J Cult Cogn Sci (2022) 6:219–228
https://doi.org/10.1007/s41809-022-00104-0

RESEARCH PAPER

Does adding an accent mark hinder lexical access? Evidence from Spanish

Melanie Labusch · Pablo Gómez · Manuel Perea

Received: 4 December 2021 / Revised: 7 April 2022 / Accepted: 4 May 2022 / Published online: 20 June 2022

Abstract Recent research has shown that omitting the accent mark in a Spanish word, which is a language in which these diacritics only indicate lexical stress, does not cause a delay in lexical access (e.g., *cárcel* [prison] ≈ *carcel*; *cárcel-CÁRCEL* ≈ *carcel-CÁRCEL*). This pattern has been interpreted as accented and nonaccented vowels sharing the abstract letter representations in Spanish. However, adding an accent mark to a nonaccented Spanish word appears to produce a reading cost in masked priming paradigms (e.g., *féliz-FELIZ* [happy] > *feliz-FELIZ*). We examined whether adding an accent mark to a non accented Spanish word slows down lexical access in two semantic categorization experiments to solve this puzzle. We added an accent mark either on the stressed syllable (Experiment 1, e.g., *cébra* for the word *cebra* [zebra]) or an unstressed syllable (Experiment 2, e.g., *cebrá*). While effect sizes were small in magnitude, adding an accent mark produced a cost relative to the intact words, especially when the accent mark was added on an unstressed syllable (*cebrá* > *cebra*). These findings favor the view that letter identity and (to a lesser extent) accent mark information are encoded during word recognition in Spanish. We also examined the practical implications of these results.

Keywords Word recognition · Lexical access · Reading · Accent marks

All data and scripts are available at the following OSF link: https://osf.io/pr42j/?view_only=fe00b77d115143f2b5d1a8cb101d9ccc.

M. Labusch · M. Perea (✉)
Departamento de Metodología, Universitat de València, Av. Blasco Ibáñez, 21, 46010 Valencia, Spain
e-mail: mperea@uv.es

M. Labusch · M. Perea
Universidad Antonio de Nebrija, Madrid, Spain

P. Gómez
California State University, San Bernardino, Palm Desert Campus, Palm Desert, USA

Introduction

Most European languages employ the Latin alphabet in writing. As there is a huge variety among these languages, most of them have elements that cannot be directly captured by the letters of Latin script (e.g., phonemes that did not exist in Latin, information on lexical stress, among others). To compensate for these idiosyncrasies, it is uncommon to create letters *ex novo*—one exception is the letter þ in Icelandic for the phoneme /θ/—or to *recycle* other Latin characters like *ſs* in the German letter ß for the phoneme /s/ after a long vowel or a diphthong. Instead, a much more common choice in Latin-based orthographies is adding a diacritical mark to some of the original letters (e.g., the diacritic letter š /ʃ/ in Czech). While not the focus here, another option is to use letter combinations (e.g., *sh* in English corresponds to the phoneme /ʃ/) (see Chetail, 2020, for a recent review on the role of graphemes in visual-word recognition).[1]

The presence of diacritical letters in most languages raises the question of the representations of these letters in the word recognition system (e.g., see Chetail & Boursain, 2019; Perea et al., 2020). However, due to the Anglo-centricity of cognitive sciences (see Share, 2008), nearly all leading visual word recognition and reading models have only focused on the English orthography (e.g., Dehaene et al., 2005; Grainger et al., 2008). Thus, most of these models remain silent about the representation of diacritical letters in the word recognition system (see Ans et al., 1998; Hutzler et al., 2004, for exceptions in French and German, respectively). An ongoing research question is how to expand models of visual word recognition to cope with the representation of diacritical letters (Chetail & Boursain, 2019; Kinoshita et al., 2021; Marcet et al., 2021b; Perea et al., 2020, 2021; Schwab, 2015). In the case of consonants, one might reasonably conceive that diacritical consonants correspond to separate abstract letter units (e.g., *š* vs. *s* in Czech, *ñ* vs. *n* in Spanish, etc.) that simply are visually similar like O and Q or i and j.

Notably, the above scenario is more complicated for vowels because the function of diacritics on vowels varies enormously across languages (e.g., vowel quality, vowel length, lexical stress, tone information, among others). In one extreme, there are languages like German in which diacritical and non-diacritical vowels correspond to different phonemes (e.g., compare *a* /a/ vs. *ä* /ɛ/). In this case, it is reasonable to assume that accented and nonaccented vowels have separate abstract letter representations (see Hutzler et al., 2004, for modeling; see Perea et al., 2021, for empirical evidence). In the other extreme, there are languages like Spanish—which is the focus of the present research—in which diacritical marks only indicate the stressed syllable without altering phonemic information. For instance, *alcázar* [fortress in Spanish] is pronounced [al.ˈka.θaɾ], whereas the word *ácrata* [libertarian in Spanish] is pronounced [ˈa.kɾa.ta]).

