Is Four Better than Two? The Influence of Bilingual and Multilingual Metrics in an Implicit Bilingual Statistical Learning Task

Anna Marie Liebelt & Davide Crepaldi

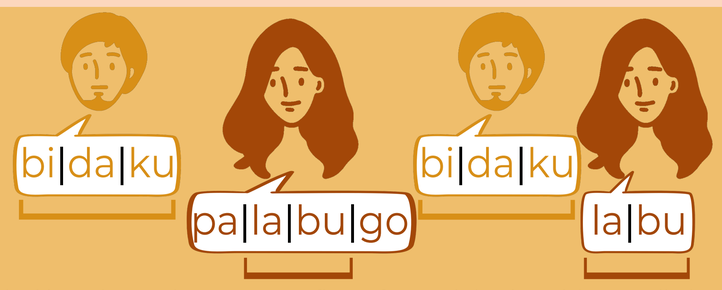
International Symposium on Bilingualism 13/06/2025

experiment GitHub & poster

Introduction

Bilingual statistical learning

- **goal**: understand structure of world
- **dual-language** learning stream - need to track regularities in two languages simultaneously
- **explicit cue** to language membership: distinct speakers
- bilinguals can track **more complex** patterns with **fewer explicit** cues[1]



Bilingual statistical learning task: a constant flow of speech is played, in two distinct voices (red for female, yellow for male) for each language. Each language is composed of individual chunks (morphemes, represented by vertical lines), with some morphemes always appearing together (represented by brackets).

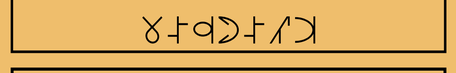
Which multilingual metrics help learning in an implicit bilingual statistical learning task?

Methods

1. Training: "pay attention to the words on the screen" language 1 language 2

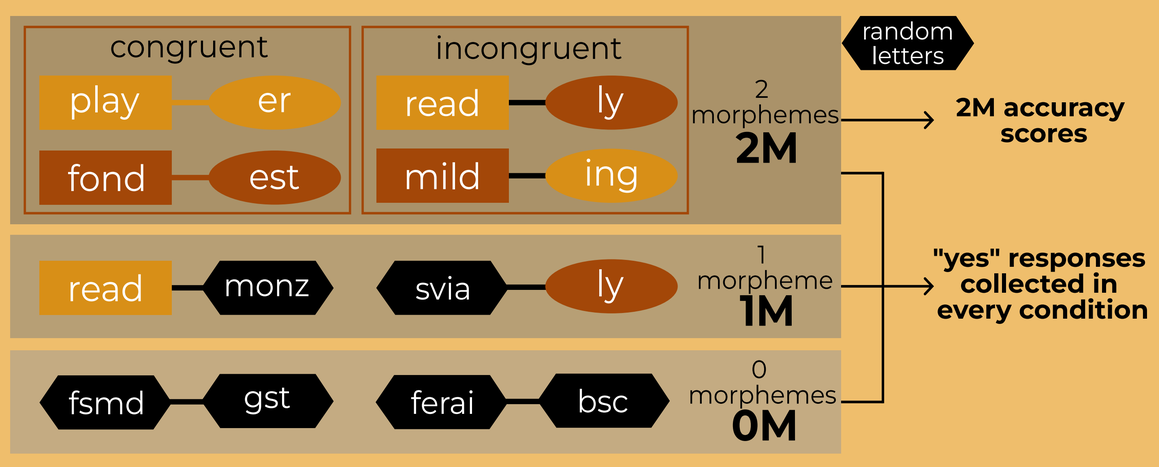
read ing · play ing · read er · mild ly · fond ly · mild est

Words shown in BACS-1 font[2]:



Participants need to
1. Segment words into parts
2. Track co-occurrence of parts

2. Testing: "does this belong to what you saw before?"

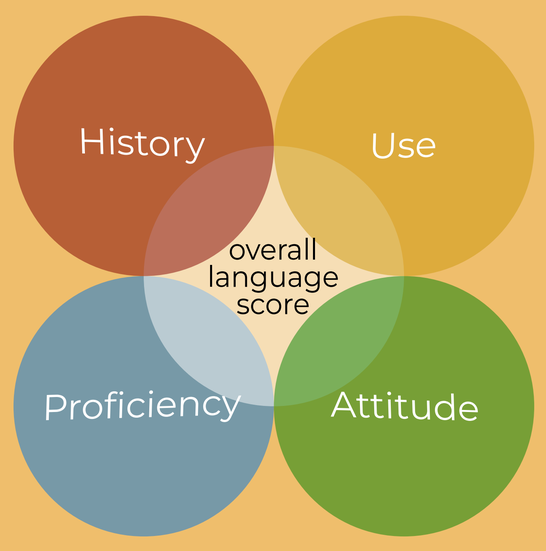


3. Familiarity: "which item have you seen before?"

read or clwd · gp or ly

4. Multilingual Language Profile
adapted from the Bilingual Language Profile[3], for up to 4 languages (L1, L2, L3, L4)

goal: **continuous** measures of multilingualism



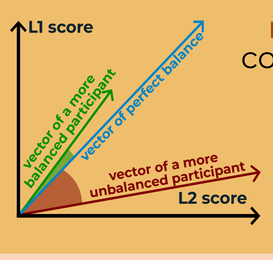
multilingual experience
sum of all scores

bilingual balance
difference of L1 & L2

multilingual balance
"entropy" of all scores

example of a high degree of multilingual balance

example of a low degree of multilingual balance



multilingual balance
cosine similarity of scores

$$\frac{A \cdot B}{\|A\| \, \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}$$

comparison of multilingual balance metrics

language scores	cosine similarity	entropy
[200 200 0 0]	0.71	1
[200 200]	1	1
[200 200 200 200]	1	2

