Universitat Pompeu Fabra Barcelona

Treball de fi de màster de recerca

Cognate Status and Dialectal Variation in Bilingual Spoken Word Recognition: The Role of Language Dominance in Catalan–Spanish Listeners

Nom i Cognoms Pau Ridameya Jan

Màster: Lingüística Teòrica i Aplicada

Edició: 2024-2025

Directors: Dr. Miquel Llompart Garcia

Any de defensa: 2025

Col·lecció: Treballs de fi de màster

Departament de Traducció i Ciències del Llenguatge

Abstract

This study investigates how cognate status, dialectal variation, and language dominance influence spoken word recognition in Catalan–Spanish bilinguals. Thirty participants completed a lexical decision task in which they listened to Catalan words presented in either Central Catalan (their native dialect) or Valencian (a non-native dialect), and categorized them as real words or not. The stimuli included both cognates and non-cognates with Spanish. Reaction times were analyzed using linear mixed-effects models, with language dominance measured continuously using the Bilingual Language Profile. Results showed that dialectal variation and language dominance modulated the effect of cognate status: while no main effect of cognates emerged in Central Catalan, a significant cognate disadvantage was found in Valencian, especially among Spanish-dominant participants. These findings suggest that cross-linguistic phonological similarity can hinder rather than facilitate recognition when combined with dialectal unfamiliarity, and highlight the importance of individual linguistic profiles in bilingual lexical access. The study contributes to a more nuanced understanding of how phonological variation across dialects interacts with bilingual processing mechanisms.

Keywords: Bilingualism, Spoken word recognition, Language dominance, Phonetics, Dialectal variation, Cognate effect, Catalan, Cross-linguistic interference, Phonological processing

Table of Contents

I. Introduction:

Spoken word recognition is a central topic of discussion in language processing, psycholinguistics and cognitive science, as it reveals how the human brain processes language in real time and manages the complexities of mapping variable acoustic signals onto stable lexical representations. In real-time conversation, listeners must rapidly and accurately locate an auditory input onto lexical representations stored in memory. This has been proved to happen before the entire word has been pronounced. This process is known as spoken word recognition, as said before, and involves the constant activation of word candidates, a process of competition between them and the final selection of the expected correct word as speech unfolds. This is already a complex cognitive task for monolingual speakers but it becomes more complicated (at a cognitive level) for bilingual speakers. As discussed by Grosjean (1989), bilingual brains form a unique and dynamic system rather than a simple sum of two independent monolingual systems. Bilinguals process and access their languages in interactive and context-dependent ways, often activating both languages to varying degrees. In spoken word recognition, that implies that candidates of both languages are constantly active when mapping the initial acoustic input, since both linguistic systems are constantly active and in competition with each other. This perspective aligns with the Language Mode Framework (Grosjean, 2001), which posits that bilinguals operate along a continuum of activation for each language depending on contextual cues. In highly bilingual environments, such as Catalonia, both languages can remain constantly active, reinforcing the likelihood of cross-linguistic interference.

More precisely, a large body of psycholinguistic research has shown that lexical access is often non-selective, which means that in bilinguals, both languages are constantly activated during speech perception and recognition, even if only one language is being spoken (Marian & Spivey, 2003). That is, when a bilingual hears a word in a language, similar-sounding words also become active and potential candidates which enter in competition with the other active candidates. As expected, this increases when the words sound similar, meaning that if the two languages share phonetic features, it is more probable to have words of the other languages being activated. For instance, this cross-linguistic overlap and competition is more probable to happen between Spanish and Catalan than Spanish and Chinese.

Among the different factors shaping this activation, language dominance plays a crucial role. As such, studying bilingual spoken word recognition offers valuable insights not only into the structure of the bilingual mental lexicon but also into how the human brain manages competing linguistic systems in real time. A widely accepted model that captures these dynamics is the Bilingual Interactive Activation Plus (BIA+) model (Dijkstra & Van Heuven, 2002). This model explains that bilinguals do not have two dictionaries within their heads, one for each language, but rather one unique shared system where the lexicons of both languages coexist within a shared mental lexicon. As said before, different factors make the brain choose one word or another between the activated possible candidates. These factors are:

- *Frequency* → The brain tends to activate with more strength more frequent words
- *Phonological overlap* → More possibilities for a word to be activated if the phonemes correspond
- *Context* → Even if words from both languages are activated, if speaking in Catalan, for instance, Catalan words have more probabilities of being activated
- *Linguistic dominance* → If the listener is more used to use Spanish, for instance, Spanish words may be activated with more strength

Bilinguals who are more dominant in Spanish, for example, may more strongly activate Spanish lexical candidates even when processing Catalan input, which can lead to increased competition or misrecognition (Simonet, 2014). This effect is particularly pronounced when Catalan lexical items resemble their Spanish equivalents in form or phonetic realization, such as with cognates, which will be discussed in the next subsection.

Complementing the BIA+ model (Dijkstra & Van Heuven, 2002), Green (1998) speaks about another cognitive task that happens during speech recognition, inhibition. He includes it in his model, the Inhibitory Control (IC) Model to explain how bilinguals manage the two languages in their mind. According to this model, both languages are often active at the same time, and the brain needs to actively suppress the one that is not being used. This process is called inhibition. It happens especially when there is competition between similar words in both languages, like cognates. The IC model also says that a part of our brain helps us stay focused on the language we want to use, especially during tasks like translation or word recognition. This inhibitory control mechanism can be weakened, due to cognitive load, fatigue, age, or low executive functioning which can make the non-target language interfere more heavily. This can result in slower or less accurate recognition of Catalan words, particularly in contexts involving high phonological similarity or low lexical frequency. This non-inhibition ability may be especially increased in "diglossic" environments like the metropolitan area of Barcelona, where Catalan and Spanish are both widely used, often interchangeably, in everyday communication and where Spanish tends to be the dominant one. As described by Grosjean (2001), such contexts place speakers in a sustained bilingual mode, in which both languages remain highly active, amplifying cross-linguistic interference during lexical access. Because Spanish lexical forms are constantly present and activated in such settings, bilinguals may show stronger phonetic convergence towards Spanish in both production and perception (Simonet, 2014). Blumenfeld and Marian (2007) highlighted that higher executive control (inhibitory capacity) in bilinguals is associated with better suppression of non-target language candidates, allowing for more efficient lexical access. This supports the view that not all bilinguals will experience the same degree of interference or facilitation, even in the same task, and justifies the inclusion of individual-level cognitive variables.

Finally, another model that helps explain bilingual word processing is the Revised Hierarchical Model (RHM) by Kroll and Stewart (1994). This model focuses more on word production than on recognition, and suggests that bilinguals have stronger connections between words and meanings in their dominant language. In early stages of learning, people tend to access the meaning of L2 words by going through the L1 equivalent. As proficiency grows, this dependency decreases, but language dominance still plays a role. Later studies (Kroll et al., 2010) confirmed that both languages can be active at the same time, and that dominant-language words can interfere with word production in the non-dominant language. These models show that even when words are similar in form, processing them is not always easier, especially if the pronunciation triggers stronger activation of the dominant language, causing interference instead of facilitation.

