Modeling Human-Like Acquisition of Language and Concepts

Peter Lindes, Steven Jones

Center for Integrated Cognition
24 Frank Lloyd Wright Dr
Ann Arbor, Michigan 48105 USA
{peter.lindes,steven.jones}@cic.iqmri.org

Introduction

Generative models have achieved much success through offline batch training of billions of parameters using huge quantities of data. Humans, on the other hand, learn expertise piecemeal and rapidly in an incremental process guided by prior experience. Here we propose a theory and a research plan to build a computational model of human-like acquisition of language and related concepts.

Human language acquisition is a lifelong process of adding increments of knowledge (Tomasello 2003). We approximate this process with three stages: a *pre-verbal stage* of concept learning (Mandler and Pagán-Cánovas 2014), an *early language stage* of attaching language to those concepts (Spinelli, Fasolo, and Mesman 2017), and an *advanced language stage* where language can drive concept learning (Lakoff and Johnson 1980). In humans acquisition and real-time use of both language and concepts are tightly integrated in all these stages.

The proposed research program is based on a theory of human-like incremental and efficient acquisition of language and related concepts in the stages mentioned. We plan to build a computational model of this theory based on our prior research on cognitive architectures (Laird, Lebiere, and Rosenbloom 2017), interactive task learning (ITL; Laird et al. 2017), and human-like language comprehension (Lindes 2022). This model will be part of an autonomous ITL agent embodied in a simulated robotic agent that learns concepts and language incrementally over an extended period of time. We plan to evaluate both the theory and the model through experiments using a large published benchmark.

A Computational Model

Many researchers have built computational models of language processing (Winograd 1972; Lewis 1993; Ball 2011), language acquisition (Anderson 1974), or learning specific language features (Nishikawa and Morita 2020). None of this work describes acquisition of language and concepts sufficient to teach new tasks to a robot.

Our research group has over several years developed an ITL agent called Rosie (Mininger 2021; Kirk 2019) using the Soar cognitive architecture (Laird 2012). As part of this work Lindes (2022) has built a human-like language comprehension system called Lucia, and Jones (2022) has explored a model of event cognition. In the research proposed here we intend to extend this computational model to enable it to acquire new language and concepts from its experience.

Since human acquisition and use are tightly integrated, a human-like computational model of acquisition requires a human-like model of real-time processing. Our research on acquisition is based on the Lucia model of comprehension, which is human-like in several respects. Comprehension is an integral part of intelligent behavior by an autonomous agent. Knowledge of language is made up of many small composable units of form-meaning mapping that can be composed in an unlimited number of ways, using a formalism called Embodied Construction Grammar (ECG; Bergen and Chang 2013). Processing is done incrementally in simulated real time with immediate interpretation. Processing is done using the general cognitive mechanisms of the Soar model of human-like cognition (Newell 1990). The units of knowledge of language were acquired in small increments from individual experiences, by a human engineer. The focus of this research is to replace the human engineer with a human-like computational model of language acquisition.

Incremental acquisition begins with what we will call *acquisition events*. Each of these events is an experience where some element of knowledge is missing to be able to fully understand a language input or a new concept. The agent uses reasoning based on situational awareness to propose a new knowledge element that would fill the gap, then acts on that proposal. As with humans, later input may require the proposal to be modified or abandoned, or it may confirm it (Trueswell et al. 2013). This fits closely with Krashen's (1985) input hypothesis and Tomasello's (2003) account of human acquisition.

As more experiences with the same elements happen, the proposal is gradually generalized into declarative knowledge (Goldberg 2019). As these elements are used over and over, this declarative knowledge is automated as procedural knowledge. Once this automization has been accomplished, this particular element of language and its associated meaning can be processed very rapidly, as humans do.

We intend to implement this process in an advanced version of our ITL agent, and experiment with training it in a way similar to the three developmental stages described.

Evaluation Using a Published Benchmark

We plan to use tasks and training data from the ALFRED benchmark (Shridhar et al. 2019) to provide a series of learning experiences in a developmental trajectory with the three stages mentioned above. ALFRED has 8,055 “expert demonstrations,” each of which has a plan generated automatically to drive a simulated robot to perform a task in an AI2-THOR (Kolve et al. 2022) simulation, images of the robot successfully performing the task by following that plan, and three “annotations” written by three different Amazon Mechanical Turk (AMT) workers describing a goal and how to perform the task.

In the pre-verbal stage, our agent will learn concepts of objects and actions from observing and attempting to imitate the operation of the demonstration agent, corresponding to the things infants learn about objects in the world and how to manipulate them during the first year of life (Mandler and Pagán Cánovas, 2014). No language will be involved in this stage. As the demonstrator performs each action, our agent will model the action in the form of an event schema, use this to create an action model for that action, and later test this new knowledge by trying to imitate the whole task. Much of this corresponds to work done by Jones (2022) on event cognition in Soar. At the same time the agent will be learning concepts for classes of objects it sees in the demonstrations.

In the early language stage, somewhat like children learning from “motherese” (Spinelli, Fasolo, and Mesman 2017; Kuhl 2000), the agent will learn simple language. We will convert the task plans from ALFRED into simple language similar to that used by Rosie, where each plan’s language has a goal statement and a series of short sentences describing the steps required to perform the task. The agent will observe demonstrations, learning lexical and syntactic constructions and their grounding to the perceived environment incrementally from each individual acquisition event. Using the knowledge learned, the agent will be able to respond quickly to test scenarios never seen before and perform the described tasks with only the language input to guide it.

