# Keynote Lecture

# Leveraging big data and AI in studies and treatment of aphasia

Author: Bonnie Nozari

Affiliation: Carnegie Mellon University

## Outline

Speech errors have been instrumental in advancing our understanding of the architecture of the language production system, the nature of its representations, and its disorders. To be most informative, researchers usually need large amounts of data. Hand-coding such data can be both cumbersome and subjective. In this talk, I will present LeCoder, the first open-source, automated error coder, which uses a data-driven approach grounded in large-scale corpora to quantify the target-response relationship, allowing it to be flexible, scalable, and generalizable across new datasets.

By testing the coder on two datasets from two aphasia labs that have been carefully coded by trained research assistants, I will show that LeCoder has high accuracy when compared to expert coders, and in certain cases, offers a more logical categorization than human coders. I will then show, using robust machine-learning approaches, that LeCoder's performance generalizes to new participants and items it has never encountered before. Collectively, these findings encourage the use of LeCoder across labs for more objective coding of speech errors, which will, in turn, increase replicability of findings in all subfields of research that use speech error analysis, including neuropsychological research.

Time permitting, I will also briefly preview a new platform developed in my lab, which uses large language models to help individuals with agrammatical aphasia produce more complete sentences in real time.

.