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. Emily was charged with the non-trivial task of helping to translate my incoherent stream of thoughts into a coherent set of ideas. She pushed me hard, believed in me, and never let me fall behind. Words can express neither the gratitude nor the debt that I owe to you, Emily.

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 many brilliant graduate students. I would especially like to thank Casey Felton, Harvey Qiu, Skyler Reese, Nicole Dodd, and Penny Pan 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 welcomed escape from the tireless work of completing a PhD.

My journey in linguistics started at the University of Oregon, and I want to thank 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 is a constant reminder that linguistics is *fun*! Had it not been for our meetings over the years that devolved into ridiculous linguistic tangents, I would have burnt out long ago. I would not be here without you, Volya.

I would also like to thank Kim 선생님. 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, and Dr. Zara Harmon. Aside from being both exceptional researchers and inspirational people, each one of them was crucial to my development 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 have been fortunate to have a strong support system in the form of of my two 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.

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.

Finally, I would like to thank Addy, Charles, Spencer, Paul, Wyatt, and 보미 for being my very strong support network. Despite being thousands of miles away I could always count on all of you when times were tough.

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

1 Introduction

1.1 Computation and Storage

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 not necessarily a reason to compute it (e.g., by combining the lexical root, *cat*, with a general plural rule, -*s*). 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 (e.g., Chomsky 1965), and understandably so.

Early theories argued for a strict division between items that are stored and items that are computed (Chomsky 1965). These theories often prioritized efficiency and minimizing memory consumption. Storing items that could be generated computationally was considered redundant.

-What is computation? What is storage?

-Chomsky accounts (minimal storage)

-Bybeean accounts (frequent items stored)

-Evidence for storage

-Evidence from phonology, learning, semantics, processing, regular vs irregular forms

-Storage at a syllable-level, word-level, phrasal-level, and sentence-level.

-Evidence for computation

-less important to show, taken for granted, can mention Berko-Gleason

1.2 What is Storage?

-What does it mean for something to be stored?

-Exemplar theories

-Abstractions

-What happens to stored items?

-How are they stored (internal structure)

-How are they processed?

-Potential competition effects

1.3 Models of Storage

-Fragment grammars

-LLMs

-What can we learn from them?

1.4 Questions

-What drives storage?

-Confident about frequency effects, but what about predictability?

-How are stored units processed?

-language processing is inherently linear, how does this interact with holistic storage?

-What happens to the internal structure of stored units?

-e.g., Kapatsinski & Radicke

2 staub_replication_and_extension

The results from our previous experiments could be a task-specific property of the maze task, since it forces participants to make a decision which may commit them to a specific interpretation in a way that more naturalistic reading may not. Thus, in this Experiment we replicate the previous experiment using eye-tracking.

2.1 Methods

2.1.1 Participants

56 native English speakers were recruited from the University of California, Davis subjects pool. They were given course credit in exchange for their participation. All participants had normal or corrected vision.

2.1.2 Materials

The materials here were identical to those in Experiment 2.

2.1.3 Procedure

We recorded participants' eye movements using the Eyelink 1000 Plus. We recorded pupil movements from the right eye. Participants were seated 850mm away from the screen. Our screen resolution was 1920x1080, 531.3mm in width, and 298.8mm in height.

Comprehension was checked for non-experimental trials and participants below 80% accuracy were excluded. Out of our 56 participants, 0 were excluded for falling below the accuracy threshold.

2.1.4 Analyses

Prior to our analyses, sentences with blinks were excluded and fixations less than 80ms in duration and within one character of the nearest fixation were merged into that fixation (following Staub et al. 2007). For our regions of interest (the first noun and the second noun independently in the compound n oun), we computed first fixation duration, first pass time, and go-past time.

For each analysis, our independent variables were plausibility (high or low) and (log) predictability (high or low) and their interaction. We also included random slopes for condition and predictability by subject and plausibility by compound noun as well as intercepts for subject and compound noun. For each of our models, we used sum-coded, where the intercept represents the grand mean and the fixed-effect coefficient estimates represent the distance from the grand mean.

