Modeling Maltese plurals with Linear Discriminative Learning

Jessica Nieder[a], Yu-Ying Chuang[b], Ruben van de Vijver[a] & Harald Baayen[b]

nieder@phil.hhu.de, yu-ying.chuang@uni-tuebingen.de, Ruben.Vijver@hhu.de, Harald.Baayen@protonmail.com

[a]Department of General Linguistics, Heinrich-Heine-Universität Düsseldorf, Germany;
[b]Department of General Linguistics, Eberhard-Karls-Universität Tübingen, Germany

DFG Research Unit FOR 2373: Project MALT & Project ART

Spoken Morphology Workshop 2021

Introduction

- How is the content of the mental lexicon structured?
  - **morpheme-based models**: unit of storage = morphemes; we need information about constituent morphemes (e.g. *Item and Arrangement* or *Item and Process*)

    (Hockett, 1954)
  - **word-based models**: unit of storage = words, we need information about the whole word form (e.g. *Word and Paradigm*)

    (Blevins, 2006)
- languages with a rich variety of inflections are a challenge for the prediction of inflectional classes → Maltese!

- 2 main ways to characterize a plural of a noun:
  12 sound plural suffixes: concatenative (traditionally called suffixation)
  annimal - *annimali* 'animal(s)'
  11 broken plural patterns: non-concatenative as changes in the syllabic structure of the plural in comparison to the singular
  ballun - *blalen* 'ball(s)'

Research Question

Is it possible to computationally model comprehension and production of Maltese word forms without morphemes?

This talk

- computational modeling of the Maltese noun plural system without morphemes using the Linear Discriminative Learner (LDL)

Experimental Background

Modeling of form without semantics

- computational modeling of the plural classes on the basis of the form of Maltese plurals with TiMBL, NDL and an Encoder-Decoder Network
- results: Plural classes can be predicted from the form of Maltese plurals without using morphemes, but information about the plural form is needed for correct predictions.
- The role of meaning is still open.
  - → What role does the semantics play for Maltese noun inflections?

- 6522 Maltese nouns manually compiled from a broken plural set collected by Schembri, 2012 and the MLRS Korpus Malti v. 2.0 and 3.0 (Gatt & Čéplö, 2013)
- distribution: 6522 nouns in total = 3382 plurals (892 broken, 2461 sound), 3140 singulars
- manually added information: CV structure, number of occurrences (based on the Korpus Malti v. 3.0), origin and gender (based on Aquilina (1987)), number, concreteness (abstract vs. concrete), verbal noun, dual noun, suppletive noun, collective noun

LDL Classification

- model of the 'discriminative lexicon' (DL) proposed by Baayen, Chuang, Shafaei-Bajestan, and Blevins (2019)
- goals of the model
    - understanding words' meanings given their forms
    - producing words' forms given their meanings
- computation of all (possible) forms and meanings simultaneously
- mapping of numerical form vectors onto numerical semantic vectors and vice versa

	Lexeme	Number	Gender
kelb	KELB	singular	M
kelba	KELB	singular	F
klieb	KELB	plural	M

Table 1: Paradigm for the Maltese noun *kelb* ‘dog’.

$$
\boldsymbol{C} = \begin{array}{c} \\ \textit{kelb} \\ \textit{kelba} \\ \textit{klieb} \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} \textit{\#ke} & \textit{kel} & \textit{elb} & \textit{lb\#} & \textit{lba} & \textit{ba\#} & \textit{\#kl} & \textit{kli} & \textit{lie} & \textit{ieb} & \textit{eb\#} \end{array} \\
\left(\begin{array}{ccccccccccc}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1
\end{array} \right)
\end{array}
$$

- 2 possibilities: simulated vectors or Corpus-based vectors using `fasttext` (Joulin, Grave, Bojanowski, & Mikolov, 2016; Joulin, Grave, Bojanowski, et al., 2016)

Modeling Maltese Plurals

Construction of semantic matrix - simulated vectors

$$\boldsymbol{S} = \begin{array}{c} \\ \textit{kelb} \\ \textit{kelba} \\ \textit{klieb} \end{array} \begin{pmatrix} S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 \\ 0.46 & 4.16 & 8.50 & -4.46 & 8.96 & -4.11 & 8.42 & 9.21 & -25.75 & 15.83 & -14.93 \\ 0.61 & -11.93 & 8.09 & 1.00 & 3.44 & -11.98 & 8.72 & -4.75 & -33.29 & 10.39 & -2.12 \\ 5.67 & 9.84 & 11.26 & 0.85 & 10.69 & -4.24 & 0.21 & 4.81 & -26.47 & 10.82 & -11.76 \end{pmatrix}$$

$$
\begin{array}{lllll}
\textit{kelb}: & \overrightarrow{\text{KLB}} & + \overrightarrow{\text{SINGULAR}} & + \overrightarrow{\text{MASCULINE}} \\
\textit{kelba}: & \overrightarrow{\text{KLB}} & + \overrightarrow{\text{SINGULAR}} & + \overrightarrow{\text{FEMININE}} \\
\textit{klieb}: & \overrightarrow{\text{KLB}} & + \overrightarrow{\text{PLURAL}} & + \overrightarrow{\text{MASCULINE}}
\end{array}
$$

- resulting semantic vectors $\overrightarrow{\text{KELB}}$, $\overrightarrow{\text{KELBA}}$, and $\overrightarrow{\text{KLIEB}}$ reflect inflectional similarities
- singulars more similar to other singulars than to plurals, masculine nouns more similar to each other than to feminine nouns...

