaneously navigating dynamic environments filled with sensory information that may compete for attention or interfere with processing. In bilingual contexts, these demands are especially relevant. Second-language (L2) reading is generally slower and more effortful than reading in the native language (L1), often involving additional cognitive load (e.g., greater working memory and attentional resources to support lexical access and semantic integration; (Paas & Van Merriënboer, 1994; Segalowitz & Hulstijn, 2005; Sweller, 1988)). This subjective sense of effort is often referred to as cognitive disfluency, a term used to describe the feeling that a task is mentally demanding or difficult to process (Alter et al., 2007; Kahneman, 2011; Oppenheimer, 2008).

In the last 30 years, the study of multilingual language processing has continued to gain momentum, particularly as the number of multilinguals in the world's population has rapidly grown. Research on language processing has aimed to elucidate whether there is a cognitive cost associated with processing an L2 in order to better understand the functional architecture of the bilingual mind. It is now well established that learning a new language leads to structural and functional brain adaptations as the cognitive system adjusts to managing multiple languages (Costa & Sebastián-Gallés, 2014). This line of research has also suggested that the cognitive disfluency produced during L2 processing may, under certain conditions, mirror perceptual disfluencies caused by challenging visual input. Advances in technology, such as Virtual Reality, now enable us to examine these parallels in unprecedented ways, inviting us to reflect on past findings while relying on innovative tools to understand bilingual cognitive processing beyond previous limitations.

In the mid-twentieth century, Cattell (1948) suggested that language processing in an L2 might not be as rapid as in a multilingual's first language. This idea is supported by more recent research, showing that L2 processing is generally less efficient than L1 processing (e.g., Stowe & Sabourin, 2005). This difficulty becomes particularly evident in reading tasks, where increased fixation durations and a greater number of fixations have been observed during L2 reading compared to L1 reading (Grotek & Ślęzak-Świat, 2024). The authors attributed this effect to different reading strategies that imply a heavier cognitive load. This finding is in line with earlier research stating that reading in L2 takes longer than in L1, even for highly proficient readers (e.g., Cop et al., 2015, 2017). Such increased processing demand has been linked to the simultaneous activation of both L2 and L1 representations during reading (Bernhardt, 2010; Koda, 2005), even in contexts explicitly designed to emphasize one language (e.g., Martin et al., 2009; Thierry & Wu, 2007; Wu & Thierry, 2010).

Different explanatory models have been proposed to understand these dynamics. For instance, the Bilingual Interactive Activation plus (BIA+) model, by Dijkstra and van Heuven (2002), suggests that upon reading a word, lexical activation is automatically triggered in every language known by the multilingual. Such involuntary interference by the L1 therefore hinders the reader during the recognition of L2 words. Furthermore, the BIMOLA model proposed by Léwy and Grosjean (2008) pivots on the idea of two different language networks in a bilingual brain and two modes of language activation: the monolingual mode in which the base language is highly activated while the guest language network is only weakly activated, and a bilingual language mode in which both language networks are strongly activated.

This heightened mental effort observed in L2 processing aligns with the broader construct of cognitive disfluency, conceived as the subjective experience of difficulty during mental processing (Oppenheimer, 2008). According to dual-process theories of reasoning (Kahneman, 2011; James, 1890), disfluency tends to shift processing from intuitive, rapid "System 1" reasoning to slower, more effortful "System 2" reasoning. Alter et al. (2007) suggested that certain types of degraded perceptual input trigger such shifts. By analogy, several studies have proposed that L2 processing also incurs higher processing costs and can reduce fluency (see Costa et al., 2019 and Vega-Mendoza et al., 2021 for review), thereby qualifying as a form of cognitive disfluency. This perspective is further supported by Segalowitz's (2010) model, which distinguishes between cognitive fluency (i.e., the efficiency of underlying cognitive processes) and verbal fluency (i.e., the observable output). According to this model, limitations in cognitive fluency, such as increased processing demands in L2, can underlie difficulties in verbal fluency commonly observed in non-native language use. Thus, cognitive disfluency may be a core mechanism con-

tributing to the effortful nature of L2 processing, especially under challenging perceptual or contextual conditions.

In parallel, perceptual disfluency, understood as a subtype of cognitive disfluency induced by visually degraded stimuli, has been shown to alter judgment and processing. For instance, [Oppenheimer and Frank \(2008\)](#) demonstrated that clearly visible items produce more accurate responses than perceptually degraded ones. In laboratory settings, perceptual disfluency is typically induced through manipulations such as altering fonts ([Alter et al., 2007](#)), blurring stimuli ([Yue et al., 2013](#)), or covering the items with a pattern mask ([Hirshman & Mulligan, 1991](#); [Mulligan, 1996](#); [Nairne, 1988](#)).

