

## Language variation influences the sensitivity to visual detail in word recognition

Jeremy Mifsud (University of Malta) & Holger Mitterer (University of Malta)

jeremy.mifsud.13@um.edu.mt & holger.mitterer@um.edu.mt

Visual-word recognition is highly sensitive to small input variations, so that removing diacritic marks from letters considerably slows down word recognition (in French: Chetail & Boursain, 2019, *Psychol Bull Rev*, in Arabic: Perea et al., 2016, *Front Psych*). This provides a stark contrast to spoken-word recognition, in which participants can recognize words relatively easily despite strong variation such as *prowly* for *probably* (Ernestus, 2014, *Lingua*). This difference might be explained by the lower amount of variation in the visual domain, in which *prowly* is considered an error. As appealing as such an account is, as difficult it is to test it. One hint in this direction is the finding that words that are likely to be misspelled tend to be recognized slower in their correct form (Rahmanian & Kuperman, 2019, *Scie Stud Reading*), possibly due to a lower frequency of exposure to that correct form. The language situation in Malta provides a possible testing ground for this hypothesis, due to the presence of letters with diacritics in Maltese, the absence of spell checkers, and the prevalence of English keyboards that lack those letters.

Our point of departure is the causal observation that the four consonant letters in Maltese that carry a diacritic (ċ, ħ, ġ, ż) are often substituted with their diacritic-free counterparts in social-media posts and newspaper comments. A corpus study based on the user comments provided by the largest Maltese newspaper (Times of Malta) substantiated this; in 95% of the cases, words containing a letter with diacritic in their canonical form were written with the diacritic-free counterpart (e.g., *ġirja* → *girja*, Engl. ‘run’ [n.]). Moreover, we observe consistency within comments, that is, a writer consistently uses diacritics or not.

Given the high likelihood of “misspelled” words, the question arises whether Maltese readers are as sensitive to missing diacritics as readers seem to be in French and Arabic. We tested this with masked priming, a method that reveals effects that arise early in visual word recognition and sentence reading to look at the natural reading process.

In the masked-priming studies, we used three conditions: identity, missing diacritic, and letter substitution (e.g., for the target *ġirja*, the primes *ġirja*, *girja*, and *dirja*, respectively, with the following read-out forms: [dʒɪja], [gɪja], and [dɪja]). With 34 participants and 240 items, the number of observations is larger than in the earlier studies cited above. We observed small but significant priming by both the identity (11 ms) and the missing-diacritic condition (15 ms) compared to the letter condition. Though there is no significant difference between these two priming conditions, the fact that the priming effect is numerically larger in the missing-diacritic makes it unlikely that this prime is a less effective than the identity prime.

We then tested how missing diacritics influence sentence reading using eye-tracking. Since our corpus study indicated that writers tend to be consistent in their use (or not) of diacritics, we also tested whether reader adapt to the “diacritic style” of a sentence. This was achieved by presenting a target word in a carrier sentence and vary the presence of diacritics on the target and the carrier sentence orthogonally. While the data replicate classical effects of target length and word frequency on fixation durations, no effect of missing diacritics or match between carrier sentence and target in “diacritic style” was found.

The results hence indicate that Maltese readers treat diacritic-missing forms as equivalent to their nominally correct forms, both in early word recognition and in sentence reading. Given the contrast with the results from other languages, it is likely that the exposure to the “wrong” forms influences the sensitivity to variation. This might also explain recent results from Spanish, in which the vowel diacritics indicating lexical stress did not lead to a slow down (Perea et al., in press), but the diacritic on ñ did (Marcel et al., in press). Spanish informants indicate that the vowel diacritics are often omitted but the ñ is usually used, given its presence on standard keyboards. Additional research is necessary to investigate whether the higher tolerance to variation observed in Maltese and Spanish vowels arises on a prelexical level—by treating letters with and without diacritics as equivalent—or on a lexical level—by representing alternative word forms for a given word, an account that is popular in the auditory domain