

Abstract

This thesis deals with POS tagging in the Talmud. We have created the PSHAT system, which handles the lack of punctuation and a language mixture of Hebrew and Aramaic in the Talmud. PSHAT employs two bidirectional long short-term memory networks, one to tag each word by language and another to tag Aramaic words with their part-of-speech. We have validated that our system is robust for varying mixtures of Hebrew and Aramaic and that it improves performance on unknown words compared to a baseline. PSHAT has been implemented such that it could be easily applied to other mixtures of languages as well. The code and data used for this project can be found at <https://github.com/nsantacruz/PSHAT>.