

IDs in L2 intuitions of phrasal frequency and association strength

Ryo Maie

The **University** of **Tokyo**

Wei Yi

Peking University

2023 Symposium on Vocabulary & Learning
October 7th, 2023

Please post ANY questions and comments here!

Event Name: MaieYi@VocabSIG
📌 Questions and Comme... ▾

My Posts

Others' Posts

Popular

Please post any questions and comments here!

Please enter text.

Name (optional)



Send



<https://app.learnwiz.one/events/4638852633>

Order the following sets of words in frequency

noun

option

selection

choice

adjective

regular

common

frequent

verb

banter

converse

talk



Order the following sets of words in frequency

noun

choice

option

selection

adjective

common

frequent

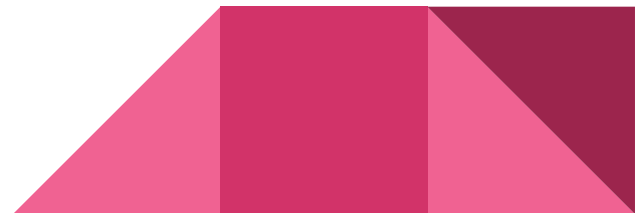
regular

verb

talk

converse

banter



Why study frequency intuition?

Theoretical motivation

- Human memory (Zacks & Hashler, 2002)
- Decision-making process (Kahneman, 2003; Tversky & Kahneman, 1974)

Practical motivation

- Language teachers can use their intuitions to select teaching materials (focusing on frequent items)



Research: Word frequency intuition

Study	Participants	Target words	Source of objective frequencies	Task	Response scale	Accuracy (correlations)
Tryk, 1968	L1	100 nouns	Thorndike-Lorge frequency tabulation	ME	Frequency of occurrence per 20 years	.74–.78
Shapiro, 1969	L1	91 words (divided for two tasks)	Thorndike-Lorge and Kučera-Francis tabulations	MRO & ME	No anchoring numbers used for ME	.92–.96 (ME) .95–.98 (Dimroth, 2018)
Carroll, 1971	L1	60 words selected from Shapiro (1969)	Thorndike-Lorge and Kučera-Francis tabulations	MRO & ME	No anchoring numbers used for ME	.97 and .93 for lexicographers and other adults, respectively
Backman, 1976	L1	50 words translated from Shapiro (1969)	Three frequency counts, one American and two Swedish	ME	No anchoring numbers used	.93
Ringeling, 1984	L1 & L2 (mixture)	24 nouns (18 selected from a word frequency book)	Word frequency book by Carroll, Davies, & Richman (1971)	MRO	NA	.61–.87 (personal) .74–.90 (general)
Schmitt & Dunham, 1999	L1 & L2	12 sets, each containing 5 near synonyms	BNC and COBUILD frequency counts	ME	Anchor words used for each lexical set	.53 and .58 for L1 & L2, respectively
Balota, Pilotti & Cortese, 2001	L1	2,938 monosyllabic words, divided into multiple sets	Kučera-Francis and CELEX frequency counts	ME	never/once a year/once a month/once a week/every two weeks/once a day/several times a day	.78–.83
Alderson, 2007	L1 & L2 (mixture)	100/50/25 words for study 1/2/3, respectively	BNC frequency counts	MRO & ME	12-point scale used for ME	.63, .68 and between .59–.71 for Study 1/2/3
McCrostie, 2007	L1	48 words (half selected from Ringeling's study)	Ringeling (1984) and BNC frequency counts	MRO	NA	.83/.84 for English instructors/college students (overall)

Note. MRO = multiple rank orders. ME = magnitude estimation. NA = not applicable. L1 = L1 speakers; L2 = L2 speakers.

