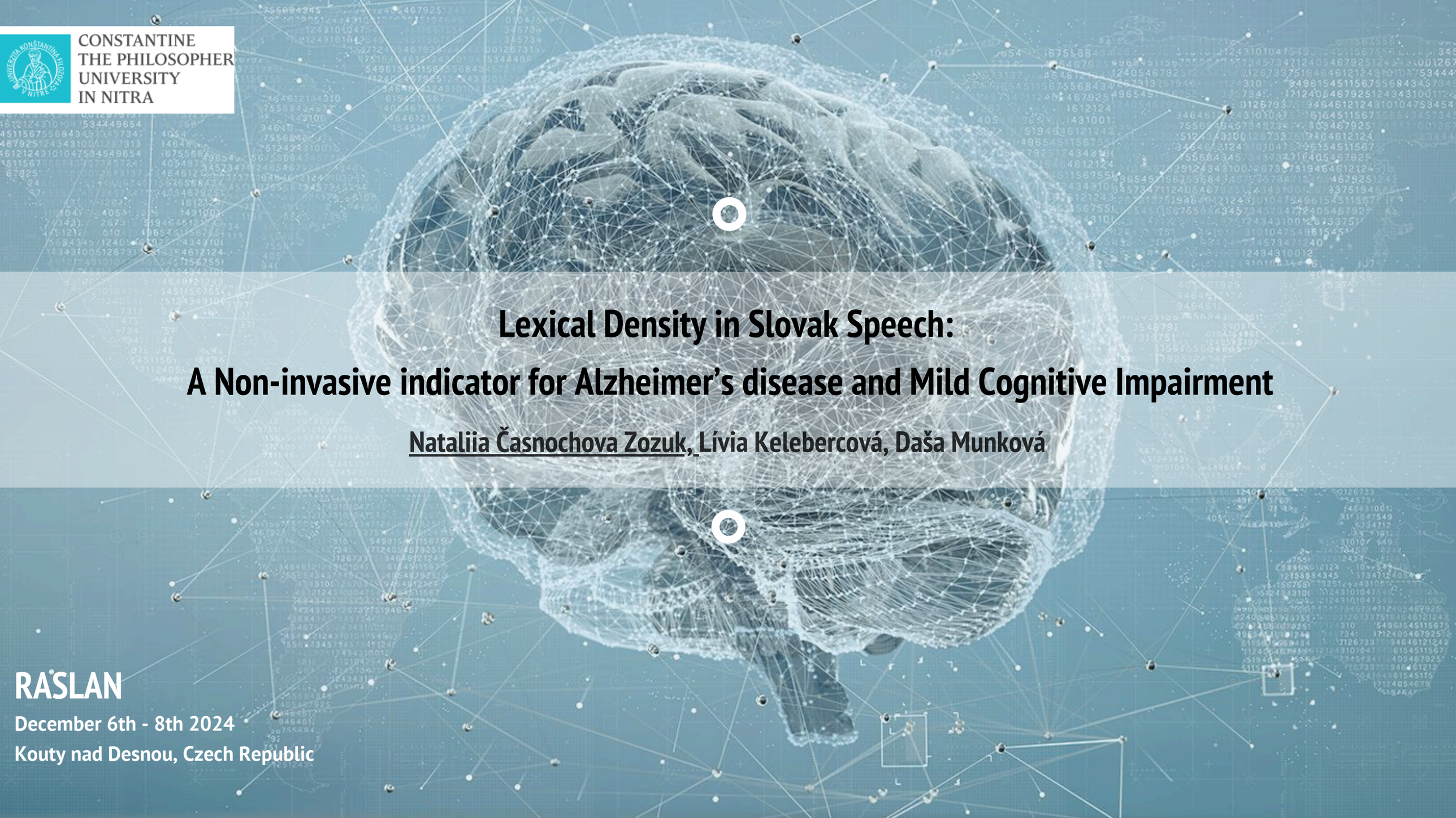


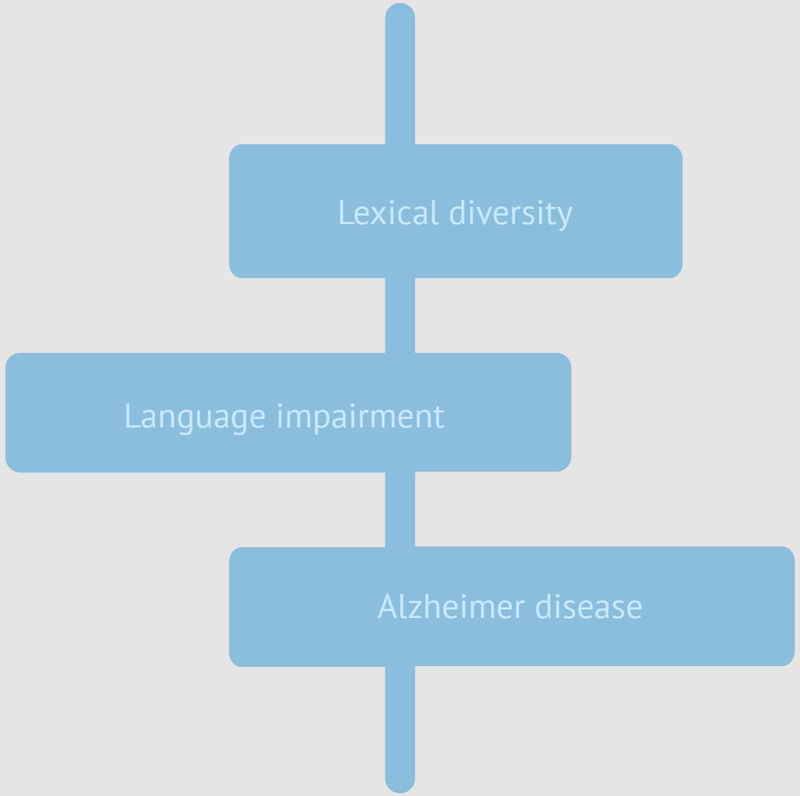
CONSTANTINE THE PHILOSOPHER UNIVERSITY IN NITRA

Lexical Density in Slovak Speech: A Non-invasive indicator for Alzheimer's disease and Mild Cognitive Impairment

Nataliia Časnochova Zozuk, Lívia Kelebercová, Daša Munková

RASLAN
December 6th - 8th 2024
Kouty nad Desnou, Czech Republic

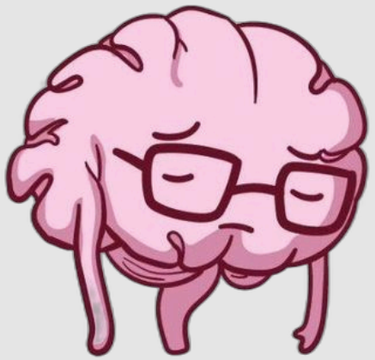
Content



- Neurodegenerative diseases (AD and MCI)
- Language disorders
- Lexical density
- Hypotheses
- Analysis
- Results

Alzheimer's disease

Memory



Thinking

Language

Behaviour

Mood and personality

Language disorders

Language disorders occur when a person has difficulty understanding spoken language and/or expressing him/herself in speech compared to people of a similar age.

Language is a valuable source of clinical information in Alzheimer's disease (**AD**) and Mild Cognitive Impairment (**MCI**) because it deteriorates concomitantly with the development of neurodegenerative disease.

Example of spontaneous speech

Healthy

chlapec stojí na stoličke naťahuje sa za banánom stolička sa mu prevracia asi padne z vodovodu tečie voda do umývadla vyteká von pozerá sa tam kocúr na to zboku otec má v ruke varechu zdvihol ju chcel trafiť muchu ale rozbil lampu ktorá je visiaca majú tam dve dve police jedna je otvorená polovica dverí sú tam priečky medzitým tam je fľaška ktorá...

