

# Language features of AD pathology

---

PCREF workshop 2020

University of Pennsylvania

Sunghye Cho, Sanjana Shellikeri, Sharon Ash,  
Mark Liberman, Murray Grossman, Naomi Nevler

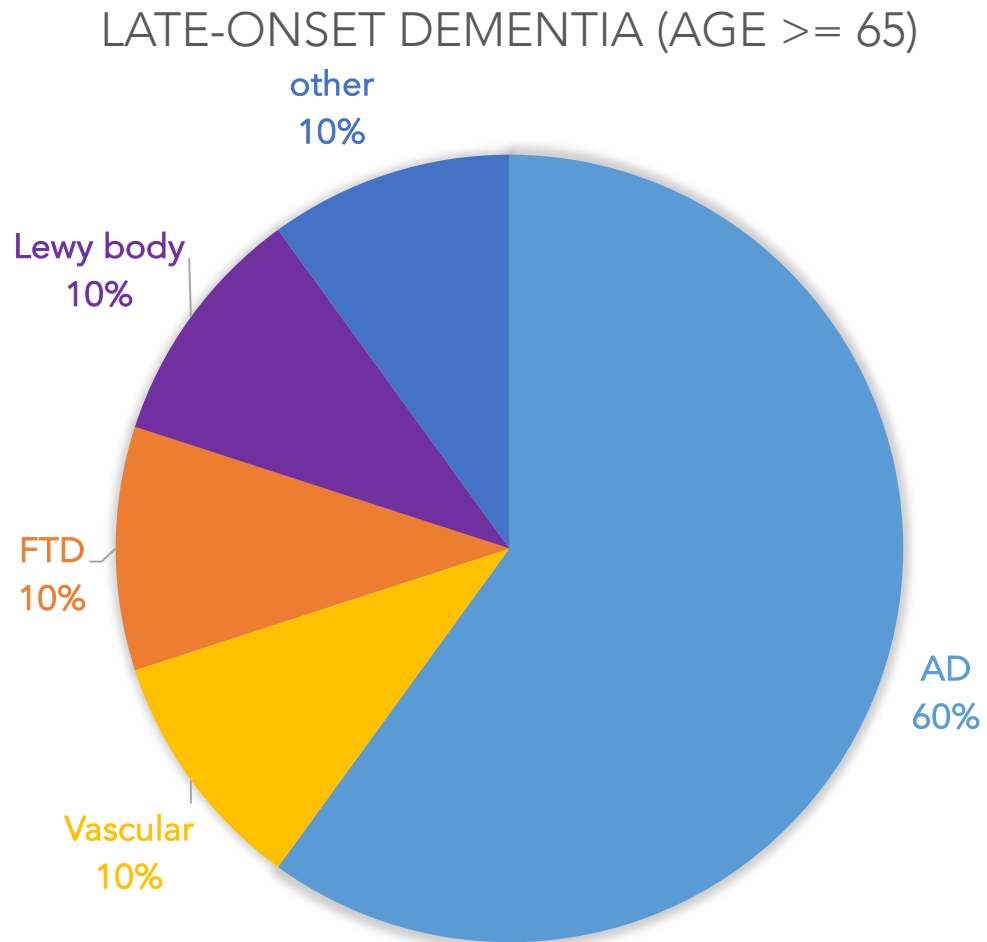


01

**INTRODUCTION**

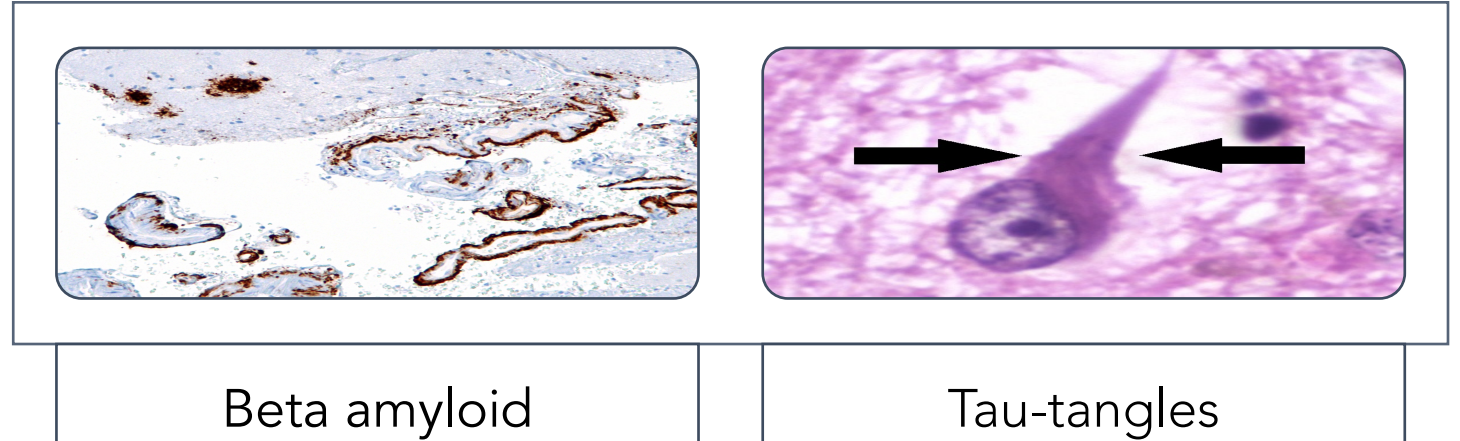
---

# Alzheimer's disease (AD)



# AD pathology

- The pathology of AD:



- AD pathology is observed not only in patients with the typical amnesic presentation, but also in patients with atypical non-amnesic presentation.
- **Logopenic variant primary progressive aphasia (lvPPA)** is one of such non-amnesic presentations.

# Limitations of previous studies

- Our understanding of linguistic features of patients with AD pathology is still relatively superficial.
- It is unclear how language of patients with AD pathology is different from that of patients with other types of neurodegenerative pathologies, such as FTLD-tau or FTLD-TDP.

# Goals of the present study

- We examined **language characteristics** of both amnestic and non-amnestic speakers with AD pathology in depth, analyzing lexical and acoustic features in narrative, natural speech.
- We directly compared **amnestic and non-amnestic AD** patients.

