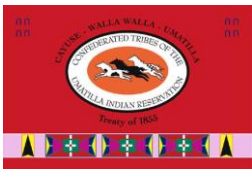# Orthographic neighborhood density effects in a Maltese visual lexical decision megastudy

Jonathan A. Geary (jonathangeary@email.arizona.edu)
Confederated Tribes of the Umatilla Indian Reservation; University of Arizona

The Words in the World (WoW) International Conference
Friday November 26, 2021

Introduction

- Using data from a Maltese visual lexical decision megastudy (Geary 2020), we analyze orthographic similarity effects on **Maltese** lexical processing.

- The number of words that are orthographically similar to a given word (i.e. its orthographic "neighbors") strongly influences lexical processing performance.
  - Neighbor (N) = Another word that differ from the target by the substitution, insertion, or deletion of a letter, or by the transposition of two adjacent letters.
  - Example – Neighbors of *trail* include: *train*, *trails*, *rail*, and *trial.*

- Lexical decision: Readers judge words with more Ns **faster + more accurately**.
  - BUT neighborhood density (ND) interacts with **frequency**: The advantage for words with more Ns is greater for low-frequency words (e.g. Andrews 1989; Hendrix and Sun 2021).
  - AND ND interacts with **target lexicality**: The ND effect is facilitatory for real words, but inhibitory for non-words (e.g. Andrews 1989, 1992; Sears et al. 1995).

Why study Maltese?

- Maltese is a **Semitic** language, spoken in Malta.
  - Approximately half of the Maltese lexicon is of Semitic origin (e.g. Bovingdon and Dalli 2006) and uses **nonconcatenative morphology** that is typical of Semitic languages.
  - Maltese is written using the **Latin alphabet**.

- To date, little research has explored **ND density effects in Semitic languages**.
  - Frost et al. (2005) obtained a facilitatory effect of ND in Hebrew lexical decision.
  - Because changes to a word's internal letters are likely to alter morphology, we may find different effects of ND/orthographic similarity in Semitic languages.

- Also, little research has explored **non-word processing in Maltese** (cf. Twist 2006).

- We address both of these gaps in the present set of analyses.

MaltLex – Maltese visual lexical decision megastudy

- We analyze data from the MaltLex visual lexical decision database (Geary 2020).

- MaltLex includes 210,960 lexical decision responses to 21,900 unique targets.
  - 104,644 responses to 10,951 real-word targets;
  - 106,316 responses to 10,949 non-word targets.

- Geary (2020) showed that contextual diversity counts (i.e. the number of documents in which a word occurs) taken from Korpus Malti v3.0 (Gatt and Čéplö 2013) outperform traditional word frequency norms in predicting lexical processing performance.

- We use MaltLex data to assess ND effects on lexical processing in Maltese for:
  - real versus non-word targets (Analysis 1; cf. Andrews 1989; Hendrix and Sun 2021);
  - real-word targets at differing levels of frequency (Analysis 2; cf. Andrews 1989, 1992).

Analysis 1 – Neighborhood density * Lexicality

- We compared orthographic neighborhood density effects on RTs and response accuracy for real versus non-word targets in the MaltLex dataset (Geary 2020).

- We fitted LMER and GLMER models with **the interaction of log orthographic neighborhood density by target lexicality** as a fixed effect.

- Participants responded faster and more accurately to real-word targets as ND increased (RT: $t(261.1) = -19.97$, $p < 0.001$; Acc: $z = 20.60$, $p < 0.001$), while the ND advantage diminished for non-word targets (RT: $t(20,760) = 28.29$, $p < 0.001$; Acc: $z = -20.48$, $p < 0.001$).
  - Participants responded slower + less accurately to non-word targets as ND increased.

- Our results are consistent with previous studies (e.g. Andrews 1989; Hendrix and Sun 2021):
  - Real Maltese words that are more similar to other words are easier to judge as words.
  - Maltese-like non-words that are more similar to real Maltese words are harder to reject.

Analysis 2 – Neighborhood density * Frequency

- We assessed ND effects on RTs and response accuracy for real-word targets differ as targets increase in frequency in the MaltLex dataset (Geary 2020).

- We fitted LMER and GLMER models with **the interaction of log orthographic neighborhood density by log contextual diversity** as a fixed effect.

- Participants responded faster and more accurately to real-word targets as ND increased (RT: $t(1,083) = -9.41$, $p < 0.001$; Acc: $z = 10.04$, $p < 0.001$), while the ND advantage on RTs diminished as CD increased (RT: $t(11,240) = 9.41$, $p < 0.001$; Acc: $z = -1.42$, $n.s.$).

