



 OPEN ACCESS



Vocabulary Learning and Instruction

ISSN 2981-9954

Volume 13, Number 1 (2024)

<https://doi.org/10.29140/vli.v13n1.1311>



Castledown

Individual Differences in L2 Speaker Intuitions of Phrasal Frequency and Association Strength of Multiword Sequences

Ryo Maie 

Tohoku University, Japan
ryo.maie.e5@tohoku.ac.jp

Wei Yi 

Peking University, China
weiyisla@pku.edu.cn

Abstract

This brief report presents the results of a re-analysis of data by Yi, Man, and Maie (2023), who investigated L1 and L2 intuitive knowledge of phrasal frequency and collocation strength in multiword sequences. We utilized an individual-differences approach and examined which participant variables (age of onset, length of residence, language use, and L2 proficiency) predicted the participants' accuracy in judging the phrasal frequency and association strength of multiword sequences in English. We found that the demographic variables were only related to the accuracy in judging association strength, but those variables differentially predicted the accuracy depending on whether the collocations were of high or medium association strength.

Keywords: multiword sequence, phrasal frequency, association strength, collocation.

Literature Review

Humans display a wide range of abilities in accumulating statistical data from everyday experiences and making decisions based on their intuitive knowledge (e.g., Kahneman, 2003; Tversky & Kahneman, 1974). Usage-based approaches claim

Copyright: © 2024 Ryo Maie and Wei Yi. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within this paper.

that language learning is primarily driven by the accumulation of statistical data, such as the frequency of linguistic units, and making abstractions of regularities within the data (Ellis & Ogden, 2017). The frequency of occurrence plays a fundamental role in determining the course of language development. Ellis (2002) describes how language learners are sensitive to frequency at every level of linguistic processing, with much empirical evidence showing that this is especially the case in second language (L2) learning (e.g., Ellis, Römer, & O'Donnell, 2016; Yi, 2018; Yi et al., 2017).

Coupled with research on frequency, many researchers have examined whether language users can accurately intuit the frequency of linguistic units (e.g., how often a single word occurs in everyday use). This line of research holds both theoretical and practical values. Theoretically, investigating language frequency intuition can contribute to our understanding of human memory (Zacks & Hashler, 2002) and the decision-making process (Kahneman, 2003; Tversky & Kahneman, 1974). From a practical perspective, if language users can accurately estimate the frequency of linguistic units, such intuition can eliminate the need for language teachers to reference corpus data when making judgments about which materials should be prioritized for teaching (McCrostie, 2007).

Existing research on language intuition has primarily focused on word frequency (see Yi et al., 2023, Appendix S1 for a summary). However, the results of this research are far from being conclusive. While initial studies reported very high correlations between objective counts of word frequency retrieved from corpora and the participants' subjective judgments of the word frequency (e.g., Balota et al., 2001; Carroll, 1971; Shapiro, 1969), more recent research failed to find evidence for language users' intuition of word frequency (Alderson, 2007; Schmitt & Dunham, 1999). The contrasting findings mainly result from the methodological differences in how the accuracy of intuition was operationalized. For instance, Alderson (2007) claimed that evidence must come through pure judgment accuracy (e.g., whether one's subjective estimation of words being in high-, medium- and low-frequency range matches objective counts) rather than correlations between subjective and objective counts.

Language comprises units of different scales, ranging from single words to multiword sequences, and the use of multiword sequences significantly contributes to learners attaining native-like proficiency (Erman & Warren, 2000). For multiword sequences, two kinds of statistical data play a vital role: phrasal frequency and association strength (Ellis, Römer, & O'Donnell, 2016). Phrasal frequency refers to how often one encounters specific word combinations, whereas association strength refers to the co-occurrence probability of the constituent words. Though few in number, researchers have examined whether (and to what extent) language users can accurately estimate the frequency of multiword sequences. For instance, Siyanova and Schmitt (2008) examined L1 and L2 speakers' judgments on the frequency of 62 adjective-noun collocations, categorized into high-, medium-, and low-frequency bands. While the researchers found only moderate correlations between the participants' subjective judgments and the objective data from corpora (0.58 and 0.44 for L1 and L2 speakers, respectively), they also found that while L1 speakers were able to distinguish high-frequency collocations from those in medium- and low-frequency bands, L2 speakers could only distinguish between high- and low-frequency collocations. Siyanova-Chanturia and Spina (2015) conducted a similar study with L1 and L2 speakers of