2.2 Results

2.2.1 First Fixation Times

2.2.1.1 N1

First, we examined the effects of plausibility and predictability on first fixation times for the first noun. Note that in our case, since plausibility was codes as -1 for plausible and 1 for implausible, a positive coefficient estimate of plausibility corresponds to longer fixation times for implausible items. Additionally, following Houghton et al. (2024) we also report the percentage of posterior samples greater than zero. Our model results can be found below in Table 2.1 and a visualization can be found in Figure 2.1.

Table 2.1: Model results examining the effect of plausibility and predictability on first fixation times for the N1 region.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept	231.016	4.433	222.356	239.502	100.00
Plausibility	1.661	2.020	-2.289	5.695	78.95
Predictability	2.782	2.870	-2.804	8.532	83.75
Plausibility:Predictability	-3.473	2.012	-7.418	0.522	4.15

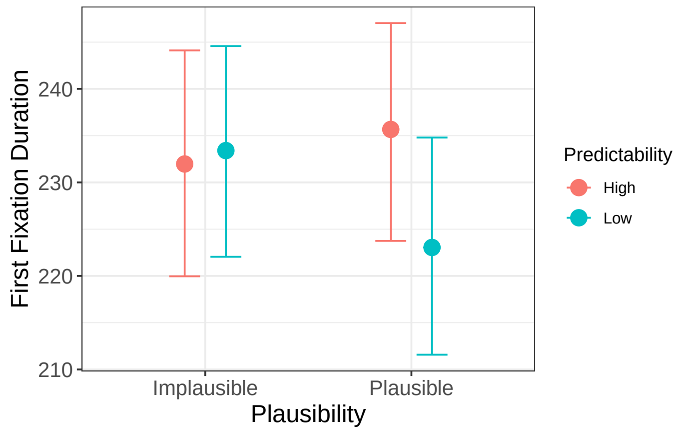

Figure 2.1: Visualization of the effects of plausibility and predictability on first fixation times for the N1 region.

Our results for first-fixation times, while non-significant, do show an interesting trend, with the effect of plausibility in the expected direction (although with only ~78% samples greater than zero). While 96% of the samples less than zero for the interaction effect, the effect size is so small that it's not particularly meaningful.

2.2.1.2 N2

Next we examine the effects of plausibility and predictability on the first fixation times of the second noun in the compound.

Table 2.2: Model results examining the effect of plausibility and predictability on first fixation times for the N2 region.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept	234.448	4.864	224.479	243.767	100.000
Plausibility	-2.151	2.973	-8.132	3.685	23.225
Predictability	-4.136	2.988	-9.887	1.887	8.575
Plausibility:Predictability	-2.928	3.012	-8.920	2.967	16.075

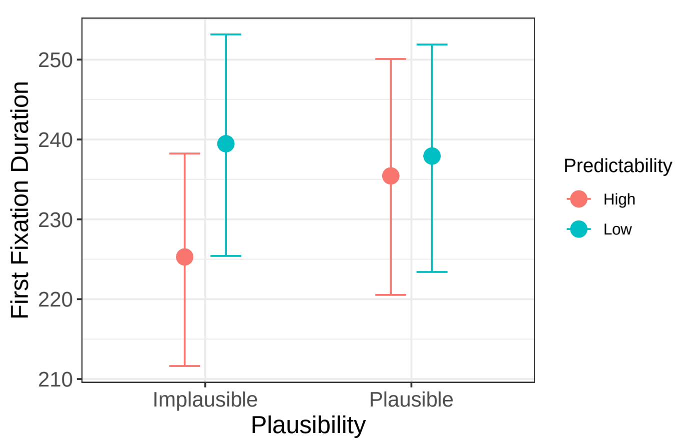

Figure 2.2: Visualization of the effects of plausibility and predictability on first

surprising because in the present study predictability is a measure of how expected the second noun is given the first. There also is not much of a plausibility effect on the N2, which is also not particularly surprising because the second noun eliminates the local implausibility. That is, both sentences are plausible at the N2 region.