Mixed evidence



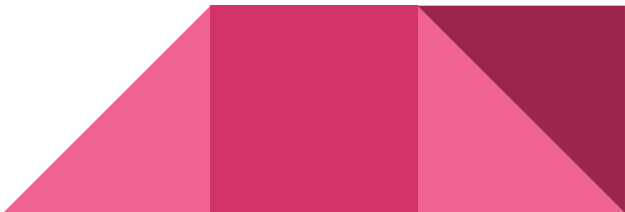
Methodological differences

Research: Phrasal frequency intuition

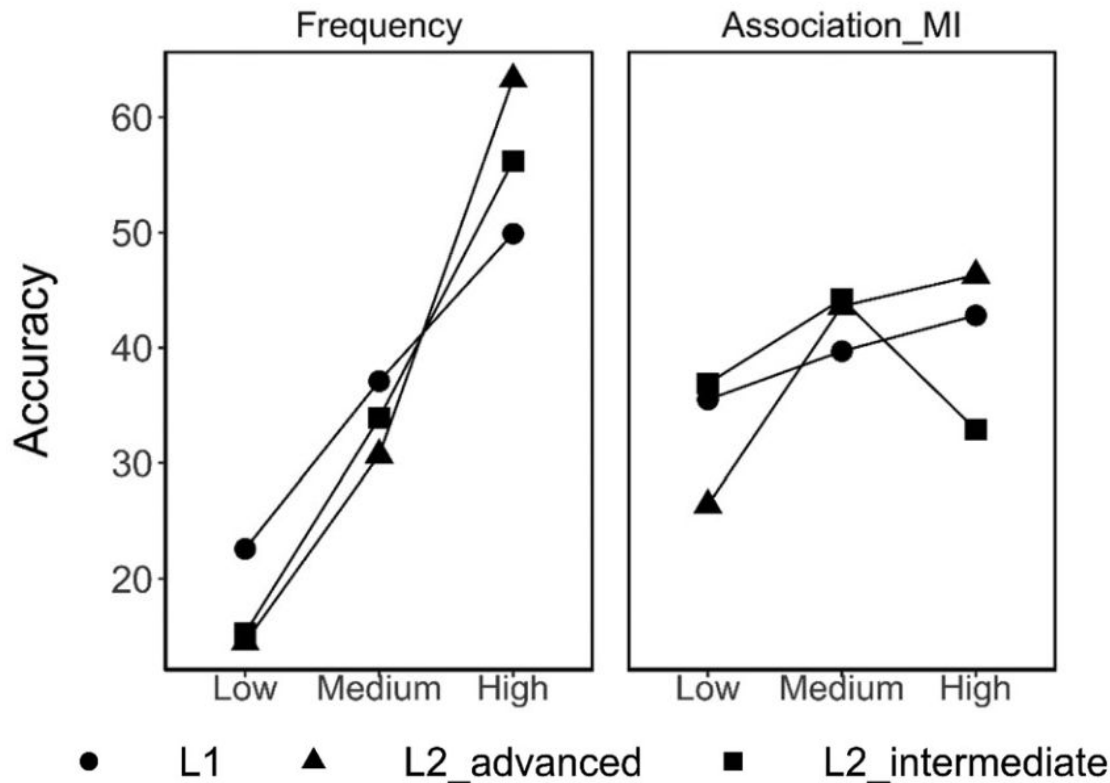
Siyanova & Schmitt (2008)

- L1 and L2 speakers' judgments on 62 English adjective-noun collocations, categorized into **high**, **medium**, and **low** frequency
- $r = .58$ for L1 speakers and $.44$ for L2 speakers
- L1 speakers: **high** vs. **medium** vs. **low** L2 speakers: **high** vs. **low**

Siyanova-Chanturia & Spina (2015)

- L1 and L2 speakers' judgments on 80 Italian adjective-noun collocations, categorized into **high**, **medium**, **low**, and **very low** frequency
 - L1 & L2 speakers only accurate on **high** vs. **very low**
- 

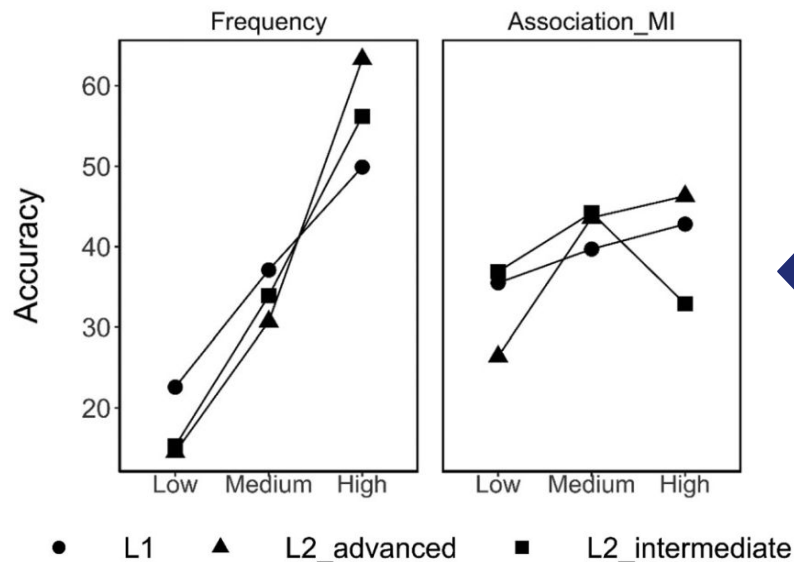
Yi, Man, & Maie (2023): frequency & association strength



Linguistic influences

- word1 length
 - word1 / word2 orthographic neighborhood size
 - word2 phonological neighborhood size
 - word1 / word2 frequency
- positive - negative

The current study - reanalysis of Yi et al. (2023)



- Age of onset
- Length of residence
- English use
- L2 proficiency

Participants: 74 L2 speakers of English

- Living in the U.S.
- Chinese international students (undergraduate)

Materials

Here are the materials!

<https://osf.io/fta9c>



180 adjective-noun collocations (Yi, 2018)

- collocations = occurring at least once per million words in the BNC and with mutual information (MI) higher than 3.0.