Results

n = 193

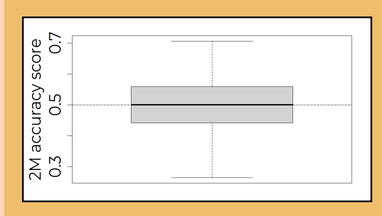


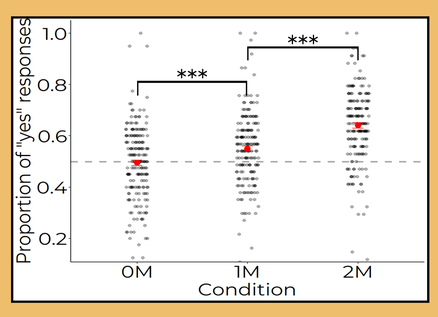
Figure 1: Boxplot of the accuracy scores in the 2M condition.

Chance level (0.5) shown as a dotted line.

Figure 2: Proportion of "yes" responses in the no morphemes (0M), one morpheme (1M), and two morphemes (2M) testing conditions.

Group means are shown by a red circle with error bars.

Chance level (0.5) is shown as a dotted line.



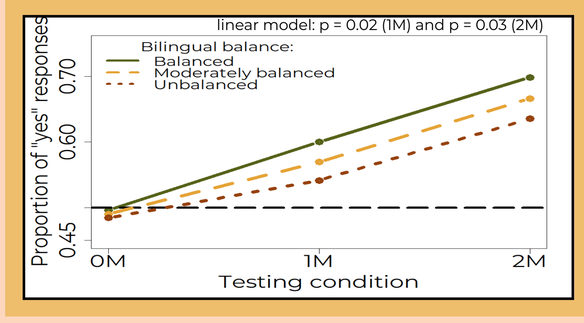


Figure 3: model fit for the interaction of the amount of "yes" responses in each condition and different amounts of imbalance of participants' scores for their first two languages.

Chance level (0.5) shown as a dotted line.

EXPERIMENT 2

IDENTICAL METHODS TO EXPERIMENT 1, WITH THE ADDITION OF:

0.5. Pre-training: "spell these word parts"

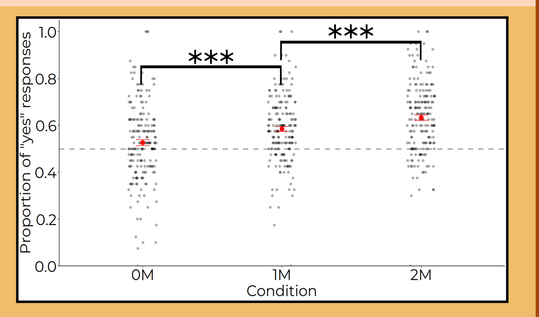
read est · mild ing

Participants only need to track co-occurrence of parts

n = 183

Figure 4: Proportion of "yes" responses in the 0M, 1M, and 2M testing conditions.

Group means are shown by a red circle with error bars. Chance level (0.5) is shown as a dotted line.



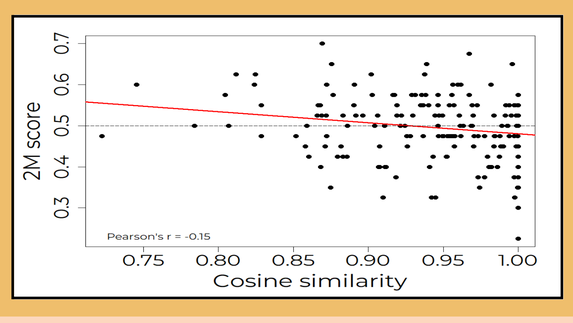


Figure 5: model fit for the interaction of scores in the 2M condition and cosine similarity. Chance level (0.5) shown as a dotted line. Cosine similarity of 1 indicates perfect balance.

cosine similarity*correct: p = 0.008

Discussion

Greater number of familiar morphemes - higher chance of responding "yes"
- expansion of Lelonkiewicz et al. (2020)[4]

Group 2M performance at chance level
- task too difficult OR no multilingual metrics with significant effect

More balanced bilinguals say "yes" more when there are more familiar morphemes
- no significant interaction with multilingual balance - strictly **bi**lingual phenomenon?

Greater number of familiar morphemes - higher chance of responding "yes"
- replication of experiment 1

Group 2M performance still at chance level
- task still too difficult to see group-level effect

More imbalanced multilinguals have higher 2M accuracy
- opposite direction of effect as in other research
- purely **multi**lingual phenomenon: L1-L2 difference non sig.
- not down to more language exposure: multilingual experience non sig.

What precisely is the role of multilingual balance? Does it depend on task difficulty?

References

1. Antovich & Graf Estes (*Developmental Science*, 2018)
2. Vidal et al. (*Behavior Research Methods*, 2017)
3. Gertken et al. (Measuring L2 Proficiency, p.208-225, 2014)
4. Lelonkiewicz et al. (*Journal of Memory and Language*, 2020)