MONOGRAPH
METODE SCIENCE STUDIES JOURNAL (2024). University of Valencia.
ISSN: 2174-3487. eISSN: 2174-9221.
Submitted: 14/11/2023. Approved: 15/04/2024
https://doi.org/10.7203/metode.15.27692

LANGUAGE EVOLUTION *IN SILICO*

From large-scale data to artificial agents creating languages from scratch

Thomas Brochhagen

We all speak a language and have intuitions about it: from its vocabulary to the way words are put together according to its grammar. However, much is still to be understood about the processes that make language even possible and those that shape its evolution. Recent computational advances have enabled us to address these issues from new angles. This article highlights methods and findings that the age of computation has given rise to, from learning from large-scale data from thousands of languages to the evolution of languages created by artificial intelligence.

Keywords: **language**, **evolution**, **artificial intelligence**, **typology**, **universals**.

By contrast to biological specimens, languages do not leave traces on the fossil record. This makes studying the evolution of language a difficult task. We do not have direct evidence about the way our ancestors' speech was structured nor do we know what changes it went through until it took its present form, after thousands of years. The most infamous reaction to these difficulties was the Linguistic Society of Paris' ban of discussions on the evolution of language in 1866 (Corballis, 2008).

Difficulties notwithstanding, the question of how languages evolve and what this tells us about ourselves has continued to fascinate scholars. Not only has research on these topics continued to this day but it has markedly begun to pick up its pace in the age of the computer. In this article, we survey a selection of methods and findings that recent advances in computation have shed light on.

With no primordial linguistic specimens to dissect, research on language evolution has traditionally been concerned with three broad types of inquiry. First, regular patterns can be inferred from the historical records of languages that have a written tradition. Second, in order to address questions like whether language arose gradually or all at once, the cognitive capabilities of modern humans are studied and compared with the cognitive capabilities our ancestors may have had; or with the cognitive capabilities of our closest living relatives, like chimpanzees and gibbons. Third, data from currently spoken languages from across the globe are gathered and compared. In this way the diversity of present-day languages serves as a window into the evolutionary processes of which they are the outcomes. Since modern languages are the expression of the evolutionary trajectories of their past, their commonalities and differences can give us important clues.

«Our current age of accessible and cheap computation has added new capacities and dimensions to the study of language evolution»

All of these three types of inquiry are still productively pursued to this day. However, our current age of accessible and cheap computation has added new capacities and dimensions to the study of language evolution. On the one hand, the aforementioned efforts are now supported by powerful algorithms that allow us to better quantify evidence for competing theories of language evolution, to build better maps of the genealogical relationship of languages, and to better

HOW TO CITE:
Brochhagen, T. (2025). Language evolution in silico: From large-scale data to artificial agents creating languages from scratch. *Metode Science Studies Journal, 15*, 19–29. **https://doi.org/10.7203/metode.15.27692**

predict future change. On the other hand, language evolution is now also studied artificially, with artificial intelligence making up languages of its own.

HOW MEANING EVOLVES ACROSS LANGUAGES, CULTURES, AND TIME

Every language has its idiosyncrasies. In English the meanings finger and toe are expressed by two different words. In Catalan this is done with a single word: dit. Surprisingly, these two meanings are also co-expressed in over 130 different languages across the globe (Rzymski et al., 2020). That is, more than 130 languages use a single word to express these two meanings: From the language Secoya, spoken in the Amazon between Ecuador and Colombia, to Papua New Guinea's Takia.
We find a lot of similar patterns across the world's vocabularies. Leaf and feather are often expressed by the same word, and so are good and beautiful, small and young, and hole and cave. These patterns cannot be explained by a shared ancestry or geographic proximity. What is to be explained then – through the study of language evolution – is what it is about the relationship between finger and toe, or that between hole and cave or leaf and feather, that attracts them to each other in so many languages. In addressing these questions we are ultimately interested in what this tells us about the way we, as humans, organize meaning.