02

**METHODS**

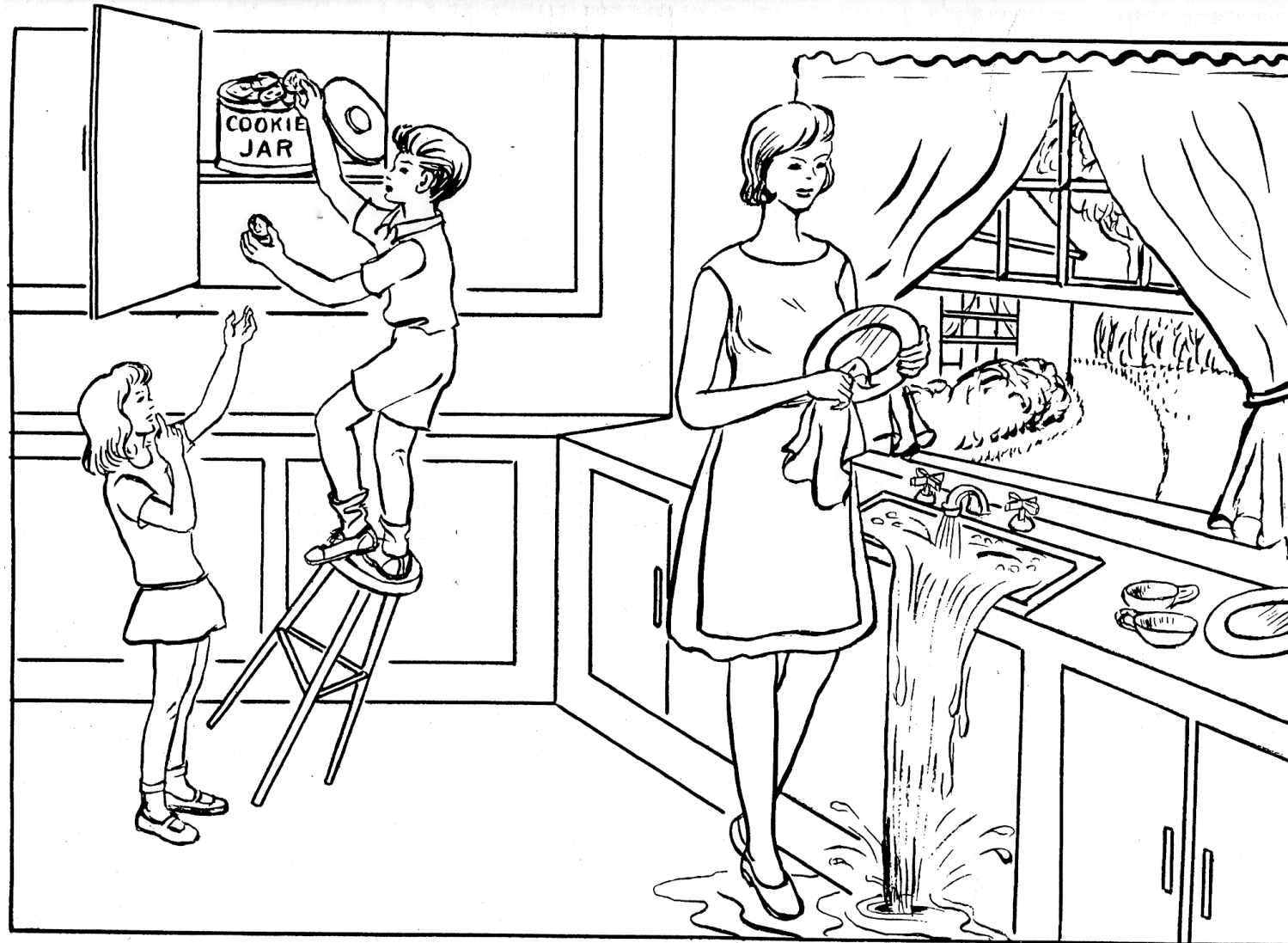
---

# Participants

	AD (N=49)	lvPPA (N=28)	FTLD-tau (N=20)	HC (N=35)	p value
Age	62.6 (7.6)	63.2 (7.1)	67.8 (7.0)	64.6 (7.0)	0.052
Education	16.0 (2.4)	16.2 (3.4)	15.6 (3.2)	15.7 (2.5)	0.832
Sex					0.741
Female	28 (57.1%)	13 (46.4%)	9 (45.0%)	18 (51.4%)	
Male	21 (42.9%)	15 (53.6%)	11 (55.0%)	17 (48.6%)	
Disease duration	3.7 (2.4)	3.5 (1.9)	3.4 (2.1)	NA	0.9
MMSE (0-30)	20.3 (5.0)	23.6 (4.5)	25.6 (3.7)	29.1 (1.1)	< 0.001



# The Cookie Theft picture (BDAE, Goodglass & Kaplan, 1972)



# Automatic part-of-speech (POS) tagging

```
62_pSubject4_4_cookie.POSout x
1 In in ADP IN prep
2 the the DET DT det
3 picture picture NOUN NN pobj
4 there there ADV EX expl
5 seems seem VERB VBZ ROOT
6 to to PART TO aux
7 be be VERB VB xcomp
8 a a DET DT det
9 middle middle ADJ JJ npadvmod
10 aged aged ADJ JJ amod
11 woman woman NOUN NN attr
12 and and CCONJ CC cc
13 a a DET DT det
14 children child NOUN NNS conj
15 uh uh INTJ UH intj
16 two two NUM CD nummod
17 children child NOUN NNS appos
18 a a DET DT det
19 boy boy NOUN NN conj