2.2.2 Gaze/First-Pass Time

2.2.2.1 N1

Our results for the effects of plausibility and predictability on gaze/first-pass times on the N1 region are presented in **?@tbl-exp1m2** and visualized in **?@fig-exp1m2**.

Table 2.3: Model results examining the effect of plausibility and predictability on Gaze/first-pass times for the N1 region.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept	264.141	8.422	246.898	280.601	100.000
Plausibility	-0.001	0.203	-0.385	0.413	49.075
Predictability	0.005	0.198	-0.394	0.387	51.475
Plausibility:Predictability	-0.010	0.202	-0.411	0.382	48.100

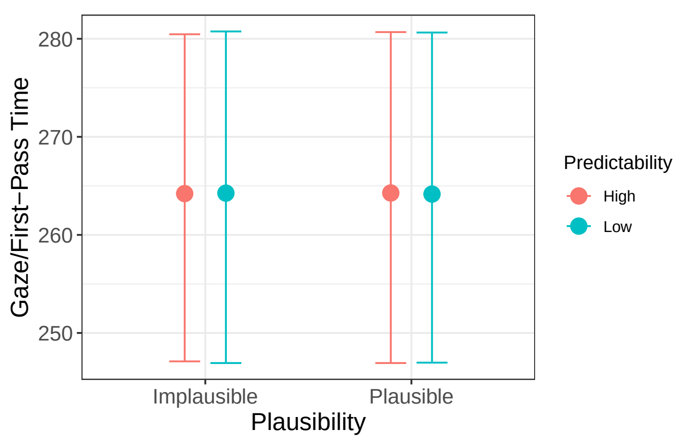

Figure 2.3: Visualization of the effects of plausibility and predictability on Gaze/first-pass times for the N1 region.

Our results demonstrate no effect of either plausibility or predictability on the gaze times at the N1 region. We further examined our filler items to rule out an error with the eye-tracker. Our filler items contain a frequency manipulation and an analysis of the filler items demonstrated that the frequency manipulation did effect gaze times in our filler items, suggesting that the results here are not due to any malfunction of the eye-tracker or cleaning procedure.

2.2.2.2 N2

Our results for the N2 region are presented in Table 2.4 and visualized in Figure 2.4.

Table 2.4: Model results examining the effect of plausibility and predictability on Gaze/first-pass times for the N2 region.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept	253.757	6.071	241.726	265.767	100.000
Plausibility	-0.003	0.100	-0.196	0.192	48.825
Predictability	-0.003	0.101	-0.207	0.190	48.625
Plausibility:Predictability	0.003	0.100	-0.199	0.202	51.625

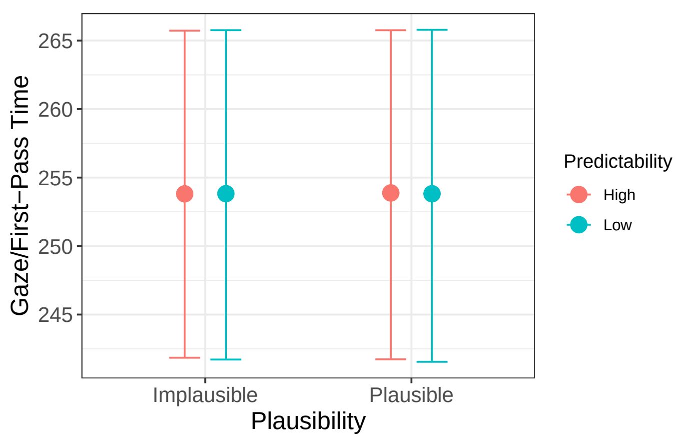

Figure 2.4: Visualization of the effects of plausibility and predictability on Gaze/first-pass times for the N2 region.

The results of the N2 region also show a lack of effect of plausibility and predictability on gaze times.

