**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:**

**Problem:**

In psycholinguistic research, it is critical to be able to find and use validated linguistic data, often found online in databases. The Psychonomic Society formerly provided a free list of such sources in one place, but this is no longer the case. Even for members, the information is spread across society journals and does not account for references to non-society journals. For this reason, the Linguistic Annotated Bibliography was created in 2018 to make validated data easier to find (Buchanan, Valentine, & Maxwell, 2018).

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 LAB is a tool to help researchers search the vast number of databases for linguistic and psycholinguistic data. However, the first edition 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 relevant resources to include in the LAB.

**Procedure:**

The goal of LAB 2.0 is to both to automate the search for new publications and to add the element of crowdsourcing. We will use Python to 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). After developing the algorithm, we will test it using abstracts from articles published after the original LAB was developed. We will hand code the abstracts to compare our results with those from the algorithm.

To incorporate crowdsourcing features, we plan to allow others in the linguistic research community to provide feedback about the publications selected for the LAB: feedback such as how useful or accessible a resource is. We will program these features into the updated Shiny app.

**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.

**Mailing address:**

?

**Submission type:** Poster