CNRS French–Danish Workshop
Reasoning in the Embedding Space
Copenhagen, Denmark

Empiricism vs. Formalism

In the Study of Embeddings

Juan Luis Gastaldi

www.giannigastaldi.com

ETH zürich

January 20, 2026

Outline

Empiricist Turn in Computer Science

Chomsky's Trap

Language Models as Formal Objects

Philosophical Consequences

Takeaways

Outline

Empiricist Turn in Computer Science

Chomsky's Trap

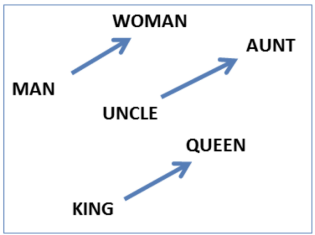
Language Models as Formal Objects

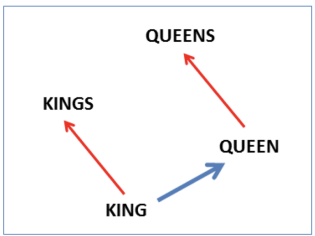
Philosophical Consequences

Takeaways

Teen Room Posters

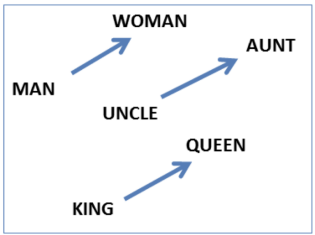
Anna's

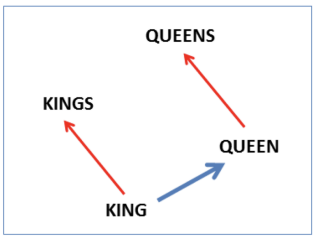




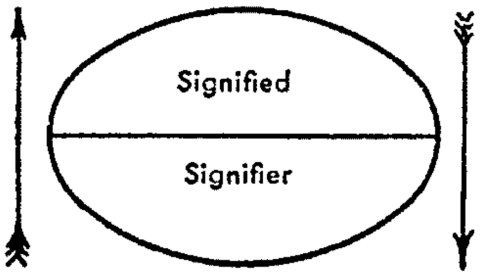
Teen Room Posters

Anna's



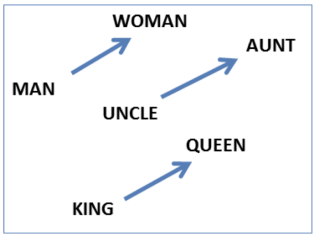


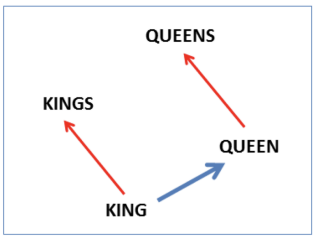
Gianni's



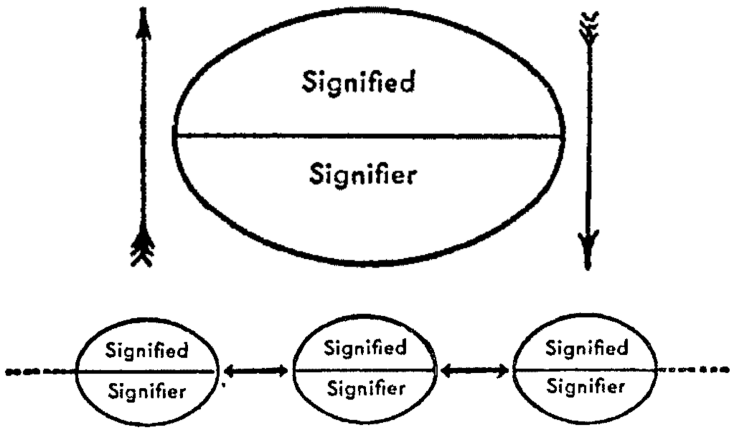
Teen Room Posters

Anna's



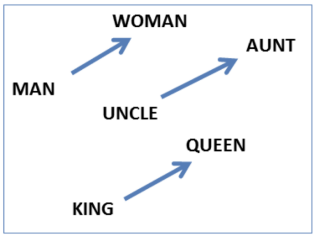


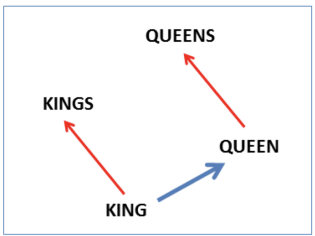
Gianni's



Teen Room Posters

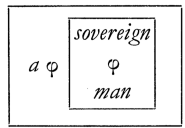
Anna's

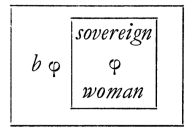


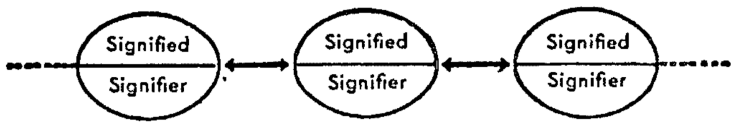


Gianni's

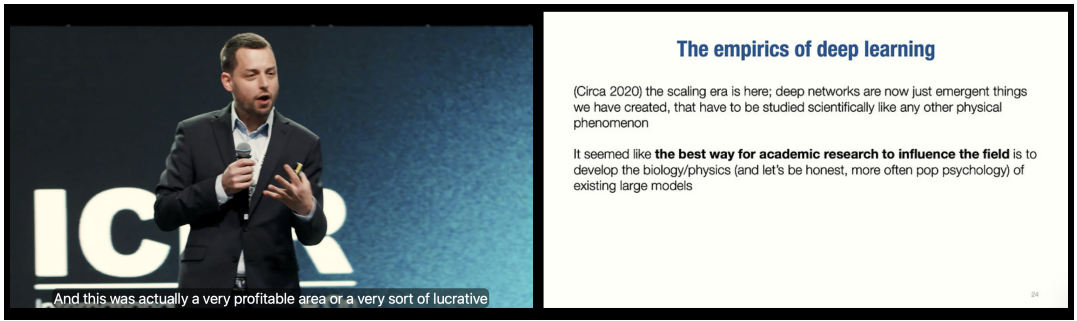
We then have a function between *a* as one terminal and the functional field *sovereign* φ *man* as the other, and likewise a function between *b* as one terminal and the functional field *sovereign* φ *woman* as the other. Diagrammatically:





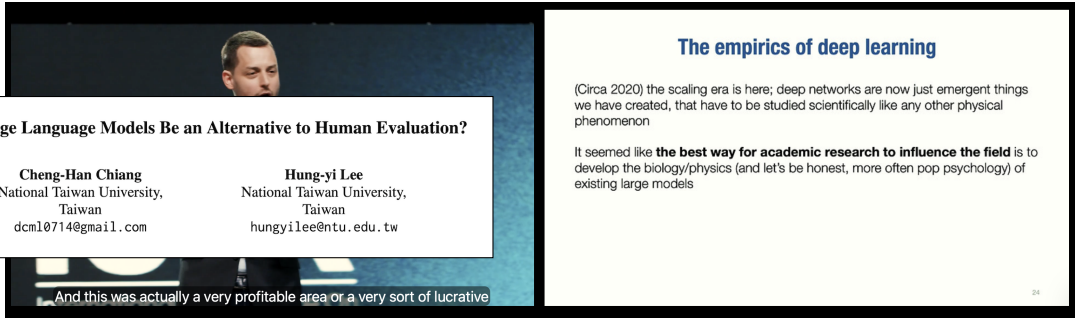


Empirical Saturnalia



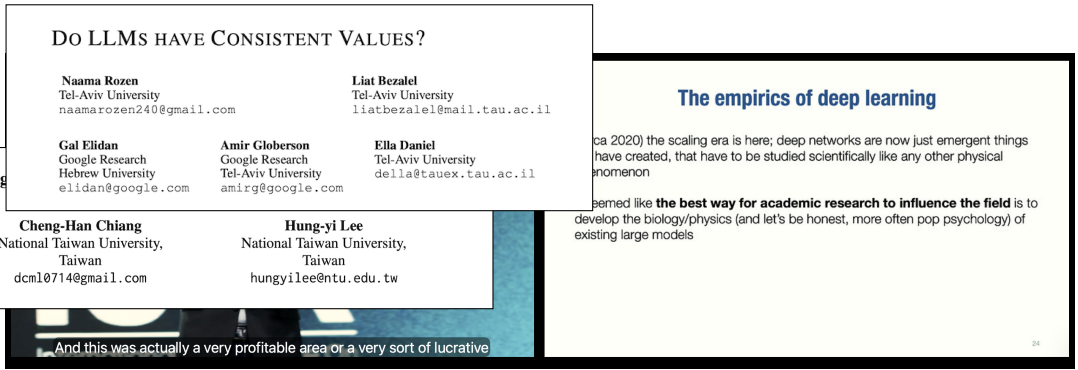
Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

Empirical Saturnalia



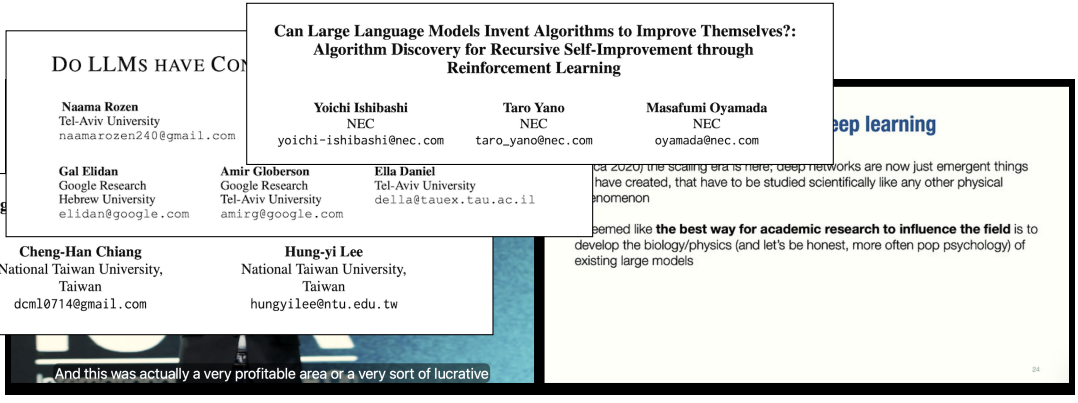
Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

Empirical Saturnalia



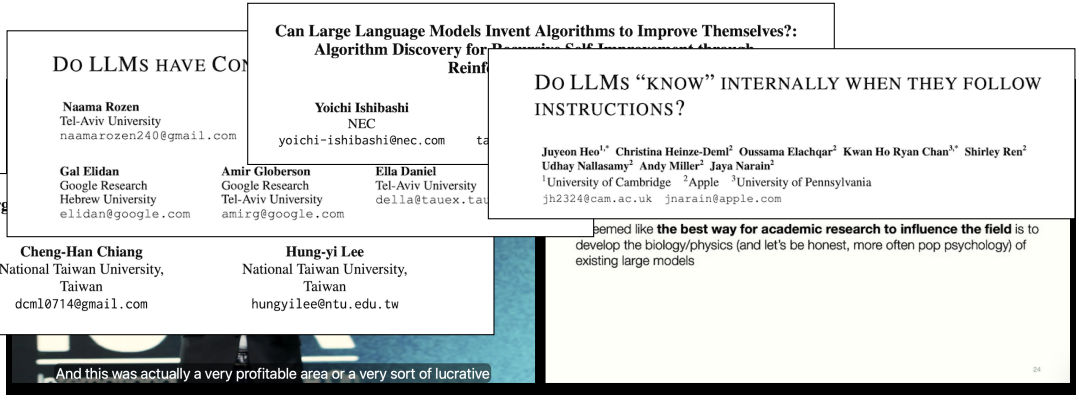
Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