	Frequency	MI
High	2.42 (1.98-3.78)	9.51 (8.18-12.71)
Medium	1.28 (1.07-1.89)	7.20 (6.17-8.10)
Low	0.33 (0.07-0.66)	5.10 (3.36-6.15)

Phrasal frequency

	Low frequency	Medium frequency	High frequency
old age	○	○	○
current situation	○	○	○
wild animals	○	○	○
basic skills	○	○	○
economic reform	○	○	○
advisory committee	○	○	○
various kinds	○	○	○
strong feeling	○	○	○
efficient way	○	○	○
extra cost	○	○	○
good thing	○	○	○
European countries	○	○	○
full year	○	○	○
new year	○	○	○
small number	○	○	○
good condition	○	○	○
good practice	○	○	○
musical instruments	○	○	○
different ways	○	○	○
varying degrees	○	○	○

Low

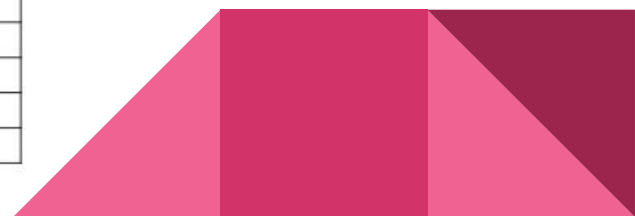
Not common

Medium

Frequent

High

Very frequent



Association strength

	Loose association	Medium association	Strong association
old age	○	○	○
current situation	○	○	○
wild animals	○	○	○
basic skills	○	○	○
economic reform	○	○	○
advisory committee	○	○	○
various kinds	○	○	○
strong feeling	○	○	○
efficient way	○	○	○
extra cost	○	○	○
good thing	○	○	○
European countries	○	○	○
full year	○	○	○
new year	○	○	○
small number	○	○	○
good condition	○	○	○
good practice	○	○	○
musical instruments	○	○	○
different ways	○	○	○
varying degrees	○	○	○

Low

One can hardly predict the other

Medium

One can predict the other to some degree

High

One can strongly predict the other



Demographic variables

☐ 4 How long have you been staying in the U.S.? _____ (number of months)

☐ 5 You started learning English at the age of (e.g., 10): _____

☐ 6 How often do you use English outside the classroom?

	never	rarely	sometimes	often	always
Frequency of English use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7 Rate your English proficiency level using the following scale.

	beginner	intermediate	advanced	very advanced	near-native
Listening	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speaking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reading	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Writing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

☐ 8 Your most recent TOEFL iBT total score is: _____

Regression Analysis

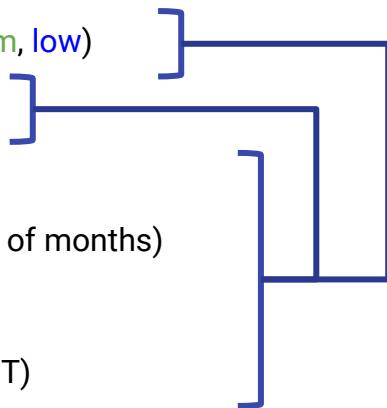
Generalized linear mixed model (binomial)

Dependent variable:

- correct (1) or incorrect (0)

Independent variable:

- frequency bands (high, medium, low)
- MI bands (high, medium, low)
- age of onset (z-scores of years)
- length of residence (z-scores of months)
- English use (z-scores of 1-5)
- Proficiency (z-scores of TOEFL iBT)



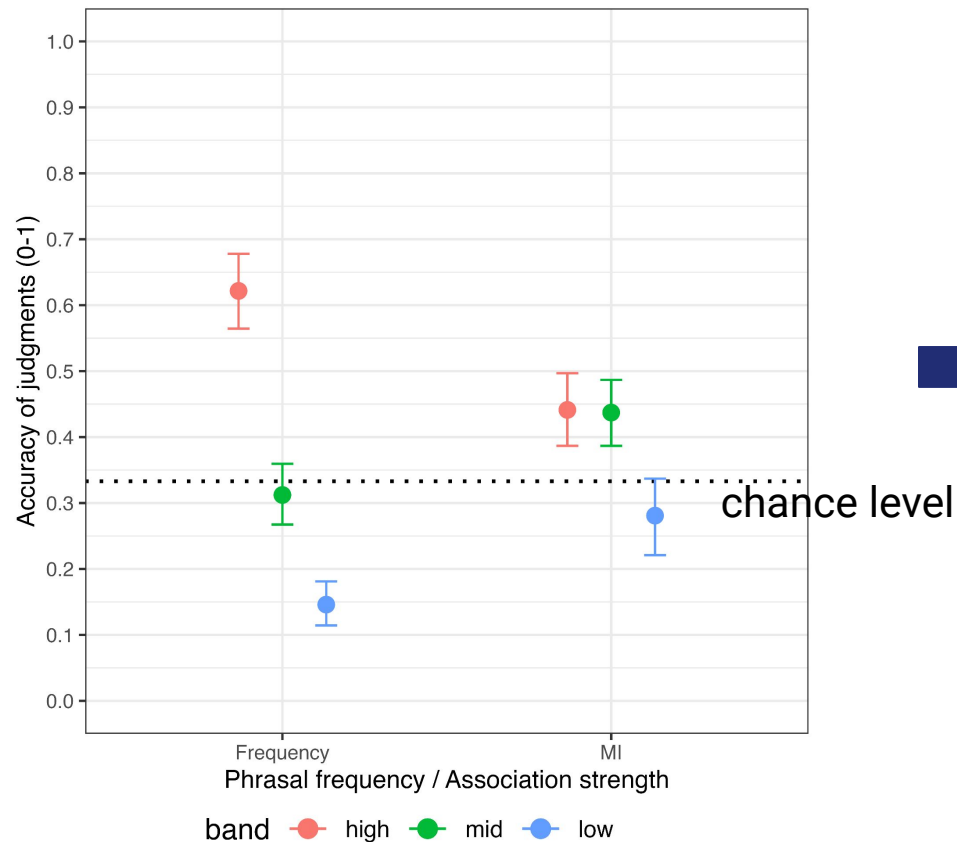
two-way
interactions

brms package

- 4 MCMC chains
- 10,000 iterations
- point estimates
- 95% credible intervals
- posterior probability