20 and and CCONJ CC cc
21 a a DET DT det
22 girl girl NOUN NN conj
23 um um INTJ UH ROOT
24 in in ADP IN ROOT
25 a a DET DT det
26 suburban suburban ADJ JJ amod
27 home home NOUN NN pobj
28 um um INTJ UH ROOT
29 The the DET DT det
30 woman woman NOUN NN poss
31 's be VERB VBZ case
32 by by ADP IN prep
33 uh uh INTJ UH intj
34 a a DET DT det
```

- spaCy (Honnibal & Johnson, 2015)
- Count of POS categories per 100 words
- Lexical measures
  - Concreteness (Brysbaert et al. 2014)
  - Semantic ambiguity (Hoffman et al. 2013)
  - Word frequency (Brysbaert & New, 2009)
  - Familiarity (Brysbaert et al. 2018)
  - Age of Acquisition (AoA; Brysbaert et al. 2018)
- Lexical diversity (MATTR; Covington & McFall 2010)
- Number of characters, phonemes, syllables with the CMU pronouncing dictionary

# Acoustic, durational feature extraction

SAD

Speech (250ms) and silence (150ms) segments

Durational

Mean duration of speech and silence segments  
Total speech and pause time  
Total pause count & total speech segment count  
Pause rate per minute  
Speech rate and articulation rate