Empirical Saturnalia



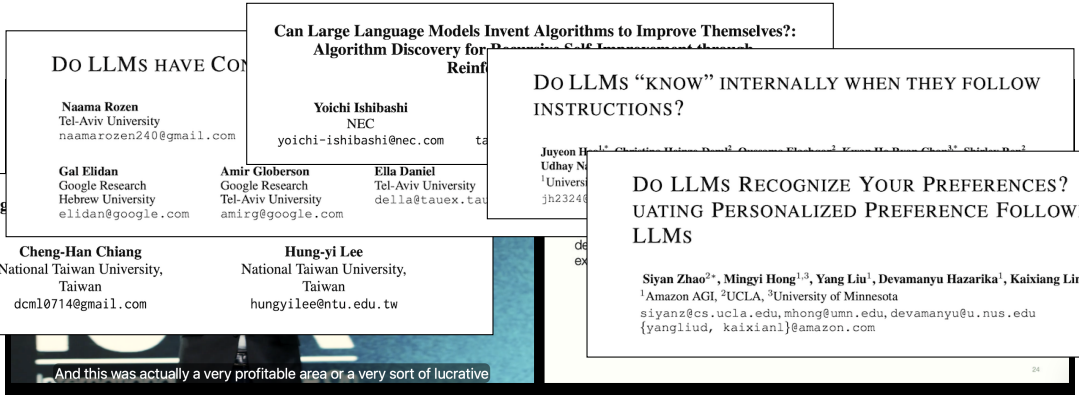
Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

Empirical Saturnalia



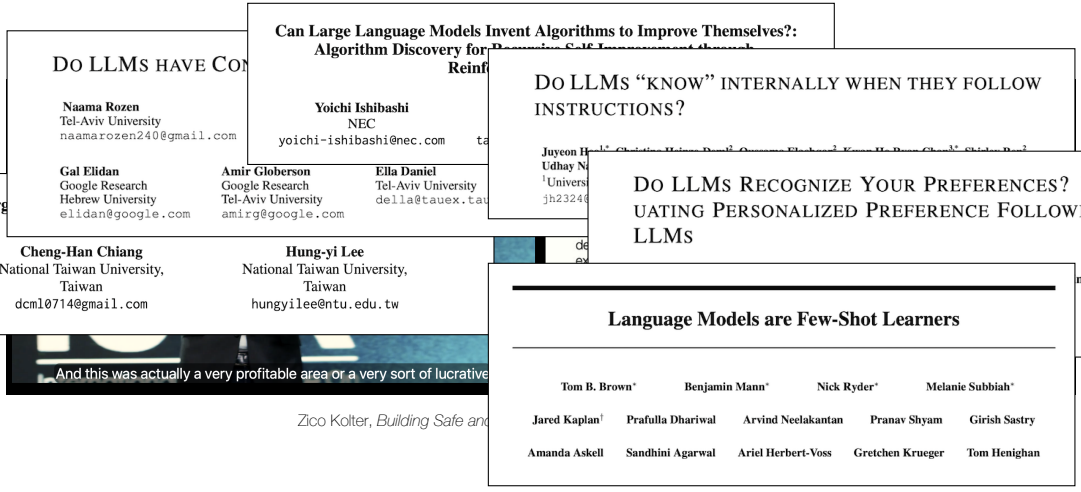
Zico Kolter, Building Safe and Robust AI Systems, Keynote at ICLR 2025.

Empirical Saturnalia



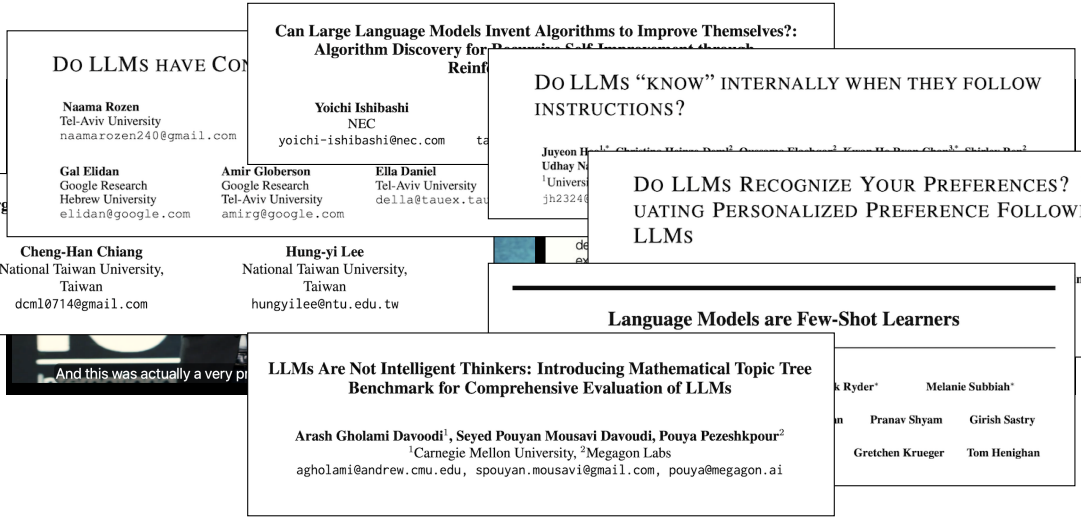
Zico Kolter, Building Safe and Robust AI Systems, Keynote at ICLR 2025.