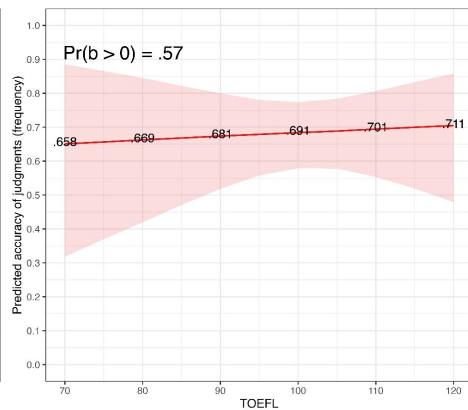
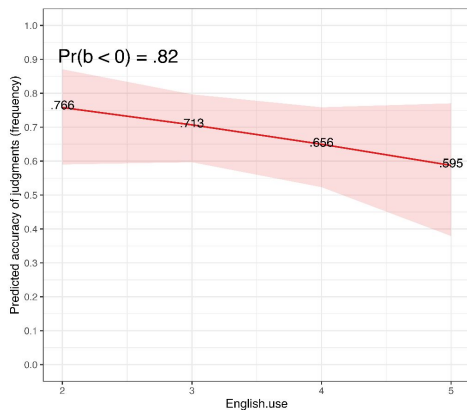
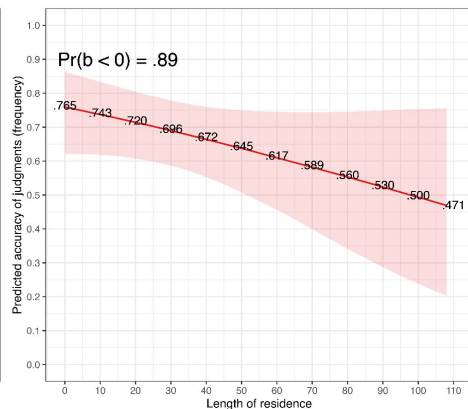
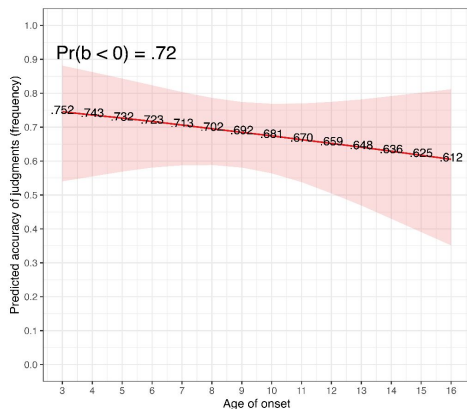
Results: Descriptive statistics



Findings

- Overall not highly accurate
- Above chance only on **high**-frequency collocations
- Above chance on **high**- and **medium**-frequency collocations