Italian judging 80 noun-adjective collocations in one of four categories: high frequency, medium frequency, low frequency, and very low frequency. They found that while both L1 and L2 speakers made accurate judgments on collocations of high and very low frequency, neither of the groups showed accurate intuitions in medium- and low-frequency bands, concluding that one's ability to make accurate judgments on phrasal frequency likely depends on the frequency range in question.

Yi et al. (2023) followed up the preceding research on phrasal frequency intuition with L1 and L2 English speakers and additionally asked the participants to make judgments on association strength. They contended that phrasal frequency and association strength are inherently interconnected and hence the effect of association strength must be controlled when one investigates participants' intuition on phrasal frequency (or vice versa). Furthermore, Yi et al. examined how linguistic variables, such as word length and concreteness, influenced participants' judgments of phrasal frequency and association strength. They found that neither L1 nor L2 speakers demonstrated accurate intuitive judgments of phrasal frequency and association strength regardless of frequency bands (high, medium, and low frequency), but the participants' judgments were indeed susceptible to the impact of linguistic characteristics such as word length, concreteness, and phonological and orthographic neighborhood size.

In this report, we re-analyzed the data from Yi et al. (2023) and examined individual differences in the participants' intuitive knowledge about phrasal frequency and association strength in English. In particular, we were interested in how four demographic variables, age of onset, length of residence, language use, and L2 proficiency, were associated with the participants' judgment accuracy. These four variables are known to influence the acquisition of L2 collocation (Granena & Long, 2013; Lundell et al., 2023). Assuming that L2 users draw on their underlying statistical knowledge when making intuitive judgments, these variables should also affect how well one can intuit phrasal frequency and association strength of L2 collocations.

Methods

Participants

We focused on 74 L2 speakers of English, whose age of onset, length of residence, English use, and L2 proficiency were available. They were Chinese international students studying in undergraduate programs in the United States. Table 1 shows the descriptive statistics of the demographic variables for the participants. In our study, age of onset referred to the age at which participants first started learning English, while length of residence pertained to how long the participants had stayed in the U.S. (in months). English use was operationalized as the frequency of participants' extracurricular English use on a scale from 1 (never) to 5 (always), and L2 proficiency was gauged through participants' self-reported scores on the TOEFL iBT.

Stimuli and Procedure

We used 180 English adjective-noun collocations adopted from Yi (2018). We defined collocations as occurring at least once per million words in the British National Corpus (BNC) and the association between a constituent adjective and noun measured by

Table 1 *Descriptive Statistics on the Four Demographic Variables*

	Mean	SD	Min–Max
Age of onset	9.01	2.69	3–16
Length of residence (months)	32.02	23.97	0–108
English use (1–5)	3.39	0.82	2–5
L2 proficiency (TOEFL iBT)	100.75	8.34	70–119

Note. For English use, 1 (never), 2 (rarely), 3 (sometimes), 4 (often), and 5 (always).

Table 2 *Mean Phrasal Frequencies and Association Strengths of the Adjective–Noun Collocations*

	Phrasal frequency			Association strength		
	Mean	SD	Min–Max	Mean	SD	Min–Max
High	2.42	0.42	1.98–3.78	9.51	1.19	8.18–12.71
Medium	1.28	0.15	1.07–1.89	7.20	0.60	6.17–8.10
Low	0.33	0.17	0.07–0.66	5.10	0.83	3.36–6.15

mutual information (MI) being higher than 3.0. We retrieved phrasal frequency of the collocations from the BNC and computed MI values. [1] L2 speakers (who were not included in the study) rated their familiarity with the collocations on a scale of 1 (totally unknown) to 5 (extremely familiar) and the mean rating was 4.5 ($SD = 0.4$). We rank-ordered the collocations from the lowest to the highest phrasal frequency or association strength and grouped them into three bands: high, medium, or low phrasal frequency/association strength. Table 2 summarizes the mean frequency and association strength values for the collocations.