Empirical Saturnalia

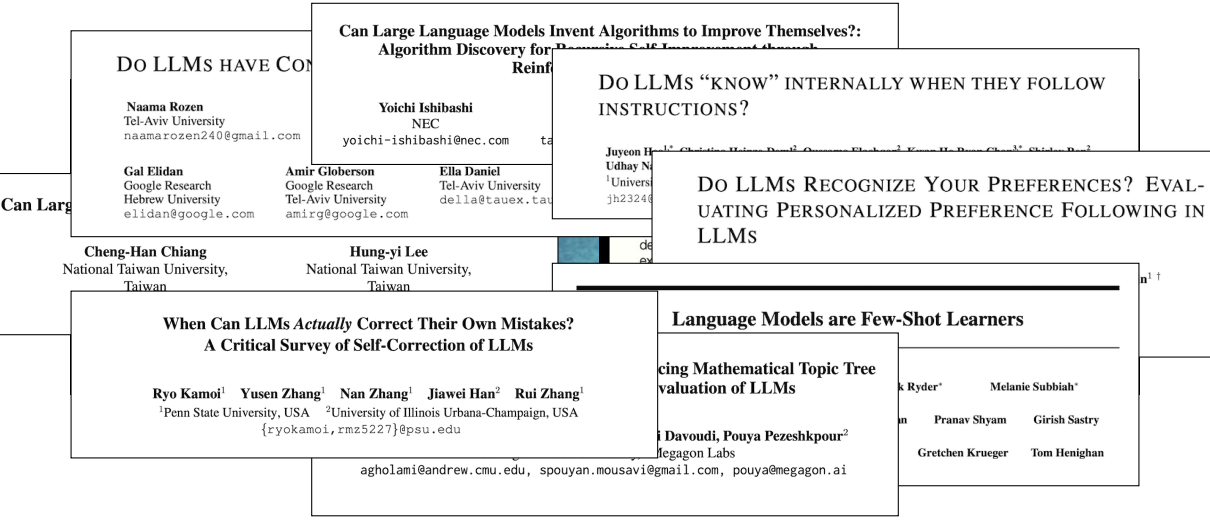


Zico Kolter, *Building Safe and*

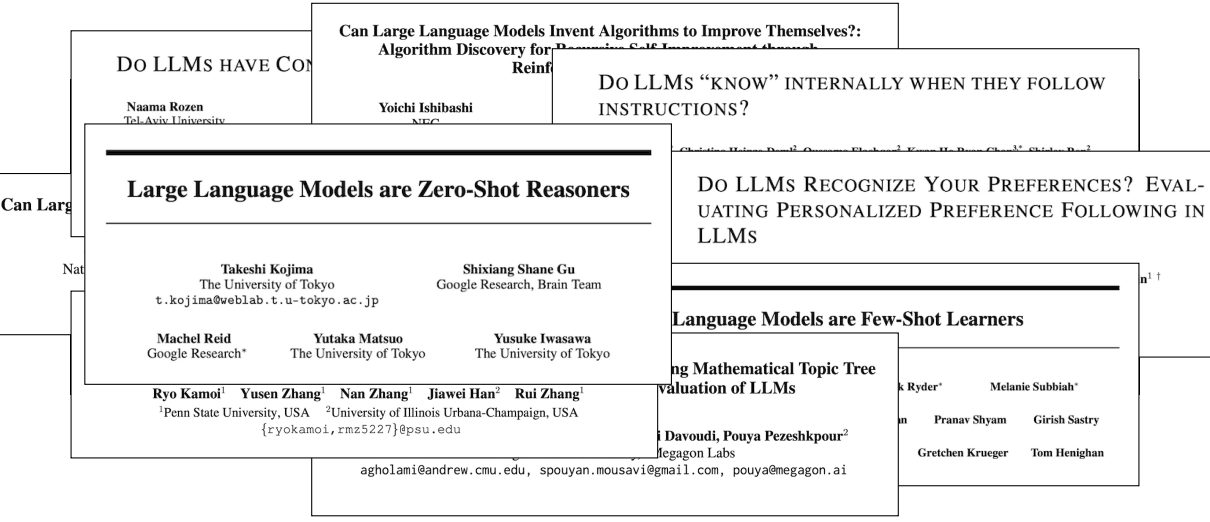
Empirical Saturnalia



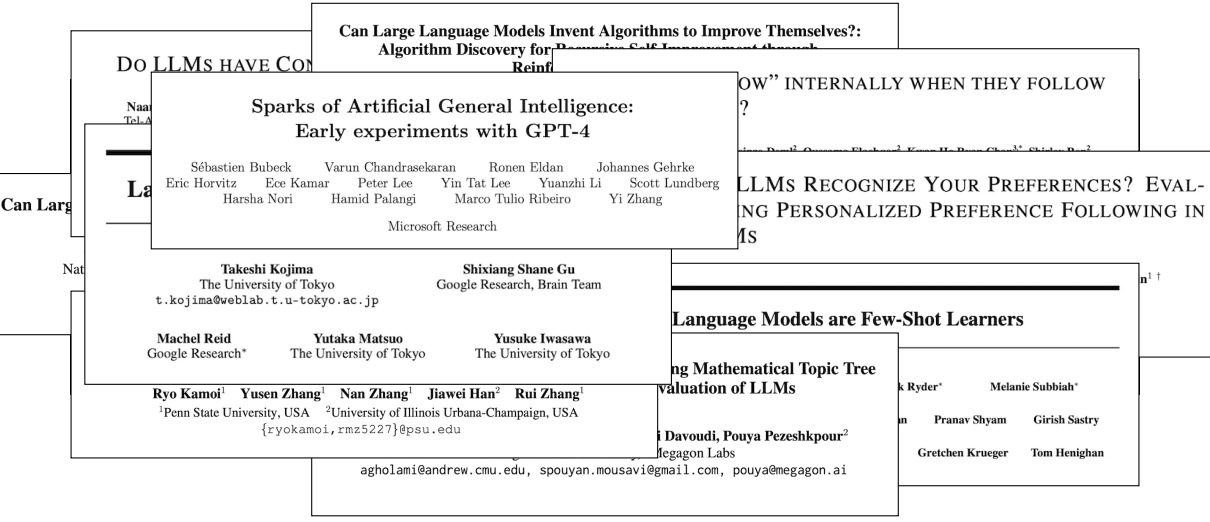
Empirical Saturnalia



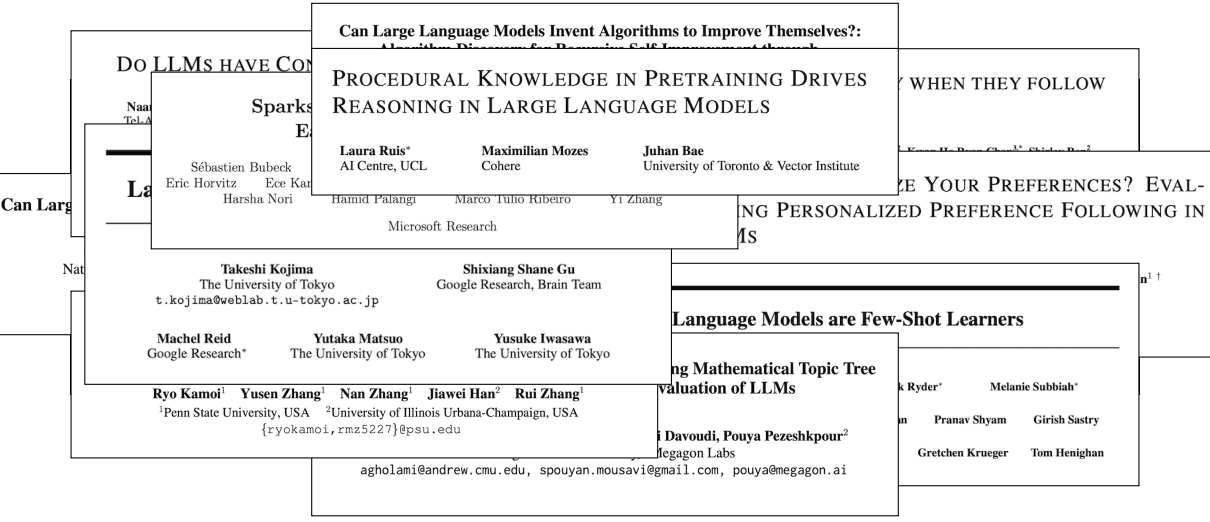
Empirical Saturnalia



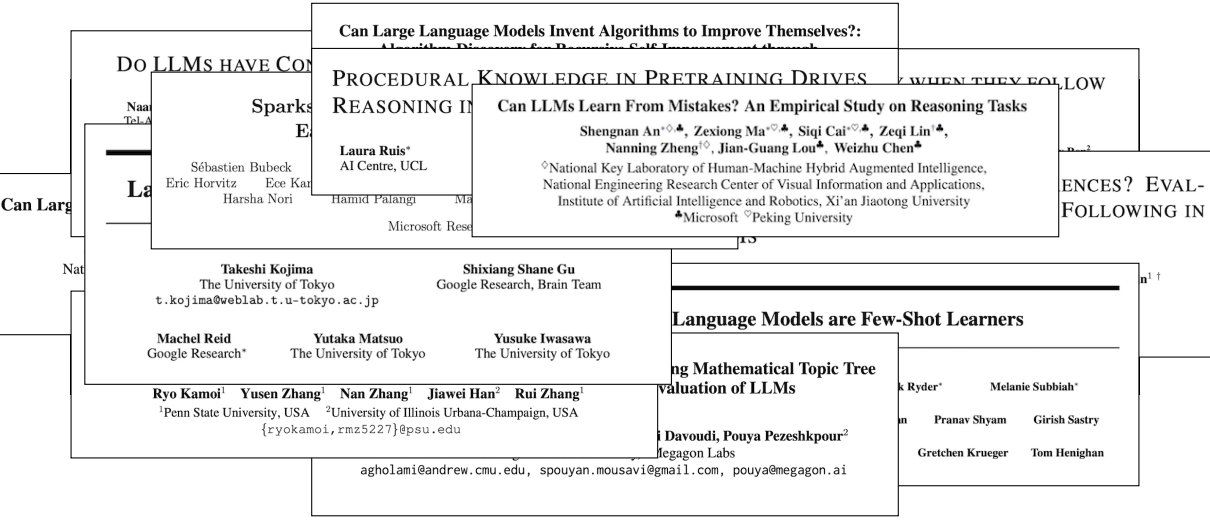
Empirical Saturnalia



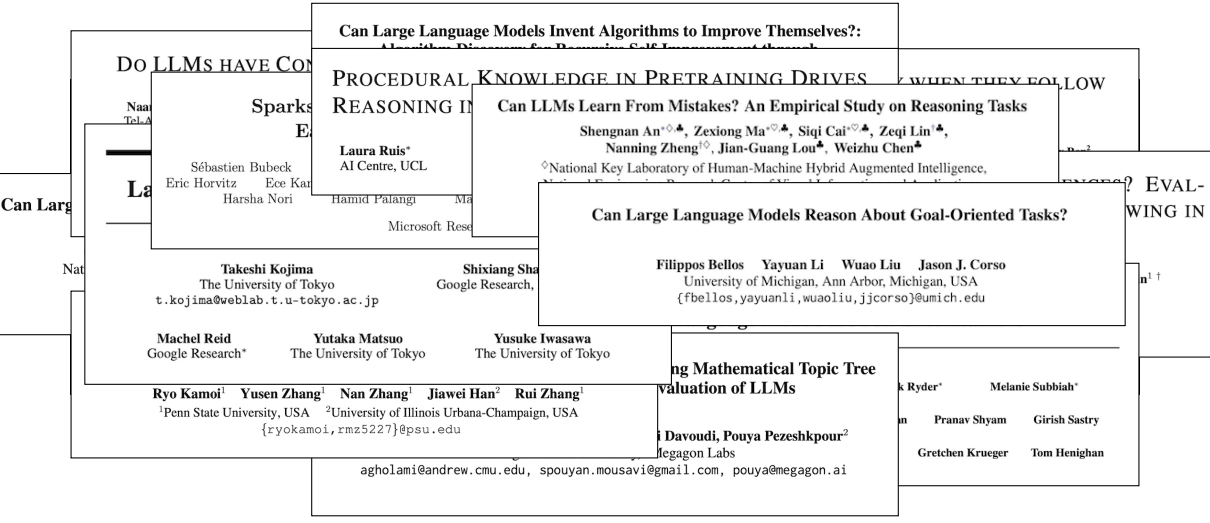
Empirical Saturnalia



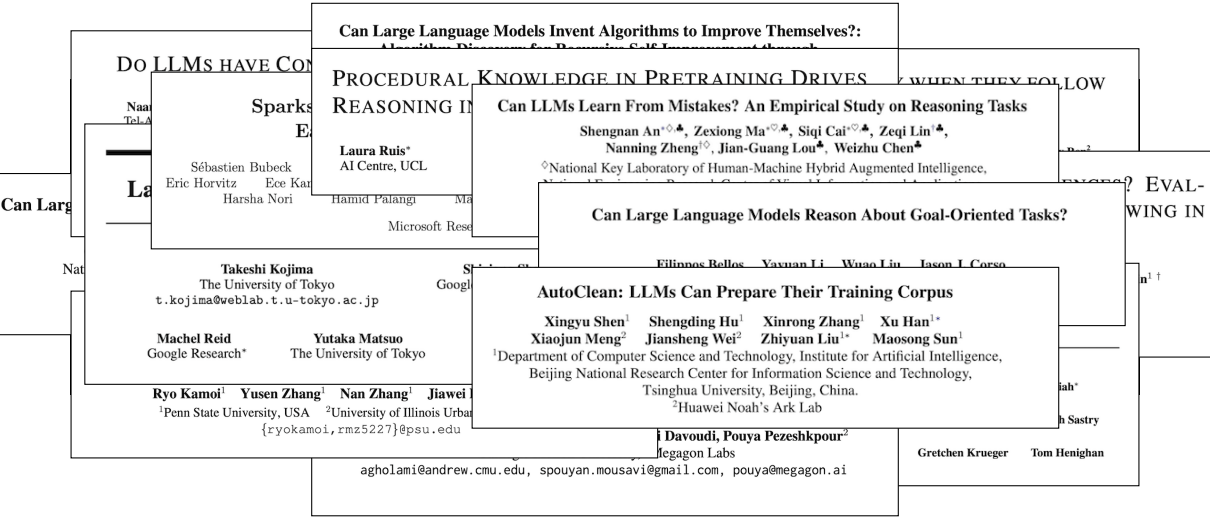
Empirical Saturnalia



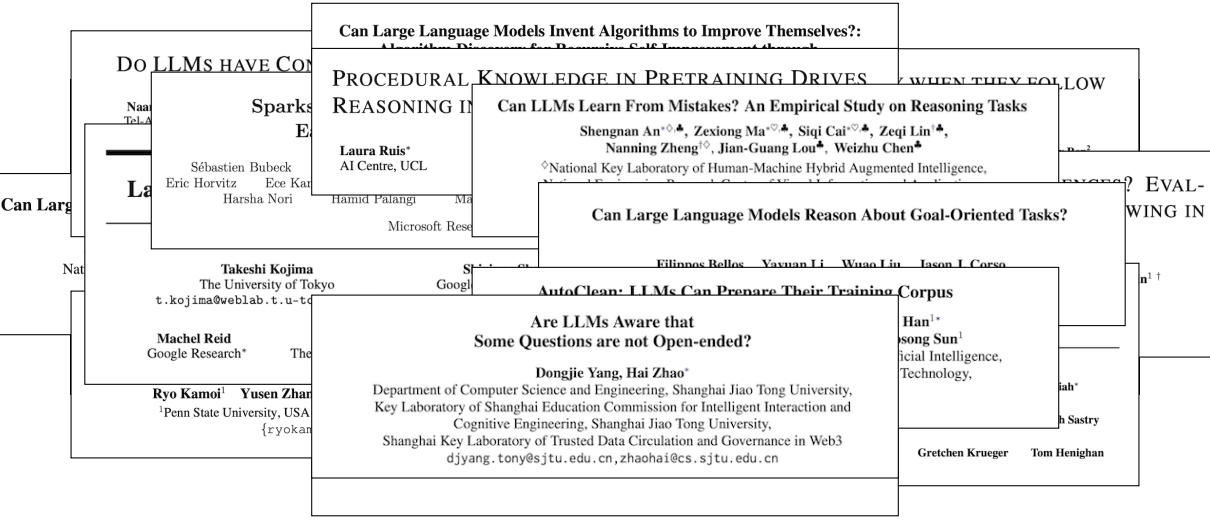
Empirical Saturnalia



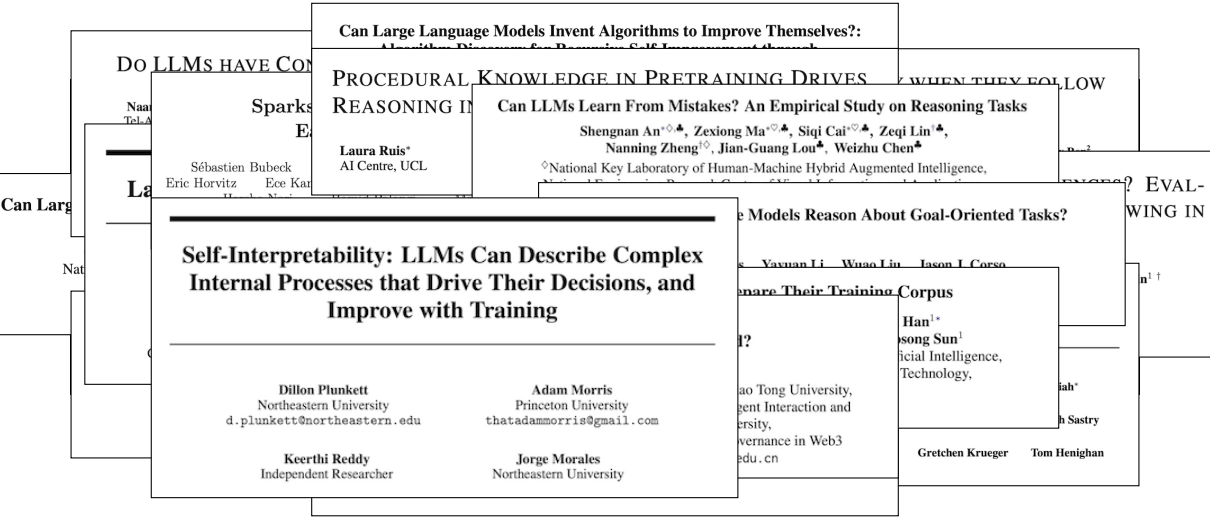
Empirical Saturnalia



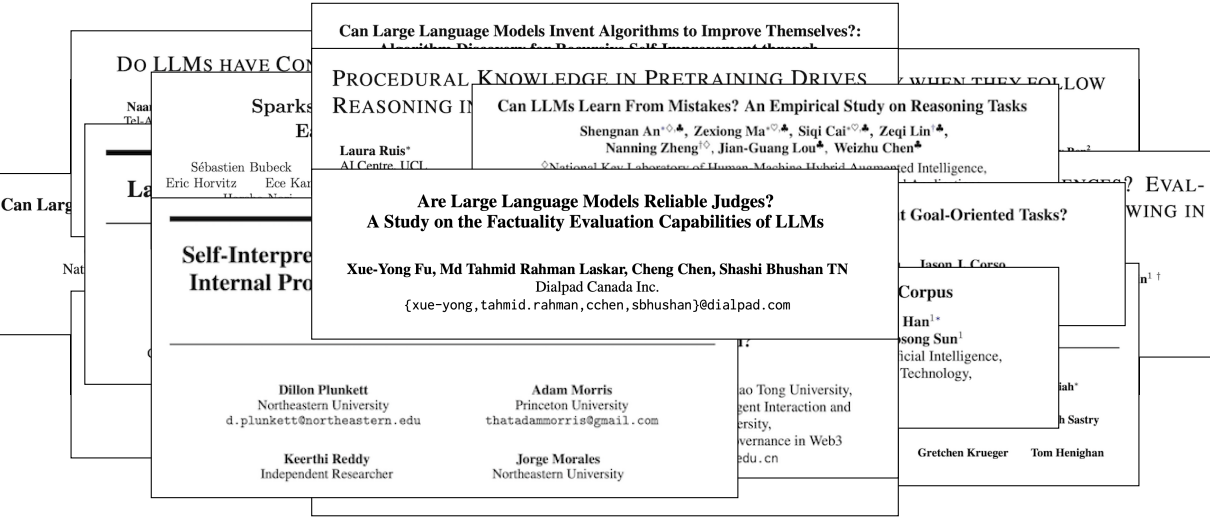
Empirical Saturnalia



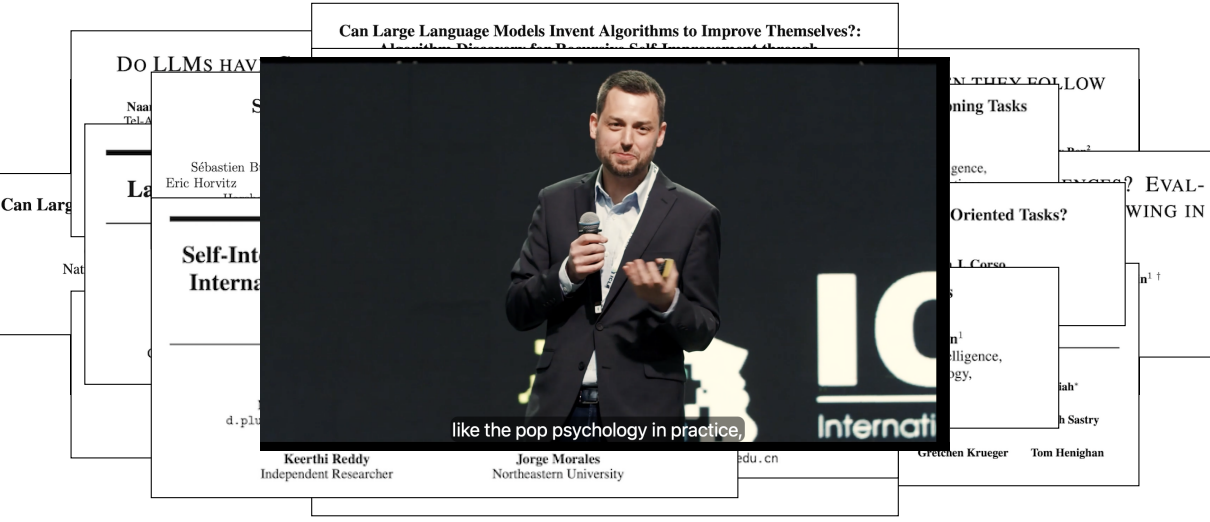
Empirical Saturnalia



Empirical Saturnalia



Empirical Saturnalia



Interpretability as a Natural Science

The Structure of Scientific Revolutions by Thomas Kuhn [42] is a classic text on the history and sociology of science. In it, Kuhn distinguishes between "normal science" in which a scientific community has a paradigm, and "extraordinary science" in which a community lacks a paradigm, either because it never had one or because it was weakened by crisis. It's worth noting that "extraordinary science" is not a desirable state: it's a period where researchers struggle to be productive.

Kuhn's description of pre-paradigmatic fields feel eerily reminiscent of interpretability today [9] There isn't consensus on what the objects of study are, what methods we should use to answer them, or how to evaluate research results. To quote a recent interview with Ian Goodfellow: "For interpretability, I don't think we even have the right definitions." [43]

One particularly challenging aspect of being in a pre-paradigmatic field is that there isn't a shared sense of how to evaluate work in interpretability. There are two common proposals for dealing with this, drawing on the standards of adjacent fields. Some researchers, especially those with a deep learning background, want an "interpretability benchmark" which can evaluate how effective an interpretability method is. Other researchers with an HCI background may wish to evaluate interpretability methods through user studies.

But interpretability could also borrow from a third paradigm: natural science. In this view, neural networks are an object of empirical investigation, perhaps similar to an organism in biology. Such work would try to make empirical claims about a given network, which could be held to the standard of falsifiability.

(Olah et al., 2020)

The Empiricization of Computer Science

MANOEL HORTA RIBEIRO
DEC 17, 2025

15 6 3 Share

