UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Comparing Acceptability in Magnitude Estimation Tests to an Unsupervised Model of Language Acquisition

Permalink
https://escholarship.org/uc/item/03v1q2jr

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Pederson, Bo
Edelman, Shimon
Solan, Zach
et al.

Publication Date
2004

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Comparing Acceptability in Magnitude Estimation Tests to an Unsupervised Model of Language Acquisition

Bo Pedersen and Shimon Edelman Zach Solan, D. Horn, E. Ruppin
Department of Psychology
Cornell University
Ithaca, NY 14853, USA
se37@cornell.edu

Faculty of Exact Sciences
Tel Aviv University
Tel Aviv, Israel 69978
{*zsolan,horn,ruppin*}*@post.tau.ac.il*

Traditionally language models have been evaluated by testing their ability to mark sentences as grammatical or ungrammatical. But with the emergence of probabilistic, connectionist models etc. on the computational side and magnitude estimation tests etc., on the linguistic side, it might make sense to go all the way and evaluate the models graded predictions.

We present a language acquisition algorithm that can learn structural regularities from raw data without any prior knowledge about the data. When trained on corpora the extracted language structures can be tested with new sentences to which a graded score is assigned.

Three experiments were conducted. The algorithm was trained on text from the English CHILDES database [MacWhinney and Snow. 1985. The child language exchange system] and then tested on linguistic acceptability data collected by Keller [Keller, Frank. 2000. Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality. PhD Thesis, University of Edinburgh] and the algorithm was partially successful on these.

A linguistic acceptability experiment was performed on a large set of well controlled data from an ESL multiple choice (English as Second Language) test and a modest but highly significant correlation with the algorithm score was found.

Finally a linguistic acceptability experiment was performed on sentences generated randomly from a small CFG. 25% of the sentences had 2 neighbor words permuted and another 25% of them had 2 random words from anywhere in the sentence permuted. Both groups got, as expected, significantly lower acceptability score but furthermore the latter had a significantly lower score and a higher variance suggesting that global permutations are more violating but also sometimes by chance get acceptable. The algorithm gives a more clear cut division of the permuted and non-permuted sentences (when trained on similar sentences) but it remains to be investigated whether it can distinguish the two different permutations.

These experiments show that our scoring function is still somewhat unstable and only performs well when variations are small or the data is highly structured as in the CFG experiment. But it also shows that the algorithm is productive under even slightly absurd circumstances like when we train it on CHILDES and test it on the more complex sentences from the ESL data. Furthermore, if we administer the ESL sentences as a multiple choice test the algorithm performs as "intermediate" according to the norms for that test.