The participants judged phrasal frequency and association strength through an online questionnaire consisting of three sections. The first section required participants to provide demographic information. In the second section, participants rated phrasal frequency on a 3-point scale: high, medium, and low frequency. Similarly, the third section prompted participants to judge association strength on a 3-point scale: high, medium, and low association strength. Phrasal frequency was explained to participants as how often each collocation occurs in English, while association strength was explained as how likely the constituent words can predict the appearance of one word given the other word, regardless of the direction of prediction. Although no time limit was set for each item, participants took approximately 25 minutes on average to complete the questionnaire. See <https://osf.io/fta9c> for the entire questionnaire.

Analysis

Following Alderson (2007), we operationalized the accuracy of judgments as binary scores of whether the participants' subjective estimations (high, medium, low) matched

the corpus-based categories. We analyzed the effects of the demographic variables using generalized linear mixed models with Bayesian inference. Since the accuracy scores were binary in nature (correct or incorrect), the models were binomial models, which regressed the probability of correct responses on the following predictor variables: phrasal frequency bands (high, medium, low), MI bands (high, medium, low), age of onset (z-score), length of residence (z-score), English use (z-score), and TOEFL (z-score). Because we were interested in the effects of the demographic variables in different frequency bands, we additionally incorporated the two-way interactions of the phrasal frequency or MI bands with each demographic variable. When modeling accuracy on phrasal frequency, we also entered the MI bands to ensure that we account for the intertwined nature of phrasal frequency and association strength. This meant that we included the phrasal frequency bands when modeling accuracy on judging association strength. Note that both models (for phrasal frequency and association strength) incorporated varying intercepts for individual participants and items as random effects. [2]

We estimated model parameters through the Markov chain Monte Carlo simulation with 4 chains of 10,000 iterations each. The first 1,000 iterations were discarded as warm-up periods. We used the R-package brms (version 2.19.0; Bürkner, 2019) to estimate the parameters, and emmeans (version 1.8.7; Lenth, 2023) and sjPlot (version 2.8.14; Lüdtke, 2023) to extract and plot model-based accuracy rates as a function of phrasal frequency or MI bands, age of onset, length of residence, English use, and L2 proficiency (TOEFL iBT).

Results

Participants' accuracy rates on judging phrasal frequency and association strength are summarized in Figure 1. Note that chance performance in the task corresponded to 0.33 (33%) due to the 3-point scale of the judgment task. As Yi et al. (2023) reported, the participants did not show accurate intuitive judgments of phrasal frequency and association strength. However, their judgments were (at least) above chance for collocations of high frequency ($M = .62$ [.56, .68]) and high ($M = .44$ [.38, .50]) and medium ($M = .43$ [.38, .49]) association strength. In the following regression analysis, we thus focused on these frequency and association bands. Our assumption was that depending on their demographic characteristics (i.e., age of onset, length of residence, English use, and L2 proficiency), the participants became more or less adept at intuitively estimating the frequency or association strength of the collocations because those characteristics affect the acquisition of statistical knowledge underlying L2 collocations, which in turn influences how well one can intuit the frequency or association strength.