Regression Analysis - Phrasal frequency



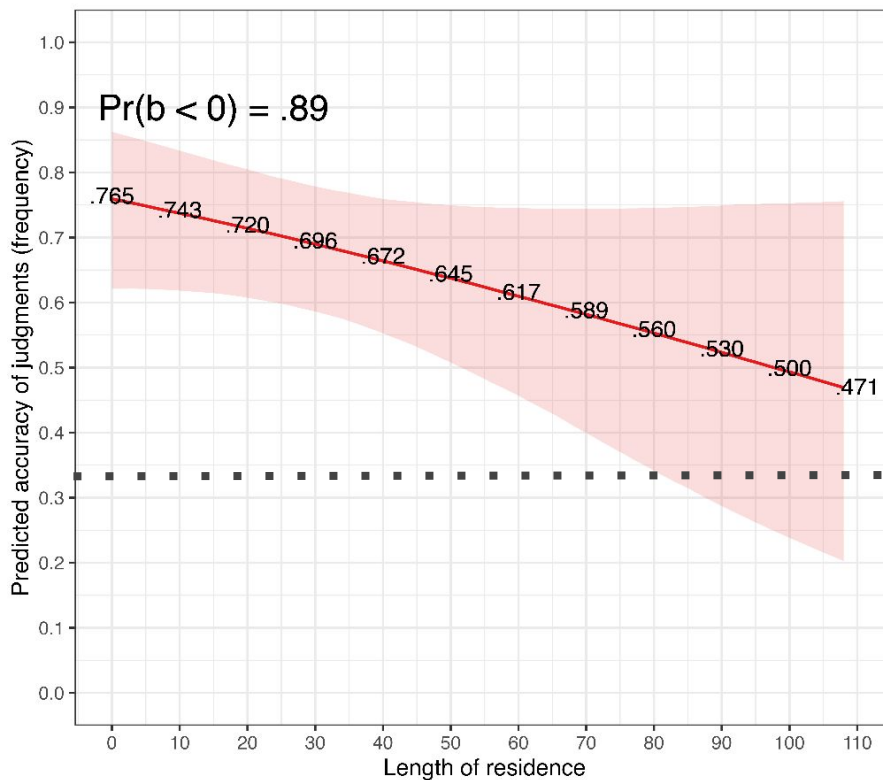
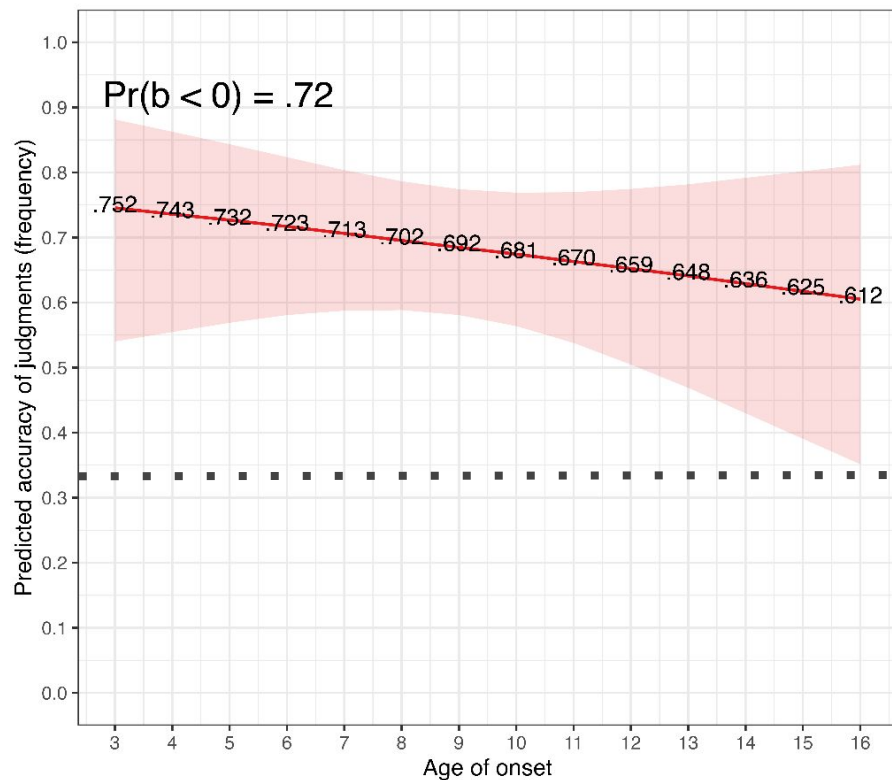
Age of onset: $\Pr(b < 0) = .72$

Length of residence: $\Pr(b < 0) = .89$

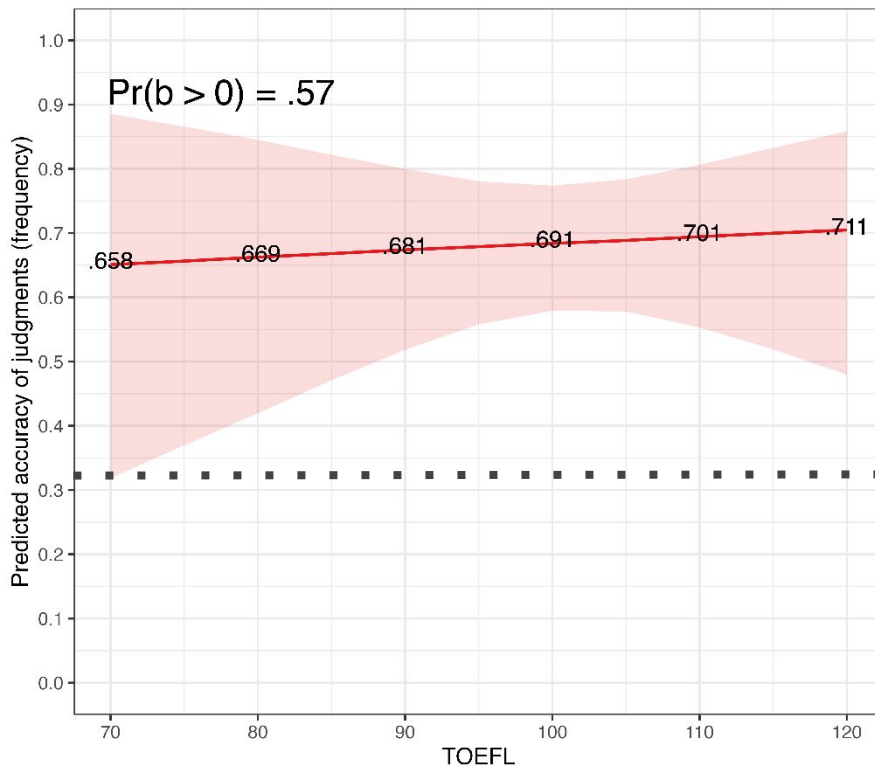
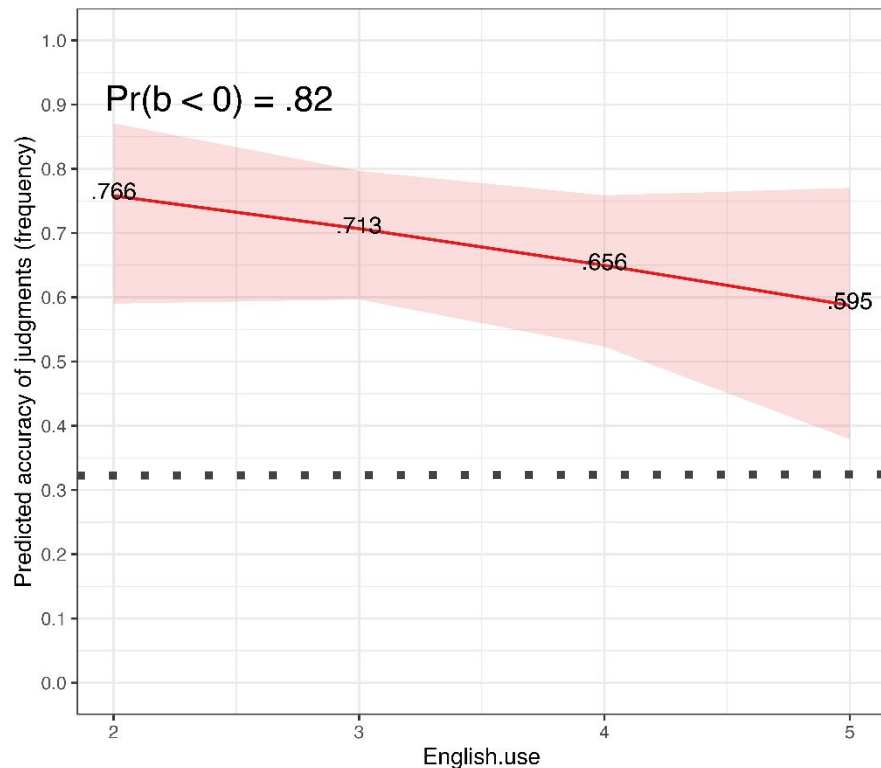
English use: $\Pr(b < 0) = .82$

