CGSC5901: Advanced Statistics for Cognitive Science

Romeo Penheiro[1]

[1] Department of Cognitive Science

Carleton University

Author Note

Graduate Student Masters in Cognitive Science

Correspondence concerning this article should be addressed to Romeo Penheiro, 1125 Colonel By Dr, Ottawa, ON K1S 5B6. E-mail: romeopenheiro@cmail.carleton.ca

CGSC5901: Advanced Statistics for Cognitive Science

Introduction

The assignment is an analysis to assess, improve, and test the robustness of data that were collected from an experiment conducted on undergraduate students in the PSYC 2001 course. The current experiment is a conceptual replication of Howes and Solomon's 1951 experiment (Howes & Solomon, 1951).

Howes and Solomon investigated the visual duration threshold as a function of word-probability. The study used Thorndike-Lorge's (1944) word count to measure word frequency because it was the largest corpus of words of that time. The results showed a strong negative relationship between word frequency and the duration threshold that was necessary to correctly identify tachistoscopically-presented words. In essence, words of low frequency (i.e., words that were used rarely) required longer presentation durations to be correctly identified compared to words of high frequency (i.e., words that were used often). This suggested that word frequency is an important variable for models of word recognition.

The current experiment is different from the original experiment in two ways. First, the experiment used a lexical decision task (Meyer & Schvaneveldt, 1971) instead of a threshold task that was used by Howes and Solomon (1951). Second, the experiment was conducted using a computer in an online setting. This design of experiment is preferable for two reasons. First, to use Howes and Solomon's (1951) task in an online setting. Second, to see whether a word frequency effect could be obtained using Howes and Solomon's stimuli (1951) in a different type of task.

The hypothesis of the experiment is to demonstrate similar word frequency as that observed in the Howes and Solomon's task (1951).

Method

Participant

The participants for the study were recruited through Carleton University's SONA system and each received a course credit in a psychology research methods course (PSYC 2001) for their participation. 102 adults ($M = 20.27$, $SD = 3.54$, *range* = 18-39, 81 female, 17 male, 3 non-binary) participated in the study. One participant was excluded from the age and gender analysis due to incomplete data (`78`).

Instrument

The experiment was created using PsychoPy software (Peirce et al., 2019) and then converted to java script to run on pavlovia. Also, the design of the experiment restricted use to laptops or desktop computers due to screen size limitations of smaller devices.

Stimuli

60 words of 6-12 letters in length were used as stimuli for the `word` condition. Unlike Howes and Solomon's (1951) study that used the Thorndike-Lorge (1944) count values, the current study used word frequency values from Kucera and Francis (1967), which is a more recent and widely-used word count. In addition, 60 non-words were used in the `non-word` condition. These were created by replacing the vowels in the words to create the non-words.

Procedure

The experiment was a lexical decision task (Meyer & Schvaneveldt, 1971). Each trial began with a fixation cross that was presented for one second. When the fixation duration ended, a string of letters were presented for 100 ms. The task involved making a judgement on the string and pressing `w` on the keyboard if it was a word or pressing `n` on the keyboard if it was not a word. The participants were asked to respond as quickly and as accurately as they could for each trial. The next trial began as soon as a response was made. If no response was made for a trial, the next trial would start automatically after ten seconds. A total of 120 trials were presented with strings uniquely randomized for each participant.

Data Cleaning

The raw data obtained from the study was cleaned in the following ways. The practice trials were removed, and a subject number was added to each participant file. The `gender` variable was renamed to female, male, non-binary, or NA (for missing values). The missing values of reaction time for each participant was imputed by using median of the reaction time across the `word` or `non-word` trials for the respective participant. The median value was used because it is less influenced by outliers compared to the mean. The trials were no responses were provided were re-coded as `NA` for accuracy to differentiate incorrect trails from missed trials. Although participant `78` had missing gender and age data, this participant was not excluded from the data analysis as the participant had complete responses. The code for data cleaning and tidying can be obtained at the following link.

Results

The results demonstrated two key findings. First, words of low frequency were incorrectly identified more often when compared to high frequency words (see Figures 1, 2). Second, words of low frequency (i.e., words that were used rarely) had longer reaction times to be correctly identified compared to high frequency words (i.e., words that were used often) (see Figures 3, 4).

