Lexical diversity

- the most widely used measure of verbal skills
- is known to be lower in discourses by people with aphasia (PWA) as compared to healthy speakers (Fergadiotis & Wright, 2011)
- Current study — apply various lexical diversity measures to speech samples from Russian speakers with and without aphasia

Lexical diversity measures

Type-token ratio (TTR)
number of words in a sample divided by the number of lexemes
+ easy to calculate
- very dependent on sample size

Measure of textual lexical diversity (MTLD)
mean length of sequential word strings in a text that maintain a given TTR value (0.72)
+ independent of sample size

Moving average type-token ratio (MATTR)
average of TTR measured in a window of n: word 1 - word n, word 2 - word (n+1), ...
  - window size 10 - can detect such properties of the text, as frequent repetitions
  - window size 100 - not sensitive to repetitions
+ independent of sample size
(Covington & McFall, 2010)

Russian CliPS corpus
- oral retellings of the Pear film (Chafe, 1980; Khudyakova et al., 2016)
- comparable samples and clear story line
- annotation on multiple levels, including lexical

Results

Great variability in text length

MATTR with window =10
- Significant difference between healthy speakers and all PWA groups
- Significant difference between speakers with dynamic aphasia and speakers with fluent types of aphasia (sensory and acoustic-mnestic)

MTLD and MATTR with window=100
- Significant difference between healthy speakers and all PWA groups
- No significant difference between speakers with different types of aphasia

Discussion

PWA have lower lexical diversity scores than healthy speakers: replicates previous findings

A specific measure of lexical diversity – MATTR with a smaller window size can detect differences between texts by speakers with dynamic aphasia and fluent aphasia types.

References