NIHR | National Institute for Health and Care Research

PROSPERO
International prospective register of systematic reviews

Optimal testing situations for the automated analysis of cognitive components in natural language : a systematic literature review.

Chiara Vantwembeke, Édith Durand, Laurent Lefebvre, Sandra Invernizzi

Citation
Chiara Vantwembeke, Édith Durand, Laurent Lefebvre, Sandra Invernizzi. Optimal testing situations for the automated analysis of cognitive components in natural language : a systematic literature review.. PROSPERO 2025 CRD420251131954. Available from https://www.crd.york.ac.uk/PROSPERO/view/CRD420251131954.

REVIEW TITLE AND BASIC DETAILS

Review title
Optimal testing situations for the automated analysis of cognitive components in natural language : a systematic literature review.

Condition or domain being studied
Speech And Language Finding; Language Stimulation; Cognitive function ; Neuropsychological Testing; Natural language processing

This review focuses on cognitive components that can be inferred from natural language production in the general and the pathological adult population. Specifically, it targets cognitive functions which are reflected in discourse characteristics. The domain includes studies employing automated or semi-automated methods to assess or model cognitive performance based on spoken language. The review aims to identify testing situations (tasks, contexts, and modalities) that best capture these cognitive components through automated analysis.

Rationale for the review
Advances in (semi)-automated analysis of speech have enabled the identification of linguistic markers associated with cognitive components. However, current studies rely on highly heterogeneous testing situations, including spontaneous narratives, picture descriptions, interviews, or task-based elicitation, which substantially influence linguistic output and the validity of the cognitive indicators derived from it.

The lack of methodological standardization limits comparability across studies and hinders the identification of optimal testing conditions for reliable cognitive assessment through automated language analysis. Moreover, understanding how outcomes from (semi)-automated analysis of

speech correspond to those obtained through traditional standardized tests is essential to evaluate the degree of convergence between these two assessment modalities. This comparison is critical to assess the ecological validity and potential clinical utility of automated language analysis.

A systematic synthesis is therefore needed to (1) map the testing situations currently used in the literature, (2) evaluate their methodological characteristics, and (3) determine which conditions most effectively reveal cognitive components in natural language production while maintaining consistency with standardized cognitive assessments.

Review objectives

1. What are the situations used for stimulating spontaneous language ?
2. What are the indicators that reflect cognitive functioning in spontaneous language ?
3. Can softwares detect cognitive markers in language analysis ? Can this detection by the software be automated ?
4. Does cognitive data extraction using software correspond to data obtained from a traditional cognitive assessment ?

Keywords
Language; Speech; Natural language processing; Machine learning; Tool; Software; Cognitive function; Neuropsychology; Cognitive assessment; Linguistic features; Automatic classification; Ecological assessment; Algorithm; Supervised machine learning

Country
Belgium; Canada

ELIGIBILITY CRITERIA

Population

Included
Adults (over 18 years old).

Intervention(s) or exposure(s)

Included
Language Stimulation; Natural language processing ; Artificial intelligence; Assessment Using Conversation Analysis Profile For People With Cognitive Impairment

- Participant's language or speech production was analyzed, not comprehension.
- Experimental or empirical data were included.
- Language data were processed using software-based, automated, or semi-automated analysis methods.
- Data counter-validity beyond the software.

Excluded

1. Studies that did not provide sufficient methodological or quantitative data for a critical analysis of the results.
2. Studies presenting significant methodological bias such as:

- Unspecified or incomplete participant inclusion and exclusion criteria.
- Insufficient description of the language elicitation task (e.g., missing details on instructions, duration, or context).
- Inadequate reporting of data collection procedures (e.g., lack of information on recording conditions, transcription methods, or preprocessing steps).
- Lack of transparency in automated analysis procedures (e.g., missing description of tools, feature extraction methods, or model parameters).

Comparator(s) or control(s)

Included
PICO tags selected: Usual Care; Active control

Studies must have a validation measure.

Study design

Only nonrandomized study types will be included.

Included
Empirical quantitative and mixed-methods studies involving adult participants and examining language production analyzed through automated or semi-automated methods. Both cross-sectional and experimental designs will be included, as long as they report sufficient methodological detail and language–cognition associations.