The correlations of the associations were calculated for mean accuracy against word frequency (*0.36*), and mean reaction time against word frequency (*-0.48*). These associations revealed a medium positive strength of association and medium negative strength of association respectively, according to Cohen's d (Cohen, 2013). In addition, the same associations were calculated but using log 10 transformed word frequency (*0.66*, *-0.77* respectively). These associations revealed a large positive strength of association and large negative strength of association respectively, according to Cohen's d (Cohen, 2013).

The overall accuracy measure across words and non-words was calculated to be 76.41 and 90.64 respectively (see Table 1). Furthermore, descriptives were calculated for

mean accuracy for each word string across participants (see Table 2), mean accuracy for each non-word string across participants (see Table 3), mean reaction time for each word string across participants (see Table 4), and mean reaction time for each non-word string across participants (see Table 5).

Discussion

One confound that may have affected the outcome is order effects due to participant fatigue (see Figure 5). This confound may be eliminated or reduced in two ways: reducing the number of trials, and introducing a break in the middle of the experiment.

References

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203771587

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, *41*(6), 401–410. https://doi.org/10.1037/h0056020

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day american english*. Dartmouth.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University. Retrieved from //catalog.hathitrust.org/Record/000987642

Table 1

Table of the overall accuracy measure across words and non-words

String type	Mean overall Accuracy (%)
Words	76
Non-words	91

Table 2

Table of the mean accuracy in percentage for each word string across participants

Words	Overall Accuracy (%)
altruistic	60.8
amicable	73.5
assets	93.1
assiduous	19.6
automobile	95.1
barrister	69.6
beatific	20.6
benign	48.0
broker	91.2
celestial	64.7
chancels	33.3
charitable	91.2
churches	97.1
condolence	89.2
conviviality	33.3
debating	99.0
earthly	97.1
economics	94.1
education	96.1
elegies	6.9
ensemble	88.2
erudition	17.6
etcher	15.7

friendly	93.1
frugality	72.5
government	95.1
heavenly	97.1
hospitality	99.0
inductive	95.1
initiative	91.2
intellectual	95.1
judiciary	79.4
knowledge	97.1
lawyer	91.2
liberties	87.2
limousine	61.8
literary	95.1
market	95.1
metaphor	92.2
mundane	80.4
orchestra	84.3
painting	95.1
pedagogue	29.4
percipience	41.2
physics	95.1
picture	92.2
poetry	95.1
psychical	72.5
rebuttal	68.6

religious	95.1
reverence	81.4
savings	98.0
scientific	91.2
service	96.1
spiritual	96.1
statistics	97.1
sympathy	96.1
theistic	46.1
uncoerced	29.4
vignette	71.6

Table 3

Table of the mean accuracy in percentage for each non-word string across participants

Non-words	Overall Accuracy (%)
acinamocs	95
alegoes	90
ancourced	75
aurthly	94
beetofoc	97
benogn	96
berroster	90
bruker	90
calestoel	88
charches	79
chencels	90
cheroteble	86
cindalence	96
convovoeloty	88
debetong	94
eltruostoc	93
emoceble	95
erudotoon	97
essaduoas	96
essets	71
eutomobole	91
frageloty	93
froandly	90

gavirnmunt	95
huevanly	94
huspoteloty	95
iducetuon	95
judocoery	93
knuwlidge	97
letarery	85
lewyor	90
lobertoes	97
lomousone	93
merket	85
mundene	64
mutephar	95
onductove	98
onotoetove	94
ontellectiel	95
otchur	98
padegigue	93
peontong	93
percopoence	88
physocs	90
pietry	80
poctere	96
psychocel	91
relogoous	94
rivurence	92

robuttel	78
sampothy	81
scoentofoc	98
servoce	90
sevongs	96
sporotuel	89
stetostocs	89
theostoc	94
unsumble	79
urchastre	90
vognette	91

Table 4

Table of the mean reaction time in seconds for each word string across participants

