

# Assessing Word Frequency by Rating Method

**Muruli Manikanta**

Research scholar, Department of Studies in Psychology, University of Mysore, Mysore

## Abstract

One of the major challenges faced by researchers in linguistics and psycholinguistics in India is the lack of corpora of linguistic information such as word frequency, letter frequency, phonological and phonetic data on sounds, and morphological data on words, etc, in Indian languages. When they do not get the stimulus material readily available in any corpus, researchers resort to preparing a stimulus set that meets the criteria for their specific study. This paper presents an attempt of preparing a list of words for an ongoing psycholinguistic study on lexical access in Kannada by a method of expert rating on selected words' frequency. A total of 60 Kannada words were selected as stimuli, and seven native Kannada experts rated them on frequency on a six-point scale. Descriptive statistics and Pearson's correlation were used for data analysis. Results revealed that the selected Kannada words have high-frequency ratings with good inter-rater reliability.

**Keywords:** Kannada, Word Frequency, Inter-Rater Reliability, Rating.

## Introduction

Reading is a complex cognitive process that is crucial to human communication and knowledge acquisition (Ehri, 1980; Frith, 1980; Frost, 2012). The cognitive processes involved in reading are influenced by linguistic and script features. Within a language also, the type of words, their length, frequency, complexity, transparency, etc. By carefully selecting stimulus types, researchers can probe the research questions effectively and gain valuable insights into the cognitive processes underlying reading and language comprehension. For a study on word processing using a sentence reading paradigm, the researcher wanted to prepare a list that would meet the frequency criterion of the target words.

Word frequency is a fundamental linguistic characteristic that influences the cognitive processing of reading. It refers to the relative occurrence or prevalence of specific words within a language or corpus, reflecting their distribution patterns (Brysaert & New, 2009). Assessing the frequency of words allows researchers to select stimuli that are representative of natural language and are more likely to elicit typical language processing mechanisms (Jones et al., 2020). Inter-rater reliability is also important when selecting stimulus words in psycholinguistics research. It ensures consistent and reliable evaluation of various linguistic aspects associated with the stimuli.

## Method

The researcher selected 60 Kannada words from Kannada text books and local newspapers. Seven experts (Native proficient Kannada Speakers) participated as Raters in the present study. They were asked to rate the prevalence of each word written from the scale of 1 to 6, rating each word based on how frequently they encountered so far in print on the rating scale, 1=Never, 2=Very Rare, 3= Rare, 4=moderately, 5=High, 6=Very High.

**Results**

Descriptive statistics were followed to find out the average frequency of words. Pearson's product-moment correlation was employed for checking the inter-rater reliability (see Tables 1 and 2). Mean rating for each word is given in the appendix.

*Table 1: Descriptive Statistics for Frequency*

	Mean	SD	N
Rater 1	4.90	0.87	60
Rater 2	4.03	0.81	60
Rater 3	4.83	0.82	60
Rater 4	4.10	0.95	60
Rater 5	5.48	0.79	60
Rater 6	4.93	1.02	60
Rater 7	4.35	0.87	60
Total	4.66	0.88	60

*Table 2: Correlation matrix showing inter-rater reliability*

	1	2	3	4	5	6
1. Rater 1	-					
2. Rater 2	.332*	-				
3. Rater 3	.582**	.366**	-			
4. Rater 4	.469**	.403**	.385**	-		
5. Rater 5	.325*	.417**	.551**	.459**	-	
6. Rater 6	.657**	.529**	.543*	.474**	.668**	-
7. Rater 7	.367**	.372**	.269*	.409**	.623**	.389**

\*Significant at 0.05 level, \*\*Significant at 0.01 level

Table 1 shows that the Mean Frequency of words was 4.66 with an SD of 0.88. Values below and above 1.0 SD could be considered as low and high frequencies respectively. Table 2 indicates the correlations between inter-rater ratings. The results revealed that the correlation coefficient is significant ( $p < 0.05$ ).

**Discussion**

The present study was aimed at preparing a set of Kannada word list with high frequency words for an ongoing study on word processing in Kannada. The rating method with seven raters has resulted in an empirically based word list. This list may also be used by other researchers, if found suitable for their research. The results have also shown high inter-rater reliability. Using such empirical based stimuli enhances the validity of studies. It also enhances replicability (Simmons, Nelson, & Simonsohn, 2011), and improves the interpretation of research results (Vasishth, Mertzen, Jäger, & Gelman, 2008).