Pitch

10<sup>th</sup> to 90<sup>th</sup> pitch percentiles from speech segments  
Normalized to St:  $\log_2(\text{pitch}/10^{\text{th}})*12$

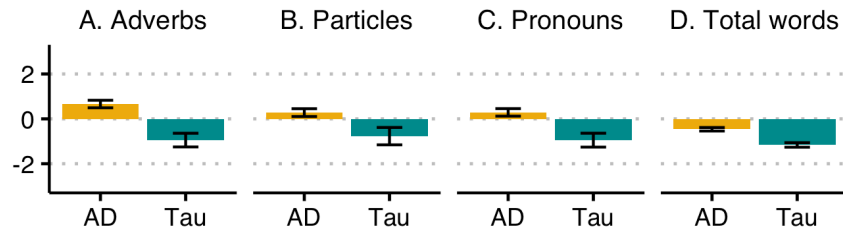
03

**RESULTS**

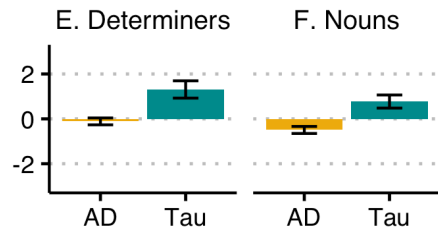
---

# POS counts

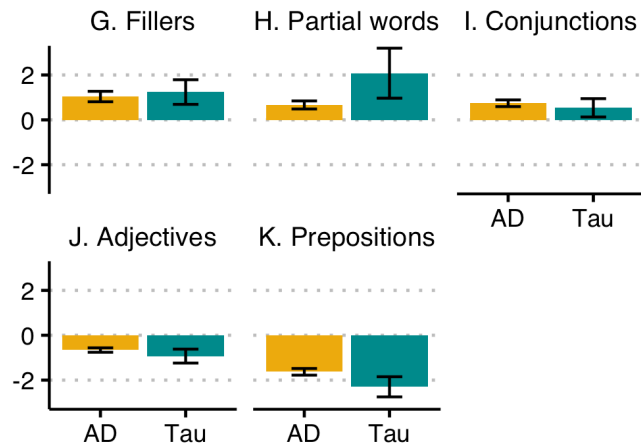
## AD > FTLD-tau



## AD < FTLD-tau



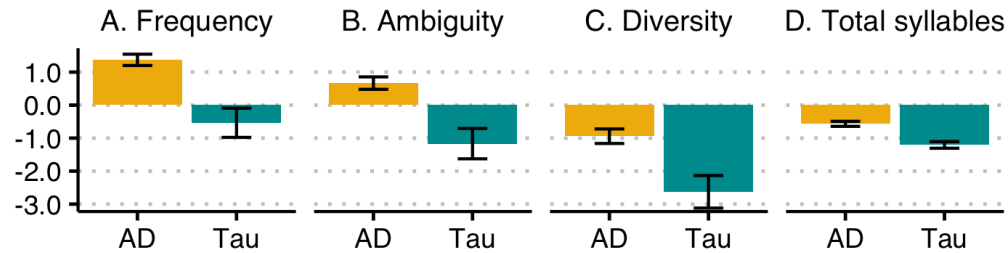
## AD = FTLD-tau



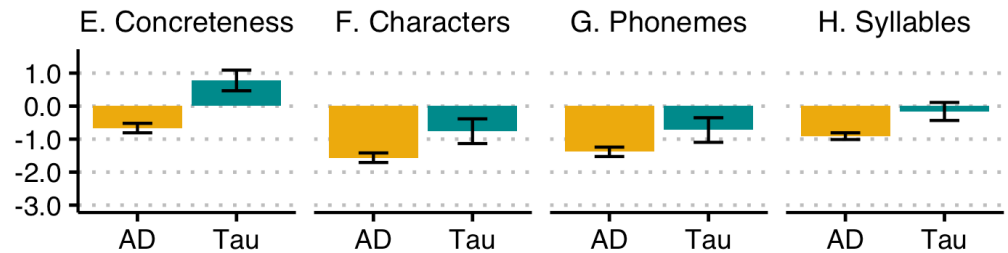
	AD (N=77)	Tau (N=20)	HC (N=35)	p value
Adverbs	7.3 (3.2)	3.8 (3.0)	5.8 (2.2)	< 0.001
Particles	3.7 (2.2)	2.2 (2.6)	3.3 (1.5)	0.023
Pronouns	8.1 (3.3)	5.3 (3.1)	7.5 (2.3)	0.002
Total words	136.6 (61.6)	74.8 (40.2)	177.6 (88.5)	< 0.001
Determiners	13.3 (3.7)	17.2 (4.8)	13.6 (2.8)	< 0.001
Nouns	18.1 (6.4)	23.9 (6.1)	20.4 (4.6)	< 0.001
Fillers	7.6 (5.3)	8.1 (6.3)	4.9 (2.6)	0.029
Partial words	1.2 (1.5)	2.6 (4.8)	0.6 (1.0)	0.044
Conjunctions	5.8 (2.6)	5.3 (3.7)	4.3 (2.0)	0.034