- Our results are consistent with previous studies (e.g. Andrews 1989, 1992; Sears et al. 1995):
  - Real Maltese words that are more similar to other words are easier to judge as words.
  - However, the advantage for high-ND targets diminishes as targets increase in frequency.

Analysis 3 – Diacritic-based word-similarity

- Three pairs of Maltese letters differ only in diacritics: "g, ġ", "h, ħ", and "z, ż".

- A handful of MaltLex non-word targets differed from real Maltese words only in whether diacritics had been included/omitted on 1–2 letters (N = 23; i.e. 0.26%).

  - Examples – *$\mathbf{g}$era* resembles *ġera* 'he ran, flowed' except in the absence of a dot above "g"; *vaġun* resembles *va$\mathbf{g}$un* 'wagon,' except in the inclusion of a dot above "ġ".

- Participants repeatedly pointed this out during the post-session debriefings, having perceived these targets to be more difficult to reject.

  - Such non-word targets were relatively few, and their real-word counterparts were of relatively low frequency (M = 0.7 occurrences per thousand contexts in Korpus Malti).

- Is it actually harder for readers to reject such non-words in lexical decision?

Analysis 3 – Diacritic-based word-similarity

- We analyzed the effect of diacritic-based word-similarity on RTs and response accuracy for non-word targets in the MaltLex dataset (Geary 2020).

- We fitted LMER and GLMER models with **the target's log neighborhood density and the real-word counterpart's log contextual diversity** as fixed effects.

- Participants responded slower and less accurately to non-word targets as ND increased (RT: $t(147.9) = 7.59$, $p < 0.001$; Acc: $z = -17.01$, $p < 0.001$).

- Participants responded slower and less accurately to non-word targets with a diacritic-based word counterpart (RT: $t(16,350) = 6.08$, $p < 0.001$; Acc: $z = -6.50$, $p < 0.001$).

- That is, participants struggled to reject non-word targets that resembled real words except in their diacritics. This effect was independent of the ND effect.

Discussion

- The results of Analyses 1–2 are consistent with prior studies of orthographic neighborhood density effects in non-Semitic languages:

  - Readers judge real words faster + more accurately as ND increases (Analyses 1–2).

  - But readers judge non-words slower + less accurately as ND increases (Analysis 1).

  - The ND advantage for words diminishes as targets increase in frequency (Analysis 2).

- Analysis 3 revealed a novel orthographic similarity effect on lexical processing:

  - Readers are slower + less accurate to reject non-words that resemble a real word except in their diacritics (e.g. *gera resembles ġera, *vaġun resembles vagun).

  - This reveals a novel characteristic which researchers must control for in selecting non-word stimuli in Maltese (and in other languages with similar use of diacritics).

Grazzi! Thank you!

Selected references

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15: 802–814. https://doi.org/10.1037/0278-7393.15.5.802.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18: 234–254. https://doi.org/10.1037/0278-7393.18.2.234.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review* 4: 439–461. https://doi.org/10.3758/BF03214334.

Bovingdon, R. and Dalli, A. (2006). Statistical analysis of the source origin of Maltese. In Wilson, A., Archer, D., and Rayson, P. (eds.), *Corpus linguistics around the world*, 63–76. New York: Rodopi.

Frost, R., Kugler, T., Deutsch, A., and Forster, K. I. (2005). Orthographic structure versus morphological structure: Principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31: 1293–1326. https://doi.org/10.1037/0278-7393.31.6.1293.

Gatt, A., and Čéplö, S. (2013). Digital corpora and other electronic resources for Maltese. In *Proceedings of the international conference on corpus linguistics*. University of Lancaster.

Geary, J. (2020). MaltLex: A database of visual lexical decision responses to 11,000 Maltese words. Poster presented at *the 33rd Annual CUNY Conference on Human Sentence Processing (CUNY 2020)*, Amherst, MA, March 20, 2020. https://osf.io/8rp3a/.

Hendrix, P., and Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 47: 157–183. http://dx.doi.org/10.1037/xlm0000819.

Sears, C. R., Hino., Y., and Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21: 876–900. https://doi.org/10.1037/0096-1523.21.4.876.

Twist, A. E. (2006). *A psycholinguistic investigation of the verbal morphology of Maltese* (PhD Thesis). University of Arizona. Tucson, AZ. http://hdl.handle.net/10150/194996.