# Lexical organization of language-ambiguous and language-specific words in bilinguals

**2 authors:**

Aina Casaponsa
Lancaster University

**7** PUBLICATIONS   **20** CITATIONS

SEE PROFILE

Jon Andoni Duñabeitia
Basque Center on Cognition, Brain and Language

**95** PUBLICATIONS   **1,684** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

Phonological and orthographic coding in skilled deaf readers View project

**Lexical organization of language-ambiguous and language-specific words in bilinguals**

*Running head:* ORTHOGRAPHIC CUES-LANGUAGE ORGANIZATION

Aina Casaponsa and Jon Andoni Duñabeitia

BCBL. Basque Center on Cognition, Brain and Language; Donostia, Spain

*Contact information:*

**Aina Casaponsa**

Basque Center on Cognition, Brain and Language (BCBL)

Paseo Mikeletegi 69, 2$^{nd}$ floor

20009 Donostia, SPAIN

*a.casaponsa@bcbl.eu*

Telephone: +34943309300

**Abstract**

Previous research has shown the importance of sub-lexical orthographic cues in determining the language of a given word when the two languages of a bilingual reader share the same script. In this study we explored the extent to which cross-language sub-lexical characteristics of words -measured in terms of bigram frequencies- constrain selective language activation during reading. In Experiment 1, we investigated the impact of language-nonspecific and language-specific orthography in letter detection using the Reicher-Wheeler paradigm in a seemingly monolingual experimental context. In Experiment 2, we used the masked translation priming paradigm in order to better characterize the role of sub-lexical language cues during lexical access in bilinguals. Results show that bilinguals are highly sensitive to statistical orthographic regularities of their languages and that the absence of such cues promotes language-nonspecific lexical access, whereas their presence partially reduces parallel language activation. We conclude that language co-activation in bilinguals is highly modulated by sub-lexical processing and that orthographic regularities of the two languages of a bilingual are a determining factor in lexical access.

The ease of language identification during single word reading depends on whether the two languages of a bilingual share a unique script or have distinctive orthographic representations. For instance, an English-Japanese bilingual can easily identify that the word "家" (the Japanese for "house") is Japanese and not English. In contrast, script-based cues may not be helpful for other types of bilinguals, such as English-French bilinguals, and language identification of the French word "maison" (the French for "house") may involve a deeper type of processing, possibly requiring access to lexical semantic representations. For instance, a Basque-Spanish bilingual easily identifies the word "etxe" (the Basque for "house") as a word of Basque and its translation equivalent "casa" as a word of Spanish, simply on the basis of the orthographic regularities of the two languages. Indeed, the bigram "tx" cannot be found in the Spanish lexicon, and the letter "c" is not used in Basque words (namely, an orthographic markedness effect).

In the current study we focused on the impact of sub-lexical cues in bilingual word identification at different stages of the processing stream. Recent studies demonstrate that bilinguals whose languages share the same script rely on sub-lexical orthographic cues to aid language and word identification, and to discriminate between real and invented words (i.e., non-words). In this line, studies using non-words with different degrees of cross-linguistic similarity in their orthographic structure have shown that this factor highly determines how fast and accurately bilingual participants will reject them in lexical decision tasks (e.g., Lemhöfer & Dijkstra, 2004; Lemhöfer & Radach, 2009). While the evidence in this regard is still scarce, the studies testing the impact of orthographic markedness in bilingual visual word processing with bilingual readers whose languages share the same alphabet have demonstrated the importance of sub-lexical orthographic cues (see Casaponsa, Carreiras and Duñabeitia, 2014; Vaid and Frenck-Mestre, 2002; Van

Kesteren, Dijkstra, and de Smedt, 2012). The purpose of the present study is to explore whether cross-linguistic orthographic regularities constrain the access to language-specific orthographic and lexical representations in seemingly monolingual language contexts (i.e., single-language tasks).

The seminal study by Vaid and Frenck-Mestre (2002) showed that French-English bilingual readers identify the language of words significantly faster when they are orthographically marked (i.e., when they include bigrams that are more plausible in one of the languages than in the other). The authors claim that orthographically unmarked words (words that are equally plausible in the two languages from a bigram frequency viewpoint) require a lexical search strategy for efficient language assignment as compared to orthographically marked words whose language can be identified on the mere basis of a perceptual-orthographic strategy. Lemhöfer, Koester and Schreuder (2011) further demonstrated the importance of orthotactic cues in bilingual visual word processing by showing that bilinguals (as well as monolinguals) are highly sensitive to the frequency and legality of the bigrams at the morphemic boundary of compound words, and that they use this piece of information to process morphologically complex words. Van Kesteren, Dijkstra, and de Smedt (2012) showed that Norwegian-English bilinguals strongly rely on sub-lexical orthographic cues such as cross-language bigram legality to access language membership and accomplish lexical access. Bilinguals responded to orthographically marked and unmarked Norwegian and English words differently in a series of lexical decision and language identification tasks, and the authors concluded that orthographic markedness aids language membership identification. More recently, Casaponsa, Carreiras and Duñabeitia (2014) also found that Basque-Spanish bilinguals make use of sub-lexical orthotactic cues to efficiently

determine the language of a word (with the language of orthographically marked words being recognized significantly faster than that of unmarked words).

As a consequence of the evidence showing a clear impact of sub-lexical cues in bilingual lexical access, it has been recently proposed that the bilingual lexicon may be organized differently depending on the degree of language-specificity of the words (e.g., based on orthographic markedness). Recent studies suggest that non-selective language access occurs during bilingual word processing in the absence of sub-lexical cues (e.g., Ng & Wicha, 2013) but, when such cues are available to the readers (e.g., Rodríguez-Fornells, Rotte, Heinze, Nösselt, & Münte, 2002), bilinguals may display some degree of language-selectivity during language processing showing reduced interference from the non-target language. In the same vein, Orfanidou and Sumner (2005) showed that language-specific orthographic cues reduce the potential interference from lexical representations of a non-target language (see also Gonzales & Lotto, 2013, and Ju & Luce, 2004, for evidence from spoken word recognition), while words with language non-selective sub-lexical units yield greater co-activation of the two languages. For instance, in the absence of language-specific sub-lexical cues, non-balanced bilingual participants take longer to reject words from their non-native language when completing a lexical decision task in their native language (e.g., Van Kesteren et al., 2012), and both balanced and non-balanced bilinguals struggle with perceptual identification of unmarked words more than marked words, possibly due to the enhanced parallel activation of competing items in the non-target language (e.g., Casaponsa et al., 2014).

Support for this assumption also stems from a recent study exploring the way in which the electrophysiological markers of the masked language switching effects vary as a function of orthographic markedness of the primes (e.g., comparing the masked switch cost effects for

primes that are clearly distinct from the non-target language due to their bigrams' low degree of plausibility in the target language and the effects for primes from the non-target language that are perfectly plausible in the target language). Using this manipulation, Casaponsa, Carreiras and Duñabeitia (submitted) have recently demonstrated that masked switch cost effects are prominently reduced for unmarked Basque prime words preceding unrelated Spanish words (as compared to a non-switch condition), while Basque prime words that are not orthographically plausible in Spanish (i.e., marked Basque words) yielded early significant and robust masked switch cost effects.

Thus, orthographic markedness appears to provide bilingual readers with a processing cue that modulates language non-selective lexical access in language combinations with partially but not completely overlapping sub-lexical representations. The effects associated with the processing of orthographically marked words (e.g., the Basque word "neska", meaning "girl"; note that "sk" is an implausible bigram in Spanish) suggest that these items are assigned language membership at visuo-orthographic stages of processing in language identification tasks, and that their degree of cross-language competition is partially reduced in tasks requiring lexical access (see Casaponsa et al., 2014; Van Kesteren et al., 2012). In contrast, the effects associated with the processing of orthographically unmarked words (e.g., the Basque word "mutil", meaning "boy"; note that the word "mutil" is orthographically legal in Spanish), suggest that these items are prone to a high degree of cross-language interference, thus leading to delayed language recognition and more effortful lexical access (see Casaponsa et al.2014; Vaid & Frenck-Mestre, 2002; Van Kesteren et al. 2012). Hence, cross-linguistic orthographic regularities may constrain access to language-specific orthographic and lexical representations (i.e., enable language-selective lexical access) in the presence of orthographically distinctive sub-lexical units. In contrast, in the absence of

such orthographic cues bilingual readers may access language-independent orthographic and lexical representations. This is precisely the hypothesis that we tested in the current study.

