MALINCA
Institut Henri Poincaré
Paris, France

Language and Mathematics in the Time of AI

Philosophical and Theoretical Perspectives

Juan Luis Gastaldi

www.giannigastaldi.com

ETH zürich

October 1, 2025

Outline

Bring-Ins

Philosophical Perspectives

~~Historical Perspectives~~

Epistemological Perspectives

Conceptual Perspectives

Theoretical Perspectives

Takeaways

Outline

Bring-Ins

Philosophical Perspectives

~~Historical Perspectives~~

Epistemological Perspectives

Conceptual Perspectives

Theoretical Perspectives

Takeaways

Tony Wu at the IPAM

Tony Wu, *Autoformalization with Large Language Models* (IPAM (UCLA), Feb 15, 2023)

Melanie Mitchell on PaLM2

Melanie Mitchell
@MelMitchell1

Weird statement from Google's Palm 2 announcement.

(from blog.google/technology/ai/...)

- **Reasoning:** PaLM 2's wide-ranging dataset includes scientific papers and web pages that contain mathematical expressions. As a result, it demonstrates improved capabilities in logic, common sense reasoning, and mathematics.

6:23 PM · May 12, 2023 · **52.9K** Views

https://blog.google/technology/ai/google-palm-2-ai-large-language-model/

Outline

Bring-Ins

Philosophical Perspectives

~~Historical Perspectives~~

Epistemological Perspectives

Conceptual Perspectives

Theoretical Perspectives

Takeaways

What's So Funny?

- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.

What's So Funny?

- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.
- Mathematical literacy was supposed to be a condition to write and read scientific papers and mathematical expressions instead of their miraculous effect.

What's So Funny?

- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.
- Mathematical literacy was supposed to be a condition to write and read scientific papers and mathematical expressions instead of their miraculous effect.
- Natural language was considered the cause of rather than the solution to the multiple problems preventing mathematics from achieving higher degrees of precision.

What's So Funny?

- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.
- Mathematical literacy was supposed to be a condition to write and read scientific papers and mathematical expressions instead of their miraculous effect.
- Natural language was considered the cause of rather than the solution to the multiple problems preventing mathematics from achieving higher degrees of precision.
- More generally, the *formal* nature of mathematics was believed to make it impassive to the strong empirical position assumed by connectionist approaches guiding the application of DNNs.

Arithmetic in GPT-3

(Brown et al., 2020)

Reasoning in ChatGPT o3

If p_n is a sequence of probability distributions on a countably infinite set X that converge pointwise to a probabability distribution p on X, do they necessarily converge uniformly?

Reasoned about uniform convergence of probability distributions for 4m 41s >

We will show that if

$$p_n(x) \to p(x) \quad \text{for every } x \in X,$$

where X is a countably infinite set and each p_n and p is a probability distribution on X (so that

$$\sum_{x \in X} p_n(x) = \sum_{x \in X} p(x) = 1),$$

then in fact the convergence is uniform; that is,

$$\sup_{x \in X} |p_n(x) - p(x)| \to 0.$$

In other words, point-wise convergence of the probability mass functions on a countable set (with the limit being a probability distribution) forces uniform convergence.

In what follows we give a proof.

Reasoning in ChatGPT o3

Proof

Let $\varepsilon > 0$ be arbitrary. Our goal is to show that for all sufficiently large n we have
$\sup_{x \in X} |p_n(x) - p(x)| < \varepsilon.$

Because p is a probability measure on the countable set X, we can "capture almost all of its mass" in a finite set. (This is sometimes called a truncation argument.) More precisely, there exists a finite subset
$F \subset X$

such that

$$p(X \setminus F) = \sum_{x \notin F} p(x) < \varepsilon/3.$$

Now, since F is finite and we know that for every fixed x we have $p_n(x) \to p(x)$, it follows that for each $x \in F$ there is an index N_x such that for all $n \geq N_x$
$|p_n(x) - p(x)| < \varepsilon/(3|F|).$

If we let
$N_1 = \max\{N_x : x \in F\},$

then for all $n \geq N_1$ and every $x \in F$ we have

$$|p_n(x) - p(x)| < \frac{\varepsilon}{3|F|}.$$

Thus, summing over $x \in F$ we obtain

Reasoning in ChatGPT o3

Conclusion

Yes. If a sequence $\{p_n\}$ of probability distributions on a countably infinite set X converges point-wise to a probability distribution p, then in fact

$$\sup_{x \in X} |p_n(x) - p(x)| \to 0,$$

so the convergence is uniform.

A Note on the Issue

It is very important that the limit p is itself a probability distribution. (For example, the sequence of "Dirac measures" δ_n on \(\Bbb N\) defined by

$$\delta_n(x) = \begin{cases} 1 & \text{if } x = n, \\ 0 & \text{otherwise}, \end{cases}$$

converges point-wise to the zero function; but the zero function is not a probability distribution, and indeed the convergence is not uniform.) The fact that no mass "escapes" ensures that the tiny probabilities in the tail force the errors there to be small uniformly.

Final Answer

Yes. If p_n is a sequence of probability distributions on a countably infinite set X that converges point-wise to a probability distribution p (so that no mass is lost), then in fact the convergence is uniform.

Reasoning in ChatGPT o3

Reasoned about uniform convergence of probability distributions for 4m 41s

> I'm curious if pointwise convergence of probability distributions on a countably infinite set implies uniform convergence. This involves examining the supremum of the absolute difference between distributions as it approaches zero.
>
> OK, let me see: pointwise convergence on a countably infinite set doesn't guarantee uniform convergence without additional constraints, like finite sets or specific continuity conditions.
>
> I'm curious about sequences of probability measures where each point's probability trends to zero. This hints at pointwise convergence but raises questions about uniform convergence due to potential "spikes."
>
> OK, let's think it through: the maximum difference of 1/n between distributions p_n and p as n increases. This difference converges to zero, suggesting uniform convergence in a bounded setup.
>
> I'm examining a counterexample of pointwise convergence without uniform convergence, where individual probabilities decline to zero but the overall distribution doesn't converge uniformly due to continuous fluctuation.
>
> I'm considering pointwise convergence with a shifting spike in a probability distribution on natural numbers, highlighting the gap between pointwise and uniform convergence due to inconsistent distribution updates.