MCI

no v kuchyni decko nejaké tam niečo pustil vodu voda do drezu vyteká z drezu voda vonku vidím tu ďalšie na kraji ešte mačku nejakú mačičku a dotyčný pán rozbil bola buchol do svetla varechou a zase na kuchynskom pulte tam je nejaké nejaký hrniec tiež niečo vyteká vonku nejaká omáčka alebo také niečo...

AD

no neviem prečo tam do toho klepe či búcha do toho svetla tam a ach je chlapček zase berie si z oného banán ale sa mu šmykla asi stolička neviem či nespadne tam tam je ešte nejaké..”

Lexical density (LD)

LD is a linguistic concept that refers to the ratio of the number of different lexical units (words) to the total number of words in a text. It is used to determine the linguistic richness and variety of a text or speech.

Metrics

- V1/W (Words that occur only once / Total number of words)
- V2/W (Words that occur twice / Total number of words)
- H1 (a measure of lexical density with hapax legomena for T)
- H2 (a measure of lexical density with hapax dislegomena for T)
- R1 (measure of lexical density with hapax legomena for N)
- R2 (measure of lexical density with hapax dislegomena for N)
- NOUN/W (nouns/total number of words)
- ADJ/W (adjectives/total number of words)
- ADV/W (adverbs/total number of words)
- VERB/W (verbs/total number of words)

$$H1 = \frac{\log(T) \cdot 100}{1 - V1/T},$$

$$H2 = \frac{\log(T) \cdot 100}{1 - V2/T},$$

$$R1 = \frac{\log(N) \cdot 100}{1 - V1/N},$$

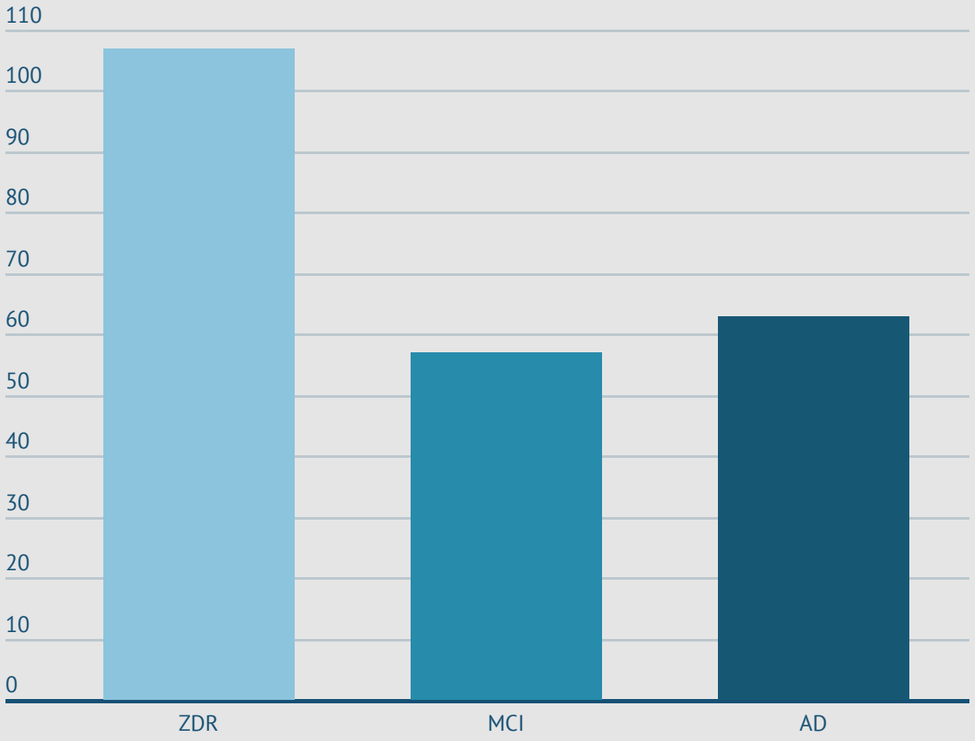
$$R2 = \frac{\log(N) \cdot 100}{1 - V2/N},$$

Hypotheses

- Patients with mild cognitive impairment describe situations with lower lexical density (as measured by the metrics V1/W, V2/W, H1, H2, R1, R2, NOUN/W, ADJ/W, ADV/W, VERB/W) compared to healthy subjects.

