

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

The next Theory of Evolution

### Permalink

<https://escholarship.org/uc/item/37t8k62b>

### Author

Begus, Gasper

### Publication Date

2026-09-30

### DOI

10.5281/zenodo.23071867

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

# The Next *Theory of Evolution*

Gašper Beguš\*

University of California, Berkeley  
Berkeley, CA 94720, United States

\*Corresponding author; [begus@berkeley.edu](mailto:begus@berkeley.edu)

September 30, 2026

## Abstract

New evidence from large language models and animal communication challenges human linguistic exceptionalism. Drawing on historical parallels between linguistics and evolutionary biology, this comment argues that new evidence from non-human linguistic behavior could transform humanity's relationship with the natural world.

For much of Western history, Greek and Latin were considered the richest and most admirable languages, while the tongues people actually spoke were dismissed as lowly vernaculars, impoverished kitchen languages. Who in their right mind could compare the language in which Homer, Aristotle, or Virgil created their masterpieces with the dialects of everyday language such as Middle French or English? Parallel to this value judgment, humans have long been considered the lords of the natural world, distinct from irrational animals. Who could say that the human with reason, language, culture, and civilization has anything to do with animals like chimpanzees, goats, or cats?

These views no longer hold. We now know that we share evolutionary ancestry with chimpanzees, goats, and cats. Reframing our position relative to animals enabled Darwin to propose his *Theory of Evolution*, which forever transformed science. This reframing was predated and enabled by a much less famous development in science: the realization that Greek and Latin might not be god-given, but developed from earlier languages that “perhaps no longer exist.”

The stage for Darwin's theory of evolution was thus set by linguists seventy years before his *On the Origin of Species*. British colonialist scholars were starting to learn the sacred text of India. One of them was Sir William Jones, a classically trained scholar. Having learned Sanskrit and realized that it is remarkably similar to Homeric Greek, in 1786 he addressed the Asiatic Society saying that “The Sanscrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin” and that “no philologist could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists”. In the decades that followed, the brightest minds started taking this hypothesis seriously. Jacob Grimm, one of the two Grimm brothers, for example, proposed

one of the first linguistic laws, suggesting that languages change in a regular fashion, and that the proto-language that perhaps no longer exists can be reverse-engineered. The evolutionary principle was first born from and tested out on linguistic data. Darwin himself uses linguistics to argue for his new theory[1].

We are on the cusp of another revolution of this type. Until recently, only humans were able to master human language perfectly. No computational model came even close. Transformer-based large language models (LLMs) are turning this assumption upside down. A growing body of work suggests that transformers are able to learn highly complex rules of language.[2] In a paper on metalinguistic abilities of LLMs, I showed that the models perform well on one of the most complex and long considered uniquely human properties of language—recursion [3]. Not only that, the models also showed metalinguistic abilities: they could reason about linguistic structure, not merely produce fluent language.

While LLMs tell us that the human might not be the only entities capable of learning human language, they fail at realistically replicating how humans actually acquire language. In my research, I build interpretable AI models that learn language more like human babies do [4]. My models learn from raw sound, by imitating and imagining data. I was able to show that these models discover symbolic-like structure in language only by virtue of being trained to informatively imitate data they hear from the environment. These models are also more interpretable than other architectures, and closely parallel the human brain, as I was able to show by comparing internal representations of the brain and internal layers [5].

Reframing how we think of language—as *informative imitation*—allowed me the next step. Given that the models were able to learn human language without supervision, I decided to train them on whale vocalizations. As part of Project CETI[6], a nonprofit organization dedicated to understanding and translating the communications of sperm whales, I used my AI interpretability technique for understanding what the model had learned [7], which returned a surprising result: sperm whale vocalizations were predicted to be informative not only in the traditional dimensions such as the number of clicks and the timing, but in the spectral domain as well, which was not considered a meaningful candidate at the time.

Sperm whale vocalizations[8] were long believed to be a Morse code-like system, which sounds alien and not very human-like. After our interpretability technique pointed me to the spectrum, I conducted extensive analysis of whale clicks—combining traditional philological, linguistic, and bioacoustic work—and uncovered previously unobserved structured patterns in sperm whale vocalizations. The crucial insight was *timing*. Sperm whale clicks are not merely clicks, they are very slow vocalic pulses. When I removed silences, spectral patterns parallel to human vowels began to appear. The seemingly alien system became much more human-like. We report two discretely distributed types, the *a*- and the *i*-coda vowel [9] and showed that these vowel-like patterns are highly structured, repeated by several whales, and systematically parallel human vowels

[10]. The same approach—stepping outside of our human-centred perception of time—that allowed me to uncover sperm whale vowels has proven useful in demonstrating that orcas have not only vowel-like but also consonantal patterns [11]. These results suggest that speech-like organization in cetacean vocalizations had been hidden in plain sight.