Main Orientations in DNN Applications to Maths

- Proof-Oriented
  - Bansal et al., 2019; Polu and Sutskever, 2020; Wu et al., 2022.
- Object-Oriented
  - Blechschmidt and Ernst, 2021; Charton, 2021; d'Ascoli et al., 2022; Lample and Charton, 2019; Li et al., 2021
- Skill-Oriented (e.g., Reasoning)
  - Brown et al., 2020; Lewkowycz et al., 2022; Shen et al., 2021
- Heuristic-Oriented
  - Davies et al., 2021; Wagner, 2021

(Alemi et al., 2016)

(Lample and Charton, 2019)

(Peng et al., 2021)

(Davies et al., 2021)

Philosophical Significance: The Return of Language

- Research orientations tend to be spontaneously organized according to the AI researchers' implicit assumptions as to what characterizes mathematical practice (i.e. what it is that we do when we do mathematics).

Philosophical Significance: The Return of Language

- Research orientations tend to be spontaneously organized according to the AI researchers' implicit assumptions as to what characterizes mathematical practice (i.e. what it is that we do when we do mathematics).

- However, practically all applications share a common philosophical assumption: *Written natural language* plays a critical role in the processing mathematical knowledge.

Philosophical Significance: The Return of Language

- Research orientations tend to be spontaneously organized according to the AI researchers' implicit assumptions as to what characterizes mathematical practice (i.e. what it is that we do when we do mathematics).

- However, practically all applications share a common philosophical assumption: *Written natural language* plays a critical role in the processing mathematical knowledge.

- The potential success of DNN methods in mathematics is inseparable from a reorientation of the epistemology of mathematics from logic and formal systems to natural language and vernacular writing practices.

Outline

Bring-Ins

Philosophical Perspectives

~~Historical Perspectives~~

Epistemological Perspectives

Conceptual Perspectives

Theoretical Perspectives

Takeaways

Outline

Bring-Ins

Philosophical Perspectives

~~Historical Perspectives~~

Epistemological Perspectives

Conceptual Perspectives

Theoretical Perspectives

Takeaways

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.

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

DO LLMS HAVE CON

Naama Rozen
Tel-Aviv University
naamarozen240@gmail.com

Gal Elidan
Google Research
Hebrew University
elidan@google.com

Amir Globerson
Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel
Tel-Aviv University
della@tauex.tau

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Recursive Self-Improvement through Reinf

Yoichi Ishibashi
NEC
yoichi-ishibashi@nec.com

DO LLMS "KNOW" INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Juyeon H
Udhay Na
[1]Univers
jh2324@

DO LLMS RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING IN LLMS

Can Larg

Cheng-Han Chiang
National Taiwan University,
Taiwan
dcml0714@gmail.com

Hung-yi Lee
National Taiwan University,
Taiwan
hungyilee@ntu.edu.tw

n[1 †]

Zico Kolter, *Building Safe and*

Language Models are Few-Shot Learners

Tom B. Brown* Benjamin Mann* Nick Ryder* Melanie Subbiah*
Jared Kaplan† Prafulla Dhariwal Arvind Neelakantan Pranav Shyam Girish Sastry
Amanda Askell Sandhini Agarwal Ariel Herbert-Voss Gretchen Krueger Tom Henighan

Empirical Saturnalia

DO LLMS HAVE CON

Naama Rozen
Tel-Aviv University
naamarozen240@gmail.com

Gal Elidan
Google Research
Hebrew University
elidan@google.com

Amir Globerson
Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel
Tel-Aviv University
della@tauex.tau

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Recursive Self-Improvement through Reinf

Yoichi Ishibashi
NEC
yoichi-ishibashi@nec.com

DO LLMS "KNOW" INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Juyeon H
Udhay Na
[1]Univers
jh2324@

Can Larg

Cheng-Han Chiang
National Taiwan University,
Taiwan
dcml0714@gmail.com

Hung-yi Lee
National Taiwan University,
Taiwan
hungyilee@ntu.edu.tw

DO LLMS RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING IN LLMS

n[1 †]

Language Models are Few-Shot Learners

k Ryder*
Melanie Subbiah*
n
Pranav Shyam
Girish Sastry
Gretchen Krueger
Tom Henighan

And this was actually a very pr

LLMs Are Not Intelligent Thinkers: Introducing Mathematical Topic Tree Benchmark for Comprehensive Evaluation of LLMs

Arash Gholami Davoodi[1], Seyed Pouyan Mousavi Davoudi, Pouya Pezeshkpour[2]
[1]Carnegie Mellon University, [2]Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Empirical Saturnalia

DO LLMS HAVE CON

Naama Rozen
Tel-Aviv University
naamarozen240@gmail.com

Gal Elidan
Google Research
Hebrew University
elidan@google.com

Amir Globerson
Google Research
Tel-Aviv University
amirg@google.com

Ella Daniel
Tel-Aviv University
della@tauex.tau

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Recursive Self-Improvement through Reinf

Yoichi Ishibashi
NEC
yoichi-ishibashi@nec.com

DO LLMS "KNOW" INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Juyeon H
Udhay Na
[1]Univers
jh2324@

DO LLMS RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING IN LLMS

Can Larg

Cheng-Han Chiang
National Taiwan University, Taiwan

Hung-yi Lee
National Taiwan University, Taiwan

Language Models are Few-Shot Learners

k Ryder*
Melanie Subbiah*
n
Pranav Shyam
Girish Sastry
Gretchen Krueger
Tom Henighan

When Can LLMs *Actually* Correct Their Own Mistakes? A Critical Survey of Self-Correction of LLMs

Ryo Kamoi[1] Yusen Zhang[1] Nan Zhang[1] Jiawei Han[2] Rui Zhang[1]
[1]Penn State University, USA [2]University of Illinois Urbana-Champaign, USA
{ryokamoi,rmz5227}@psu.edu

cing Mathematical Topic Tree
valuation of LLMs

i Davoudi, Pouya Pezeshkpour[2]
Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Empirical Saturnalia

DO LLMS HAVE CON

Naama Rozen
Tel-Aviv University

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Recursive Self-Improvement through Reinf

Yoichi Ishibashi
NEC

DO LLMS "KNOW" INTERNALLY WHEN THEY FOLLOW INSTRUCTIONS?

Can Larg

Nat

DO LLMS RECOGNIZE YOUR PREFERENCES? EVALUATING PERSONALIZED PREFERENCE FOLLOWING IN LLMS

Large Language Models are Zero-Shot Reasoners

Takeshi Kojima
The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Shane Gu
Google Research, Brain Team

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

Yusuke Iwasawa
The University of Tokyo

Ryo Kamoi[1] Yusen Zhang[1] Nan Zhang[1] Jiawei Han[2] Rui Zhang[1]
[1]Penn State University, USA [2]University of Illinois Urbana-Champaign, USA
{ryokamoi,rmz5227}@psu.edu

Language Models are Few-Shot Learners

Melanie Subbiah*

Pranav Shyam

Girish Sastry

Gretchen Krueger

Tom Henighan

ng Mathematical Topic Tree
valuation of LLMs

i Davoudi, Pouya Pezeshkpour[2]
Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?: Algorithm Discovery for Recursive Self-Improvement through Reinf

DO LLMS HAVE CON

Naa
Tel-A

Can Larg

Nat

Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research

OW" INTERNALLY WHEN THEY FOLLOW
?