Excluded
Theoretical papers, methodological reports without participant data, literature reviews, meta-analyses, conference abstracts without full text, and case studies will be excluded. Studies based solely on simulated or synthetic data without human participants will also be excluded. Unpublished doctoral dissertations and master's theses (grey literature) will not be considered in this review.

Context

The review will include studies conducted in experimental, laboratory, clinical, or naturalistic settings where adults produced spoken language samples for cognitive or linguistic assessment. Eligible studies may involve healthy adults or adults with cognitive difficulties, provided that the focus remains on language production and automated analysis of cognitive components. Studies using human participants are required; research based solely on simulated or synthetic data will be excluded. The context may include in-person data collection, online tasks, or pre-existing language corpora. No restrictions will be applied regarding country, publication date, or socioeconomic context.

TIMELINE OF THE REVIEW

Date of first submission to PROSPERO

02 December 2025

Review timeline

Start date: 9 October 2025. End date: 1 September 2026.

Date of registration in PROSPERO

10 December 2025

AVAILABILITY OF FULL PROTOCOL

Availability of full protocol
A full protocol has been written but is not available because:

A full review protocol has been written but is not yet available on OSF. The PDF will be uploaded and made publicly accessible as soon as it becomes available.

SEARCHING AND SCREENING

Search for unpublished studies
Only published studies will be sought.

Main bibliographic databases that will be searched
The main databases to be searched are *PsycInfo*, *PubMed* and *Scopus*.

Other important or specialist databases that will be searched
ProQuest

Search language restrictions
The review will only include studies published in English and French.

Search date restrictions
There are no search date restrictions.

Other methods of identifying studies
No other methods will be used.

Link to search strategy
A full search strategy has been uploaded to PROSPERO. The PDF may be accessed through this link
https://www.crd.york.ac.uk/PROSPEROFILES/9ebcf3f7f400f632d8eae7e89be63b8c.pdf.

Selection process
Studies will be screened independently by at least two people (or person/machine combination) with a process to resolve differences.

Other relevant information about searching and screening
None

DATA COLLECTION PROCESS

Data extraction from published articles and reports
Data will be extracted by one person (or a machine) and checked by at least one other person (or machine).

Authors will be asked to provide any required data not available in published reports.

Study risk of bias or quality assessment
Risk of bias will be assessed using:

Two authors will independently analyse the included studies and assess their quality using a dichotomous custom risk-of-bias grid developed for this review, based on ROBINS-I. In case of disagreement, a third author will review the study and the final decision will be reached through team discussion.

Data will be assessed independently by at least two people (or person/machine combination) with a process to resolve differences.

Additional information will be sought from study investigators if required information is unclear or unavailable in the study publications/reports.

Reporting bias assessment

Risk of bias due to missing results will be assessed

Certainty assessment

Certainty of evidence will be assessed using a qualitative approach adapted from the GRADE framework. Each outcome will be evaluated across key domains, including precision of the effect estimate (or sample size), consistency of findings across studies, study design limitations and missing results (risk of bias), and the directness of the evidence relative to the review question.

Criteria from GRADE domains (imprecision, inconsistency, risk of bias, and others) will be applied systematically. An overall certainty rating—high, moderate, low, or very low—will be assigned for each outcome based on explicit decision rules. Adaptations specific to this review (e.g., assessment of transparency and reproducibility of automated language analysis procedures) will be reported.

Two reviewers will independently assess certainty, with disagreements resolved through discussion or third-reviewer arbitration. Additional information will be sought from study authors if needed. Certainty assessments will be documented explicitly to ensure transparency, replicability, and alignment with international reporting standards.

OUTCOMES TO BE ANALYSED

Main outcomes

The main outcomes will be the measures derived from automated or semi-automated analyses of spontaneous or elicited language that reflect cognitive functioning in adults (healthy or pathological).

Outcomes of interest include linguistic, semantic, syntactic, pragmatic, or discourse-level indicators extracted using natural language processing (NLP) tools, automated linguistic feature analysis software, or similar computational methods.

Acceptable measures will include any quantitative or categorical linguistic metrics (e.g., lexical diversity, syntactic complexity, coherence, fluency, informativeness) reported in relation to cognitive processes such as executive functions, attention, memory, reasoning, or overall cognitive performance.

