Can measures of processing complexity predict progressive aphasia from speech?

Kathleen C. Fraser[1], Marten van Schijndel[2], Naida Graham[3,4], Elizabeth Rochon[3,4], Sandra Black[5,6]

[1]Department of Computer Science, University of Toronto, [2]Department of Linguistics, Ohio State University,
[3]Department of Speech Language Pathology, University of Toronto, [4]Toronto Rehabilitation Institute,
[5]Department of Medicine (Neurology), University of Toronto, [6]Sunnybrook Health Sciences Centre

Introduction

Primary progressive aphasia (PPA) is a progressive language impairment without other notable cognitive impairment [1].

- **Semantic variant (svPPA)**
  - Word-finding difficulty
  - Empty speech
  - Spared fluency, grammar
- **Nonfluent variant (nfPPA)**
  - Effortful, nonfluent speech
  - Agrammatism
  - Spared single-word comprehension

Previous computational work did not uncover syntactic complexity differences between subtypes, or consider word use in context [2].

Can we distinguish between PPA patients and controls, and between the two subtypes?

- **New approach:** use contextual features (*n*-grams) and psycholinguistic measures of processing complexity

Data

Narrative speech elicited using *Cinderella* story-telling task.

	svPPA	nfPPA	Control
n	11	17	23
MMSE	24.8	25.2	29.2
Age	65.9	53.5	67.8
Education	17.5	14.6	16.5
Sex (M/F)	8/3	10/7	12/11

Methods

Data split **50-50** into development and testing partitions.

Logistic mixed regression used to:
- Separate control from PPA narratives
- Separate svPPA from nfPPA narratives

Evaluation baseline:
- Random intercepts for each word
- Fixed effects: sentence position, word length, word frequency, all 2-way interactions

Features

Feature	Motivation
Sentence position (proxy for sentence length)	Expect nfPPA patients will use shorter sentences due to reduction in fluency.
Word length in characters	Expect nfPPA patients will use shorter words.
Word frequency (obtained from SUBTL norms [3])	Expect svPPA patients will use more high-frequency words as a result of word-finding difficulty.
5-gram probability (obtained from Gigaword 4.0 [4])	Expect PPA patients will combine words in less probable combinations.
Syntactic surprisal [5]	Expect nfPPA narratives will show higher syntactic surprisal due to syntactic difficulties.
Lexical surprisal [5]	Expect svPPA narratives will show higher lexical surprisal due to semantic difficulties.
Entropy reduction [6]	Expect PPA patients may show increase.
Embedding depth [7]	Expect nfPPA sentences will show shallower embedding, reflecting syntactic simplification.

Results

PPA vs Controls
- 5-grams improved accuracy ($p < 0.001$)
- Syntactic surprisal and entropy reduction helped in dev set, but not in test set ($p > 0.1$)
- PPA patients use:
  - Shorter sentences
  - High-frequency words in unusual lexical contexts (interaction effect)
  - Short words which are also low-frequency (interaction effect)

svPPA vs nfPPA
- 5-grams improved accuracy ($p < 0.001$)
- Syntactic surprisal plus all 2-way interactions also improved accuracy ($p = 0.012$)
- Embedding depth helped in dev set but not in test set ($p > 0.1$)
- nfPPA patients use:
  - Longer sentences, possibly due to repairs and false starts
  - Long, low-frequency words (interaction effect)
- svPPA patients use:
  - More contextually probable words
  - High-frequency words late in the sentence (interaction effect)

Weak evidence for syntactic surprisal and embedding depth effects.

Strongest predictors related to word probability and sentence length.

References:

[1] Gorno-Tempini, M. L. et al. Cognition and anatomy in three variants of primary progressive aphasia. Annals Neurol. 55, 335–346 (2004). **[2]** Fraser, K. C. et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex 55, 43–60 (2014). **[3]** Brysbaert M. et al. 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 Res. Methods 41(4), 977—90 (2009). **[4]** Parker R. et al. English Gigaword LDC2009T13 (2009). **[5]** Roark, B. et al. Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. Proc EMNLP 324–333 (2009). **[6]** Hale, J. Uncertainty about the rest of the sentence. Cognitive Science 30, 609–642 (2006). **[7]** Chomsky, N. et al. Introduction to the formal analysis of natural language. In Handbook of Mathematical Psychology, 269—321 (1963).

Acknowledgements:

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1343012, the Natural Sciences and Engineering Research Council of Canada Grant No. RGPIN-2014-06020, and the Canadian Institutes of Health Research Grant No. MOP-82744.