Analyzing Brain Age through Entity-Centric Analysis of Narratives in Picture Description Tasks

Benjamin Ascoli
Computer Science
Emory University
Atlanta, GA, 30322, USA
bascoli@emory.edu

Ezequiel Gleichgerrcht
Neurology
Emory University
Atlanta, GA, 30322, USA
eze.gleich@emory.edu

Jinho D. Choi
Computer Science
Emory University
Atlanta, GA, 30322, USA
jinho.choi@emory.edu

Abstract

This study leverages advanced NLP tools to introduce a novel, three-dimensional framework for analyzing linguistic changes associated with brain aging through the "Cookie Theft" picture description task. We assess lexical richness and diversity, syntactic complexity, and semantic content to uncover subtle markers of cognitive decline. Our analysis reveals that aging is associated with several of these metrics, suggesting adaptive strategies to maintain semantic coherence despite cognitive changes.

1 Introduction

In the rapidly evolving field of Natural Language Processing (NLP), the pursuit of novel applications has led to groundbreaking intersections between technology and neuroscience (Crema et al., 2022). Among these, the study of brain aging through the lens of linguistic change presents a frontier rich with potential insights. This study leverages advanced NLP techniques to introduce a novel, three-dimensional framework for analyzing discourse, focusing on lexical richness and diversity, syntactic complexity, and semantic content. By applying this framework to picture description tasks—specifically the 'Cookie Theft' picture, seen in figure 1—our research aims to identify subtle markers of brain aging. This study aims to redefine how discourse analysis can contribute to our understanding of cognitive decline and facilitate more comprehensive assessments of brain health.

2 Related Work

The use of picture description tasks for textual analysis has provided a robust platform to elicit spontaneous narratives. Studies such as those by Fraser et al. 2016, Gleichgerrcht et al. 2021, and Wilmskoetter et al. 2019 have utilized this task as a predictor for Alzheimer's Disease and aphasia. However, existing research has focused on conventional metrics such as parts of speech tagging, leaving several unexplored avenues in discourse structure and entity relationships. Discourse analysis in general has seen extensive exploration, including methodologies for analyzing lexical richness by Lu 2012, but the application of these techniques to picture description tasks and brain aging remains largely untouched. While previous studies have contributed significantly to the fields of discourse analysis and cognitive linguistics, there remains a notable gap in employing advanced NLP technologies to explore brain aging and picture description tasks. This study seeks to bridge this divide by utilizing state-of-the-art NLP tools like openAI's GPT-4 (OpenAI, 2023) for an in-depth exploration of linguistic patterns.

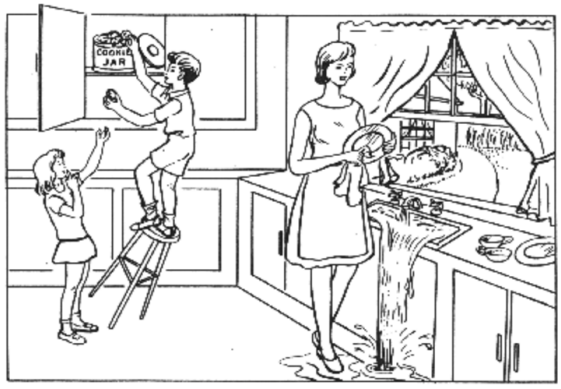

Figure 1: Cookie Theft picture (Goodglass and Kaplan, 1972). Participants are told to "describe everything you see in this picture", and their response is transcribed for later analysis.

3 Approach

This study uses the Aging Brain Cohort (ABC) dataset, which collects sociodemographic, neuroimaging, and language tasks in healthy participants, including verbatim transcripts from participants engaged in the "Cookie Theft" picture description task (Newman-Norlund et al.,

2021). Each participant's MRI T1-weighted sequences was input into BrainAgeR, a machine learning pipeline for computing estimated brain age (EBA) based on a model trained on 3377 healthy participants (Clausen et al., 2022). Participants are grouped according to their EBA into 10 consecutive, non-overlapping age groups.

We adopt a three-dimensional approach to analyze discourse in the context of brain aging:

1. **Lexical Richness.** This involves analyzing word-level metrics including frequency, imageability, familiarity, and age of acquisition, helping us measure the sophistication of language use among participants. We extract these metrics from the Medical Research Council (MRC) Psycholinguistic Database (Clark and Coltheart, 1997).