We tested a group of balanced simultaneous Basque-Spanish bilinguals in two experiments aimed at demonstrating the role that orthographically marked words play in bilingual lexical access, and whether or not words containing language-specific orthographic cues yield language-selective lexical access at initial stages of bilingual visual word recognition. In Experiment 1 we employed an orthographic task sensitive to both lexical and sub-lexical factors: the forced-choice letter identification task (namely, the Reicher-Wheeler paradigm; see Reicher, 1969; Wheeler, 1970). In this task, participants are presented with a letter string for a brief period of time, and then asked to indicate which of two letters appearing immediately afterwards was present in the string. Typically, letters embedded within real words are recognized better than letters embedded in strings that have no existing orthographic representations but whose orthographic regularities nevertheless match those from the target language (i.e., legal non-words). This effect has been termed as the word superiority effect (e.g., McClelland, 1976; Prinzmetal, 1992; see also Grainger, Bouttevin, Truc, Bastien, & Ziegler, 2003). Similarly, previous studies using the Reicher-Wheeler paradigm have also shown that readers struggle less when identifying letters embedded in legal non-words than those embedded in illegal non-words (i.e., the pseudo-word superiority effect; see Grainger & Jacobs, 1994; Grainger et al., 2003), demonstrating that orthographic coding is highly influenced by the degree of familiarity with the orthographic structure of the target language. Interestingly, Grossi, Murphy and Boggan (2008) extended these effects to bilingual readers, showing that non-balanced Italian-English bilinguals display sizeable (pseudo-)word superiority effects when tested in their two languages.

Here, in Experiment 1 we tested balanced simultaneous Basque-Spanish bilinguals in an adapted version of the Reicher-Wheeler paradigm using Basque words and non-words legal in Basque. Critically, half of the Basque words and non-words corresponded to language-specific orthographically marked strings (i.e., their bigram combinations were implausible in Spanish). The other half of the items in each set was made of unmarked Basque words and non-words without language-specific orthographic cues (i.e., in which all bigram combinations were also legal in Spanish).

Considering previous evidence from monolinguals and bilinguals (see Grossi et al., 2008, for review), we predicted significant word superiority effects for both orthographically marked and unmarked Basque words as compared to marked and unmarked non-words. Our critical question of interest in Experiment 1, however, relates to the extent to which Basque-Spanish bilingual readers would show significant differences in letter identification for marked and unmarked strings. Following the assumption that orthographic coding is modulated by the degree of familiarity of the reader with the orthographic structure of language (see Grainger et al., 2003), we hypothesized that bilingual participants would display significantly better performance in letter identification for Basque words and non-words that are orthographically legal in Spanish (i.e., orthographically unmarked items) than for Basque words and non-words that feature orthographic cues implausible in Spanish (i.e., orthographically marked items). Under the assumption that orthographically unmarked words lead to language non-selective activation, the predicted enhanced recognition of letters embedded in orthographically unmarked strings stems from the increased feedback received by these letters from higher-level lexical representations in the two languages (as predicted by interactive-activation models of bilingual and monolingual word recognition; e.g., Grainger & Jacobs, 1994; Grainger, Midgley, & Holcomb, 2010;

McClelland & Rumelhart, 1981; van Heuven & Dijkstra, 2010). According to the hypothesis that orthographically marked words would lead to language-selective lexical-orthographic activation, letters embedded in language-specific Basque strings would receive a lesser degree of feedback activation from higher levels of processing (they would not co-activate Spanish words), thus resulting in impoverished recognition of sub-lexical units.

In Experiment 2, a stronger test of our guiding hypothesis was implemented. If orthographically unmarked words elicit language non-selective lexical access while orthographically marked words primarily activate language-specific lexical units, it could be predicted that the former type of words would be more prone than the latter to effects that depend on cross-language co-activation. The same group of balanced Basque-Spanish bilinguals completed a masked translation priming experiment with words without extensive orthographic overlap across languages. Spanish targets were preceded by Basque primes that could be the translation equivalents of the targets (e.g., ilargi-LUNA, the Basque and Spanish words for "moon"), or that were unrelated in meaning to the targets (e.g., ipuin-LUNA, where "ipuin" is the Basque for "tale"). Critically, half of the Basque translation equivalents corresponded to language-specific words (i.e., marked Basque words with bigrams that were implausible in Spanish), while the other half of the Basque primes corresponded to language-nonspecific words (i.e., unmarked Basque words with bigram combinations legal in Spanish).

Cross-script masked translation priming effects for non-cognates have been repeatedly shown for non-balanced sequential bilinguals, mainly occurring when primes belong to the native language and targets belong to the nonnative language (i.e., L1-to-L2 direction; see, among many others, Gollan, Forster, & Frost, 1997, for Hebrew-English combinations; Jiang, 1999, Jiang & Forster, 2001, and Witzel & Forster, 2012, for Chinese-English bilinguals; Kim & Davis, 2003, for

Korean-English bilinguals; Dimitropoulou, Duñabeitia, & Carreiras, 2011a, and Voga & Grainger, 2007, for Greek-English bilinguals; Dimitropoulou, Duñabeitia, & Carreiras, 2011b, for Greek-Spanish bilinguals). Somewhat parallel effects have also been found in non-balanced bilinguals in the L1-to-L2 translation direction in language combinations that share the same script, even though fewer masked priming studies have reported such effects (see Basnight-Brown & Altarriba, 2007, for Spanish-English bilinguals; de Groot & Nas, 1991, for Dutch-English bilinguals; Duyck & Warlop, 2011, for Dutch-French bilinguals; Williams, 1994, for Italian-English, French-English and German-English combinations). Interestingly for the purposes of Experiment 2, balanced simultaneous bilinguals that share the same script have been found to show highly similar effects irrespectively of the translation direction. Duñabeitia, Perea, and Carreiras (2010) for instance showed that balanced Basque-Spanish bilinguals permanently exposed to their two languages display significant and comparable masked translation priming effects in both language directions (see also Duñabeitia, Dimitropoulou, et al., 2010, for an EEG replication of these effects). The question under scrutiny in Experiment 2 is whether these masked translation priming effects in balanced simultaneous bilinguals (i.e., faster recognition of targets preceded by their translation equivalents in the other language as compared to unrelated primes) are modulated by the orthographic markedness of the briefly displayed masked words.

Hence, following the line of reasoning sketched above, we expected larger masked translation priming effects in Experiment 2 for Spanish targets preceded by orthographically unmarked translation equivalents in Basque (i.e., Basque primes with language-nonspecific orthography) as compared to Spanish targets preceded by Basque translation equivalents that are orthographically distinctive or marked (i.e., Basque primes with language-specific orthographic regularities). Such a result would demonstrate that orthographic markedness modulates bilingual lexical access,

suggesting that lexical-orthographic representations of words from two languages that share basic sub-lexical orthographic distributional information are stored closer in lexical semantic memory than representations of words with highly distinctive language-selective orthotactics.

## Experiment 1: Forced-choice letter identification task

**Methods**

**Participants.** Fifty undergraduate students (30 females; mean age=24.26, *SD*= 3.72) with normal or corrected-to-normal vision participated in this experiment in exchange for monetary compensation. All of them were balanced simultaneous Basque-Spanish bilinguals (mean age of Spanish acquisition: 1.12 years, *SD*=1.57; mean age of Basque acquisition: 0.54, *SD*=0.81). All participants were perfectly fluent in both Spanish and Basque, as demonstrated by their proficiency self-ratings, their performance in a vocabulary test in Basque and Spanish, and by an extensive interview to test their overall fluency in each language (see Table 1)[Footnote 1].

*(Table 1 about here)*

**Materials.** One hundred Basque words (79 nouns, 21 adjectives) were selected from E-Hitz (length range: 5-8; Perea et al., 2006). We used the length-corrected orthographic Levenshtein distance in order to restrict the cross-linguistic similarity between these words and their Spanish translation equivalents. This measure ranges between 0 and 1, where 0 corresponds to the minimum possible value in the cognate continuum (i.e., completely different translation equivalents at the form level), and 1 corresponds to fully overlapping cognates; see Casaponsa, Antón, Pérez, & Duñabeitia, in press; Duñabeitia et al., 2013; Schepens, Dijkstra, & Grootjen,

2012). We only selected words in the range from 0 to .6, avoiding extensive overlap across languages (mean=.17, *SD*=.13). Half of these Basque words were made of bigrams combinations that were also plausible in Spanish (i.e., unmarked words; 43 nouns and 7 adjectives), while the other half were Basque words including Basque-specific bigram combinations (i.e., marked words; 36 nouns and 14 adjectives). Marked words were always formed by at least one implausible bigram when measured according to the Spanish vocabulary. An additional set of one hundred Basque-like pseudo-words (length range: 5-8) were generated with Wuggy (Keuleers & Brysbaert, 2010). The same procedure than the one used for words was used to select the marked and unmarked conditions for the non-words. All possible sub-lexical and lexical factors were balanced across conditions (see Table 2 for detailed description), as well as the cross-linguistic orthographic overlap of the translations equivalents (unmarked: mean=.17, *SD*=.14; marked: mean=.18, *SD*=.13; p>.69).

