

Towards an automated method for quantifying pedantic language in children with Autism Spectrum Disorder

Grace Lawley, B.A.¹, Jan van Santen, Ph.D.¹, Jill Dolata Ph.D., CCC-SLP^{1,2}, Steven Bedrick, Ph.D.¹, & Eric Fombonne, M.D.³

OHSU Research Week
June 12th, 2020

¹ Center for Spoken Language Understanding, OHSU; ² Department of Pediatrics, OHSU; ³ Department of Psychiatry, OHSU



Autism Spectrum Disorder (ASD)

Core characteristics

- Restricted, repetitive interests
- Difficulties with social communication

These language difficulties can appear in few different forms...

Pedantic Speech

- Inappropriately formal, adult-like, overly specific
- A hole in a sock vs. “a temporary loss of knitting” (Wing, 1981)

Pedantic Speech in ASD

- Definitions are vague
- Various interpretations
- *Can be pedantic in many different ways*

Vocabulary choice

- “I ate crustaceans for lunch”
- “I ate shrimp for lunch”

Level of detail

- “First you need to check the expiration date, then get the can opener...”
- “Mix in the tomato paste”

How is this currently measured?

Autism Diagnostic Observation Schedule (ADOS)

- Standard ASD assessment tool
- Series of semi-structured, examiner led activities
- Coding scheme for behaviors characteristic of ASD

A4. Stereotyped/Idiosyncratic Use of Words or Phrases

“Use of words or phrases tends to be more repetitive or formal than that of most individuals at the same level of expressive language, but not obviously odd...”

Limitations

- Subjective
- Inconsistent across examiners



Good candidate for automation

Can we quantify how “pedantic” a child’s vocabulary choice is?

- We are interested in the outlier words since pedantic words will most likely be infrequent ones (Prud’hommeaux et al., 2010)
- Calculate a pedantic-index score for a given word using two reference corpora:
 - Wall Street Journal (WSJ) corpus = adult-like speech
 - Child Language Data Exchange System (CHILDES) corpus = child-like speech
- Then calculate an overall **pedantry score** for each kid

Data

- Transcribed ADOS sessions of 4-8 year old kids
- Four diagnostic groups
 - **Typically Developing (TD)**, n = 44
 - **ASD with Language Normal (ALN)**, n = 23
 - **ASD with Language Language Impairment*** (ALI), n = 20
 - **Specific Language Impairment*** (SLI), n = 17
- All participants
 - Full-scale IQ > 70, MLU > 3.0

Hypothesis

- Some (but not all) kids in the the ASD group will be pedantic
- ALN group will be more pedantic than both the ALI and TD groups
- SLI group will be the least pedantic

**Language impairment determined by a CELF Core Language Score lower than one standard deviation below the mean (standard score < 85)*

