Psychonomic Bulletin & Review (2023) 30:1530–1538
https://doi.org/10.3758/s13423-022-02244-4

BRIEF REPORT

Just a mark: Diacritic function does not play a role in the early stages of visual word recognition

András Benyhe[1] · Melanie Labusch[2,3] · Manuel Perea[2,3]

Accepted: 29 December 2022 / Published online: 12 January 2023

Abstract

A very common feature in most writing systems is the presence of diacritics: distinguishing marks that are added for various linguistic reasons. Most models of reading, however, have not yet captured the nature of these marks. Recent priming experiments in several languages have attempted to resolve how diacritical letters are represented in the visual word recognition system. Since the function and appearance of diacritics can change from one language to the other, it is hard to interpret the accumulated evidence. With this in mind, we conducted two masked priming lexical decision experiments in Hungarian, a transparent orthography with a wide use of diacritic vowels that allows for clear-cut manipulations. In the two experiments, we manipulated the presence or absence of the same diacritic (i.e., the acute accent) on two specific sets of letters that behave differently. In Experiment 1, the manipulation changed only the length of vowels, whereas in Experiment 2, it also changed the quality (e.g., a⤳/ɒ/ vs. á⤳/aː/). In both experiments, we found that primes with an omitted diacritic work just as good as the identity primes (nema→NÉMA = néma→NÉMA [mute]), whereas the addition of a diacritic comes with a cost (mése→MESE > mese→MESE [tale]). This asymmetry favors a purely perceptual account of the very early stages of word recognition, making it blind to the function of diacritics. We suggest that the linguistic functions of diacritics originate at later processing stages.

Keywords Masked priming · Orthographic processing · Lexical decision · Visual word recognition

Introduction

Through the centuries, writing has evolved to be a remarkably useful invention for communicating highly complex information in a reproducible way (Pinheiro et al., 2020). To achieve this goal, writing systems must embody the variability of the language they represent. Although most European languages are written with the Latin alphabet, many have added small markings to some letters (i.e., diacritics) to reflect the features of the language (for review, see Protopapas & Gerakaki, 2009). Importantly, the role of diacritics can change from one language to the other. For vowels, they might be present due to purely etymological reasons, like in certain French words (hôte⤳/ot/); they could represent the location of lexical stress, as in Spanish (anfitrión⤳/anfiˈtɾjon/); in yet other cases, diacritics can change the phonetic value of the affected letters, like in Hungarian (házigazda⤳/ˈhaːzigɒzdɒ/). Notably, all the above examples have vowels with diacritics sharing the same meaning (host), but the type of information that their diacritics indicate is completely different.

Importantly, an open question is whether the representation of vowel diacritics depends on their linguistic function. In the present study, we investigated the effects of omitting and adding diacritics using a masked priming paradigm (Forster & Davis, 1984) in Hungarian. Thanks to its shallow orthography (Borgwaldt et al., 2005), vowel diacritics in Hungarian have a well-defined role in signaling phonetic values that may have a contrastive role: vowel length for some vowels (e.g., kor⤳/kor/ [age] vs. kór⤳/koːr/ [disease]), and vowel length and quality for other vowels (kar⤳/kɒr/ [arm] vs. kár⤳/kaːr/ [pity]). Therefore, Hungarian provides an excellent scenario to test whether the function of diacritics affects their processing.

It is generally believed that during the processing of written stimuli, successive populations of detector nodes process

✉ András Benyhe
andras.benyhe@szte.hu

1 Department of Physiology, Albert Szent-Györgyi Medical School, University of Szeged, Dóm tér 10, Szeged 6720, Hungary

2 Universidad Antonio de Nebrija, Madrid, Spain

3 Universitat de València, Valencia, Spain

the increasingly complex and abstract information chunks (Dehaene et al., 2005; Grainger et al., 2008). From the detected simple features (e.g., ⌜, ⊦, and ⊃), the left ventral visual stream compiles letters (e.g., P) and interprets them in the context of their neighborhood, pointing at words. Note that, having been proposed for the English orthography, these models remain agnostic as to the encoding of diacritics. This network is understood to be invariant to features of the script such as size, CASE, *font*, etc. However, in the first stages of processing, the visual word form detectors can be fooled when we replace letters with similar-looking characters (e.g., obiect, cornputer), and this property can be exploited with the masked priming technique (Forster & Davis, 1984). If a replaced letter is sufficiently similar to the original one, modified stimuli can successfully facilitate target processing in behavioral experiments (i.e., letter similarity effects; e.g., object→OBJECT ≈ obiect→OBJECT < obaect→OBJECT; Marcet & Perea, 2017, 2018).