Given that lexical stress in Spanish falls around 80% of the time in the last-but-one syllable and to apply the principle of economy, the general logic of the accentuation rules in Spanish is that the number of Spanish words with an accent mark should be as few as possible (see Appendix A for an overview of accentuation rules in Spanish). For instance, most Spanish words ending with the letter "n" are pronounced in the last-but-one syllable (e.g., *joven* [young]) and the accent mark only applies to words ending in "n" with lexical stress on the final syllable (e.g., *común* [common]) or the second-to-last syllable (e.g., *oxímoron* [oxymoron]). Thus, the proportion of words that require an accent mark in Spanish is relatively small. This figure is less than 12.5% of the 500 most frequent words in Spanish (Davis & Perea, 2005). That is, in most Spanish words, the designation of the stressed syllable is most often not marked explicitly.

Due to the particular role of diacritics in Spanish vowels, it is generally assumed that both accented and nonaccented vowels activate the same abstract letter representations (Chetail & Boursain, 2019; Schwab, 2015). That is, lexical access for the word *cárcel* (prison) and its nonaccented counterpart *carcel* would be essentially the same. Indeed, for accented Spanish words, the omission of their accent marks only has a minimal cost on word recognition. For instance, when omitting the accent mark in a masked priming lexical decision task, Perea et al. (2020) found that word identification times to an accented target word like *CÁRCEL* (prison, in Spanish) were essentially the same when preceded by the primes *cárcel* or *carcel*. Likewise, in a semantic categorization task ("is the word an animal name?"), *cárcel* and *carcel* also produced similar response times (Perea et al., 2021). A similar pattern occurred in an unprimed lexical decision task ("is the stimulus a word?") when the experimental blocks contained either accented items (e.g., the word *cárcel*, the pseudoword *nárdil*) or nonaccented items (e.g., *carcel*, *nardil*) (Marcet et al., 2021b; Schwab, 2015). Finally, a similar pattern occurs when omitting the accent mark from a target word during sentence reading (Marcet & Perea, 2022): first-pass fixation times (e.g., the sum of the fixations during the first pass) were virtually the same for words like *cárcel*

[1] The letters j, u, and w are not native to Latin. They were created as variations of the letters i and v, and they are considered an integral part of Latin script since the XVII century.

and *carcel*. Overall, this pattern is consistent with the idea that vowels like *a* and *á* share the same abstract letter representations in Spanish.

However, the role of accent marks in Spanish may not be as straightforward as it appears. Prior empirical evidence concerning adding an accent mark to an otherwise nonaccented word in Spanish appears to show a reliable reading cost. For instance, in a masked priming lexical decision experiment, Perea et al. (2020) found that a nonaccented word like *FELIZ* [happy]—which is pronounced [fe.ˈliθ]—was responded faster when preceded by the identity prime *feliz* than when preceded by a prime that was the same as the identity prime except for an added accent mark on an unstressed syllable (*féliz*, pronounced as [ˈfe.liθ]). Similarly, in other masked priming experiments, Domínguez and Cuetos (2018) and Perea et al. (2021) found some reading cost due to the primes having an accent mark (e.g., *rasgo-RASGO* faster than *rasgó-RASGO*; *obaeto-OBJETO* faster than *obáeto-OBJETO*). These findings may be difficult to reconcile with the idea that the vowels *á* and *a* could be considered variants of the same abstract representation while accent mark information is lost. Two potential accounts manage to explain this pattern, where one account is purely perceptual, and the other is linguistic.

The perceptual explanation would claim that changing a letter from *á* to *a* is not the same as changing a letter from *a* to *á*. Perea et al. (2021) suggested that the word recognition system may quickly encode a letter's accent marks (e.g., the letter *é* in the prime *féliz*). The underlying idea is that the added mark to the base letter is a salient perceptual element (e.g., F is more similar to E than E is to F; see Tversky, 1977, for a model of perceptual asymmetries). Consequently, the resulting percept from the prime *féliz* would be less perceptually similar to the target word *FELIZ* than the identity prime *feliz*, thus being a less effective prime. Therefore, adding an accent mark to a nonaccented word could have a disruptive effect on lexical access.

The linguistic explanation would claim that, for those languages in which diacritical marks indicate lexical stress (for instance, Spanish, Italian), the accent mark is processed in parallel to abstract letter identity (Cubelli & Beschin, 2005). In this account, the accent mark of *á* would be processed on another route than the identity of its abstract letter representation. The idea is that the accent mark may serve as an orthographic cue that pre-activates words with similar syllabic stress patterns in the orthographic lexicon (Cubelli & Beschin, 2005).[2] Under this framework, when an accent mark is placed on an unstressed syllable in a Spanish word, the processing of *féliz* [ˈfe.liθ] may produce some phonological mismatch with the stored representation of the word *feliz* [fe.ˈliθ], thus slowing down lexical access. In contrast, this cost would be minimal or nonexistent had the accent mark been added to the stressed syllable (for example, *felíz* [fe.ˈliθ]).