Letter identification was tested at all positions except for first, last or exact middle locations (in order to avoid saliency effects; see Tygdat & Grainger, 2009). The critical letters at test could be either vowels or consonants, and they were selected so that they never formed part of the critical bigrams of the marked words (i.e., Basque-specific bigrams). The alternative letter given in the two-alternative forced-choice procedure was never part of the string. The amount of vowels and consonants, as well as the test positions within the strings and the location of the presentation of the alternative letters (left/right on the screen) were balanced across all four sets of strings (Marked words, Unmarked words, Marked non-words, Unmarked non-words). The full set of materials can be found in Appendix 1.

*(Table 2 about here)*

**Procedure.** Participants were tested individually in a quiet room using DMDX software (Forster & Forster, 2003) on a 15" CRT monitor set at 100 Hz. Stimuli were presented in Courier New white letters on a black background. Each trial started with a forward mask (e.g., ####) displayed for 500 ms that was of the same length of the subsequently presented letter string. Next, the referent word or non-word was presented in uppercase for 200 ms, immediately followed by a backward mask that was presented together with the two alternative letters (target and foil) in lowercase for 2500 ms or until a response was given (see Figure 1 for a schematic representation of a trial). Participants were asked to respond with the right hand when the letter that was previously embedded in the word appeared on the right side of the screen using a response box and with the left hand when the correct letter appeared on the left side of the screen. Trial presentation order was randomized across participants so that no participant was presented with the strings in the exact same order. 5 practice trials were included prior to the 200 experimental trials (50 Marked words, 50 Unmarked words, 50 Marked non-words, 50 Unmarked non-words). The experimental session approximately lasted for approximately 10 minutes.

### Results and Discussion

Timeouts (0.42%), erroneous responses (8.73%) and responses above or below 2.5 standard deviations from the participants-based and items-based means in each condition (4.16%) were excluded from the latency analysis. Mean latencies for correct responses and error rates are presented in Table 3. ANOVAs based on participant and item response latencies and error rates were conducted including the factors Lexicality (Word, Non-word) and Markedness (Marked, Unmarked).

*(Table 3 about here)*

13

ANOVAs on the reaction times revealed a main effect of Lexicality. Letters previously embedded in words were identified faster than letters embedded in non-words [F1(1,49)=26.85, MSE=2513.12, p<.001; F2(1,196)=12.11, MSE=7654.74, p<.001]. The main Markedness effect was also significant, showing that letters embedded in strings containing Spanish-plausible bigrams (i.e., unmarked language-unspecific string) being responded faster than letters embedded in strings containing Spanish-implausible bigrams (i.e., Basque-specific marked strings) [F1(1,49)=25.34, MSE=1287.12, p<.001; F2(1,196)=4.22, MSE=7654.74, p=.04]. These two factor did not interact with each other [F1/F2<.1, ps>.75].

ANOVAs on the error rates revealed a main effect of Lexicality, showing that letters previously embedded in words were detected more accurately than letters embedded in non-words [F1(1,49)=65.37, MSE=24.45, p<.001; F2(1,196)=31.14 MSE=51.10, p<.001]. The main Markedness effect was not significant [F1|F2<.2, ps>.65] and a marginal interaction between the two main factors was found only in the by-participants analysis [F1(1,49)=3.59 MSE=12.92, p=.06; F2(1,196)=.90 MSE=51.10, p=.34]. Planned comparisons did not show any significant difference between conditions.

As expected, results from Experiment 1 showed a significant word superiority effect. Bilingual participants identified target letters that were previously embedded in real Basque words significantly faster and more accurately than those that were embedded in non-words (e.g., McClelland, 1976; Prinzmetal, 1992; see Grossi et al., 2008, for supporting evidence with bilinguals). Critically, significant orthographic markedness effects were also found for both words and non-words. Basque-Spanish bilinguals recognized letters previously embedded in Basque-unspecific strings (namely, in strings whose orthographic regularities also matched Spanish orthotactics) faster than letters embedded in Basque-specific orthographically marked

strings. Hence, these results showed significant differences in letter identification for marked and unmarked strings, in accordance with the hypothesis that bilingual participants display significantly better performance in letter identification for Basque words and non-words that are orthographically plausible in Spanish. These results are in line with the idea that orthographically unmarked items lead to language non-selective activation, receiving enhanced feedback to the orthographic units forming these strings from higher-level lexical representations in the two languages (as predicted by models based on interactive-activation principles; see Grainger et al., 2003).

In Experiment 2, the same Basque-Spanish balanced simultaneous bilinguals completed a lexical decision task on Spanish targets that could be preceded either by their Basque orthographically marked or unmarked translation equivalents, or by unrelated Basque words. According to our guiding hypothesis and the results from Experiment 1, orthographically unmarked Basque words grant enhanced language-unspecific lexical access, thus yielding higher levels of cross-language interactions. Hence, we predicted that Spanish target words preceded by Basque primes with Basque-unspecific orthographic structures (i.e., Unmarked Basque primes) would result in significantly larger masked translation priming effects than Spanish targets following Basque-specific orthographically marked Basque primes.


### Experiment 2: Masked translation priming

**Methods**

**Participants.** The same participants who completed Experiment 1 took part in this experiment[Footnote 2].

**Materials.** Two hundred and forty Spanish words (178 nouns and 62 adjectives) and their corresponding non-cognate translation equivalents were selected from B-Pal and E-Hitz (Davis & Perea, 2005 and Perea et al., 2006, respectively). Spanish words were used as targets and Basque words as primes. As in Experiment 1, words were selected according to their length-corrected orthographic Levenshtein distance in order to avoid extensive orthographic overlap across translation equivalents (mean=.20, *SD*=.15, range: 0 to .58). Critically, half of the Basque translation equivalents followed the orthotactic rules of the Spanish orthography (i.e., Unmarked words; e.g. ilargi-LUNA), and the other half were Basque words with Basque-specific orthography (i.e., Marked words; e.g., txapel-BOINA). Marked Basque words (75 nouns and 45 adjectives) were always formed by at least one illegal bigram when measured according to the Spanish vocabulary, while Unmarked Basque words (103 nouns and 17 adjectives) respected Basque and Spanish orthotactics (see Table 4 for further details). Unrelated priming conditions were created by rearranging the primes and the targets within each markedness condition and avoiding orthographic and semantic overlap between primes and targets. This way, four conditions were created with 60 items per condition and list (i.e., Related Unmarked, Unrelated Unmarked, Related Marked, Unrelated Marked). All possible sub-lexical and lexical factors were equated across and within sets (see Table 4), as well as the orthographic overlap across languages for the translation equivalents (Unmarked: mean=.19, *SD*=.15; Marked: mean=.20, *SD*=.15; p>.7). For the purposes of the lexical decision an additional set of 240 Spanish non-words was created with Wuggy (Keuleers & Brysbaert, 2010). Half of the pseudowords were preceded by Unmarked Basque primes and the other half were preceded by Marked Basque primes. Both sets of words were also matched for the most critical factors. Two lists were created, so that each target word only appeared once in each list, but each time in a different

priming condition (Related, Unrelated). Items were randomly distributed. The full set of materials can be found in Appendix 2.

*(Table 4 about here)*

**Procedure.** Participants were tested individually in a quiet room using DMDX software (Forster & Forster, 2003) on a 15" CRT monitor set at 100 Hz. Stimuli were presented in Courier New white letters on a black background. Each trial started with a forward mask (e.g., ####) displayed for 500 ms. Next, the prime word was presented in lowercase for 50 ms (5 cycles), immediately followed by the target presented in uppercase for 2500 ms or until a response was given. Feedback was provided only when an erroneous response was given. Participants were asked to respond with the right hand to real words and with the left hand to invented words using a response box. Trial presentation order was randomized across participants. 10 practice trials were included prior to the 480 experimental trials. The experimental session approximately lasted for approximately 15 minutes.

## Results and Discussion

Timeouts (0.04%), erroneous responses (3.54%), and responses falling beyond or above the mean ± 2.5 standard deviations for each item and participant in each condition (4.67%) were excluded from the latency analysis. Mean latencies and error rates are shown in Table 5. Reaction times and error rates were analyzed in separate participant- and item-based ANOVAs following a 2 (Relatedness: Related, Unrelated) by 2 (Markedness: Marked, Unmarked) x 2 (List: List 1, List 2) design. The factor List was included as a dummy factor (see Pollatsek & Well, 1995).