Modeling Maltese Plurals

Construction of semantic matrix - `fasttext` vectors

- pre-trained word vectors for 157 languages, trained on *Common Crawl* and *Wikipedia*
- construction of semantic vectors for words from semantic vectors of substrings of words
- 300 dimensional vectors
- available at: https://fasttext.cc/docs/en/crawl-vectors.html

- we extracted `fasttext` vectors for 4,056 Maltese nouns (2266 singulars, 1781 plurals)

- triphones vs. syllable cues
- for syllable cues all word forms in the data set were manually syllabified
- LDL available as package `JudiLing` for the Julia environment (Luo, 2021)

	simulated	fasttext
triphone	93.8%	95.6%
syllable	99.95%	99.93%

Table 2: Comprehension accuracy.

	simulated	fasttext
triphone	92.7%	91.2%
syllable	100%	96.3%

Table 3: Production accuracy.

- to validate the `fasttext` vectors we reduced the given 300 dimensions to a 2-dimensional plane: Principal Components Analysis

Modeling Maltese Plurals

Validation of `fasttext` vectors

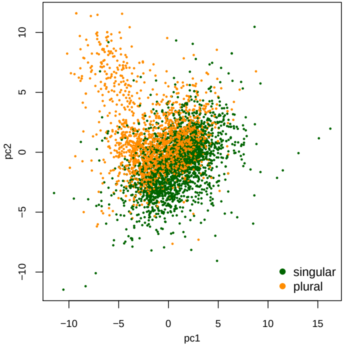


Figure 1: Results of PC analysis: singular vs. plural.

- clear clusters of singulars and plurals
- LDA accuracy (correct classification of number): 89%

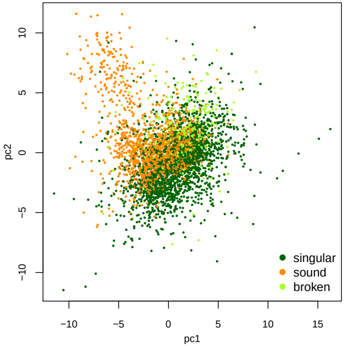


Figure 2: Results of PC analysis: broken vs. sound plural.

- LDA accuracy for sound plurals: 90%
- for broken plurals: 63%

Modeling Maltese Plurals

Validation of `fasttext` vectors

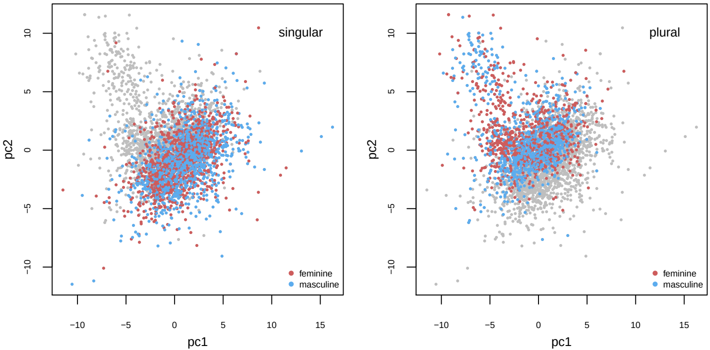


Figure 3: Results of PC analysis: semantic vectors gender

- for singulars, LDA accuracy of correct gender classification: 84%
- for plurals, LDA accuracy of correct gender classification: 76%

Summary results

- we modeled comprehension and production of Maltese plurals with several LDL models $\rightarrow$ excellent predictions for simulated and for `fasttext` semantic vectors
- `fasttext` vectors are the more realistic semantic vectors
- syllables are better predictors than triphones

- are the `fasttext` vectors predictive for modeling experimental reaction time data?

- experimental data set from Nieder, van de Vijver, and Mitterer (accepted)
- reaction time experiment with 59 participants
- based on selection of two frequent and two infrequent sound plural suffixes (*-i* and *-ijiet*,*-a* and *-at*) and broken plural patterns (*CCVVCVC* and*CCVVC*, *CCVjjVC*and *CCVVCV*)
- experimental conditions:
  - pluralType: sound vs. broken
  - patternFreq: frequent vs. infrequent

- model setup
  - bi-syllable as cues
  - `fasttext` vectors
- stimuli for which `fasttext` vectors are not available were excluded
- statistical analyses
  - generalized additive mixed models (Wood, 2017)
  - current analyses are preliminary