Words	Mean reaction time (s)
altruistic	0.93
amicable	0.91
assets	0.70
assiduous	0.89
automobile	0.69
barrister	0.94
beatific	0.92
benign	0.93
broker	0.75
celestial	0.79
chancels	0.93
charitable	0.74
churches	0.68
condolence	0.82
conviviality	0.94
debating	0.70
earthly	0.71
economics	0.69
education	0.65
elegies	0.83
ensemble	0.78
erudition	0.85
etcher	0.79

friendly	0.64
frugality	0.86
government	0.71
heavenly	0.71
hospitality	0.74
inductive	0.74
initiative	0.71
intellectual	0.71
judiciary	0.81
knowledge	0.63
lawyer	0.63
liberties	0.80
limousine	0.86
literary	0.74
market	0.67
metaphor	0.75
mundane	0.76
orchestra	0.71
painting	0.66
pedagogue	0.87
percipience	0.96
physics	0.71
picture	0.65
poetry	0.69
psychical	0.95
rebuttal	0.80

religious	0.66
reverence	0.80
savings	0.64
scientific	0.73
service	0.64
spiritual	0.64
statistics	0.64
sympathy	0.65
theistic	0.91
uncoerced	0.82
vignette	0.80

Table 5

Table of the mean reaction time in seconds for each non-word string across participants

Non-words	Mean reaction time (s)
acinamocs	0.74
alegoes	0.83
ancourced	0.94
aurthly	0.79
beetofoc	0.69
benogn	0.76
berroster	0.80
bruker	0.86
calestoel	0.84
charches	0.80
chencels	0.78
cheroteble	0.79
cindalence	0.87
convovoeloty	0.86
debetong	0.76
eltruostoc	0.74
emoceble	0.73
erudotoon	0.74
essaduoas	0.70
essets	0.89
eutomobole	0.83
frageloty	0.86
froandly	0.76

gavirnmunt	0.82
huevanly	0.72
huspoteloty	0.76
iducetuon	0.75
judocoery	0.75
knuwlidge	0.73
letarery	0.78
lewyor	0.80
lobertoes	0.84
lomousone	0.80
merket	0.77
mundene	0.81
mutephar	0.81
onductove	0.79
onotoetove	0.67
ontellectiel	0.81
otchur	0.75
padegigue	0.79
peontong	0.75
percopoence	0.81
physocs	0.80
pietry	0.91
poctere	0.86
psychocel	0.83
relogoous	0.75
rivurence	0.80

robuttel	0.87
sampothy	0.84
scoentofoc	0.69
servoce	0.78
sevongs	0.72
sporotuel	0.79
stetostocs	0.78
theostoc	0.75
unsumble	0.95
urchastre	0.81
vognette	0.75

Figure 1

Scatterplot of the accuracy in percentage against Kucera and Francis word frequency for trials with correctly spelt words. The regression line is in blue with 95% confidence intervals.