Recently, the importance of more naturalistic methods for investigations into the processing of language has reached more prominence ([Hasson et al., 2018](#)). Advances in technology, such as the progressive generalization of the use of Virtual Reality (VR), have provided the opportunity to merge perceptually rich experiences and close-to-real contexts with a controlled setting for experiments ([Eichert et al., 2018](#); [Shin et al., 2021](#)). VR is an ecologically valid tool for creating immersive virtual settings that can help researchers gain a deeper understanding of human cognition ([Eichert et al., 2018](#); [Heyselaar et al., 2021](#); [Rocabado & Duñabeitia, 2022](#)), facilitating exploration of the interplay between environmental context and cognition ([Shin et al., 2021](#); [Rocabado et al., 2022](#); [Titus et al., 2024](#)), which is otherwise difficult to achieve in a laboratory setting. In this line, recent research has turned to immersive virtual environments to simulate naturalistic disfluency. For instance, [Rocabado et al. \(2024\)](#) investigated the impact of visual conditions on reading performance, showing that rainy weather in VR increases fixation rates and reading times, whereas sunny weather facilitates reading. Similar claims were also raised by [Rocabado and Duñabeitia \(2024\)](#) in a study that examined emotional valence evaluation under different weather simulations, finding that rainy conditions modestly prolonged response times but did not alter the perceived emotional valence of words. Collectively, these studies provide clear support that simulated weather can reliably serve as an ecologically valid disfluency manipulation (see [Rocabado et al., 2025a](#), for recent evidence).

Importantly, from a theoretical standpoint in the field of second-language acquisition, [Bernhardt \(2010\)](#) referred to the L2 as a “degraded channel” in contrast to the L1, which is conceived as a “clear channel”, thus suggesting that L2 processing might be governed by some form of perceptual disfluency (see also [Segalowitz, 2016](#)). Furthermore, if they are not, this discrepancy may reflect deeper differences between lab-based and ecologically embedded stimuli. Thus, our study seeks to answer whether these two disfluency types produce equivalent effects and whether they interact with the disfluency associated with L2 processing.

To address these questions, in the present study, we employed a language decision task in a VR setting to compare naturalistic and artificial disfluency inducers. This task was chosen for its sensitivity to lexical and sub-lexical processing demands, especially when using orthotactically unmarked words ([Lecerf et al., 2024](#)). Unlike lexical decision tasks, language decision tasks require not just recognizing a word but also identifying the language it belongs to. This is an ability that hinges on successful integration of phonotactic, orthotactic, and semantic cues. Previous research has shown that such tasks can detect subtle effects of cognitive load and language interference (e.g., [Duyck & De Houwer, 2008](#); [Lemhöfer & Dijkstra, 2004](#)), making it a theoretically motivated tool for testing cognitive disfluency in bilinguals.

Perceptual disfluency was created through the inclusion of a rainy weather condition, following [Rocabado and Duñabeitia \(2024\)](#) and [Rocabado et al. \(2024\)](#). Additionally, to test this naturalistic method in comparison with a typical laboratory setting, the current study also employed a visual mask comprising static Gaussian visual noise superimposed

on the text. This method is inspired by traditional masking techniques used in perceptual research, where static or patterned visual noise is superimposed on stimuli to hinder early perceptual processing (e.g., [Hirshman & Mulligan, 1991](#) and [Steindorf et al., 2023](#) for recent examples). The rainy condition degrades the visibility of the linguistic stimuli, and it was expected to serve as a naturalistic alternative to the use of masks in laboratory settings. Participants were Spanish–English bilinguals presented with orthotactically unmarked words in both languages and asked to decide which language the word belonged to. In this context, we examined whether the naturalistic tool of virtually simulated rain would yield similar results as the laboratory tool of a visual mask. Additionally, we explored whether there is a possible similarity between the effects of perceptual disfluency and the so-called L2 processing disfluency in a way that the latter can hence be considered as real type of disfluency. Moreover, the current study investigated whether there is an interconnection between the different phenomena or if they function independently of each other. We hypothesized that if both types of perceptual disfluency—masking and rain—function similarly to L2-induced disfluency, their effects might be additive or even interactive, compounding the cognitive load. Alternatively, differences in their mechanisms might lead to dissociable patterns of interference.

2. Methods

2.1. Participants

A total of 37 students and employees from Nebrija University participated in the experiment for a monetary reward, all of them being Spanish L1 users and English L2 users. A minimum sample size of 28 participants was estimated using G*Power 3.1. ([Faul et al., 2007](#)) to achieve a medium effect size ($\alpha = 0.05$; $1-\beta = 0.95$) with the current study design. Participants' English proficiency level was assessed using LexTALE ([Lemhöfer & Broersma, 2012](#)). The average level of proficiency ($M = 69.72$, $SD = 11.44$) was equivalent to a B2 level on the European Common Framework of Reference for Languages. Twenty-four of these participants self-identified as female (mean age = 21.25, $SD = 3.31$) and thirteen participants self-identified as male (mean age = 25.00, $SD = 6.07$). All had normal or corrected-to-normal visual acuity and hearing, and none reported any form of cognitive dysfunction, assessed with a computerized cognitive battery (CogniFit Inc., San Francisco, CA, USA). Participants granted written informed consent before the experiment, and the experimental procedures were approved by the Research Ethics Committee from Nebrija University (approval code UNNE-2022-0017).