Yet the nature of computer science has changed tremendously: it is now an empirical science, driven by observation and experiment as much as by theory and construction. If you don't believe me, spend 5 minutes going over the best paper awards for three conferences in different subfields in CS. If you did so in 2025, you might have found papers like "Characterizing and Detecting Propaganda-Spreading Accounts on Telegram" (USENIX; Security and Privacy), or "Examining Mental Health Conversations with Large Language Models through Reddit Analysis" (CSCW; Human Computer Interaction), or "Scaling Depth Can Enable New Goal-Reaching Capabilities" (NeuRIPS; Machine Learning). The first two papers describe online user traces, to understand sociotechnical phenomena (Online Propaganda, people using LLMs to talk about mental health issues), while the latter is a full-blown examination of what happens when you do self-supervised RL with very deep neural networks.

This empiricization of Computer Science seems to have happened independently across subdisciplines, and I argue that the reasons are two-fold.

(Horta Ribeiro, 2025)

Outline

Empiricist Turn in Computer Science

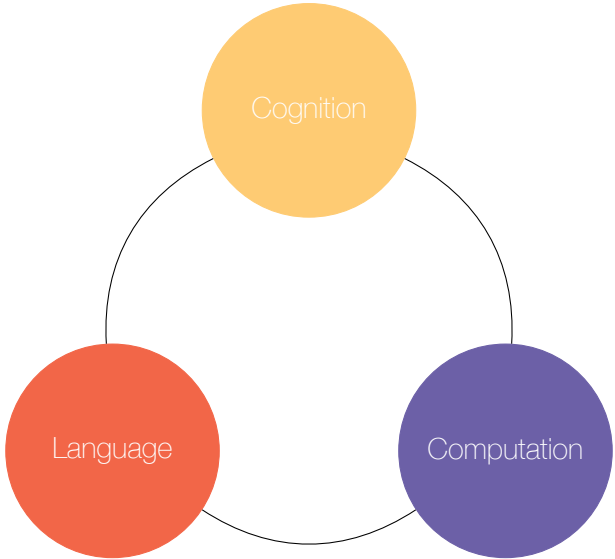
Chomsky's Trap

Language Models as Formal Objects

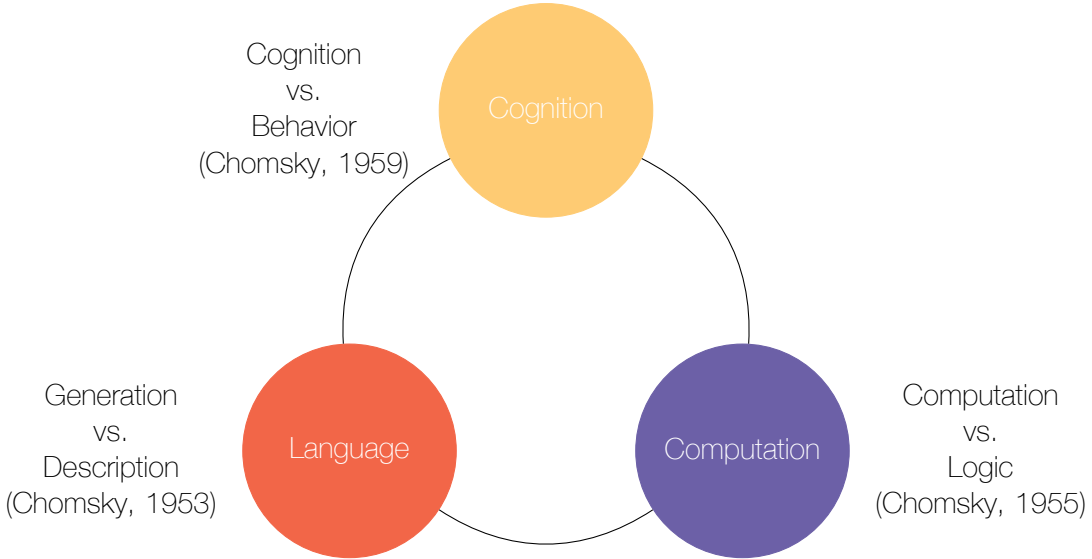
Philosophical Consequences

Takeaways

Chomsky's Generativist Program and the Cognitive Revolution



Chomsky's Generativist Program and the Cognitive Revolution



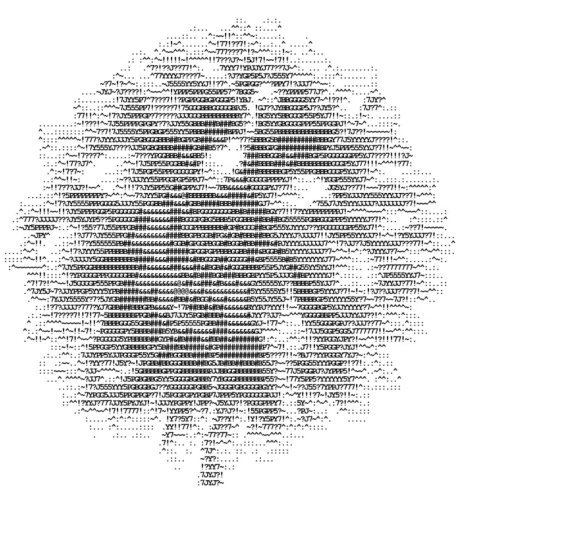
The New York Times

OPINION

GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023



Chomsky against Abstraction in Principle

"Pick the properties that you like for a set of processors. Pick the criteria you like for success, whether in terms of performance or structure or whatever. Consider the class of all organisms, *abstracting in principle* from the existing world, that satisfy those things. And then you can ask whether they have some property of things in the material world. Do they breathe? Do they grow? Do they think? Do they talk? Do they walk? Do they enjoy themselves? Do they have moral rights?"

(Chomsky, 1992)

Chomsky against Abstraction in Principle

"All of these questions are stupid. And the reason they're stupid is because you've departed from naturalism. Once you've departed from naturalism, you have an algorithm for constructing stupid questions."

(Chomsky, 1992)

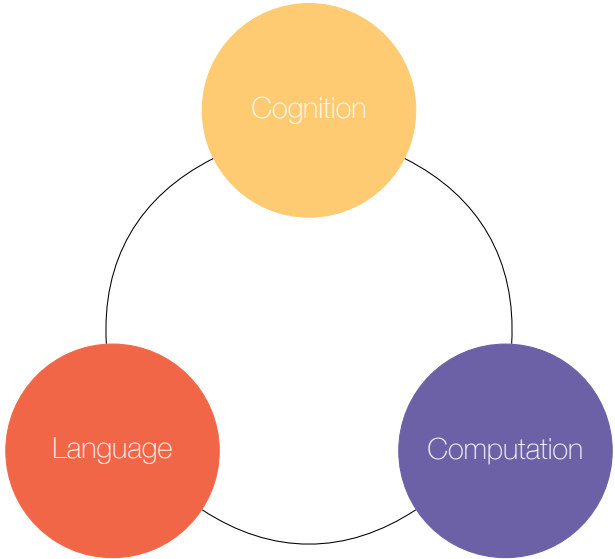
Chomsky against Abstraction in Principle

"There's nothing wrong with principled abstraction. In fact, one might think of large areas of mathematics as that. But here we have something new, principled abstraction in an empirical discipline."