Altogether, a delay in lexical access in Spanish words with an additional accent mark could be explained by perceptual or linguistic factors. Critically, it remains to be shown whether this cost is limited to the masked priming paradigm. Keep on mind that priming effects reflect an integration of information between the prime and the target (see Gómez et al., 2021). Clearly, it is fundamental to study whether the same phenomenon occurs when directly testing the impact of adding an accent mark to a word in an unprimed scenario. To get the complete picture of how accent marks are represented in the word recognition system in Spanish, it is necessary to examine the nature of the potential reading cost when adding an accent mark to a nonaccented word (e.g., *cebra* [zebra], pronounced as [ˈθe.bra]). In the present experiments, we selected a set of nonaccented Spanish words. We added an accent mark either on the stressed syllable (e.g., *cébra*; Experiment 1) or the unstressed syllable (e.g., *cebrá*; Experiment 2). We chose to employ a semantic categorization task (see Perea et al., 2021, for the same procedure). We decided to use this task because it requires lexical access while being less influenced by phonology or visual familiarity than other word recognition tasks like naming or lexical decision (see Labusch et al., 2021, for discussion).

We can think of three possible scenarios in the present experiments. The first scenario assumes that adding an accent mark has a general perceptual cost when processing a nonaccented word (Kinoshita et al., 2021; Perea et al., 2021). As this perceptual effect would occur regardless of whether the accent

2 These ideas originated from the insights of a study with individuals with left neglect dyslexia in Italian (Cubelli & Beschin, 2005).

mark is added to the stressed syllable or not, one would expect some reading cost for the words with the added accent mark in both experiments (i.e., *cébra* > *cebra*; *cebrá* > *cebra*). The second scenario follows the assumption that the vowels *é* and *e* activate the same letter units during letter/word recognition in Spanish (Schwab, 2015; see also Marcet et al., 2021b; Perea et al., 2020, 2022, for similar claims). As the addition of an accent mark to a word (e.g., *é*, as in *cébra*, or *á* as in *cebrá*) would not play a role during lexical access, one might expect a negligible reading cost in both experiments (i.e., Experiment 1: *cébra* ≈ *cebra*; Experiment 2: *cebrá* ≈ *cebra*). Finally, the third scenario is that, during printed word recognition in languages where accent marks indicate lexical stress, the word recognition system would process both a word's abstract letter identities and the accent mark information related to the syllabic stress (Cubelli & Beschin, 2005). In this case, a mismatch between the word's stored information (e.g., lexical, phonological, semantic) and the presented visual input would delay lexical access. A mismatch in phonological information would occur in Experiment 2: the accent mark in *cebrá* explicitly indicates that the lexical stress falls on the final syllable (i.e., [θeˈbra]), but the accent in the word *cebra* falls on the next-to-last syllable (i.e., [ˈθe.bra]). In contrast, no such cost would occur in Experiment 1 because both *cebrá* and *cebra* would share the same position of lexical stress (i.e., [θeˈbra] in both cases). Thus, in this third scenario, one would expect a different pattern of effects in the two experiments (i.e., *cébra* ≈ *cebra* in Experiment 1; *cebrá* > *cebra* in Experiment 2).

Experiment 1 (added accent mark to the stressed syllable)

Methods

Participants We recruited 50 individuals (30 women, 19 men, 1 preferred not to say) with a mean age of 28.8 years (SD=9.1, age range=19–50 years) via the online platform Prolific Academia (http://prolific.ac). With this sample size, we obtained 2400 observations in the critical comparison (*cebra* vs. *cébra*), which is in line with Brysbaert and Stevens (2018) for small-sized effects. We used Prolific Academia's recruitment filters and a questionnaire to ensure that all participants were native Spanish speakers with normal/corrected vision (46% normal vision, 54% corrected-to-normal vision) and no reading difficulties. The majority of participants completed a university degree (82%), whereas the rest of the participants (18%) completed at least the Spanish equivalent of high school. It is important to note here that recent research has shown that, in word recognition experiments, the findings from Prolific participants produce the same pattern of effects as those conducted in the lab with undergraduate students (e.g., see Angele et al., 2022). The experiments were approved by the Research Ethics Committee of the Universitat de València, following the Helsinki convention. Prior to the beginning of the experiments, all participants signed a written consent form. They received a small monetary compensation for their participation.