2.2. Materials

Two hundred and forty words were used as stimuli. Half of them were Spanish words taken from the EsPal Database ([Duchon et al., 2013](#)) and the other half were English words taken from the English Lexicon Project ([Balota et al., 2007](#)). To ensure task ambiguity and discourage heuristic-based decisions, only orthotactically unmarked words were included (i.e., words with valid bigram sequences in both languages). For instance, words like “shine” or “back” were excluded due to the presence of “sh” or “ck,” which are not legal bigrams in Spanish (see [Casaponsa & Duñabeitia, 2016](#)). Additionally, words containing letters that do not exist in the other language were also excluded (e.g., *niña* [girl], due to “ñ”). In addition, none of the words was a perfect cognate, given that this would obviously interfere with the language decision task.

The word stimuli were divided into two lists of 120 words each (60 Spanish, 60 English), one for each weather condition (sunny and rainy). The lists were matched across key lexical variables: word frequency (Zipf scale), word length (letters), orthographic neighborhood

(OLD20), within-language bigram frequency, and between-language bigram frequency (see Table 1).

Table 1. Descriptive statistics of characteristics of the materials.

Word Properties	English		Spanish	
	List 1	List 2	List 1	List 2
Word frequency	2.56 (0.39)	2.52 (0.44)	2.55 (0.33)	2.54 (0.32)
Word length	5.62 (0.49)	5.48 (0.50)	5.48 (0.50)	5.45 (0.50)
Within-language bigram frequency	0.81 (0.33)	0.79 (0.31)	0.87 (0.28)	0.82 (0.34)
Between-language bigram frequency	0.67 (0.23)	0.72 (0.21)	0.67 (0.28)	0.63 (0.27)
Orthographic neighborhood	1.85 (0.28)	1.82 (0.25)	1.55 (0.24)	1.48 (0.24)

Values reported are means with standard deviation in parentheses for word frequency (Zipf scale), word length (number of letters), within-language bigram frequency (percentage per million), between-language bigram frequency (percentage per million), and orthographic neighborhood, measured through the average orthographic distance to the 20 nearest neighbors to indicate neighborhood density (OLD20).

2.3. Virtual Reality Setting and Apparatus

The setting of the experiment was created using VR via a head-mounted display (HMD). The items were presented in a 3D open street residential neighborhood that served as the main scenario due to its high quality of realism and the familiarity and openness of the simulated space. In this environment, the participants could experience simulated weather in a more realistic context (see Figure 1 and Supplementary Materials). Model editions and 3D model implementation to the main environment were made with the Vizard inspector (Worldviz, 2019). It was used to remove redundant 3D objects and to integrate a white canvas for the presentation of the experimental materials. Lastly, ambient sounds were added to the VR setting to improve the immersiveness in both weather conditions. This entailed rain sounds for the rainy condition and sounds of a fountain and pigeons for the sunny condition. Additionally, the background sky was animated to comply with both weather conditions (see Supplementary Materials for a video demonstration). Finally, to induce visual masking, a static Gaussian visual noise mask was superimposed on the white canvas in the appropriate trials. The use of visual noise followed standard methods for perceptual degradation and cognitive disfluency manipulations (e.g., Steindorf et al., 2023).



Figure 1. Illustration of the main scenario from the participant's viewpoint under sunny (left image) and rainy (right image) weather conditions.

Python 2.7 (Van Rossum & Drake, 1995) and Vizard 6 (Worldviz, 2019) were used to program and design the VR task. The 3D environments, including all experiment-related content, were displayed through an HTC VIVE Pro HMD at a rendering resolution of 2880×1600 pixels (1440×1600 pixels per eye). The built-in display of the headset provides a 90 Hz refresh rate as well as a 110° field of view. Participants' viewpoint was

continuously anchored throughout the experiment regardless of changes in their position in the real world.

2.4. Task and Procedure

Participants were provided with the HMD while seated on a rotating chair to immerse them in the abovementioned 3D virtual setting, enabling a full 360° view from a stationary perspective. After the placement and calibration of the headset, participants were equipped with two controllers, simulating two hands in the virtual setting. They were then presented with the instructions for the language decision task on a floating canvas. The instructions informed the participants of the two stages of the experiment: a practice phase followed by an experimental phase. Participants were asked to assign the linguistic stimuli to either Spanish, by pressing the button on the left controller, or English, by pressing that on the right controller. They were advised to react to the items as fast as they could and, hence, to rely on initial impressions. The target items were displayed centrally in black Courier New monospaced font on a simulated white canvas, ensuring readability. Each trial started with a central fixation cross presented for 500 ms, immediately followed by the target string, which remained visible for a maximum of 3000 ms or until a response was registered (see Figure 2). Subsequently, an inter-stimulus interval blank space was visible for 500 ms before the next trial began.