Regression Analysis

Figure 2 displays the model-based prediction of accuracy rates for phrasal frequency judgments based on demographic variables. Again, the result only pertains to the high-frequency band as the participants did not perform above chance on the medium- or low-frequency collocations. The figure also shows the posterior probability

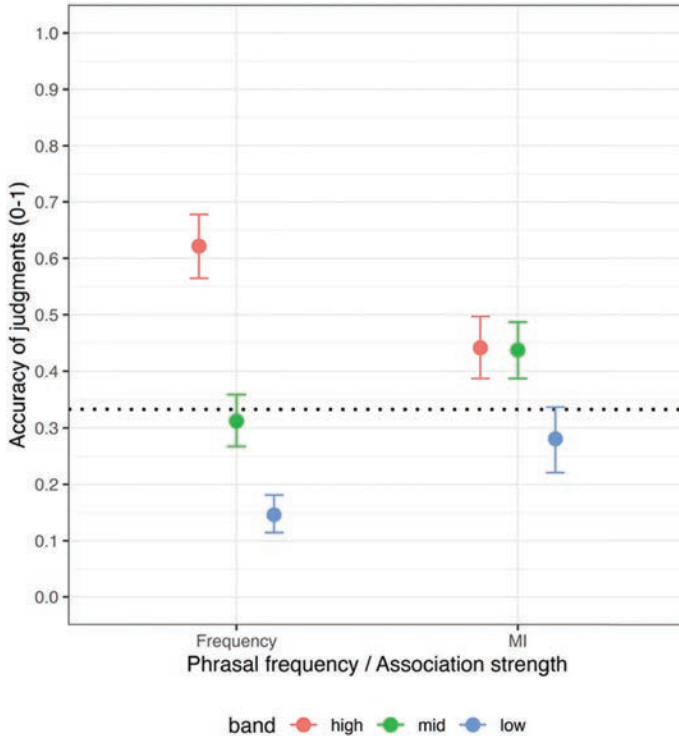


Figure 1 Accuracy rates on judging phrasal frequency and association strength.
Note. The error bars show 95% confidence intervals.

of whether the slope for each variable is greater or smaller than 0. Following Yi et al. (2023), we interpret effects as reliable if the posterior probability for the variables equals or exceeds .95.

For high-frequency collocations, none of the demographic variables appeared to consistently predict individual differences in participants' intuitive knowledge of phrasal frequency. The predictor with the largest effect size was length of residence, yet surprisingly, it displayed a negative correlation with accuracy rates. Note that even for this variable, the effect size was negligible, as a 10-month increase in length of residence was associated with only a 2.2% decrease in accuracy. Overall, there remains scientific uncertainty surrounding the impact of demographic variables on the accuracy of phrasal frequency intuition.

Figure 3 shows the predicted accuracy rates for association strength judgments as a function of the participants' individual differences in the demographic variables and the association strength bands (high and medium). Note that for association strength, we focus on both high (red lines) and medium bands (blue lines). Age of onset, English use, and L2 proficiency (TOEFL) seemed to be associated with the accuracy rates, but the presence or the direction of the effect depended on whether the collocations were of high or medium association strength. First, age of onset was negatively predictive