Recent masked priming research on diacritics has shown an intriguing pattern of letter similarity effects. Masked priming of non-diacritical targets that were primed with an added diacritic was found to hinder letter detection as much as priming with an unrelated letter, compared to the identity condition (a→A < â→A = z→A; Chetail & Boursain, 2019). This pattern was then replicated with the same conditions embedded into French words for lexical decision (taper→TAPER < tâper→TAPER = tuper→TAPER [to type][1]). Chetail and Boursain interpreted these findings as evidence for separate representation of vowels with and without diacritics in French (see Ans et al., 1998, for a model of word recognition in French making this assumption).

An analogous effect was detectable using Spanish letters and words (e.g., a→A < á→Á < é→A; feliz→FELIZ < féliz→FELIZ ≤ fáliz→FELIZ [happy]; Perea et al., 2020). Importantly, Perea et al. also tested the effects in the other direction, showing that priming diacritical targets with their non-diacritical versions results in a priming effect comparable to the identity condition in both alphabetical and lexical decision tasks (á→Á ≈ a→Á < e→Á; fácil→FÁCIL ≈ facil→FÁCIL < fecil→FÁCIL [easy]). Since the main function of diacritics on vowels in Spanish is the indication of lexical stress (i.e., no phonetic distinction), the abstract representations of diacritical and base letters would be shared. The observed asymmetry in priming effects (a⇒Á, á⇏A) points at some divergence, though.

Notably, a Finnish study by Perea, Hyönä, and Marcet (2022b) did not find any cost of primes with omitted diacritics (pöytä→PÖYTÄ ≈ poytä→PÖYTÄ < paytä→PÖYTÄ [table], in both lexical decision and word naming), mimicking the Spanish results. This outcome may appear surprising considering that Finnish diacritics have a distinctive phonological function: umlauts distinguish between front and back vowels (e.g., o⇝/o/ vs. ö⇝/ø/) that can never appear together in the same word, a feature known as *vowel harmony*. In their experiments, the primes with omitted diacritics always produced disharmonious words and yet managed to enhance target processing. However, Perea et al. did not examine the effects of diacritical primes on non-diacritical target words (e.g., päita→PAITA vs. paita→PAITA [shirt]).

The above language-independent findings of diacritic perception are consistent with recent computational models of letter and word recognition (see Norris, 2013, for review). These models generally consider how that the addition of information differs from the absence of information. Extra features (e.g., diacritics) could be incompatible with the original identity of a letter and interfere with letter detectors. Inversely, a missing feature does not carry such information, thus such partial stimuli would still be compatible with the original letter. An elegant analogy of this hypothesis is presented in the noisy-channel model (Norris & Kinoshita, 2012). This model states that our minds are prepared to fill out missing details lost to the uncertainty of perception, as if we were picking up a weak signal from a noisy channel. Hence, in such model, features could easily be missed in the very early stages of processing, but once perceived, they exhibit strong evidence against their absence. Indeed, this asymmetry can be observed in the priming patterns with visually similar letters (F⇒E vs. E⇏F; A⇒Á vs. Á⇏A; Kinoshita et al., 2021). Importantly, this account can also capture the asymmetric similarities reported with diacritical consonants (e.g., n/ñ in Spanish; Marcet et al., 2020) and diacritical katakana characters (カ/ガ; Kinoshita et al., 2021). It should be added here that the noisy-channel model does not make any specific claims about the underlying neural mechanisms. Although feature-letter inhibition would seem to be an apt explanation for these asymmetries, Rey et al. (2009) reported electrophysiological evidence against the existence of such connections in letter perception. Rather, it is likely that lateral inhibition between the activated letter detectors is the main mode of competition among visually similar letters.

Of note, this asymmetry based on the presence versus absence of features has also been found outside the realm of visual-word recognition, namely in visual search tasks (for reviews, see Treisman & Gormican, 1988; Wolfe, 2001). In these cases, targets with certain defining features are found more quickly and efficiently from a group of distractors without that feature than the other way round. These search asymmetries are understood to arise from early perceptual

[1] These findings also occurred for diacritical primes that altered the phonology (e.g., païre → PAIRE [pair]; /pa'iʁ/ and /pɛʁ/ respectively).

representations of stimuli, thus suggesting that the above-described effects with diacritics could also have a strong perceptual component.

However, there is evidence showing that, in the case of diacritical vowels, masked priming effects can be modulated by the linguistic function. In English, a writing system without diacritics, there is a small but detectable cost of diacritical primes relative to identity primes (7 ms; nórth→NORTH) (Perea, Gomez, & Baciero, 2022a). In the Spanish study by Perea et al. (2020), priming with an added diacritic introduced a more sizable cost (17 ms; féliz→FELIZ), and Chetail and Boursain observed a much higher effect in French (50 ms; néveu→NEVEU). Although the presence of a disadvantage in the diacritical condition in these experiments can be explained by computational models of visual-word recognition (see Norris & Kinoshita, 2012), a linguistic approach is still required to predict the different sizes of these effects. Furthermore, in a masked priming alphabetic decision task with diacritical and non-diacritical letters in Catalan, Marcet et al. (2022) found a pattern inconsistent with the perceptual account (i.e., e→E ≈ é→E; é→É < e→É) – note that grapheme-phoneme mapping for non-diacritical vowels in Catalan is rather complex (e.g., e⇝/ɛ/, /e/, or /ə/; whereas é⇝/e/). These findings suggest that in the first stages of processing, at least under some circumstances, the linguistic value of diacritical vowels may play a greater role than perceptual similarity alone.