2. **Syntactic Complexity.** We evaluate the complexity of sentence structures produced by participants using the Emory Language and Information Toolkit (ELIT) (He et al., 2021). We quantify syntactic complexity by counting the number of descendants from each noun and verb in both dependency and constituency parse trees generated by ELIT.

3. **Semantic Content.** Through an entity-centric approach, we focus on the identification and linking of entities and predicates, aiming to uncover changes in the ability to convey ideas. We employ GPT-4 to facilitate the identification and linking of entities mentioned, such as characters and objects within the "Cookie Theft" image. We then extract predicates associated with these entities, cluster them using Sentence-BERT (Reimers and Gurevych, 2019), and perform predicate linking.

Each dimension—lexical, syntactic, and semantic—serves as a pillar in understanding language use. Our study aims to reveal nuanced patterns of cognitive decline, offering a holistic view of how aging impacts linguistic capabilities.

4 Analysis

Our analysis revealed key insights into the linguistic manifestations of brain aging. Our word-level metrics indicated that participants in older age groups tended to use words with lower familiarity scores, suggesting a capability to use more advanced language as cognitive processing capabilities evolve. Contrastingly, the number of descendants from each noun and verb generally

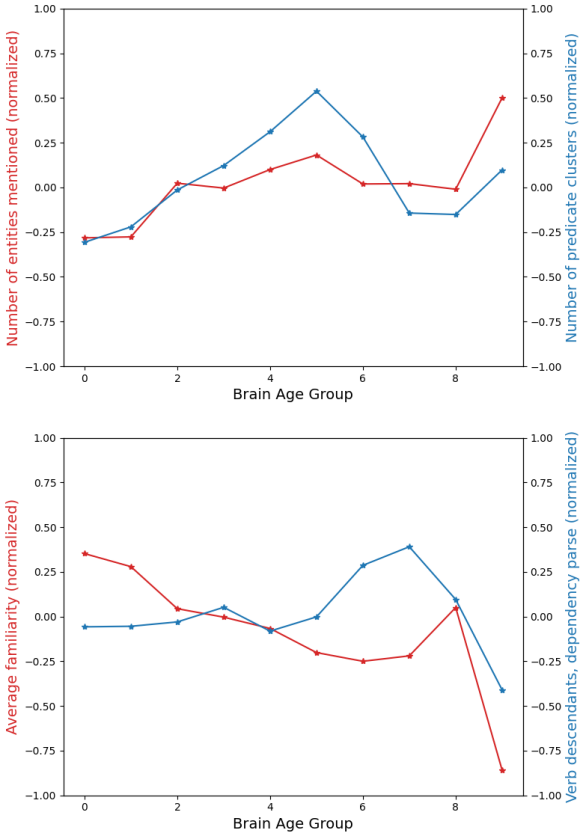

Figure 2: Various metrics vs. Brain age groups. Top: number of entities and predicate clusters mentioned. Bottom: Average familiarity and verb descendants from a dependency parse tree.

increased with age, and then plummeted, pointing to a reduction in syntactic complexity. This trend suggests that cognitive decline may manifest in straightforward linguistic constructions. Older age groups used a broader range of entities, but fewer predicates in their descriptions. This could indicate that they discuss more entities without the same amount of detail.

These findings underscore the impact of brain aging on linguistic capabilities and the interplay between cognitive decline and language use.

5 Conclusion

This study employs a novel framework to analyze linguistic changes associated with brain aging, uncovering trends in lexical diversity, syntactic complexity, and semantic coherence. Our work suggests that these linguistic markers could inform the development of diagnostic tools for cognitive decline. Expanding this research to include longitudinal datasets could offer deeper insights into how linguistics evolve with aging, and enhance our understanding of the relationship between cognition and language in aging.

Acknowledgements

E.G. was supported by the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH) under Award Numbers UL1TR002378 and KL2TR002381. We would like to thank the Aging Brain Cohort (ABC) team for providing access to the dataset that made this research possible.

Ethics Statement

This research adhered to ethical guidelines for the responsible conduct of research. The study exclusively used anonymized transcripts, maintaining the confidentiality and anonymity of all participants. The use of NLP tools, including openAI's GPT-4, was solely for the purpose of linguistic analysis within the bounds of ethical research practices. The authors declare no conflicts of interest in the conduct of this research.