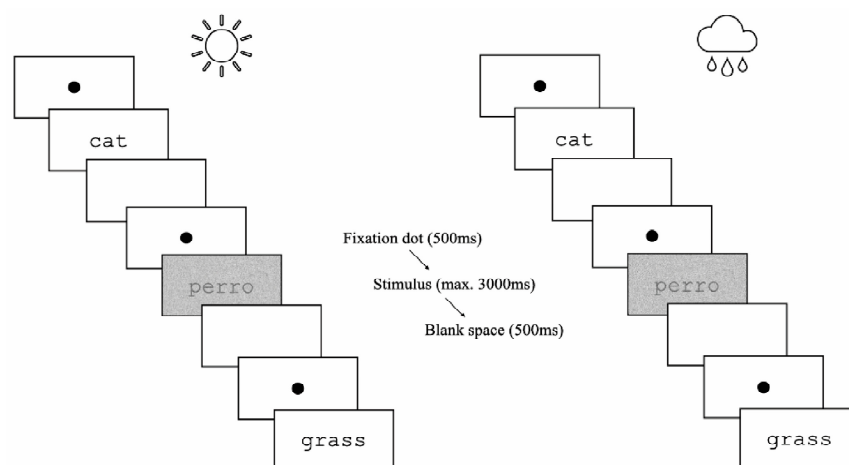


Figure 2. Representation of the structure of three sequential trials in the sunny and rainy blocks. Unmasked and masked conditions and English and Spanish stimuli are shown for illustration purposes.

Participants started in one of the weather conditions (rainy or sunny), which was randomly assigned, and completed all 120 trials from the first list of words, comprising both Spanish and English words, with or without a visual mask superimposed (i.e., 30 items per condition in each weather context). After completion of the first block corresponding to one weather context, they were granted a 5 min break, and they were subsequently presented with the other weather conditions with a new set of 120 words.

3. Results

Data were preprocessed and cleaned using R 4.3 (R Core Team, 2022) within the RStudio environment (RStudio Team, 2022). Reaction times (RTs) that fell outside ± 2.5 standard deviations from the mean RT per participant and condition were excluded, following standard practice in psycholinguistics (Baayen & Milin, 2010; Ratcliff, 1993). This method aims to remove extreme responses, whether too fast (e.g., anticipations) or too slow (e.g., lapses in attention), without biasing results toward fixed thresholds. As a result, 3.21% ($n = 136$) of trials were removed in the rainy condition and 3.12% ($n = 133$) in the sunny condition.

All excluded outliers were from the upper tail of the RT distribution. Additionally, trials with incorrect or missing responses were excluded from the RT analyses, yielding 4098 and 4127 valid observations for rainy and sunny weather conditions, respectively. The percentage (%) of correct responses per participant during the language decision task was determined as accuracy. Exploratory analyses revealed that the estimated average likelihood of accuracy rates for words was nearly at ceiling and very similar across conditions (see Table 2). Due to this, only reaction time data were analyzed.

Table 2. Descriptive analysis of accuracy proportions and reaction times (in milliseconds) across language, mask, and weather conditions. Means are reported with standard deviations in parentheses.

Language	Mask Condition	Weather Condition	Accuracy M (SD)	Reaction Time M (SD)
Spanish	Masked	Rainy	0.94 (0.23)	636 (192)
		Sunny	0.95 (0.21)	623 (183)
	Unmasked	Rainy	0.96 (0.20)	599 (194)
		Sunny	0.97 (0.17)	588 (161)
English	Masked	Rainy	0.96 (0.19)	625 (174)
		Sunny	0.97 (0.18)	613 (151)
	Unmasked	Rainy	0.96 (0.20)	618 (171)
		Sunny	0.96 (0.20)	600 (166)

Linear mixed-effects models were used to analyze RT data in Jamovi 2.3 ([The Jamovi Project, 2022](#)), using the GAMLj module ([Gallucci, 2019](#)). The model included language (Spanish, English), mask (masked, unmasked), and weather (rainy, sunny) as fixed effects, along with all two- and three-way interactions. Random intercepts were included for both participants and items. The model formula (R notation) was as follows: $RT \sim \text{Language} \times \text{Mask} \times \text{Weather} + (1 | \text{Subject}) + (1 | \text{Item})$. Initial models tested all possible two- and three-way interactions and more complex random structures; however, their inclusion did not substantially improve model fit, as assessed by Akaike Information Criterion (AIC) comparisons. Nevertheless, their inclusion did not result in a significant detriment to model performance. For this reason, the full factorial model was retained and reported to ensure transparency and theoretical completeness. Importantly, more parsimonious models yielded qualitatively equivalent results, reinforcing the robustness of the findings.

A significant main effect of mask was observed, $F(1, 224) = 32.37, p < 0.001$, indicating that masked stimuli elicited slower responses than unmasked ones (21 ms difference). The weather effect was also significant, $F(1, 8030) = 18.66, p < 0.001$, with slower responses under rainy conditions (14 ms difference). No significant main effect of language was found, $F(1, 224) = 0.55, p = 0.459$.

The interaction between language and mask was significant, $F(1, 224) = 10.01, p = 0.002$ (see Figure 3). Post hoc comparisons (Bonferroni corrected) revealed that Spanish masked words elicited significantly slower responses than Spanish unmasked words (33 ms difference), $t(226) = 6.25, p < 0.001$. In contrast, no significant difference was found between English masked and unmasked words ($t(223) = 1.79, p = 0.449$). While Spanish and English masked words did not differ ($t(225) = 1.71, p = 0.529$), Spanish unmasked words were processed faster than English unmasked ones (15 ms difference), $t(224) = -2.76, p = 0.037$.

The interactions between language and weather, $F(1, 8027) = 0.002, p = 0.963$, mask and weather, $F(1, 8026) = 0.029, p = 0.864$, and the three-way interaction language and mask and weather, $F(1, 8027) = 0.003, p = 0.958$, were not significant.