LLMS RECOGNIZE YOUR PREFERENCES? EVAL-
ING PERSONALIZED PREFERENCE FOLLOWING IN
MS

Takeshi Kojima
The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Shane Gu
Google Research, Brain Team

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

Yusuke Iwasawa
The University of Tokyo

Ryo Kamoi[1] Yusen Zhang[1] Nan Zhang[1] Jiawei Han[2] Rui Zhang[1]
[1]Penn State University, USA [2]University of Illinois Urbana-Champaign, USA
{ryokamoi,rmz5227}@psu.edu

Language Models are Few-Shot Learners

ng Mathematical Topic Tree
valuation of LLMs

k Ryder* Melanie Subbiah*
n Pranav Shyam Girish Sastry
Gretchen Krueger Tom Henighan

i Davoudi, Pouya Pezeshkpour[2]
Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?:

PROCEDURAL KNOWLEDGE IN PRETRAINING DRIVES REASONING IN LARGE LANGUAGE MODELS

Laura Ruis* — AI Centre, UCL
Maximilian Mozes — Cohere
Juhan Bae — University of Toronto & Vector Institute

DO LLMS HAVE CON

Sparks

Sébastien Bubeck
Eric Horvitz Ece Kar
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang
Microsoft Research

WHEN THEY FOLLOW

ZE YOUR PREFERENCES? EVAL-
ING PERSONALIZED PREFERENCE FOLLOWING IN
MS

Can Larg

Nat

Takeshi Kojima — The University of Tokyo — t.kojima@weblab.t.u-tokyo.ac.jp
Shixiang Shane Gu — Google Research, Brain Team
Machel Reid — Google Research*
Yutaka Matsuo — The University of Tokyo
Yusuke Iwasawa — The University of Tokyo

Language Models are Few-Shot Learners

k Ryder* Melanie Subbiah*
n Pranav Shyam Girish Sastry
Gretchen Krueger Tom Henighan

ng Mathematical Topic Tree
valuation of LLMs

i Davoudi, Pouya Pezeshkpour[2]
Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Ryo Kamoi[1] Yusen Zhang[1] Nan Zhang[1] Jiawei Han[2] Rui Zhang[1]
[1]Penn State University, USA [2]University of Illinois Urbana-Champaign, USA
{ryokamoi,rmz5227}@psu.edu

n[1 †]

Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?:

Are Large Language Models Reliable Judges?
A Study on the Factuality Evaluation Capabilities of LLMs

Xue-Yong Fu, Md Tahmid Rahman Laskar, Cheng Chen, Shashi Bhushan TN
Dialpad Canada Inc.
{xue-yong,tahmid.rahman,cchen,sbhushan}@dialpad.com

Laura Ruis*
AI Centre, UCL

Sébastien Bubeck
Eric Horvitz Ece Kar
Harsha Nori Hamid Palangi
Microsoft Rese

Takeshi Kojima
The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Shane Gu
Google Research, Brain Team

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

Yusuke Iwasawa
The University of Tokyo

Ryo Kamoi[1] Yusen Zhang[1] Nan Zhang[1] Jiawei Han[2] Rui Zhang[1]
[1]Penn State University, USA [2]University of Illinois Urbana-Champaign, USA
{ryokamoi,rmz5227}@psu.edu

Language Models are Few-Shot Learners

Melanie Subbiah*
Pranav Shyam Girish Sastry
Gretchen Krueger Tom Henighan

Davoudi, Pouya Pezeshkpour[2]
Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?:

DO LLMS HAVE CON

PROCEDURAL KNOWLEDGE IN PRETRAINING DRIVES REASONING IN LLOW

Laura Ruis*
AI Centre, UCL

Sparks
E

Sébastien Bubeck
Eric Horvitz Ece Kar
Harsha Nori Hamid Palangi Ma
Microsoft Rese

Are Large Language Models Reliable Judges?
A Study on the Factuality Evaluation Capabilities of LLMs

Xue-

Can Larg

La

? EVAL-
WING IN

Nat

n[1 †]

TEACHING ARITHMETIC TO SMALL TRANSFORMERS

Nayoung Lee*
University of Wisconsin-Madison
nayoung.lee@wisc.edu

Kartik Sreenivasan*
University of Wisconsin-Madison
ksreenivasa2@wisc.edu

Jason D. Lee
Princeton University
jasonlee@princeton.edu

Kangwook Lee
University of Wisconsin-Madison
kangwook.lee@wisc.edu

Takeshi Kojima
The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Sha
Google Research,

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

Yusuke
The Univer

ng Mathematical Topic Tree
valuation of LLMs

k Ryder* Melanie Subbiah*

n Pranav Shyam Girish Sastry

Gretchen Krueger Tom Henighan

Ryo Kamoi[1] Yusen Zhang[1] Nan Zhang[1] Jiawei Han[2] Rui Zhang[1]
[1]Penn State University, USA [2]University of Illinois Urbana-Champaign, USA
{ryokamoi,rmz5227}@psu.edu

i Davoudi, Pouya Pezeshkpour[2]
Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?:

DO LLMS HAVE CON

PROCEDURAL KNOWLEDGE IN PRETRAINING DRIVES REASONING IN

Laura Ruis*
AI Centre, UCL

Sparks

Sébastien Bubeck
Eric Horvitz Ece Kar
Harsha Nori Hamid Palangi
Microsoft Rese

Are Large Language Models Reliable Judges?
A Study on the Factuality Evaluation Capabilities of LLMs

Xue-

TEACHING ARITHMETIC TO SMALL TRANSFORMERS

Nayoung Lee*
University of Wisconsin-Madison

Kartik Sreenivasan*
University of Wisconsin-Madison

Takeshi Kojima
The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

IS YOUR MODEL REALLY A GOOD MATH REASONER? EVALUATING MATHEMATICAL REASONING WITH CHECKLIST

Zihao Zhou[12*] Shudong Liu[3*] Maizhen Ning[126] Wei Liu[4] Jindong Wang[5]
Derek F. Wong[3] Xiaowei Huang[2] Qiufeng Wang[1†] Kaizhu Huang[6]
[1]Xi'an Jiaotong-liverpool University [2]University of Liverpool [3]University of Macau
[4]HKUST [5]William & Mary [6]Duke Kunshan University

