Embedding-Based Graded Scoring of Neuropsychological Language Tests

Authors: Dimitrios Kokkinakis, University of Gothenburg
Fredrik Boglind, University of Stockholm
Dana Dannélls, University of Gothenburg
Abstract ID: 276
Event: Linguistics 2026 Athens
Subject: 21. Language technology
Presenter Name: Dimitrios Kokkinakis

Objectives

Language is dynamic, meanings converge, diverge, and form evolving semantic fields. In clinical neuropsychology, however, this variability is typically reduced to fixed categories. Linguistic ability and impairment are commonly assessed using standard neuropsychiatric instruments such as the Semantic Verbal Fluency test (SVF), the Phonological Verbal Fluency test (FAS), and the Boston Naming Test (BNT). Most often, responses on these measures are scored dichotomously as *correct* or *incorrect*. This binary scoring obscures semantically related, approximate, or deviant responses. The objective of this study is to develop and evaluate a reproducible computational method for continuous, semantically informed scoring of these tests in Swedish. The primary research question is whether modern vector-based language models can generate stable and interpretable continuous semantic scores that capture graded variation beyond binary classification.

Methodology

The study applies a computational linguistic framework grounded in distributional semantics, where word meaning is represented as position in a high-dimensional semantic space. Anonymised, synthetically generated lexical responses were used to enable controlled methodological development without sensitive data. Text preprocessing, including normalisation and lemmatisation, was performed using tools from *Språkbanken's text infrastructure*. Responses and target words were represented using Swedish-adapted BERT- based vector embeddings. BERT ("Bidirectional Encoder Representations from Transformers") is a transformer-based language model that learns contextual word representations by analysing large corpora of text and modelling how words relate to surrounding words in both left and right contexts. In this framework, lexical meaning is encoded as numerical vectors in a high-dimensional semantic space, where semantically similar words are positioned closer to one another. This representation enables graded measurement of semantic proximity rather than categorical judgments of correctness. For the verbal fluency tests (SVF and FAS), semantic dispersion was also computed to quantify how responses are distributed within the semantic space. In this context, semantic dispersion denotes the quantitative distribution of response vectors within a high-dimensional embedding space, operationalised as the extent to which lexical items diverge from one another in semantic representation.

Results

Vector-based representations generated stable and interpretable continuous scores. The method captured fine-grained variation among semantically related responses that is lost under binary scoring. Systematic differences in response structure were observed across the Boston Naming Test (BNT), Semantic Verbal Fluency (SVF), and Phonological Verbal Fluency (FAS). Linguistic performance could thus be modelled as movement within a semantic space rather than as a series of discrete outcomes.

Discussion

The study demonstrates the feasibility of continuous semantic scoring for Swedish language assessment. The proposed method provides a methodological foundation for future clinical validation and contributes to research on how meaning is structured and dynamically organised in cognitive processes. By reconceptualising test performance as graded semantic movement, the study advances computational approaches to linguistic assessment in neuropsychology. Importantly, this framework enables the quantification of latent semantic structure in a manner that is theoretically grounded, statistically scalable, and reproducible across datasets. Such an approach may facilitate more sensitive detection of subtle linguistic deviations, potentially improve early identification of cognitive decline and supporting longitudinal monitoring of semantic change over time.