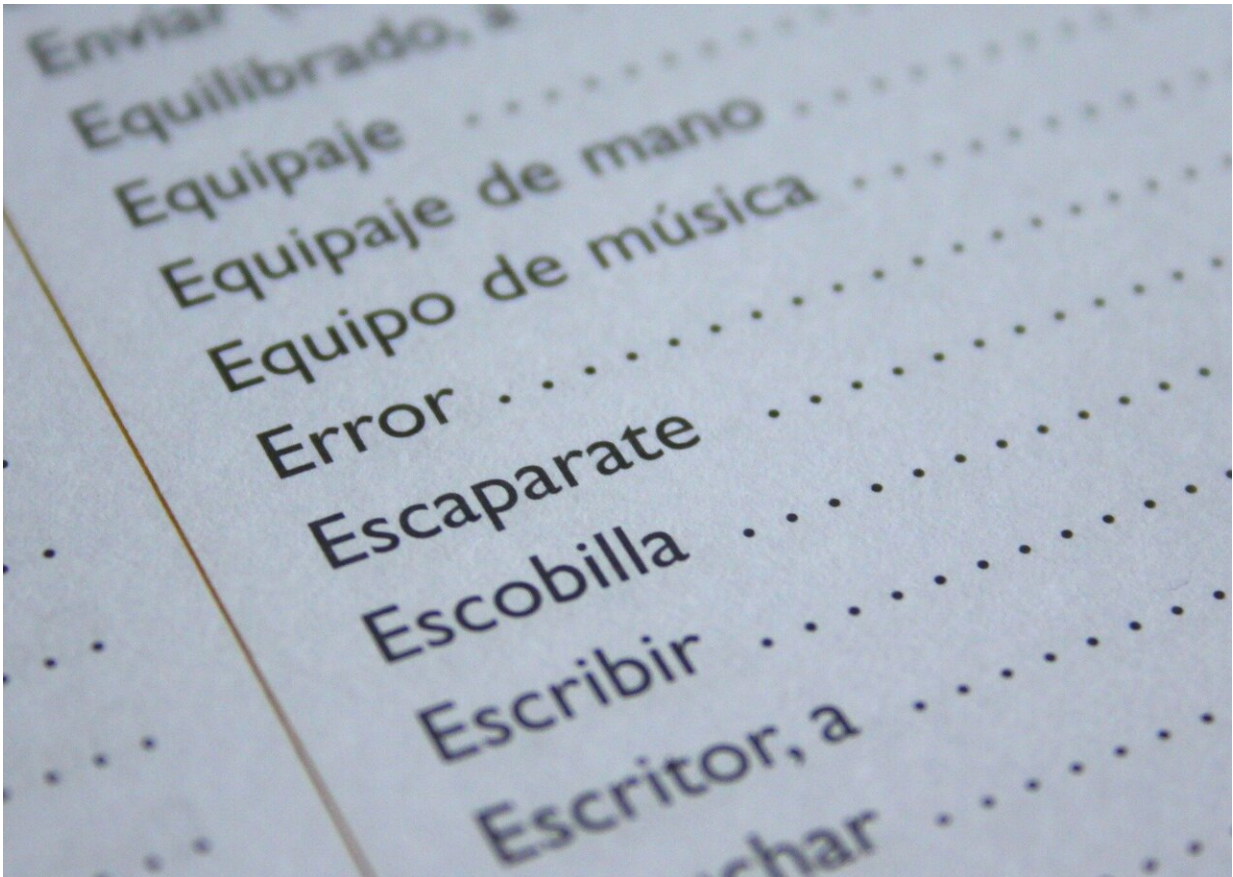


AI agents can learn to communicate effectively

July 16 2020



Credit: CC0 Public Domain

A multi-disciplinary team of researchers from Chalmers and University of Gothenburg has developed a framework to study how language

evolves as an effective tool for describing mental concepts. In a new paper, they show that artificial agents can learn how to communicate in an artificial language similar to human language. The results have been published in the scientific journal *PLOS ONE*.