*(Table 5 about here)*

ANOVAs on the reaction times showed negligible main Markedness and Relatedness effects [Markedness: $F_1(1,48)=.49$, MSE=377.13, p=.49; $F_2(1,237)=.19$, MSE=6299.78, p=.66; Relatedness: $F_1(1,48)=1.92$, MSE=496.74, p=.17; $F_2(1,237)=2.93$, MSE=871.22, p=.09]. Importantly, a significant interaction between Relatedness and Markedness was found [$F_1(1,48)=9.38$, MSE= 253.49, p=.004; $F_2(1,237)=6.87$, MSE=871.22, p=.009], showing that the translation priming effect significantly differed in magnitude for Marked and Unmarked conditions (-3 ms and 11 ms, respectively). Planned pairwise comparisons showed significant masked translation priming effect for Spanish words preceded by Unmarked Basque primes [$F_1(1,48)=7.42$, MSE=427.80, p=.009; $F_2(1,118)=10.36$, MSE=786.97, p=.002], but not for Spanish targets preceded by Marked Basque primes [$F_1|F_2<1$, p>.5]. A further confirmation of the presence of a significant masked translation priming effect for Unmarked primes was carried out by computing the Bayes Factors ($BF_{10}$) in a series of Bayesian paired-sample t-tests on the data at the participant- and item-level following the hypothesis that the Related items would elicit shorter RTs than the Unrelated items (see Rouder et al., 2012, and Wetzel et al., 2011, for reviews). Results demonstrated that the significant masked translation priming effect found for these items was associated with a $BF_{10}$ of 7.87 in the by-participants analysis and with a $BF_{10}$ of 26.09 in the by-items analysis (indicating that the alternative hypothesis was around 8 and 26 times more probable than the null in the two analyses, respectively). In sharp contrast, a parallel analysis on the data for the Marked primes showed a $BF_{10}$ of 0.10 in the by-participants analysis and of 0.07 in the by-items analysis, demonstrating that the alternative hypothesis was not more likely than the null.

ANOVAs on the error rates revealed a main effect of Markedness that was significant only in the analysis by participants [$F_1(1,48)=7.96$, MSE=5.07 p=.007; $F_2(1,237)=2.42$, MSE=39.77,

p=.12], suggesting that targets preceded by Marked Basque primes were recognized slightly more accurately than targets preceded by Unmarked Basque words. The rest of effects or interactions did not reach significance.

Results from Experiment 2 demonstrated that Basque-Spanish balanced simultaneous bilinguals showed significant masked translation priming effects for Spanish targets following Basque-unspecific (i.e., orthographically unmarked) Basque primes, while no such translation priming effects were found for Spanish targets following Basque words including Basque-specific bigram combinations that were not plausible in Spanish (i.e., orthographically marked). These results demonstrate that the degree of cross-language interactions and co-activation is significantly higher for words that are orthographically plausible in bilinguals' both languages than for words that are orthographically marked.

**General Discussion**

The goal of the present study was to investigate the role of sub-lexical orthographic regularities during bilingual word recognition. We explored how the presence or absence of language-specific orthographic cues modulate cross-language sub-lexical and lexical activation and the extent to which orthotactic rules between different languages sharing the same script lead to language-(non)selective lexical access. Basque-Spanish balanced bilinguals were tested in a forced-choice letter identification task using a modified version of the Reicher-Wheeler paradigm (Experiment 1) and in a lexical decision task combined with the masked translation priming paradigm (Experiment 2).

In Experiment 1, we explored the influence of sub-lexical language information at the sub-lexical and lexical levels of visual word recognition when bilinguals are immersed in a single-language experimental context (i.e., Basque). We presented Basque words and Basque-like non-words to Basque-Spanish bilingual participants who had to decide which of two letters displayed after the presentation of the stimulus was embedded in the string. Critically, half of the items (Basque words and non-words) were language-specific strings (i.e., strings with bigram combinations that did not exist in bilinguals' other language, that is, orthographically marked strings). The other half of the items corresponded to strings made of sub-lexical units (i.e., bigrams) that were legal in both of their languages (i.e., orthographically unmarked words). First, in line with previous evidence, we found that letter recognition was significantly faster and more accurate for words than non-words (i.e., the word superiority effect; see Coch & Mitra, 2010; Grainger et al., 2003; Grossi et al., 2008; McClelland, 1976, among others). This effect was not modulated by orthographic markedness (i.e., language-common or language-specific orthography) but we found that letters embedded in strings that respect the orthotactic rules of both languages (i.e., unmarked language-unspecific strings) were recognized faster and more accurately than letters embedded in strings containing language-specific sub-lexical cues (namely, an orthographic markedness effect).

Results from Experiment 1 show that orthographic information (here measured as the plausibility of bigrams in bilinguals' two languages) modulates cross-language interactions at the sub-lexical and lexical levels when participants read strings in a single-language experimental context. Unmarked Basque words and non-words activate language-unspecific sub-lexical units that in turn activate lexical representations in the bilinguals' two languages. In contrast, Basque words and Basque-like non-words whose letter combinations include language-specific cues activate

language-selective lexical orthographic representations. As a consequence, and following interactive activation principles, letter detection for orthographically unmarked (non-)words is faster and more accurate than that for orthographically marked (non-)words, mainly due to the increased feedback received by the former group of items from higher-level representations from both languages (namely, due to the larger lexical space of the former group). It should be noted that we specifically avoided using letters from language-selective bigrams as target letters. The letters to be identified always corresponded to bigram combinations that were common to both languages, and language-specificity was imposed by the context (non-target) letters. Even in these circumstances, the mere presence of language-specific sub-lexical units resulted in impoverished interaction between sub-lexical and lexical orthographic levels in a purely monolingual experimental context. Thus, the delay in response latencies and the increased error rates for orthographically marked words can be only understood as the result of the lesser degree of cross-language activation, which in turn leads to reduced top-down feedback due to language-selective lexical access (see Casaponsa et al., 2014; Orfanidou & Sumner, 2005; Rodríguez-Fornells et al., 2002).

Previous studies have shown that bilinguals are highly sensitive to the statistical regularities of words in their two languages and that they use this information in an automatic fashion while reading (see Casaponsa et al., 2014; Vaid & Frenck-Mestre, 2002; Van Kesteren et al., 2012, among others). It thus seems that bilingual readers whose languages share the same script develop fine-grained sensitivity to language-specific sub-lexical information, and that orthotactic cues lead to a different organization of lexical semantic representations for orthographically marked and unmarked words.

In Experiment 2, we then used the masked priming translation paradigm to test the former assumption and determine the extent to which bilingual lexical access is mediated by language-dependent orthographic regularities. We hypothesized that if unmarked words are organized closer in the bilingual lexicon (in a language non-selective manner) as opposed to orthographically marked words (that prompt language-selective lexical access), this would yield greater translation priming effects for the former type of words than for the latter. In other words, if sub-lexical regularities between languages modulate lexical access in bilinguals, the magnitude of the masked translation priming should be different for marked and unmarked words. Basque-Spanish bilinguals we presented with Spanish target words briefly preceded by translation equivalents in Basque or unrelated Basque masked primes. Critically, half of the primes corresponded to language-specific (i.e., marked) Basque words, and the other half corresponded to language-nonspecific (i.e., unmarked) Basque words. We observed a significant masked translation priming effect only for primes that share their sub-lexical orthographic regularities between languages (i.e., unmarked words). Indeed, no significant masked translation priming effect was found when Spanish target words are preceded by Basque primes with Basque-specific orthographic regularities[footnote 3].

Together Experiments 1 and 2 converge in showing that the degree of cross-language co-activation is notably higher for unmarked words than for marked words. The results from the masked translation priming experiment cannot be accounted for by a lack of statistical power given the large number of participants tested and the number of items used. It has been shown previously that bilinguals are highly sensitive to the statistical orthographic regularities between languages and that they use this critical information during reading (Casaponsa et al., 2014; Grainger & Beauvillain, 1987; Lemhöfer et al., 2011; Ng & Wicha, 2013; Vaid & Frenck-

Mestre, 2002; Van Kesteren et al., 2012), and it is thus reasonable to assume that words sharing similar orthographic regularities between languages are organized differently to words with clearly distinctive orthographic cues. We interpret these results as showing that marked words are not as close to their translation equivalents as unmarked words in bilingual lexical semantic memory, probably due to a (sub-)lexical organization mainly shaped by the intrinsic orthographic characteristics of each language. Indeed, it is well established that words sharing some sort of overlap, either at the orthographic, the phonological, the lexical and/or the semantic level, are closer in the lexicon than words that are unrelated at most levels of representations (see McClelland & Rumelhart, 1981; Paap, Newsome, McDonald & Schvaneveldt, 1982). Therefore, is not entirely surprising that unmarked words are closer to their translation equivalents than marked words, given that they share orthographic features (orthotactic regularities; see Grainger & Beuavillain, 1987; Dijsktra & van Heuven, 2002; Dijsktra, 2005; van Heuven & Dijkstra, 2010).