L2 proficiency: $\Pr(b > 0) = .57$

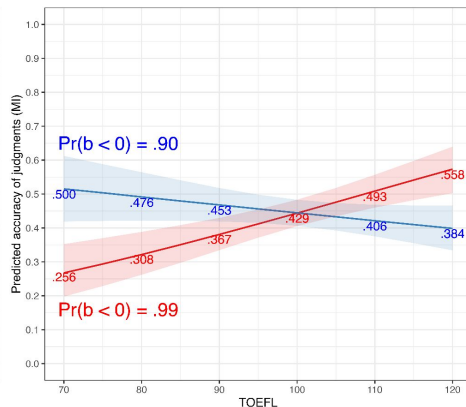
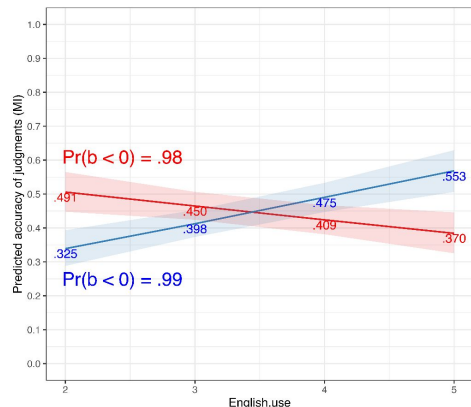
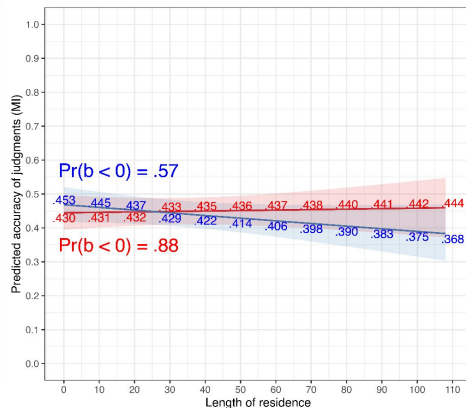
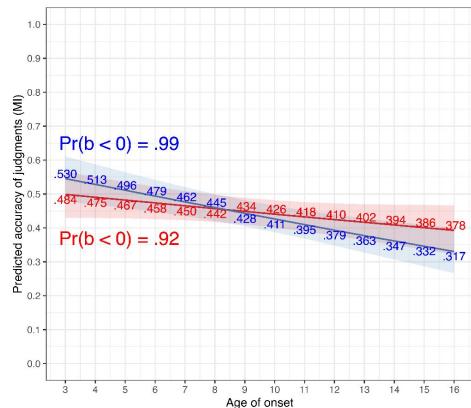
Regression Analysis - Phrasal frequency



Regression Analysis - Phrasal frequency



Regression Analysis - Association strength



High

Medium

Age of onset:

$Pr(b < 0) = .92$ $Pr(b < 0) = .99$

Length of residence:

