**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

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

EBuchanan@HarrisburgU.edu

**Short abstract:** Stuff.

**Full abstract (250 words):**

**Problem:**

In psycholinguistic research, it is critical to be able to find and use validated linguistic data. Recent trends show that there have been a large number of publications about linguistic datasets, and with new publications becoming available at this rate, it can be difficult for researchers to sort through them and find the resources most relevant to them. The Linguistic Annotated Bibliography (LAB) was created in 2018 to help researchers search among the vast number of databases for linguistic and psycholinguistic data (Buchanan, Valentine, & Maxwell, 2018). However, the first version of the LAB depended on manual search to bring in new publications. At the rate of publication in this area, a more efficient solution would be to automate the search for and inclusion of relevant resources.

**Procedure:**

To address this opportunity, we are re-designing the LAB to automate the search for new publications and to add the element of crowdsourcing. We will create a script that searches relevant journals and databases for abstracts containing key words and phrases. The training data to create the algorithm will consist of abstracts currently existing in the LAB (for the “accept” data) as well as articles from the same time period that are not relevant (for the “reject” data). We will test the algorithm by using abstracts from articles published after the original LAB was developed and manually coding the abstracts to compare our results with those from the algorithm. Additionally, we plan to allow others in the linguistic research community to provide feedback about the publications selected for the LAB.

**Conclusions:**

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 usefulness of data. With these features, the LAB can more effectively make validated psycholinguistic data easy to find.

**Submission type:** Spoken or Poster