Another (complementary) interpretation of the data from Experiments 1 and 2 is that orthographically marked words (e.g., Basque masked primes with language-specific orthography) would lead to faster language detection through the use of a sub-lexical route to access language membership information (see Casaponsa et al., 2014; Vaid & Frenck-Mestre, 2000; Van Kesteren et al., 2012). In this case, the lack of masked translation priming effects for marked words could be attributed to faster inhibition of the neighboring lexical forms of language-specific words (see the BIA+ extended model; Van Kesteren et al., 2012). In the absence of such orthotactic cues, unmarked words (Basque primes) would yield lower levels of lateral inhibition, and thus their neighboring representations (e.g., Spanish translation equivalents) would still be active, leading to the observed masked translation priming effects.

Recently, Casaponsa et al. (submitted) reported ERP language switch cost priming effects only for marked Basque prime words followed by unrelated Spanish targets (as compared to a non-switching condition) in a masked language-switching paradigm, suggesting that when reading words in one of their languages, bilinguals do not show any noticeable conflict for word pairs with language-nonspecific orthotactic cues. The code-switching conflict was only evident when primes violated the orthotactic regularities of the target language. These results, together with those from Experiment 2, suggest that a common sub-lexical organization of orthographic units across the languages of a bilingual is necessary for lexical semantic effects to emerge, at least in masked priming studies. Further research and reanalysis of the materials used in previous experiments should help testing this assumption in the future.

It should be noted that Spanish and Basque are considered relatively transparent languages with high degree of similarity in their grapheme-to-phoneme correspondences. In this context, teasing apart the potentially different roles of cross-language phonotactics and orthotactics is certainly difficult. Further research on this topic testing languages with different orthographic-to-phonological correspondences are needed in order to shed light on the independent contribution of phonotactics and othotactics in bilingual lexical organization.

Both the interpretations presented above are not mutually exclusive and both phenomena could co-exist in bilingual reading, i.e., unmarked words would engage prominent cross-language interactions at the same time as orthographically marked words would yield a greater degree of language-specific lexical access. Hence, we can hypothesize that the organization of the bilingual lexicon and the modalities of access to stored mental representations are shaped by sub-lexical distributional probabilities within and between languages. Further research should explore not only the way in which sub-lexically marked and unmarked words are represented in the bilingual

lexicon of native or native-like bilinguals, but also the way in which these different types of representations are learned, integrated and consolidated in the process of second language acquisition.

**References**

Altarriba, J., & Basnight-Brown, D. M. (2007). Methodological considerations in performing semantic-and translation-priming experiments across languages. *Behavior Research Methods, 39*(1), 1-18. doi: 10.3758/BF03192839

Casaponsa, A., Antón, E., Pérez, A., & Duñabeitia, J.A. (in press). Foreign language comprehension achievement: insights from the cognate facilitation effect. *Frontiers in Psychology.*

Casaponsa, A., Carreiras, M., & Duñabeitia, J. A. (2014). Discriminating languages in bilingual contexts: the impact of orthographic markedness. *Frontiers in Psychology, 5*. doi: 10.3389/fpsyg.2014.00424

Casaponsa, A, Carreiras, M, & Duñabeitia, J. A. (submitted). How do bilinguals detect the language of the words they read?

Coch, D., & Mitra, P. (2010). Word and pseudoword superiority effects reflected in the ERP waveform. *Brain Res, 1329*, 159-174. doi: 10.1016/j.brainres.2010.02.084

Davis,C.J.,& Perea,M.(2005).BuscaPalabras: a program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavioral Research Methods* 37, 665–671.doi:10.3758/BF03192738

De Groot, A., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language, 30*(1), 90-123. doi:10.1016/0749-596X(91)90012-9

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(03), 175–197. doi: 10.1017/s1366728902003012

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. De Groot (Eds.), Handbook of bilingualism: Psycholinguistic approaches (pp. 178–201). Oxford, UK: Oxford University Press.

Dimitropoulou, M., Dunabeitia, J. A., & Carreiras, M. (2011a). Masked translation priming effects with low proficient bilinguals. *Memory & Cognition, 39*(2), 260-275. doi: 10.3758/s13421-010-0004-9

Dimitropoulou, M., Dunabeitia, J. A., & Carreiras, M. (2011b). Two words, one meaning: evidence of automatic co-activation of translation equivalents. *Frontiers in Psychology, 2,* 188. doi: 10.3389/fpsyg.2011.00188

Dunabeitia, J. A., Dimitropoulou, M., Uribe-Etxebarria, O., Laka, I., & Carreiras, M. (2010). Electrophysiological correlates of the masked translation priming effect with highly proficient simultaneous bilinguals. *Brain Research, 1359,* 142-154. doi: 10.1016/j.brainres.2010.08.066

Dunabeitia, J. A., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology, 57*(2), 98-107. doi: 10.1027/1618-3169/a000013

Duñabeitia, J. A., Cholin, J., Corral, J., Perea, M., & Carreiras, M. (2010). SYLLABARIUM: An online application for deriving complete statistics for Basque and Spanish orthographic syllables. *Behavior research methods*, 42(1), 118-125. doi: 10.3758/BRM.42.1.118

Duñabeitia, J. A., Dimitropoulou, M., Morris, J., & Diependaele, K. (2013). The role of form in morphological priming: Evidence from bilinguals. *Language and Cognitive Processes* 28,

967-987.  doi: 10.1080/01690965.2012.713972

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language: new evidence from Dutch-French bilinguals. *Experimental Psychology, 56*(3), 173-179. doi: 10.1027/1618-3169.56.3.173

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1122. doi: 10.1037/0278-7393.23.5.1122

Grainger, J., Bouttevin, S., Truc, C., Bastien, M., & Ziegler, J. (2003). Word superiority, pseudoword superiority, and learning to read: a comparison of dyslexic and normal readers. *Brain Language, 87*(3), 432-440.  doi:10.1016/S0093-934X(03)00145-7

Grainger, J. & Jacobs, A.M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1158-1176. doi: 10.1037/0096-1523.20.6.1158

Grainger, J. Midgley, K.J., & Holcomb, P.J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail and M. Hickman(Eds), *Language Acquisition across linguistic and cognitive system*s. Philadelphia: John Benjamins (pp. 267-284)

Grossi, G., Murphy, J., & Boggan, J. (2008). Word and pseudoword superiority effects in Italian–English bilinguals. *Bilingualism: Language and Cognition, 12*(01), 113. doi: 10.1017/s1366728908003891

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-

language priming. *Bilingualism: Language and Cognition*, *2*(01), 59-75.

Jiang, N., & Forster, K. I. (2001). Cross-Language Priming Asymmetries in Lexical Decision and Episodic Recognition. *Journal of Memory and Language, 44*(1), 32-51. doi: 10.1006/jmla.2000.2737

Ju, M., & Luce, P. A. (2004). Falling on Sensitive Ears Constraints on Bilingual Lexical Activation. *Psychological Science*, *15*(5), 314-318. doi: 10.1111/j.0956-7976.2004.00675.x

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* 42(3), 627-633. doi: 10.3758/BRM.42.3.627

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language*, *49*(4), 484-499. doi:10.1016/S0749-596X(03)00093-7

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition, 32*(4), 533-550. doi: 10.3758/BF03195845

Lemhofer, K., Koester, D., & Schreuder, R. (2011). When bicycle pump is harder to read than bicycle bell: effects of parsing cues in first and second language compound reading. *Psychon Bulletin Review, 18*(2), 364-370. doi: 10.3758/s13423-010-0044-y

Lemhöfer, K., & Radach, R. (2009). Task context effects in bilingual nonword processing. *Experimental Psychology*, *56*(1), 41. doi: 10.1027/1618-3169.56.1.41.