$Pr(b > 0) = .88$ $Pr(b < 0) = .57$

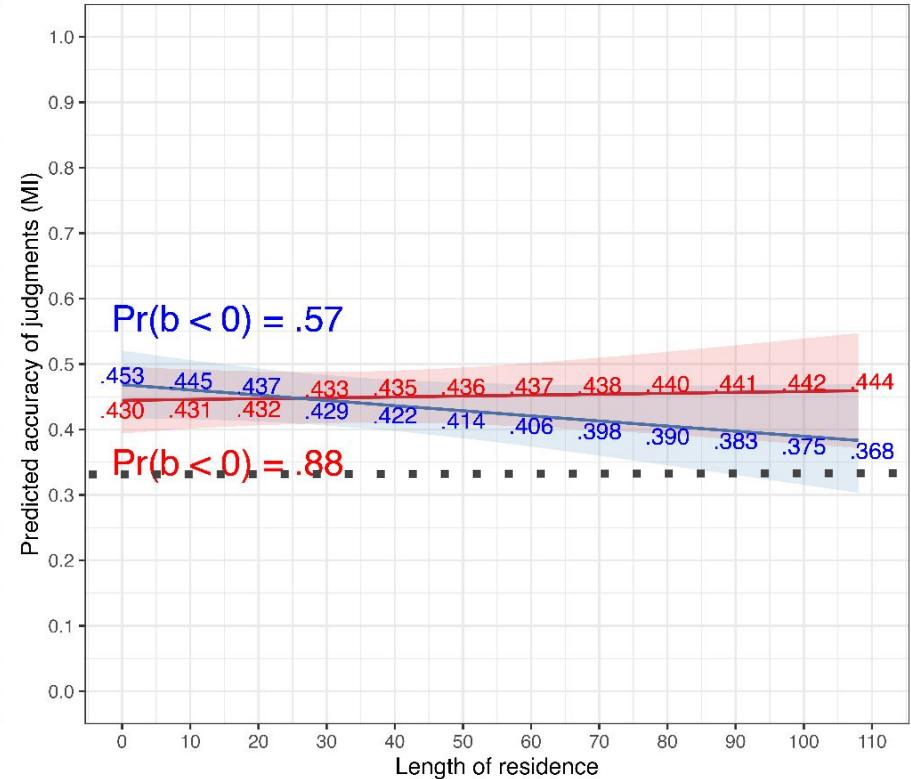
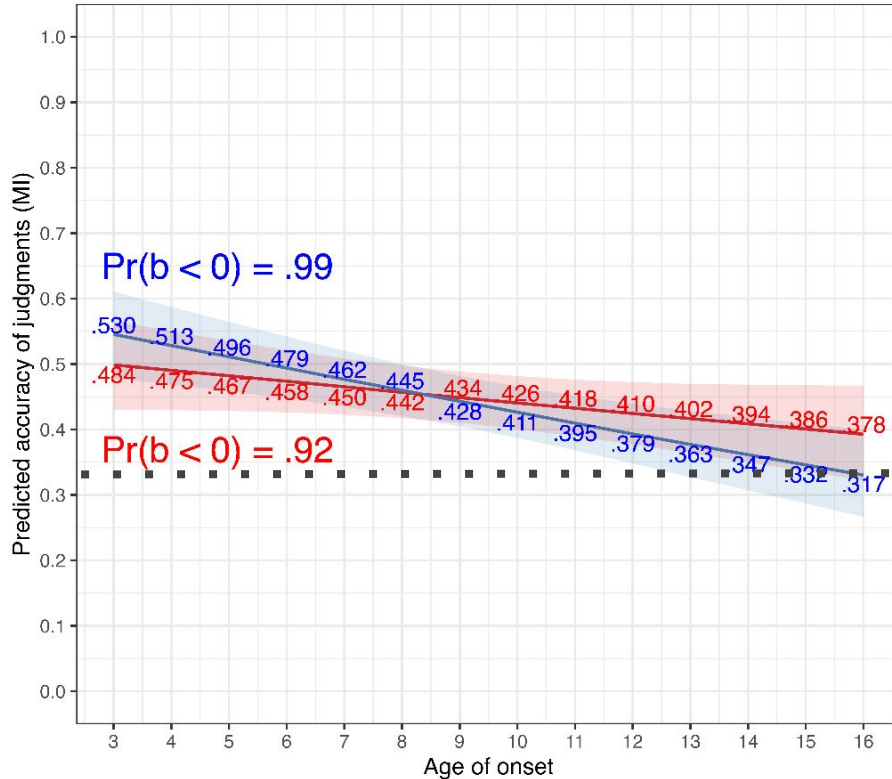
English use:

