Wesleyan Word Experience Project Preliminary Analysis

Wesleyan University

Shakeel Jessa, Wesleyan University

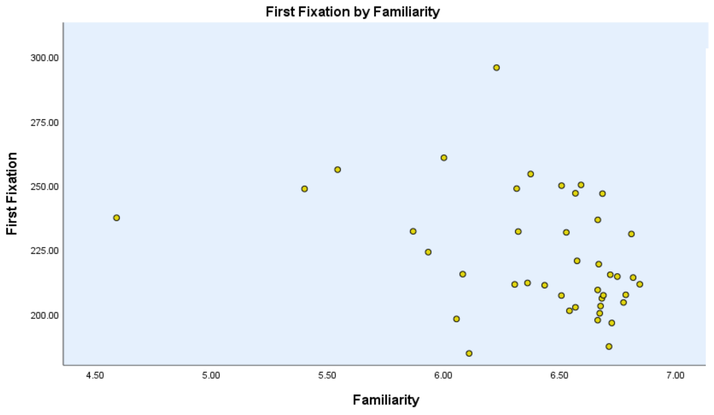
WESLEYAN EYE LAB

Introduction

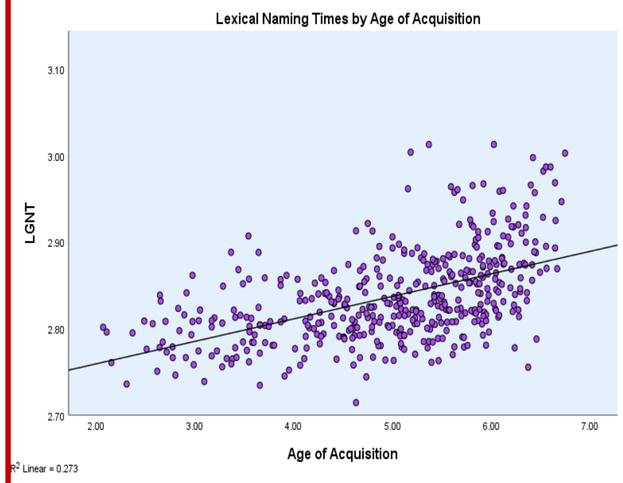
The Wesleyan Eye Movement and Reading lab collected ratings on words related to their age-of-acquisition (AoA) and familiarity using an online survey of Psyc 105 students over the past year. Rated familiarity is a measure of subjective frequency, it indexes the experience that an individual has with a word (Juhasz, 2014). Rated familiarity has an early and long-lasting effect on reading behavior (Juhasz, 2014). AoA refers to a scale about when a word is learned in life and has shown the effect that words learned early in life are processed faster than words learned later in life, it is correlated with word frequency and familiarity. But, its influence on word processing has been found to be distinct from these variables (Juhasz, 2014).This project is known as the Wesleyan Word Experience Project (WWEP). A good portion of the words the words were chosen from the English Lexicon Project (Balota et al., 2007), which is a large online database that includes lexical decision and naming times for these words.. The aim of the project is to track changes in experiences to words over time and how that relates to their processing in standard word recognition tasks. This poster's purpose is twofold, first to add to this project we used the Ghent Eye Movement Corpus (GECO- Cop et al., 2017) to extract eye movement measures for the words used in our rating study. Second, analyses were conducted to explore how the AoA and familiarity ratings collected this year relate to the word recognition time