2.2.3 Go-Past Time

2.2.3.1 N1

Our results for the effects of plausibility and predictability on go-past times are presented in Table 2.5 and visualized in Figure 2.5.

Table 2.5: Model results examining the effect of plausibility and predictability on go-past times for the N1 region.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept	357.207	16.452	325.205	388.957	100.00000
Plausibility	0.018	0.197	-0.367	0.400	54.05000
Predictability	0.005	0.200	-0.390	0.394	50.10000
Plausibility:Predictability	-0.000	0.200	-0.392	0.393	49.76667

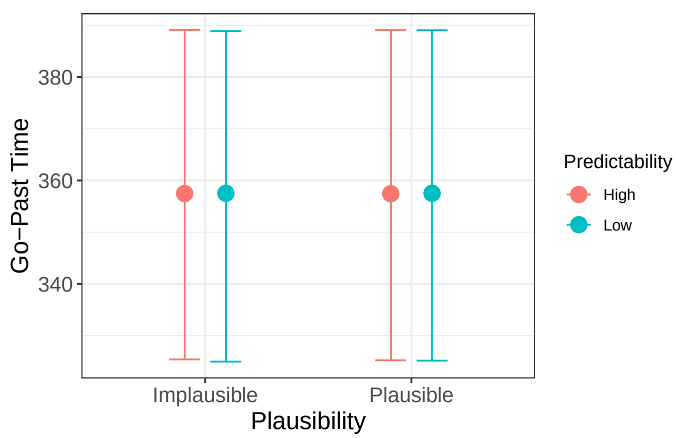

Figure 2.5: Visualization of the effect of plausibility and predictability on go-past times for the N1 region.

Our results for go-past times similarly show no effect of predictability and plausibility.

2.2.3.2 N2

Our results for the N2 region are presented in Table 2.6 and visualized in Figure 2.6.

Table 2.6: Model results examining the effect of plausibility and predictability on go-past times for the N2 region.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept	342.737	12.866	317.685	367.649	100.000
Plausibility	-0.005	0.099	-0.202	0.195	47.600
Predictability	-0.001	0.100	-0.198	0.195	50.000
Plausibility:Predictability	-0.002	0.100	-0.192	0.198	49.025

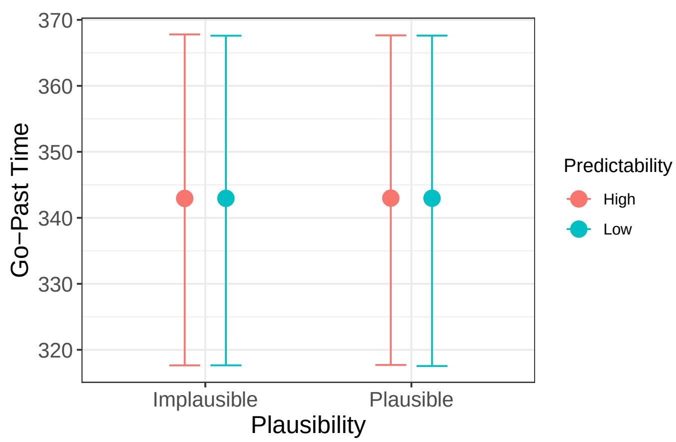

Figure 2.6: Visualization of the effect of plausibility and predictability on go-past times for the N2 region.

Our results for the N2 region similarly show no effect of predictability and plausibility on go-past times.

2.2.4 First-Pass Regression

2.2.4.1 N1

Our results for the effects of predictability and plausibility on the first-pass regression times on the N1 region are presented in Table 2.7 and visualized in Figure 2.7.