### Implications

Present study results has implications for future studies as this provides a list of Kannada words with high frequency that could be readily used in future research. The present study offers valuable insights to future researchers concerning the significance of validating procedures for stimuli before incorporating them into a study. Further investigations can delve deeper into the psychometric properties of different types of stimulus words by conducting thorough analyzes.

### References:

1. Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977-990.
2. Costa, A., & Santesteban, M. (2004). Bilingual word perception and production. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 417-439). Oxford University Press.
3. Ehri, L. C. (1980). The development of orthographic images. *Cognitive processes in spelling*, 495, 516.
4. Frith, U. (1980). Unexpected spelling problems. *Cognitive processes in spelling*, 495-515.
5. Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, 35(5), 263-329.
6. Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161-163.
7. Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38.
8. Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1-2), 71-102.
9. Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, 76(1p2), 1.
10. Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.
11. Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2013). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Journal of Memory and Language*, 69(4), 511-526.

**Appendix**  
**Word List with mean frequency**

Kannada Word	Average Frequency	Kannada Word	Average Frequency	Kannada Word	Average Frequency	Kannada Word	Average Frequency
ಪವನಶಕ್ತಿ	4	ವೈಚಾರಿಕತೆ	3.5	ಮನಮೋಹಕ	4.5	ದ್ವೇಷರಹಿತ	4.3
ದೇಶಾಭಿಮಾನ	5.3	ಚಟುವಟಿಕೆ	4	ದೂರಾಲೋಚನೆ	4.3	ಮೂಕವೇದನೆ	4.5
ಗುರುಕಾಣಿಕೆ	4.8	ಭೋಗಜೀವನ	4.3	ನದಿಬಯಲು	3.8	ಕಲಾಮಂದಿರ	4.8
ಗಣಕಯಂತ್ರ	4.6	ವಾತಾವರಣ	4.6	ಸೈನ್ಯಾಧಿಪತಿ	4.1	ಸಮಾಜವಾದಿ	4.6
ಯಶಸ್ವಿಯುತ	4.5	ದೇವನಾಗರಿ	5.2	ಗಾನಕೋಗಿಲೆ	5.4	ಸರ್ವಾನುಮತ	5
ಕಾಲಹರಣ	4.3	ಹಣವಂತರ	3.8	ಪ್ರಕಾಶಮಾನ	4.3	ವಿಚಾರಶೀಲ	5
ಹಯವದನ	4.1	ಪರಮಗುರಿ	4	ಮಾನವೀಯತೆ	4.6	ದೂರದರ್ಶಕ	4.3
ಸಮತಾವಾದ	5.2	ಜೀವನಶೈಲಿ	5	ಪರಿವರ್ತನೆ	5.3	ಗುರುವಂದನೆ	4.3
ಸಮತೋಲನ	4.3	ಸರ್ವಾಧಿಕಾರ	5	ಪ್ರಾಣಿಸಂಕುಲ	4.6	ಪ್ರಾಣಿಸ್ನೇಹಿತ	5
ರಾಸಾಯನಿಕ	4.6	ಧರ್ಮಾಧಿಕಾರಿ	4.6	ಸಂಗೀತಗಾರ	4.5	ಮದ್ಯಕಾಲೀನ	4
ಪ್ರತಿಭಾವಂತ	4.8	ವಸಂತಕಾಲ	4.5	ಭಾವಸಾಗರ	5.2	ಕ್ಷೇಮಪಾಲನೆ	4.5
ಮಾರ್ಗದರ್ಶನ	5.3	ಭವಿಷ್ಯವಾಣಿ	4.6	ದಯಾಮರಣ	4	ಸಹಾನುಭೂತಿ	5.3
ಸಿರಿವಂತಿಕೆ	4.3	ಪತ್ರಿಕೋದ್ಯಮ	4.3	ಸಹೋದರತ್ವ	4.5	ಬರವಣಿಗೆ	4.5
ಪರಿಕಲ್ಪನೆ	4.6	ಸುರ್ವಾರಥ	5	ದೋಷಪೂರಿತ	4.8	ಫಲದಾಯಕ	3.8
ಸಾಹಸಗಾಥೆ	4.5	ಮಹಾಭಾರತ	4.5	ಪ್ರತಿಫಲನ	5	ಸಹನಾಶಕ್ತಿ	5.4