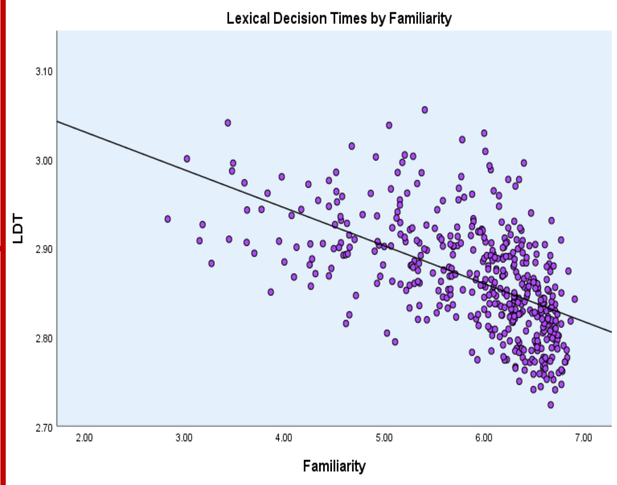
G.E.C.O

The data extracted from the GECO Corpus contains observations from 14 English monolingual undergraduates from the University of Southampton. The participants read the novel *The Mysterious Affair at Styles* by Agatha Christie, with the participants reading a total of 5,031 sentences. The three major variables obtained from this data were the participants first fixation duration, gaze duration, and total reading time for a given word. First fixation duration is defined as the duration (in ms) of the first fixation on a word during the first pass through the text, gaze duration is the sum of all first-pass fixations on a word, and total reading time is the sum of all fixations on a word, including those that occur after regressions. By matching the words with viable data to the WWEP it was possible to conduct linear regression analysis on the 42 matching words. The only significant relationships that were found were with length and word frequency.



English Lexicon Project





Using measures from the English Lexicon project, detailed correlation and linear regression analyses were conducted on the WWEP. The measures tested were Lexical Decision Time, Naming Time, and each of these measures' accuracy. The correlation analysis found a very strong relationship between the variables which is exactly what we had hypothesized and is supported by past research. Linear regression analysis was then performed in a tiered process, beginning with a baseline of length and word frequency and then adding the familiarity and age of acquisition data for the past year of the WWEP. Significance for familiarity and age of acquisition was found. These findings suggest that as familiarity increases LDT and NT times decreases, and accuracy increases. While as AoA increases these times increases and accuracy decreases. This analysis supports Perfetti's (2007) Lexical Quality Hypothesis, words that are lower in *lexical quality* are more difficult to recognize relative to words higher in lexical quality. Lexical quality refers to a variety of factors though the two most vital are Age of Acquisition and Familiarity.

Table 1.

Correlations Between Standardized Lexical Decision Times and Word Naming Times from the English Lexicon Project (Baltoa et al., 2007) and Key Predictors from the Wesleyan Word Experience Project

Predictor	1	2	3	4	5	6	7	8	9	10
1. Length	1	-.374**	-0.062	.327**	-0.058	.292**	.524**	0.069	.516**	-.116*
2. WF	-.374**	1	.622**	-.652**	-.599**	-.650**	-.666**	-.438**	-.530**	.298**
3. F19Fam	-0.062	.622**	1	-.666**	-.975**	-.691**	-.565**	-.647**	-.481**	.492**
4. F19AoA	.327**	-.652**	-.666**	1	-.660**	-.983**	-.602**	-.438**	-.522**	.347**
5. S20Fam	-0.058	.599**	.975**	-.660**	1	-.688**	-.546**	-.650**	-.479**	.493**
6. S20AoA	.292**	-.650**	-.691**	.983**	-.688**	1	-.591**	-.449**	-.516**	.357**
7. LDT	.524**	.666**	.565**	-.602**	-.546**	-.591**	1	.476**	.683**	-.370**
8. LDT Accuracy	0.069	-.438**	-.647**	.438**	-.650**	.449**	.476**	1	-.465**	.581**
9. NT	.516**	.530**	.481**	-.522**	-.479**	-.516**	-.683**	-.465**	1	-.532**
10. NT Accuracy	-.116*	.298**	.492**	-.347**	-.493**	-.357**	-.370**	-.581**	-.532**	1

Note. WF= word frequency; F19Fam/AoA & S20Fam/AoA= familiarity and age of acquisition ratings for fall 2019 and spring 2020; LDT= lexical decision times; NT= Naming times
** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table 2.