Nevertheless, several studies show that cognates, words that share form and meaning in both languages facilitate recognition. For example, Spivey and Marian (1999) demonstrated through eye-tracking that bilinguals look at objects whose names sound similar in both languages, showing that phonological candidates from both lexicons are activated. To do that, Russian–English bilinguals were asked to follow spoken instructions in Russian while viewing a set of objects on a screen. When they heard a Russian word like *марка* [mářka], meaning *stamp*, they were more likely to also fixate on a picture of a marker, a word that sounds similar in English but is unrelated in meaning. This showed that even irrelevant words from the non-target language, in this case English, were partially activated based on phonological overlap. The use of eye-tracking allowed the researchers to observe this activation in real time, before any verbal response was given. This evidence supports the idea that bilinguals do not fully suppress their other language, and that lexical candidates from both languages are active and competing even when only one language is being used (Marian & Spivey, 2003). Using the same methodology, Weber and Cutler (2004) provided evidence of cross-linguistic competition in Dutch-English bilinguals, where English lexical candidates with a similar pronunciation than the Dutch target word were momentarily activated during Dutch speech processing.

Several studies have investigated how these effects unfold in bilingual lexical access. For example, Sherkina (2003) shows that bilinguals tend to recognize and produce cognates faster than non-cognates, especially in tasks involving picture naming and lexical decision, supporting the view that overlapping representations across languages boost access. Costa et al. (2000) tested Spanish-Catalan bilinguals across several production tasks to explore the cognate facilitation effect. In their main experiment, participants performed a picture naming task, where they had to name images in their second language (L2, Catalan) as quickly and accurately as possible. The stimuli included both cognates and non-cognates. They measured reaction times and error rates, finding that cognates were named faster and with fewer errors than non-cognates. This effect remained robust even when the task involved naming pictures in the first language (L1) or translating words from one language to another. Their results strongly support the idea that shared phonological and semantic representations across languages enhance lexical access in bilinguals. Dijkstra et al. (2010) further examined the interaction between orthographic and phonological similarity in bilingual visual and auditory tasks using a lexical decision task with Dutch-English bilinguals and showed that the extent of cognate facilitation depends on the degree of cross-linguistic similarity. Similarly, Lijewska (2020) compiles a series of findings supporting the cognate facilitation effect happens not just in listening, but also in reading and speaking tasks, suggesting that the reason for the cognate facilitation effect is that both languages activate similar meanings and sounds at the same time. That is why studies on bilingual speech recognition tend to include both cognates and non-cognates in their analyses and experimental designs to explore the extent to which phonological and semantic overlap between languages facilitates or interferes with lexical access. Cognates are particularly useful in revealing how bilinguals handle phonological similarity across systems, while non-cognates are useful to isolate language-specific effects. Including both types of items allows for a more precise understanding of cross-linguistic activation mechanisms, especially in contexts where formal overlap may trigger unintended lexical competitors.

On a separate but relevant front, dialectal variation has also been shown to impact speech processing. Larraza et. al. (2017), in a cross-dialectal study with Basque–Spanish bilinguals, found that even subtle dialectal mismatches, such as fricative mergers, can delay word recognition and increase the likelihood of misidentification, especially among speakers whose native dialect lacks the phonemic contrast present in the target variety. This phenomenon is not exclusive to Basque-Spanish bilinguals. Floccia et al. (2006) found that even native speakers of the same language show delays when processing unfamiliar regional accents. They conducted their study with English native speakers from the UK showing them and proved that phonetic variation within a language, even without any change in vocabulary or grammar, can interfere with word recognition. Sumner and Samuel (2009) showed that listeners adapt to dialectal variation over time, but accented input tends to elicit longer reaction times and reduced recognition accuracy in the initial phases of exposure. They did this by asking participants to perform a lexical decision task and a shadowing task (repeating words as quickly as possible after hearing them). Their results showed that in the early trials, responses to the unfamiliar accented speech were significantly slower and less accurate than to the standard accent. However, as exposure increased, participants' performance improved, suggesting perceptual adaptation. This study highlights how listeners need time to adjust to unexpected phonetic patterns, and that initial unfamiliarity with a dialect can hinder word recognition.

While these studies provide valuable insights into cognate effects and dialectal processing separately, to our knowledge, no research has yet explored how cognate facilitation might be modulated by dialectal variation. This gap is particularly relevant in bilingual communities where regional dialects differ phonetically in ways that interact with cross-linguistic similarity, as is the case in Catalan–Spanish bilinguals exposed to dialects with varying degrees of vowel reduction. The present study also complements a small but growing body of research examining how phonetic similarity across languages and dialects modulates lexical access in bilinguals. For example, Amengual (2016) found that exposure to Spanish phonology influenced the production of Catalan vowels among Catalan–Spanish bilinguals, suggesting that bilingual speakers form integrated phonological categories shaped by both languages. Although this study focuses on cross-linguistic influence rather than dialectal variation per se, it highlights how the interaction between a bilingual's two phonological systems can affect language production. This is particularly relevant in regions where dialectal forms in one language closely resemble phonological patterns in the other, as such overlap may impact recognition and processing.

In the context of Catalan and Spanish, another important variable is the dialectal variation within Catalan itself. Catalan exhibits notable differences across its dialects in the realization of unstressed vowels. In Central Catalan, for instance, stressed vowels [a], [e] and [ɛ] become [ə] in unstressed position and stressed vowels [o] and [ɔ] become [u] in unstressed position, undergoing vowel reduction. Vowels [i] and [u] remain the same in both stressed and unstressed positions. While much has been written about Catalan dialectology (IEC, 2016), fewer studies have linked it explicitly to bilingual lexical processing. The role of vowel reduction and final consonant pronunciation patterns, as seen in Central Catalan vs. Valencian, has potential implications not just for production but also for the perception of phonetic forms that align more or less with Spanish cognates.

In contrast, western Catalan (spoken from the western Pyrenees to the southern tip of the Valencian Country) maintains (to some degree) the original vowel quality in unstressed positions (Jiménez, 2019). In North-Western Catalan (spoken in the region of Lleida), it is common for the final unstressed /a/ to be raised or fronted, frequently realized as [ɛ] in informal speech. This feature is widespread and has even been incorporated into the spoken standard for this variety, as recognized by the Institut d'Estudis Catalans (IEC, 2016). As a result, many final unstressed /a/ vowels in words such as casa or Lleida may be pronounced as [kázɛ] and [ʎéjðɛ], introducing a surface form that is more distant from the Spanish equivalent, which preserves the [a] vowel in such positions.

In Valencian, another western catalan dialect, we find ourselves nevertheless with a much more restricted pattern of final /a/ modification. The final vowel only suffers a modification after post-tonic /ɛ́/ and /ɔ́/. Only in those contexts of harmony the vowel /a/ assimilates the tonic vowel and gets articulated in the same way. Therefore, cosa and terra become [kɔ́zɔ] and [tɛ́rɛ]. Even in these cases, the phenomenon is considered regionally marked and socially dispreferred, having been historically described as a "local defect" (Jiménez, 2019). Outside of these contexts, final /a/ in Valencian is typically preserved as [a], maintaining a strong phonetic similarity with its Spanish counterpart.

A. This Present Study

Building on these theoretical foundations, the current study aims to address that gap by testing whether the typical cognate advantage persists when words are presented in a dialect that diverges from the listener's own. Therefore, this study explores how bilingual Catalan-Spanish speakers who are used to the Central Catalan dialect recognize words that are presented either in their own dialect or in Valencian. By including both cognates and non-cognates, the experiment investigates whether cross-linguistic phonological similarity with Spanish affects lexical access, and whether this effect is modulated by language dominance. To account for individual variability in language use and experience, this study will assess participants' language dominance using the Bilingual Language Profile (Birdsong et al., 2012), a widely used and validated instrument that captures the essential information about language exposure, proficiency, usage, and age of acquisition. By incorporating this measure, the study can more accurately interpret differences in lexical activation and interference as a function of language dominance, rather than attributing them solely to phonological similarity.