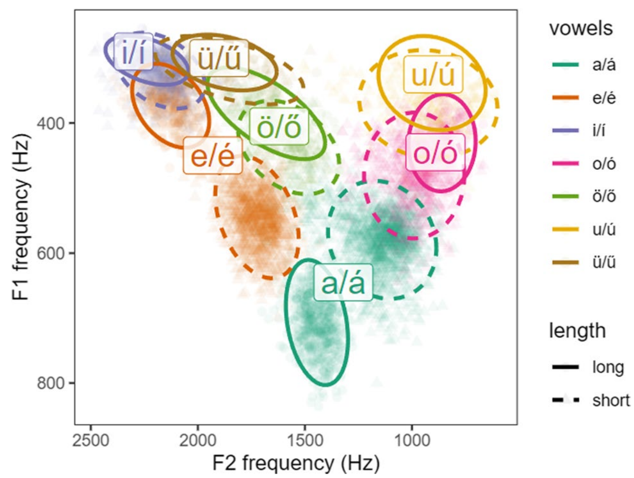

Fig. 1 Formant frequency distributions of Hungarian short-long vowel pairs, with ellipses drawn at the 95% level. Data are redrawn from the Hungarian formant database of Abari et al. (2011), with permission. F1 and F2 designate the first and second formants, respectively. In Hungarian, vowels have a one-to-one grapheme-phoneme mapping, with letters corresponding to the following IPA symbols: a /ɒ/, á /aː/, e /ɛ/, é /eː/, i /i/, í /iː/, o /o/, ó /oː/, ö /ø/, ő /øː/, u /u/, ú /uː/, ü /y/, ű /yː/

Table 1 Stimulus word examples from Experiments 1 and 2. D+ stands for targets with diacritics and D- for targets without diacritics

		Prime			Target
		Identity	Similar	Dissimilar	
Exp.1 (o,u)	D-	mozi /mozi/	mózi /moːzi/	múzi /muːzi/	MOZI [cinema]
		kapu /kɒpu/	kapú /kɒpuː/	kapá /kɒpaː/	KAPU [gate]
	D+	róka /roːkɒ/	roka /rokɒ/	reka /rɛkɒ/	RÓKA [fox]
		túra /tuːrɒ/	tura /turɒ/	tera /tɛrɒ/	TÚRA [hike]
Exp.2 (a,e)	D-	alak /ɒlɒk/	alák /ɒlaːk/	alók /ɒloːk/	ALAK [shape]
		mese /mɛʃɛ/	mése /meːʃɛ/	máse /maːʃɛ/	MESE [tale]
	D+	néma /neːmɒ/	nema /nɛmɒ/	nama /nɒmɒ/	NÉMA [mute]
		házi /haːzi/	hazi /hɒzi/	huzi /huzi/	HÁZI [home]

To resolve the role of linguistic versus perceptual factors in the early encoding of diacritical vowels during word recognition, we chose a language, Hungarian, with a dual function of diacritics: vowel length and vowel quality. The Hungarian writing system differentiates 14 vowels, nine of which are diacritical versions of the five Latin base vowels[2] (Fig. 1; see Szende, 1994, for details). Accents, the most abundant of diacritical marks, always increase vowel length (e.g., o⇝/o/, ó⇝/oː/; as in olló⇝/olːoː/ [scissors]), but, for the letters ‘a’ and ‘e’, the accent also changes vowel quality (a⇝/ɒ/ vs. á⇝/aː/; e⇝/ɛ/ vs. é⇝/eː/; e.g., elég⇝/ɛleːg/ [enough]). Thus, the letters ‘ó’ and ‘á’ differ perceptually from their base letters in the same way; however, in the latter case, the diacritic carries more phonetic information (length plus quality). The dual role of diacritics in Hungarian allowed us to design two masked priming experiments testing whether the encoding of diacritics depends on the linguistic function within the same language. Following the design from the Perea et al. (2020) lexical decision experiment, we compared identity primes, visually similar primes (only differing in

[2] It is important to note that Hungarian writing consistently employs diacritics in both lowercase and uppercase text, whereas in Spanish or French, uppercase text often omits the mandatory diacritics.

the presence/absence of the diacritics (acute accents, as in Spanish)), and visually dissimilar controls (i.e., changing the base letter). In Experiment 1, the letter modifications only affected vowel length, whereas in Experiment 2, they affected both length and vowel quality (see Table 1).