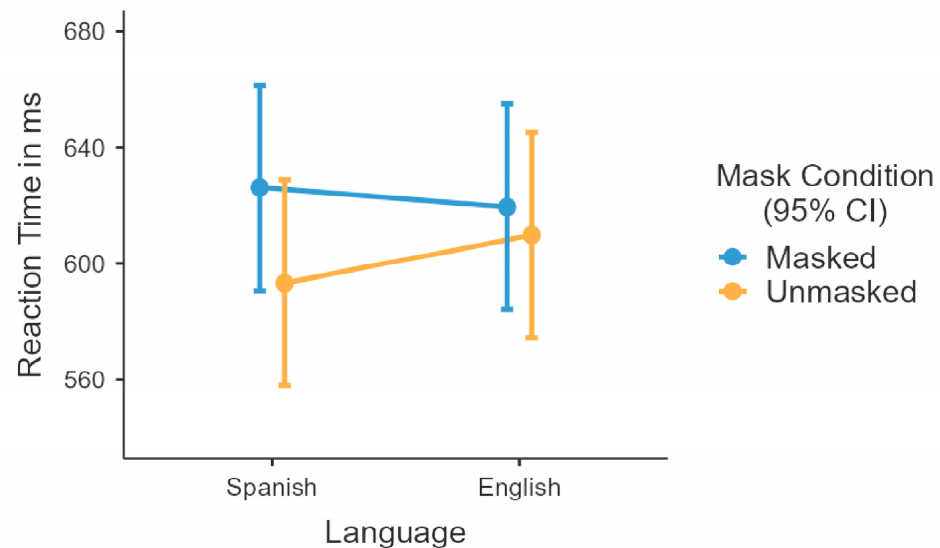


Figure 3. Interaction effect between language and mask in the reaction time analysis. Bars represent 95% confidence intervals.

4. Discussion

The current study aimed to explore the differences in cognitive processing elicited by the almost simultaneous retrieval of information from both the L1 and L2, as well as by simulated real-world perceptual disfluency conditions—operationalized through VR-induced meteorological scenarios and visual stimuli degradation via masking. Specifically, this study sought to determine the extent to which L2 processing might be conceptualized as a form of cognitive disfluency, comparable to that elicited by physical distortions typically used to manipulate processing fluency. To this end, we examined how visual challenges, whether derived from naturalistic weather simulations or artificial laboratory masks, affect cognitive load during a bilingual language decision task. By leveraging the immersive capabilities of VR, the present research aimed to provide a more ecologically valid framework for investigating the interplay between environmental context and bilingual language processing. In doing so, it revisited foundational questions in bilingual cognition through the lens of a naturalistic and controlled methodology that more closely reflects real-world language use.

Contrary to the initial hypothesis that L2 reading would inherently require more cognitive effort, similar to a form of perceptual disfluency, no significant main effect was found for language on response times in the language decision task. The additional cost associated with L2 reading was exclusively observed under conditions in which the target items were not visually distorted by masking. Put differently, the language effect exclusively emerged in conditions of intact presentation of the stimuli, in line with preceding reports of such an effect. In other words, L2 processing only required more cognitive effort than L1 processing when items were presented without a pattern mask. These findings nuance prior evidence linking L2 reading with higher processing demands (Cop et al., 2017; Koda, 2005) by showing that such effects may be contingent on stimulus clarity. When fluent processing is disrupted by perceptual manipulation, such as visual masking, the cognitive load typically associated with L2 processing appears less distinct.

The significant effects of visual masking, which led to longer response times compared to unmasked conditions, are consistent with research on perceptual disfluency (Alter et al., 2007) and reinforce the idea that visually degraded stimuli require greater cognitive effort. Similarly, the simulated rainy weather condition resulted in longer response times than the sunny condition (see Rocabado et al., 2024; Rocabado & Duñabeitia, 2024), supporting

the notion that environmental manipulations in VR can effectively simulate real-world perceptual disfluency. Importantly, while weather conditions had a general effect on performance regardless of language, the masking condition revealed an interaction with language, indicating differential effects. These results suggest that perceptual disfluency, whether naturally or artificially induced, increases cognitive load (Oppenheimer & Frank, 2008) but not uniformly across contexts or linguistic conditions. A complementary, though tentative, explanation is that individuals may be more perceptually adapted to the visual irregularities introduced by natural phenomena like rainfall, which could explain why its effects appear to be more uniform and less tied to language-specific processing.

