



Longitudinal Predictors of Cognitive Impairment using Connected Speech

Davida Fromm, Brian MacWhinney, Ziyang Wang, Xinfei Cen, Jin Yu Kim, Jack Shen, Joel Greenhouse
Carnegie Mellon University

OBJECTIVE

To identify linguistic variables that are associated with FUTURE cognitive status using the Cookie Theft picture description task.

BACKGROUND

Speech and language abilities are important factors in detecting cognitive impairments in older adults [1, 2, 3].

- This is well established in both heuristic studies that measure language behaviors and machine learning studies that develop algorithms.

But what about BEFORE dementia is diagnosed?

- After all, the neuropathological processes begin years to decades before the onset of symptoms [4].
- Studies also show that language differences appear before the onset of dementia – e.g., in semantic content and speech fluency [3].

RESEARCH QUESTIONS

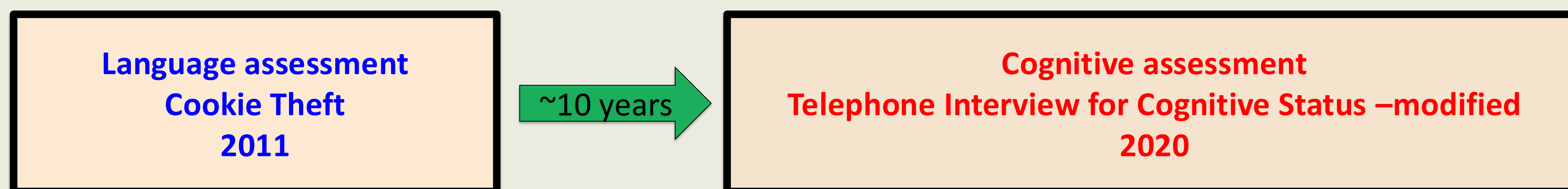
- Can language samples predict the likelihood of cognitive impairment a decade later?
- If so, what are the characteristics in those language samples that differ from those who did not experience cognitive impairment?

METHODS

This project uses a large DementiaBank dataset from the **Wisconsin Longitudinal Study** (WLS, [5]). Media files, transcripts, demographic data, and test results: <https://talkbank.org/dementia/access/English/WLS.html>

The WLS corpus is a long-term, longitudinal study of a random sample of men and women who graduated from Wisconsin high schools in 1957 and one of their siblings.

The sample for this project (**n = 801**) includes all those who described the Cookie Theft picture [6] in 2011 and were assessed in 2020 for cognitive status.



TICS-m [8] -- screening instrument that broadly assesses orientation, attention, working memory, conceptual reasoning, language skills, and episodic memory for words

Definition of Cognitive Impairment: IF TICS-m score ≤ 28

- Comprehensive home- or clinic-based assessment** aligned with the National Alzheimer's Coordinating Center's (NACC) neuropsychological battery [9]
- Clinical diagnosis made by consensus by interdisciplinary team**

Sample demographics (n = 801)

- mean age = 69.9 years (s.d. = 4.1)
- 430 females, 371 males
- normal cognitive status: n=698**
- cognitively impaired: n=103**

- Transcribed in CHAT format
- Analyzed using 3 composite CLAN [7] commands
 - EVAL – 18 general linguistic variables
 - FLUCALC – 17 fluency-related variables
 - CORELEX – 15 core lexicon variables

METHODS, cont.

Statistical analyses: Logistic regression models were used to predict the log odds of future cognitive impairment.

- Variable reduction:** Stepwise regression was used to select variables from each category: **linguistic**, **fluency**, and **core lexicon**

RESULTS

These variables emerged as significant ($p < .05$) from the variable selection procedures:

- complexity index** – ratio of grammatical relations indicating true clause-based complexity to total number of grammatical relations
- duration** – total time for participant's picture description
- type token ratio** – ratio of unique words to total words, measure of lexical diversity
- total words** – excluding repetitions and revisions
- MLU** – mean length of utterance in words
- words per minute** – rate of speech
- # of omitted words** -- # of words omitted from established Cookie Theft core lexicon list [10]

The final logistic regression model controlled for **age and sex** and included the following variables except **total words** which was excluded due to significant collinearity.