$Pr(b < 0) = .98$ $Pr(b > 0) = .99$

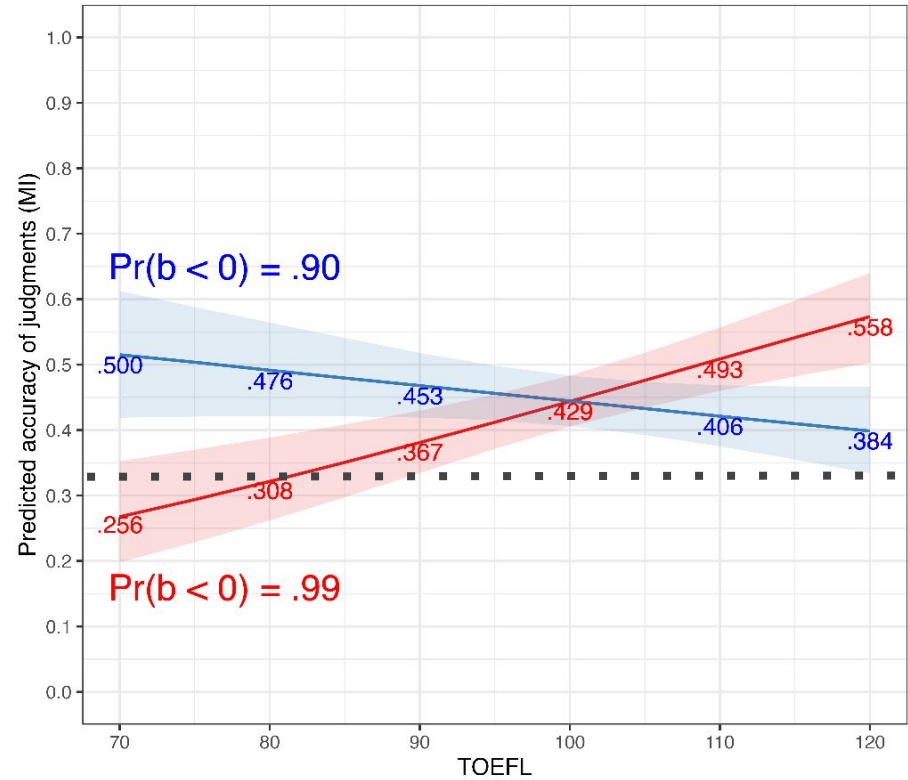
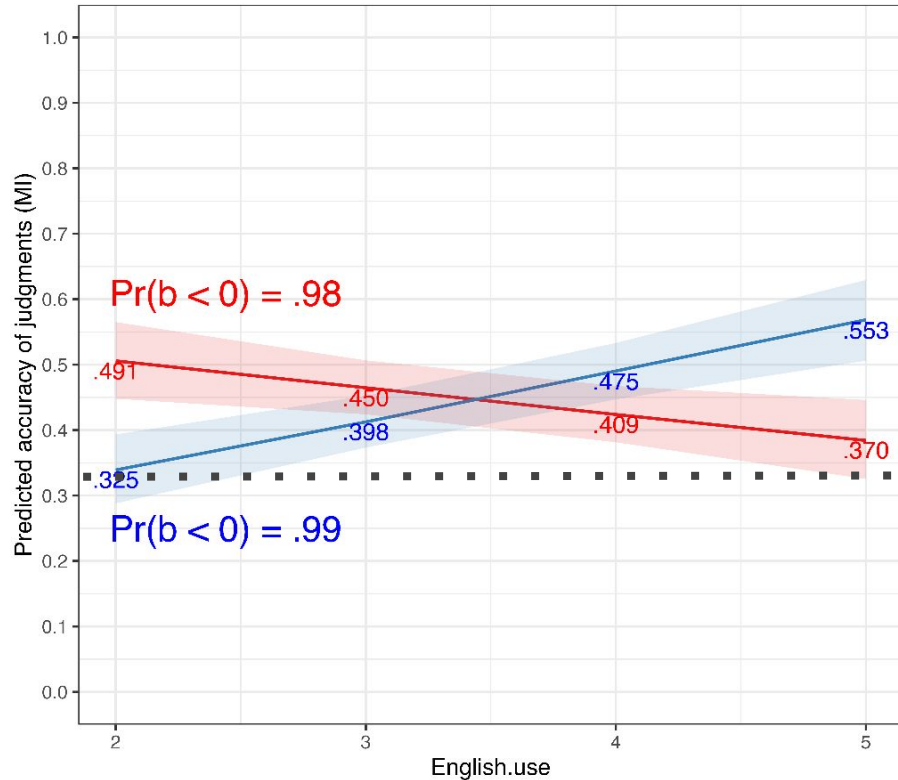
L2 proficiency:

$Pr(b > 0) = .99$ $Pr(b < 0) = .90$

Regression Analysis - Association strength



Regression Analysis - Association strength



Discussion

Accuracy

- Neither L1 nor L2 speakers were accurate on judging phrasal frequency and association strength - even for **highly** frequent or associated collocations!

Individual differences

- No predictor seemed to reliably predict IDs in phrasal frequency intuition
- (a) Age of onset, (b) English use, (c) L2 proficiency predicted IDs in association strength intuition



Discussion - Association strength

Age of onset

- Age effects even for intuiting association strength (**high**- and **medium**-bands)
- When starting at 16 years of age, you may not judge at the above-chance level
- Previous findings for age effects on L2 collocations may be related to aspects concerning learning association strength (Granena & Long, 2012; Yi, 2017)


English use

- Negative predictor for **high** association strength Positive predictor for **medium** association strength

L2 proficiency

- Positive predictor for **high** association strength
- Intuition can improve with proficiency at least for highly associated collocations

Limitations and future directions

- Sample size ($n = 74$)
 - Bigger sample size + replications necessary
 - Finer-grained and systematic measurements of L2 use
 - We used a 5-point scale - what are alternatives?
 - Participant majors
 - We do not know about their majors
 - Previous studies compared people from different disciplines (e.g., linguists vs. non-linguists)
 - Not genre specific
 - Language knowledge is genre-specific and so should statistical intuition!
 - Aptitude measures
 - Statistical learning ability, language analytic ability, etc.
- 

Thank you!

Please post any questions and comments here:

<https://app.learnwiz.one/events/4638852633>



Slides are available here:



Correlations among demographic variables

