**Automating the Linguistic Annotated Bibliography (LAB)**

Ari L. Cunningham, B.S., Missouri State University

Erin M. Buchanan, Ph.D., Harrisburg University of Science and Technology

Nikhil Chate, B.S., Harrisburg University of Science and Technology

[Arielle924@live.missouristate.edu](mailto:Arielle924@live.missouristate.edu)

EBuchanan@HarrisburgU.edu

[NChate@my.harrisburgu.edu](mailto:NChate@my.harrisburgu.edu)

**Short abstract:** Stuff. (Don’t think we need this)

**Full abstract (250 words):**

In psycholinguistic research, it is critical to be able to find and use validated linguistic data. The rate of publication for linguistic norms is exponential, and it can be difficult for researchers to sort through resources and find the datasets most relevant to their research. The Linguistic Annotated Bibliography (LAB), which supplemented an existing resource (The Language Goldmine: LG), was published in 2018 to help researchers search the vast number of linguistic and psycholinguistic databases. However, both the LAB and LG depended on manual search to add new publications. At the current rate of publication, it would be more efficient to automate the search for and inclusion of relevant resources. This presentation will focus on a computer algorithm that searches new abstracts across popular journals for inclusion in the LAB or LG. The existing data in the LAB was used as a “gold standard” to train a classifier that detects the featural components of publications. The classifier was used to detect new articles for inclusion, which were manually checked and compared with the algorithm’s results. We plan to present a new gamified website displaying articles tagged by the algorithm. Researchers can vote on their inclusion for further script development. Automation will allow the LAB to stay up to date with a vast number of resources as they are published, while crowdsourcing features will allow users to provide information to others about the usability of data. With these features, the LAB can more effectively make validated psycholinguistic data easy to find.

**Submission type:** Spoken or Poster