Being able to address this issue based not only on a handful of languages but based on evidence from hundreds to thousands of them is a major recent development in the field. Some examples of new large-scale resources include the Database of Cross-Linguistic Colexifications (Rzymski et al., 2020), registering the way meanings are expressed in over 3,000 languages; or Kinbank, Database of Kinship Terminology (Passmore et al., 2023) which collects data on how languages express kinship (e.g., whether they have a word for uncle; or if they differentiate between paternal and maternal grandfather with different words). To be clear, while these resources are new, they build on the fundamental field work of linguists who went – and still go – into the world to document languages. The novelty is that this kind of data is now conveniently available as digitalized resources and in unified formats, and that we now have the computational power and methods to process them automatically.

If we want to know what makes some meanings more likely to be expressed by the same word than others then the next step comes in capturing the relationships between them. For this to work, meanings need to be somehow represented. This can be done

Imagen de Freepik

Every language has its own peculiarities. Many languages use the same word to express two different concepts, such as good and beautiful, small and young, or hole and cave. This pattern cannot be explained by common ancestry or geographical proximity.

«Meaning is organized in a regular fashion across human languages, with the evolution of all languages following predictable patterns»

by leveraging modern computational techniques. For instance, the special bond between finger and toe as well as that between hole and cave is likely to be – at least to some extent – based on their visual similarity. This idea can be operationalized with modern vision models, which process large amounts of images to arrive at a computational representation for them. In this way, we can get a measure of how similar fingers and toes are, visually. Other resources can analogously be leveraged to approximate, for instance, how similar the context of use of different meanings is, how close they are in associative memory, and so on. In a nutshell, we can computationally represent different ways in which two meanings may be (dis)similar through modern techniques, building on interdisciplinary work from psychology, artificial intelligence, natural language processing, and statistics.

What we end up with is a resource that tells us how meanings are expressed across many different languages and in what relationship these meanings stand. That is, finger and toe are visually similar; they appear in similar contexts; and they are close associates (if I tell you finger you may automatically think toe). By contrast, while a finger may be visually similar to

a sausage, they certainly do not appear in the same context nor are they closely associated with one another. A number of important recent findings have been made based on this kind of information. First, words that are similar in meaning are universally attracted to each other (Xu et al., 2020). That is, they are more likely to be expressed by the same word, irrespective of whether it is Mandarin or Dutch. This is likely so because expressing similar meanings together in a single word makes it easier to learn them. Second, this universal tendency has a limit: meanings that are so similar that they may be confused for one another are not attracted to each other (Brochhagen & Boleda, 2022). For instance, using the same word for Thursday and Wednesday would not survive the test of time. Instead, doing so for finger and toe works just fine because it is usually clear which one we mean in context. Third, these patterns may be due to a universal tendency for languages to be both simple («use as few words as possible») but effective («use different words for meanings that we care to distinguish», like Tuesday and Thursday). In other words, languages are shaped by the need to maintain a balance for the need to be simple but informative (e.g., Kemp & Regier, 2012; Zaslavsky et al., 2018). A language too simple is not useful to talk to others. A language too complex is unwieldy or impossible to learn. The evolution of language strikes a balance between the two, explaining how meaning is organized across languages. Lastly, the same factors that predict whether two meanings are attracted to each other across languages also explain small children's language use (Brochhagen et al., 2023): A child that calls a boat a «car», or calls a cow a «dog», or calls a lamp «sun» builds on the same relationship between meanings that is universally reflected by the languages we speak. When a small child lacks a word and uses another to get its point across (like calling a cow a «doggy»), at a fundamental level, it is doing something analogous to what Catalan or Chechen speakers do by calling fingers and toes by the same name.

Taken together, the above indicates that meaning is organized in a regular fashion across human languages, with the evolution of all languages following predictable patterns that can be distilled from data. These patterns are a result of the forces that shape language. Particularly, a push toward efficient languages that are both simple and informative: simple enough to learn and use them but informative enough so that we can understand each other.

