SAND2021-7281C

# Using Machine Learning to Predict Bilingual Language Proficiency from Reaction Time Priming Data

Laura E. Matzen, Christina L. Ting & Mallory C. Stites
Sandia National Laboratories, Albuquerque, NM, USA

Contact information: lematze@sandia.gov

## Abstract

Studies of bilingual language processing typically assign participants to groups based on their language proficiency and average across participants in order to compare the two groups. This approach loses much of the nuance and individual differences that could be important for furthering theories of bilingual language comprehension. In this study, we present a novel use of machine learning (ML) to develop a predictive model of language proficiency based on behavioral data collected in a priming task. The model achieved 75% accuracy in predicting which participants were proficient in both Spanish and English. Our results indicate that ML can be a useful tool for characterizing and studying individual differences.

## Methods

### Materials

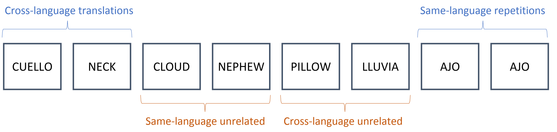
- 30 Spanish nouns and their translations in English
- No special characters, cognates or false cognates
- Matched on length, frequency, and orthographic neighborhood size
- The words were paired in 8 types of pairings: repetitions in the same language, translations across languages, and unrelated pairings within and across languages. There were a total of 240 pairs with every word appearing in every possible pair type, 4 times as a prime and 4 times as a target
- The word pairs were divided into 4 blocks of 60 pairs each with each word appearing twice in each block
  - Once as part of a related pair and once as part of an unrelated pair

| | English | Spanish |
|---|---|---|
| **Avg. Length** | 5.37 | 5.47 |
| **Avg. Frequency** | 106.93 | 99.04 |
| **Avg. English Orthographic Neighborhood Size** | 6.77 | 0.77 |
| **Avg. Spanish Orthographic Neighborhood Size** | 2.27 | 5.43 |

### Procedure

- Participants completed a short language proficiency questionnaire based on the Language Experience and Proficiency Questionnaire (Marian, Blumenfeld & Kaushanskaya, 2007). They were asked:
  - To list up to four languages they know, in order of dominance
  - To list up to four languages they know, in order of acquisition
  - The age at which the acquired each language (infant, child, teen, adult or never)
  - What factors contributed to them learning the language (family, friends, formal language classes, TV/movies, self-instruction, etc.)
  - What percentage of the time they are currently exposed to English and Spanish
  - How much time they have spent living in areas where English and Spanish are the dominant language
  - Their level of proficiency in English and Spanish on a 0-10 scale ("None" to "Perfect")
- Participants were then given instructions and examples of the task
  - They were asked to press the "B" key for words that had 5 letters or fewer and the "N" key for words that had 6 letters or more
  - They were told that words would often be repeated or appear in translated pairs
  - Participants responded to every word. Each word remained on the screen until the participant responded
- The task was completed in four blocks with self-paced breaks in between. Each block took about 2 minutes to complete

**An example of a sequence of stimuli within the experiment. Participants saw the words one at a time and responded to each one.**




### Participants

- 95 participants recruited via Amazon Mechanical Turk
- One group (40 participants) had Mechanical Turk's Spanish fluency qualification
- The other group (55 participants) was not required to have Mechanical Turk's Spanish fluency qualification
  - 13 participants from this group were rejected because they did not complete the task or failed to follow the task instructions

Group without Spanish Fluency Qualification

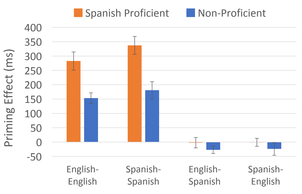
All said English was their dominant language
41 said English was their first language (1 said Mandarin)
All had lived in predominately English-speaking countries (mean 38 years)
7 had also lived in predominately Spanish-speaking countries (mean 3 years)
21 reported no exposure to Spanish
15 said their Spanish was 3 or below on the 0-10 scale
3 said their Spanish was intermediate (5-7 rating)
3 said they were proficient in Spanish (8-10 rating)

Group with Spanish Fluency Qualification

33 said Spanish was their dominant language and the first language they acquired
3 said English was their dominant language and the first language they acquired
2 said Spanish was their first language and English was their dominant language
2 said English was their first language and Spanish was their dominant language
All had lived in predominately Spanish-speaking countries (mean 29 years)
23 had also lived in English-speaking countries (mean 9 years)
One person rated their Spanish proficiency as a 7 on the 0-10 scale
All other participants said they were proficient in Spanish (8-10 rating)

- We used a self-reported proficiency of 8 or higher on the 0-10 scale as our cutoff for considering someone to be proficient in Spanish
  - Three participants who did not have Mechanical Turk's fluency qualification were moved to the "Spanish Proficient" group
  - One participant who had Mechanical Turk's fluence qualification was moved to the "Non-Proficient" group
  - The participants indicated with asterisks in the graphic above were moved to the other group for the analysis and modeling

## Behavioral Results




**Summary:**

- **There were priming effects for same-language pairs but not cross-language pairs**
- **Priming effects were larger for the Spanish Proficient group**

Details:

A 2x4 (Spanish Proficiency x Priming Condition) ANOVA showed:

- Significant effects of proficiency group ($F(1,240) = 16.77$, $p < .001$) and condition ($F(3,240) = 90.04$, $p < .001$)
- A significant interaction between the two ($F(3, 240) = 5.11$, $p < .01$).