Method

Each kid is represented as a set of unique words (types) $\{w_0, \dots, w_n\}$

Calculate a *pedantic index* for each word

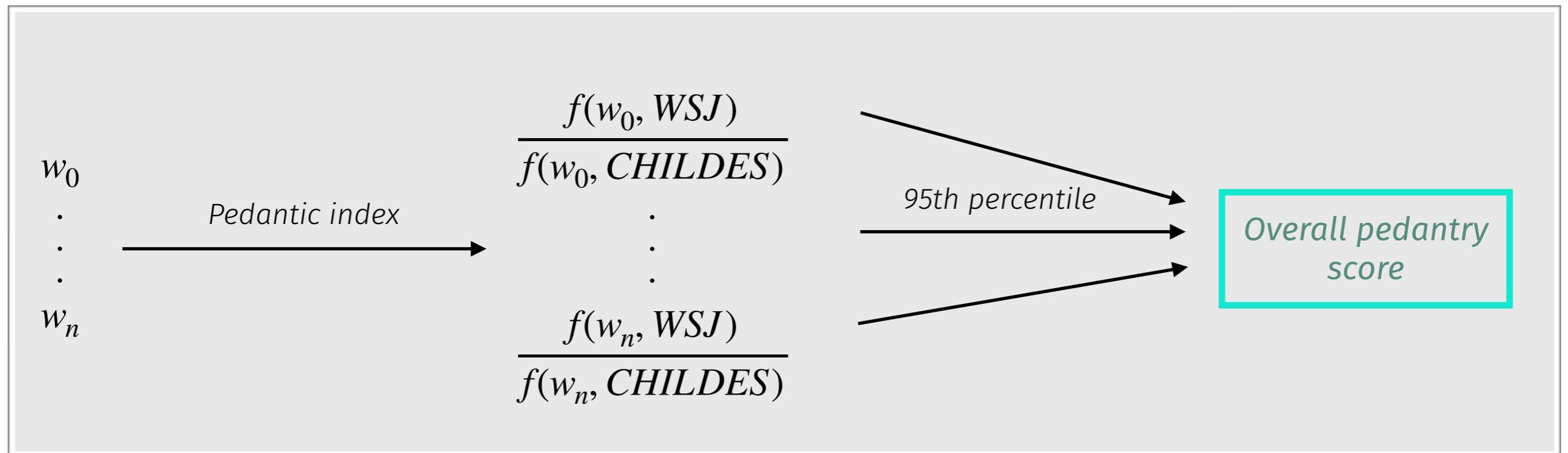
- For each word a given kid used, calculate the frequency of the word in the two reference corpora
 - Wall Street Journal Corpus = adult-like speech
 - CHILDES Corpus = child-like speech
- Transform frequencies; motivation:
 - A majority of the words will have very small frequencies that are close to zero (Zipf's law)
 - The rare words will have frequencies that are very, very close to zero
- Pedantic index of a word is the ratio of the transformed WSJ frequency over the transformed CHILDES frequency
 - Higher score = more adult-like; more frequent in WSJ than CHILDES
 - Lower score = more child-like; more frequent in CHILDES than WSJ