	Variable	Coefficient (SE)	p-value	Goodness of Fit (Hosmer-Lemeshow)
Final model	Grammatical Complexity Index	0.05 (0.03)	0.089	p-value= 0.21
	log(duration)	-1.19 (0.42)	0.005	
	log (MLU in words)	-0.21 (0.32)	0.506	
	log(lexical diversity)	-3.37 (1.24)	0.006	
	log(words/minute)	-1.14 (0.52)	0.027	
	# of omitted words	0.06 (0.07)	0.369	

DISCUSSION

- Measurements of linguistic ability hold predictive value for cognitive impairment occurring a decade later. This underscores the potential of linguistic assessments as early indicators of future cognitive decline.
- After controlling for age and sex, the significant variables are: **time spent on task** (in seconds), **lexical diversity**, and **words per minute**. More specifically, increased likelihood of future cognitive impairment is significantly associated with:
 - ↓ **productivity** ----- **shorter sample length**
 - ↓ **semantic content** -- **less lexical diversity**
 - ↓ **speech fluency** ----- **slower rate of speech**
- Measures of syntax were less predictive.
- These findings should be replicated with another sample of individuals as well as with other discourse tasks.

References & Acknowledgment

- [1] De la Fuente Garcia, S., Ritchie, C. W., & Luz, S. (2020). Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: a systematic review. *Journal of Alzheimer's Disease*, 78(4), 1547-1574.
- [2] Lanzi, A. M., Saylor, A. K., Fromm, D., Liu, H., MacWhinney, B., & Cohen, M. L. (2023). DementiaBank: Theoretical rationale, protocol, and illustrative analyses. *American Journal of Speech-Language Pathology*, 32(2), 426-438.
- [3] Mueller, K. D., Kosciak, R. L., Hermann, B. P., Johnson, S. C., & Turkstra, L. S. (2018). Declines in connected language are associated with very early mild cognitive impairment: Results from the Wisconsin Registry for Alzheimer's Prevention. *Frontiers in Aging Neuroscience*, 9, 437.
- [4] Golde, T. E., Schneider, L. S., & Koo, E. H. (2011). Anti-Aβ therapeutics in Alzheimer's disease: the need for a paradigm shift. *Neuron*, 69(2), 203-213.
- [5] Herd, P., Carr, D., & Roan, C. (2014). Cohort profile: Wisconsin Longitudinal Study (WLS). *International Journal of Epidemiology* 43, 34-41.
- [6] Goodglass, H., Kaplan, E., & Barresi, C. (2001). Boston Diagnostic Aphasia Examination (3rd ed.). Lippincott Williams & Wilkins.
- [7] MacWhinney, B., & Fromm, D. (2024). TalkBank methods for studying spoken discourse. In *Spoken discourse impairments in the neurogenic populations: A state-of-the-art, contemporary approach* (pp. 97-109). Cham: Springer International Publishing.
- [8] Welsh, K. A., Breitner, J. C., & Magruder-Habib, K. M. (1993). Detection of dementia in the elderly using telephone screening of cognitive status. *Cognitive and Behavioral Neurology*, 6(2), 103-110.
- [9] Besser, L., Kukull, W., Knopman, D. S., Chui, H., Galasko, D., Weintraub, S., ... & Morris, J. C. (2018). Version 3 of the national Alzheimer's coordinating center's uniform data set. *Alzheimer Disease & Associated Disorders*, 32(4), 351-358.
- [10] Dalton, S. G., AL Harbi, M., Berube, S., & Hubbard, H. I. (2024). Development of main concept and core lexicon checklists for the original and modern Cookie Theft stimuli. *Aphasiology*, 38(12), 1975-1999.

This work was funded by National Institute of Aging of the National Institutes of Health under award # RF1AG083823 (PIs: MacWhinney and Lanzi).