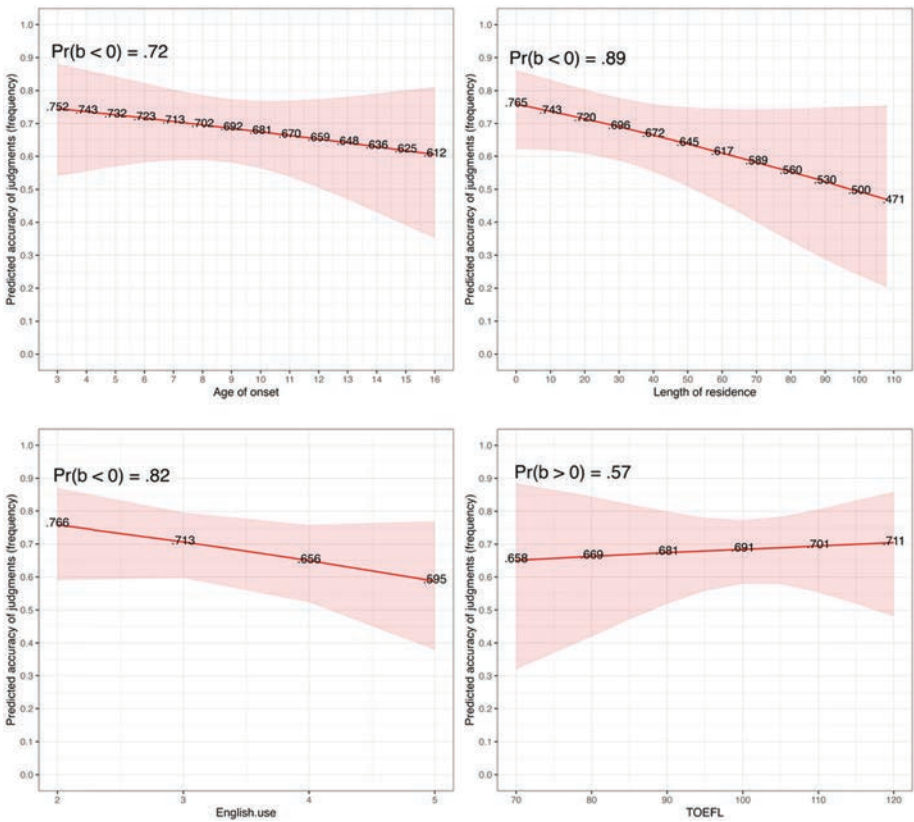


Figure 2 Predicted accuracy for accuracy on phrasal frequency in the high-frequency band. Note. The error bands show 95% credible intervals.

of accuracy for the collocations of medium association strength, which indicated the presence of age effects on how accurately one can estimate the association of the constituent words. One year increase in age of onset was associated with a 1.7% decrease in accuracy. When one begins learning English at the age of 16, one may not even be able to judge collocations of medium frequency at the above-chance level. Note that there is also a trend of age of onset negatively impacting judgment accuracy for highly associated collocations, but the estimate seems less certain probably due to a smaller effect size on these collocations.

English use was the only variable that was consistently linked to accuracy rates across the high and medium association strength bands, but the direction of prediction was different for the two bands. For highly associated collocations, English use was negatively associated with accuracy, while the relationship was positive for collocations of medium association strength. It remains unclear why English use, which operationalized the frequency of English use outside the classroom, manifested as a negative predictor of accuracy. For instance, a shift in English use from a rating of 2 (rarely) to