Critically, the significant interaction between language and mask indicates that different forms of disfluency operate at distinct levels, with unique impacts on language processing. Specifically, masked Spanish words led to longer response times than masked English words. The effect of masking increased response times by approximately 33 ms in Spanish and by a negligible time of 9 ms in English, suggesting that the perceptual difficulty introduced by masking overrides linguistic advantages, effectively equalizing the processing demands of L1 and L2 and mainly affecting the native language. This increased perceptual difficulty may necessitate the engagement of System 2 processing—a more deliberate and effortful cognitive system (Kahneman, 2011)—for both languages, thereby diminishing the typical advantages of L1 processing under masked conditions. Note that System 2 is expected to be the system operating in L2 processing as a default, and thus it comes as no surprise that the only language sensitive to a system change is L1. Moreover, the presence of a language and mask interaction suggests that participants were not merely adopting an “L1/not-L1” decisional heuristic. If decisions had relied on a strategy of detecting L1 membership and rejecting everything else as “not L1”, the language-specific modulation of the visual masking would not have been expected. Supporting this, previous findings by Rocabado et al. (Rocabado et al., 2024, Exp. 1) using a monolingual lexical decision task showed that weather-induced disfluency had no significant effects on pseudowords, which would represent the stimuli most compatible with a “not-L1” strategy. In sharp contrast, the effects on word processing were robust. Thus, a “not-L1” decisional criterion would have predicted a reduced and possibly negligible weather effect with L2 words in the current study, and the results do not support such a view.

In contrast, the effect of rain was additive and did not interact with language. Rainy conditions increased response times by approximately 12 ms in Spanish and 15 ms in English. These results replicate our previous findings and confirm that naturalistic visual disfluency affects L1 and L2 similarly. The lack of interaction indicates that environmental disfluency of this type imposes a general increase in cognitive load rather than modulating language-specific processing (see Rocabado et al., 2025a for recent evidence of similar results). This distinction supports the broader theoretical implication that not all disfluencies operate equally: while some, like masking, selectively disrupt L1 processing, others, such as weather-related degradation, impact all linguistic input equally. Such dissociations reveal important boundaries in how perceptual and linguistic systems interact.

These results align with broader literature describing how increased perceptual difficulty and linguistic interference can differentially impact language processing under challenging conditions. For instance, Van Engen and Bradlow (2007) showed that speech recognition is more adversely affected by background noise in the same language as the target compared to a different language, suggesting that linguistic interference from a familiar language can amplify perceptual challenges. Similarly, Ng et al. (2015) found that native language background noise had a stronger masking effect on memory for speech than non-native background noise, indicating that semantic interference from the native language intensifies processing demands. However, Arslan et al. (2022) found that under

conditions of severe perceptual degradation—as with cochlear implant users attempting to distinguish high-pass-filtered musical sounds—linguistic distinctions become less relevant, as extreme perceptual limitations reduced participants' ability to discern between stimuli, regardless of language. These findings collectively suggest that while linguistic factors can indeed heighten word language recognition, conditions where the quality of the linguistic material is compromised, such as classical laboratory masking conditions where the quality of the stimuli is markedly degraded, can override these language-based effects at higher rates compared to recreating natural environmental conditions. In our study, this was evident as masking-induced perceptual disfluency reduced the typical processing advantage of the L1, effectively leveling the cognitive load across languages. This supports the idea that under highly challenging perceptual conditions, native language processing demands can approach those of the L2, illustrating the complex interplay between perceptual and linguistic factors. This reinforces the hypothesis that perceptual challenges can neutralize the usual cognitive benefits associated with L1 processing, particularly when perceptual clarity is severely compromised. In such cases, the typical cognitive cost associated with L2 processing—often attributed to disfluency in lexical access or decision-making (Segalowitz, 2010; Costa et al., 2019)—becomes less distinguishable, as both languages are processed under conditions that force a shift toward more effortful, System-2-like mechanism reasoning (Kahneman, 2011).

Beyond these empirical findings, the use of VR in this study demonstrates its potential for cognitive research. VR allows researchers to simulate ecologically valid environments while maintaining experimental control, see Rocabado et al. (2025b) for a recent demonstration. Our study exemplifies how VR can be leveraged to investigate language processing in contexts that closely approximate everyday challenges. This approach aligns with growing trends in cognitive science that emphasize naturalistic paradigms and complements prior work showing how embodied, situated cognition can inform theoretical models (Rocabado et al., 2024).

The present study has several limitations that warrant consideration. First, although our design aimed to simulate realistic disfluency conditions, it focused on a relatively constrained task: language decision using unmarked words. Future research should explore whether these effects generalize to more complex linguistic materials, such as sentences, narratives, or interactional speech. Second, while the use of unmarked words reduced orthographic bias, it may also have elevated cognitive demands by requiring decisions based on higher-level lexical–semantic processing rather than perceptual cues (Lecerf et al., 2024; Duñabeitia et al., 2020). This may partly explain the absence of a main effect of language. Future studies could directly manipulate orthotactic markedness to examine how surface features facilitate or hinder bilingual language processing under disfluent conditions. Third, although word familiarity was not directly assessed, this variable was not deemed critical for the aims of the current study. The stimuli were deliberately selected to be lexically simple, unmarked, and visually balanced across conditions, minimizing variability related to prior exposure. Nonetheless, future studies could explicitly explore how familiarity may influence processing ease by manipulating it. Fourth, although the weather manipulation was intended to induce perceptual disfluency, previous research has shown that mood can modulate affective word processing (Akram et al., 2020; Beedie et al., 2005) and that weather conditions are linked to fluctuations in emotional state (Keller et al., 2005; Kööts et al., 2011). However, in a recent study using the same simulated environments, weather conditions did not significantly alter the emotional evaluation of valenced words (Rocabado & Duñabeitia, 2024), suggesting that emotional state is unlikely to have been a major driver in the present findings. Moreover, the randomized block design likely minimized any systematic mood differences across conditions. Even so,

future studies would benefit from incorporating manipulation checks of mood or arousal to empirically disentangle emotional influences from perceptual ones. Finally, our sample size may have shown limited power to detect subtle effects. Larger samples and complementary paradigms (e.g., EEG, pupillometry) could offer a more granular understanding of how disfluency and language interact.