Adjectives	4.0 (2.0)	3.4 (3.2)	5.5 (2.3)	< 0.001
Prepositions	7.5 (2.5)	6.2 (3.9)	10.7 (1.9)	< 0.001
Verbs	22.9 (5.2)	20.5 (4.0)	22.5 (3.6)	0.126
Ratio of content words (%)	52.2 (5.8)	51.6 (9.0)	54.3 (4.1)	0.177

# Lexical measures of content words

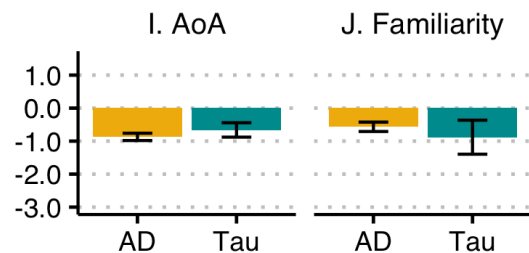
## AD > FTLD-tau



## AD < FTLD-tau



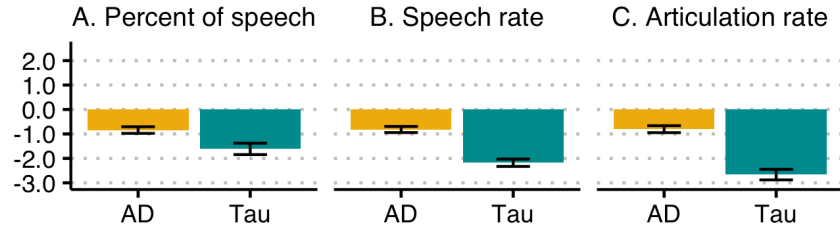
## AD = FTLD-tau



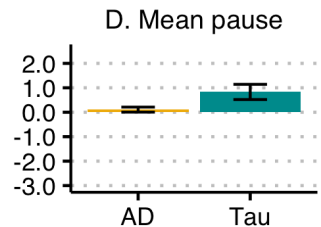
	AD (N=77)	Tau (N=20)	HC (N=35)	p value
Frequency	4.5 (0.3)	4.1 (0.4)	4.2 (0.2)	< 0.001
Ambiguity	2.0 (0.1)	1.9 (0.1)	2.0 (0.0)	< 0.001
Diversity	0.8 (0.1)	0.7 (0.1)	0.8 (0.0)	< 0.001
Total Syllables	158.1 (73.2)	88.2 (49.1)	220.0 (109.3)	< 0.001
Concreteness	3.0 (0.3)	3.4 (0.4)	3.2 (0.3)	< 0.001
Characters	4.3 (0.3)	4.5 (0.4)	4.6 (0.2)	< 0.001
Phonemes	3.4 (0.2)	3.5 (0.3)	3.7 (0.2)	< 0.001
Syllables	1.4 (0.1)	1.4 (0.1)	1.4 (0.1)	< 0.001
AoA	4.5 (0.2)	4.5 (0.2)	4.7 (0.2)	< 0.001