The advanced language stage involves learning to use the broader range of language in the task descriptions produced by the human AMT workers. The agent will reason about how to match each demonstration to the language provided. The agent will seek the knowledge it needs from several possible sources: using its own internal search mechanisms, asking questions of a human using previously-developed ITL techniques (Mininger 2021; Kirk 2019), and consulting a large language model (LLM) as we have done in previous experiments (Kirk, Wray, and Lindes 2023). We plan to explore the trade-offs and ways to combine these different knowledge sources as in previous work (Kirk et al. 2023).

Based on past experience with similar models, we expect that we can build this agent and design its experiments to show that while learning from only a small subset of the demonstrations provided in the benchmark the agent has much better success in executing test tasks based on learned knowledge when compared to systems based on training large neural networks on the whole benchmark data set.

Potential Broader Impact

If this effort succeeds, it will both shed additional light on the Cognitive Science of how human language acquisition works, and show much greater learning efficiency as compared to deep-learning approaches using the same ALFRED benchmark.

The proposed research program has many limitations. The implementation is in a simulated environment rather than a full physical embodiment, the scope of language involved is only a small part of full human language ability, and it includes no attempt to compare results to actual human data.

Nevertheless, motivating this approach with ideas from Cognitive Science and computationally implementating our theory will provide opportunities for extending our model’s performance and comparing it to real human data. Success in achieving substantially better task performance based on much less training data than needed by a deep learning approach will provide a powerful argument to the AI community that human-like learning approaches can make substantial contributions toward building artificial embodied agents that can collaborate with humans while constantly increasing their knowledge of how to communicate with a human partner and how to perform tasks in the world.

Furthermore, this proposed project incorporates substantial overlap between AI and Cognitive Science, which can help increase collaboration between these two historically connected fields.

Acknowldgements

The authors would like to express their gratitude to John Laird for his insightful ideas in preparing this abstract. We also express appreciation to the Center for Integrated Cognition for providing the environment we have needed to do this work.

References

Anderson, J. R. 1974. Language Acquisition by Computer and Child. Technical Report 55, Human Performance Center, The University of Michigan.

Ball, J. T. 2011. A Pseudo-Deterministic Model of Human Language Processing. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 495–500.

Goldberg, A. E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton University Press. ISBN 978-0-691-17425-9.

Jones, S. 2022. *A Cognitive Architecture Realization of Event Cognition Capabilities*. Ph.D. thesis, University of Michigan.

Kirk, J. R. 2019. *Learning Hierarchical Compositional Task Definitions through Online Situated Interactive Language Instruction*. Ph.D. thesis, University of Michigan.

Kirk, J. R.; Wray, R. E.; and Lindes, P. 2023. Improving Knowledge Extraction from LLMs for Task Learning through Agent Analysis. ArXiv:2306.06770 [cs].

Kirk, J. R.; Wray, R. E.; Lindes, P.; and Laird, J. E. 2023. Integrating Diverse Knowledge Sources for Online One-shot Learning of Novel Tasks. ArXiv:2208.09554 [cs].

Kolve, E.; Mottaghi, R.; Han, W.; VanderBilt, E.; Weihs, L.; Herrasti, A.; Deitke, M.; Ehsani, K.; Gordon, D.; Zhu, Y.; Kembhavi, A.; Gupta, A.; and Farhadi, A. 2022. AI2-THOR: An Interactive 3D Environment for Visual AI.

Krashen, S. 1985. *The Input Hypothesis: Issues and Implications*. New York: Longman.

Kuhl, P. K. 2000. A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97: 11850–11857.

Laird, J. E. 2012. *The Soar Cognitive Architecture*. Cambridge, MA: The MIT Press.

Laird, J. E.; Anderson, J.; Forbus, K. D.; Lebiere, C.; Salvucci, D.; Scheutz, M.; Thomaz, A.; Trafton, G.; Wray, R. E.; Mohan, S.; and Kirk, J. R. 2017. Interactive Task Learning. *IEEE Intelligent Systems*, 32: 6–21 (invited).

Laird, J. E.; Lebiere, C.; and Rosenbloom, P. S. 2017. A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine*, 38: 13.

Lakoff, G.; and Johnson, M. 1980. *Metaphors We Live By*. University of Chicago Press.

Lewis, R. L. 1993. *An Architectually-based Theory of Human Sentence Comprehension*. Ph.D. thesis, Carnegie Mellon University.

Lindes, P. 2022. *Constructing Meaning, Piece by Piece: A Computational Cognitive Model of Human Sentence Comprehension*. Ph.D. thesis, University of Michigan.

Mandler, J. M.; and Pagán-Cánovas, C. 2014. On defining image schemas. *Language and Cognition*, 6: 510–532.

Mininger, A. 2021. *Expanding Task Diversity in Explanation-Based Interactive Task Learning*. Ph.D. thesis, University of Michigan.

Newell, A. 1990. *Unified Theories of Cognition*. Harvard University Press.

Nishikawa, J.; and Morita, J. 2020. The Cognitive Modeling of Errors During the Japanese Phonological Awareness Formation Process. In *Proceedings of the 18th Annual Meeting of the International Conference on Cognitive Modelling (ICCM 2020)*.

Shridhar, M.; Thomason, J.; Gordon, D.; Bisk, Y.; Han, W.; Mottaghi, R.; Zettlemoyer, L.; and Fox, D. 2019. ALFRED: A Benchmark for Interpreting Grounded Instructions for Everyday Tasks.

Spinelli, M.; Fasolo, M.; and Mesman, J. 2017. Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, 44: 1–18.

Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Trueswell, J. C.; Medina, T. N.; Hafri, A.; and Gleitman, L. R. 2013. Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66: 126–156.

Winograd, T. 1972. Understanding Natural Language. *Cognitive Psychology*, 3: 1–191.