Materials We selected 96 non-diacritical nouns in Spanish from the subtitled-based EsPal database (Duchon et al., 2013). The words corresponded either to non-animal names or animal names. For the set of 64 non-animal names, the mean Zipf frequency was 3.48 (range 2.72–3.96), the mean number of letters was 6.2 (range 4–8), and the mean OLD20 was 1.76 (range 1–2.85). None of these words were associated with animal names (e.g., the word *granja* [farm] would not have been selected). For the set of 32 animal names, the mean Zipf frequency was 3.42 (range 2.24–3.99), the mean number of letters was 6.1 (range 4–8), and the mean OLD20 was 1.91 (range 1–3.8). The values of these three variables (Zipf frequency, Number of letters, OLD20) were comparable in the two sets of words (all $ps > 0.13$). In addition, the proportion of words with stress in the last-but-one and final syllables was similar for the sets of animals and non-animal words. In the experiment, each word could be presented as-is (i.e., without diacritics; e.g., the animal name *cebra* [zebra], the non-animal name *sidra* [cider]) or with an added accent mark on the stressed vowel (e.g., *cébra*, *sídra*). To counterbalance the materials across the critical manipulation, we created two lists following a Latin square design: if List 1 included *cebra* (without diacritics), List 2 would include *cébra* (with the added diacritic). Each list was composed of 96 words (48 without diacritics, 48 with an added diacritic). The list of stimuli is presented in Appendix B.

Procedure The experiment was conducted online using a script written in PsychoPy 3 (Peirce & MacAskill, 2018) on the Pavlovia server (www.pavlovia.org). Our demographic data were obtained from a questionnaire in the platform Limesurvey (www.limesurvey.org). At the beginning, all participants filled out a form with demographic information (age, gender, level of education). Afterwards, they were redirected to the experiment. Participants were presented with a word that could be an animal name or not on each trial. When the presented word referred to an animal, participants had to press the "m" button on their keyboard, whereas the word did not refer to an animal, they had to press the "z" button on their keyboard (i.e., a semantic categorization task). Participants were instructed to take this decision independently of the possible extra accent mark on the word: Both *cébra* and *cebra* were to be classified as "animal". Further, they were instructed to respond as fast and accurately as possible. All participants went through 16 practice trials to get familiarized with the task. Before the presentation of each stimulus, a fixation cross was presented for 500 ms in the center of the screen. Then, the stimulus word was presented in the center of the screen until the participant responded—note that there was a timeout after 2000 ms. Participants were instructed to complete the task in a quiet room without distractions. They were allowed to take a break every 60 trials. In total, the experiment took about 6–8 min.

Results and discussion

For the latency analyses, we removed the error responses (3.0% of trials) and the very short response times (less than 250 ms; 0 trials)—note that response times above 2 s (i.e., the deadline for responding) were automatically categorized as errors (8 trials; 0.17% of trials). Two names of animals from the Americas (*chacal* [jackal] and *tapir* [tapir]) were quite unfamiliar to our all European participants (the error rates were 46% and 78%, respectively), and were removed from the analyses—the pattern of findings was the same had we included these two words in the analyses.

To analyze each of the two dependent variables (response times [RTs], accuracy), we employed Bayesian linear mixed-effects models with the *brms* package (Bürkner, 2017) in the R statistical environment (R Core Team, 2021). The distribution of the RTs was modeled with the ex-Gaussian distribution, and the accuracy data was modeled with the Bernoulli distribution—note that the accuracy data involved "1" (correct) and "0" (incorrect) responses (see Marcet et al., 2021a, 2021b; Perea et al., 2022, for the same procedure). The fixed factors were Format (standard vs. added accent mark, encoded as – 0.5 and 0.5, respectively) and Word type (animal vs. non-animal, encoded as − 0.5 and 0.5) and the random-factor structure was the maximal allowed by the design (Barr et al., 2013). Thus, the models were:

$$\begin{aligned}&\text{Dependent variable}\\&= \text{format} * \text{wordtype}\\&+ (1 + \text{format} * \text{wordtype} \mid \text{subject})\\&\quad + (1 + \text{format} \mid \text{item})\end{aligned}$$

We conducted 5000 iterations in both the latency and accuracy models—the first 1000 iterations were for warm-up. The models produced good fits, and the value of the $\hat{R}$ statistic was 1.00 for all estimates. The output of Bayesian linear mixed-effects models provides a value of each estimate, together with its standard error and a 95% credible interval (95% CrI) rather than a frequentist *p*-value. We interpreted positive evidence of an effect when its corresponding 95% CrI did not cross zero.

Response Times The analyses of the latency data showed faster responses to the non-animal than to the animal words, b=− 13.52, SE=5.69, 95% CrI (− 24.73, − 2.36). We did not find evidence of an effect of Format, b=1.58, SE=4.19, 95% CrI (− 6.68, 9.87), or an interaction between the two factors, b=5.03, SE=5.10, 95% CrI (− 5.01, 14.85).

Accuracy The analyses on the accuracy data showed higher accuracy to the non-animal than to the animal words, b=1.20, SE=0.49, 95% CrI (0.28, 2.22). We found no evidence of a main effect of format, b=-− 0.25, SE=0.32, 95% CrI (− 0.90, 0.38), or an interaction between format and type of word, b=0.73, SE=0.53, 95% CrI (− 0.28, 1.81).