We built our hypotheses from two perspectives. Leading computational models of visual-word recognition (e.g., Norris & Kinoshita, 2012) would predict that Experiment 1 should yield asymmetrical priming effects parallel to the Spanish study of Perea et al. (2020). From this account, similar primes should facilitate diacritical targets as much as their identity primes do (róka→RÓKA = roka→RÓKA < reka→RÓKA); similar primes of non-diacritical targets, however, should behave differently, more resembling the dissimilar condition (mozi→MOZI < mózi→MOZI ≤ múzi→MOZI). On the other hand, a linguistic approach makes different predictions. Given that in Hungarian, vowels that differ in length are treated as separate letters (e.g., learning to read, dictionaries, etc.) and have contrastive value (e.g., kor⤳/kor/ [age] vs. kór⤳/koːr/ [disease]), they could be encoded as separate letter units. In this case, we would expect an advantage of identity primes over similar primes for both diacritical and non-diacritical targets (róka→RÓKA < roka→RÓKA < reka→RÓKA; mozi→MOZI < mózi→MOZI < múzi→MOZI).

Experiment 1: Vowel length

Methods

This study was pre-registered on the Open Science Framework (OSF) via the following link: https://osf.io/2kwja. The materials, raw data, scripts, and outputs are available at https://osf.io/5qunx/?view_only=f445a8065cf847b6b202d3a0ade6b8a1.

Participants

We recruited a total of 72 Hungarian native speakers (mean age = 24.72 years, 41 women, 29 men, two rather not say) via the online recruitment platform Prolific Academia (www.prolific.co). Following the pre-registration protocol, the number of total observations with the current sample size would be above 1,800 observations per condition, following the guidelines proposed by Brysbaert and Stevens (2018) for masked priming experiments. All the participants were native Hungarian speakers without any reading and/or writing problems and with normal or corrected-to-normal vision. Before the beginning of the experiment, all participants gave informed consent to participate. The experiment was approved by the Ethics Committee of the University of València and followed the guidelines of the Helsinki convention. Participants were compensated for their participation based on Prolific's average participant salary.

Materials

The target words were 180 words selected from the Hungarian National Corpus (Oravecz et al., 2014). We chose to include four- to six-character-long two-syllable Hungarian lemmas, containing the critical letters 'o' and 'u' for non-diacritical or 'ó' and 'ú' for diacritical target stimuli (see the rationale for choosing these letters in the *Introduction*). Importantly, the other syllable never had diacritics. Words with the critical letter in the first position were excluded, along with words that produced real words when changing their target vowel's diacritical status. After selection, the diacritical and non-diacritical words had mean Zipf-frequencies of 3.96 and 4.49, respectively, and mean target letter positions of 3.5 and 3.1, respectively. Primes were constructed by modifying the critical letters of target words, according to the following rules: (1) identity primes are the same as the target, (2) visually similar primes have a flipped diacritical status at the selected vowel, resulting in a pseudoword, and (3) visually dissimilar primes have a different base letter at the selected position with a flipped diacritical status, also producing a pseudoword (see Table 1 for examples). For the lexical decision task, 180 pseudowords had to be created. To achieve this, a pseudoword counterpart was generated for each target by leaving the vowels in place and filling in new consonants according to relative positional letter frequencies in the Hungarian language (e.g., the target word 'orvos' [doctor] had the pseudoword counterpart 'opros'). Finally, three counterbalanced lists were created in a Latin square manner. Every target was presented only once, paired with a different priming condition in each list. The lists contained 30 items per condition (diacritical status, priming condition, and word status), summing up to a total of 360 trials. The word lists and MATLAB scripts necessary for their generation are available via the OSF and can be accessed at: https://osf.io/5qunx/?view_only=f445a8065cf847b6b202d3a0ade6b8a1.

Procedure

We constructed the masked priming experiment with PsychoPy 3 (Peirce et al., 2022) and presented it online via Psychopy's online servers Pavlovia (www.pavlovia.org) and LimeSurvey (www.limesurvey.org) (see Angele et al., 2022, for a demonstration of the validity of the use of online masked priming with PsychoPy). Participants initially filled out a form in LimeSurvey with demographic information (age, gender, medical conditions, and education level) and were then redirected to the beginning of the experiment to Pavlovia. We instructed participants to conduct the experiment in a quiet place, without any interruptions. We employed the standard instructions for a lexical decision task: Participants had to decide as fast and accurately as possible whether a Hungarian string of letters was a word or not. If the presented string of letters was a word, participants were instructed to press the 'M' key on their keyboard and if it was not a word, they had to press the 'X' key. In a given trial, the participant saw a pattern mask (#####) for 500 ms, then, a brief prime stimulus was presented for 50 ms, which was replaced by the target item for 2,000 ms or until the participant gave a response. Before the beginning of the actual experiment, all participants went through 20 practice trials, where they received feedback for their responses. During the experimental phase, they did not receive feedback. Altogether, the experiment consisted of 360 trials with a break every 60 trials. In general, participants took about 20 min to complete the task.