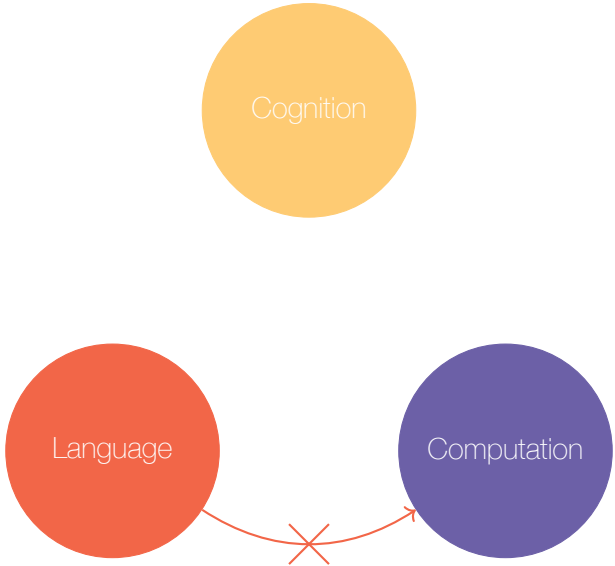
"I don't think we should cross that border, because there's no empirical claim. It is just a question of how to extend the metaphor."

(Chomsky, 1992)

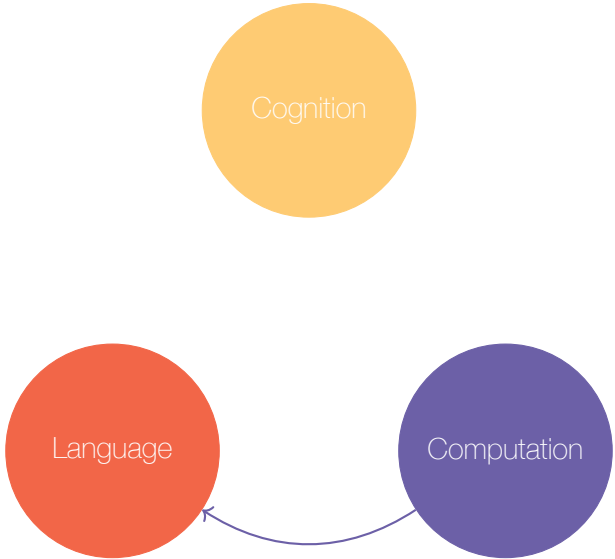
The Trap



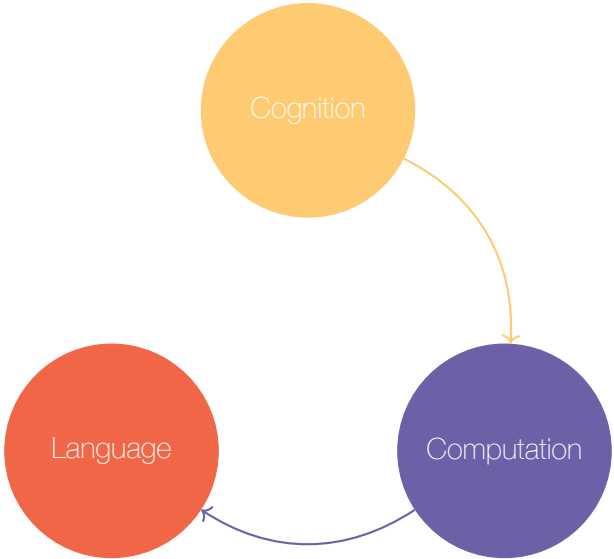
The Trap



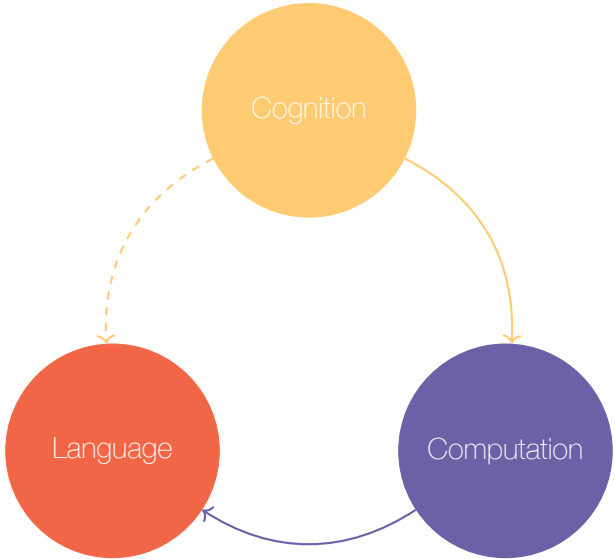
The Trap



The Trap

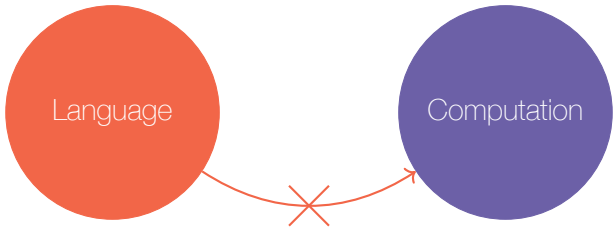


The Trap



Necessary Condition?

- Inadequacy of distributional models (Chomsky, 1953)
- Limited expressive power of FSAs (Chomsky, 1956)
- The probability of a sentence is useless (Chomsky, 1957, 1959)
- Poverty of stimulus (Chomsky, 1959)



Necessary Condition?

- Inadequacy of distributional models (Chomsky, 1953)

  Inconclusive

- The probability of a sentence is useless (Chomsky, 1957, 1959)

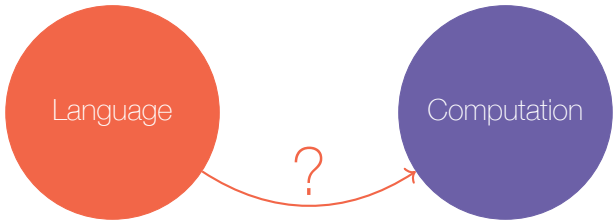
  Empirically challenged

- Limited expressive power of FSAs (Chomsky, 1956)

  The relevance is unclear

- Poverty of stimulus (Chomsky, 1959)

  Assumes what is to be proved



Outline

Empiricist Turn in Computer Science

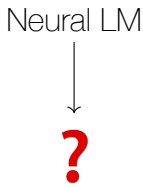
Chomsky's Trap

Language Models as Formal Objects

Philosophical Consequences

Takeaways

LLMs are computable functions



LLMs are computable functions

Neural LM

LLMs are computable functions

Neural LM

LLMs are computable functions

Neural LM

LLMs are computable functions

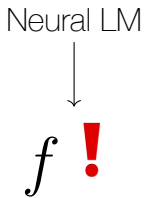
Neural LM

LLMs are computable functions

Neural LM

?

LLMs are computable functions



LLMs are computable functions

Neural LM

Function

$$f\colon \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

LLMs are computable functions

Neural LM

Computable

Function

$$f\colon \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

LLMs are computable functions

Neural LM

↓

Computable

Function

$$f\colon \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

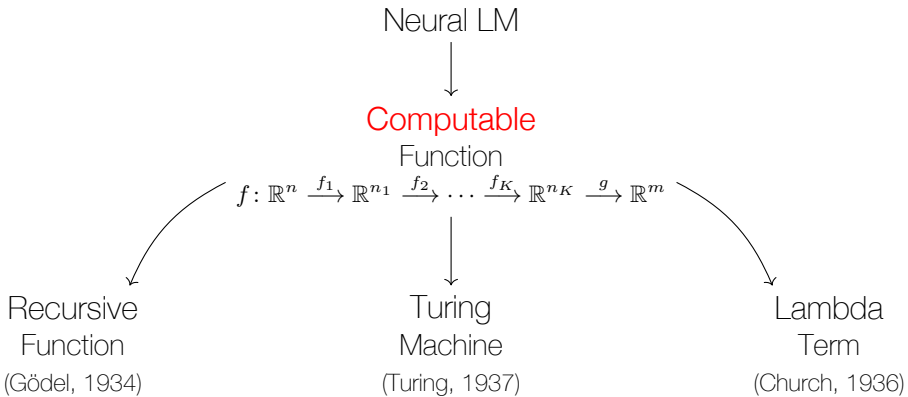
↓

Turing

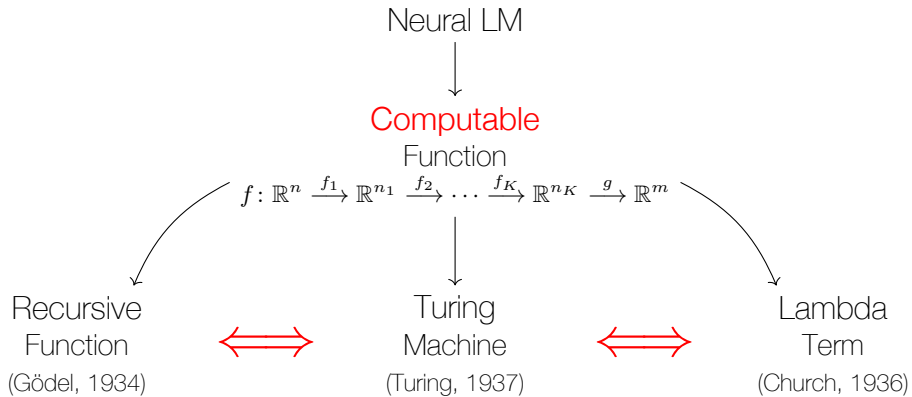
Machine

(Turing, 1937)

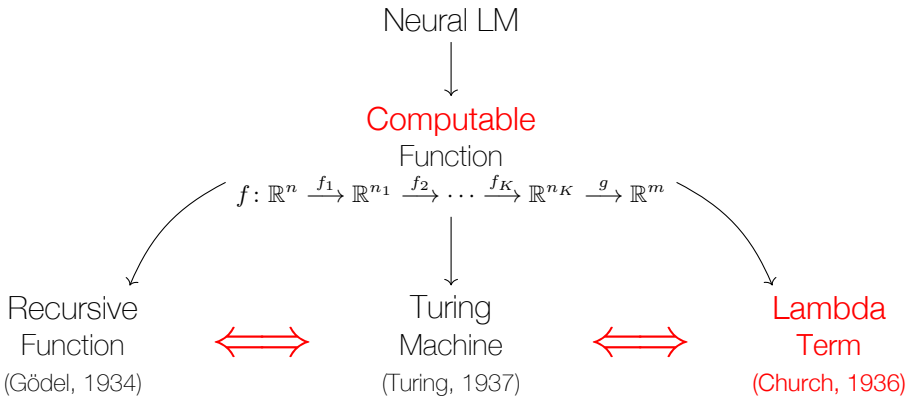
LLMs are computable functions



LLMs are computable functions



LLMs are computable functions



LLMs are computable functions

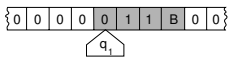
Neural LM

↓

Computable Function

$f : \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$

Recursive Function (Gödel, 1934) ⟺ Turing Machine (Turing, 1937) ⟺ Lambda Term (Church, 1936)



credit: Nynexman4464

$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$

$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$

Empirical evaluation

$$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$$

0: $\lambda f.\lambda x.x$

1: $\lambda f.\lambda x.fx$

2: $\lambda f.\lambda x.f(fx)$

3: $\lambda f.\lambda x.f(f(fx))$

4: $\lambda f.\lambda x.f(f(f(fx)))$

5: $\lambda f.\lambda x.f(f(f(f(fx))))$

$\ldots$

n: $\lambda f.\lambda x.\underbrace{f(\ldots(f}_{n \text{ times}}x)\ldots)$

Empirical evaluation

$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$

0: $\lambda f.\lambda x.x$

1: $\lambda f.\lambda x.fx$

2: $\lambda f.\lambda x.f(fx)$

3: $\lambda f.\lambda x.f(f(fx))$

4: $\lambda f.\lambda x.f(f(f(fx)))$

5: $\lambda f.\lambda x.f(f(f(f(fx))))$

$\ldots$

n: $\lambda f.\lambda x.\underbrace{f(\ldots(f}_{n \text{ times}}x)\ldots)$

$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$

Empirical evaluation

$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$

0: $\lambda f.\lambda x.x$
1: $\lambda f.\lambda x.fx$
2: $\lambda f.\lambda x.f(fx)$
3: $\lambda f.\lambda x.f(f(fx))$
4: $\lambda f.\lambda x.f(f(f(fx)))$
5: $\lambda f.\lambda x.f(f(f(f(fx))))$
...
n: $\lambda f.\lambda x.\underbrace{f(\ldots(f}_{n \text{ times}}x)\ldots)$