This study focuses on how bilingual speakers of Catalan and Spanish, specifically those familiar with the Central Catalan dialect, recognize spoken words presented in a non-native Catalan variety, namely Valencian. While dialects such as North-Western Catalan or Valencian preserve more conservative vowel systems that more closely resemble Spanish phonology, the goal here is not to compare which dialect is easier to process. Instead, the aim is to investigate whether presenting words in a different dialect affects recognition, particularly when those words are cognates with Spanish. Participants will hear words in either their native dialect (Central Catalan) or in Valencian. Some of these words will be cognates (e.g., *proposta* - *propuesta*), while others will be non-cognates (e.g., trencar –

romper). Importantly, all stimuli will be in Catalan, and no Spanish words will be used. The study seeks to assess whether dialectal variation and cognate status interact during word recognition, and to what extent cross-linguistic phonological similarity with Spanish influences recognition—especially among Spanish-dominant bilinguals.

Although Valencian may show more lexical differences from Central Catalan than North-Western Catalan, which could complicate the selection of shared lexical items for the experiment, this issue can be resolved by modifying and adapting the stimulus list to guarantee lexical familiarity and functional equivalence. Crucially, when the aim is to enhance phonetic resemblance to Spanish, Valencian stands out as a better candidate. In addition to preserving the final unstressed /a/ as [a], Valencian also pronounces final /r/ in infinitive verbs, as in *votar* [boˈtáɾ], mirroring the Spanish realization. In contrast, in North-Western Catalan, the final /r/ is often dropped, resulting in surface forms such as *votar* pronounced as [boˈtá]. This further reduces the phonetic similarity between the Catalan and Spanish forms, particularly in verb recognition. By choosing lexical items that steer clear of vowel harmony conditions in Valencian, the list of words can be chosen specifically to ensure the phonetic similarity with Spanish and facilitating a more precise exploration of how cross-linguistic similarity influences spoken word recognition in bilingual individuals. In doing so, this study contributes to our understanding of how dialectal and cross-linguistic variation jointly shape lexical access in bilingual speech perception.

Based on previous findings, my hypothesis is that cognates presented in Central Catalan would show faster reaction times than non-cognates. When presented in Valencian, however, non-cognates are expected to produce longer reaction times, while cognates are expected to remain relatively unaffected due to their cross-linguistic similarity with Spanish.

Furthermore, language dominance is expected to modulate these effects: participants with stronger Spanish dominance may show increased facilitation for cognates and greater difficulty with non-cognates in Valencian especially, whereas more Catalan-dominant individuals may show less variation across conditions.

II. Methodology

As briefly explained before, this study investigates how dialectal variation and cognate status influence spoken word recognition in Catalan–Spanish bilinguals. The study involves participants completing an auditory word recognition task in which they listened to Catalan words and non-words presented in their own dialect (Central Catalan) or in a different dialect (Valencian). The stimulus set included both cognates and non-cognates, allowing for the examination of cross-linguistic phonological similarity effects. Participants' language dominance was also measured to explore individual differences in lexical access. The following sections describe the participants, materials, procedure, and analysis plan in detail.

A. Participants

Thirty Catalan–Spanish bilinguals were chosen to participate in the study. They were chosen ensuring two conditions were fulfilled: 1) that they were all born and raised in Barcelona and its metropolitan area, to ensure that their main Catalan dialect was Central Catalan, and 2) that 15 of them were more prone to use Spanish and the other 15 were more prone to use Catalan, in order to allow for an even distribution of language dominance across the sample. Gender was not a selection criterion, and participants of all gender identities were welcome, although no significant gender-related effects were expected or analyzed in this study.

Participant’s language dominance was assessed using the Bilingual Language Profile (BLP), a validated self-report questionnaire that measures language exposure, usage, proficiency, and age of acquisition. Based on their BLP scores, participants varied in their degree of Catalan and Spanish dominance, allowing for an exploration of individual differences in cross-linguistic activation during the word recognition task. All participants gave informed consent before beginning the study.

Finally, in order to minimize variability related to cognitive processing speed and language exposure across different life stages, participants were selected to be within a narrow age range of 18 to 30 years old. This age group was chosen because it represents young adults who have fully developed cognitive and linguistic systems, while also reducing the likelihood of age-related decline in processing speed or inhibitory control. Furthermore, limiting age helps control for generational differences in exposure to Catalan and Spanish, as sociolinguistic contexts in Catalonia and Barcelona have changed significantly over recent decades. This ensures that all participants were socialized in relatively similar linguistic environments during their formative years.

B. Materials

Regarding the materials needed for the experiment, 60 Catalan words were selected, equally divided into two categories: cognates (n = 30) and non-cognates (n = 30). As explained before, cognates were defined as words that share both form and meaning with their Spanish equivalents (e.g., *proposta – propuesta*), whereas non-cognates shared meaning but showed low or no formal similarity (e.g., *trencar – romper*). Special attention was given to the phonetic characteristics of the words to avoid phonological contexts that would trigger vowel harmony or vowel raising in Valencian, ensuring that final vowels remained stable. This helped preserve phonetic similarity with Spanish, when the words are pronounced in

Spanish, especially for cognate items, and allowed for a more precise examination of cross-linguistic phonological influence. These words were also selected to be familiar words in Catalan, avoiding low-frequency or regionally marked vocabulary. This was done drawing frequency information from the NIM corpus (Guasch et al., 2013). The mean relative frequency of the cognates was 45,403 per million tokens while the mean relative frequency of the non-cognates was 42,777 per million tokens.

To support the assumption that Valencian pronunciations exhibit greater phonetic similarity to Spanish than Central Catalan's, a phonological distance analysis was conducted using the Levenshtein distance metric. This measure calculates the minimal number of phoneme insertions, deletions, or substitutions required to transform one word into another, providing a quantitative estimate of phonetic proximity. In this case, Levenshtein distances were computed only for the 30 cognate items, as non-cognates are too different to compare across languages (in many cases, the distance was equivalent to the number of phonemes of the word itself). For each cognate, two distances were calculated: 1) between the Valencian pronunciation and its Spanish equivalent, and 2) between the Central Catalan pronunciation and the same Spanish equivalent. All words were transcribed phonetically prior to analysis. The mean Levenshtein distance between Valencian and Spanish pronunciations was 0,97, while the mean distance between Central Catalan and Spanish was substantially higher, at 2,17. This quantitative difference shows once again that Valencian pronunciation is phonetically closer to Spanish, which is really noticeable in the case of cognates, and therefore more likely to trigger cross-linguistic activation in Spanish-dominant bilinguals. All words can be found in Annex 1 with their respective transcriptions, length, spanish equivalent, frequencies and with the computed Levenshtein distances.

As for the non-words, they were generated using the pseudoword generator UniPseudo (Barra et al., 2024) in which the user can choose the length of the pseudowords they want, and create their pseudowords based on a pasted list of words (or phonological words) having the characteristics they want. In order for all words and non-words to be similar in number of letters and syllables, they were all chosen to contain between 4 and 7 letters while always consisting of two syllables. All non-words can be found in Annex 2 with their respective transcriptions and length.

Each word and non-word was recorded in two dialectal variants: Central Catalan and Valencian. Recordings were made by two native female speakers to ensure a similar voice, one from the Barcelona metropolitan area (Central Catalan) and one from the province of Valencia (Valencian). The 2 speakers were instructed to produce the words in a natural, neutral intonation, avoiding hyperarticulation or deliberate slowing. Recordings were carried out in an acoustically appropriate environment, ensuring clear, high-quality audio suitable for perceptual experiments. All audio files were normalized for amplitude and trimmed to remove background noise or unintended pauses. The overall duration of the words was preserved as naturally produced to maintain ecological validity.