Data analysis

The statistical analyses followed the pre-registration protocol (https://osf.io/2kwja). Incorrect responses and extremely short response times (RTs; less than 250 ms) were omitted from the latency analyses.

The inferential analyses on the word targets employed Bayesian linear-mixed effects models (brms package Bürkner, 2017) in R (R Core Team, 2021). The fixed factors were diacritical status (accented vs. non-accented; coded as 0.5 and -0.5) and priming condition (identity, similar, dissimilar; the reference was "similar") with maximal random-effect structure:

```
Dependent_Variable = DiacriticalStatus * PrimingCondition +
(1 + DiacriticalStatus * PrimingCondition | subject) + (1 +
DiacriticalStatus * PrimingCondition | item)
```

RTs were modeled with a Gaussian distribution after a -1,000/RT transformation and accuracy data were modeled with a Bernoulli distribution. The number of iterations for each chain was 5,000 (1,000 for warm-up).[3] We employed the default priors from the model. The quality of the fits was good and, in all cases, $\hat{R} < 1.01$. The output of the models offers, for each effect, a coefficient (i.e., the mean of its posterior distribution) together with its standard error and its 95% credible interval (CrI). Following the pre-registered protocol, we interpreted evidence of an effect when its 95% CrI did not cross zero.

[3] In the analyses of the accuracy data, the program indicated some warnings with this set-up, so that the reported analyses were on 10,000 iterations (2,000 as warm-up).

Results and discussion

Table 2 displays the average RTs and accuracy in each condition. The posterior RT distributions of the model fits are shown in Fig. 2.

Table 2 Mean response times in ms and accuracy (rate correct) in Experiment 1

	Identity	Similar	Dissimilar
Word targets			
Non-diacritical	632 (.947)	643 (.952)	656 (.948)
Diacritical	665 (.926)	661 (.923)	681 (.932)
Pseudoword targets			
Non-diacritical	704 (.963)	707 (.954)	712 (.965)
Diacritical	727 (.958)	732 (.957)	734 (.953)

Similar versus identity

RTs were faster in the identity than in the similar condition, b = -0.03, SE = 0.01, 95%CrI (-0.05, -0.01). This effect was modulated by diacritical status (interaction, b = 0.03, SE = 0.01, 95% CrI [0.00, 0.05]): the advantage occurred for accented target words (11 ms faster for mozi→MOZI than mózi→MOZI) but not for non-accented target words (4 ms difference between róka→RÓKA and roka→RÓKA).

Accuracy was higher in the identity than in the similar condition, b = 0.79, SE = 0.20, 95%CrI (0.40, 1.20), and this effect was modulated by diacritical status, b = -0.85, SE = 0.26, 95% CrI (-1.37, -0.34). This interaction reflected higher accuracy in the identity than in the similar condition for the targets without diacritics (mozi→MOZI more accurate than mózi→MOZI) but not for the targets with diacritics (róka→RÓKA = roka→RÓKA).

Similar versus dissimilar

RTs were faster in the similar than in the dissimilar condition, b = 0.03, SE = 0.01, 95%CrI (0.01, 0.05). This advantage did not depend on diacritical status (interaction, b = 0.00, SE = 0.01, 95% CrI [-0.02, 0.03]) (13 ms for mózi→MOZI over múzi→MOZI; 20 ms for roka→RÓKA over reka→RÓKA).

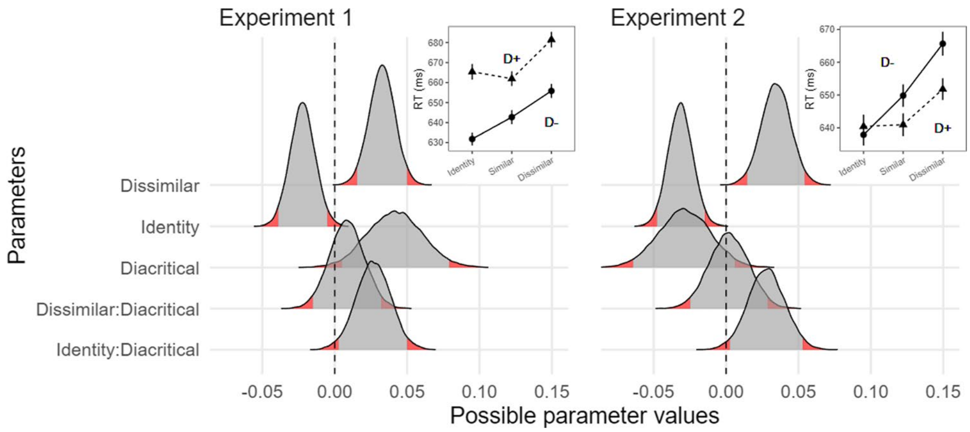

Fig. 2 Posterior distributions of Experiment 1 (left panel) and Experiment 2 (right panel). The gray areas correspond to the 95% credible interval (CrI) for each parameter. Note that the reference level was the Similar priming condition. The insets show the plots of mean response times ± 1 S.E.M. for the two experiments, respectively. D+: diacritical target, D-: non-diacritical target

The analyses of accuracy did not show evidence of a difference between similar and dissimilar conditions, b = -0.05, SE = 0.16, 95% CrI (-0.35, 0.28) or an interaction with diacritical status, b = -0.36, SE = 0.24, 95% CrI (-0.83, 0.11).