In conclusion, the present findings provide evidence that L2 processing, though more effortful under ideal visual conditions, becomes indistinguishable from L1 processing when perceptual disfluency is introduced. The differential effects of masking and weather highlight the importance of distinguishing between artificial and natural sources of disfluency in bilingual language processing. The use of VR enhances ecological validity and offers new avenues for studying how everyday environments modulate cognitive load. These findings contribute to both theoretical debates and methodological innovation in bilingualism research, paving the way for future work examining how perceptual, linguistic, and contextual factors jointly shape language comprehension.

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://osf.io/5rgwj> (accessed on 4 November 2024).

Author Contributions: Conceptualization: J.A.D. and F.R.; Data curation: F.R.; Formal analysis: F.R. and J.A.D.; Funding acquisition: J.A.D.; Methodology: J.A.D. and F.R.; Project administration: F.R. and G.S.; Resources: J.A.D.; Supervision: J.A.D.; Validation: J.A.D.; Writing—original draft: G.S. and F.R.; Writing—review and editing: G.S., F.R. and J.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partially funded by grant PID2021-126884NB-I00 by the MCIN/AEI/10.13039/501100011033.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Nebrija University (approval code: UNNE-2022-0017, 2 December 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data supporting this study can be found at the following link: <https://osf.io/5rgwj> (accessed on 4 November 2024).

Acknowledgments: The authors express their gratitude to all participants from Nebrija University for their essential contribution to this study. Their involvement was crucial to making this research possible.

Conflicts of Interest: The authors declare no conflicts of interest to disclose.

References

- Akram, U., Drabble, J., Cau, G., Hershaw, F., Rajenthiran, A., Lowe, M., Trommelen, C., & Ellis, J. G. (2020). Exploratory study on the role of emotion regulation in perceived valence, humour, and beneficial use of depressive internet memes in depression. *Scientific Reports*, *10*, 899. [CrossRef] [PubMed]
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*, 569–576. [CrossRef]
- Arslan, N. Ö., Akbulut, A. A., Köse, B., Karaman-Demirel, A., & Derinsu, U. (2022). Sound quality perception of cochlear implant recipients: Low-frequency information and foreign-language effect. *International Journal of Audiology*, *61*, 1045–1053. [CrossRef]
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychology Research*, *3*, 12–28. [CrossRef]
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, *39*, 445–459. [CrossRef] [PubMed]
- Beedie, C., Terry, P., & Lane, A. (2005). Distinctions between emotion and mood. *Cognition and Emotion*, *19*, 847–878. [CrossRef]
- Bernhardt, E. (2010). *Understanding advanced second-language reading*. Routledge. ISBN 978-0-203-85240-8.

- Casaponsa, A., & Duñabeitia, J. A. (2016). Lexical organization of language-ambiguous and language-specific words in bilinguals. *Quarterly Journal of Experimental Psychology*, 69, 589–604. [CrossRef]
- Cattell, J. M. (1948). Experiments on the association of ideas, 1887. In *Readings in the history of psychology* (pp. 329–335). Century Psychology Series. Appleton-Century-Crofts.
- Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An Eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*, 49, 602–615. [CrossRef]
- Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE*, 10, e0134008. [CrossRef]
- Costa, A., Duñabeitia, J. A., & Keysar, B. (2019). Language context and decision-making: Challenges and advances. *Quarterly Journal of Experimental Psychology*, 72, 1–2. [CrossRef]
- Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience*, 15, 336–345. [CrossRef] [PubMed]
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175–197. [CrossRef]
- Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-Stop shopping for spanish word properties. *Behavior Research Methods*, 45, 1246–1258. [CrossRef]
- 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, 1691. [CrossRef]
- Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word processing: Evidence from the semantic simon paradigm. *Psychonomic Bulletin & Review*, 15, 961–966. [CrossRef]
- Eichert, N., Peeters, D., & Hagoort, P. (2018). Language-driven anticipatory eye movements in virtual reality. *Behavior Research Methods*, 50, 1102–1115. [CrossRef]
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. [CrossRef]
- Gallucci, M. (2019). *GAMLj: General analyses for the linear model in jamovi*. Available online: https://gamlj.github.io/gamlj_legacy/ (accessed on 4 November 2024).
- Grotek, M., & Ślęzak-Świat, A. (2024). The perceived and measured difficulty of texts and tasks in L1 and L2. *Reading in a Foreign Language*, 36, 1.
- Hasson, U., Egidi, G., Marelli, M., & Willems, R. M. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition*, 180, 135–157. [CrossRef] [PubMed]
- Heyselaar, E., Peeters, D., & Hagoort, P. (2021). Do we predict upcoming speech content in naturalistic environments? *Language, Cognition and Neuroscience*, 36, 440–461. [CrossRef]
- Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 507–513. [CrossRef]
- James, W. (1890). *The principles of psychology* (Vol. II, pp. 188–205). Henry Holt and Company.
- Kahneman, D. (2011). Thinking, fast and slow. In *Thinking, fast and slow* (p. 499). Farrar, Straus and Giroux. ISBN 978-0-374-27563-1.
- Keller, M. C., Fredrickson, B. L., Ybarra, O., Côté, S., Johnson, K., Mikels, J., Conway, A., & Wager, T. (2005). A warm heart and a clear head: The contingent effects of weather on mood and cognition. *Psychological Science*, 16, 724–731. [CrossRef]
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge Applied Linguistics. Cambridge University Press. ISBN 978-0-521-54513-6.
- Kööts, L., Realo, A., & Allik, J. (2011). The influence of the weather on affective experience. *Journal of Individual Differences*, 32, 74–84. [CrossRef]
- Lecerf, M.-A., Casalis, S., & Commissaire, E. (2024). New Insights into bilingual visual word recognition: State of the art on the role of orthographic markedness, its theoretical implications, and future research directions. *Psychonomic Bulletin & Review*, 31, 1032–1056. [CrossRef]
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of english. *Behavior Research Methods*, 44, 325–343. [CrossRef]
- Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, 32, 533–550. [CrossRef]
- Léwy, N., & Grosjean, F. (2008). The Léwy and Grosjean BIMOLA model. In *Studying bilinguals* (pp. 201–210). Oxford University Press. ISBN 978-0-19-928128-2.
- 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, 326–333. [CrossRef]
- Mulligan, N. W. (1996). The effects of perceptual interference at encoding on implicit memory, explicit memory, and memory for source. *The Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1067–1087. [CrossRef]