In sum, the present experiment has revealed that adding an accent mark on the stressed syllable in Spanish (e.g., *cébra* for the word *cebra*; both pronounced [ˈθe.bra]) does not produce a reliable reading cost on lexical access (i.e., *cebra* ≈ *cébra*; see Table 1). One might argue that there was a small cost

Table 1 Mean response times (in ms) and error rates (in percentages) for non-animal and animal names with the standard format or the added accent on the stressed syllable (Experiment 1)

	Standard format		Accent added	
	Response time	% Errors	Response time	% Errors
Non-animal	636	2.6	644	1.8
Animal	632	4.3	632	5.1

of 8 ms of the words with the added diacritic for the non-animal words, but this difference must be taken with caution because the accuracy data showed an opposite 0.8 trend. Thus, we prefer to interpret this data pattern as reflecting a negligible reading cost from the added accent mark on the stressed syllable.

The issue now is whether there is a more robust reading cost when the added accent mark on a word occurs in a non-stressed syllable (e.g., *cebra* vs. *cebrá*). This was the goal of Experiment 2.

Experiment 2 (added accent mark to a non-stressed syllable)

Method

Participants We selected a new sample of 50 individuals (17 women, 31 men, 2 preferred not to say) with a mean age of 31.32 years (SD=9.9, age range=19–58 years) via Prolific Academia. Participants had to fulfill the same requirements as in Experiment 1, that is, we ensured that all participants were native Spanish speakers with normal/corrected-to-normal vision (50% normal vision, 50% corrected-to-normal vision) and no reading difficulties. The vast majority of participants 84% completed university education, whereas the rest (16%) completed at least the Spanish equivalent of high school. All participants also signed a consent form.

Materials We used the same set of words as in Experiment 1, except that we placed the accent mark on an unstressed syllable of the words instead of the stressed syllable. Thus, the animal name *cebra* [zebra] could be presented as *cebrá,* and the non-animal word *sidra* [cider] could be presented as *sidrá*.[3]

Procedure It was the same as in Experiment 1.

Results and discussion

The statistical analyses were the same as in Experiment 1. The percentage of error responses was 3.9%, the percentage of response times faster than 250 ms was 0.0% (0 trials), and the percentage of trials that reached the deadline for responding was 0.35% (17 trials). As in Experiment 1, we excluded the words *chacal* with an error rate of 58% and *tapir* with 90%—again, when including these error-prone words, the cost of adding accent marks for animal words produced the same findings as in the reported analyses (indeed, the cost for animal words was numerically the same, 7 ms). Table 2 displays the mean response time (in milliseconds) and error rate (in percentage) in each condition.

Response times As in Experiment 1, response times were faster for non-animal than animal words, $b=-14.81$, $SE=6.65$, 95% CrI (-27.68, -1.47). Critically, responses were longer when the word had an added mark on an unstressed syllable than when presented in its standard (non-accented) format, $b=8.86$, $SE=4.00$, 95% CrI (0.99, 16.76). There were no signs of an interaction between these two factors, $b=-2.30$, $SE=4.96$, 95% CrI (-12.01, 7.34).

Accuracy The analyses on the accuracy data showed higher accuracy to the non-animal than to the animal words, $b=1.19$, $SE=0.39$, 95% CrI (0.45, 1.96). In addition, there were no trends on an effect of format, $b=-- 0.16$, $SE=0.27$, 95% CrI (-0.69, 0.36), or an interaction between format and type of word, $b=0.17$, $SE=0.38$, 95% CrI (-0.58, 0.92).

The present experiment revealed a small but reliable cost of adding an accent mark on an

[3] Four non-animal words (*asfalto* [asphalt], *albergue* [hostel], *copo* [flake], *taladro* [drill]) were replaced because they formed past verbal forms when an accent mark was added to the last syllable. The replacement words were *usuario* [user], *huerta* [orchard], *esquema* [scheme], and *cascada* [waterfall].

Table 2 Mean response times (in ms) and error rates (in percentages) for non-animal and animal names with the standard format or the added accent on an ustressed syllable (Experiment 2)

	Standard format		Accent Added	
	Response time	% Errors	Response time	% Errors
Non-Animal	611	2.6	620	2.9
Animal	615	5.7	622	6.7

unstressed syllable (i.e., *cebrá* produced longer response times than *cebra*).[4] As discussed below, this finding offers empirical support to Cubelli and Beschin's (2005) claim that letter identity and accent mark information are processed together in languages where accent marks indicate lexical stress.

As suggested by a Reviewer, it may be of interest whether this cost is modulated by word-frequency. To that end, we computed the Pearson correlation, on an item-by-item basis, between the cost of adding an accent mark and the Zipf frequency of each item from the EsPal database (Duchon et al., 2013). The value of the Pearson correlation coefficient was small, $r=0.122$, $p=0.241$. Although the value of this index in Spanish must be taken with caution (i.e., the overall cost was quite small), we believe that, in future research, it will be important to run parallel analyses in languages where diacritics modify vowel quality.