$$P_{index}(w_i) = \frac{f(w_i, WSJ)}{f(w_i, CHILDES)}$$

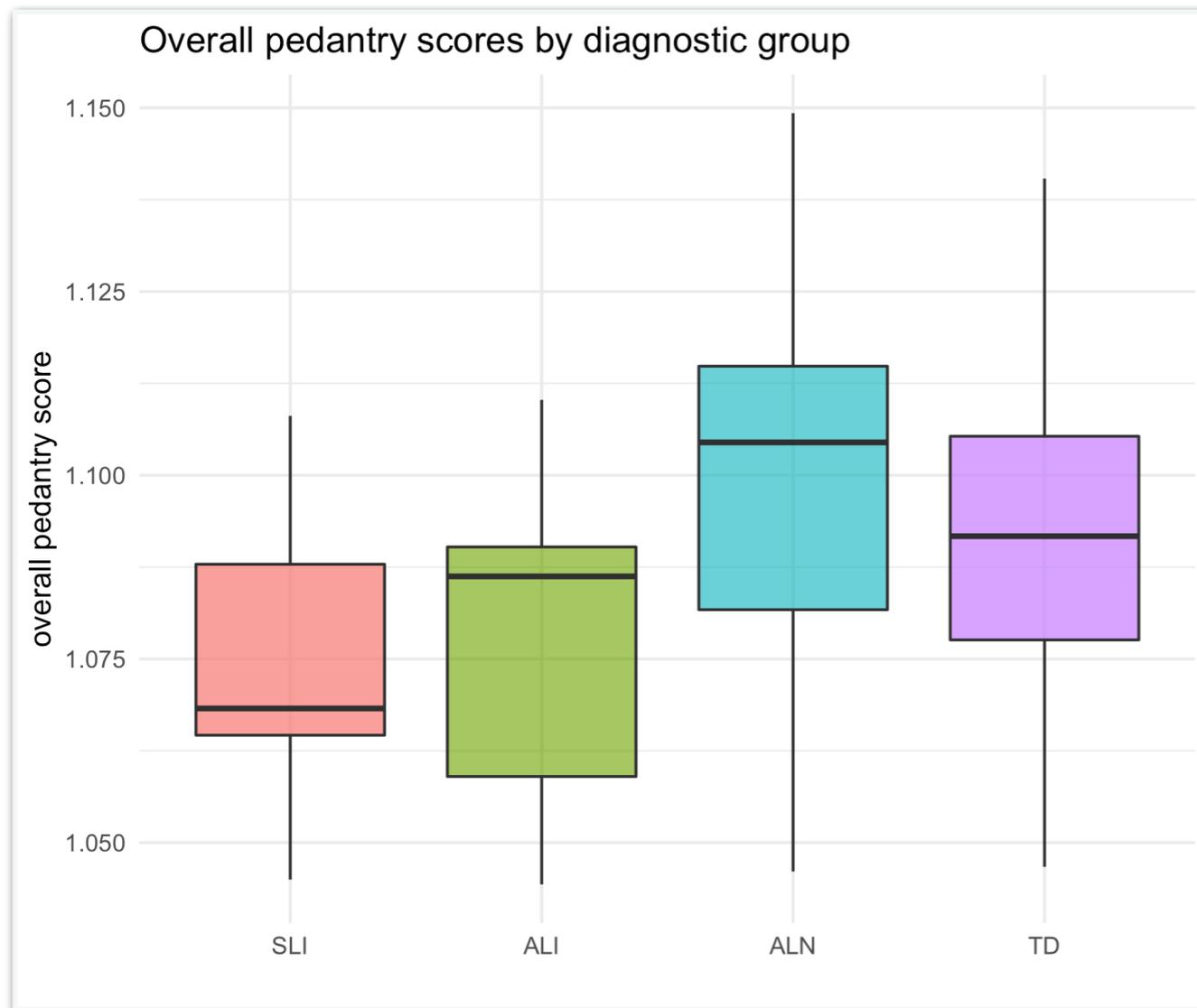
Method

- A child's **overall pedantry score** is the 95th percentile of the transformed frequencies for every unique word they said

For each kid...



Results



	Overall pedantry score mean (sd)
SLI	1.07 (0.0188)
ALI	1.08 (0.0189)
ALN	1.10 (0.0279)
TD	1.09 (0.0214)

- Kruskal-Wallis rank sum test to check for significant effect of dx group
 - chi-squared = 16.35, df = 3, $p < 0.001$
- Post-hoc Dunn's tests to compare amongst dx levels
 - SLI < TD and SLI < ALN (p adj. < 0.001)
 - ALI < ALN (p adj. = 0.013)
 - ALI < TD (p = 0.044)

Next Steps

- Control for differences in age, FSIQ, gender, etc.
- Validate that the *pedantic index and overall pedantry score* are capturing what we think they are capturing
- Are there other ways to quantify whether speech is pedantic?
 - The fillers “um” and “uh” are important pragmatic markers in conversation — beneficial to listener’s on-line speech comprehension (Fox Tree, 2001)
 - High usage rates of “um” alongside low usage rates of “uh” might contribute to individuals with ASD sounding pedantic (Irvine et al., 2016, Parish-Morris et al., 2017)

Thank you

References

Fox Tree, J.E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition*, 29(2), 320-326. <https://doi.org/10.3758/BF03194926>

Irvine, C. A., Eigsti, I.-M., & Fein, D. A. (2015). Uh, Um, and Autism: Filler Disfluencies as Pragmatic Markers in Adolescents with Optimal Outcomes from Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 46(3), 1061–1070. <http://doi.org/10.1007/s10803-015-2651-y>

Parish-Morris, J., Liberman, M. Y., Cieri, C., Herrington, J. D., Yerys, B. E., Bateman, L., et al. (2017). Linguistic camouflage in girls with autism spectrum disorder, 8(1), 48. <http://doi.org/10.1186/s13229-017-0164-6>

Prud'hommeaux, E. T., van Santen, J. P. H., Black, L. M., & Roark, B. (2010). Automatic detection of idiosyncratic word use in autism spectrum disorders.

Wing, L. (1981). Asperger's Syndrome: a Clinical Account. *Psychological Medicine*, 11, 115–129.

This research was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award R01DC012033