Available online at www.sciencedirect.com
ScienceDirect
Procedia - Social and Behavioral Sciences 00 (2010) 000–000
Procedia Social and Behavioral Sciences
www.elsevier.com/locate/procedia

Academy of Aphasia 2012

Comprehensive Statistical Analysis of Aphasia Types

MacWhinney B.[a,*], Fromm D.[a], Forbes M.[a], Holland A.[b], Makris C.[a], Todd T.[a], Greenhouse J.[a]

[a] *Carnegie Mellon University*
[b] *University of Arizona*

Comprehensive Statistical Analysis of Aphasia Type

Classifying people with aphasia (PWAs) into aphasia types is important for clinical research purposes, treatment method decisions, and neurolinguistic theory. This presentation will describe how standardized assessment measures can be linked to data from conversational interactions to further inform and differentiate syndrome classifications.

Samples and scores for this analysis come from the AphasiaBank database that now provides transcriptions of test sessions with 176 PWAs, along with standardized test scores. Participants from 14 different sites around the United States were tested with a standardized protocol, comprising samples of free speech (stroke story and recovery), three picture descriptions, story telling (Cinderella), and procedural discourse (making a peanut butter and jelly sandwich). We also administered the Aphasia Quotient subtests from the Western Aphasia Battery-Revised (WAB; Kertesz, 2007), the short form Boston Naming Test-Second Edition (Kaplan, Goodglass, & Weintraub, 2001), the Northwestern Verb Naming Test (Thompson, in preparation), and a non-standardized test developed to assess word-level and sentence-level repetition. All discourse tasks and testing, with the exception of the WAB, were recorded on video. Extensive demographic data (51 fields) were collected on each participant as well.

The discourse samples went through a detailed process of transcription, coding, and checking. Transcription was done using the CHAT format, which operates closely with the CLAN programs (MacWhinney, 2000). These programs allow for the analysis of a wide range of linguistic and discourse structures. We coded word repetitions, revisions, fillers, gestures, sound fragments, and unintelligible output. Word-level errors were coded in six categories: phonology, semantics, neologism, dysfluency, morphology, and formal lexical features. The CLAN programs produced automatic morphosyntactic analysis of the transcripts from which we calculated total words, total utterances, duration of PWA response, and vocabulary diversity (VOCD, Malvern et al. 2005). We also tabulated usage proportions for 23 part-of-speech classes and 16 grammatical features. These measures were calculated for each section of the discourse protocol (free speech, pictures, story, procedure).

Our first statistical analysis was conducted when the database contained only 76 PWAs. We applied principal components analysis verified with k-means clustering from which we extracted two principal components. The first had heavy loading on WAB subtest scores. The second had heavy loadings from AphasiaBank measures, particularly the repetition test.

We then used model-based clustering that maximized the BIC (Bayesian Information Criterion) with three clusters and a VEI (Variable Shape, Equal Volume, Diagonal Orientation) density. This method showed that participants classified as Broca by clinicians fell into two different clusters in the statistical analysis, as suggested earlier by Sundet and Engvik (1985), and that those classified as Anomic also split into two separable types.

With the current set of 176 PWAs, we will repeat the earlier analyses and use several additional clustering approaches, including agglomerative, spectral, and random forest to potentially reveal new group relationships. Preliminary results suggest refined groupings of PWAs by clinically meaningful factors, such as fluency, speech output, and word error rates.