C. Procedure

The experimental design included two within-subject factors: Dialect (Central Catalan vs. Valencian) and Cognate Status (Cognate vs. Non-cognate). Each participant was exposed to all experimental conditions, with stimuli presented in the two dialectal variants and including both cognate and non-cognate items The primary dependent variable was reaction time, measured from the onset of the auditory stimulus to the participant's lexical decision response. Accuracy was also recorded to ensure task compliance. Finally, and as explained before, individual differences in bilingual experience were accounted for by including language dominance scores, obtained through the Bilingual Language Profile (BLP), as a continuous covariate.

The experiment was designed using PsychoPy, and participants were tested individually in a quiet room using headphones connected to a laptop. Upon arrival, participants received an explanation of the task and were instructed to respond both quickly and accurately. The task was, as briefly explained before, an auditory lexical decision task in which participants listened to Catalan audio stimuli, consisting of both real words and non-words, pronounced in either Central Catalan or Valencian.

Each trial began with a fixation cross displayed on the screen, followed by the auditory presentation of a stimulus. Participants were asked to determine whether the word they had just heard was a real Catalan word or not. They were instructed to press the "Z" key if they believed the stimulus was a real word, and the "M" key if they believed it was an invented non-word. To ensure participants did not forget the keys they had to press, a green tick was displayed in the left downer corner of the screen, close to the "Z" key, and a red cross was displayed on the right downer corner of the screen, close to the "M" key. They were explicitly informed before starting that some of the words they would hear were invented words. The experimental block included 120 randomized trials: 60 real words (30 cognates, 30 non-cognates) and 60 non-words, each balanced across dialects.

To control word-specific effects, the assignment of dialect to each word was counterbalanced across participants, so that the same word was not always presented in the same dialect. This ensured that any observed effects could not be attributed to particular word/dialect pairings. To implement this counterbalancing, participants were assigned to Group 1 or 2. Each group received a different allocation of dialectal variants for the same items: for example, if Group 1 heard a word (e.g. *proposta*) in Central Catalan, Group 2 would hear the same item in Valencian. This ensured that each participant heard each lexical item only once, while across the whole sample, all items were tested in both dialects.

Reaction time was measured from the beginning of the audio stimulus to the participant's keypress. This was achieved using PsychoPy's internal timing functions, which recorded the interval between the beginning of the auditory stimulus and the participant's response key press, ensuring precise millisecond-level accuracy. Accuracy was also recorded. The influence of dialect (Central Catalan vs. Valencian) and cognate status (Cognate vs. Non-cognate) on processing speed was assessed, taking into account individual differences in language dominance, obtained with the BLP responses of participants, as a covariate. The Python script used for the PsychoPy experiment can be found in the link in Annex 3.

Results:

This section presents the results of the lexical decision task that the 30 participants took which examined the effects of cognate status, dialect, and language dominance on reaction times. First, descriptive statistics are provided to illustrate overall patterns in accuracy and reaction time. Then, results from the multiple regression analysis are reported to examine the interaction between dialect, cognate status, and dominance.

After each participant completed the lexical decision task in PsychoPy, a CSV file was automatically saved. Therefore, after collecting data from all 30 participants, 30 individual CSV files were obtained and saved. These files were then merged into a single dataset using Python. During this cleaning step, all non-word were filtered out in order to retain only lexical decision responses to real words. This was done because the original CSV had a column called *is_word* (0 or 1).

As for the BLP scores, they were saved automatically in a spreadsheet after participants responded to the questionnaire in Google Forms. The 30 scores were merged into one unique

CSV file and participants' names were changed by their ID to match the IDs on the lexical decision task CSV. Once that was done, the two CSVs were merged to be able to have one final file which contained Participants' ID, their answers and reaction time to the lexical decision task, and their language dominance in Catalan and Spanish.

After computing the descriptive statistics with R, the overall mean reaction time across all conditions was approximately 1.30 seconds (SD ≈ 0.27). Reaction times below 0.4 seconds and above 2.5 seconds were excluded from the analysis, as they were likely due to accidental key presses (too fast) or momentary distractions (too slow). When analyzed by dialect, responses were slightly faster for stimuli presented in Central Catalan (M = 1.26 s) compared to Valencian Catalan (M = 1.34 s). The effect of cognate status varied by dialect: while the mean RTs for cognates and non-cognates differed by only 60 milliseconds in the Central Catalan condition (1.29 vs. 1.23 s), in the Valencian condition, cognates were noticeably slower than non-cognates (1.40 vs. 1.28 s), with a difference of 120 milliseconds. These patterns suggest a possible interaction between dialect and cognate status, particularly in the Valencian condition. Accuracy was extremely high across all experimental conditions, with participants answering correctly in over 97% of trials. Due to this ceiling effect and the resulting lack of variability, accuracy was not included in the inferential analyses and is reported here descriptively. The descriptive statistics can be observed in Table 1 below.