McClelland, J. L. (1976). Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 80. doi: 10.1037/0096-1523.2.1.80

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context

effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375. doi: 10.1037/0033-295X.88.5.375

Ng, S., & Wicha, N. Y. (2013). Meaning first: a case for language-independent access to word meaning in the bilingual brain. *Neuropsychologia, 51*(5), 850-863. doi: 10.1016/j.neuropsychologia.2013.01.017

Orfanidou, E., & Sumner, P. (2005). Language switching and the effects of orthographic specificity and response repetition. *Memory & Cognition, 33*(2), 355-369. doi: 10.3758/BF03195323

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological review*, *89*(5), 573. doi: 10.1037/0033-295X.89.5.573

Perea, M., Urkia, M., Davis, C. J., Agirre, A., Laseka, E., & Carreiras, M. (2006). E-Hitz: A word-frequency list and a program for deriving psycholinguistic statistics in an agglutinative language (Basque). *Behavior Research Methods*, 38, 610-615. doi: 10.1016/j.jml.2008.01.003

Pollatsek, A., and Well, A. (1995). On the use of counterbalanced designs in cognitive research: a suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory & Cognition*. 21, 785–794. doi: 10.1037/0278-7393.21.3.785

Prinzmetal, W. (1992). The word-superiority effect does not require a T-scope. *Perception & Psychophysics*, *51*(5), 473-484. doi: 10.3758/BF03211643

Reicher, G. M. (1969). Perceptual recognition as a function of meaninfulness of stimulus material. *Journal of Experimental Psychology, 81*(2), 275-280. doi: 10.1037/h0027768

Rodríguez-Fornells, A., Rotte, M., Heinze, H. J., Nösselt, T., & Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one

brain. *Nature*, *415*(6875), 1026-1029. doi:10.1038/4151026a

Rouder, J. N., Speckman P. L., Sun D., Morey R. D., & Iverson G. (2009). Bayesian t-Tests for Accepting and Rejecting the Null Hypothesis. *Psychonomic Bulletin & Review, 16*, 225-237. doi: 10.3758/PBR.16.2.225

Schepens, J., Dijkstra T., & Grootjen, F. (2011). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition* 15: 157-166. doi: 10.1017/s1366728910000623

Sebastián-Gallés, N., Martí, M., Carreiras, M., & Cuetos, F. (2000). LEXESP: Una base de datos informatizada del español. *Universitat de Barcelona, Barcelona.*

Tydgat, I., & Grainger, J. (2009). Serial position effects in the identification of letters, digits and symbols. *Journal of Experimental Psychology: Human Perception and Performance, 35(2),* 480-498. doi: 10.1037/a0013027

Vaid, J., & Frenck-Mestre, C. (2002). Do orthographic cues aid language recognition? A laterality study with French–English bilinguals. *Brain and Language, 82*(1), 47-53. doi: 10.1016/S0093-934X(02)00008-1

van Heuven, W. J., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Review, 64*(1), 104-122. doi: 10.1016/j.brainresrev.2010.03.002

Van Kesteren, R., Dijkstra, T., & de Smedt, K. (2012). Markedness effects in Norwegian-English bilinguals: task-dependent use of language-specific letters and bigrams. *Q J Exp Psychol (Hove), 65*(11), 2129-2154. doi: 10.1080/17470218.2012.679946

Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition, 35*(5), 938-952.  doi: 10.3758/BF03193467

Wetzels, R., Matzke D., Lee M. D., Rouder J. N., Iverson G. J., & Wagenmakers E. J. (2011).  Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on Psychological Science, 6*, 291-298.  doi: 10.1177/1745691611406923

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology,1*(1), 59-85. doi:10.1016/0010-0285(70)90005-8

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: eye tracking, moving-window, and maze. *Joural of Psycholinguistic Research, 41*(2), 105-128. doi: 10.1007/s10936-011-9179-x

# Appendices

**Appendix 1.** Materials used in Experiment 1.

| NON-WORDS | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Marked** | | | **Unmarked** | | | **Marked** | | | **Unmarked** | | |
| **Prime** | **Target** | **Foil** | **Prime** | **Target** | **Foil** | **Prime** | **Target** | **Foil** | **Prime** | **Target** | **Foil** |
| azats | z | h | gornen | r | d | akats | k | n | emari | m | d |
| betsa | e | i | olduri | l | z | lotsa | o | i | ageri | g | h |
| bruka | r | l | betain | e | o | truke | r | g | geldi | e | a |
| aliko | l | r | soren | o | i | hauts | a | i | hodei | o | u |
| pirtsa | r | d | gagur | a | e | ohiko | h | p | ipuin | p | r |
| nenko | e | i | idora | d | n | altxor | l | s | isuri | s | g |
| areko | r | z | idore | d | m | erdiko | d | t | egungo | u | i |
| ezets | z | p | palbai | l | m | mutiko | t | b | adimen | d | s |
| bezkor | z | t | hargile | r | d | norako | r | p | belaun | l | p |
| etrako | t | n | uduri | d | m | pentsu | n | b | bidaia | d | t |
| partsi | r | n | kigare | g | t | ahaltsu | h | m | gailur | a | o |
| andiko | d | r | ordari | d | m | aldakor | d | s | garden | r | l |
| zinbra | i | u | ozapin | z | l | antzeko | n | p | ilargi | l | d |
| adezko | d | t | erapi | r | s | aurkako | u | i | masail | s | r |
| arozkun | r | b | blodatu | o | e | egiazko | g | n | sorgin | o | a |
| araleko | r | n | etuarpe | u | i | egonkor | g | z | alargun | l | d |
| lamazka | m | t | lagerre | g | t | gatazka | t | d | garesti | a | o |
| entriko | n | l | esmaera | s | p | hildako | l | g | hasiera | s | t |
| mirreko | i | a | erliezi | r | n | jatetxe | a | o | ontasun | t | p |
| ultazko | l | r | neuratu | e | o | mingots | n | k | ostiral | t | p |
| aistako | s | r | datiera | t | z | neketsu | k | g | ateraldi | t | s |
| lorinzko | r | d | ulmorio | m | z | ordezko | r | n | ezaguera | z | l |
| haidekoz | d | l | moxitari | x | s | azoka | z | b | hasperen | p | t |
| litzazko | t | d | saildare | l | m | ereduzko | d | h | hondamen | d | t |
| hungako | n | r | gergarri | e | o | goranzko | r | d | ibilaldi | b | z |
| hastixka | t | z | lazgeren | g | t | jakintsu | k | l | jostailu | s | r |
| bersinbo | s | t | fatorri | t | z | maitekor | t | n | orrialde | i | u |
| trantsus | r | k | armatigo | m | s | amets | m | h | sendagai | n | l |
| atxela | l | s | jentu | t | s | kontu | t | d | berme | m | t |
| txisu | s | l | arapi | p | z | txalo | l | r | gernu | n | k |

33

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| txanak | n | z | biton | o | s | bizkar | a | o | onura | r | b |
| kosti | t | n | dahaila | l | n | jainko | n | b | esaldi | d | h |
| atxona | o | e | hermen | m | t | txano | n | l | ubide | d | r |
| lonpru | r | s | irruta | t | m | kopuru | r | s | gosari | r | k |
| uzkinu | n | r | sapuri | r | l | txapel | p | s | inguru | r | t |
| txandu | d | z | lilobe | b | z | bokazio | z | t | beldur | d | m |
| uzarok | r | d | errilba | b | z | errauts | u | o | errota | t | n |
| itsatun | u | e | narela | l | g | gerriko | i | a | gidari | r | s |
| sanbeta | t | k | ontapin | p | z | hozkada | d | n | salgai | g | m |
| pukiela | l | d | anirpen | p | t | jokaera | r | d | zorion | i | e |
| taltsio | i | e | serrezi | z | t | lanbide | d | r | altuera | r | z |
| dokioti | t | s | errazai | z | p | etsai | a | o | lodiera | r | s |
| lukandi | n | t | soizane | n | b | txosten | e | a | bisitari | t | d |
| kokasun | s | t | birregi | g | m | aginpide | d | t | deigarri | a | u |
| ezjikan | k | r | batapin | p | s | alokairu | r | s | epaitegi | g | l |
| bitzenba | e | o | arribine | n | t | ekoizpen | z | g | hedadura | r | s |
| koirrisa | s | n | nentagun | g | s | jokabide | b | z | mirespen | p | t |
| emoketa | t | s | baleu | e | i | koadrila | r | h | omenaldi | l | g |
| baluketa | e | i | garaiman | m | s | korapilo | l | d | sarraila | l | z |
| izenpeta | t | r | errizoi | z | p | tximinia | n | k | sorbalda | d | k |