Spanish Proficient group had significantly larger priming effects:

- English-English condition ($t(67) = 3.50$, $p < .001$)
- Spanish-Spanish condition ($t(67) = 3.59$, $p < .001$).

## Model of Bilingual Language Proficiency

### Can we use reaction time data to make predictions about individual differences such as bilingual language proficiency?

- We tried to use machine learning to learn a function that maps the RTs of known participants to their corresponding proficiency labels. Then we can use the RTs of new participants to predict their proficiency.

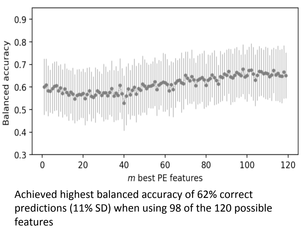
Linear support vector machine (SVM)

- Tries to find a hyperplane that separates labeled data into two (or more) classes
- New data items are labeled based on where they fall relative to the hyperplane
- The model's weights that tell us the relative importance of different features

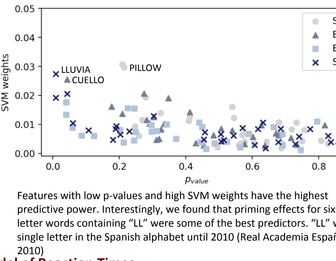
Model validation

- *k*-fold cross validation
- Data are split into *k* consecutive folds. *k*-1 sets used for training, last one used for testing
- We used *k* = 5 and 10 runs of the cross-validation experiment
- Created one model for the priming effects and another for the raw response times for every word

### Results for Model of Priming Effects

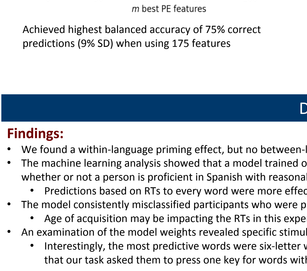



Achieved highest balanced accuracy of 62% correct predictions (11% SD) when using 98 of the 120 possible features




Features with low p-values and high SVM weights have the highest predictive power. Interestingly, we found that priming effects for six-letter words containing "LL" were some of the best predictors. "LL" was a single letter in the Spanish alphabet until 2010 (Real Academia Española, 2010).

### Results for Model of Reaction Times




Achieved highest balanced accuracy of 75% correct predictions (9% SD) when using 175 features

| | | Predicted Group: Spanish Proficient | Predicted Group: Non-proficient |
|---|---|---|---|
| **Actual Group** | Spanish Proficient | 0.74 (0.15) | 0.26 (0.15) |
| | Non-proficient | 0.23 (0.13) | 0.77 (0.13) |

- There were 5 proficient Spanish speakers who reported learning English beginning in infancy and Spanish as children or teens. All 5 of those participants were sometimes or always misclassified by the model.
- There were 5 participants in the non-proficient group who were consistently misclassified by the model. One had poor performance overall (only 75% correct on the word length task). The other 4 were all multilingual, including 2 who reported some knowledge of Spanish.

## Discussion

### Findings:

- We found a within-language priming effect, but no between-language priming effect for the Spanish Proficient group
- The machine learning analysis showed that a model trained on response time data (either priming effects or RTs to each word) can predict whether or not a person is proficient in Spanish with reasonably good accuracy
  - Predictions based on RTs to every word were more effective than predictions based on priming effects in this case
- The model consistently misclassified participants who were proficient in Spanish but learned it later in life
  - Age of acquisition may be impacting the RTs in this experiment
- An examination of the model weights revealed specific stimuli that were the most predictive of individuals' Spanish proficiency
  - Interestingly, the most predictive words were six-letter words containing "LL" which may have been tricky for Spanish speakers given that our task asked them to press one key for words with 5 letters or fewer and another for words with 6 letters or more

### Limitations:

- The lack of cross-language priming effects suggests that the word length task did not encourage semantic priming
  - Having participants respond to every word and intermixing cross- and between-language pairs may also have reduced the effects of semantic priming
- Relying on self-reports of language proficiency from anonymous online participants is not ideal
- Participants in our non-proficient group are not monolingual. Many have some knowledge of Spanish or are proficient in other languages

### Conclusions: Machine learning can enable new approaches to the study of bilingualism and other individual differences.

- Allows for the discovery of differences between participants based on their patterns of performance, instead of averaging across those differences through group-level analyses
- Can be used predictively if individual characteristics are not known in advance
- Can identify specific stimuli that provide the most differentiation, enabling new approaches to item analyses

**References:**

- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50(4),* 940-967.
- Real Academia Española (2010). *Ortografía de la lengua Española.* Espasa.

This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.