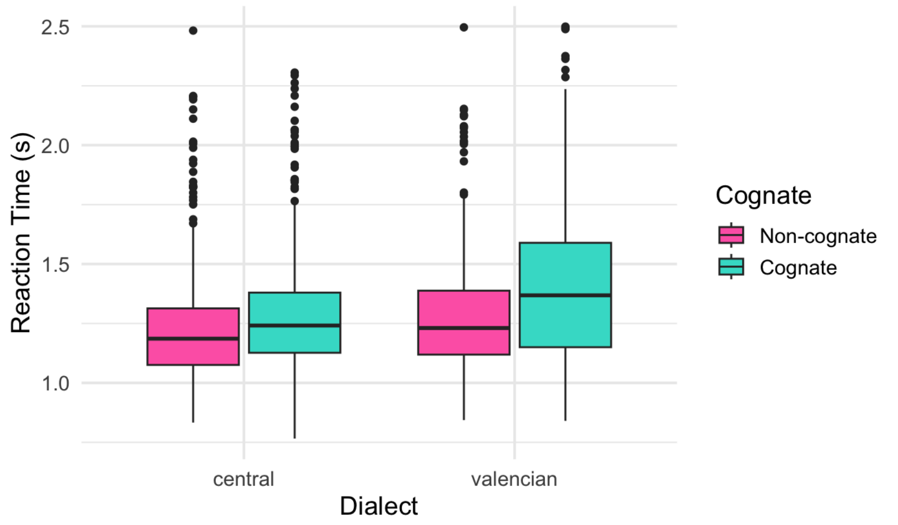
Table 1

Descriptive Statistics for Reaction Time by Conditions

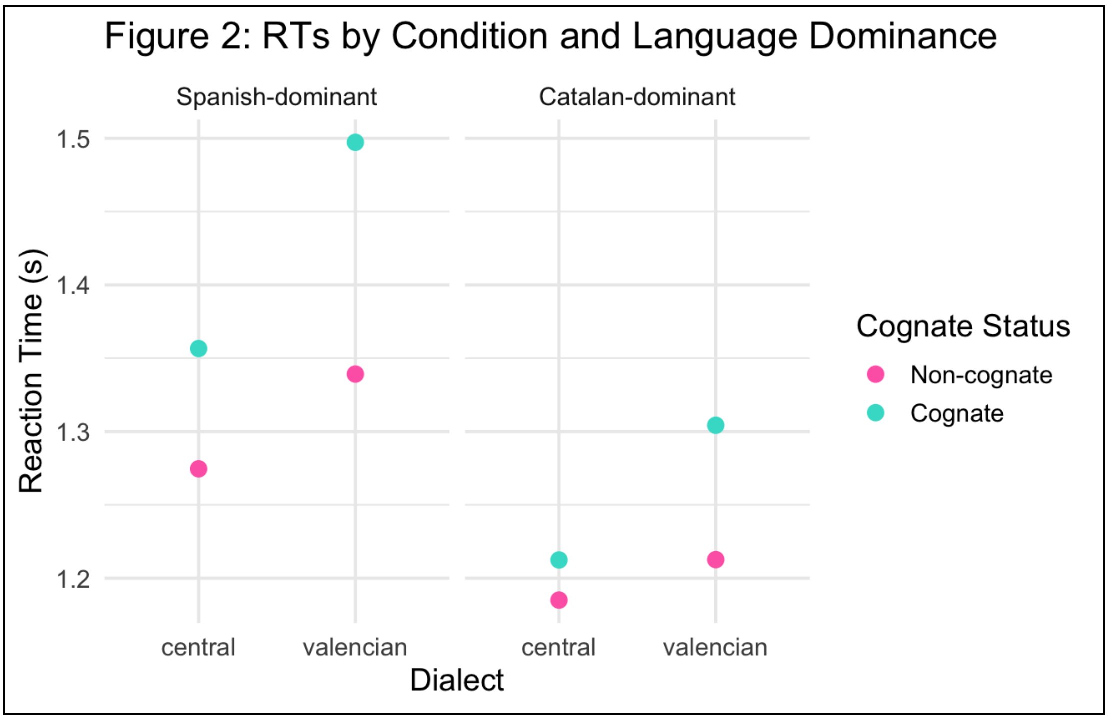
Is-cognate	Dialect	Mean-rt	Sd-rt	N
0	central	1.23	0.237	428
0	valencian	1.28	0.232	448
1	central	1.29	0.261	444
1	valencian	1.40	0.322	444

To visualize the distribution of reaction times across conditions, a boxplot was generated(Figure 1). The plot shows reaction times as a function of dialect (Central vs. Valencian Catalan) and cognate status (cognate vs. non-cognate). Across conditions, reaction times appeared slightly lower for non-cognates than for cognates, particularly in the Valencian dialect. The dispersion of values was notably greater in the cognate-Valencian condition, suggesting more variability in responses for that group.

Figure 1: Reaction Time by Dialect and Cognate Status



Before moving on to the inferential statistics, an exploratory plot was created to visualize the interaction between cognate status, dialect, and language dominance. Figure 2 below displays mean reaction times as a function of dialect (Central vs. Valencian Catalan) and cognate status (non-cognates vs. cognates), separated by two levels of language dominance: Spanish-dominant and Catalan-dominant. The dominance score was calculated by subtracting the Catalan score from the Spanish score on the Bilingual Language Profile (BLP), which ranges from –218 (maximum Spanish dominance) to +218 (maximum Catalan dominance). Based on this score, participants were grouped as either Spanish-dominant (dominance score < 0) or Catalan-dominant (dominance score > 0). In the present sample, the dominance values ranged from –79.6 to +81.4, indicating that all participants were bilingual to some degree, but with varying degrees of dominance. The updated plot reveals distinct patterns across dominance groups, which are explored further in the inferential analysis.



As shown, Catalan-dominant participants (right panel) were generally faster across all conditions, as expected, and responded slightly faster to non-cognates than to cognates in both dialects, although the difference between the cognate status is more evident when words were presented in Valencian. In contrast, Spanish-dominant participants (left panel) showed a clearer cognate disadvantage, particularly in the Valencian condition, where reaction times for cognates were substantially higher. This visual pattern suggests a potential interaction between language dominance, dialect, and cognate status, which is tested formally in the mixed-effects regression analysis.

Moving on to the inferential statistics, a linear mixed-effects regression model was implemented using R to investigate how cognate status, dialectal variation, and language dominance influence spoken word recognition. The dependent variable was the log-transformed reaction time (logRT). The raw RT were log-transformed in order to correct for the typical positive skew found in response time distributions. Even after trimming extreme outliers (responses faster than 0.4 seconds or slower than 2.5 seconds), reaction times remained asymmetrically distributed, with a longer tail of slower responses. Log transformation helps to normalize this distribution, satisfying the assumptions of linear modeling and producing more reliable estimates.

The model included three fixed effects: 1) cognate status (cognate vs. non-cognate), 2) dialect (Valencian vs. Central Catalan), and 3) language dominance, calculated as the difference between participants' BLP scores in Catalan and Spanish, offering a continuous dominance score. All two-way and three-way interaction terms among these predictors were also included to test whether the effect of one variable depended on the level of another. To account for repeated measures and inter-individual variability, random intercepts were included for both participants and items.

The reference levels for the categorical variables were set to non-cognate, Central Catalan, and a dominance score of 0, corresponding to balanced bilinguals. This allows the intercept to represent the predicted logRT in the baseline condition, and the other coefficients to reflect changes relative to this baseline. Overall, this model structure shows a comprehensive test of whether reaction time differences in lexical decision tasks can be explained by lexical overlap (cognates), phonological variation (dialect), and individual language profiles (dominance), as well as how these factors may interact. The summary of this mixed-effects model can be seen in Table 2 below.

Table 2

Fixed Effects from the Linear Mixed-Effects Regression Model Predicting Log-Transformed Reaction Times

Predictor	Slope	Std. Error	T Value	P Value
Intercept	0.2085	0.0239	8.71	< .001
Cognate (*Cognate vs. Non-cognate*)	0.0146	0.0312	0.47	.640
Dialect (*Valencian vs. Central*)	0.0211	0.0312	0.68	.500
Dominance (*Catalan > Spanish*)	-0.000676	0.000202	-3.35	.001 **
Cognate × Dialect	0.0754	0.0440	1.71	.089 (•)
Cognate × Dominance	-0.000185	0.000168	-1.11	.269
Dialect × Dominance	-0.000077	0.000167	-0.46	.646
Cognate × Dialect × Dominance	0.000201	0.000235	0.86	.393

Note: Double asterisk (**) indicates statistical evidence. Median point (•) indicates a trend

Although no statistically significant interaction was observed, the Cognate × Dialect interaction approached significance ($p = .089$), suggesting a potential trend. To better understand how the effect of cognate status may affect each dialect, and whether language dominance plays a distinct role in each, the dataset was subsequently split by dialect (Central vs. Valencian Catalan), and separate models were run for each subset.

The results for the Central Catalan model, observable in Table 3, showed a significant main effect of language dominance ($p = .002$), indicating that higher Catalan dominance was associated with faster response times. However, neither cognate status nor its interaction with dominance reached significance ($ps > .25$). This suggests that, in the Central Catalan condition, cognate status did not substantially influence reaction times, and dominance affected responses uniformly regardless of cognate status.

Table 3

Linear Mixed-Effects Regression Results – Central Catalan

Predictor	Slope	Std. Error	T Value	P Value
Intercept (Non-cognate, balanced)	0.2094	0.0248	8.43	< .001
Cognate (Cognate vs. Non-cognate)	0.0138	0.0323	0.43	.671
Dominance (Catalan > Spanish)	-0.000681	0.000203	-3.36	.002**
Cognate × Dominance	-0.000178	0.000155	-1.15	.251