**Appendix 2.** Materials used in Experiment 2.

| MARKED | | UNMARKED | |
|---|---|---|---|
| **Related prime** | **TARGET** | **Related prime** | **TARGET** |
| antzeko | SEMEJANTE | ipuin | CUENTO |
| jakintsu | SABIO | berme | GARANTÍA |
| baserriko | CAMPESINO | errementari | HERRERO |
| hozkada | BOCADO | epaitegi | JUZGADO |
| ezkongai | SOLTERO | aberastasun | RIQUEZA |
| aurkako | CONTRARIO | ostiral | VIERNES |
| koba | CUEVA | orrialde | PÁGINA |
| maitekor | CARIÑOSO | inguru | CERCO |
| arratsalde | TARDE | margolari | PINTOR |
| doako | GRATUITA | heldutasun | MADUREZ |
| jokabide | PROCEDER | arintasun | LIGEREZA |
| goranzko | ASCENDENTE | nobelagile | NOVELISTA |
| aginpide | AUTORIDAD | adimen | INTELIGENCIA |
| korridore | PASILLO | maisutasun | MAESTRÍA |
| atseden | DESCANSO | onargarri | ACEPTABLE |
| abokatu | ABOGADO | gorespen | EXALTACIÓN |
| kanpai | CAMPANA | ibilaldi | EXCURSIÓN |
| zuku | JUGO | handitasun | GRANDEZA |
| mendebaldeko | OCCIDENTAL | bateragarri | COMPATIBLE |
| ordezkari | DELEGADO | arreba | HERMANA |
| erreakzio | REACCIÓN | belaun | RODILLA |
| errauts | CENIZA | geldi | QUIETO |
| elkarrekiko | MUTUO | mugimendu | MOVIMIENTO |
| soka | CUERDA | ilargi | LUNA |
| mendekotasun | DEPENDENCIA | estutasun | ANGUSTIA |
| neketsu | AGOTADOR | goraipamen | ELOGIO |
| zinpeko | JURADO | iluntasun | OSCURIDAD |
| mingots | AMARGO | santutegi | SANTUARIO |
| lokarri | CORDÓN | larrialdi | URGENCIA |
| norako | DESTINO | zabal | ANCHO |
| oinordeko | HEREDERO | argitasun | CLARIDAD |
| korapilo | NUDO | errebolta | REVUELTA |
| kontalari | NARRADOR | errentagarri | RENTABLE |
| koadrila | PANDILLA | garden | TRANSPARENTE |
| konplizitate | COMPLICIDAD | sorbalda | HOMBRO |
| ezinezko | IMPOSIBLE | epel | TIBIO |
| itxurazko | APARENTE | apaltasun | HUMILDAD |

| | | | |
|---|---|---|---|
| baliokide | EQUIVALENTE | gernu | ORINA |
| bokazio | VOCACIÓN | udal | AYUNTAMIENTO |
| ohiko | HABITUAL | usaimen | OLFATO |
| lanbide | PROFESIÓN | hasperen | SUSPIRO |
| ekialdeko | ORIENTAL | zale | AFICIONADO |
| hizketa | DISCURSO | seinale | INDICACIÓN |
| burrunba | ZUMBIDO | naturaltasun | NATURALIDAD |
| sakontasun | PROFUNDIDAD | porrot | QUIEBRA |
| hirukoitz | TRIPLE | ondorio | CONSECUENCIA |
| eguneroko | COTIDIANO | lasai | TRANQUILO |
| ezkontza | MATRIMONIO | mirespen | ADMIRACIÓN |
| mutiko | CHIQUILLO | altuera | TALLA |
| kontzentratu | CONCENTRADO | isurialde | VERTIENTE |
| ezker | IZQUIERDA | alai | ALEGRE |
| garrantzizko | IMPORTANTE | basamortu | DESIERTO |
| aurrekontu | PRESUPUESTO | garesti | CARO |
| tximinia | CHIMENEA | adeitasun | AMABILIDAD |
| aldizkari | REVISTA | isun | MULTA |
| pentsu | PIENSO | hodei | NUBE |
| orokor | GENERAL | harridura | EXTRAÑEZA |
| minbera | SUSCEPTIBLE | baliotasun | VALIDEZ |
| faltsukeria | FALSEDAD | zorion | DICHA |
| egonkortasun | ESTABILIDAD | aireportu | AEROPUERTO |
| jainko | DIOS | masail | MEJILLA |
| mehatxu | AMENAZA | segurtasun | SEGURIDAD |
| orrazketa | PEINADO | haurdunaldi | EMBARAZO |
| txapel | BOINA | isiltasun | SILENCIO |
| bertsio | VERSIÓN | errota | MOLINO |
| ordezko | SUSTITUTO | eramaile | PORTADOR |
| kopuru | CUANTÍA | sendagai | MEDICAMENTO |
| bitxi | CURIOSO | garapen | DESARROLLO |
| hauts | POLVO | iraingarri | HUMILLANTE |
| ekoizpen | PRODUCCIÓN | harri | PIEDRA |
| gonbidatu | INVITADO | beldur | MIEDO |
| txalo | APLAUSO | ondorengo | SUCESOR |
| bazkide | SOCIO | desadostasun | DISCREPANCIA |
| jokaera | ACTUACIÓN | bidaia | VIAJE |
| egiturazko | ESTRUCTURAL | sudur | NARIZ |
| gatazka | CONFLICTO | lapurreta | ROBO |
| zorioneko | AFORTUNADO | lasaitasun | TRANQUILIDAD |
| aldakor | CAMBIANTE | larruazal | PIEL |
| adeitsu | CORTÉS | jostailu | JUGUETE |
| txosten | INFORME | egungo | ACTUAL |

| | | | |
|---|---|---|---|
| edukitze | POSESIÓN | senidetasun | FRATERNIDAD |
| ikastetxe | COLEGIO | malgutasun | FLEXIBILIDAD |
| nahitaezko | OBLIGATORIO | zailtasun | DIFICULTAD |
| bizkar | ESPALDA | hondamen | RUINA |
| ustekabeko | INESPERADO | garagardo | CERVEZA |
| kanpoko | EXTERIOR | egiatasun | VERACIDAD |
| aukeraketa | SELECCIÓN | geldialdi | PARADA |
| altxor | TESORO | gidari | CONDUCTOR |
| erdiko | CENTRAL | salgai | MERCANCÍA |
| marrazki | DIBUJO | gailur | CIMA |
| akats | DEFECTO | barealdi | CALMA |
| harrokeria | VANIDAD | esaldi | FRASE |
| kode | CLAVE | sorgin | BRUJA |
| kutxatila | ESTUCHE | bisitari | VISITANTE |
| ahaltsu | POTENTE | zaletasun | AFICIÓN |
| gerriko | CINTURÓN | alargun | VIUDA |
| idazkari | SECRETARIO | isuri | FLUJO |
| alokairu | ALQUILER | ontasun | BONDAD |
| txantxa | BROMA | gosari | DESAYUNO |
| zoriontsu | FELIZ | deigarri | LLAMATIVO |
| indarreko | VIGENTE | moraltasun | MORALIDAD |
| egiazko | VERDADERO | onura | BENEFICIO |
| bizkortasun | AGILIDAD | hasiera | PRINCIPIO |
| erruki | PIEDAD | sarraila | CERRADURA |
| kontzertu | CONCIERTO | ageri | MANIFIESTO |
| konposatu | COMPUESTO | lodiera | ESPESOR |
| funtsezko | FUNDAMENTAL | testuinguru | CONTEXTO |
| zalantzazko | DUDOSO | arteraldi | OCURRENCIA |
| iruzkin | COMENTARIO | garbitasun | LIMPIEZA |
| zuzeneko | DIRECTA | emari | CAUDAL |
| egonkor | ESTABLE | ezaguera | CONOCIMIENTO |
| boteretsu | PODEROSO | errealitate | REALIDAD |
| ereduzko | EJEMPLAR | indar | FUERZA |
| bizkarralde | RESPALDO | susmo | SOSPECHA |
| hildako | FALLECIDO | hedadura | EXTENSIÓN |
| jatetxe | RESTAURANTE | zabaltasun | AMPLITUD |
| nazkagarri | REPUGNANTE | liluragarri | FASCINANTE |
| pentsakor | PENSATIVO | urruntasun | LEJANÍA |
| kontzientzia | CONCIENCIA | eraso | ATAQUE |
| sutsu | ARDIENTE | omenaldi | HOMENAJE |

# Tables

**Table 1.** Mean Spanish and Basque proficiency scores calculated according to participants' self-ratings (in a 1-to-10 scale), vocabulary knowledge (maximum of 77) and judges' assessment in an interview (in a 1-to-5 scale). Standard deviations are provided in parentheses.