Ryo Kamoi[1] Yusen Zhang[1] Nan Zhang[1] Jiawei
[1]Penn State University, USA [2]University of Illinois Urba
{ryokamoi,rmz5227}@psu.edu

Megagon Labs
agholami@andrew.cmu.edu, spouyan.mousavi@gmail.com, pouya@megagon.ai

Gretchen Krueger Tom Henighan

Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?:

DO LLMS HAVE CON

PROCEDURAL
REASONING IN

Laura Ruis*
AI Centre, UCL

Sparks

Sébastien Bubeck
Eric Horvitz Ece Kar
Harsha Nori Hamid Palangi
Microsoft Rese

Are Large Language Models Reliable Judges?
A Study on the Factuality Evaluation Capabilities of LLMs

Xue-

TEACHING ARITHMETIC TO SMALL TRANSFORMERS

Nayoung Lee*
University of Wisconsin-Madison

Kartik Sreenivasan*
University of Wisconsin-Madison

Takeshi Kojima
The University of To
t.kojima@weblab.t.u-to

Machel Reid
Google Research*

Ryo Kamoi[1] Yusen Zhan
[1]Penn State University, USA
{ryokan

Learning Formal Mathematics
From Intrinsic Motivation

Gabriel Poesia[1] David Broman[4] Nick Haber[1,3] Noah D. Goodman[1,2]
{poesia,nhaber,ngoodman}@stanford.edu dbro@kth.se
Departments of Computer Science[1], Psychology[2] and Education[3], Stanford University, USA
EECS and Digital Futures[4], KTH Royal Institute of Technology, Sweden

Gretchen Krueger Tom Henighan

Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?:

DO LLMS HAVE CON

PROCEDURAL KNOWLEDGE IN PRETRAINING DRIVES REASONING IN

Laura Ruis*
AI Centre, UCL

Sparks

Sébastien Bubeck
Eric Horvitz Ece Kar
Harsha Nori Hamid Palangi
Microsoft Rese

Are Large Language Models Reliable Judges?
A Study on the Factuality Evaluation Capabilities of LLMs

Xue-

TEACHING ARITHMETIC TO SMALL TRANSFORMERS

Nayoung Lee*
University of Wisconsin-Madison

Kartik Sreenivasan*
University of Wisconsin-Madison

Takeshi Kojima

(How) Can Transformers Predict Pseudo-Random Numbers?

Tao Tao*[1] Darshil Doshi*[1] Dayal Singh Kalra*[2] Tianyu He*[1] Maissam Barkeshli[1,3]
{tao2021, ddoshi, dayal, tianyuh, maissam}@umd.edu

{ryokan

Gabriel Poesia[1] David Broman[4] Nick Haber[1,3] Noah D. Goodman[1,2]
{poesia,nhaber,ngoodman}@stanford.edu dbro@kth.se
Departments of Computer Science[1], Psychology[2] and Education[3], Stanford University, USA
EECS and Digital Futures[4], KTH Royal Institute of Technology, Sweden

Gretchen Krueger Tom Henighan

Empirical Saturnalia

Can Large Language Models Invent Algorithms to Improve Themselves?:

DO LLMS HAVE CON

PROCEDURAL
REASONING IN

Are Large Language Models Reliable Judges?
A Study on the Factuality Evaluation Capabilities of LLMs

Sparks

Sébastien Bubeck
Eric Horvitz Ece Kar
Harsha Nori

Can Larg

Takeshi Kojima

(How) Can

Tao Tao *[1] Darsh Doshi Dayal Singh Kalra Tianyu He Maissam Barkeshli
{tao2021, ddoshi, dayal, tianyuh, maissam}@umd.edu

{ryokan

LLOW

LL TRANSFORMERS

Kartik Sreenivasan*
University of Wisconsin-Madison
.edu

Madison
.edu

? EVAL-
WING IN

n‡ †

H REASONER?
ONING WITH

dong Wang[5]
ng[6]
ersity of Macau

iah*

h Sastry

Gretchen Krueger Tom Henighan

GPT Can Solve Mathematical Problems Without a Calculator

Zhen Yang[†*], Ming Ding[§†*], Qingsong Lv[§], Zhihuan Jiang[†], Zehai He[†], Yuyi Guo[†], Jinfeng Bai[◇], Jie Tang[†‡]

[†]Tsinghua University [◇]TAL AI Lab [§]Zhipu.AI

Gabriel Poesia[1] David Broman[4] Nick Haber[1,3] Noah D. Goodman[1,2]
{poesia,nhaber,ngoodman}@stanford.edu dbro@kth.se
Departments of Computer Science[1], Psychology[2] and Education[3], Stanford University, USA
EECS and Digital Futures[4], KTH Royal Institute of Technology, Sweden

Empirical Saturnalia

Programs As Formal Objects

Programs As Formal Objects

Programs As Formal Objects

Programs As Formal Objects

Programs As Formal Objects

Programs As Formal Objects

New Research Orientation

- Why and how, *irrespective of the model*, a phenomenon under study can be captured through the distributional analysis of data at all?

- What are, *irrespective of the model*, the general principles for capturing distributional phenomena?

Outline

Bring-Ins

Philosophical Perspectives

~~Historical Perspectives~~

Epistemological Perspectives

Conceptual Perspectives

Theoretical Perspectives

Takeaways

Distributionalism and Word Embeddings

- Distributional Hypothesis
  (Harris, 1960; Saussure, 1959)
  - "You shall know a word by the company it keeps!" (Firth, 1935)
  - "Words which are similar in meaning occur in similar contexts" (Rubenstein & Goodenough, 1965)
  - The content of a linguistic unit is determined by its distribution over a corpus (i.e., the other units appearing in its context)

Distributionalism and Word Embeddings

- Distributional Hypothesis
  (Harris, 1960; Saussure, 1959)
  - "You shall know a word by the company it keeps!" (Firth, 1935)
  - "Words which are similar in meaning occur in similar contexts" (Rubenstein & Goodenough, 1965)
  - The content of a linguistic unit is determined by its distribution over a corpus (i.e., the other units appearing in its context)

- Computational version:
  Word Embeddings

(https://projector.tensorflow.org)

Distributionalism and Word Embeddings

- Distributional Hypothesis
  (Harris, 1960; Saussure, 1959)
  - "You shall know a word by the company it keeps!" (Firth, 1935)
  - "Words which are similar in meaning occur in similar contexts" (Rubenstein & Goodenough, 1965)
  - The content of a linguistic unit is determined by its distribution over a corpus (i.e., the other units appearing in its context)