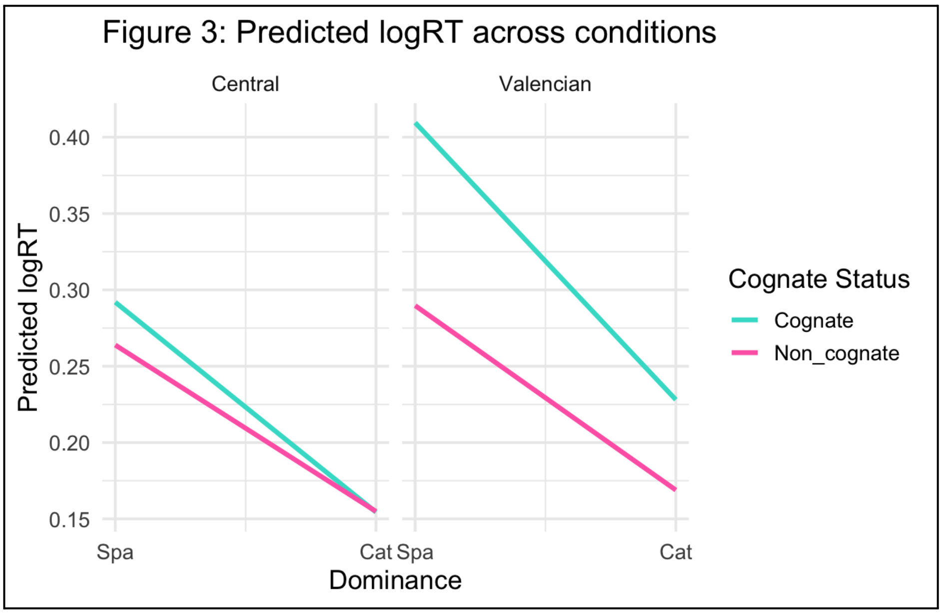
However, the Valencian Catalan model, shown in Table 4, revealed a more complex pattern. In this dialect condition, all predictors were statistically significant. A significant main effect of cognate status was found ($p = .004$), showing that participants responded significantly more slowly to cognates than non-cognates in Valencian. Additionally, language dominance again had a significant effect ($p < .001$), with higher Catalan dominance linked to faster responses. Most notably, the Cognate × Dominance interaction was also significant ($p = .032$), suggesting that the slowdown observed for cognates was modulated by participants' language dominance: the cognate disadvantage was strongest among Spanish-dominant individuals and decreased with greater Catalan dominance.

Table 4

Linear Mixed-Effects Regression Results – Valencian Catalan

Predictor	Slope	Std. Error	T Value	P Value
Intercept (Non-cognate, balanced)	0.2293	0.0229	10.02	< .001
Cognate (Cognate vs. Non-cognate)	0.0895	0.0300	2.98	.004 **
Dominance (Catalan > Spanish)	-0.000755	0.000195	-3.88	< .001 **
Cognate × Dominance	-0.000379	0.000176	-2.15	.032 **

These results confirm that the influence of cognate status on word recognition is not uniform across dialects and is particularly sensitive to the listener's linguistic profile in the Valencian condition. This pattern is observed more easily in Figure 3, which displays the model's predicted log-transformed reaction times as a function of language dominance and cognate status, separately for each dialect. The visual trend highlights the stronger cognate disadvantage in the Valencian condition, particularly among Spanish-dominant participants, and its gradual reduction as Catalan dominance increases.



The R script used to run the statistical analysis can be found in the link Annex 4.

Discussion:

The present study investigated how cognate status, dialectal variation (Central vs. Valencian Catalan), and individual language dominance influence lexical access in Catalan–Spanish bilinguals. While the full mixed-effects model did not reveal statistically significant main effects for cognate status or dialect, nor a three-way interaction, a marginally significant interaction between cognate status and dialect ($p = .089$) suggested a potential trend worth further exploration. Follow-up analyses, based on separate models for each dialect, revealed that the effects of cognate status were not uniform across dialects and were particularly modulated by language dominance in the Valencian condition.

One possible explanation for the dialect-specific cognate disadvantage observed in the Valencian condition lies in the phonological similarity between Valencian and Spanish, particularly for cognates. Although the overall model did not show a significant interaction, the separate analysis for Valencian Catalan revealed a clear cognate disadvantage, especially among Spanish-dominant participants. This suggests that the phonetic closeness of Valencian cognates to their Spanish equivalents may have triggered increased cross-linguistic activation. However, instead of facilitating recognition, this overlap may have led to lexical interference or uncertainty, especially considering that participants were explicitly asked to judge whether the stimulus was a Catalan word. In cases where the input closely resembled the Spanish form (e.g., votar or doctor), Spanish-dominant listeners may have doubted its belonging to the Catalan lexicon, slowing down their responses. This interpretation aligns with prior studies suggesting that phonological similarity can either facilitate or hinder lexical access depending on the listener's expectations, the task, and their language dominance (Dijkstra et al., 2010).

These results also align with findings from accent processing research, particularly the work of Floccia et al. (2006), who demonstrated that unfamiliar regional accents induce temporary processing delays in lexical access. In their study on French regional accents, lexical decision times increased when words were presented in an unfamiliar accent, suggesting that listeners require a short period of exposure to adjust their phonological expectations. Similarly, in the present study, participants were not accustomed to hearing Valencian in a prior exposure phase. As such, the phonological unfamiliarity of the Valencian dialect may have caused a perturbation in processing, especially for cognates that already activate both languages in bilinguals. This suggests that even within the same language, accent and dialect variation can impose measurable cognitive costs, particularly when listener expectations are misaligned with the auditory input.

These findings also align with Kroll and Stewart's (1994) Revised Hierarchical Model, which emphasizes asymmetries between dominant and non-dominant languages in bilingual lexical access. According to this model, bilinguals develop stronger conceptual links to their dominant language, especially in early stages of acquisition. In the present study, the phonological similarity between Valencian and Spanish may have led to stronger activation of Spanish lexical candidates among Spanish-dominant participants, increasing the need for inhibition and potentially causing interference. This could help explain why cognates in Valencian, despite their cross-linguistic similarity, were processed more slowly than non-cognates in this group. Thus, even though cognates are typically associated with facilitation, under certain conditions such as dialectal unfamiliarity or increased L2 interference, they may actually hinder processing.

The results partially support the original hypotheses. It was initially predicted that cognates would elicit faster reaction times than non-cognates in Central Catalan due to their phonological and semantic overlap with Spanish, and that this cognate facilitation would remain stable in the Valencian condition. However, the results did not support this pattern. In Central Catalan, the expected facilitation for cognates was not observed and no interaction with dominance emerged. More unexpectedly, in Valencian, cognates were processed significantly slower than non-cognates, particularly among Spanish-dominant participants.

This finding contradicts the assumption that cognates always facilitate lexical access and instead suggests that phonological similarity may increase lexical competition in unfamiliar dialects. The anticipated difficulty for non-cognates in Valencian did not occur; instead, cognates were more problematic, likely due to their phonetic overlap with Spanish triggering cross-linguistic interference, competition and inhibition. Overall, the results emphasize that cognate effects are not uniform across dialects and are heavily modulated by the listener's language dominance. This highlights the importance of phonological context and sociolinguistic experience in shaping bilingual spoken word recognition.

To summarise, while no main effects of dialect or cognate status were found in the general model, both variables showed meaningful effects when considered in interaction with language dominance, particularly within the Valencian dialect. This suggests that neither dialect nor cognate status uniformly influence lexical access across all bilinguals. Instead, their impact is shaped by the listener's language dominance profile and the phonological characteristics of the input. In other words, lexical processing in bilinguals appears to be context-sensitive, with effects emerging only under specific combinations of linguistic experience and phonological variation.

While the study yielded meaningful insights into the interaction between cognate status, dialectal variation, and language dominance, several limitations should be acknowledged. First, the sample size was relatively small (N = 30), which is sufficient to be able to detect large effects, but may have limited the statistical power needed to detect more subtle effects or generalize findings to a broader bilingual population. A larger and more diverse sample would strengthen the reliability of the results and allow for finer-grained analyses, such as comparisons across different age groups or levels of exposure to Valencian.

Second, the study did not control for phonotactic probability, which refers to the likelihood of a given sequence of sounds occurring in a specific language. Words with high phonotactic probability (i.e., sequences that are more common or expected in a language) are typically processed more quickly and accurately than those with lower probability. Failing to control for this variable may have introduced unintended variance in reaction times.