Table 2.7: Model results examining the effect of plausibility and predictability on first-pass regression for the N1 region.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept	-1.645	0.151	-1.948	-1.358	0.000
Plausibility	0.199	0.080	0.041	0.357	99.375
Predictability	-0.049	0.086	-0.214	0.123	27.750
Plausibility:Predictability	0.128	0.075	-0.019	0.275	96.025

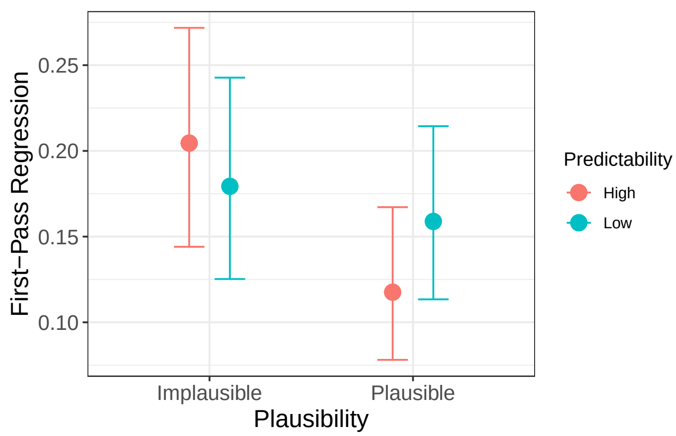

Figure 2.7: Visualization of the effect of plausibility and predictability on first-pass regression for the N1 region.

Our results suggest that readers are more likely to regress after the first fixation in the implausible condition compared to the plausible condition. Further, this plausibility effect is larger for high-predictability items than low predictability items. This is surprising because predictability is a measure of the N2, not the N1 and if readers are anticipating the N2 then it should alleviate the local implausibility at the N1 (which would result in a negative interaction effect, i.e. the opposite trend from what we see here). Instead, we see …

2.2.4.2 N2

Table 2.8: Model results examining the effect of plausibility and predictability on first-pass regression for the N2 region.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept	-2.061	0.176	-2.426	-1.728	0.000
Plausibility	-0.026	0.106	-0.241	0.184	40.625
Predictability	-0.022	0.102	-0.223	0.177	41.100
Plausibility:Predictability	0.049	0.097	-0.136	0.239	69.975

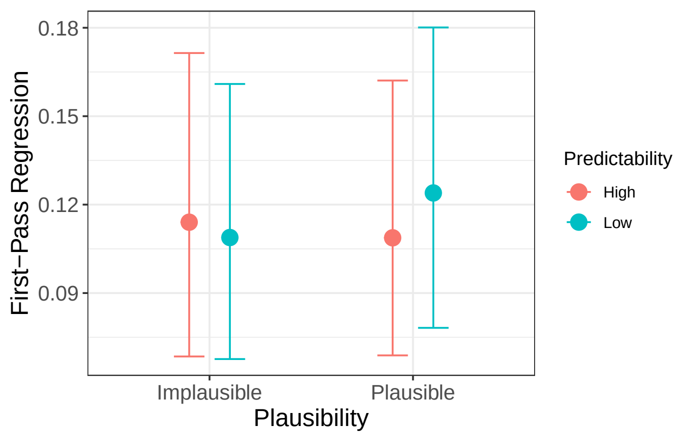

Figure 2.8: Visualization of the effect of plausibility and predictability on first-pass regression for the N2 region.

Our results at the N2 region show no effect of predictability or plausibility.

2.3 Discussion

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.

Chomsky, Noam. 1965. "Aspects of the Theory of Syntax Special Technical Report No. 11." https://ntrs.nasa.gov/api/citations/19670002070/downloads/19670002070.pdf.

Houghton, Zachary, Misaki Kato, Melissa Baese-Berk, and Charlotte Vaughn. 2024. "Task-Dependent Consequences of Disfluency in Perception of Native and Non-Native Speech." *Applied Psycholinguistics*, January, 1–17. https://doi.org/10.1017/S0142716423000486.

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.

Staub, Adrian, Keith Rayner, Alexander Pollatsek, Jukka Hyönä, and Helen Majewski. 2007. "The Time Course of Plausibility Effects on Eye Movements in Reading: Evidence from Noun-Noun Compounds." *Journal of Experimental Psychology: Learning Memory and Cognition* 33 (6): 1162–69. https://doi.org/10.1037/0278-7393.33.6.1162.

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.