ARTIFICIAL AGENTS INVENTING LANGUAGES FROM SCRATCH

A radically different way to approach the question of how languages evolve is through the lens of artificial intelligence. There are two major motivations to study language evolution in silico. First, human languages are a product of our biology, ecology, and culture. The way we process, perceive and interact with the world is what ultimately determines the properties of human language. However, there is much debate about precisely which biological, ecological or cultural factors are responsible for particular linguistic properties. Studying artificial agents and their languages is accordingly a promising venue to explore. By contrast to humans, we know and control their «biology», «ecology», and «culture» to the last detail. Second, although artificial intelligence has made impressive progress in the last couple of years – most prominently, with the rise of ChatGPT and its kind – artificial languages remain far from human. They lack the flexibility and open-endedness of our speech. In studying the evolution of artificial language the hope is thus that we can learn both something about language but also how to enable artificial agents to come up with better, more human-like, languages (Lazaridou & Baroni, 2020).

«We should not assume that the way we humans solve communicative tasks is the only way to go about it»

A popular setup to encourage artificial agents to create their own languages is called a reference game. Many variations exist. The simplest setup consists of a sender and a receiver. The sender's task is to make the receiver pick out a particular image out of an array of candidate images. For instance, as illustrated in Figure 1, the task may be to convey the image with the white puppy. The sender then sends a message of their creation to the receiver. The receiver interprets it and selects an image (e.g., one of the three in Figure 1). They then both receive feedback about whether they were successful or not. The game then repeats for many more rounds.

How can language emerge from this kind of situation? At the beginning the sender has no established way to convey anything to the receiver. The best they can do is to send a random message. Conversely, the receiver has no way to interpret the message and is forced to randomly guess the intended image. No information is transferred. However, over time, the initially random messages will acquire meaning through sheer force of repetition: If the receiver happens to guess the image correctly, the agents will be more likely to use that same message for that same kind of image in the future. This situation

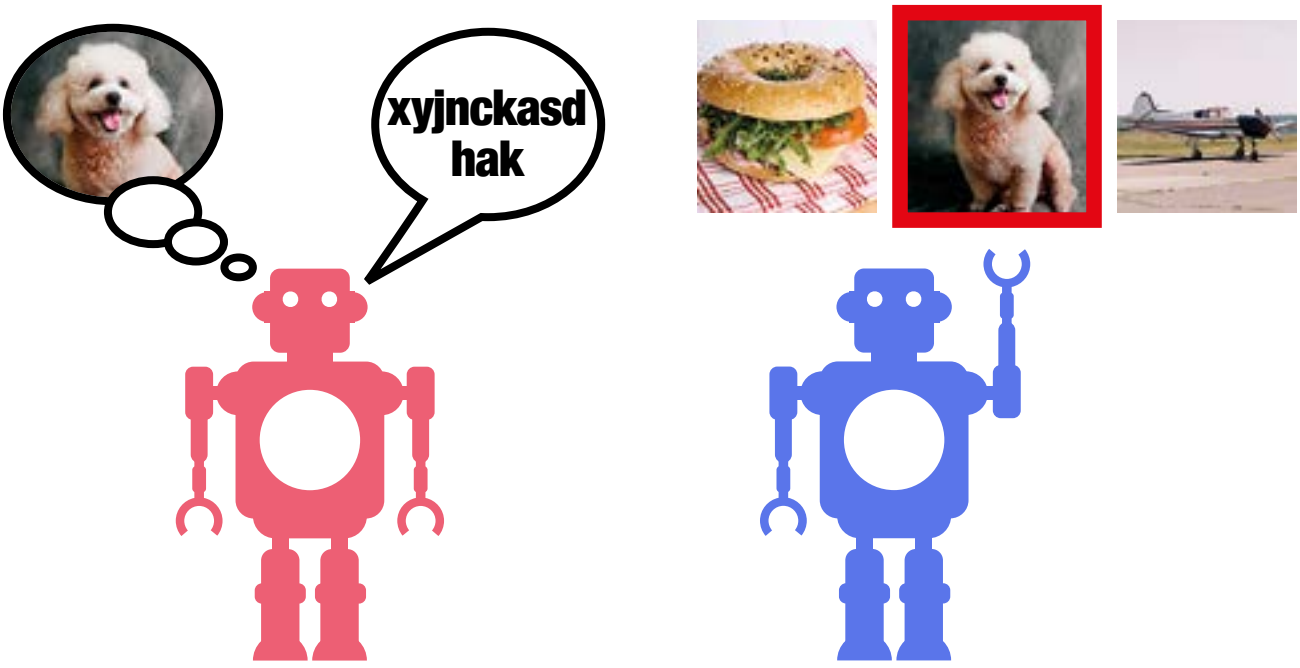

Figure 1. Two artificial agents talking about images similar to those in the ImageNet database (Deng et al., 2009). The sender (left) wants the receiver (right) to pick out an image (the white dog). To do so, the sender sends a message. The receiver then guesses the image the sender had in mind, and the game repeats with a different set of images. In most setups, over time, agents invent languages that enable them to communicate with high accuracy.

