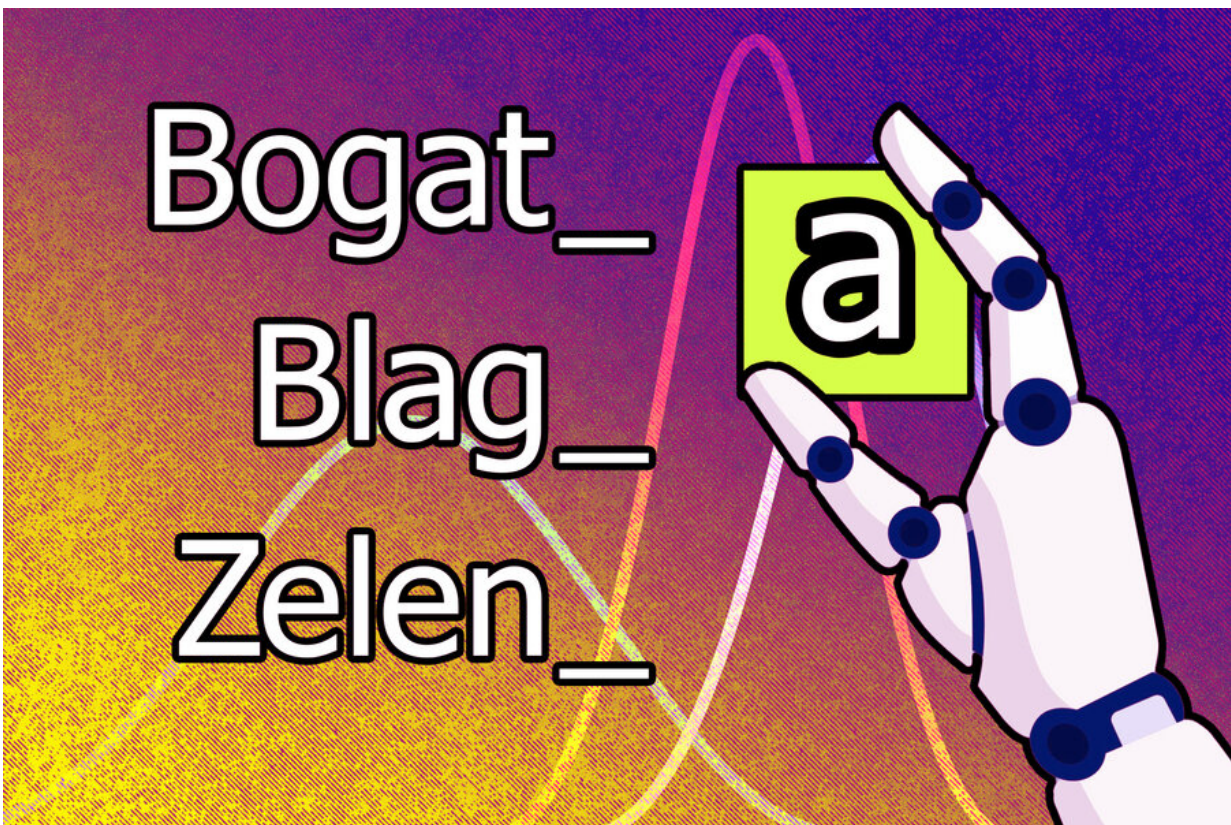# AI that can learn the patterns of human language

August 31 2022, by Adam Zewe

A new machine learning model might learn that the letter "a" must be added to the end of a word to make the masculine form feminine in Serbo-Croatian. For instance, the masculine form of the word "bogat" becomes the feminine "bogata." Credit: Jose-Luis Olivares, MIT

Human languages are notoriously complex, and linguists have long

thought it would be impossible to teach a machine how to analyze speech sounds and word structures in the way human investigators do.

But researchers at MIT, Cornell University, and McGill University have taken a step in this direction. They have demonstrated an artificial intelligence system that can learn the rules and patterns of human languages on its own.

When given words and examples of how those words change to express different grammatical functions (like tense, case, or gender) in one language, this machine-learning model comes up with rules that explain why the forms of those words change. For instance, it might learn that the letter "a" must be added to end of a word to make the masculine form feminine in Serbo-Croatian.

This model can also automatically learn higher-level language patterns that can apply to many languages, enabling it to achieve better results.

The researchers trained and tested the model using problems from linguistics textbooks that featured 58 different languages. Each problem had a set of words and corresponding word-form changes. The model was able to come up with a correct set of rules to describe those word-form changes for 60% of the problems.

This system could be used to study language hypotheses and investigate subtle similarities in the way diverse languages transform words. It is especially unique because the system discovers models that can be readily understood by humans, and it acquires these models from small amounts of data, such as a few dozen words. And instead of using one massive dataset for a single task, the system utilizes many small datasets, which is closer to how scientists propose hypotheses—they look at multiple related datasets and come up with models to explain phenomena across those datasets.