A specific objective of the review is to identify the testing situations that most optimally elicit language samples capable of reflecting cognitive components through automated analysis.

Where available, data comparing automated language-based measures with standardized cognitive or neuropsychological test scores will be collected.

The main outcomes will be summarized descriptively and, if possible, quantitatively synthesized according to the cognitive domain investigated and the linguistic indicators used.

Additional outcomes

none

PLANNED DATA SYNTHESIS

Strategy for data synthesis

A formal data synthesis will be conducted following PRISMA guidelines.

Data from the included studies will be systematically extracted and organised into analytical categories based on:

1. The type of testing situation used to elicit language (e.g., narrative production, picture description, conversational tasks, procedural discourse).
2. The linguistic or computational indicators generated by automated analysis tools.
3. The cognitive domains assessed (e.g., executive functions, attention, memory, reasoning).
4. The number of participants in the studies and the characteristics of the participants (e.g., age, gender, education).
5. The participant group (healthy controls vs. clinical populations), including details of the clinical condition when applicable (e.g., neurodegenerative disorders, acquired brain injury).
6. The correspondence between automated linguistic measures and standardized cognitive assessments.

A structured framework will be used to compare studies, including tabulation of extracted variables, and identification of converging and diverging findings.

CURRENT REVIEW STAGE

Stage of the review at this submission

Review stage	Started	Completed
Pilot work		
Formal searching/study identification		
Screening search results against inclusion criteria		
Data extraction or receipt of IPD		
Risk of bias/quality assessment		
Data synthesis		

Review status

The review is currently planned or ongoing.

Publication of review results

Results of the review will be published in English.

REVIEW AFFILIATION, FUNDING AND PEER REVIEW

Review team members

Mrs Chiara Vantwembeke. Université de Mons. Belgium.

No conflict of interest declared.

Professor Édith Durand. Université du Québec à Trois-Rivières. Canada.

No conflict of interest declared.

Professor Laurent Lefebvre. Université de Mons. Belgium.

No conflict of interest declared.

Dr Sandra Invernizzi (review guarantor). Université de Mons. Belgium.

No conflict of interest declared.

Named contact

Mrs Chiara Vantwembeke (chiara.vantwembeke@umons.ac.be). Université de Mons. Belgium.

Review affiliation

Université de Mons (UMONS), Département de Psychologie Cognitive et Neuropsychologie

Université du Québec à Trois-Rivières (UQTR), Département d'Orthophonie

Funding source

Review has no specific/external funding but is supported by guarantor/review team (non-commercial) institutions.

Peer review

The review process will include a peer review stage based on inter-judge agreement.

Study selection, data extraction, and quality appraisal will be independently performed by at least two reviewers.

Discrepancies will be discussed and resolved by consensus, and a third reviewer will be consulted if needed.

This procedure aims to ensure methodological rigor and reliability in the inclusion and evaluation of studies.

ADDITIONAL INFORMATION

Review conflict of interest

Declared individual interests are recorded under team member details.. No additional interests are recorded for this review.

Medical Subject Headings

Humans; Adult; Cognition; Speech; Language; Software; Supervised Machine Learning; Natural Language Processing

SIMILAR REVIEWS

Check for similar records already in PROSPERO

PROSPERO identified a number of existing PROSPERO records that were similar to this one (last check made on 30 October 2025). These are shown below along with the reasons given by that the review team for the reviews being different and/or proceeding.

- A Systematic Review of Natural Language Processing Techniques for Early Detection of Cognitive Decline [published 30 September 2024] [CRD42024592875]. The review was judged **not to be similar**
- Early Detection of Cognitive Decline in Elderly Patients Using Natural Language Processing Applied to Clinical Notes: A Systematic Review [published 25 October 2024] [CRD42024601303]. The review was judged **not to be similar**
- Text-Based Depression Detection Using NLP With Gold Standard Labels: A Systematic Review and Meta-Analysis [published 25 May 2025] [CRD420251056902]. The review was judged **not to be similar**

PROSPERO version history

- Version 1.0, published 10 Dec 2025

Disclaimer