General discussion

An ongoing question in the field of word recognition and reading is how letters with and without accent marks are represented in the word recognition system for languages where accent marks only indicate lexical stress, such as Spanish (e.g., Perea et al., 2020, 2022). Previous research investigating the omission of accent marks in accented Spanish words (e.g., *carcel* for the word *cárcel*) has found little or no cost on lexical access in various paradigms. This pattern has been taken to suggest that accented and nonaccented vowels (e.g., *á* and *a*) share the letter representations during word processing. However, when an accent mark is added to an otherwise nonaccented Spanish word, there seems to be a processing cost in masked priming paradigms (e.g., Domínguez and Cuetos, 2018; Perea et al., 2020, 2022). Here, we further investigated the cost of processing when adding an accent mark to an otherwise nonaccented Spanish word in a semantic categorization task. Critically, we placed the accent marks either on the stressed (*cébra*) or on the unstressed syllable (*cebrá*) of the otherwise nonaccented word (*cebra*). We found out that: 1) when an accent mark was added to the stressed syllable of a nonaccented Spanish word, there was only a minimal difference between the accented and nonaccented word (e.g., *cébra* ≈ *cebra;* Experiment 1); 2) when the accent mark was added to the unstressed syllable, the cost relative to the intact words was reliable, but it was of a small magnitude, around 8.5 ms (e.g., *cebrá* > *cebra;* Experiment 2).

The reading cost when adding an accent mark on the unstressed syllable of a nonaccented Spanish word (Experiment 2) generalizes previous findings with the masked priming task. As noted in the Introduction, in masked priming experiments, there is a cost of processing when a nonaccented target word like *FELIZ* ([fe.ˈliθ], happy in Spanish) is preceded by the diacritical prime *féliz* ([ˈfe.liθ]) relative to the identity prime *feliz* (e.g., Perea et al., 2020; see also Domínguez & Cuetos, 2018). Interestingly, when adding an accent mark on the stressed syllable of an unaccented word (Experiment 1), we found a more negligible difference in the same direction.

Taken together, our findings provide empirical support to the idea that when processing words in languages like Spanish—where accent marks only indicate the stressed syllable—accent mark information are processed in parallel to abstract letter information (Cubelli & Beschin, 2005). In line with previous observations by Cubelli and Beschin (2005), we propose that accent marks and letter identity are processed in parallel. The accent marks would serve as orthographic cues to activate words with similar properties (see Peressotti et al., 2003, for an analogous claim concerning the initial capitalization for proper nouns). In the case of Spanish words, the orthographic cue of the accent mark activates words with similar lexical stress properties. When placing

[4] As suggested by a Reviewer, we conducted a parallel analysis comparing the intact words and the words with an extra accent mark using a frequentist approach. Results corroborated the findings from the Bayesian linear mixed effects, $t(49)=2.014$, $p < 0.05$.

an accent mark on the stressed syllable of a word (e.g., *cébra*), words with a similar lexical stress get activated (in this case, *cebra*), thus producing only a minimal cost relative to the intact word. In turn, when placing an accent mark on an unstressed syllable, as in *cebrá*, the word recognition system would detect some phonological mismatch between the presented item and the stored representation in the mental lexicon, thus producing a delay in lexical access.

How can we reconcile the above-discussed account, in which accent mark information and abstract letter identities are processed in parallel, with previous evidence showing that the omission of accent marks does not delay lexical access? We believe that the differences may be more apparent than real. When looking closely at previous findings with omitted accent marks in Spanish, they hint towards the same direction as the present experiments. For instance, Perea et al. (2022) found a nonsignificant 7-ms difference when comparing *cárcel* (prison) and *carcel* in a semantic categorization task. A similar numerical trend (namely, 9 ms) occurred in the lexical decision experiment conducted by Marcet et al. (2021b) and in the first-pass fixation durations of the sentence reading experiment conducted by Marcet and Perea (2022) (e.g., 4 ms in the first-fixation durations on the target word). Thus, all these experiments suggest a small—but most likely real in an eventual meta-analysis—reading cost when omitting or adding an accent mark in a Spanish word. This cost may be statistically more robust when the phonological information from the word (e.g., as in *cebrá*) mismatches that of the stored representations (*cebra*), as occurred in Experiment 2. Thus, all these findings strongly suggest that accent marks are not entirely left out during lexical access. Nonetheless, it must be stressed that, while accent mark information can be processed in parallel to the letter identity, it only produces a minimal impact in word recognition in Spanish.