- Computational version:
  Word Embeddings

- Interpretability efforts on vector representations of mathematical expressions
  - (e.g. d'Ascoli et al., 2022; Gao et al., 2017; Krstovski & Blei, 2018; Purgał et al., 2021; Ryskina & Knight, 2021)

(d'Ascoli et al., 2022)

Peano Axioms

1. 0 is a number.
   $0 \in \mathbb{N}$
2. If n is a number, the successor of n is a number.
   $n \in \mathbb{N} \implies \text{succ}(n) \in \mathbb{N}$
3. 0 is not the successor of a number.
   $\forall n \in \mathbb{N}, 0 \neq \text{succ}(n)$
4. Two numbers of which the successors are equal are themselves equal.
   $\forall n, m \in N, \text{succ}(x) = \text{succ}(y) \implies x = y$
5. If a set $\mathbf{S}$ of numbers contains 0 and also the successor of every number in $\mathbf{S}$, then every number is in $\mathbf{S}$ (induction axiom).
   $0 \in S \wedge (\forall n, n \in S \implies \text{succ}(s) \in \mathbf{S}) \implies \forall n \in \mathbf{S}, n \in \mathbb{N}$

The Distributional Properties of Characters

$$A_{i,j} = pmi(c_i; c_j) = \log \frac{p(c_i, c_j)}{p(c_i)p(c_j)}$$

The Characteristic Content of Digits

$$O := \{=, -, \text{-}, /\}$$

$$D := \{0, 9, 8, 3, 4, 5, 6, 7, 1, 2\}$$

$$V := \{\mathtt{u}, \mathtt{i}, \mathtt{e}, \mathtt{a}, \mathtt{o}\}$$

$$C := \{\mathtt{y}, \mathtt{k}, \mathtt{z}, \mathtt{x}, \mathtt{s}, \mathtt{d}, \mathtt{g}, \mathtt{n}, \mathtt{l}, \mathtt{r}, \mathtt{h}, \mathtt{c}, \mathtt{t}, \mathtt{q}, \mathtt{v}, \mathtt{f}, \mathtt{m}, \mathtt{p}, \mathtt{w}, \mathtt{b}, \mathtt{j}\}$$

The Syntactic Content of Digits

The Syntactic Content of Digits

The Syntactic Content of Digits

The Characteristic Content of Digit Pairs

Benford's Law

TABLE I

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit									Count
		1	2	3	4	5	6	7	8	9	
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	$n^{-1}, \sqrt{n}, \cdots$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	*Digest*	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average.....		30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Probable Error		±0.8	±0.4	±0.4	±0.3	±0.2	±0.2	±0.2	±0.2	±0.3	—

Informational Content and Order

δ	$p(\delta)$	$n_{10}(\delta)$
1	0.301	1
2	0.176	2
3	0.125	3
4	0.097	4
5	0.079	5
6	0.067	6
7	0.058	7
8	0.051	8
9	0.046	9

$$D := \{\mathtt{0}, \mathtt{1}, \mathtt{2}, \mathtt{3}, \mathtt{4}, \mathtt{5}, \mathtt{6}, \mathtt{7}, \mathtt{8}, \mathtt{9}, \mathtt{0}\}$$
$$N := D^*$$
$$\delta \in D$$
$$\nu_1, \nu_1', \nu_2, \nu_2' \in N$$
$$n_b : N \to \mathbb{N}$$
$$\nu \mapsto \text{valeur numérique de } \nu \text{ en base } b$$
$$p : D \to [0, 1]$$
$$\delta \mapsto \log\left(1 + \frac{1}{n_b(\delta)}\right)$$

Informational Content and Order

δ	$p(\delta)$	$n_{10}(\delta)$
1	0.301	1
2	0.176	2
3	0.125	3
4	0.097	4
5	0.079	5
6	0.067	6
7	0.058	7
8	0.051	8
9	0.046	9

$D := \{\mathtt{0}, \mathtt{1}, \mathtt{2}, \mathtt{3}, \mathtt{4}, \mathtt{5}, \mathtt{6}, \mathtt{7}, \mathtt{8}, \mathtt{9}, \mathtt{0}\}$

$N := D^*$

$\delta \in D$

$\nu_1, \nu_1', \nu_2, \nu_2' \in N$

$n_b : N \to \mathbb{N}$

$\nu \mapsto$ valeur numérique de ν en base b

$p : D \to [0, 1]$

$$\delta \mapsto \log\left(1 + \frac{1}{n_b(\delta)}\right)$$

$$p(\nu\nu') \stackrel{?}{=} \mathcal{C}(p(\nu), p(\nu'))$$

$$n_b(\nu_1\nu_1') \leq n_b(\nu_2\nu_2') \iff \mathcal{C}(p(\nu_1), p(\nu_1')) \leq \mathcal{C}(p(\nu_2), p(\nu_2'))$$

Constructing $\mathcal{C}$

$$p(\delta) = \mathcal{B}(\delta) = \log_b \left(1 + \frac{1}{n_b(\delta)}\right)$$

Constructing $\mathcal{C}$

$$p(\delta) = \mathcal{B}(\delta) = \log_b\left(1 + \frac{1}{n_b(\delta)}\right)$$

$$n_b(\delta) = \mathcal{B}^{-1}(\delta) = \frac{1}{b^{p(\delta)} - 1}$$

Constructing $\mathcal{C}$

$$p(\delta) = \mathcal{B}(\delta) = \log_b\left(1 + \frac{1}{n_b(\delta)}\right)$$
$$n_b(\delta) = \mathcal{B}^{-1}(\delta) = \frac{1}{b^{p(\delta)} - 1}$$

$$\mathcal{B}(\nu\nu') = \log_b(1 + \frac{1}{n_b(\nu\nu')})$$

Constructing $\mathcal{C}$

$$p(\delta) = \mathcal{B}(\delta) = \log_b\left(1 + \frac{1}{n_b(\delta)}\right)$$
$$n_b(\delta) = \mathcal{B}^{-1}(\delta) = \frac{1}{b^{p(\delta)} - 1}$$

$$\begin{aligned}\mathcal{B}(\nu\nu') &= \log_b(1 + \frac{1}{n_b(\nu\nu')}) \\ &= \log_b(1 + \frac{1}{n_b(\nu) \cdot b^{|\nu'|} + n_b(\nu')})\end{aligned}$$

