# Multi-Word Representations in Minds and Models: Investigating Storage Mechanisms in Humans and Large Language Models

Zachary Nicholas Houghton

2024-11-25

# Table of contents

# List of Figures

# List of Tables

# Abstract

This is my abstract.

# Acknowledgements

I started this journey in the middle of a pandemic that persisted through much of my program. It is no exaggeration to say that my success in this program is due in large, or perhaps completely, to the people below.

Though I would like to take some of the credit for myself.

First and foremost, I would not be here if it weren't for my incredible advisor, Dr. Emily Morgan. Emily has been a never-ending source of knowledge, a guiding light, and a constant source of reassurance. If you're familiar with me you probably know that I don't stop moving. I have often struggled to linearize my endless thoughts into a coherent manner and Emily was charged with the seemingly impossible task of translating my incoherent stream of thoughts into a coherent set of ideas. She believed in me for my entire time here, even at times when I didn't quite believe in myself. Words can express neither the gratitude nor the debt that I owe to you, Emily.

In addition to all of the professors here in both the linguistics and psychology departments here at UC Davis, I'd also like to thank many of the other brilliant minds here who have been crucial to my development as a researcher. Specifically, I'd like to thank Dr. Fernanda Ferreira, Dr. Kenji Sagae, Dr. Santiago Barreda, Dr. Georgia Zellou, and Dr. Masoud Jasbi. Over my years at UC Davis, each of these professors has volunteered countless hours of their time and wisdom to me, indulging my endless stream of questions.

Many of the ideas presented here have benefited in some form or another from feedback from the many graduate students here. I would especially like to thank Casey Felton, Harvey Qiu, Skyler Reese, and Nicole Dodd for their feedback on much of the work included here.

I'd also like to thank Casey, Felix, and Nora for being a strong support system during my time here. Our Sunday shenanigans were a constant escape from the tireless work of completing a PhD.

My journey in linguistics started at the University of Oregon, and I want to thank my undergraduate cohort in linguistics as well as all of the professors that supported the beginning of my journey. I particularly want to acknowledge Dr. Vsevolod Kapatsinski. Volya has donated countless hours of his time to me even after his role as my undergraduate thesis advisor was long over. He continues to be an endless source of knowledge and inspiration and much of my knowledge and interest in language learning comes from him. Perhaps more importantly, however, he has always reminded me that linguistics is *fun*! Had it not been for our meetings over the years that devolved into random and fun tangents, I would have burnt out long ago. I would not be here without you, Volya.

I would also like to thank Kim . Her words of encouragement and faith in me helped me believe in myself.

In addition, I want to thank Dr. Melissa Baese-Berk, Dr. Misaki Kato, Dr. Zara Harmon. Aside from being both exceptional researchers and inspirational people, each one of them was crucial to developing my skills as a researcher, as a linguist, and as a person. If I can become even half the linguist they are, I'll be incredibly proud of myself.

Along with the technical and academic guidance, it also would have been impossible to complete this PhD without the unending support I received from my many close friends. It would take up too much space to name all of them, but they surely know who they are.

I also want to acknowledge my sisters, Kayla and Lily. We've been through so much together. I don't know where I would be, not just academically, but more generally in life, had you two not been by my side. I love you both.

This work would also have not been completed without the influence of my parents. Specifically, I want to thank my mom for teaching me that the ability to find the answer is far more important than knowing the answer, and my dad, for teaching me the discipline and practical skills to achieve my goals.

The number of people who have been inconsequential for me getting here is undoubtedly larger than is feasible to include here. To those that I have inevitably left out, I apologize.

# 1 Introduction

From a young age, humans are capable of generating sentences that they've never encountered before (Kapatsinski 2018; Berko 1958). This ability is largely enabled by our ability to store forms that we've learned and compute new forms by applying knowledge of the grammar to these stored forms (Joseph P. Stemberger and MacWhinney 2004; Joseph Paul Stemberger and MacWhinney 1986; Morgan and Levy 2016, 2015; Berko 1958). In theory, these can be complementary forces: if a form is stored, it does not need to be computed, and if a form can be computed, it does not have to be stored. For example, if the word *cats* is stored, then there is no reason to compute it. On the other hand, if it can be computed (e.g., we have learned the word *cat*, and we have learned how to make regular forms plural in English), then there may be no reason to store it. This has been the story told by many of the early linguistic theories, and understandably so.

## 1.1 Computation and Storage

# References


Berko, Jean. 1958. "The Child's Learning of English Morphology." WORD 14 (2-3): 150–77. https://doi.org/10.1080/00437956.1958.11659661.

Kapatsinski, Vsevolod. 2018. *Changing Minds Changing Tools: From Learning Theory to Language Acquisition to Language Change*. MIT Press. https://books.google.com/books?hl=en&lr=&id=YZxjDwAAQBAJ&oi=fnd&pg=PR5&dq=kapatsinski+changing+minds&ots=9bGhgkCaY0&sig=MHfWF9cbhbtMmx33a0FYSM6AMAs.

Morgan, Emily, and Roger Levy. 2015. "Modeling Idiosyncratic Preferences : How Generative Knowledge and Expression Frequency Jointly Determine Language Structure," 1649–54.

———. 2016. "Abstract Knowledge Versus Direct Experience in Processing of Binomial Expressions." *Cognition* 157: 384–402. https://doi.org/10.1016/j.cognition.2016.09.011.

Stemberger, Joseph Paul, and Brian MacWhinney. 1986. "Frequency and the Lexical Storage of Regularly Inflected Forms." *Memory & Cognition* 14 (1): 17–26. https://doi.org/10.3758/BF03209225.

Stemberger, Joseph P., and Brian MacWhinney. 2004. "Are Inflected Forms Stored in the Lexicon." *Morphology: Critical Concepts in Linguistics* 6: 107122. https://books.google.com/books?hl=en&lr=&id=bGl0aKBld3cC&oi=fnd&pg=PA107&dq=stemberger+2004+inflected&ots=RdvzVaC_NS&sig=0DJV8gUVaoZv_COZqcLXOu5_evU.