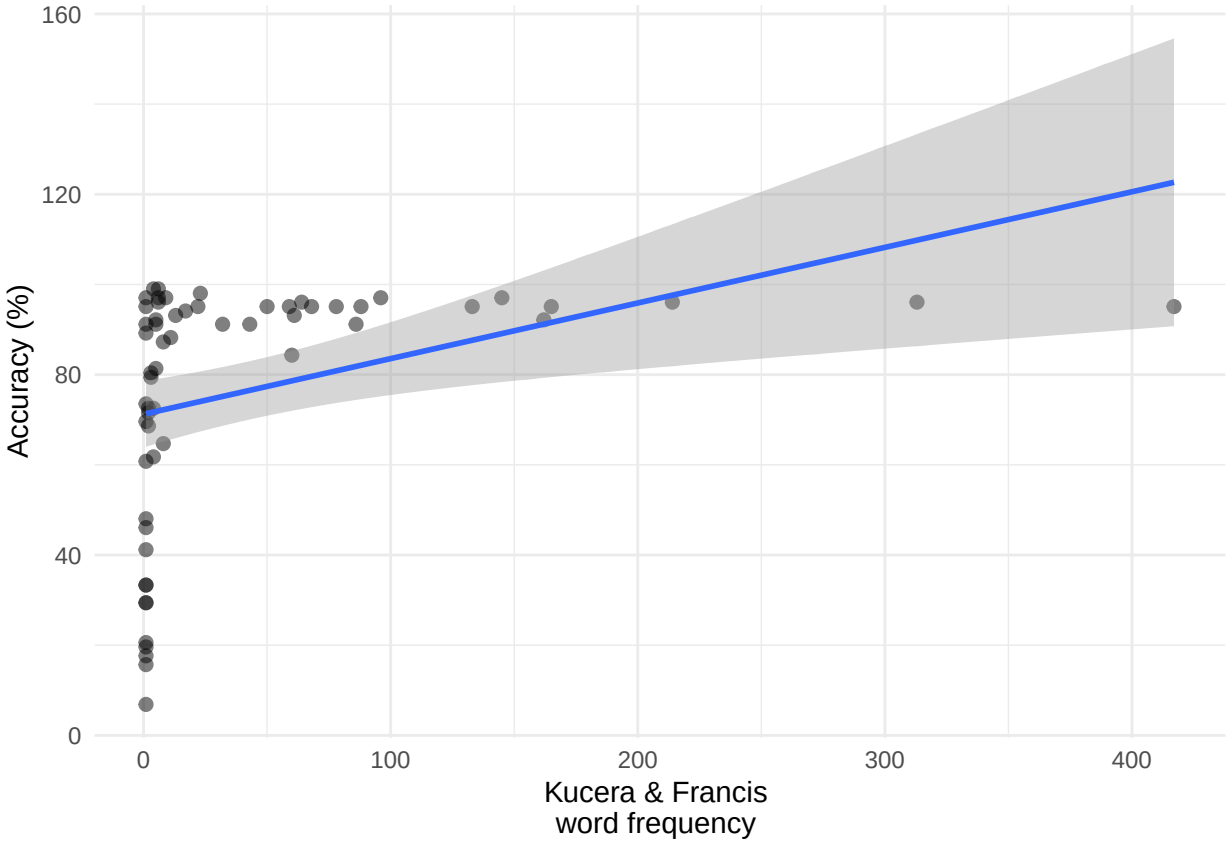

Figure 2

Scatterplot of the accuracy in percentage against log 10 of Kucera and Francis word frequency for trials with correctly spelt words. The regression line is in blue with 95% confidence intervals.