Constructing $\mathcal{C}$

$$p(\delta) = \mathcal{B}(\delta) = \log_b\left(1 + \frac{1}{n_b(\delta)}\right)$$
$$n_b(\delta) = \mathcal{B}^{-1}(\delta) = \frac{1}{b^{p(\delta)}-1}$$

$$\begin{aligned}
\mathcal{B}(\nu\nu') &= \log_b(1 + \frac{1}{n_b(\nu\nu')}) \\
&= \log_b(1 + \frac{1}{n_b(\nu)\cdot b^{|\nu'|} + n_b(\nu')}) \\
&= \log_b(1 + \frac{1}{\mathcal{B}^{-1}(\nu)\cdot b^{|\nu'|} + \mathcal{B}^{-1}(\nu')}) \\
&= \log_b(1 + \frac{1}{\frac{1}{b^{p(\nu)}-1}\cdot b^{|\nu'|} + \frac{1}{b^{p(\nu')}-1}})
\end{aligned}$$

$\left(\text{with } |\nu| = \left\lfloor \log_b\left(\frac{1}{b^{p(\nu)}+1}\right)\right\rfloor + 1\right)$

Constructing $\mathcal{C}$

$$p(\delta) = \mathcal{B}(\delta) = \log_b\left(1 + \frac{1}{n_b(\delta)}\right)$$
$$n_b(\delta) = \mathcal{B}^{-1}(\delta) = \frac{1}{b^{p(\delta)}-1}$$

$$\begin{aligned}
p(\nu\nu') = \mathcal{B}(\nu\nu') &= \log_b(1 + \frac{1}{n_b(\nu\nu')}) \\
&= \log_b(1 + \frac{1}{n_b(\nu)\cdot b^{|\nu'|} + n_b(\nu')}) \\
&= \log_b(1 + \frac{1}{\mathcal{B}^{-1}(\nu)\cdot b^{|\nu'|} + \mathcal{B}^{-1}(\nu')}) \\
&= \log_b(1 + \frac{1}{\frac{1}{b^{p(\nu)}-1}\cdot b^{|\nu'|} + \frac{1}{b^{p(\nu')}-1}}) \\
&=: \mathcal{C}(p(\nu), p(\nu'))
\end{aligned}$$

$\left(\text{with } |\nu| = \left\lfloor \log_b\left(\frac{1}{b^{p(\nu)}+1}\right)\right\rfloor + 1\right)$

Benford in Data (δ)

Benford in Data (δ)

Benford in Data (δ)

Benford in Data ($\delta\delta'$)

Benford in Data ($\delta\delta'$)

Outline

Bring-Ins

Philosophical Perspectives

~~Historical Perspectives~~

Epistemological Perspectives

Conceptual Perspectives

Theoretical Perspectives

Takeaways

Embeddings in LLMs

Epistemology of Machine Learning
Distributional Language Models

Epistemology of Machine Learning
Distributional Language Models

(https://tiktokenizer.vercel.app)

Embeddings in LLMs

Epistemology of Machine Learning
Distributional Language Models

Epistemology of Machine Learning
Distributional Language Models

(https://tiktokenizer.vercel.app)

(https://projector.tensorflow.org)

Embeddings in LLMs

Epistemology of Machine Learning
Distributional Language Models

Tokenization
(Sennrich et al., 2016)

Embedding
(Mikolov et al., 2013)

Attention
(Vaswani et al., 2017)

Epistemology of Machine Learning
Distributional Language Models

(https://tiktokenizer.vercel.app)

(https://projector.tensorflow.org)

(https://github.com/jessevig/bertviz)

Embeddings in LLMs

```
Epistemology of Machine Learning
Distributional Language Models
```

(https://tiktokenizer.vercel.app)

(https://projector.tensorflow.org)

Formal Explainability

```
Epistemology of Machine Learning
Distributional Language Models
```

Formal Explainability

Structure

?

Epistemology of Machine Learning
Distributional Language Models

Embedding

Data

Embeddings: Similarity and Analogy

(https://projector.tensorflow.org)

Embeddings: Other Applications

(Mikolov, Sutskever, Chen, Corrado, Dean, et al., 2013)

(Belrose et al., 2024)

(https://nlp.stanford.edu/~johnhew/structural-probe.html)

Formal Explainability

```
Epistemology of Machine Learning
Distributional Language Models
```

Formal Explainability

Structure

?

Epistemology of Machine Learning
Distributional Language Models

Embedding

Data

Formal Explainability

Structure

?

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

Embedding

Data

word2vec Explained

(Levy & Goldberg, 2014)

$$\ell = \sum_{w \in V_w} \sum_{c \in V_c} \#(w,c)\left(\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D}[\log \sigma(-\vec{w} \cdot \vec{c}_N)]\right)$$

$$\frac{\partial \ell}{\partial(\vec{w} \cdot \vec{c})} = 0 \quad \text{when} \quad \vec{w} \cdot \vec{c} \quad = \log\left(\frac{\#(w,c) \cdot |D|}{\#(w) \cdot \#(c)}\right) - \log k$$

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.

word2vec Explained

(Levy & Goldberg, 2014)

$$\ell = \sum_{w \in V_w} \sum_{c \in V_c} \#(w,c)\left(\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D}[\log \sigma(-\vec{w} \cdot \vec{c}_N)]\right)$$

$$\frac{\partial \ell}{\partial(\vec{w} \cdot \vec{c})} = 0 \quad \text{when} \quad \vec{w} \cdot \vec{c} \quad = \log\left(\frac{\#(w,c) \cdot |D|}{\#(w) \cdot \#(c)}\right) - \log k$$

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.
- The Singular Value Decomposition (SVD) provides an exact solution to this problem.

Embedding structure

Structure

?

Data

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

Embedding

Embedding structure

Structure

$\{-, /, 0, 1, 2, \ldots, 8, 9, =,$

$a, b, c, \ldots, w, x, y, z, é\}$

Embedding

Data

SVD

We need something more formal

4 Why does this produce good word representations?

Good question. We don't really know.

The distributional hypothesis states that words in similar contexts have similar meanings. The objective above clearly tries to increase the quantity $v_w \cdot v_c$ for good word-context pairs, and decrease it for bad ones. Intuitively, this means that words that share many contexts will be similar to each other (note also that contexts sharing many words will also be similar to each other). This is, however, very hand-wavy.

Can we make this intuition more precise? We'd really like to see something more formal.