The superiority of visually similar primes over dissimilar ones was manifest in the RTs, independently of diacritical status (i.e., a visual similarity effect). Critically, the advantage of the identity condition over the visually similar condition was only present for non-diacritical targets, thus mimicking the pattern in Spanish (Perea et al., 2020). This finding favors a perceptual explanation of masked priming effects with diacritical letters (o⇒Ó but ó⇏O).

Notably, only vowel length was manipulated by the diacritics in Experiment 1. One might argue that the letters 'o'⤳/o/ and 'ó'⤳/oː/ share virtually the same phonetic qualities, except for length, and this might be insufficient for these letters to be encoded separately in the recognition system. To test this hypothesis, we conducted Experiment 2, which paralleled Experiment 1 except for the critical letters (a/á, e/é). For these letters, the addition of diacritics changes not only vowel length but also vowel quality (a⤳/ɒ/; á⤳/aː/; e⤳/ɛ/; é⤳/eː/), granting these diacritics a more prominent linguistic function.

Models of visual-word recognition like the noisy-channel model would predict identical results as in Experiment 1 (néma→NÉMA = nema→NÉMA < nama→NÉMA; mese→MESE < mése→MESE ≤ máse→MESE). In contrast, a linguistic account would predict an advantage of the identity primes for both diacritical and non-diacritical targets (néma→NÉMA < nema→NÉMA < nama→NÉMA; mese→MESE < mése→MESE < máse→MESE).

Experiment 2: Vowel length and quality

Methods

Participants

We recruited an additional sample of 72 participants (mean age = 24.6 years, 36 women, 34 men, two rather not say) via Prolific Academia. We used the same recruitment filters as in Experiment 1.

Materials

Stimulus lists were created the same way as for Experiment 1, except that the critical vowels were 'a' and 'e' for non-diacritical, and 'á' and 'é' for diacritical words. The mean Zipf-frequencies for diacritical and non-diacritical targets

Table 3 Mean response times in ms and accuracy (rate correct) in Experiment 2

	Identity	Similar	Dissimilar
Word targets			
Non-diacritical	638 (.971)	650 (.949)	666 (.944)
Diacritical	640 (.964)	641 (.970)	652 (.955)
Pseudoword targets			
Non-diacritical	709 (.967)	718 (.975)	719 (.978)
Diacritical	731 (.956)	743 (.954)	733 (.964)

were 4.75 and 4.87, respectively, and the average position of the modified letter was 3.2 for both sets.

Procedure

This was the same as in Experiment 1.

Results and discussion

The analysis plan was the same as in Experiment 1. The average RTs and accuracy per condition are shown in Table 3. Posterior distributions are shown in Fig. 2.

Similar versus identity

RTs were faster in the identity than in the similar condition, b = -0.02, SE = 0.01, 95%CrI (-0.04, -0.01). This effect was modulated by diacritical status (interaction, b = 0.03, SE = 0.01, 95% CrI [0.00, 0.05]): the advantage occurred for non-accented target words (12 ms faster for mese→MESE than mése→MESE) but not for accented target words (1 ms difference between néma→NÉMA and nema→NÉMA).

Accuracy did not differ between the identity and similar conditions, b = -0.10, SE = 0.16, 95%CrI (-0.41, 0.21); there were no signs of an interaction with diacritical status either, b = -0.09, SE = 0.21, 95%CrI (-0.31, 0.49).

Similar versus dissimilar

RTs were faster in the similar than in the dissimilar condition, b = 0.03, SE = 0.01, 95%CrI (0.02, 0.05). This advantage did not diverge based on diacritical status (interaction, b = 0.01, SE = 0.01, 95%CrI [-0.02, 0.03]) (16 ms for mése→MESE over máse→MESE; 11 ms for nema→NÉMA over nama→NÉMA).

The analyses of accuracy did not show evidence of a difference between similar and dissimilar conditions, b = -0.01, SE = 0.17, 95%CrI (-0.34, 0.35) or an interaction with diacritical status, b = 0.17, SE = 0.22, 95%CrI (-0.26, 0.59).