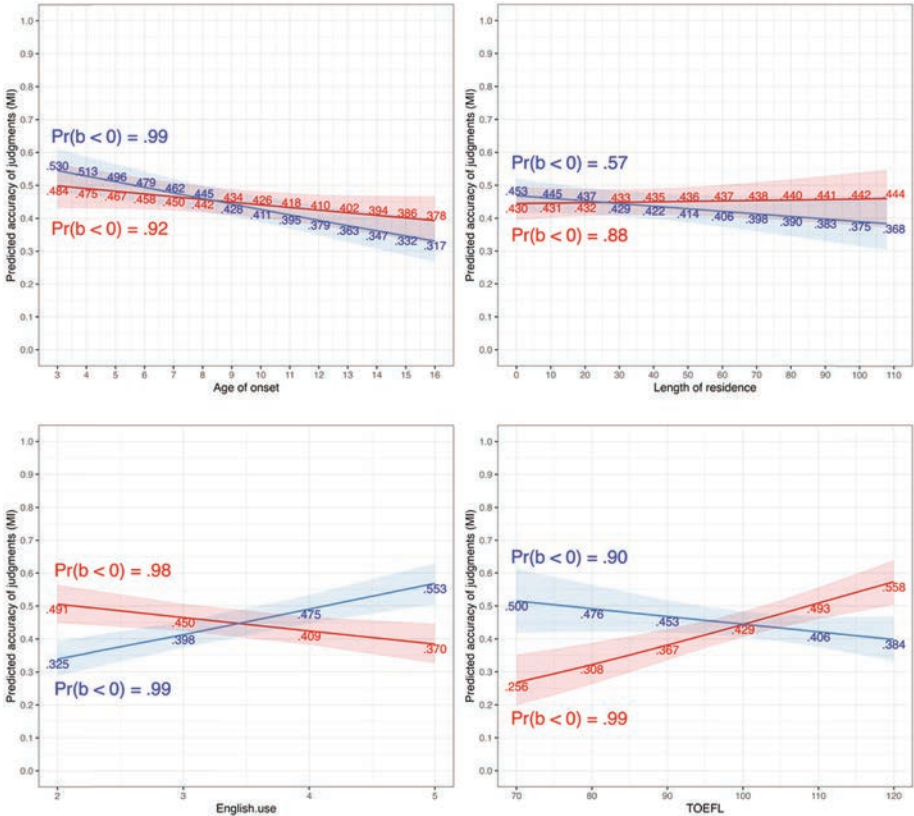


Figure 3 Predicted accuracy for accuracy on association strength in the high- and medium-frequency bands.

Note. The error bands show 95% credible intervals.

3 (sometimes) corresponded to a 4.1% decline in accuracy. When participants indicated that they always used English outside the classroom, accuracy dropped to 37%, which barely exceeded the chance rate. Conversely, for collocations of medium association strength, we found an anticipated positive association between English use and accuracy rates. When participants reported that they rarely used English outside the classroom, their performance did not exceed chance levels. However, consistent use of English helps learners achieve as high as 55.3%.

Lastly, L2 proficiency showed a positive relationship with accuracy, although this effect was limited to collocations within the medium association band. At TOEFL scores of 70 and 80, the participants did not exhibit performance above chance levels (with accuracy rates of .256 and .308, respectively). To accurately judge collocations of medium association above the chance level, it appeared that one needs a TOEFL score close to 90, which converts to the proficiency level of C1 (advanced) on the CEFR scale. [3] Note that there was also a trend that L2 proficiency may be negatively related

to accuracy on highly associated collocations, but a 10-point increase in TOEFL scores was only associated with a 2.4% decrease in accuracy, which is negligible as an effect size.

Discussion and Conclusions

In this report, we investigated individual differences in L2 English users' ability to intuitively estimate phrasal frequency and association strength of adjective-noun collocations. We focused on four demographic variables, age of onset, length of residence, English use, and L2 proficiency, as they relate to the acquisition of L2 collocations. As reported in Yi et al. (2023), neither L1 nor L2 speakers accurately judged the target L2 collocations. However, their judgements were above the chance level for collocations of high frequency or high and medium association strength.

Our individual differences analysis revealed that the demographic variables were only associated with the participants' accuracy in judging association strength. Furthermore, these variables differentially predicted accuracy depending on whether the collocations had high or medium association strength. For instance, age of onset negatively predicted accuracy for items within the medium-association band. This finding, coupled with the weaker effect for collocations in the high-association band, suggests the presence of age effects in developing the ability to judge collocations of medium association strength. However, the same process does not constrain the development of intuitions for highly associated collocations. This interpretation is further supported by the fact that L2 proficiency was positively associated with accuracy in highly associated collocations only, which suggests that one's ability can at least improve as one's overall proficiency increases and as long as the collocations are of high association strength. The correlation between age of onset and accuracy rates on association strength also implies that previous findings for age effects on the acquisition of L2 collocations may be due to a decline in an individual's ability to learn statistical data related to association strength rather than phrasal frequency.

Finally, the frequency of English use outside the classroom emerged as a positive predictor of accuracy in judging medium-association collocations. This suggests that participants who engaged with English more frequently in extracurricular contexts tended to exhibit higher accuracy levels. Surprisingly, we encountered a reversed negative relationship concerning high-association collocations, which, to the best of our knowledge, lacks a clear explanation. This finding implies that participants who utilized English more frequently outside the classroom demonstrated a comparatively poorer ability to judge highly associated collocations. Our present analysis was additional to the original investigation conducted by Yi et al. (2023). Consequently, further empirical experimentation is essential to determine whether this effect, along with the other anticipated findings, genuinely relates to the actual knowledge or process underlying the judgment of association strength in L2 collocations.