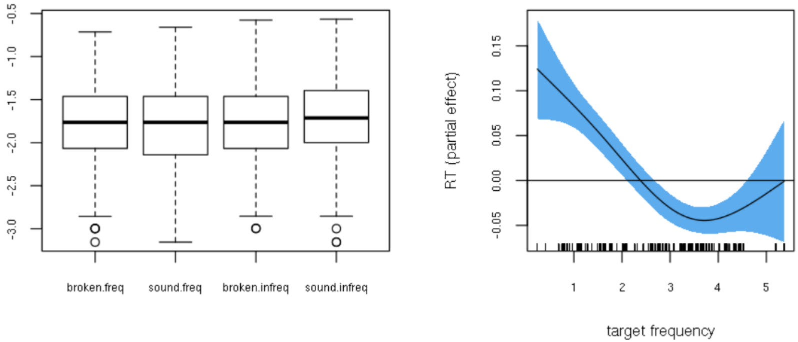


Figure 4: baseline GAM model results

- effect of priming conditions is not strong
- target frequency is predictive

pre-activation distance

- closely related to the NDL measure of PrimeToTargetPreActivation (Baayen & Smolka, 2020)
- Euclidean distance between the predicted semantic vector of prime and the gold standard semantic vector of target: $dist(\hat{s}_{prime}, s_{target})$
- interpretation: smaller value, larger priming effect

prime support

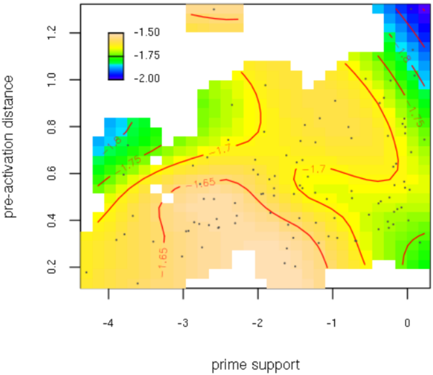
- sum of the semantic support that the bi-syllable cues of the prime word receives
- eg., *bar.ka*: 0.02 (#.bar) + 0.05 (bar.ka) + 0.12 (ka.#) = 0.19
- interpretation: the degree to which sublexical cues are supported by semantics

Baayen and Smolka (2020), page 19:

> *To model the effect of priming, we presented the triphones of the prime to the network, and summed the weights on the connections from these triphones to the pertinent target to obtain a measure of the extent to which the prime pre-activates its target (henceforth PrimeToTargetPreActivation).*

Modeling Maltese Plurals

RT model - LDL measures as predictors



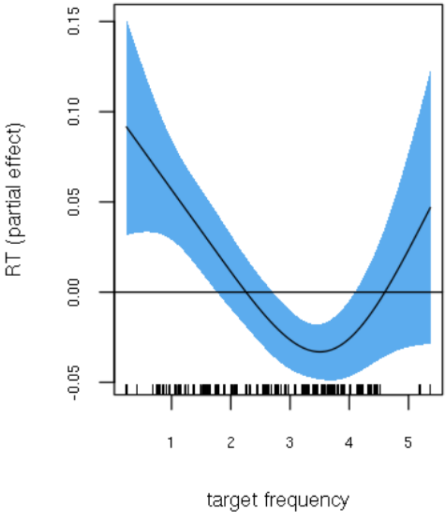
warmer color, longer RTs

general patterns:

- larger prime support, shorter RTs
- larger pre-activation, shorter RTs

Modeling Maltese Plurals

RT model - LDL measures as predictors



- target frequency is again predictive

- the model with LDL measures has better model fit
- prime support: when prime words' sublexical cues receive strong semantic support, RTs are shorter
- pre-activation distance: when prime words' meanings are farther away from target words' meanings, RTs are shorter.
  $\rightarrow$ Contrary to prediction!
  $\rightarrow$ interference effect when prime meaning is too close to target meaning?

- modeling Maltese plurals with LDL gives us excellent results for comprehension and production
- syllables are better predictors than triphones
- LDL can be used to model reaction times for Maltese plurals
- results support a word-based model of morphology (no morphemes!)

Grazzi ħafna!

References I

Aquilina, J. (1987). *Maltese-English Dictionary*. Midsea Books Ltd.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. *Complexity*, *2019*.

Baayen, R. H., & Smolka, E. (2020). Modeling morphological priming in german with naive discriminative learning. *Frontiers in Communication, section Language Sciences*, 1–40.

Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics*, *42*, 531–573.

Gatt, A., & Čéplö, S. (2013). Digital corpora and other electronic resources for Maltese. In A. Hardie & R. Love (Eds.), *Corpus linguistics* (pp. 96–97). UCREL.

References II

Hockett, C. F. (1954). Two models of grammatical description. *Word*, *10*(2-3), 210–234.

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). Fasttext.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.

Luo, X. (2021). *JudiLing: An implementation for Linear Discriminative Learning in Julia* (tech. rep.). https://github.com/MegamindHenry/JudiLing.jl

Nieder, J., van de Vijver, R., & Mitterer, H. (accepted). Priming Maltese Plurals: Representation of sound and broken plurals in the mental lexicon. To appear in: *Mental Lexicon*, *16*(1).

Schembri, T. (2012). *The Broken Plural in Maltese: A Description*. Univ.-Verl. Brockmeyer.

References III

Wood, S. (2017). *Generalized additive models: An introduction with r* (2nd ed.). Chapman; Hall/CRC.