To deepen the knowledge around bilingual spoken word recognition, future studies could benefit from incorporating additional temporal and neurocognitive measures, such as eye-tracking. This tool would provide greater temporal resolution, allowing researchers to track the precise time course of lexical activation and competition during speech perception. Such methods could reveal whether the interference observed in reaction times arises early in processing (e.g., during phonological decoding) or later at the decision-making stage.

Another valuable direction would be to replicate this experiment using auditory stimuli recorded in Majorcan Catalan, a dialect that is phonologically even more distant from Spanish than either Central or Valencian Catalan. Majorcan Catalan features phonetic properties such as strong vowel reduction, diphthongisation, and specific consonant articulations that diverge substantially from Spanish phonology. Using this variety would allow researchers to explore whether greater phonological distance reduces lexical interference from Spanish, particularly in cognates, and whether this leads to clearer facilitation effects. Comparing results across dialects, Valencian, Central, and Majorcan, would offer a more comprehensive understanding of how phonetic similarity and dialectal familiarity shape lexical access in bilinguals, helping disentangle the effects of linguistic experience from those of phonological overlap.

Conclusion:

This study investigated how cognate status, dialectal variation, and language dominance influence lexical access in Catalan–Spanish bilinguals. The key finding was a significant three-way interaction, showing that the effects of cognate status and dialect depend on the speaker's dominance in either language. While cognates are generally thought to facilitate word recognition, this effect was reversed in certain conditions, particularly for Spanish-dominant participants hearing Valencian stimuli, suggesting that phonological similarity alone does not guarantee processing advantages.

By highlighting the role of dialectal unfamiliarity and listener expectations, the study contributes to a more nuanced understanding of bilingual lexical access. It challenges simplified models of cognate facilitation and underlines the importance of sociolinguistic and phonetic context in bilingual processing. These findings have implications for both theoretical models of bilingualism and practical applications, such as language teaching and speech recognition tools in multilingual contexts.

References:

Amengual, M. (2014). The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. *International Journal of Bilingualism, 20*(2), 133-152. https://doi.org/10.1177/1367006914544988

Birdsong, D., Gertken, L.M., & Amengual, M. (2012). Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin. https://blp.coerll.utexas.edu.

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition, 17*(3), 610-629. DOI: 10.1017/S1366728913000564

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283. DOI: 10.1037//0278-7393.26.5.1283

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and language, 62*(3), 284-301. DOI: https://doi.org/10.1016/j.jml.2009.12.003

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing?. *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1276. DOI: 10.1037/0096-1523.32.5.1276

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition, 1*(2), 67-81. DOI: https://doi.org/10.1017/S1366728998000133

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and language, 36*(1), 3-15. DOI: https://doi.org/10.1016/0093-934X(89)90048-5

Grosjean, F. (2020). The bilingual's language modes. *In The bilingualism reader* (pp. 428-449). Routledge. DOI: https://doi.org/10.1017/9781009210409.004

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A Web-based Swiss Army knife to select stimuli for psycholinguistic studies. Behavior Research Methods, 45, 765-771. DOI: https://psico.fcep.urv.cat/utilitats/nim/eng/valorescat.php

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

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373-381. DOI: https://doi.org/10.1017/S136672891000009X

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, *82*(13), 1-26. doi:10.18637/jss.v082.i13 <https://doi.org/10.18637/jss.v082.i13

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. Bilingualism: *Language and cognition, 6*(2), 97-115. DOI: https://doi.org/10.1017/S1366728903001068

New, B., Bourgin, J., Barra, J., & Pallier, C. (2024). UniPseudo: A universal pseudoword generator. *Quarterly Journal of Experimental Psychology, 77*(2), 278-286. DOI: https://doi.org/10.1177/17470218231164373

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. Behavior Research Methods. 10.3758/s13428-018-01193-y

Peirce, J. W., Hirst, R. J. & MacAskill, M. R. (2022). Building Experiments in PsychoPy. 2nd Edn London: Sage.

Python Software Foundation. (2023). Python (Version 3.10) [Programming language]. https://www.python.org

R Core Team. (2024). R: A language and environment for statistical computing (Version 4.x) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological science, 10*(3), 281-284. DOI: https://doi.org/10.1111/1467-9280.00151

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of memory and language, 60*(4), 487-501. DOI: https://doi.org/10.1016/j.jml.2009.01.001

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of memory and language, 50*(1), 1-25. DOI: https://doi.org/10.1016/S0749-596X(03)00105-0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr

Wickham, H., Vaughan, D., Girlich, M. (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1, https://CRAN.R-project.org/package=tidyr

Annex 1:

Word	IPA Cat	IPA Val	Relative F	Letters	POS	Word ESP	IPA Esp	Cog/No cog	Levenshtein distance cat-esp	Levenshtein distance val-esp
carro	kˈaru	kˈaro	33,364	5	sus	carro	kˈaro	cog	1	0
botó	butˈo	botˈo	7,668	4	sus	botón	botˈon	cog	2	1
doctor	duktˈo	doktˈoɾ	86,648	6	sus	doctor	doktˈoɾ	cog	2	0
escut	əskˈut	eskˈut	25,286	5	sus	escudo	eskˈuðo	cog	3	2
error	ərˈoɾ	erˈoɾ	39,197	5	sus	error	erˈoɾ	cog	1	0
volar	bulˈa	bolˈaɾ	19,257	5	vrb	volar	bolˈaɾ	cog	2	0
acte	ˈaktə	ˈakte	139,21	4	sus	acto	ˈakto	cog	1	1
color	kulˈo	kolˈoɾ	195,83	5	sus	color	kolˈoɾ	cog	2	2
piano	piˈanu	pjˈano	24,603	5	sus adv	piano	pjˈano	cog	4	3
content	kuntˈen	kontˈent	39,8	7	adj	contento	kontˈento	cog	3	1
votar	butˈa	botˈaɾ	8,682	5	vrb	votar	botˈaɾ	cog	2	0
cotxe	kˈotʃə	kˈotʃe	52,094	5	sus	coche	kˈotʃe	cog	1	0
moto	mˈotu	mˈoto	3,922	4	sus	moto	mˈoto	cog	1	0
consell	kunsˈɛʎ	konsˈej	116,07	7	sus	consejo	konsˈexo	cog	4	2
tigre	tˈiɣrə	tˈiɣre	7,141	5	sus	tigre	tˈiɣre	cog	1	0
regle	rˈeglə	rˈeɣle	2,068	5	sus	regla	rˈeɣla	cog	2	1
córrer	kˈorə	kˈoreɾ	66,161	6	vrb	correr	korˈeɾ	cog	2	0
portal	purtˈal	portˈal	44,309	6	sus	portal	portˈal	cog	1	0
costós	kustˈos	kostˈos	2,653	6	adj	costoso	kostˈoso	cog	2	1
timbre	tˈimbrə	tˈimbre	13,853	6	sus	timbre	tˈimbre	cog	1	0
brodar	bruðˈa	broðˈaɾ	2,283	6	vrb	bordar	borðˈaɾ	cog	3	2
feliç	fəlˈis	felˈis	53,733	5	adj	flexible	felˈiθ	cog	2	1
enviar	əmbiˈa	emvjˈaɾ	28,349	6	vrb	enviar	embjˈaɾ	cog	3	1
zero	zˈɛru	zˈɛro	12,994	4	sus	cero	θˈero	cog	3	2
esforç	əsfˈɔrs	esfˈɔrs	110,88	6	sus	esfuerzo	esfwˈerθo	cog	5	4
vestit	bəstˈit	vestˈit	102,28	6	sus	vestido	bestˈiðo	cog	3	3
molí	mulˈi	molˈi	18,34	4	sus	molino	molˈino	cog	3	2
destí	dəstˈi	destˈi	47,997	5	sus	destino	destˈino	cog	3	2
perfil	pərfˈil	perfˈil	16,877	6	sus	perfil	perfˈil	cog	1	0
comptar	kumtˈa	kontˈaɾ	40,543	7	vrb	contar	kontˈaɾ	cog	3	0
colom	kulˈom	kolˈom	10,38	5	sus	paloma	palˈoma	no cog	/	/
cordar	kurðˈa	korðˈaɾ	1,405	6	vrb	atar / abrochar	atˈaɾ / aβrotʃˈaɾ	no cog	/	/
forat	furˈat	forˈat	29,364	5	sus	agujero	aɣuxˈero	no cog	/	/
menjar	mənʒˈa	mendʒˈaɾ	114,548	6	sus vrb	comer	komˈeɾ	no cog	/	/

