

Syntactic Rule Frequency as a measure of Syntactic Complexity: Insights from Primary Progressive Aphasia

Neguine Rezaii (Harvard Medical School), Rachel Ryskin (University of California), Kyle Mahowald (University of California), Bradford Dickerson (Harvard Medical School), Edward Gibson (Massachusetts Institute of Technology)

We investigate syntactic processing in the language of patients with primary progressive aphasia (PPA), a neurodegenerative clinical syndrome where language is the predominant initial impairment. Depending on the primary region of brain atrophy, PPA can have different psycholinguistic presentations. The nonfluent variant of PPA (nfvPPA) is characterized by simple and impoverished syntactic structures and/or effortful speech. In contrast, the other two variants of PPA are described based on lexico-semantic deficits: Individuals with the logopenic variant of PPA (lvPPA) exhibit difficulty with sentence repetition and lexical retrieval. In the semantic variant of PPA (svPPA), difficulties in object naming and word comprehension are the hallmark of the disorder (Gorno-Tempini 2011). In this work, we aim to use the frequency of syntactic rules as a measure of syntactic complexity based on the psycholinguistic literature suggesting that language comprehension is sensitive to the probability distribution of words and syntactic rules. The consequence of this finding is that the complexity of any utterance corresponds to the probability of the utterance in context. Thus one production complexity metric is one based on the frequency of combinatory syntactic rules: compared to control participants, nfvPPA patients might have relatively weaker access to lower frequency syntactic rules.

Methods. Clinical and language assessments and MRI scans were used to characterize 79 patients with PPA and its subtypes (29 nfvPPA, 26 lvPPA, and 24 svPPA). We also included 51 age matched healthy controls. Participants were asked to describe a drawing of a family having a picnic from the Western Aphasia Battery–R (Kertesz, 2007) using as many full sentences as they could. The recorded responses were transcribed by a researcher blind to the subtypes. Disfluencies were removed from the analyses. These language samples were then parsed using the Stanford Probabilistic Context-Free Grammar (PCFG) parser (Klein and Manning, 2003). We examined binary syntactic rules using the output from the dependency grammar parse. For this metric, we take each dependency in the dependency structure as a separate rule (e.g., amod-NOUN) (Figure 1).

Results. Figure 2 shows the 20 most common binary rules for binary dependency grammar in nfvPPA and healthy controls. Fitting a maximal mixed effect model with random effects for subject and sentence that predicts log syntactic rule frequency with patient subtype and sentence length as predictors, we found a main effect of *patient subtype* while controlling for sentence length, with higher binary syntactic rules likely to occur in nfvPPA than other subtypes ($\beta=0.18$, $SE=0.06$, $t(16234)=2.78$, $p<0.01$). To better control for the effect of sentence length, we sampled sentences from each of the four groups so that all groups would have similar sentence length distributions. We continued to find a main effect of *patient subtype*, with nfvPPA patients producing higher frequency syntactic rules ($\beta=0.29$, $SE=0.06$, $t(16234)=4.40$, $p<0.001$).

Conclusions. Language production in nfvPPA is characterized by use of high frequency syntactic rules, when compared with control, svPPA, and lvPPA language production. Our results suggest a syntactic rule-specific locus of impairment, in line with proposals of a syntax-specific component of language production (Garrett, 1980; Bock, 1995), perhaps localized to a particular brain area (Fedorenko, Williams & Ferreira, 2018).

Figure 1. Constructing dependency rules. Sentences were parsed using the Stanford Lexicalized Parser Package (v3.9.2). Dependency heads, underlined, were identified by the parser. The combination of a head and its dependent was considered as a binary dependency rule.