- Patients with Alzheimer's disease describe a situation of even lower lexical density (as measured by the metrics V1/W, V2/W, H1, H2, R1, R2, NOUN/W, ADJ/W, ADV/W, VERB/W) compared to patients with mild cognitive impairment.

Participants



Speech recordings were made of healthy and ill, diagnosed people within the EWA project.

Recordings from over 2000 participants were collected throughout Slovakia.

Only **214** participants were included in our research to keep the average age the same across all groups (Healthy, MCI, AD).

Results

Tab. 1. Lexikal density – Gamma coefficients

Metric	Valid N	Gamma	Z-score	p-value
V1/W & dgn	214	0.203221***	3.48637	0.000490
V2/W & dgn	214	0.053228	0.91216	0.361683
H1 & dgn	214	0.145262**	2.49158	0.012718
H2 & dgn	214	-0.028523	-0.48882	0.624968
R1 & dgn	214	-0.064127	-1.10081	0.270980
R2 & dgn	214	-0.555665***	-9.53960	0.000000
NOUN/W & dgn	214	-0.014862	-0.25463	0.799012
ADJ/W & dgn	214	-0.278981***	-4.78137	0.000002
ADV/W & dgn	214	-0.155380**	-2.66690	0.007655
VERB/W & dgn	214	0.074857	1.28342	0.199345

Note: *** - $p < 0.001$, ** - $p < 0.01$, * - $p < 0.05$

Tab. 1. Multiple comparisons of lexical density measures x dgn

Metric	Valid	Sum of Ranks	Mean Rank	0	1	2
V1/W: Kruskal-Wallis test: $H(2, N = 214) = 16.67294$, $p = 0.0002$						
0	108	9915.5	91.8102		**0.000272**	0.053241
1	44	5947.0	135.1591	**0.000272**		0.306096
2	62	7142.5	115.2016	0.053241	0.306096	
H1: Kruskal-Wallis test: $H(2, N = 214) = 11.92536$, $p = 0.0026$						
0	108	10293.5	95.3102		**0.001852**	0.372791
1	44	5862.0	133.2273	**0.001852**		0.186970
2	62	6849.5	110.4758	0.372791	0.186970	
R2: Kruskal-Wallis test: $H(2, N = 214) = 68.12768$, $p = 0.0000$						
0	108	15221.0	140.9352		**0.000007**	**0.000000**
1	44	3901.5	88.6705	**0.000007**		0.098482
2	62	3882.5	62.6210	**0.000000**	0.098482	
ADV/W: Kruskal-Wallis test: $H(2, N = 214) = 7.080144$, $p = 0.0290$						
0	108	12407.5	114.8843		1.000000	**0.033711**
1	44	5025.5	114.2159	1.000000		0.138280
2	62	5572.0	89.8710	**0.033711**	0.138280	
ADJ/W: Kruskal-Wallis test: $H(2, N = 214) = 16.71594$, $p = 0.0002$						
0	108	13370.5	123.8009		0.093404	**0.000202**
1	44	4397.0	99.9318	0.093404		0.616234
2	62	5237.5	84.4758	**0.000202**	0.616234	

Conclusion

V1/W	H1	R2	ADV/W	ADJ/W
Healthy vs MCI	Healthy vs MCI	Healthy vs MCI Healthy vs AD	Healthy vs AD	Healthy vs AD

Future research directions

- analysis of other metrics for investigating language disorders
- applying appropriate metrics
- training a classifier to detect symptoms of neurodegenerative diseases

Thank you for your attention!

email: nataliia.casnochova.zozuk@ukf.sk