(Goldberg and Levy, 2014)

From matrices to distributional operators

$$X = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$Y = X \times X = \{\,\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\,\}$$

$$M\colon X \times Y \to \mathbb{R}$$

$$(x, y) \mapsto \mathrm{pmi}(x, y)$$

From matrices to distributional operators

$$X = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$Y = X \times X = \{\,\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\,\}$$

$$M\colon X \times Y \to \mathbb{R}$$

$$(x, y) \mapsto \mathrm{pmi}(x, y)$$

$$M_x\colon X \to \mathbb{R}^Y$$

$$x \mapsto M(x, -)$$

From matrices to distributional operators

$$X = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$Y = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

$$M\colon X \times Y \to \mathbb{R}$$
$$(x, y) \mapsto \mathrm{pmi}(x, y)$$

$$M_x\colon X \to \mathbb{R}^Y$$
$$x \mapsto M(x, -)$$

$$M_y\colon Y \to \mathbb{R}^X$$
$$y \mapsto M(-, y)$$

From matrices to distributional operators

$$X = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$Y = X \times X = \{\,\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\,\}$$

$$M\colon X \times Y \to \mathbb{R}$$
$$(x, y) \mapsto \mathrm{pmi}(x, y)$$

$$M_x\colon X \to \mathbb{R}^Y$$
$$x \mapsto M(x, -)$$

$$M_y\colon Y \to \mathbb{R}^X$$
$$y \mapsto M(-, y)$$

$$X \xrightarrow{\;M_x\;} \mathbb{R}^Y$$

$$\mathbb{R}^X \xleftarrow[\;M_y\;]{} Y$$

From matrices to distributional operators

$$X = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$Y = X \times X = \{\,\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\,\}$$

$$M\colon X \times Y \to \mathbb{R}$$
$$(x, y) \mapsto \mathrm{pmi}(x, y)$$

$$M_x\colon X \to \mathbb{R}^Y$$
$$x \mapsto M(x, -)$$

$$M_y\colon Y \to \mathbb{R}^X$$
$$y \mapsto M(-, y)$$

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \\
\mathbb{R}^X & Y \arrow[l, "M_y"] \arrow[u, hook]
\end{tikzcd}$$

From matrices to distributional operators

$$X = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$Y = X \times X = \{\,\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\,\}$$

$$\begin{aligned} M \colon X \times Y &\to \mathbb{R} \\ (x, y) &\mapsto \mathrm{pmi}(x, y) \end{aligned}$$

$$\begin{aligned} M_x \colon X &\to \mathbb{R}^Y \\ x &\mapsto M(x, -) \end{aligned}$$

$$\begin{aligned} M_y \colon Y &\to \mathbb{R}^X \\ y &\mapsto M(-, y) \end{aligned}$$

$$\begin{tikzcd} X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[dl, dashed, shift right, "M_*"'] \\ \mathbb{R}^X \arrow[ur, dashed, shift right, "M^*"'] & Y \arrow[l, "M_y"] \arrow[u, hook] \end{tikzcd}$$

$$M^* \colon \mathbb{R}^X \to \mathbb{R}^Y$$
$$M_* \colon \mathbb{R}^Y \to \mathbb{R}^X$$

From matrices to distributional operators

$M_* M^* : \mathbb{R}^X \to \mathbb{R}^X$

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[d, hook] \arrow[dl, dashed, shift right, "M_*"'] \\
\mathbb{R}^X \arrow[ur, dashed, shift right, "M^*"'] & Y \arrow[l, "M_y"]
\end{tikzcd}$$

From matrices to distributional operators

$$M_* M^* : \mathbb{R}^X \to \mathbb{R}^X$$

$$M^* M_* : \mathbb{R}^Y \to \mathbb{R}^Y$$

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[dl, dashed, shift right, "M_*"'] \\
\mathbb{R}^X \arrow[ur, dashed, shift right, "M^*"'] & Y \arrow[l, "M_y"] \arrow[u, hook]
\end{tikzcd}$$

From matrices to distributional operators

$$M_* M^* \colon \mathbb{R}^X \to \mathbb{R}^X$$
$$M^* M_* \colon \mathbb{R}^Y \to \mathbb{R}^Y$$

$$\{u_1, \ldots, u_m\} \subset \mathbb{R}^X$$
$$\{v_1, \ldots, v_n\} \subset \mathbb{R}^Y$$
$$\{\lambda_1, \ldots, \lambda_{\min(m,n)}, 0, \ldots, 0\}$$

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[d, hook] \arrow[dl, dashed, shift right, "\boldsymbol{M_*}"'] \\
\mathbb{R}^X \arrow[ur, dashed, shift right, "\boldsymbol{M^*}"'] & Y \arrow[l, "M_y"]
\end{tikzcd}$$

From matrices to distributional operators

$$M_* M^* \colon \mathbb{R}^X \to \mathbb{R}^X$$
$$M^* M_* \colon \mathbb{R}^Y \to \mathbb{R}^Y$$

$$\{u_1, \ldots, u_m\} \subset \mathbb{R}^X$$
$$\{v_1, \ldots, v_n\} \subset \mathbb{R}^Y$$
$$\{\lambda_1, \ldots, \lambda_{\min(m,n)}, 0, \ldots, 0\}$$

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[d, hook] \arrow[dl, dashed, shift right, "M_*"'] \\
\mathbb{R}^X \arrow[ur, dashed, shift right, "M^*"'] & Y \arrow[l, "M_y"]
\end{tikzcd}$$

$$M = U \Sigma V^T \qquad \begin{aligned} U &:= [u_1, \ldots, u_m] \\ V &:= [v_1, \ldots, v_n] \\ \Sigma &:= \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_r} \end{bmatrix} \end{aligned}$$

From matrices to distributional operators

$$M_* M^* \colon \mathbb{R}^X \to \mathbb{R}^X$$
$$M^* M_* \colon \mathbb{R}^Y \to \mathbb{R}^Y$$

$$\{u_1, \ldots, u_m\} \subset \mathbb{R}^X$$
$$\{v_1, \ldots, v_n\} \subset \mathbb{R}^Y$$
$$\{\lambda_1, \ldots, \lambda_{\min(m,n)}, 0, \ldots, 0\}$$

$$M_* M^* u_i = \lambda_i u_i$$
$$M^* M_* v_i = \lambda_i v_i$$

The u_i and v_i are (linear) fixed points!