# Acoustic measures

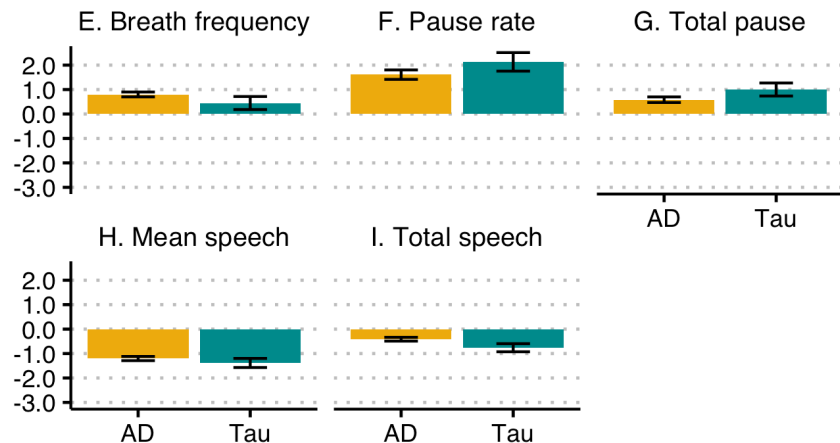
## AD > FTLD-tau



## AD < FTLD-tau



## AD = FTLD-tau



	AD (N=77)	Tau (N=20)	HC (N=35)	p value
Breath frequency per minute	25.8 (5.1)	23.8 (6.9)	21.2 (5.8)	< 0.001
Pause rate per minute	50.0 (18.2)	55.6 (18.2)	32.7 (10.7)	< 0.001
Total pause time (sec)	31.0 (13.1)	36.5 (15.7)	23.3 (13.2)	0.002
Total speech time (sec)	40.0 (15.6)	32.0 (17.2)	49.5 (23.1)	0.002
Mean speech duration (sec)	1.3 (0.4)	1.2 (0.5)	2.0 (0.6)	< 0.001
Mean pause duration (sec)	1.1 (0.6)	1.6 (0.9)	1.0 (0.7)	0.009
Speech rate (wpm)	115.4 (40.0)	66.0 (24.6)	145.2 (36.3)	< 0.001
Articulation rate (sps)	4.0 (0.8)	2.8 (0.6)	4.4 (0.6)	< 0.001
Percent of speech time (%)	56.0 (15.9)	45.7 (14.0)	67.3 (13.4)	< 0.001
Pitch range (st)	5.1 (2.3)	4.3 (1.9)	5.7 (2.6)	0.130
Total time (sec)	71.0 (17.4)	68.5 (24.0)	72.8 (26.1)	0.759

# Comparison of amnestic and non-amnestic AD

- Out of 33 features, amnestic and non-amnestic AD groups only differed in 4 features:

	AD (N=49)	lvPPA (N=28)	p value
Determiners	12.4 (3.4)	14.9 (3.8)	0.004
Fillers	6.5 (5.0)	9.6 (5.2)	0.013
Particles	4.1 (2.2)	3.0 (2.2)	0.039
Ratio of content words (%)	53.5 (5.4)	49.9 (5.8)	0.008



04

**DISCUSSION**

---

# Semantic knowledge impairment in AD

- AD produced content words that were **less concrete, more ambiguous, more frequent and shorter** than the other groups.
- Also, amnesic and non-amnesic AD patients did not significantly vary in these measures.

# AD vs. FTLD-tau

- Previous findings: AD produce more pronouns with a lower lexical diversity compared to MCI or HC.
- Pronouns: HC = AD > FTLD-tau
- Lexical diversity: HC > AD > FTLD-tau
- Pronouns and lexical diversity are helpful in distinguishing AD speech, but those are **not the most robust, distinctive features of AD speech.**

# Pause duration

- The total pause time and pause rate:  $AD = FTLD\text{-tau} < HC$
- Duration of speech segments and total speech time:  
 $AD = FTLD\text{-tau} < HC$
- Patients with neurodegenerative disease in general showed similar patterns (Nevler, Ash, Irwin, Liberman, & Grossman, 2019; Nevler et al., 2017).
- These features seem to be important and useful measures in distinguishing neurodegenerative **patients' speech from controls**.

# Selected references

- Goodglass, H., Kaplan, E., & Weintraub, S. 1983. Boston Diagnostic Aphasia Examination. Lea & Febiger.
- Ratnavalli E, Brayne C, Dawson K, Hodges JR. 2002. The prevalence of frontotemporal dementia. *Neurology* 58: 1615–21.
- Honnibal & Johnson. 2015. An improved non-monotonic transition system for dependency parsing. *EMNLP 2015: Conference on empirical methods in natural language processing*, 1373–1378.
- Brysbaert et al. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911.
- Hoffman et al. 2013. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45(3), 718–730.
- Brysbaert & New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990.
- Brysbaert et al. 2018. Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, 51(2), 467–479.

# Thank you!

---

Correspondence: [csunghye@ldc.upenn.edu](mailto:csunghye@ldc.upenn.edu)