We again found that the dissimilar condition performed less effectively than the visually similar condition, independently of diacritical status. More importantly, as in Experiment 1, we found the identity advantage over visually similar primes vanished for the diacritical targets, while being present for the non-diacritical ones. This imitates the pattern seen in Spanish (Perea et al., 2020) and reproduces the effects reported by Perea, Hyönä, and Marcet (2022b) in Finnish, where omitting diacritics, despite changing vowel quality, did not impair early processing.

General discussion

We designed two masked priming experiments that tested whether the linguistic function of diacritics (vowel length in Experiment 1; vowel length plus quality in Experiment 2) affects the early stages of word processing in Hungarian, an extremely transparent orthography. The two experiments yielded remarkably similar results. For diacritical targets, the visually similar condition performed as well as the diacritic condition regardless of functional differences, and more effectively than the visually dissimilar condition (e.g., roka-RÓKA ≈ róka-RÓKA < reka-RÓKA). For non-diacritical targets, the visually similar condition performed more effectively than the visually dissimilar condition, but less effectively than the identity condition, again independently of language function (e.g., mozi→MOZI < mózi→MOZI < múzi→MOZI).

These findings suggest that the early encoding of diacritical vowels during word recognition does not contain phonological features, thus posing problems for a linguistic account of diacritic processing. Instead, our findings favor those models of visual-word recognition that assume that, at least in isolated contexts, perceptual information plays a preeminent role in the early stages of word recognition (e.g., noisy-channel model; Norris & Kinoshita, 2012). In these accounts, the diacritics (or, in general, certain features of the prime stimulus) could be initially missed in the first stages of letter and word encoding; note that this process would be blind to phonology. As a result, for a diacritical target such as RÓKA, the non-diacritical prime roka can be as effective as the identity prime róka (e.g., o⇒Ó). This same pattern has also been reported not only in Spanish (Perea et al., 2020), where diacritics do not alter vowel quality or length, but also in Finnish (Perea, Hyönä, & Marcet, 2022b), where diacritics unambiguously alter vowel quality. Furthermore, the same findings occur for diacritical consonants (n⇒Ñ; Marcet et al., 2020) and katakana characters (カ⇒ガ; Kinoshita et al., 2021).

The noisy-channel model explicitly postulates that once the diacritics in the prime are encoded, they provide evidence against their absence (Kinoshita et al., 2021). Thus, for non-diacritical targets, this account predicts that the visually similar (diacritical) condition would produce a cost in processing regardless of phonology (ó⇏O), as occurred in the present experiments (mozi→MOZI < mózi→MOZI). Notably, the same pattern has been obtained not only in French (Chetail & Boursain, 2019), where diacritics may have a linguistic function, but also in Spanish (Perea et al., 2020). We should note that the differences between the identity and visually similar conditions are substantially smaller in Hungarian (11–12 ms) and Spanish (17 ms) than in French (50 ms). A potential reason for this discrepancy is that while the Hungarian and Spanish experiments only included acute accent marks (´), the French experiment often (72 of 104 pairs) included more visually complex diacritics (i.e., circumflexes, as in tâper-TAPER) that may be less easily missed during prime processing (see Perea et al., 2021, for discussion).

Importantly, our findings do not necessarily disprove the idea of separate letter representations for diacritical vowels. Our claim is that this issue cannot be easily resolved with masked priming experiments because the obtained effects are conflated with the effects of visual similarity, which are asymmetric for diacritical and non-diacritical vowels. Instead, a better option to answer this question would be to focus on unprimed paradigms, such as semantic categorization. Using this task, Perea, Labusch, and Marcet (2022c) found a reading cost when the diacritics were omitted in German, where diacritics indicate vowel quality, but not in Spanish.[4] Furthermore, to have a better resolution of the time course of the effects, behavioral tasks could be combined with the recording of event-related potentials or presenting the words in a more natural setting such as sentence reading (see Marcet & Perea, 2022). Another route to take from here would be to conduct visual search tasks (see Treisman & Gormican, 1988; Wolfe, 2001) with diacritical letters (e.g., searching "a" in a group of "á" or vice versa). The perceptual account described here would predict an asymmetry (i.e., presence of a feature stands out more than its absence). Future research should examine whether the function of diacritics in the language modulates this asymmetry.

In sum, masked priming effects with diacritical letters in word recognition tasks do not reflect the linguistic function of diacritics. Rather, priming effects with diacritical primes can be considered instances of letter similarity effects (e.g., cornputer→COMPUTER). This analogy explains the asymmetrical pattern of the effects (o⇒Ó vs. ó⇏O), independently of language or function, as occurs with F⇒E vs. E⇏F. Thus, whether diacritical and non-diacritical vowels correspond to the same letter representations in the visual-word recognition system should be addressed by techniques that reflect later processing stages than those tapped by masked priming.

[4] Note that lexical decisions may be dependent on strategies based on visual similarity (see Marcet et al., 2021, for discussion).