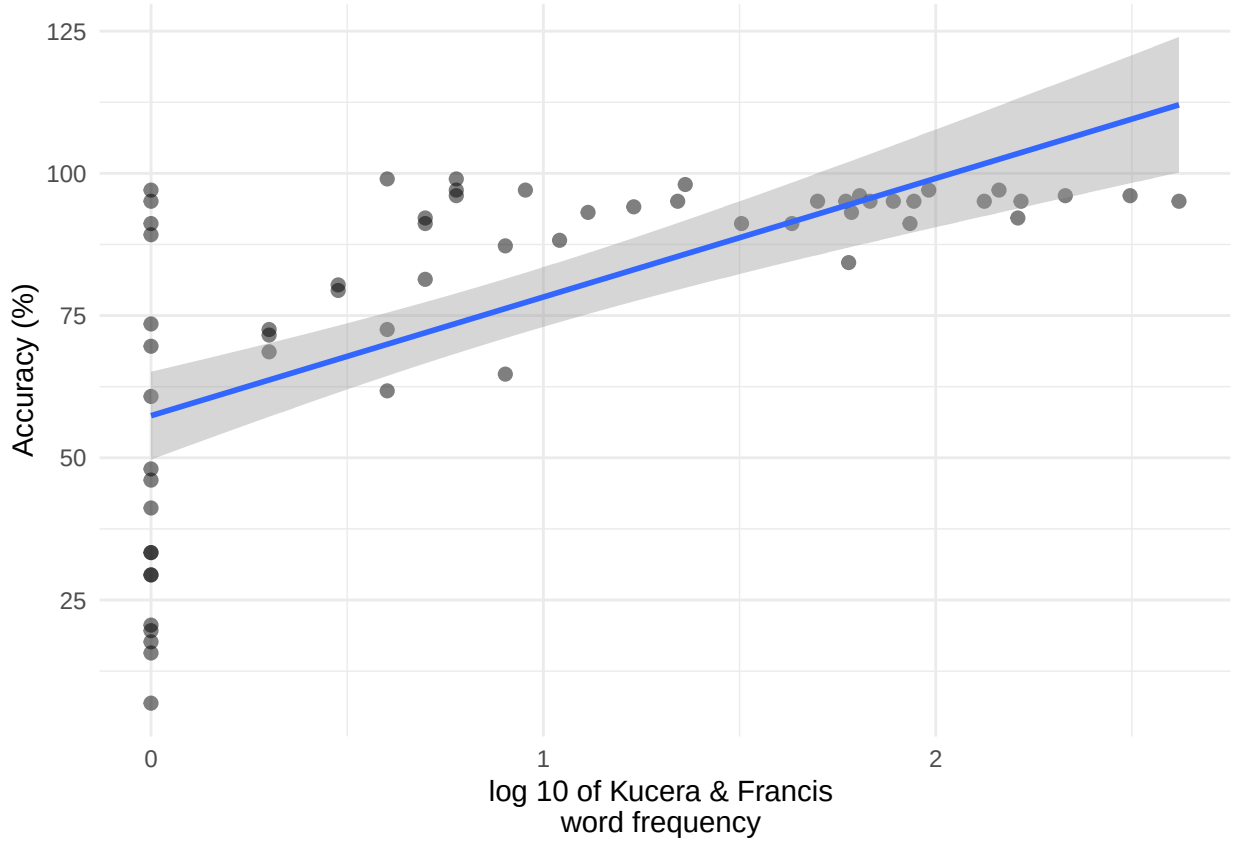

Figure 3

Scatterplot of the mean reaction time in seconds against Kucera and Francis word frequency for trials with correctly spelt words. The regression line is in blue with 95% confidence intervals.

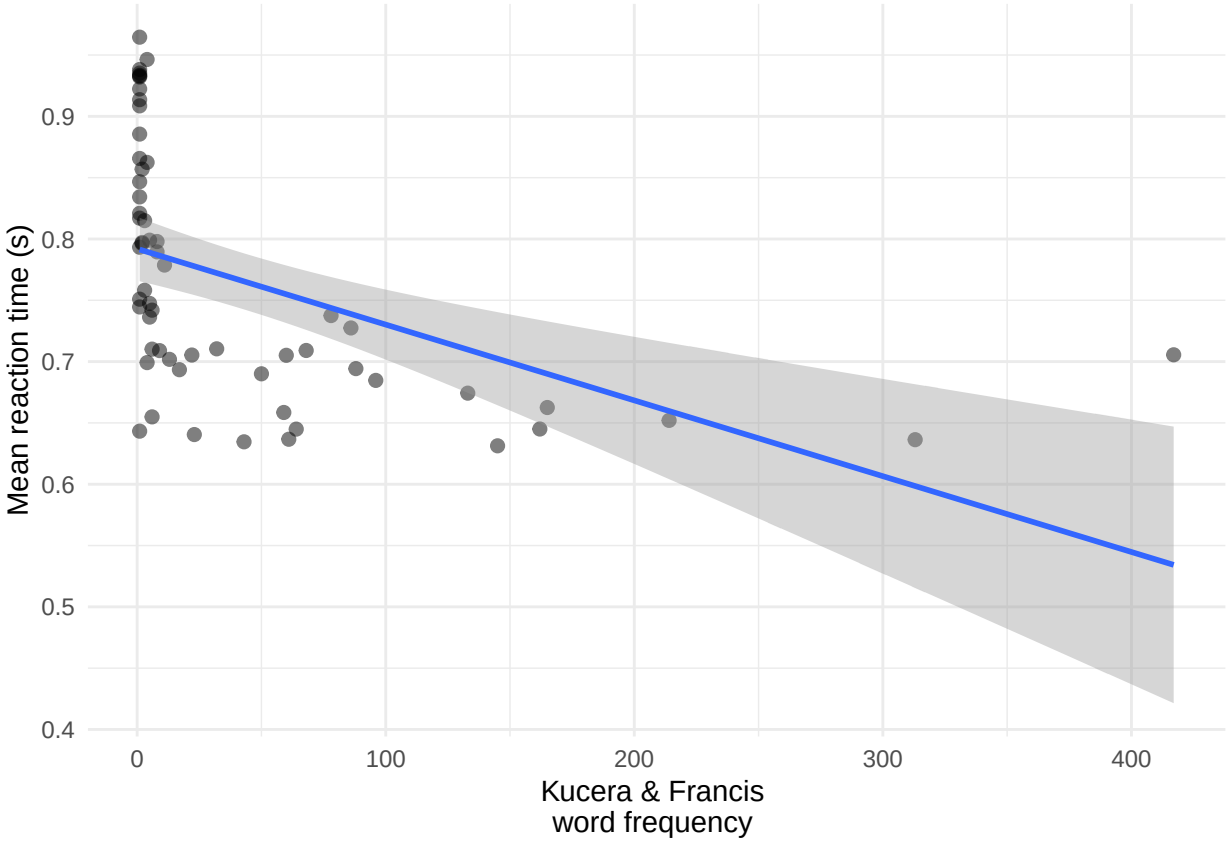

Figure 4

Scatterplot of the the mean reaction time in seconds against log 10 of Kucera and Francis word frequency for trials with correctly spelt words. The regression line is in blue with 95% confidence intervals.