Both AI and new advances in animal research[12] are forcing us to rethink human exclusivity once again. Almost 250 years ago, dethroning Latin and Greek and comparing the two to Sanskrit allowed us to discover the evolutionary principle. Seventy years later, expanding such thinking to animals allowed us to answer one of the deepest questions in science: where humanity came from. Today, the primacy of human language as a whole is being questioned by AI models. Our ability to rethink the exclusivity and uniqueness of humans will likely be crucial for future scientific discovery. Perhaps the next *Theory of Evolution* is the realization that we are just one of many possible and diverse intelligences and that our legal, societal, and value systems should be adjusted accordingly. We form our identities always in relation to the *other*. The other used to be the animals, now the other is artificial, silicon intelligence. This provides an opportunity to identify as a biological intelligence that will unite the natural world. At a time when AI is starting to outperform us in many cognitive domains, building alliances and identities together with other biological intelligences might be more important than ever.

## References

- [1] C. Darwin, *On the Origin of Species* (John Murray, London), 1 ed. (1859), [https://darwin-online.org.uk/converted/published/1859\\_Origin\\_F373/1859\\_Origin\\_F373.html](https://darwin-online.org.uk/converted/published/1859_Origin_F373/1859_Origin_F373.html).
- [2] R. Futrell, K. Mahowald, How linguistics learned to stop worrying and love the language models. *Behavioral and Brain Sciences* **49**, e198 (2026), doi:10.1017/S0140525X2510112X.
- [3] G. Beguš, M. Dabkowski, R. Rhodes, Large Linguistic Models: Investigating LLMs’ Metalinguistic Abilities. *IEEE Transactions on Artificial Intelligence* **6** (12), 3453–3467 (2025), doi:10.1109/TAI.2025.3575745.
- [4] G. Beguš, CiwGAN and fiwGAN: Encoding information in acoustic data to model lexical learning with Generative Adversarial Networks. *Neural Networks* **139**, 305–325 (2021), doi:<https://doi.org/10.1016/j.neunet.2021.03.017>, <https://www.sciencedirect.com/science/article/pii/S0893608021001052>.
- [5] G. Beguš, A. Zhou, T. C. Zhao, Encoding of speech in convolutional layers and the brain stem based on language experience. *Scientific Reports* **13** (1), 6480 (2023), doi:10.1038/s41598-023-33384-9, <https://doi.org/10.1038/s41598-023-33384-9>.

- [6] J. Andreas, *et al.*, Toward understanding the communication in sperm whales. *iScience* **25** (6), 104393 (2022), doi:<https://doi.org/10.1016/j.isci.2022.104393>, <https://www.sciencedirect.com/science/article/pii/S2589004222006642>.
- [7] G. Beguš, Generative Adversarial Phonology: Modeling Unsupervised Phonetic and Phonological Learning With Neural Networks. *Frontiers in Artificial Intelligence* **3**, 44 (2020), doi:10.3389/frai.2020.00044, <https://www.frontiersin.org/article/10.3389/frai.2020.00044>.
- [8] W. A. Watkins, W. E. Schevill, Sperm whale codas. *The Journal of the Acoustical Society of America* **62** (6), 1485–1490 (1977), doi:10.1121/1.381678, <https://doi.org/10.1121/1.381678>.
- [9] G. Beguš, R. L. Sprouse, A. Leban, M. Silva, S. Gero, Vowel- and Diphthong-Like Spectral Patterns in Sperm Whale Codas. *Open Mind* **9**, 1849–1874 (2025), doi:10.1162/OPMI.a.252, <https://doi.org/10.1162/OPMI.a.252>.
- [10] G. Beguš, M. Dabkowski, R. L. Sprouse, D. F. Gruber, S. Gero, The phonology of sperm whale coda vowels. *Proceedings of the Royal Society B: Biological Sciences* **293** (2069), 20252994 (2026), doi:10.1098/rspb.2025.2994.
- [11] G. Beguš, M. Holt, B. Wright, D. F. Gruber, Orca vowels and consonants: convergent spectral structures across cetacean and human speech. *bioRxiv* (2026), doi:10.64898/2026.02.27.708287, <https://www.biorxiv.org/content/early/2026/03/02/2026.02.27.708287>.
- [12] W. T. Fitch, The Biology and Evolution of Speech: A Comparative Analysis. *Annual Review of Linguistics* **4**, 255–279 (2018), doi:10.1146/annurev-linguistics-011817-045748.