$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$

$\lambda f.\lambda x.f(f(f(f(fx))))$

Empirical evaluation

$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$
$P' := \lambda r.\lambda s.\lambda f.\lambda x.f(f(f(f(fx))))$

0: $\lambda f.\lambda x.x$
1: $\lambda f.\lambda x.fx$
2: $\lambda f.\lambda x.f(fx)$
3: $\lambda f.\lambda x.f(f(fx))$
4: $\lambda f.\lambda x.f(f(f(fx)))$
5: $\lambda f.\lambda x.f(f(f(f(fx))))$
...
n: $\lambda f.\lambda x.\underbrace{f(\ldots(f}_{n \text{ times}}x)\ldots)$

$\lambda r.\lambda s.\lambda f.\lambda x.f(f(f(f(fx))))(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$

$\lambda f.\lambda x.f(f(f(f(fx))))$

Empirical interpretability

$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$

0: $\lambda f.\lambda x.x$

1: $\lambda f.\lambda x.fx$

2: $\lambda f.\lambda x.f(fx)$

3: $\lambda f.\lambda x.f(f(fx))$

4: $\lambda f.\lambda x.f(f(f(fx)))$

5: $\lambda f.\lambda x.f(f(f(f(fx))))$

$\ldots$

n: $\lambda f.\lambda x.\underbrace{f(\ldots(f}_{n \text{ times}}\, x)\ldots)$

$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$

$\vdots$

$\lambda f.\lambda x.f(f(f(f(fx))))$

Empirical interpretability

$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$

0: $\lambda f.\lambda x.x$

1: $\lambda f.\lambda x.fx$

2: $\lambda f.\lambda x.f(fx)$

3: $\lambda f.\lambda x.f(f(fx))$

4: $\lambda f.\lambda x.f(f(f(fx)))$

5: $\lambda f.\lambda x.f(f(f(f(fx))))$

$\ldots$

n: $\lambda f.\lambda x.\underbrace{f(\ldots(f}_{n \text{ times}}x)\ldots)$

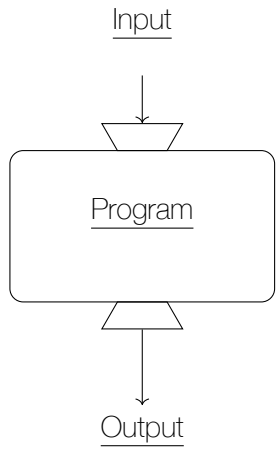
$$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$$
$$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda g.\lambda y.g(gy))(\lambda h.\lambda z.h(h(hz)))$$
$$\lambda n.\lambda f.\lambda x.(\lambda g.\lambda y.g(gy))f(nfx)(\lambda h.\lambda z.h(h(hz)))$$
$$\lambda n.\lambda f.\lambda x.(\lambda g.\lambda y.g(gy))f(nfx)(\lambda h.\lambda z.h(h(hz)))$$
$$\lambda f.\lambda x.(\lambda g.\lambda y.g(gy))f((\lambda h.\lambda z.h(h(hz)))fx)$$
$$\lambda f.\lambda x.(\lambda y.f(fy))((\lambda h.\lambda z.h(h(hz)))fx)$$
$$\lambda f.\lambda x.(\lambda y.f(fy))((\lambda z.f(f(fz)))x)$$
$$\lambda f.\lambda x.(\lambda y.f(fy))(f(f(fx)))$$
$$\lambda f.\lambda x.f(f(f(f(fx))))$$

Empirical interpretability

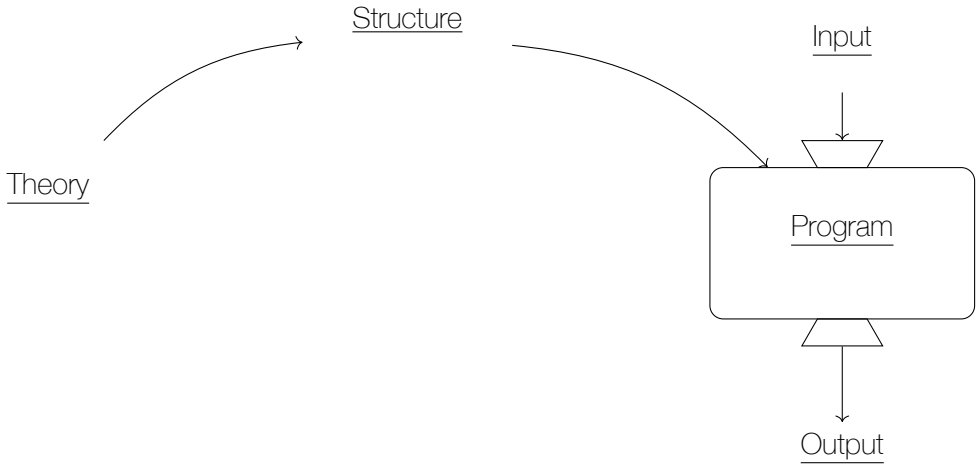
$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$

$P'' :=$ λRófÄÒêÑ5È|Àxñ=∞ù ÿmWf286ëy'SÒú>v&ìÂ ¬2 óÉ7öç∞{ã>2flB°µG#À9ÇU ∞btYBô¥ Ù ë‰3;5 å[l–èuô¨Ü ˚ 7–Ù.λ:ˆ4mÓØ¥´è—+ÌsÖ,$+gï„B™÷o–#ïŸê Ûv –gÓÿ/ëiijO‡Œfi • J1«€ø˛Ï hãêt‡æY$ˆ6 FïW»RÙKg¢¨.λ‡d¯… D2÷˛ò ˚ xêÈy. Ó”¢b Bé£NÈ1Ê‡/Û9Ñµ–/JYÇõË9ÿÀÈ.λÁÍ À^öÇ,»fq∞±îˇB5Ì>O˜g™“6Ωe“æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨.λÆà€fUòfEÙ·Í m#„˛4\r√¬÷Îpò»y*vtÄJÃF1ûÁóz«ñM”DjŒ BËèÍT _ Êa‰ÁÇΩ @\Ø^˜]Î›h‡.:ˆ4mÓØ¥´è—+ÌsÖ,$+gï„B™÷o–#ïŸê Ûv–gÓÿ /ëiijO‡Œfi • J1«€ø˛Ï hãêt‡æY$ˆ6 FïW»RÙKg¢¨ÁÍ À^öÇ,»fq∞±îˇB5Ì>O˜g™“6 Ωe“æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨(‡d¯… D2÷˛ò ˚ xêÈy. Ó”¢bBé£NÈ1Ê‡/Û9Ñµ–/ JYÇõË9ÿÀÈÁÍ À^öÇ,»fq∞±îˇB5Ì>O˜g™“6Ωe“æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨Æ à€fUòfEÙ·Í m#„˛4\r√¬÷Îpò»y*vtÄJÃF1ûÁóz«ñM”DjŒBËèÍT _ Êa‰ÁÇΩ @\ Ø^˜]Î›h‡)(ËÍÛeí4WµÍ }w„$Ω“K5èÃ¶%3[m ’˜BÃfifÒ; óJçCËîõŸØcB,ñ$Åå}ÓÅØ3¡ ˇ?ö¯oŒ@ fl8ˇR CÆoˇ* &<Ÿ ¬o12À‰ àÓÜ#ï ”,ú¨ «ô˛„∞Îâä“øÃd|ˆÑ’Ê yØ;ˆW ›″wò[}\»ÖÈúw’6<ùˇ=ãÕ ¯Ì‘Dz?2±|é‘ 3Â/rxµ∞µ$ÀëÂ*lf˜îû’+ÍVìy^{a}Gæßägô/,uÑ)

