

6. Landrigan, J., Zhang, F., & Mirman, D. (2021) A data-driven approach to post-stroke aphasia classification and lesion-based prediction, *Brain*, awab010, <https://doi.org/10.1093/brain/awab010>

Aphasia is an acquired impairment in the production or comprehension of language, typically caused by left hemisphere stroke. The subtyping framework used in clinical aphasiology today is based on the Wernicke-Lichtheim model of aphasia formulated in the late 19th century, which emphasizes the distinction between language production and comprehension. The current study used a data-driven approach that combined modern statistical, machine learning, and neuroimaging tools to examine behavioural deficit profiles and their lesion correlates and predictors in a large cohort of individuals with post-stroke aphasia. First, individuals with aphasia were clustered based on their behavioural deficit profiles using community detection analysis (CDA) and these clusters were compared with the traditional aphasia subtypes. Random forest classifiers were built to evaluate how well individual lesion profiles predict cluster membership. The results of the CDA analyses did not align with the traditional model of aphasia in either behavioural or neuroanatomical patterns. Instead, the results suggested that the primary distinction in aphasia (after severity) is between phonological and semantic processing rather than between production and comprehension. Further, lesion-based classification reached 75% accuracy for the CDA-based categories and only 60% for categories based on the traditional fluent/non-fluent aphasia distinction. The results of this study provide a data-driven basis for a new approach to classification of post-stroke aphasia subtypes in both research and clinical settings.