References

Craig Clark and Max Coltheart. 1997. Mrc psycholinguistic database. Available from the MRC Psycholinguistic Database website.

Ashley N. Clausen, Kelene A. Fercho, Molly Monsour, Seth Disner, Lauren Salminen, Courtney C. Haswell, Emily Clarke Rubright, Amanda A. Watts, M. Nicole Buckley, Adi Maron-Katz, Anika Sierk, Antje Manthey, Benjamin Suarez-Jimenez, Bunmi O. Olatunji, Christopher L. Averill, David Hofmann, Dick J. Veltman, Elizabeth A. Olson, Gen Li, Gina L. Forster, Henrik Walter, Jacklynn Fitzgerald, Jean Théberge, Jeffrey S. Simons, Jessica A. Bomyea, Jessie L. Frijling, John H. Krystal, Justin T. Baker, K. Luan Phan, Kerry Ressler, Laura K. M. Han, Laura Nawijn, Lauren A. M. Lebois, Lianne Schmaal, Maria Densmore, Martha E. Shenton, Mirjam van Zuiden, Murray Stein, Negar Fani, Raluca M. Simons, Richard W. J. Neufeld, Ruth Lanius, Sanne van Rooij, Saskia B.J. Koch, Serena Bonomo, Tanja Jovanovic, Terri deRoon Cassini, Timothy D. Ely, Vincent A. Magnotta, Xiaofu He, Chadi G. Abdallah, Amit Etkin, Christian Schmahl, Christine Larson, Isabelle M. Rosso, Jennifer Urbano Blackford, Jennifer S. Stevens, Judith K. Daniels, Julia Herzog, Milissa L. Kaufman, Miranda Olff, Richard J. Davidson, Scott R. Sponheim, Sven C. Mueller, Thomas Straube, Xi Zhu, Yuval Neria, Lee A. Baugh, James H. Cole, Paul M. Thompson, and Rajendra A. Morey. 2022. Assessment of brain age in posttraumatic stress disorder: Findings from the enigma ptsd and brain age working groups. *Brain and Behavior*, 12(1):e2413.

Claudio Crema, Giuseppe Attardi, Daniele Sartiano, and Alberto Redolfi. 2022. Natural language processing in clinical neuroscience and psychiatry: A review. *Frontiers in Psychiatry*, 13.

K.C. Fraser, J.A. Meltzer, and F. Rudzicz. 2016. Linguistic features identify alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2):407–422. PMID: 26484921.

E. Gleichgerrcht, R. Roth, J. Fridriksson, D. den Ouden, J. Delgaizo, B. Stark, G. Hickok, C. Rorden, J. Wilmskoetter, A. Hillis, and L. Bonilha. 2021. Neural bases of elements of syntax during speech production in patients with aphasia. *Brain and Language*, 222:105025. Epub 2021 Sep 20. PMID: 34555689; PMCID: PMC8546356.

Harold Goodglass and Edith Kaplan. 1972. *The Assessment of Aphasia and Related Disorders*, 1 edition. Lea Febiger, Philadelphia.

Han He, Liyan Xu, and Jinho D. Choi. 2021. Elit: Emory language and information toolkit.

Xiaofei Lu. 2012. The relationship of lexical richness to the quality of esl learner's oral narratives. *The Modern Language Journal*.

Roger D. Newman-Norlund, Sarah E. Newman-Norlund, Sara Sayers, Samaneh Nemati, Nicholas Riccardi, Chris Rorden, and Julius Fridriksson. 2021. The aging brain cohort (abc) repository: The university of south carolina's multimodal lifespan database for studying the relationship between the brain, cognition, genetics and behavior in healthy aging. *Neuroimage: Reports*, 1(1):100008.

OpenAI. 2023. Gpt-4 Technical Report.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks.

J. Wilmskoetter, J. Fridriksson, E. Gleichgerrcht, B.C. Stark, J. Delgaizo, G. Hickok, K.I. Jr. Vaden, A.E. Hillis, C. Rorden, and L. Bonilha. 2019. Neuroanatomical structures supporting lexical diversity, sophistication, and phonological word features during discourse. *NeuroImage: Clinical*, 24:101961. Epub 2019 Jul 27. PMID: 31398554; PMCID: PMC6699249.