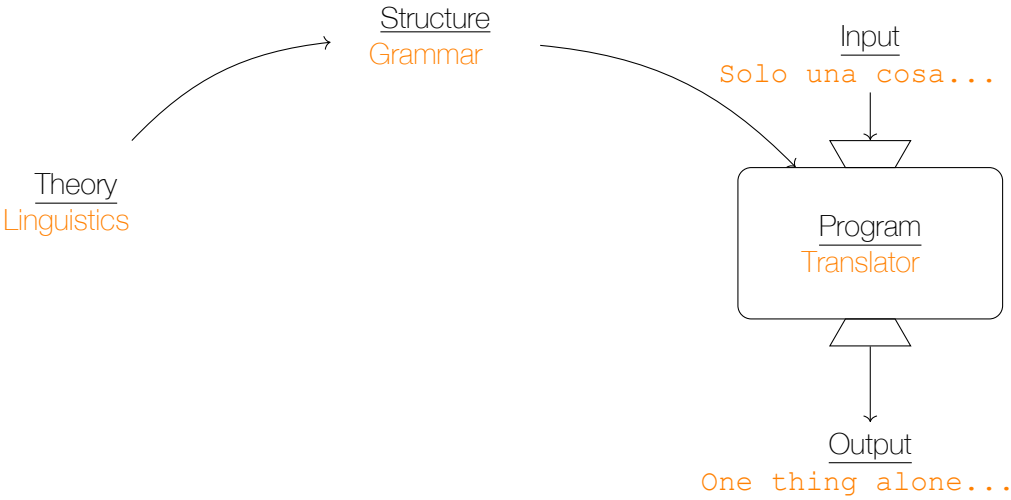
Implicit Structure



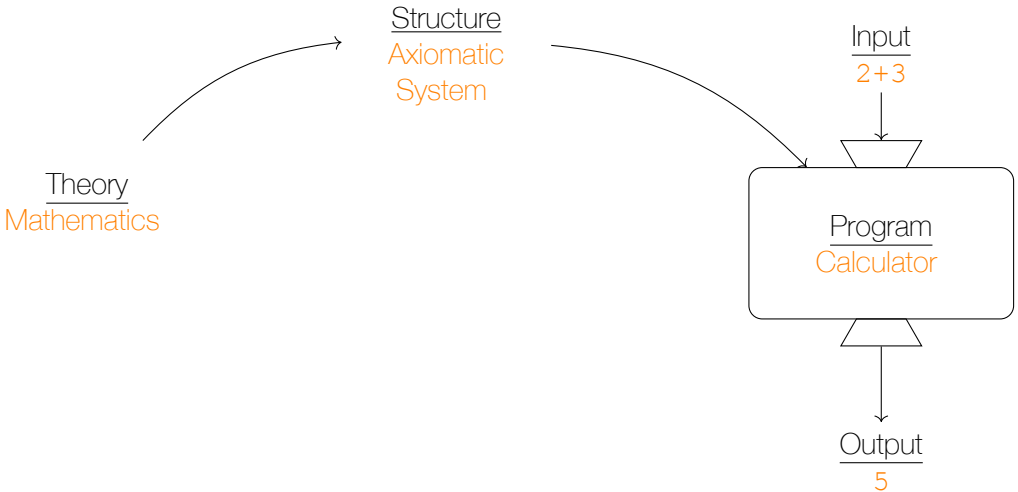
Implicit Structure



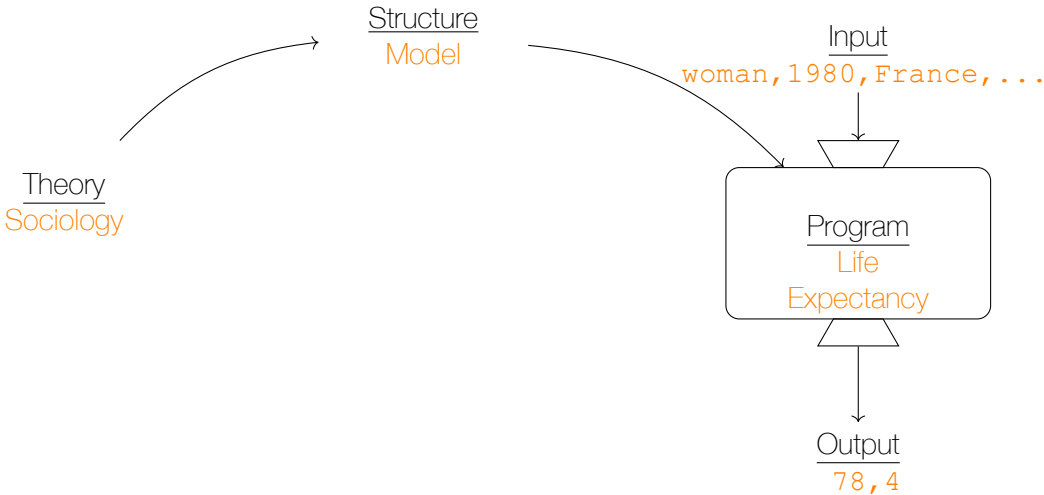
Implicit Structure



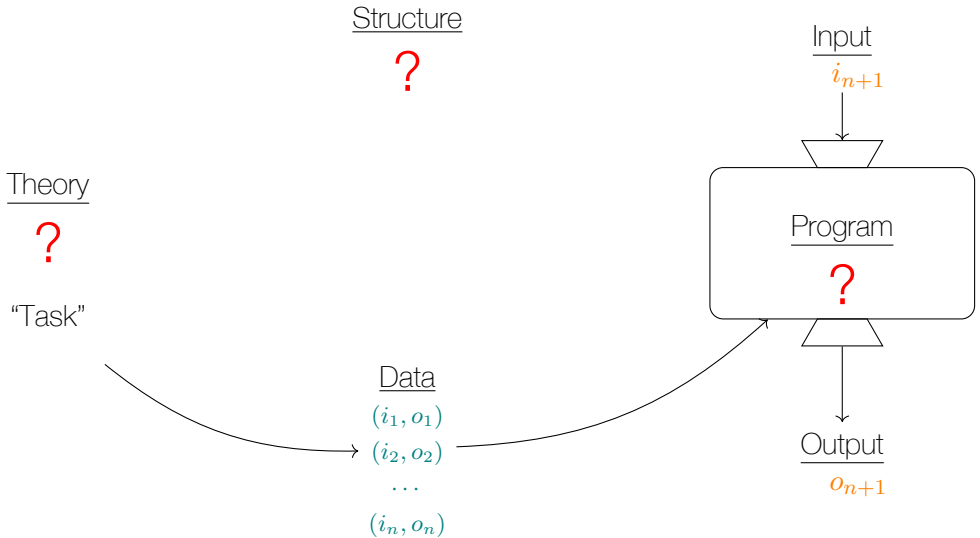
Implicit Structure



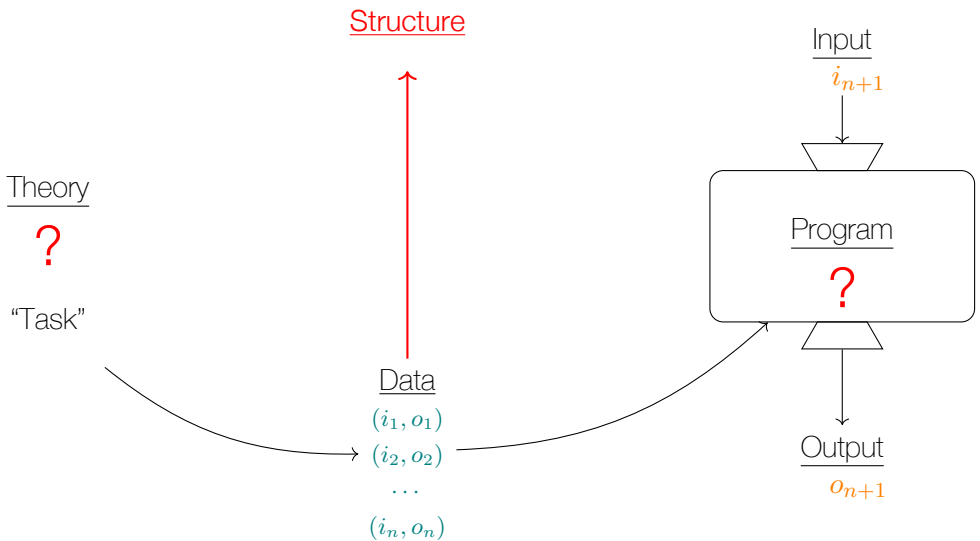
Implicit Structure



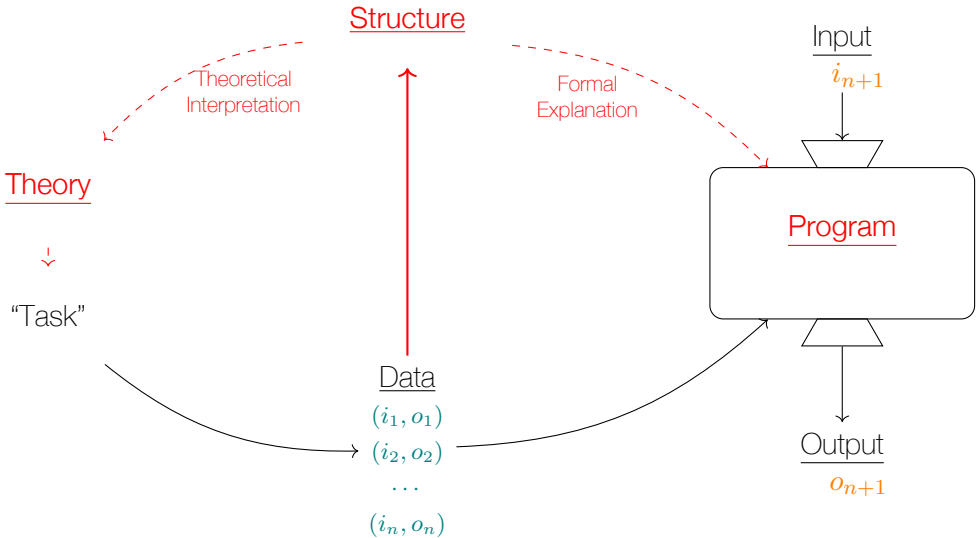
Implicit Structure



Making It Explicit



Making It Explicit

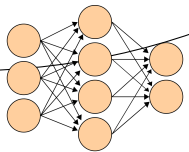


Embedding structure

Structure

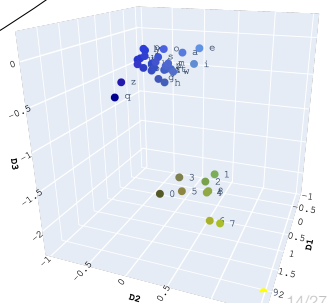
?

Data



{-, /, 0, 1, 2, . . . , 8, 9, =,
a, b, c, . . . , w, x, y, z, é}

Embedding



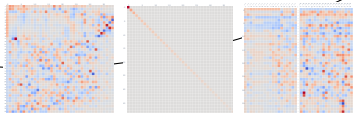
Embedding structure

Structure

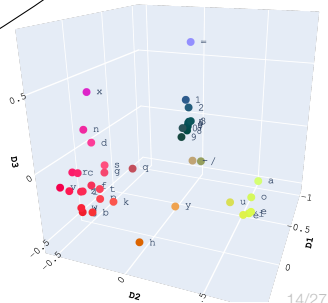
{-, /, 0, 1, 2, . . . , 8, 9, =,
a, b, c, . . . , w, x, y, z, é}

Embedding

Data



SVD



Structure

?

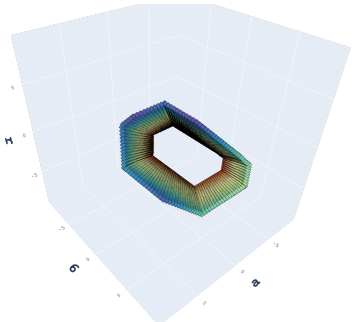
$$\mathsf{C}^{\mathrm{op}} \times \mathsf{D} \longrightarrow \bar{\mathbb{R}}$$

$$\Downarrow$$

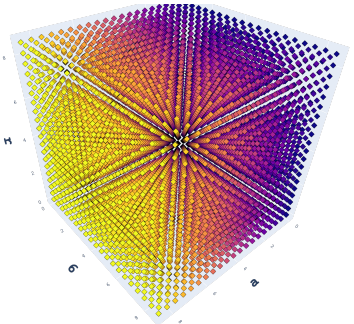
$$\mathcal{M}^* \colon \bar{\mathbb{R}}^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows \left(\bar{\mathbb{R}}^{\mathsf{D}}\right)^{\mathrm{op}} \colon \mathcal{M}_*$$

Structure

?



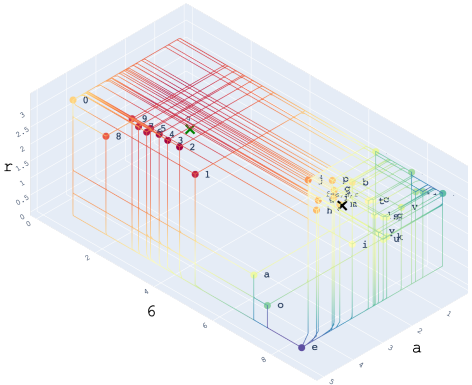
$\mathcal{M}_*\mathcal{M}^*$



$$\mathsf{C}^{\mathrm{op}} \times \mathsf{D} \to \bar{\mathbb{R}}$$

$$\mathcal{M}^* \colon \bar{\mathbb{R}}^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows (\bar{\mathbb{R}}^{\mathsf{D}})^{\mathrm{op}} \colon \mathcal{M}_*$$

Structure

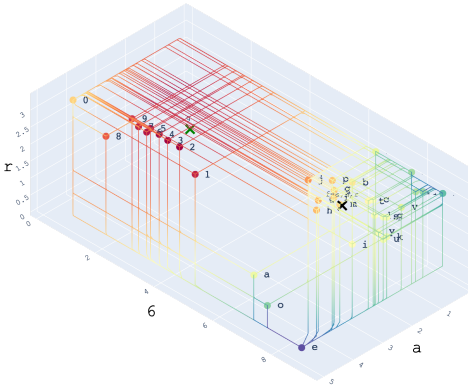


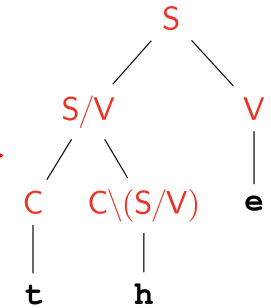
$$\mathsf{C}^{\mathrm{op}} \times \mathsf{D} \longrightarrow \bar{\mathbb{R}}$$

$$\Downarrow$$

$$\mathcal{M}^* : \bar{\mathbb{R}}^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows \left(\bar{\mathbb{R}}^{\mathsf{D}}\right)^{\mathrm{op}} : \mathcal{M}_*$$

Structure





$$\mathsf{C}^{\mathrm{op}} \times \mathsf{D} \longrightarrow \bar{\mathbb{R}}$$

$$\Downarrow$$

$$\mathcal{M}^* : \bar{\mathbb{R}}^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows \left(\bar{\mathbb{R}}^{\mathsf{D}}\right)^{\mathrm{op}} : \mathcal{M}_*$$

Outline

Stochastic parrots vs. AGI

LLMs are not like us,
therefore they do not and can not have any relation to natural language.

LLMs have a relation to natural language,
therefore they are like us.

...I supposed that all the objects (presentations) that had ever entered into my mind when awake, had in them no more truth than the illusions of my dreams. But immediately upon this I observed that, whilst I thus wished to think that all was false, it was absolutely necessary that I, who thus thought, should be something; And as I observed that this truth, I think, therefore I am, was so certain and of such evidence that no ground of doubt, however extravagant, could be alleged by the Sceptics capable of shaking it, I concluded that I might, without scruple, accept it as the first principle of the philosophy of which I was in search.

Descartes, *Meditations on First Philosophy* (1641)

But I was persuaded that there was nothing in all the world, that there was no heaven, no earth, that there were no minds, nor any bodies: was I not then likewise persuaded that I did not exist? Not at all; of a surety I myself did exist since I persuaded myself of something [or merely because I thought of something]. But there is some deceiver or other, very powerful and very cunning, who ever employs his ingenuity in deceiving me. Then without doubt I exist also if he deceives me, and let him deceive me as much as he will, he can never cause me to be nothing so long as I think that I am something. So that after having reflected well and carefully examined all things, we must come to the definite conclusion that this proposition: I am, I exist, is necessarily true each time that I pronounce it, or that I mentally conceive it.