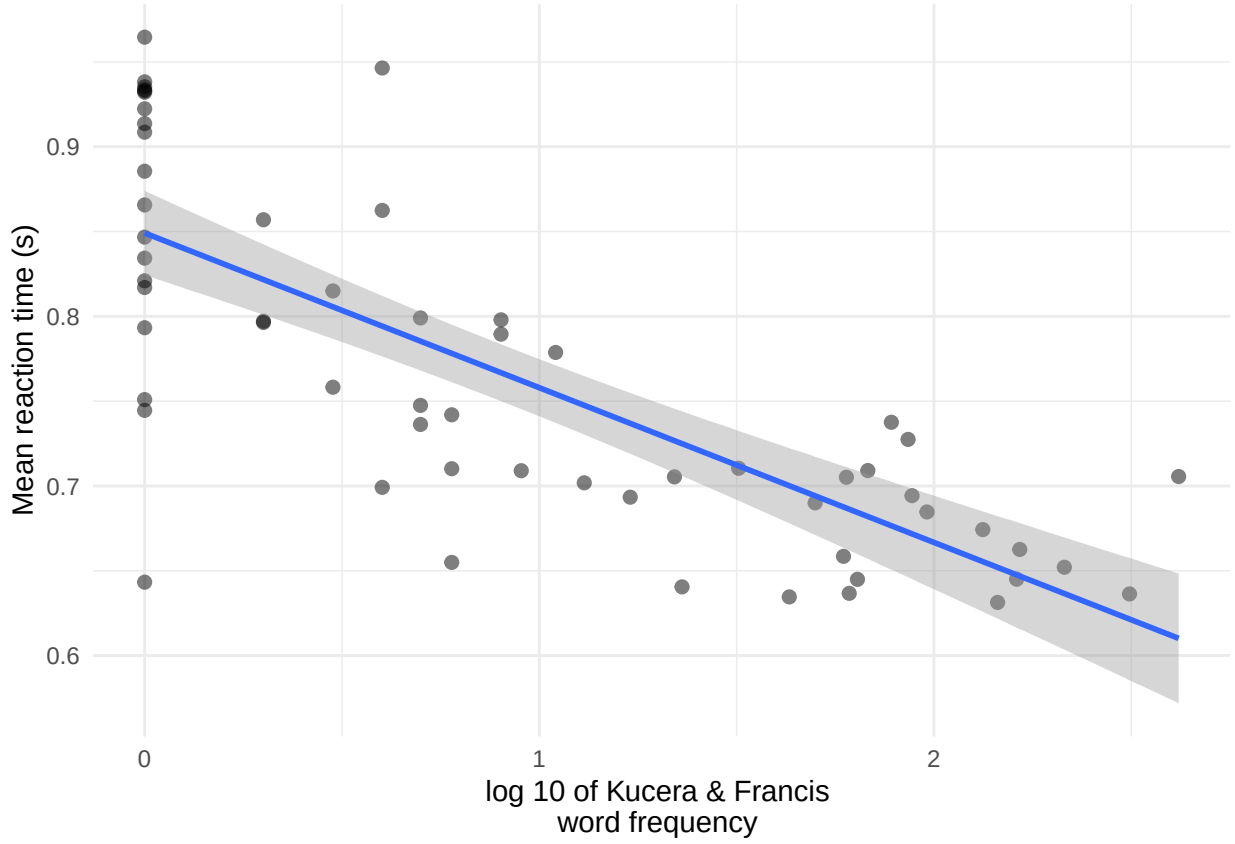

Figure 5

Scatterplot of the the mean reaction time in seconds against trial index for trials with words (left) and non-words (right). The regression line is in blue with 95% confidence intervals.