ocell	usˈeʎ	osˈej	35,646	5	sus	pájaro	pˈaxaɾo	no cog	/	/
vespre	bˈesprə	vˈespre	79,62	6	sus	atardecer	ataɾðeθˈeɾ	no cog	/	/
foscor	fuskˈo	foskˈoɾ	18,301	6	sus	oscuridad	oskuɾiðˈad	no cog	/	/
sortida	surtˈiðə	sortˈiða	64,503	7	sus	salida	salˈiða	no cog	/	/
voler	bulˈɛ	volˈeɾ	98,081	5	sus vrb	querer	keɾˈeɾ	no cog	/	/
sostre	sˈɔstrə	sˈɔstre	34,066	6	sus	techo	tˈetʃo	no cog	/	/
trencar	trəŋkˈa	treŋkˈaɾ	30,964	7	vrb	romper	rompˈeɾ	no cog	/	/
quelcom	kəlkˈɔm	kelkˈɔm	103,895	7	adv pro	algo	ˈalɣo	no cog	/	/
portar	purtˈa	portˈaɾ	132,79	6	vrb	llevar / traer	ɟeβˈaɾ / traˈeɾ	no cog	/	/
forat	fuɾˈat	foɾˈat	29,364	5	sus	agujero	aɣuxˈeɾo	no cog	/	/
estiu	əstiˈw	estiˈw	92,15	5	sus	verano	beɾˈano	no cog	/	/
treure	trˈɛwrə	trˈewre	119,933	6	vrb	sacar / quitar	sakˈaɾ / kitˈaɾ	no cog	/	/
petó	pətˈo	petˈo	11,219	4	sus	beso	bˈeso	no cog	/	/
oncle	ˈoŋklə	ˈoŋkle	58,845	5	sus	tío	tˈio	no cog	/	/
sopar	supˈa	sopˈaɾ	49,031	5	sus vrb	cena	θˈena	no cog	/	/
pernil	pərnˈil	pernˈil	7,765	6	sus	jamón	xamˈon	no cog	/	/
genoll	ʒənˈɔʎ	dʒenˈɔj	13,189	6	sus	rodilla	roðˈiɟa	no cog	/	/
llençol	ʎənsˈɔl	jensˈɔl	8,097	7	sus	sábana	sˈaβana	no cog	/	/
enciam	ənsiˈam	ensjˈam	2,614	6	sus	lechuga	letʃˈuɣa	no cog	/	/
préssec	pɾˈesək	pɾˈesek	2,946	7	sus	melocotón	melokotˈon	no cog	/	/
bessons	bəsˈons	besˈons	5,561	7	sus adj	gemelos / mellizos	xemˈelos / meɟˈiθos	no cog	/	/
omplir	umplˈi	omplˈiɾ	30,61	6	vrb	llenar	ɟenˈaɾ	no cog	/	/
cosí	kuzˈi	kozˈi	15,082	4	sus	primo	pɾˈimo	no cog	/	/
pebrot	pəβɾˈɔt	peβɾˈɔt	3,414	6	sus	pimiento	pimjˈento	no cog	/	/
feble	fˈeblə	fˈeβle	32,973	5	sus adj	débil	dˈeβil	no cog	/	/
vermell	bərmˈɛʎ	vermˈej	46,963	7	sus adj	rojo	rˈoxo	no cog	/	/

Annex 2:

Word	Transcription Cat	Transcription Val	Letters
tolú	tulˈu	tolˈu	4
cegil	səʒˈil	sedʒˈil	4
odal	uðˈal	oðˈal	4
coví	kuβˈi	kovˈi	4
medó	məðˈo	meðˈo	4
elal	əlˈal	elˈal	4
obor	uβˈo	oβˈoɾ	4
lesís	ləzˈis	lezˈis	5
rebic	rəβˈik	reβˈik	5
bofoc	bufˈɔk	bofˈɔk	5
sirle	sˈirlə	sˈirle	5
gàcil	gˈasil	gˈasil	5
creda	krˈɛðə	krˈɛða	5
còlim	kˈɔlim	kˈɔlim	5
pralo	prˈalu	prˈalo	5
ampes	ˈampəs	ˈampes	5
domià	dumiˈa	domjˈa	5
eniat	əniˈat	enjˈat	5
narro	nˈaru	nˈaro	5
paiat	pəjˈat	pajˈat	5
libar	liβˈa	liβˈaɾ	5
plupa	plˈupə	plˈupa	5
desta	dˈestə	dˈesta	5
bogès	buʒˈɛs	bodʒˈes	5
migil	miʒˈil	midʒˈil	5
moric	muɾˈik	moɾˈik	5
obsem	upsˈɛm	opsˈem	5
cunda	kˈundə	kˈunda	5
jancle	ʒˈaŋklə	dʒˈaŋkle	6
solsat	sulzˈat	solzˈat	6
mustet	mustˈɛt	mustˈet	6
flopès	flupˈɛs	flopˈes	6
tèrdia	tˈɛrðiə	tˈɛrðja	6
crover	kruβˈe	kroβˈeɾ	6
mellés	məʎˈes	mejˈes	6
indalt	indˈal	indˈalt	6
tresta	trˈɛstə	trˈɛsta	6

cronís	krunˈis	kronˈis	6
colter	kultˈe	koltˈeɾ	6
mombès	mumbˈɛs	mombˈes	6
rebell	rəβˈeʎ	reβˈej	6
presat	prəzˈat	prezˈat	6
rínvol	rˈimbul	rˈimvol	6
taulèn	təwlˈɛn	tawlˈen	6
sunyal	suɲˈal	suɲˈal	6
feutic	fəwtˈik	fewtˈik	6
meixir	məʃˈi	mejʃˈiɾ	6
lirtal	lirtˈal	lirtˈal	6
crossot	krusˈɔt	krosˈɔt	6
espurt	əspˈur	espˈurt	6
pullars	puʎˈas	pujˈars	7
esprest	əsprˈest	esprˈest	7
fitreig	fitrˈɛtʃ	fitrˈetʃ	7
vostrem	bustrˈɛm	vostrˈem	7
marpell	mərpˈeʎ	marpˈej	7
sorprit	surprˈit	sorprˈit	7
clíptic	klˈiptik	klˈiptik	7
gualdic	gwəldˈik	gwaldˈik	7
prempar	prəmpˈa	prempˈaɾ	7
doncalt	duŋkˈal	doŋkˈalt	7

Annex 3

Link to the repository that contains the Python script used to run the Lexical Decision Task:
https://github.com/paaueetee/Python-file-for-the-lexical-deicison-task

Annex 4

Link to the repository that contains the R script used to compute the statistics:
https://github.com/paaueetee/R-script-for-statistics-for-the-Lexical-Decision-Task