The Acquisition of Case Systems in Typologically Diverse Languages: Children Gradually Generalize Abstract Grammatical Rules

Children initially develop a language that is item-specific, where grammatical phenomena are centered around rote-learned lexical items (e.g. Lieven & Tomasello 2008; Tomasello 2000; Goldberg 2006). The question how children become productive after this item-specific phase is, however, still under-researched (Granlund et al., 2019).

We ask how children become productive in the use of case-marking. We hypothesize that productivity emerges gradually, independent of the case system of their native language. We test this hypothesis in two steps: In Study 1, we simulate the learning process implementing different learning hypotheses, i.e. we test item-specific learning against instant generalization (rule-based learning). In Study 2 we test the hypothesis in naturalistic longitudinal corpora of three typologically diverse languages: Russian (Indo-European), Japanese and Chintang (Sino-Tibetan).

In Study 1, we simulate the different learning mechanisms computationally, in order to establish a theoretical baseline to compare against. Conceptually, we "create a child" that learns STEM-CASE co-occurrences, under different assumptions about the underlying learning process, i.e., item-specific learning followed by gradual emergence of productivity or full productivity from early on. In order to do this, we simulate monthly recordings that contain gradually more stems (to reflect vocabulary learning), which can be marked for case according to the different theoretical expectations. To judge the number of new stems and case frequency (types and tokens for both variables) we used the Chintang corpus for empirical estimates. We use the information-theoretic measure of conditional entropy as probe of STEM-CASE productivity. **We find** a constant level of productivity in the rule based-learning simulation, and no productivity (during the item specific phase) with a following gradual increase in the second simulation. These patterns provide a baseline for comparison for the second study.

In Study 2, we measure the productivity of case markers in longitudinal corpora of the three languages, which contain 13 children covering an age span from one year to six years old (see table **??**) (Moran et al., 2019). We mine for nominal stems marked for case, in order to test how children actually behave under real learning scenarios. To assess the development of productivity of case-marking, we measure conditional entropy, calculated on STEM-CASE co-occurrences. We predict case productivity in two ways, applying Bayesian multi-level models: (1) In language specific analyses we predict case productivity with a first order interaction between child age and speaker type (child or adult) and a main effect for speaker type. We also added random intercepts for speaker type and adult-child dyads. (2) In a cross-linguistic analysis we predict productivity with the same model specifications, but add random intercepts for language. **We find** in all analyses the same gradual increase in case productivity by children in these languages as in the simulation of gradually emerging case productivity (see table 1). Our results indicate that there is language specific variation when children become productive in case marking, which can be explained by case prevalence in the input and complexity of the case paradigm.

Results of these studies provide evidence that STEM-CASE combinations are gradually generalized towards more abstract schemata and are not subject to immediate generalization as assumed in rule-based approaches.

Effects	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	-0.05	1.75	-3.46	3.66
Child Speaker	-0.49	2.52	-5.59	6.04
Adult Speaker : Age	0.10	0.03	0.04	0.16
Child Speaker : Age	**0.26**	**0.03**	**0.20**	**0.31**

Table 1: We use a Bayesian linear model to predict STEM-CASE variability with a first order interaction between (child) age and speaker type (child or adult) and a main effect for speaker type. We also added random intercepts for speaker type, adult-child dyads and language.

References

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language.* Oxford University Press on Demand.

Granlund, Sonia, Joanna Kolak, Virve Vihman, Felix Engelmann, Elena V.M. Lieven, Julian M. Pine, Anna L. Theakston & Ben Ambridge. 2019. Language-general and language-specific phenomena in the acquisition of inflectional noun morphology: A cross-linguistic elicited-production study of Polish, Finnish and Estonian. *Journal of Memory and Language* 107. 169 - 194.

Lieven, Elena & Michael Tomasello. 2008. *Children's first language acquistion from a usage-based perspective.* Routledge/Taylor & Francis Group.

Moran, Steven, Robert Schikowski, Danica Pajović, Cazim Hysi & Sabine Stoll. 2019. The ACQDIV Corpus: a comparative longitudinal language acquisition corpus. Version 1.0.

Tomasello, Michael. 2000. The item-based nature of children's early syntactic development. *Trends in cognitive sciences* 4(4). 156–163.