

Simulating the developmental pattern of finiteness marking in English, Dutch, German, French and Spanish using MOSAIC

Julian M. Pine, Daniel Freudenthal (University of Liverpool) and Fernand Gobet (Brunel University)

One of the challenges facing computational approaches to language development is to develop models whose behaviour can be directly compared with that of language-learning children. Another is to show that these models can be extended beyond English-speaking children to simulate the behaviour of children learning a variety of different languages. MOSAIC (Model of Syntax Acquisition in Children) is a computational model of language acquisition that attempts to meet these challenges by using exactly the same learning mechanism to simulate the behaviour of children learning several different languages. MOSAIC takes as input corpora of orthographically-transcribed child-directed speech and learns to produce as output ‘child-like’ utterances that become progressively longer as learning proceeds. As a result of these characteristics, MOSAIC can be used to generate corpora of utterances at different stages of development, and hence to model the behaviour of children learning different languages across a range of MLU values.

In this paper, we describe how MOSAIC has been used to simulate developmental changes in the cross-linguistic pattern of finiteness marking across 5 different languages.

In a first set of simulations we show how one identical version of MOSAIC is able to simulate developmental changes in the proportion of OI errors in Dutch, English, Spanish and German by learning to produce and generate from progressively longer utterance-final strings.

In a second set of simulations we show how a new version of MOSAIC that learns from both edges of the utterance is able to simulate the developmental pattern of finiteness marking in both declaratives and Wh- questions. In particular, we show that MOSAIC is able to simulate differences in the proportion of OIs in Wh-questions in English and German as a result of the interaction between its edge-first learning mechanism and differences in the way Wh- questions are formed in these two languages.

In a final set of simulations we compare MOSAIC and Legate and Yang’s (2007) Variational Learning Model in terms of their ability to predict cross-linguistic variation in the rate of OI errors in English, Dutch, German, French and Spanish at $MLU = 2.0$. We conclude that, although both accounts fit the cross-linguistic patterning of OI errors surprisingly well, only MOSAIC is able to explain why verbs that occur more frequently as infinitives than as finite verb forms in the input also occur more frequently as OI errors than as correct finite verb forms in the children’s output.