is reminiscent of playing many rounds of charades: charades that were successful in the past will be reused, making the game much smoother. If you play a lot of charades with your friends you will likely develop an intricate gestural vocabulary that allows you to communicate fast and well. The same principle is at work for the artificial agents.

What is interesting is not only that the agents are able to develop their own language ex nihilo, but the question of what kinds of languages they come up with. The results in this field have been intriguing, to say the least. A major lesson from artificial language emergence is that we should not assume that the way we humans solve communicative tasks is the only way to go about it. An infamous example is found in Bouchacourt and Baroni (2018), using a setup akin to that in Figure 1. After the agents had come up with a language of their own, Bouchacourt and Baroni tested them on how well they could talk about the images they were trained on (e.g., puppies, airplanes and foodstuff like those in Figure 1) as well as on how well they could talk about strongly distorted versions of the images that they had not seen before. These distortions essentially looked like colorful TV static noise. Surprisingly, the artificial agents could communicate almost as well about the distorted images than about their original counterparts. How is it that an artificial language that is good to talk about images of dogs, food and airplanes is also useful to talk about distorted versions that look nothing like the original? This puzzling result is explained by not thinking like humans. Humans would likely create languages to talk about the objects found in the pictures: dog, bagel, and so on. The artificial agents, instead, had apparently come up with a language to talk about shallow visual features of the images. That is to say, they were talking about something like the hue of certain pixels found in the images, rather about the bagels, dogs, or airplanes that were present. These shallow features were still preserved in the distorted images but impossible to pick out by humans. The lesson is that, while we humans naturally interpret and describe images based on the objects or scenes that they depict, artificial agents are quite happy to talk about pixels and hues. What is natural to us can thus be quite different to what is natural to them. Therefore, if we want them to come up with similar languages to ours, we have to engineer them to view the world as we view it.

«Languages are shaped by the need to maintain a balance for the need to be simple but informative»

Another finding along these lines is that artificial languages created in this way also do not have

the tendency of human languages for succinctness. A universal statistical fingerprint common to all human languages is that the most frequent words are the shortest. For instance, in English the words the and a are very frequent and short compared to romboidal, a rather infrequent word. This is efficient in the sense of saving effort for speakers and listeners since the words we use the most are the shortest and easiest to produce. Artificial agents have the opposite tendency (Chaabouni et al., 2019). Frequent meanings tend to be expressed with the longest possible words; and all words tend to be longer than necessary. The reason for this behavior is simply that while we humans care if words or sentences are very long, a machine does not. Using longer messages instead allows them to make messages that are easier to interpret by other artificial agents since it gives them more characters to encode their intended meaning.

The above are just two examples out of a wealth of findings from this nascent field. At first, it may seem obvious that in order for artificial languages to be more human-like their users need to care about the same things we do (in the above examples: reduce word length and talk about objects and not pixels). However, it is far less evident – before conducting these studies – what it is exactly about how we interact and process information that matters. In other words, these results teach us valuable lessons about which parts of the human experience shape language. This makes them an ideal testbed for the evolution of language.

THE FUTURE OF LEARNING FROM THE PAST

This article touched on a few novel ways in which the evolution of language is being studied: through large-scale resources and through artificial language emergence experiments. Their use is enabled by computational resources and methods that were not available to us even a few years ago. The old challenges still remain though. The evolution of language remains abductive and hypothesis driven; and it is unlikely that we will ever isolate the precise ingredients and processes that constitute it. After all, human language is a complex product of many intertwined factors. Notwithstanding, the ever-expanding toolbox that we have at our disposal enables us to continue to refine and challenge current theories.

