

Empirical Investigations of Generative AI in Experimental Linguistics

The Estimation of Word Metrics

Dr. Enis OĞUZ



Dr. Enis OĞUZ

**EMPIRICAL INVESTIGATIONS OF GENERATIVE AI IN
EXPERIMENTAL LINGUISTICS
The Estimation of Word Metrics**

ISBN 978-625-8897-55-5

Responsibility of the contents belongs to its author.

© 2026, PEGEM AKADEMİ

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage or retrieval system, without prior permission in writing from the publishers. Pegem Academy Publishing Training and Consultancy Services Tic. Ltd. Şti. It is a publishing house founded in 1998 in Ankara/Turkey which has been operating regularly for 28 years. Published books, it is included in the catalogs of higher education institutions. Pegem Academy has over 2000 publications in the same field from different authors. Detailed information about our publications can be found <http://pegem.net>.

1st Edition: October 2026, Ankara

Publication-Project: Selcan Özgürel
Typesetter-Graphic Designer: Tuğba Kaplan
Cover Design: Pegem Akademi

Printed by: Sonçağ Yayıncılık Matbaacılık Reklam San Tic. Ltd. Şti.
İstanbul Street İstanbul Bazaar 48/48 İskitler - Ankara

Publishing House Certificate Number: 51818
Printing House Certificate Number: 47865

Contact

Pegem Akademi: Shira Trade Center
Macun Neighbourhood 204th Street Number: 141/A-33 Yenimahalle/ANKARA
Publishig House: 0312 430 67 50
Distribution: 0312 434 54 24
Preparatory Courses: 0312 419 05 60
Internet: www.pegem.net
E-mail: yayinevi@pegem.net
WhatsApp Line: 0538 594 92 40

FOREWORD

Research in experimental linguistics often relies on balanced word lists that are controlled for objective (e.g., frequency) and subjective (e.g., emotional arousal) word metrics. However, such values are not available for all words in a language, and notoriously, they are limited especially in non-Indo-European languages such as Turkish. This situation often leaves researchers with two choices: collect word metric data for specific word lists or simply ignore the variation in word metrics. The first option is challenging due to the time and resource costs, which become even more significant when the primary focus of the research is not the investigation of word metrics. The second option brings the risk of misleading results and conclusions, introducing noise rather than providing clear directions for the literature.

Empirical Investigations of Generative AI in Experimental Linguistics raises a crucial question: Can Generative AI reliably estimate word metrics even in a language that is typologically different from the languages that dominate its primary training data? Using controlled experiments, the chapters of this book compare human baseline ratings with AI-generated estimates for frequency, familiarity, concreteness, imageability, emotional arousal, and emotional valence. Each experimental chapter includes a second experiment with a “few-shot” anchoring strategy, examining whether providing reference values can calibrate the models and improve their accuracy. The results are promising for some metrics, whereas others should continue to be based primarily on human ratings.

Specifically designed for psycholinguistics and computational linguistics, this book offers a statistically tested framework for integrating Generative AI into experimental linguistics research. It offers a highly efficient way to reduce the traditional data collection costs for certain word metrics, allowing researchers to create more balanced word lists without compromising research design quality or undertaking extensive data collection procedures.

Dr. Enis Oğuz
Ankara / Türkiye

CONTENTS

Foreword.....	iii
---------------	-----

CHAPTER 1

GENERATIVE AI AND WORD METRICS IN PSYCHOLINGUISTIC RESEARCH

Psycholinguistic Word Metrics	2
Generative AI Models as Automated Annotators	3
Estimating Word Metrics by Generative AI Models	4
Reliability of AI-Estimates: Word Frequency and Lexical Richness	5
Reliability of AI Estimates: Subjective Metrics	5
Strategies in Estimating Word Metrics with GenAI Models	6
Diverse Language Challenge	7
Testing the Accuracy of GenAI Estimations	7
Assessing Rank-Order and Monotonicity with Spearman's	8
Assessing Linear Variance with Pearson's	8
Assessing Accuracy and Bias with Error Metrics	9
Structure of the Book	10
References	11

