From audio to cognitive markers : an automated verbal fluency analysis tool

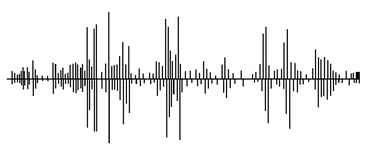
Léa Henriette[1], Florian Dubois[2], Laurent Lefebvre[3] & Sandra Invernizzi[3]
[1] Department of Computational Medicine and Neuropsychiatry, University of Mons
[2] Department of Mathematics and Operational Research, University of Mons
[3] Department of Cognitive Psychology and Neuropsychology, University of Mons

INTRODUCTION

- The verbal fluency test is used in neuropsychological assessment. It mainly includes phonological and semantic modalities, prompting participants to generate words based on a letter or category in two minutes.[1]
- Responses often form clusters, with transitions between them (called “switches”), providing insight into lexical retrieval strategies. [2]
- Traditional scoring focuses on the number of valid words, but this overlooks rich psycholinguistic and temporal data, such as word frequency and inter-word latency. These features can reveal subtle cognitive patterns, especially when analyzed dynamically. Manual analysis is time-consuming and subject to bias [3,4,5]

Research aim : develop an automated method to transcript and extract lexical and temporal markers from audio recordings.

METHODOLOGY



Audio sample → Whisper [6] → Retranscription → Lexique.org [7]

Whisper → Inter-word latency, Initiation time of the first word

Variables

Retranscription → Lexique.org → Lexical frequency, Word length, Orthographic/phonological neighborhood size, Clusters, Switches

Population : The analysis are based on recordings of verbal fluency tests conducted with three distinct groups: a control group (CTL), individuals with early-stage Alzheimer’s disease (AD), and adults over the age of 65 diagnosed with depression (DEP).

Table 1. Socio-demographic characteristics of the study sample

Groupe	AD	CTL	DEP
N	27	33	11
Age	72.8 (± 4.15)	68.2 (± 4.11)	66 (± 7.42)
% ♀	25.93% (n=7)	30.31% (n=10)	27.27% (n=3)
≤12 ans	66.66% (n=18)	72.73% (n=24)	63.43% (n=7)
MMSE	25.40 (± 2.91)	27.60 (± 2.25)	27.63 (± 2.46)

RESULTS

Table 2. Descriptive statistics and Kruskal–Wallis comparisons across groups.

mean x̄ (σ)	AD PF	AD SF	CTL PF	CTL SF	DEP PF	DEP SF	Kruskal-Wallis PF	Kruskal-Wallis SF
Nb of words	15.81 (±7.34)	13.11 (±5.29)	19.96 (±8.79)	26.12 (±8.60)	19.36 (±4.65)	24.91 (±8.02)		***
Word length	6.25 (±1.98)	6.10 (±1.79)	6.61 (±2.21)	6.08 (±1.73)	7.32 (±2.43)	6.09 (±1.83)	***	
Inter-word latency	4.51 (±6.06)	5.40 (±8.23)	3.41 (±5.20)	2.79 (±4.01)	3.52 (±4.34)	2.99 (±3.92)		***
Initiat° 1er	0.90 (±1.63)	0.61 (±0.64)	0.78 (±1.18)	1.06 (±1.39)	0.33 (±0.50)	0.56 (±0.70)		
Nb clusters	11.30 (±6.17)	9.74 (±6.94)	14.42 (±6.77)	17.00 (±5.44)	13.45 (±4.80)	18.00 (±5.95)		***
Nb switches	10.30 (±6.17)	8.81 (±3.94)	13.42 (±6.73)	16.03 (±5.43)	12.45 (±4.80)	17.09 (±6.04)		***
Frequency (films)	62.80 (±136.96)	29.59 (±40.81)	44.00 (±90.66)	19.19 (±31.37)	33.96 (±88.66)	21.22 (±33.15)	***	***
Frequency (books)	59.70 (±131.97)	23.13 (±33.09)	49.86 (±103.24)	17.12 (±26.79)	37.60 (±96.73)	18.41 (±26.65)	*	***
Nb ortho	4.15 (±4.79)	3.68 (±3.90)	3.72 (±4.40)	3.10 (±3.76)	2.98 (±4.42)	3.69 (±4.23)	*	*
Nb phono	9.39 (±9.20)	10.19 (±9.00)	7.62 (±8.42)	8.91 (±8.72)	5.73 (±7.73)	9.50 (±9.05)	***	*

Table 3. Results of the Dunn post-hoc comparisons with Holm correction

Dunn Post-hoc Holm correction	PF	FS
Nb of words		CTL & DEP *** > AD
Word length	DEP * > CTL * > AD (DEP *** > AD)	
Inter-word latency		CTL ** < AD
Nb clusters		CTL & DEP *** > AD
Nb switches		CTL & DEP *** > AD
Frequency (films)	CTL & DEP ** < AD	CTL & DEP ** < AD
Frequency (books)	CTL & DEP * < AD	CTL & DEP * < AD
Nb voisin ortho	DEP * < AD	
Nb voisin phono	CTL * & DEP ** < AD	CTL * & DEP ** < AD

Note. SF : Semantic Fluency ; PF : Phonological Fluency
* p < 0.05 ; ** p < .01 ; *** p < .001

Table 4. Error types observed in the retranscription

Error types	Hallucinations	Orthographical errors	Omissions	Intrusions
%	18.51%	18.85%	0.34%	11.25%

Table 5. Word error rates (WER) depending on the Whisper model

WER [6]	small	medium	large
%	16.2%	8.9%	8.9%

CONCLUSION

DEP: ↗ Word length (PF)

AD: ↘ Nb clusters; ↗ Frequency; ↗ phonological neighbors; ↗ orthographical neighbor (PF)

→ Integrating temporal and psycholinguistic variables offers new avenues for refining diagnostic distinction

Phase 2 : Operationalize the tool and release it as an open-access resource.

CHECK THE BIBLIOGRAPHY !

LET’S CONNECT !