Standardized Regression Coefficients from the Linear Regression Analyses on Lexical Decision Times, Lexical Naming Times, and Accuracy from the English Lexicon Project (Baltoa et al., 2007)

Variable	LDT	LDT Accuracy	NT	NT Accuracy
Step 1: Onsets				
Adjusted R Squared	0.53	0.252	0.395	0.085
Step 2: Baseline				
Length	0.320***	0.271***	0.369***	-0.005
Word Frequency	0.320***	0.540***	-0.392***	0.296***
Step 3: Familiarity and AoA				
F19Fam	-0.355***	0.566***	-0.369***	0.530***
F19AoA	0.248***	-0.309***	0.257***	-0.268***
S20Fam	-0.329***	0.561***	-0.365***	0.518***
S20AoA	0.249***	-0.308***	0.267***	-0.284***

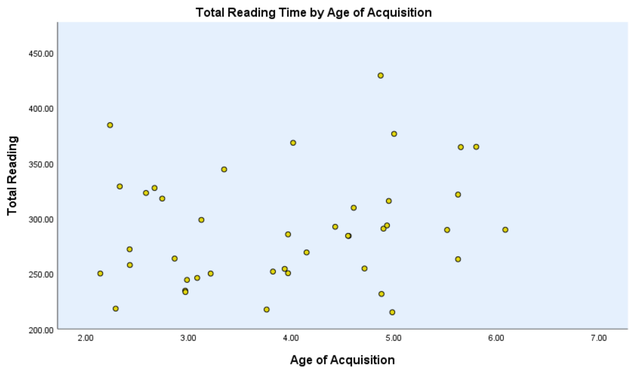
Note. F19Fam/AoA & S20Fam/AoA= familiarity and age of acquisition ratings for fall 2019 and spring 2020; LDT= lexical decision times; NT= Naming times
*** p < 0.001

Table 3.

Standardized Regression Coefficients from the Linear Regression Analyses on Lexical Decision Times, Lexical Naming Times, and Accuracy from the English Lexicon Project (Baltoa et al., 2007)

Variable	First Fixation	Gaze Duration	Total Reading Time
Step 1: Onsets			
Adjusted R Squared	0.171	0.114	0.069
Step 2: Baseline			
Length	-0.342***	-0.008	-0.248
Word Frequency	-0.446**	-0.399*	0.16
Step 3: Familiarity and AoA			
F19Fam	-0.177	0.044	-0.079
F19AoA	0.305	0.111	-0.104
S20Fam	-0.347*	-0.068	-0.097
S20AoA	0.345*	0.115	-0.058

Note. F19Fam/AoA & S20Fam/AoA= familiarity and age of acquisition ratings for fall 2019 and spring 2020; LDT= lexical decision times; NT= Naming times
* p < 0.10 ** p < 0.05 *** p < 0.005



Discussion

The preliminary data analysis on the WWEP and the ELP data yielded interesting results. The goal of the work with the GECO corpus was to create a database of the content words (nouns, verbs, adjectives, adverbs) that can be used moving forward in the Wesleyan Eye Movement and Reading Lab. For each content word, many important eye movement measures were recorded in the database (first fixation duration, gaze duration, total fixation duration). These variables will serve as the dependent measures for regression analyses in the future. Overall, this project was successful in indicating the research being conducted in the lab is moving in a significant direction and in creating avenues for future research through the management of the GECO corpus.

References

Cop, U., Dirix, N., Drieghe, D. *et al.* Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behav Res* **49,** 602–615 (2017). https://doi.org/10.3758/s13428-016-0734-0

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Studies in Written Language and Literacy Precursors of Functional Literacy,* 189-213. doi:10.1075/swll.11.14per

Balota, D.A., Yap, M.J., Hutchison, K.A. *et al.* The English Lexicon Project. *Behavior Research Methods* **39,** 445–459 (2007). https://doi.org/10.3758/BF03193014

Why study eye movements? (n.d.). Retrieved July 27, 2020, from http://www.pitt.edu/~perfetti/Eye Movements During Reading.htm

Juhasz, Barbara J., et al. "A Database of 629 English Compound Words: Ratings of Familiarity, Lexeme Meaning Dominance, Semantic Transparency, Age of Acquisition, Imageability, and Sensory Experience." *Behavior Research Methods*, vol. 47, no. 4, 2014, pp. 1004–1019., doi:10.3758/s13428-014-0523-6.