April 23, 2019

Dear Dr. Cooper,

We would like to submit our manuscript entitled "Modeling the influence of language input statistics on children's speech production" to *Cognitive Science* for consideration as a regular article. In this paper, we test whether one proposed statistical learning method for language acquisition (Backward Transitional Probability, as implemented in McCauley & Christiansen's (2011; 2014; 2019) Chunk Based Learner (CBL) model) is able to consistently and accurately account for children's speech production in the first four years of life. Prior work in this domain suggests that statistical learning for language might be special; while visual and non-verbal statistical learning improves with age, the same tasks with linguistic stimuli stay stable (Shufaniya & Arnon, 2018 *Cognitive Science*). Prior results, however, have been limited to experiment-based data. Is there evidence for age-invariance in statistical learning for children's natural language input and speech production?

In our paper, we first replicate prior results with the CBL model using a longitudinal corpus of natural at-home child language recordings. Then, using a new, corrected measure of utterance reconstruction accuracy, we demonstrate that the CBL *indeed* yields age-invariant performance in accounting for children's utterances between ages 1;0 and 4;0. We discuss the results with respect to developmental change in memory, linguistic ability, and linguistic input. We believe our paper adds an important new datapoint to the debate on age-invariance in statistical learning and our results have broader implications for thinking about the role of developmental change in the use and outcomes of statistical learning mechanisms.

The main text of the manuscript is 6056 words long (excluding abstract, references, and the supplementary materials) and includes 6 figures, placed near where they are referred to in the text for ease of review. We also submit a set of Supplementary Materials that include statistical analyses using an alternative implementation of the CBL model. We confirm that the study is original and is not under review or published elsewhere. The data on which the model was trained is publicly available through the CHILDES website; no further ethical approval was needed.

Model code and analysis scripts are available at https://github.com/marisacasillas/CBL-Roete. For blind review, we link to an anonymous OSF repository in the manuscript: https://osf.io/ca8ts/?view_only=daaa1bcc71654842b0d70efe785a26b9. Finally, none of the coauthors have any interests that may influence the research, and all authors have agreed to the byline order and to submission of the manuscript.

We would recommend the following person as handling editor:
Padraic Monaghan p.j.monaghan@uva.nl

We also suggest the following people as reviewers:
Inbal Arnon inbal.arnon@gmail.com
Jennifer Misyak jennifer.misyak@wbs.ac.uk
Franklin Chang chang.franklin@gmail.com
Elise Hopman hopman@wisc.edu

Please note that we would prefer *not* to have the following scholars as reviewers since these individuals could have a potential conflict of interest with this project: Rebecca L. A. Frost, Evan Kidd, Limor Raviv, Stewart McCauley, and Morten Christiansen.

Sincerely,

Ingeborg Roete, Stefan Frank, Paula Fikkert, and Marisa Casillas

Marisa Casillas (Corresponding Author)
Max Planck Institute for Psycholinguistics
Wundtlaan 1
P.O. Box 310
6500 AH Nijmegen
The Netherlands
024-3521446
marisa.casillas@mpi.nl