CHAPTER 2

GENERATIVE AI AND WORD FREQUENCY

Introduction	14
What is Word Frequency?	14
How to Use Generative AI to Calculate Word Frequency	15
The Present Study	16
Experiment 1	17
Results of Experiment 1	19
Discussion for Experiment 1	22
Experiment 2	23
Results of Experiment 2	24
Discussions for Experiment 2	28
Conclusions	29
References.....	29

CHAPTER 3 GENERATIVE AI AND WORD FREQUENCY FOR A SPECIFIC CONTEXT AND TIMELINE

Introduction	32
The Present Study	33
Experiment 1	33
Results for Experiment 1	35
Discussion for Experiment 1	38
Experiment 2	39
Results for Experiment 2	41
Discussions for Experiment 2	44
Conclusions.....	45
References.....	46

CHAPTER 4 GENERATIVE AI AND WORD FAMILIARITY

Introduction	48
The Distinction between Frequency and Familiarity	48
Generative AI Models in Estimating Word Familiarity	48
Present Study.....	49
Experiment 1	49
Results for Experiment 1	50
Discussions for Experiment 1	53
Experiment 2	54
Results for Experiment 2	55
Discussions for Experiment 2	58
Conclusion	59
References.....	60

CHAPTER 5 GENERATIVE AI AND WORD IMAGEABILITY

Introduction.....	62
The present study	63
Experiment 1.....	63
Results for Experiment 1	64
Discussions for Experiment 1	67
Experiment 2.....	68
Results for Experiment 2	69
Discussions for Experiment 2	71
Conclusions.....	72
References.....	72



CHAPTER 6
GENERATIVE AI AND WORD CONCRETENESS

Introduction 74

The Present Study 75

 Experiment 1 75

 Results for Experiment 1 76

 Discussions for Experiment 1 79

 Experiment 2 80

 Results for Experiment 2 81

 Discussions for Experiment 2 84

Conclusions 85

References 85

CHAPTER 7
GENERATIVE AI AND WORD EMOTION METRICS

Introduction 87

The present study 88

 Experiment 1 89

 Results for Experiment 1 90

 Discussions for Experiment 1 95

 Experiment 2 95

 Results for Experiment 2 96

 Discussions for Experiment 2 101

Conclusions 101

References 102

CHAPTER 8
**THE POWER OF GENERATIVE AI IN ESTIMATING WORD METRICS:
PERFORMANCE, LIMITATIONS, AND FUTURE DIRECTIONS**

Introduction 104

Performance of AI Models in Estimating Word Metrics 104

Current Limitations in Estimating Word Metrics by GenAI Models 106

Future Directions 107

References 108

CHAPTER 1

GENERATIVE AI AND WORD METRICS IN PSYCHOLINGUISTIC RESEARCH

Abstract

This introductory chapter examines the fundamental concepts in estimating word metrics (e.g., frequency) using Generative AI (GenAI) models. The chapter, and the book, focuses on a morphologically rich and productive non-Indo-European language, Turkish. Consequently, a critical question was investigated: Can AI reliably estimate word metrics in a language that is typologically different from the language it was mostly trained on? The investigation is of utmost importance since calculating word metrics can be costly and time-consuming, especially when the data for that particular language is limited and requires new data collection processes. The chapter discusses different strategies in prompting, as well as the statistical evaluations to determine the performance of AI models. In particular, Spearman's rank correlation (ρ), Pearson's correlation coefficient (r), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) were explained along with their formulas. The definitions in this chapter set the scene for upcoming chapters, giving essential information regarding statistical metrics and methodological choices.

Keywords: Generative AI (GenAI), psycholinguistics, word metrics, Turkish, computational linguistics