Endnote

[1] Although Yi et al. (2023) additionally analyzed logDICE as a complimentary measure of association strength, we restricted our focus to MI because accuracy

- rates on judging association strength analyzed with those two measures were highly correlated: $r = .97$ [.96, .98].
- [2] Note that because the participants only performed above chance for collocations of high frequency and high and medium association strength, we only focused on these frequency or association strength bands in the regression analysis. Hence, the frequency model was fit to a dataset that only contained the participants responses on the high-frequency collocations, whereas the MI model was fit to a dataset that contained responses on the high- and medium-MI bands only.
- [3] The conversion table for TOEFL iBT scores and the CEFR scale scores can be found at: <https://www.ets.org/toefl/score-users/ibt/compare-scores.html>.

Correspondence concerning this article should be addressed to Wei Yi, *Peking University, China*. Email: weiyisla@pku.edu.cn

References

- Alderson, J. C. (2007). Judging the frequency of English words. *Applied Linguistics*, 28(3), 383–409. <https://doi.org/10.1093/applin/amm024>
- Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, 29(4), 639–647. <https://doi.org/10.3758/BF03200465>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Carroll, J. B. (1971). Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning and Verbal Behavior*, 10(6), 722–729. [https://doi.org/10.1016/S0022-5371\(71\)80081-6](https://doi.org/10.1016/S0022-5371(71)80081-6)
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, N. C., & Ogden, D. C. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science*, 9(3), 604–620. <https://doi.org/10.1111/tops.12256>
- Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Wiley-Blackwell.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, 20(1), 29–62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Granena, G., & Long, M. H. (2012). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(2), 311–343. <https://doi.org/10.1177/0267658312461497>
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720. <https://psycnet.apa.org/doi/10.1037/0003-066X.58.9.697>

- Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means*. (R package version 1.8.7.) [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- Lüdecke, D. (2023). *sjPlot: Data visualization for statistics in social science*. (R package version 2.8.14) [Computer software]. <https://CRAN.R-project.org/package=sjPlot>
- Lundell, F. F., Arvidsson, K., & Jemstedt, A. (2023). The importance of psychological and social factors in adult SLA: The case of productive collocation knowledge in L2 Swedish of L1 long-term residents. *Studies in Second Language Acquisition*, 45(2), 558–570. <https://doi.org/10.1017/S0272263122000419>
- McCrostie, J. (2007). Investigating the accuracy of teachers' word frequency intuitions. *Regional Language Centre Journal*, 38(1), 53–66. <https://doi.org/10.1177/0033688206076158>
- Schmitt, N., & Dunham, B. (1999). Exploring native and non-native intuitions of word frequency. *Second Language Research*, 15(4), 389–411. <https://doi.org/10.1191/026765899669633186>
- Shapiro, B. J. (1969). Subjective estimation of relative word frequency. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 248–251. [https://doi.org/10.1016/S0022-5371\(69\)80070-8](https://doi.org/10.1016/S0022-5371(69)80070-8)
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429–458. <https://doi.org/10.3138/cmlr.64.3.429>
- Siyanova-Chanturia, A., & Spino, S. (2015). Investigation of native and second language learner intuition of collection frequency. *Language Learning*, 65(3), 533–562. <https://doi.org/10.1111/lang.12125>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://psycnet.apa.org/doi/10.1126/science.185.4157.1124>
- Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, 40(4), 831–856. <https://doi.org/10.1017/S0272263118000141>
- Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, 33(4), 519–549. <https://doi.org/10.1177/0267658317708009>
- Yi, W., Man, K., & Maie, R. (2023). Investigating L1 and L2 speaker intuitions of phrasal frequency and association strength of multiword sequences. *Language Learning* 73(1), 266–300. <https://doi.org/10.1111/lang.12521>
- Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 21–36). Oxford University Press.