The current findings have implications for future computational models of visual word recognition in those languages, like Spanish or Italian, where accent marks only provide information on lexical stress. Previous research has suggested that vowels with and without accent marks share the letter representations in Spanish. This conclusion should be slightly qualified when considering together the experiments when adding or omitting an accent mark in Spanish. As first suggested by Cubelli and Beschin (2005), accent mark information may not be entirely lost but instead processed in parallel to abstract letter information. While a fully comprehensive computational model of visual word recognition in Spanish does not require implementing the letters *a* and *á* as different letter units, the model should also encode accent mark information in the mapping onto higher levels of processing. Nonetheless, the reading cost when adding or omitting an accent mark in Spanish is very small (i.e., less than 10 ms across experiments). Thus, researchers can safely run the simulations of word recognition experiments in Spanish with an English-based orthography (e.g., easyNet platform, Adelman et al., 2018) by omitting the accent marks (e.g., see Conrad et al., 2010).

Altogether, the present experiments have shown only a minor but reliable reading cost when adding an accent mark to an otherwise nonaccented Spanish word, especially when the accent mark is put on an unstressed syllable. The small but consistent reading cost of adding or deleting an accent mark across experiments in Spanish suggests that adult skilled readers can use accent mark information in the course of visual word processing. Future research using techniques such as event-related potentials (see Grainger & Holcomb, 2009) is necessary to unveil the time-course of diacritical letters in Spanish and other languages (see Marcet et al., 2021a, for cross-language differences when processing accented vowels).

Another implication of the current finding relates to reform in the orthography of Spanish: Adding an accent mark in the incorrect vowel produces a small but reliable cost; adding an accent mark that is not prescribed in the stressed vowel yields a minimal cost; furthermore, this reading cost seems to be even less reliable in cases where a prescribed accent mark is removed. This begs the question, why should Spanish preserve the accent marks? They tend to serve more as a marker of education and social class (as the public and somewhat pedantic debates on social media about the abolition of some grammatical accent marks demonstrate) than a tool to facilitate word recognition. Indeed, in an influential speech in Zacatecas (Mexico) in 1997, Nobel-prize winner García-Márquez asserted that we had to reduce the “martial laws” of the accentuation rules in Spanish. Indeed, Italian, a language germane to Spanish, has much simpler accentuation rules (see Colombo & Sulpicio, 2021). So, we conclude this article with a suggestion that will upset other famous writers but will be cheered by millions of second-grade children in Spain and Latin America: little will be lost if we retire the *á*ccent m*á*rks in Spanish.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study was supported by Grant PID2020-116740 GB-I00 funded by MCIN/AEI/10.13039/501100011033.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix A: Accentuation rules in Spanish

The current general accentuation rules in Spanish (Real Academia Española, 2010) are the following:

(1) for those words in which lexical stress falls on the last syllable, an accent mark is only required when the ending letter is -n or -s (e.g., *ratón* [ra.ˈton], but not *andar* [an.ˈdar]);

(2) for those words in which lexical stress falls on the last-but-one syllable, an accent mark is required only when the final letter is a vowel other than -n or -s (e.g., *cárcel* [ˈkar.θel], but not *comes* [ˈko.mes]);

(3) when lexical stress falls on the second-to-last syllable or earlier, an accent mark is always required.

There are, however, numerous exceptions to these general rules. For instance, the adverb *lentamente* [ˈlen.ta.men.te] (slowly), which is a composed of the stem *lenta* and the suffix *-mente*, does not have an accent mark. Similarly, *bíceps* (biceps) has an accent mark despite an ending in -s and having its lexical stress on the last-but-one syllable. There are also relatively complex rules concerning diphthongs and hiatuses (e.g., *distraído* [distracted] is accented but *imbuido* [imbued] is not accented).

Overall, the number of pages of the accentuation rules in Spanish in the last edition of the dictionary is above 5.

Appendix B: Materials in the experiments

(*animal names*) cisne, koala, cuervo, paloma, lombriz, mosca, cangrejo, jirafa, gorila, hormiga, gacela, panda, avispa, avestruz, lagarto, caracol, medusa, tapir, perdiz, pulga, abeja, chacal, oveja, ballena, ganso, lechuza, cebra, codorniz, calamar, rana, ardilla, trucha; (*nonanimal names*) contexto, textura, pedal, cajero, faro, cuchara, cargador, carril, mantel, turismo, eficaz, nudo, capucha, sidra, secador, trastero, ruina, tramo, pincel, bazar, talco, fachada, igualdad, pipa, burbuja, taberna, bombilla, chalet, asfalto, albergue, jarra, alambre, probador, neblina, mechero, zumo, taladro, poncho, vinagre, cemento, lino, ceniza, vacuna, vereda, barril, carpa, variedad, tecla, impuesto, pintor, corcho, suceso, alergia, ocio, trofeo, trapo, blusa, sandalia, copo, espuma, novedad, muralla, escoba, rapidez.