Descartes, *Meditations on First Philosophy* (1641)

Language vs. thought

…the philosopher has to say: "When I dissect the process expressed in the proposition 'I think,' I get a whole set of bold claims that are difficult, perhaps impossible, to establish, – for instance, that I am the one who is thinking, that there must be something that is thinking in the first place, that thinking is an activity and the effect of a being who is considered the cause, that there is an 'I,' and finally, that it has already been determined what is meant by thinking, – that I know what thinking is. [...]

Nietzsche, *Beyond Good and Evil*, §16 (1886)

Language vs. thought

…Because if I had not already made up my mind what thinking is, how could I tell whether what had just happened was not perhaps 'willing' or 'feeling'? Enough: this 'I think' presupposes that I compare my present state with other states that I have seen in myself, in order to determine what it is: and because of this retrospective comparison with other types of 'knowing,' this present state has absolutely no 'immediate certainty' for me." – In place of that "immediate certainty" which may, in this case, win the faith of the people, the philosopher gets handed a whole assortment of metaphysical questions, genuinely probing intellectual questions of conscience, such as: "Where do I get the concept of thinking from? Why do I believe in causes and effects? What gives me the right to speak about an I, and, for that matter, about an I as cause, and, finally, about an I as the cause of thoughts?" [...]

Nietzsche, *Beyond Good and Evil*, §16 (1886)

Language vs. thought

Now in order to cognize ourselves, there is required in addition to the act of thought, which brings the manifold of every possible intuition to the unity of apperception, a determinate mode of intuition, whereby this manifold is given; it therefore follows that although my existence is not indeed appearance (still less mere illusion), the determination of my existence can take place only in conformity with the form of inner sense, according to the special mode in which the manifold, which I combine, is given in inner intuition. Accordingly I have no cognition of myself as I am but merely as I appear to myself

Kant, *Critique of Pure Reason* (1781)

Language vs. thought

But, isn't thinking a kind of speaking? How is it possible for thinking to be engaged in a struggle with speaking? Wouldn't that be a struggle in which thinking was at war with itself? Doesn't this spell the end to the possibility of thinking?

Frege, *Sources of Knowledge of Math. and the math. natural Sc.* (1924-25)

Language vs. thought

It is sometimes said: animals do not talk because they lack the mental abilities. And this means: "They do not think, and that is why they do not talk." But — they simply do not talk.

Wittgenstein, *Philosophical Investigations*, 1953, § 25

Language vs. thought

The perennial man in the street believes that when he speaks he freely puts together whatever elements have the meanings he intends; but he does so only by choosing members of those classes that regularly occur together, and in the order in which these classes occur. [...] the restricted distribution of classes persists for all their occurrences; the restrictions are not disregarded arbitrarily, e.g. for semantic needs.

Harris, *Distributional Structure*, pp. 775-776, (1954).

Formal Content

(Gastaldi and Pellissier, 2021)

Form ~~vs.~~ and ~~Meaning~~ Content

Kant, Hegel, Frege, Russian formalists, Saussure, Hjelmslev, etc.

Formal Content: The dimension of content which finds its source in the internal relations holding between the expressions of a language.

Formal Content

(Gastaldi and Pellissier, 2021)

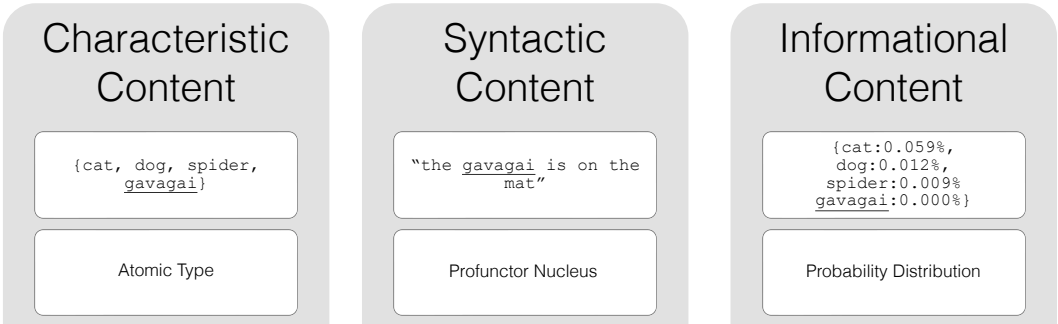
Form ~~vs.~~ and ~~Meaning~~ Content

Kant, Hegel, Frege, Russian formalists, Saussure, Hjelmslev, etc.

Formal Content: The dimension of content which finds its source in the internal relations holding between the expressions of a language.

- *Characteristic Content*: The content resulting from the inclusion of a unit in a class of other units by which it accepts to be substituted in given contexts
- *Syntactic Content*: The content a unit receives as a result of the multiple dependencies it can maintain with respect to other units in its context
- *Informational Content*: The content related to the non-uniform distribution of units within those substitutability classes

Illustration of Formal Contents



(Gastaldi & Pellissier, 2021)

Outline

Empiricist Turn in Computer Science

Chomsky's Trap

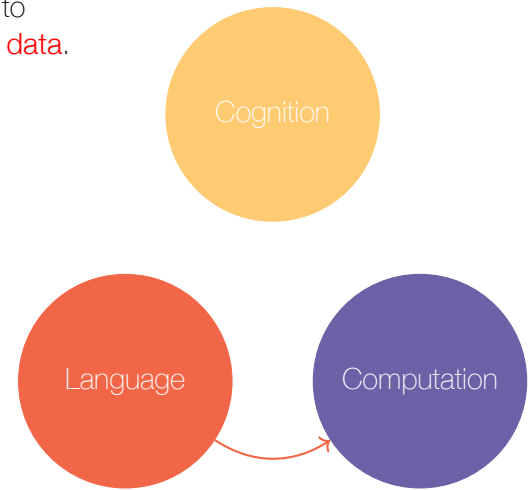
Language Models as Formal Objects

Philosophical Consequences

Takeaways

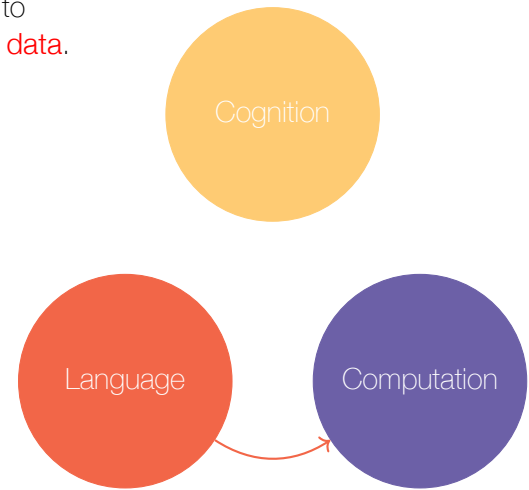
Takeaways

- A formal approach to data analysis can contribute to inferring symbolic language models from linguistic data.



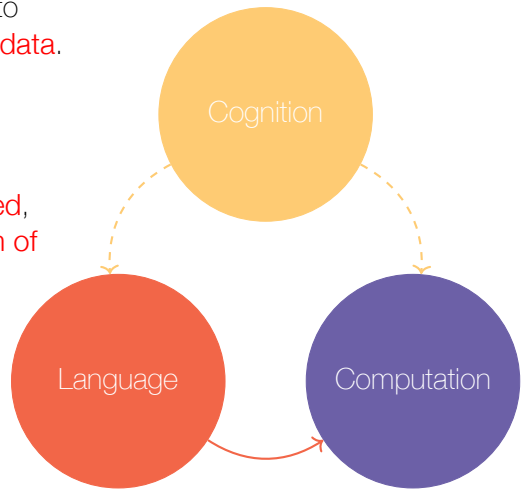
Takeaways

- A formal approach to data analysis can contribute to inferring symbolic language models from linguistic data.
- Resulting models are, a priori, models of the data.



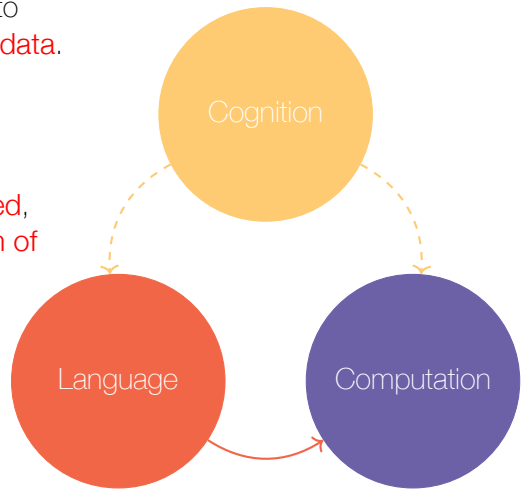
Takeaways

- A formal approach to data analysis can contribute to inferring symbolic language models from linguistic data.
- Resulting models are, a priori, models of the data.
- The cognitive content of such models is suspended, and cannot be restored without raising the problem of the data.



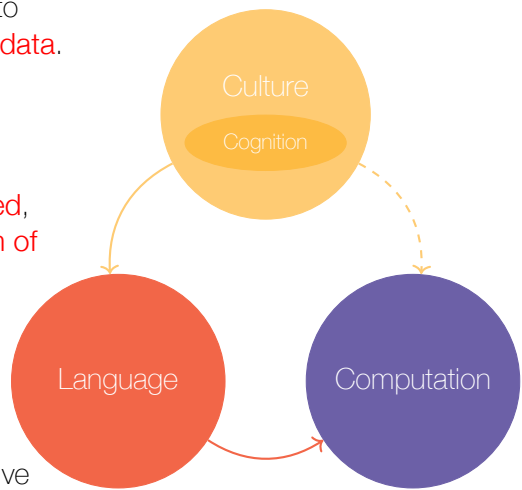
Takeaways

- A formal approach to data analysis can contribute to inferring symbolic language models from linguistic data.
- Resulting models are, a priori, models of the data.
- The cognitive content of such models is suspended, and cannot be restored without raising the problem of the data.
- The scale of the data for such models exceeds the individual scale.



Takeaways

- A formal approach to data analysis can contribute to inferring symbolic language models from linguistic data.
- Resulting models are, a priori, models of the data.
- The cognitive content of such models is suspended, and cannot be restored without raising the problem of the data.
- The scale of the data for such models exceeds the individual scale.
- Cultural conditions of data production become constitutive in the relation between cognitive contents and language models.



References I

Chomsky, N. (1953).Systems of syntactic analysis. *Journal of Symbolic Logic*, *18*(3), 242–256. https://doi.org/10.2307/2267409

Chomsky, N. (1955).Logical syntax and semantics: Their linguistic relevance. *Language*, *31*(1), 36–45.

Chomsky, N. (1956).Three models for the description of language. *IRE Transactions on Information Theory*, *2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813

Chomsky, N. (1957). *Syntactic structures*. Mouton; Co.

Chomsky, N. (1959). *Language*, *35*(1), 26–58. Retrieved July 7, 2025, from http://www.jstor.org/stable/411334

Chomsky, N. (1992, November). Language and the “cognitive revolutions” [Delivered November 23–27, 1992].

Church, A. (1936).An unsolvable problem of elementary number theory. *American Journal of Mathematics*, *58*(2), 345–363.

Gastaldi, J. L., & Pellissier, L. (2021).The calculus of language: explicit representation of emergent linguistic structure through type-theoretical paradigms. *Interdisciplinary Science Reviews*, *46*(4), 569–590. https://doi.org/10.1080/03080188.2021.1890484

Gödel, K. (1934). On undecidable propositions of formal mathematical systems. In *Collected works* (pp. 346–371). Clarendon Press Oxford University Press.

Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S. (2020).Zoom in: An introduction to circuits [https://distill.pub/2020/circuits/zoom-in]. *Distill*. https://doi.org/10.23915/distill.00024.001

Turing, A. (1937).On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230–265. https://doi.org/10.1112/plms/s2-42.1.230

Wittgenstein, L. (1953). *Philosophical investigations* (4th ed.). Wiley-Blackwell.

CNRS French–Danish Workshop
Reasoning in the Embedding Space
Copenhagen, Denmark

Empiricism vs. Formalism

In the Study of Embeddings

Juan Luis Gastaldi

www.giannigastaldi.com

ETH zürich

January 20, 2026