Verb and Sentence Impairment in Aphasia: Insights from Cluster Analysis

Akinina, Yu. [1, 2] jakinina@hse.ru, Bastiaanse, R.[1, 2], Buivolova, O.[2], Iskra, E.[2, 3], Soloukhina,O.[2]

[1]University of Groningen, Groningen, Netherlands; [2]National Research University Higher School of Economics, Moscow, Russia; [3]Center for Speech Pathology and Neurorehabilitation, Moscow, Russia

1. Introduction

Verb and sentence deficits are prevalent in aphasia, in comprehension and production, both in fluent and non-fluent aphasia (Bastiaanse et al., 2003; Cho-Reyes & Thompson, 2012).

However, it is not entirely clear if there are consistent patterns of verb and sentence impairment profiles, and how these patterns correspond to aphasia types.

We try a data-driven **exploratory** approach to these questions, applying *k*-means cluster analysis to the results of four subtests of Verb and Sentence Test, Russian version (VAST-ru) obtained from a large group of individuals with aphasia (IWA).

2. Materials and Procedure

VAST-ru is an adaptation of VAST (Bastiaanse et al., 2000) to the Russian language and consists of several subtests that assess production and comprehension of verbs at single word and sentence level.

For this study, we use four comprehension subtests of VAST-ru.

Minimal Pairs (28 items): an oral same-different judgment task that assesses vowel discrimination at the end of the word. In Russian, it is crucial for case, and consequently for morpho-syntactic processing.

Verb Comprehension (40 items): a word-to-picture matching task that assesses comprehension of verb semantics.

Sentence Comprehension (40 items): a sentence-to-picture matching task that assesses comprehension of semantically reversible sentences.

Plausibility Judgment (60 items): an oral plausible-implausible judgment task that estimates the ability of morpho-syntactic parsing of semantically irreversible sentences.

All the subtests were presented via a tablet application (Ivanova et al., 2016); the responses were registered automatically.

3. Control group data

Control group:
22 non brain-damaged participants, 15 female, mean age 46.73 yo (± 14.74, range = 25-77), education level from professional to advanced university degree.

	Mean	SD	Median	Min	Max
Minimal Pairs	0,99	0,01	1	0,96	1
Verb Comprehension	1	0,01	1	0,98	1
Sentence Comprehension	0,98	0,03	0,98	0,92	1
Plausibility Judgement	0,99	0,01	0,99	0,97	1

Descriptive statistics of control group performance on the four VAST-ru subtests

4. Analysis 1

Participants

32 IWA with various etiology, 13 female, mean age 56.47 yo (±11.66, range 26 – 73), mean time post onset 31.44 m (± 35.18, range = 2.07 – 114.17; data missing for two participants), education level from secondary school to university degree.

12 with fluent, 17 with non-fluent aphasia, 3 with mixed (fluent and non-fluent) aphasia.

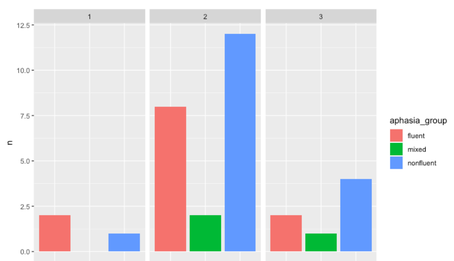
Data analysis 1

k-means clustering method: an algorithm that minimizes the squared error between the empirical mean of a cluster and the points of a cluster in a multi-dimensional space. The number of clusters is user-specified (Jain, 2010).

Preliminary analysis: three clusters to reflect three-fold partitioning of our patient sample into fluent, non-fluent and mixed group.

Results:

Patient partitioning into three clusters:



Cluster	N	Minimal Pairs, Mean (SD)	Verb Comprehension, Mean (SD)	Sentence Comprehension, Mean (SD)	Plausibility Judgment, Mean (SD)
1	3	0.92 (0.02)	0.72 (0.27)	0.46 (0.08)	0.68 (0.14)
2	22	0.96 (0.08)	0.97 (0.04)	0.94 (0.06)	0.91 (0.07)
3	7	0.97 (0.05)	0.9 (0.09)	0.68 (0.13)	0.73 (0.16)

Results of the four VAST-Ru subtests, by cluster

5. Analysis 2

A problem with analysis 1

In *k*-means, the number of clusters is user-specified. An automatic algorithm for choosing the optimal number of clusters for a given dataset is needed.

Data analysis 2

Before clustering the data with k-means, we used three standard methods for choosing the optimal number of clusters for a given dataset: the elbow plot, the silhouette method and the gap statistic method.

The analysis was performed on a larger group of participants. Participants with mixed aphasia were excluded from the analysis.

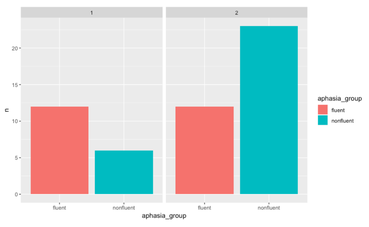
Participants

53 IWA with various etiology, 25 female, mean age 54.57 yo (±11.6, range = 20-72), mean time post onset 30.18 m (± 35.31, range = 1.77 – 176.46), education level from secondary to university degree.

24 with fluent, 29 with non-fluent aphasia.

Results:

Patient partitioning into two clusters:



Chi-squared test: a tendency to significance (p = 0.051).

The optimal number of clusters determined by all three methods applied was **two**.

Cluster	N	Minimal Pairs, Mean (SD)	Verb Comprehension, Mean (SD)	Sentence Comprehension, Mean (SD)	Plausibility Judgement, Mean (SD)
1	18	0.93 (0.08)	0.87 (0.15)	0.58 (0.12)	0.72 (0.15)
2	35	0.96 (0.07)	0.97 (0.05)	0.92 (0.07)	0.89 (0.07)

Results of the four VAST-Ru subtests, by cluster

6. Discussion

Summary

- Analysis 1 resulted in partitioning of IWA into three groups that did not correspond to aphasia types;
- Analysis 2 was performed on an extended IWA group and with user-independent determination of number of clusters;
- In Analysis 2, the results of Analysis 1 were not replicated;
- Analysis 2 partitioned the IWA group into two clusters;
- In both clusters, there were fluent and non-fluent IWA;
- The difference between aphasia types distribution across types approached significance (Chi-squared test, p = 0.051);
- Performance on all subtests except Minimal Pairs was significantly worse in cluster 1 (independent samples t-test, $p < 0.05$).

Conclusions

- The *k*-means analysis split the IWA into two groups based on severity of impairment in verb and sentence comprehension;
- There is no one-to-one correspondence between aphasia type and patterns of comprehension impairment;
- However, a tendency was observed: fluent IWA were, generally, more impaired in comprehension than non-fluent IWA.
- An individual-level analysis is needed to see if the patterns differ within clusters;
- The results should be validated and refined with other clustering techniques, such as hierarchical cluster analysis;
- Data of the production subtests of VAST-ru will be taken into account in further analyses.