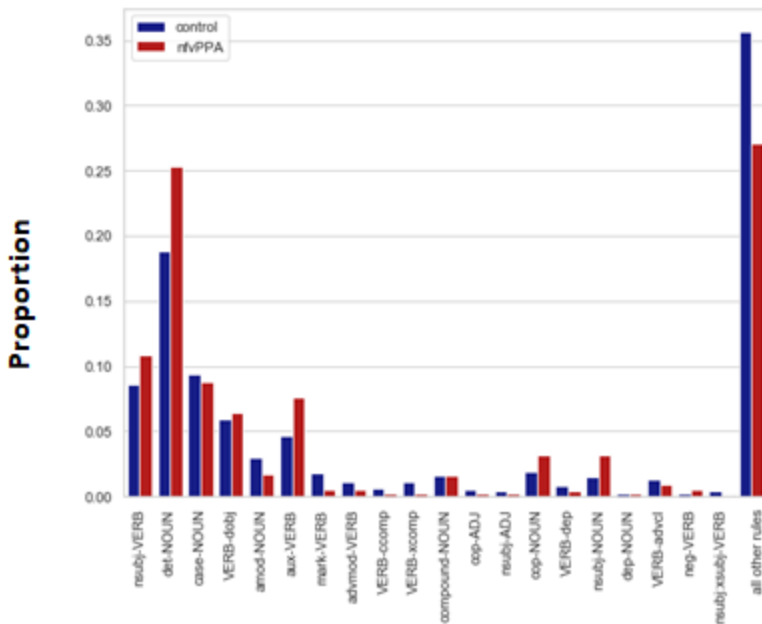
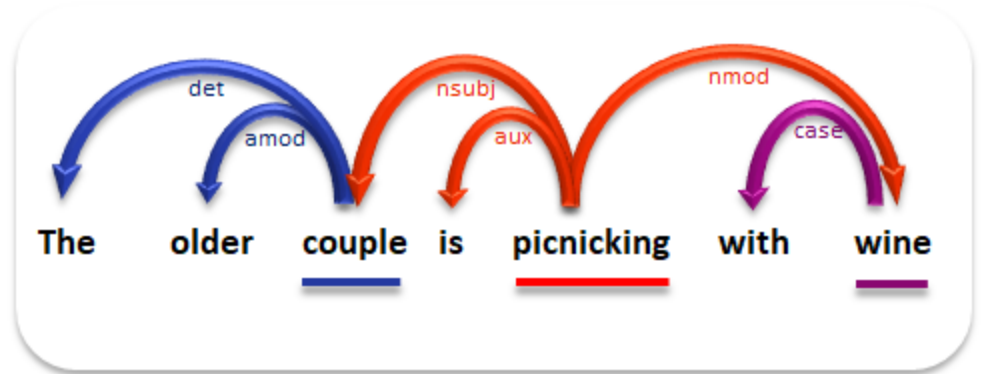


Figure 2. The proportion of 20 most common binary dependency grammar rules in nfvPPA and healthy controls. The last two bars show the proportion of all other binary dependency grammar rules combined.

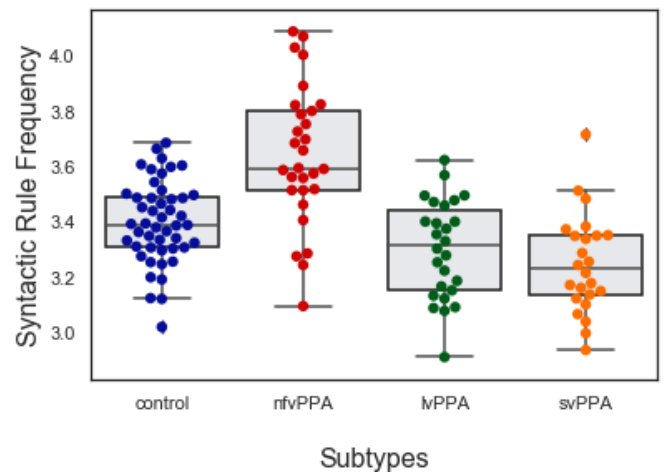


Figure 3. The box plots of syntactic rule frequency for the four groups

Reference

Fedorenko, E., Williams, Z.M. & Ferreira, V.S. (2018). Remaining puzzles about morpheme production in the posterior temporal lobe. *Neuroscience*, 392, 160-163.

Kertesz A. (2007). *Western Aphasia Battery (Revised)*. San Antonio: PsychCorp;

Gorno-Tempini et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, 76(11), 1006-1014

Klein, D. & Manning, C. D. (2003) Accurate unlexicalized parsing. *ACL*. 423–430

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production*, Vol. 1, (pp. 177-220). London: Academic Press.

Bock, J. K. (1995). Sentence production: From mind to mouth. In J. L. Miller, & P. D. Eimas (Eds.), *Handbook of perception and cognition*.