- Nairne, J. S. (1988). The mnemonic value of perceptual identification. *The Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 248–255. [CrossRef]
- Ng, E. H. N., Rudner, M., Lunner, T., & Rönnerberg, J. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear and Hearing*, 36, 82. [CrossRef]
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, 12, 237–241. [CrossRef]
- Oppenheimer, D. M., & Frank, M. C. (2008). A Rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, 106, 1178–1194. [CrossRef]
- Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6, 351–371. [CrossRef]
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510–532. [CrossRef]
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rocabado, F., Alonso-Bernal, N., & Duñabeitia, J. A. (2025a). Word recognition during movement under simulated weather conditions. *PLoS ONE*, 20, e0326945. [CrossRef]
- Rocabado, F., & Duñabeitia, J. A. (2022). Assessing inhibitory control in the real world is virtually possible: A virtual reality demonstration. *Behavioral Sciences*, 12, 444. [CrossRef]
- Rocabado, F., & Duñabeitia, J. A. (2024). Clouded judgments? The role of virtual weather in word valence evaluations. *Cognition and Emotion*, 1–11. [CrossRef]
- Rocabado, F., González Alonso, J., & Duñabeitia, J. A. (2022). Environment context variability and incidental word learning: A virtual reality study. *Brain Sciences*, 12(11), 1516. [CrossRef]
- 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, 1433781. [CrossRef]
- Rocabado, F., Muntini, L., Jubran, O. F., Lachmann, T., & Duñabeitia, J. A. (2025b). Transforming language research from classic desktops to virtual environments. *Scientific Reports*, 15, 23118. [CrossRef]
- RStudio Team. (2022). *RStudio: Integrated development environment for R*. RStudio Team.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge. ISBN 978-0-203-85135-7.
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54, 79–95. [CrossRef]
- Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University Press. ISBN 978-0-19-515177-0.
- Shin, Y. S., Masis-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2021). Context-dependent memory effects in two immersive virtual reality environments: On Mars and underwater. *Psychonomic Bulletin & Review*, 28, 574–582. [CrossRef]
- Steindorf, L., Pink, S., Rummel, J., & Smallwood, J. (2023). When there is noise on sherlock holmes: Mind wandering increases with perceptual processing difficulty during reading and listening. *Cognitive Research: Principles and Implications*, 8, 31. [CrossRef]
- Stowe, L. A., & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching*, 43, 329–353. [CrossRef]
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285. [CrossRef]
- The Jamovi Project. (2022). *Jamovi* (Version 2.3) [Computer software]. The Jamovi Project.
- Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 12530–12535. [CrossRef]
- Titus, A., Dijkstra, T., Willems, R. M., & Peeters, D. (2024). Beyond the tried and true: How virtual reality, dialog setups, and a focus on multimodality can take bilingual language production research forward. *Neuropsychologia*, 193, 108764. [CrossRef]
- Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, 121, 519–526. [CrossRef]
- Van Rossum, G., & Drake, F. L., Jr. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.
- Vega-Mendoza, M., Hansson, P., Sörman, D. E., & Ljungberg, J. K. (2021). Testing the foreign language effect on cognitive reflection in older adults. *Brain Sciences*, 11, 1527. [CrossRef]
- Worldviz. (2019). *Vizard* (Version 6.0) [Computer software]. Worldviz.
- Wu, Y. J., & Thierry, G. (2010). Chinese–English bilinguals reading english hear chinese. *The Journal of Neuroscience*, 30, 7646–7651. [CrossRef] [PubMed]
- Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—A desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, 41, 229–241. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.