"One of the motivations of this work was our desire to study systems that learn models of datasets that is represented in a way that humans can understand. Instead of learning weights, can the model learn expressions or rules? And we wanted to see if we could build this system so it would learn on a whole battery of interrelated datasets, to make the system learn a little bit about how to better model each one," says Kevin Ellis, an assistant professor of computer science at Cornell University and lead author of the paper.

Joining Ellis on the paper are MIT faculty members Adam Albright, a professor of linguistics; Armando Solar-Lezama, a professor and associate director of the Computer Science and Artificial Intelligence Laboratory (CSAIL); and Joshua B. Tenenbaum, the Paul E. Newton Career Development Professor of Cognitive Science and Computation in the Department of Brain and Cognitive Sciences and a member of CSAIL; as well as senior author Timothy J. O'Donnell, assistant professor in the Department of Linguistics at McGill University, and Canada CIFAR AI Chair at the Mila—Quebec Artificial Intelligence Institute.

The research is published today in *Nature Communications*.

Looking at language

In their quest to develop an AI system that could automatically learn a model from multiple related datasets, the researchers chose to explore the interaction of phonology (the study of sound patterns) and morphology (the study of word structure).

Data from linguistics textbooks offered an ideal testbed because many languages share core features, and textbook problems showcase specific linguistic phenomena. Textbook problems can also be solved by college students in a fairly straightforward way, but those students typically have

prior knowledge about phonology from past lessons they use to reason about new problems.

Ellis, who earned his Ph.D. at MIT and was jointly advised by Tenenbaum and Solar-Lezama, first learned about morphology and phonology in an MIT class co-taught by O'Donnell, who was a postdoc at the time, and Albright.

"Linguists have thought that in order to really understand the rules of a human language, to empathize with what it is that makes the system tick, you have to be human. We wanted to see if we can emulate the kinds of knowledge and reasoning that humans (linguists) bring to the task," says Albright.

To build a model that could learn a set of rules for assembling words, which is called a grammar, the researchers used a machine-learning technique known as Bayesian Program Learning. With this technique, the model solves a problem by writing a [computer program](#).

In this case, the program is the grammar the model thinks is the most likely explanation of the words and meanings in a linguistics problem. They built the model using Sketch, a popular program synthesizer which was developed at MIT by Solar-Lezama.

But Sketch can take a lot of time to reason about the most likely program. To get around this, the researchers had the model work one piece at a time, writing a small program to explain some data, then writing a larger program that modifies that small program to cover more data, and so on.

They also designed the model so it learns what "good" programs tend to look like. For instance, it might learn some general rules on simple Russian problems that it would apply to a more complex problem in

Polish because the languages are similar. This makes it easier for the model to solve the Polish problem.

Tackling textbook problems

When they tested the model using 70 textbook problems, it was able to find a grammar that matched the entire set of words in the problem in 60% of cases, and correctly matched most of the word-form changes in 79% of problems.

The researchers also tried pre-programming the model with some knowledge it "should" have learned if it was taking a linguistics course, and showed that it could solve all problems better.

"One challenge of this work was figuring out whether what the model was doing was reasonable. This isn't a situation where there is one number that is the single right answer. There is a range of possible solutions which you might accept as right, close to right, etc.," Albright says.

The model often came up with unexpected solutions. In one instance, it discovered the expected answer to a Polish language problem, but also another correct answer that exploited a mistake in the textbook. This shows that the model could "debug" linguistics analyses, Ellis says.

The researchers also conducted tests that showed the model was able to learn some general templates of phonological rules that could be applied across all problems.

"One of the things that was most surprising is that we could learn across languages, but it didn't seem to make a huge difference," says Ellis. "That suggests two things. Maybe we need better methods for learning across problems. And maybe, if we can't come up with those methods,

this work can help us probe different ideas we have about what knowledge to share across problems."

In the future, the researchers want to use their model to find unexpected solutions to problems in other domains. They could also apply the technique to more situations where higher-level knowledge can be applied across interrelated datasets. For instance, perhaps they could develop a system to infer differential equations from datasets on the motion of different objects, says Ellis.

"This work shows that we have some methods which can, to some extent, learn inductive biases. But I don't think we've quite figured out, even for these textbook problems, the inductive bias that lets a linguist accept the plausible grammars and reject the ridiculous ones," he adds.

"This work opens up many exciting venues for future research. I am particularly intrigued by the possibility that the approach explored by Ellis and colleagues (Bayesian Program Learning, BPL) might speak to how infants acquire language," says T. Florian Jaeger, a professor of brain and cognitive sciences and computer science at the University of Rochester, who was not an author of this paper.

"Future work might ask, for example, under what additional induction biases (assumptions about universal grammar) the BPL approach can successfully achieve human-like learning behavior on the type of data infants observe during language acquisition. I think it would be fascinating to see whether inductive biases that are even more abstract than those considered by Ellis and his team—such as biases originating in the limits of human information processing (e.g., memory constraints on dependency length or capacity limits in the amount of information that can be processed per time)—would be sufficient to induce some of the patterns observed in human languages."

Provided by MIT Computer Science & Artificial Intelligence Lab