| Language proficiency | Spanish | Basque |
| --- | --- | --- |
| Vocabulary knowledge | 76.42 (1.03) | 74.20 (2.45) |
| Interview | 5.00 (0) | 5.00 (0) |
| *Self-perception scores* | | |
| Speaking | 9.50 (.79) | 9.50 (.65) |
| Understanding | 9.62 (.67) | 9.69 (.55) |
| Writing | 9.16 (1.04) | 9.36 (.88) |
| Reading | 9.51 (.87) | 9.57 (.65) |
| General self-perception | 9.40 (.88) | 9.48 (.68) |

**Table 2.** Characteristics of the materials used in Experiment 1 by condition. Standard deviations are provided in parentheses.

| | WORDS | | NON-WORDS | |
|---|---|---|---|---|
| *Task Language (Basque)* | Marked | Unmarked | Marked | Unmarked |
| Word Frequency | 39.65 (47.59) | 40.94 (58.65) | - | - |
| Word Length | 6.60 (1.07) | 6.58 (1.14) | 6.56 (1.01) | 6.52 (.99) |
| Number of Orthographic Neighbors | 1.20 (1.39) | 1.10 (1.50) | 1.04 (1.80) | .90 (1.36) |
| OLD 20 | 2.43 (.45) | 2.43 (.41) | 2.41 (.49) | 2.49 (.47) |
| Age of Acquisition | 3.25 (0.48) | 3.23 (0.54) | - | - |
| Word Concreteness | 3.97 (0.91) | 4.12 (0.88) | - | - |
| Log10 Bigram Frequency | 2.91 (.14) | 2.92 (.16) | 2.93 (.16) | 2.95 (.12) |
| Bigram Frequency | 2.11 (.34) | 2.14 (.25) | 2.09 (.29) | 2.09 (.35) |
| *Hidden Language (Spanish)* | | | | |
| Number of Orthographic Neighbors | .02 (.14) | .08 (.27) | .18 (.72) | .08 (.44) |
| OLD 20 | 3.12 (.67) | 2.96 (.55) | 3.05 (.64) | 2.95 (.52) |
| Log 10 Bigram Frequency | 2.10 (0.37)* | 3.08 (0.19) | 2.16 (.31)* | 3.06 (.20) |
| Bigram Frequency | 1.31 (0.41)* | 1.79 (.50) | 1.35 (.41)* | 1.84 (.47) |
| Number of Implausible Bigrams | 1.08 (0.27)* | 0 (0) | 1.12 (.32)* | 0 (0) |

*Note: Log10 Bigram Frequency corresponds to the logarithmic transformation (base 10) of bigram counts from the Basque words contained in the Syllabarium database (Duñabeitia et al., 2010) and from the LEXESP database (Sebastián-Gallés et al, 2000) for Spanish. Bigram Frequency corresponds to the logarithmic transformation of bigram frequencies dependent of word length and position taken from E-Hitz database (Perea et al., 2006) and B-Pal (Davis & Perea, 2005). Asterisks indicate significant statistical differences between Marked and Unmarked conditions.*

**Table 3.** Mean latencies (in milliseconds) and error rates (in percentage) in the four conditions tested in Experiment 1. Standard deviations are provided in parenthesis.

| | Words | | Non-words | | Word superiority effect | |
|---|---|---|---|---|---|---|
| | RT | Error rate | RT | Error rate | RT | Error rate |
| Unmarked | 828 (124) | 6.26 (6.12) | 863 (151) | 10.95 (8.61) | **35** | 4.69 |
| Marked | 852 (131) | 5.53 (5.80) | 890 (157) | 12.15 (8.68) | **38** | 6.62 |
| Plausibility effect | **24** | -0.73 | **27** | 1.20 | | |

**Table 4.** Characteristics of the materials used in Experiment 2 by condition. Standard deviations are provided in parentheses.

| | PRIMES | | TARGETS | |
|---|---|---|---|---|
| | Marked | Unmarked | Marked-Pair | Unmarked-Pair |
| Word Frequency | 30.10 (36.71) | 36.04 (57.28) | 34.79 (46.22) | 31.28 (39.37) |
| Word Length | 7.91 (1.91) | 7.77 (2.06) | 7.87 (1.80) | 7.73 (2.13) |
| Age of Acquisition | 3.34 (.44) | 3.29 (.53) | 3.28 (.51) | 3.20 (.57) |
| Word Concreteness | 3.92 (.81) | 3.89 (.86) | 3.82 (.81) | 3.84 (.87) |
| Spanish Orthographic Neighbors | .14 (1.02) | .18 (.59) | 1.00 (1.44) | 1.20 (2.30) |
| Spanish OLD 20 | 3.87 (1.08) | 3.65 (1.20) | 2.61 (.62) | 2.62 (.73) |
| Spanish Log10 Bigram Frequency | 2.17 (.42)* | 3.10 (.17) | 3.14 (.19) | 3.10 (.22) |
| Spanish Bigram Frequency | 1.20 (.42)* | 1.67 (.48) | 2.51 (.29) | 2.48 (.29) |
| Number of Spanish-Implausible bigrams | 1.23 (.48)* | 0 (0) | - | - |
| Basque Orthographic Neighbors | .84 (.182) | .97 (.162) | - | - |
| Basque OLD 20 | 2.83 (.70) | 2.73 (.68) | - | - |
| Basque Log10 Bigram Frequency | 2.96 (.13) | 2.97 (.15) | - | - |
| Basque Bigram Frequency | 2.07 (.31) | 2.04 (.29) | - | - |

*Note: Asterisks indicate significant differences between Marked and Unmarked conditions.*

**Table 5.** Mean latencies (in milliseconds) and error rates (in percentage) in each condition tested in Experiment 2. Standard deviations are provided in parenthesis.

| | Related | | Unrelated | | Translation effect | |
|---|---|---|---|---|---|---|
| | RT | Error rate | RT | Error rate | RT | Error rate |
| Marked | 648 (96) | 2.84 (3.23) | 645 (94) | 3.10 (3.33) | **-3** | .27 |
| Unmarked | 643 (86) | 4.07 (4.37) | 654 (93) | 3.67 (4.16) | **11** | -.40 |

**Figure 1.** Schematic representation of a trial in Experiment 1.

Fixation (500 ms)

Referent (200 ms)

EMARI

Targets (max.2500 ms)
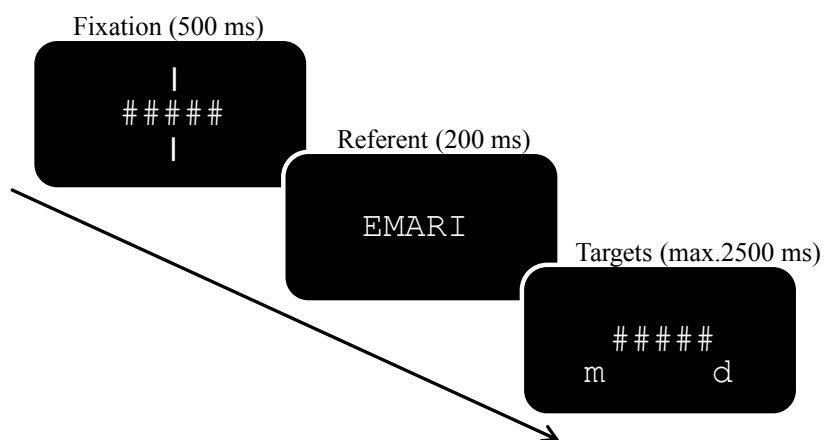
# Footnotes

1. The vocabulary test was an adaptation of the Boston Naming Test with 77 pictures corresponding to 77 non-cognate Basque-Spanish translation equivalents. Participants had to name each of the pictures in each of their two languages (see Gollan et al., 2012, for a similar procedure and validation). The structured individual interviews were conducted by a native Basque-Spanish bilingual linguist in order to assess participants' communicative skills in each language (see also Gollan et al.).

2. The words used in Experiment 1 were also included within the set of items used in Experiment 2. In order to minimize any potential effect of word, participants completed Experiment 2 first, immediately followed by Experiment 1. This way, the items that were repeated were first presented as unconsciously perceived masked primes, and then presented as consciously processed targets, minimizing any potential influence of long-lag repetition effects.

3. In order to explore whether masked translation priming effects for marked words changed as a function of the position of the critical bigrams within the string, we run a correlation analysis between the priming effects and bigrams' position for the marked words containing a single distinctive bigram (96 out of 120 items). The correlation was not significant [r=-.097, p=.35], suggesting that the lack of masked translation priming effects for marked words does not depend on the position of the critical bigram within the string. Besides, we also explored whether the masked translation priming effects could vary as a function of the consonant-vowel structure of the critical bigrams in the marked condition. The mean priming effect for bigrams made of a consonantal cluster was -7ms, and it was 6ms for bigrams including a consonant and a vowel, but this difference was not significant (p>.17).