Funding Grant from the Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital, Generalitat Valenciana (CIAICO/2021/172) to Manuel Perea, and Grant from the Ministerio de Ciencia y Universidades (PID2020-116740GB-I00 funded by the MCIN/AEI/10.13039/501100011033) to Manuel Perea.

Data availability This study was pre-registered at following OSF link: https://osf.io/2kwja.

The materials, raw data, scripts, and outputs is available at https://osf.io/5qunx/?view_only=f445a8065cf847b6b202d3a0ade6b8a1.

References

Abari, K., Rácz, Z. Z., & Olaszy, G. (2011). Formant maps in Hungarian vowels—Online data inventory for research, and education. *Twelfth Annual Conference of the International Speech Communication Association* (pp. 1609–1612). Retrieved March 28, 2022 from https://www.isca-speech.org/archive/pdfs/interspeech_2011/abari11_interspeech.pdf

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.

Borgwaldt, S. R., Hellwig, F. M., & de Groot, A. M. B. (2005). Word-initial entropy in five languages: Letter to sound, and sound to letter. *Written Language & Literacy, 7*(2), 165–184.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1), 9.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1–28.

Chetail, F., & Boursain, E. (2019). Shared or separated representations for letters with diacritics? *Psychonomic Bulletin & Review, 26*(1), 347–352.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences, 9*(7), 335–341.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680–698.

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in Cognitive Sciences, 12*(10), 381–387.

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.

Marcet, A., & Perea, M. (2017). Is nevtral NEUTRAL? Visual similarity effects in the early phases of written-word recognition. *Psychonomic Bulletin & Review, 24*(4), 1180–1185.

Marcet, A., & Perea, M. (2018). Can I order a burger at rnacdonalds.com? Visual similarity effects of multi-letter combinations at the early stages of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(5), 699–706.

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*(1), 148–155.

Marcet, A., Ghukasyan, H., Fernández-López, M., & Perea, M. (2020). Jalapeno or jalapeño: Do diacritics in consonant letters modulate visual similarity effects during word recognition? *Applied Psycholinguistics, 41*(3), 579–593.

Marcet, A., Fernández-López, M., Labusch, M., & Perea, M. (2021). 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., Fernández-López, M., Baciero, A., Sesé, A., & Perea, M. (2022). What are the letters e and é in a language with vowel reduction? The case of Catalan. *Applied Psycholinguistics, 43*(1), 193–210.
Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences, 17*(10), 517–524.
Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review, 119*(3), 517–545.
Oravecz, C., Váradi, T., & Sass, B. (2014). The Hungarian gigaword corpus. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (pp. 1719–1723). Retrieved March 28, 2022 from http://www.lrec-conf.org/proceedings/lrec2014/pdf/681_Paper.pdf
Peirce, J., Hirst, R., & MacAskill, M. (2022). *Building experiments in PsychoPy*. Sage.
Perea, M., Fernández-López, M., & Marcet, A. (2020). What is the letter é? *Scientific Studies of Reading, 24*(5), 434–443.
Perea, M., Baciero, A., & Marcet, A. (2021). Does a mark make a difference? Visual similarity effects with accented vowels. *Psychological Research, 85*(6), 2279–2290.
Perea, M., Gomez, P., & Baciero, A. (2022a). Do diacritics entail an early processing cost in the absence of abstract representations? Evidence from masked priming in English. *Language and Speech*. https://doi.org/10.1177/00238309221078321
Perea, M., Hyönä, J., & Marcet, A. (2022b). Does vowel harmony affect visual word recognition? Evidence from Finnish. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(12), 2004–2014. https://doi.org/10.1037/xlm0000907
Perea, M., Labusch, M., & Marcet, A. (2022c). How are words with diacritical vowels represented in the mental lexicon? Evidence from Spanish and German. *Language, Cognition and Neuroscience, 37*, 457–468.
Pinheiro, S., Mota, N. B., Sigman, M., Fernández-Slezak, D., Guerreiro, A., Tófoli, L. F., Cecchi, G., Copelli, M., & Ribeiro, S. (2020). The history of writing reflects the effects of education on discourse structure: Implications for literacy, orality, psychosis and the axial age. *Trends in Neuroscience and Education, 21*, 100142. https://doi.org/10.1016/j.tine.2020.100142
Protopapas, A., & Gerakaki, S. (2009). Development of processing stress diacritics in reading Greek. *Scientific Studies of Reading, 13*(6), 453–483.
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing https://www.R-project.org/
Rey, A., Dufau, S., Massol, S., & Grainger, J. (2009). Testing computational models of letter perception with item-level event-related potentials. *Cognitive Neuropsychology, 26*(1), 7–22.
Szende, T. (1994). Hungarian. *Journal of the International Phonetic Association, 24*(2), 91–94.
Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*(1), 15–48.
Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics, 63*(3), 381–389.