Providing a Multi-Agent Computational Laboratory for Language Acquisition Experiments

David Ph. Shakouri[1,2], Crit Cremers[1], Niels O. Schiller[1,2,3]
[1] Leiden University Centre for Linguistics (LUCL), Leiden University; [2] Leiden Institute for Brain and Cognition (LIBC), Leiden University; [3] City University of Hong Kong (CityU), Hong Kong
d.p.shakouri@hum.leidenuniv.nl

This presentation discusses experiments performed using a multi-agent computational laboratory environment for language acquisition experiments. This system is based on the interaction between two agents: (1) an adult language model employing a symbolic generator and parser [1] and (2) a daughter language model, which sets out to learn the mother language [2, 3]. Both agents implement a model of many aspects of grammar such as syntactic as well as semantic representations. Crucially, the daughter agent performs *unsupervised learning*: The daughter does not have access to the internal linguistic knowledge of the mother agent but only to the language exemplars the mother produces during the conversation. To this end, the daughter language model employs a hybrid approach combining statistical as well as rule-based techniques to acquire the target language that constitutes (a fragment of) Dutch. Details on the learning mechanisms, the representations of the acquired grammatical knowledge, and how this framework relates to other models of language acquisition and evolution are provided in [2]. As soon as the daughter agent has acquired new grammatical knowledge, this is used to take part in the conversation with the mother. The presented experiments illustrate how the MODOMA can be used to acquire explicit abstract grammatical knowledge and substantiate that the MODOMA project resulted in a viable tool for language acquisition simulations.

The properties of this system provide novel possibilities for modelling language acquisition and psycholinguistic experiments from a natural language processing perspective. Moreover, both the mother and daughter agents are knowledge-based language models: All aspects of the system are parametrized and can straightforwardly be consulted by researchers employing the system. As most computational models of language acquisition are based on corpus data such as the CHILDES database [4], the multi-agent conversational framework such that two language models take part in an interaction but only one of these agents provides samples of the target language while the other acquires the mother language, enables addressing new research questions by conducting computational language acquisition experiments. Crucially, this design allows the adult language model to give feedback to the daughter.

In particular, the presented experiments demonstrate that the MODOMA provides an additional tool to perform language acquisition experiments from a cognitive perspective adding to experimentations with on the one hand human children and on the other hand adult subjects. These experiments show that the daughter language model can successfully acquire discrete grammatical categories such as function and content words [2] and noun, adjective and verb [3] in an unsupervised fashion. As part of the hybrid approach, mother exemplars are classified. For example, Figure 1 reproduced from [2, Figure 7] visualizing that functional and content words differ in frequency, illustrates that quantitative data can be used by the daughter language model to acquire these discrete grammatical categories. This procedure can be employed to represent structures that are similar to grammatical categories proposed by linguists for natural languages and acquired by human language learners. Thus, it is established that non-trivial grammatical knowledge has been acquired.

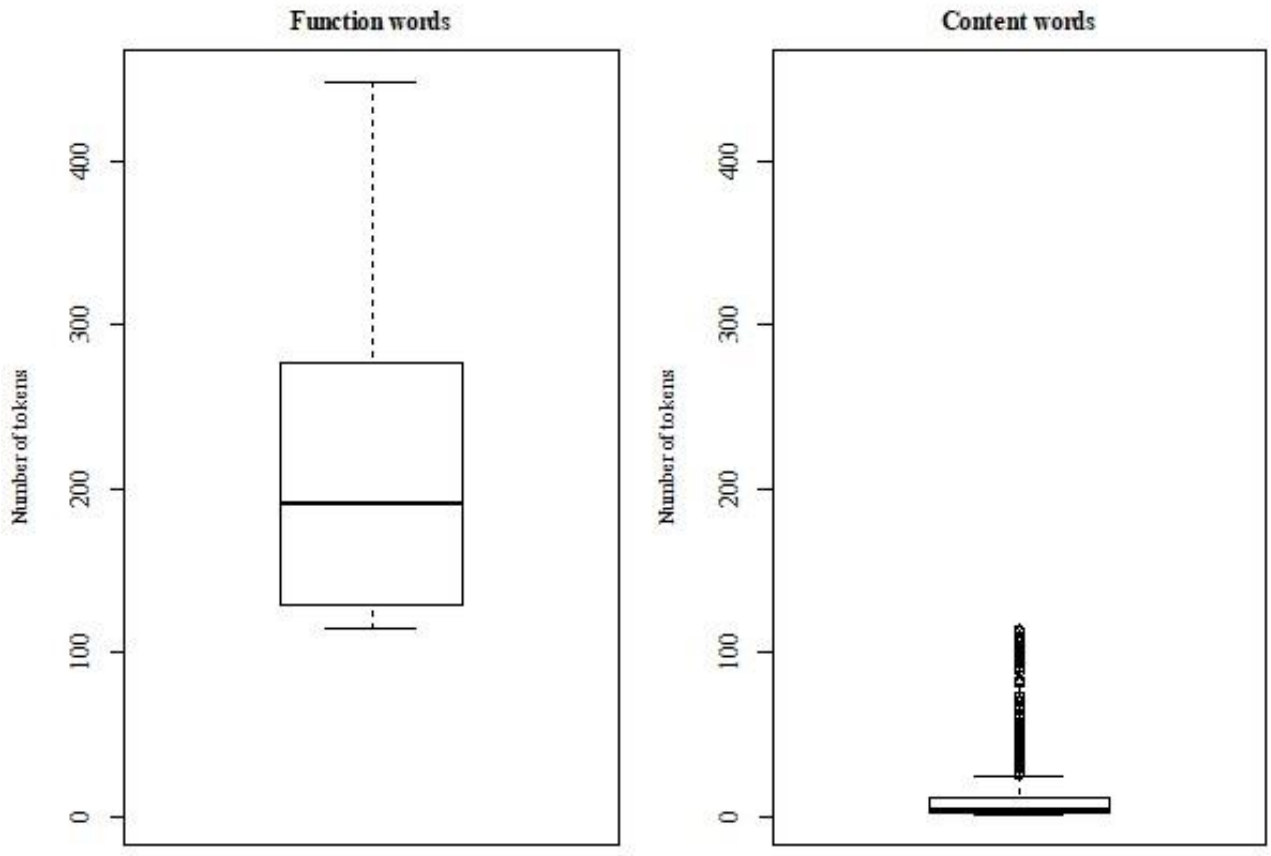

Figure 1: Boxplots depicting functional and content categories acquired by the daughter language agent based on 10,000 exemplar utterances generated by the adult agent reproduced from [2, Figure 7]

References

[1] Cremers, C., P.M. Hijzelendoorn, and H.G.B. Reckman. (2014). *Meaning versus grammar*. Leiden: Leiden University Press.

[2] Shakouri, D.P., C. Cremers, and N.O. Schiller. (2025). A knowledge-based language model: Deducing grammatical knowledge in a multi-agent language acquisition simulation. *Computational Linguistics in the Netherlands Journal 14*, 167-189. Retrieved from https://www.clinjournal.org/clinj/article/view/193

[3] Shakouri, D.P., C. Cremers, and N.O. Schiller. (2025). Unsupervised acquisition of discrete grammatical categories. *Computing Research Repository (CoRR), cs.CL/2503.18702v1*. doi: https://doi.org/10.48550/arXiv.2503.18702

[4] MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk* (3rd ed.). New York, NY: Routledge.