As with studying the past, predictions about the future developments of this field are hard to make. One thing we can be relatively certain of is that linguistic diversity will play an increasingly important role in the next few years. Current computational approaches tend to be very data-hungry. This has meant that most current large-scale research is based on a few select languages that are well represented in recorded text and speech. However, efforts to provide a less biased picture of the world's linguistic diversity at a scale, such as the DoReCo corpus (Seifart et al., 2022) or BLOOM (BigScience Workshop, 2023), signal that things are changing also in this respect. The continued digitalization of underrepresented languages and dialects will thus likely usher in a wealth of new data, enabling us to test new and old ideas. The field keeps evolving, just as languages do.

REFERENCES

Bouchacourt, D., & Baroni, M. (2018). How agents see things: On visual representations in an emergent language game. In E. Riloff, D. Chiang, J. Hockenmaier & J. Tsujii, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Progressing (p. 981–985). Association for Computational Linguistics.

BigScience Workshop. (2023). BLOOM: A 176B-parameter open-access multilingual language model. arXiv. https:/doi.org/10.48550/arxiv.2211.05100

Brochhagen, T., & Boleda, G. (2022). When do languages use the same word for different meanings? The Goldilocks principle in colexification. Cognition, 226, 105179. https://doi.org/10.1016/j.cognition.2022.105179

Brochhagen, T., Boleda, G., Gualdoni, E., & Xu, Y. (2023). From language development to language evolution: A unified view of human lexical creativity. Science, 381(6656), 431–436. https://doi.org/10.1126/science.ade7981

Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2019). Anti-efficient encoding in emergent communication. In Proceedings of NeurIPS 2019 (33d Conference on Neural Information Processing Systems) (p. 6290–6300). Curran Associates.

Corballis, M. C. (2008). Not the last word. American Scientist, 96(1), 68–70.

Deng, J., Dong, W., Socher, R., Li, L.-J.,Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In IEEE Computer Vision and Pattern Recognition (CVPR) (p. 248–255). https://doi.org/10.1109/CVPR.2009.5206848

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. Science, 336(6084), 1049–1054. https://doi.org/10.1126/science.1218811

Lazaridou, A., & Baroni, M. (2020). Emergent multi-agent communication in the deep learning era. arXiv. https://doi.org/10.48550/arXiv.2006. 02419

Rzymski, C., Tresoldi, T., Greenhill, S. J., Wu, M.-S., Schweikhard, N. E., Koptjevskaja-Tamm, M., Gast, V., Bodt, T. A., Hantgan, A., Kaiping, G. A., Chang, S., Lai, Y., Morozova, N., Arjava, H., Hübler, N., Koile, E., Pepper, S., Proos, M., Van Epps, B., ... List, J.-M. (2020). The database of cross-linguistic colexifications, reproducible analysis of cross- linguistic polysemies. Scientific Data, 7, 13. https://doi.org/10.1038/s41597-019-0341-x

Seifart, F., Paschen, L., & Stave, M. (2022). Language Documentation Reference Corpus (DoReCo) 1.2. [Archive material]. Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2). https://doi.org/10.34847/nkl.7cbfq779

Passmore, S., Barth, W., Greenhill, S. J., Quinn, K., Sheard, C., Argyriou, P., Birchall, J., Bowern, C., Calladine, J., Deb, A., Diederen, A., Metsäranta, N. P., Araujo, L. H., Schembri, R., Hickey-Hall, J., Honkola, T., Mitchell, A., Poole, L., Rácz, P. M., ... Jordan, F. M. (2023). Kinbank: A global database of kinship terminology. PLOS ONE, 18(5), e0283218. https://doi.org/10.1371/journal.pone.0283218

Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. Cognition, 201, 104280. https://doi.org/10.1016/j.cognition.2020.104280

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. Proceedings of the National Academy of Sciences, 115(31), 7937–7942. https://doi.org/10.1073/pnas. 1800521115

THOMAS BROCHHAGEN. Tenure-track professor in computational cognitive science at the Universitat Pompeu Fabra's Department of Translation and Language Sciences (Spain). His research interests include language evolution, artificial intelligence, Bayesian models, and statistics.
✉ thomas.brochhagen@upf.edu