References

Adelman, J. S., Gubian, M., & Davis, C. J. (2018). *easyNet: A computational modeling software package for cognitive science, bridging the gap between novices and experts* [Computer software]. http://adelmanlab.org/easyNet/

Angele, B., Baciero, A., Gómez, P., & Perea, M. (2022). Does online masked priming pass the test? The effects of prime exposure duration on masked identity priming. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01742-y

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review, 105*(4), 678–723. https://doi.org/10.1037/0033-295x.105.4.678-723

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*. https://doi.org/10.5334/joc.10

Bürkner, P.-C. (2017). BRMS: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Chetail, F. (2020). Are graphemic effects real in skilled visual word recognition? *Journal of Memory and Language, 111*, 104085. https://doi.org/10.1016/j.jml.2019.104085

Chetail, F., & Boursain, E. (2019). Shared or separated representations for letters with diacritics? *Psychonomic Bulletin & Review, 26*, 347–352. https://doi.org/10.3758/s13423-018-1503-0

Colombo, L., & Sulpizio, S. (2021). The role of orthographic cues to stress in Italian visual word recognition. *Quarterly Journal of Experimental Psychology, 74*(9), 1631–1641. https://doi.org/10.1177/17470218211006062

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. M. (2010). Simulating syllable frequency effects within an interactive activation framework. *European Journal of Cognitive Psychology, 22*(5), 861–893. https://doi.org/10.1080/09541440903356777

Cubelli, R., & Beschin, N. (2005). The processing of the right-sided accent mark in left neglect dyslexia. *Brain and Language, 95*(2), 319–326. https://doi.org/10.1016/j.bandl.2005.02.002

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

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences, 9*, 335–341. https://doi.org/10.1016/j.tics.2005.05.004

Domínguez, A., & Cuetos, F. (2018). The contrastive value of lexical stress in visual word recognition: Evidence from Spanish. *Psicothema, 30*, 276–282. https://doi.org/10.7334/psicothema2018.30

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for spanish word properties. *Behavior Research Methods, 45*, 1246–1258. https://doi.org/10.3758/s13428-013-0326-1

Gómez, P., Marcet, A., & Perea, M. (2021). Are better young readers more likely to confuse their mother with their mohter? *Quarterly Journal of Experimental Psychology, 74*(9), 1542–1552. https://doi.org/10.1177/17470218211012960

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistic Compass, 3*, 128–156. https://doi.org/10.1111/j.1749-818X.2008.00121.x

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in Cognitive Sciences, 12*, 381–387. https://doi.org/10.1016/j.tics.2008.06.006

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition, 91*(3), 273–296. https://doi.org/10.1016/j.cognition.2003.09.006

Kinoshita, S., Yu, L., Verdonschot, R. G., & Norris, D. (2021). Letter identity and visual similarity in the processing of diacritic letters. *Memory & Cognition, 49*(4), 815–825. https://doi.org/10.3758/s13421-020-01125-2

Labusch, M., Kotz, S. A., & Perea, M. (2021). The impact of capitalized German words on lexical access. *Psychological Research Psychologische Forschung*. https://doi.org/10.1007/s00426-021-01540-3

Marcet, A., Fernández-López, M., Baciero, A., Sesé, A., & Perea, M. (2021a). What are the letters e and é in a language with vowel reduction? The case of Catalan. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716421000497

Marcet, A., Fernández-López, M., Labusch, M., & Perea, M. (2021b). The omission of accent marks does not hinder word recognition: Evidence from Spanish. *Frontiers in Psychology, 12*, 794923. https://doi.org/10.3389/fpsyg.2021.794923

Marcet, A., & Perea, M. (2022). Does omitting the accent mark in a word affect sentence reading? Evidence from Spanish. *Quarterly Journal of Experimental Psychology, 75*, 148–155. https://doi.org/10.1177/17470218211044694

Peirce, J. W., & MacAskill, M. R. (2018). *Building experiments in PsychoPy*. Sage.

Perea, M., Baciero, A., & Marcet, A. (2021). Does a mark make a difference? Visual similarity effects with accented vowels. *Psychological Research Psychologische Forschung, 85*, 2279–2290. https://doi.org/10.1007/s00426-020-01405-1

Perea, M., Fernández-López, M., & Marcet, A. (2020). What is the letter é? *Scientific Studies of Reading, 24*, 434–443. https://doi.org/10.1080/10888438.2019.1689570

Perea, M., Labusch, M., & Marcet, A. (2022). How are words with diacritical vowels represented in the mental lexicon? Evidence from Spanish and German. *Language, Cognition, and Neuroscience.* https://doi.org/10.1080/23273798.2021.1985536

Peressotti, F., Cubelli, R., & Job, R. (2003). On recognizing proper names: The orthographic cue hypothesis. *Cognitive Psychology, 47*(1), 87–116. https://doi.org/10.1016/s0010-0285(03)00004-5

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Real Academia Española. (2010). *Ortografía de la lengua española* [Orthography of Spanish Language]. Espasa.

Schwab, S. (2015). Accent mark and visual word recognition in Spanish. *Loquens, 2*(1), e018. https://doi.org/10.3989/loquens.2015.018

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*, 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352. https://doi.org/10.1037/0033-295x.84.4.327