This research lies on the border between cognitive science and [machine learning](#). There has been an influential proposal from cognitive scientists that all [human languages](#) can be viewed as having evolved as a means to communicate concepts in a near-optimal way in the sense of classical information theory. The Gothenburg researchers' method for training the artificial agents is based on reinforcement learning, which is an area of machine learning where agents gradually learn by interacting with an environment and getting feedback. In this case, the agents start without any linguistic knowledge and learn to communicate by getting feedback on how well they succeed in communicating a mental concept.

Reconstructing colors

"In our paper we have studied how agents learn to name mental concepts and communicate by playing a several rounds of a referential game consisting of a sender and a listener. We have especially focused on the color-domain which is well studied in cognitive science. The game works as follows; the sender sees a color and describes it by uttering a word from a glossary to the listener which then tries to reconstruct the color.

Both agents receive a shared reward based on how precise the listener's reconstruction was. The words in the glossary have no meaning at the outset; it is up to the agents to agree on the meaning of the words during multiple rounds of the game. We see that the resulting artificial languages are near-optimal in an information-theoretic sense and with similar properties as found in human languages," says Mikael Kågebäck, researcher at Sleepcycle, and whose Ph.D. dissertation at Chalmers contained some of the results presented in the paper.

Together with Asad Sayeed, researcher in computer linguistics at the Centre for Linguistic Theory and Studies in Probability (CLASP) at University of Gothenburg, and Devdatt Dubhashi, professor, and Emil Carlsson, Ph.D. student, in the Data Science and AI division at the Department of Computer Science and Engineering, he has now published the results.

"From a practical viewpoint, this research provides the [fundamental principles](#) to develop conversational agents such as Siri and Alexa that communicate with human language," says Asad Sayeed.

The underlying idea of learning to communicate through [reinforcement learning](#) is also interesting for research in social and cultural fields, for example for the project GRIPES, which studies dogwhistle politics, led by Asad Sayeed.

Useful in future research studies

"Cognitive experiments are very time consuming, as you often need to carry out careful experiments with human volunteers. Our approach provides a very powerful, flexible and inexpensive approach to investigate these fundamental questions. The experiments are fully under our control, repeatable and totally reliable. Thus our computational framework provide a valuable tool to investigate fundamental questions in cognitive science, language and interaction. For computer scientists it is a fertile area to explore the effectiveness of various learning mechanisms," says Devdatt Dubhashi.

"In the future, we want to investigate whether agents can develop communication similar to human language in other areas as well. One example is if our agents are able to reconstruct the hierarchical structures we observe in human language," says Emil Carlsson.

Long-standing question

The study stems from a long-standing central question in [cognitive science](#) and linguistics: whether, in all of the vast diversity of human languages, there are common universal principles. Classic work from the 20th century indicated that there were common properties in different languages in words to describe colors. Are there underlying principles accounting for these common properties?

A recent influential proposal from cognitive scientists is that there are indeed such common universal principles when viewed from the lens of information theory when languages are viewed as a means to communicate mental concepts making the most efficient use of resources.

A series of talks given at CLASP by Ted Gibson from MIT back in 2016, where he described results from experiments on human subjects chosen from different societies and cultures around the world, led to the question "what if the human subjects were substituted by artificial computer agents? Would they develop a [language](#) with similar universal properties?"

More information: Mikael Kågebäck et al. A reinforcement-learning approach to efficient communication, *PLOS ONE* (2020). [DOI: 10.1371/journal.pone.0234894](https://doi.org/10.1371/journal.pone.0234894)

Provided by University of Gothenburg

Citation: AI agents can learn to communicate effectively (2020, July 16) retrieved 10 April 2024 from <https://techxplore.com/news/2020-07-ai-agents-effectively.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.