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[d, hook] \arrow[dl, dashed, shift right, "M_*"'] \\
\mathbb{R}^X \arrow[ur, dashed, shift right, "M^*"'] & Y \arrow[l, "M_y"]
\end{tikzcd}$$

$$M = U \Sigma V^T \qquad U := [u_1, \ldots, u_m] \qquad V := [v_1, \ldots, v_n] \qquad \Sigma := \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_r} \end{bmatrix}$$

The operator $M_* M^*$ is a covariance matrix

Eigenvectors as Fixed Points

$$M_* M^* u = \lambda u$$

$\times$

$=$

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}$$

Eigenvectors of $M_* M^*$:

Eigenvalues of $M_* M^*$ and $M^* M_*$:

Eigenvectors of $M^* M_*$:

Plot

$$\hat{U} \times \hat{\Sigma}$$

Words

	-5	-4	-3	-2	-1	0	1	2	3	4	5
D 1	church	university	field	house	centre	...	held	used	offered	found	made
D 2	use	leave	keep	buy	meet	...	boy	club	sun	uk	hotel
D 3	show	boy	project	move	play	...	production	size	interests	activities	nature
D 4	used	expected	made	considered	allowed	...	london	europe	scotland	france	england
D 5	used	expected	food	water	england	...	during	couple	under	series	lot
D 6	perhaps	indeed	under	during	in	...	cup	bit	series	couple	lot
D 7	difficult	hard	easy	necessary	close	...	won	gave	started	saw	took
D 8	europe	scotland	england	france	lot	...	middle	want	needs	army	could
D 9	wish	tried	seem	seemed	began	...	received	established	won	published	produced
D 10	10	15	20	30	3	...	from	on	black	into	through

Embeddings as Functors Over Categories

$$X = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$Y = X \times X = \{\, \texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)} \,\}$$

$$M \colon X \times Y \to \mathbb{R}$$
$$(x, y) \mapsto \mathrm{pmi}(x, y)$$

$$M_x \colon X \to \mathbb{R}^Y$$
$$x \mapsto M(x, -)$$

$$M_y \colon Y \to \mathbb{R}^X$$
$$y \mapsto M(-, y)$$

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d] & \mathbb{R}^Y \arrow[d, hook] \arrow[dl, shift right, dashed, "M_*"'] \\
\mathbb{R}^X \arrow[ur, shift right, dashed, "M^*"'] & Y \arrow[l, "M_y"]
\end{tikzcd}$$

$$M^* \colon \mathbb{R}^X \to \mathbb{R}^Y$$
$$M_* \colon \mathbb{R}^Y \to \mathbb{R}^X$$

Embeddings as Functors Over Categories

$\mathsf{C} = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$

$\mathsf{D} = \mathsf{C} = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$

Profunctor

$$\mathcal{M} \colon \mathsf{C}^{\mathrm{op}} \times \mathsf{D} \to \mathsf{Set}$$
$$(c, d) \mapsto \mathcal{M}(c, d)$$

$$\mathcal{M}_c \colon \mathsf{C} \to \left(\mathsf{Set}^{\mathsf{D}}\right)^{\mathrm{op}}$$
$$c \mapsto \mathcal{M}(c, -)$$

$$\mathcal{M}_d \colon \mathsf{D} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$$
$$d \mapsto \mathcal{M}(-, d)$$

$$\begin{tikzcd}
\mathsf{C} \arrow[r, "\mathcal{M}_c"] \arrow[d, "Yoneda"'] & (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \\
\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \arrow[ur, dashed, shift left, "\mathcal{M}^*"] & \mathsf{D} \arrow[l, "\mathcal{M}_d"] \arrow[u, "Yoneda"'] \arrow[ul, dashed, shift left, "\mathcal{M}_*"]
\end{tikzcd}$$

$$\mathcal{M}^* \colon \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \to (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}}$$
$$\mathcal{M}_* \colon (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$$

Embeddings as Functors Over Categories

Isbell Adjunction

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

$$\mathcal{M}_*\mathcal{M}^* \colon \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$$

$$\mathcal{M}^*\mathcal{M}_* \colon (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \to (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}}$$

$$\begin{tikzcd}
\mathsf{C} \arrow[r, "\mathcal{M}_c"] \arrow[d, hook, "Yoneda"'] & (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \arrow[dl, shift right, dashed, "\mathcal{M}_*"] \\
\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \arrow[ur, shift right, dashed, "\mathcal{M}^*"'] & \mathsf{D} \arrow[l, "\mathcal{M}_d"] \arrow[u, hook, "Yoneda"']
\end{tikzcd}$$

Embeddings as Functors Over Categories

Isbell Adjunction

$$\mathcal{M}^* : \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} : \mathcal{M}_*$$

$$\mathcal{M}_*\mathcal{M}^* : \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$$

$$\mathcal{M}^*\mathcal{M}_* : (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \to (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}}$$

$$\mathrm{Fix}(\mathcal{M}_*\mathcal{M}^*) := \{f \in \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} | \mathcal{M}_*\mathcal{M}^*(f) \cong f\}$$

$$\mathrm{Fix}(\mathcal{M}^*\mathcal{M}_*) := \{g \in (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} | \mathcal{M}^*\mathcal{M}_*(g) \cong g\}$$

$$\begin{tikzcd}
\mathsf{C} \arrow[r, "\mathcal{M}_c"] \arrow[d, hook, "Yoneda"'] & (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \arrow[dl, shift left, dashed, "\mathcal{M}_*"] \\
\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \arrow[ur, shift left, dashed, "\mathcal{M}^*"] & \mathsf{D} \arrow[l, "\mathcal{M}_d"] \arrow[u, hook, "Yoneda"']
\end{tikzcd}$$

Nucleus of $\mathcal{M} = \{(f_i, g_i)\}$, such that:

$\mathcal{M}^* f_i \cong g_i$ and $\mathcal{M}_* g_i \cong f_i$

Embeddings as Functors Over Categories

Isbell Adjunction

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

$$\mathcal{M}_*\mathcal{M}^* \colon \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$$

$$\mathcal{M}^*\mathcal{M}_* \colon (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \to (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}}$$

$$\mathrm{Fix}(\mathcal{M}_*\mathcal{M}^*) := \{f \in \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} | \mathcal{M}_*\mathcal{M}^*(f) \cong f\}$$

$$\mathrm{Fix}(\mathcal{M}^*\mathcal{M}_*) := \{g \in (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} | \mathcal{M}^*\mathcal{M}_*(g) \cong g\}$$

Nucleus of $\mathcal{M} = \{(f_i, g_i)\}$, such that:

$\mathcal{M}^* f_i \cong g_i$ and $\mathcal{M}_* g_i \cong f_i$

The nucleus is a category complete and cocomplete

Embeddings as Functors Over Categories

Isbell Adjunction

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

$$\mathcal{M}_*\mathcal{M}^* \colon \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$$

$$\mathcal{M}^*\mathcal{M}_* \colon (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \to (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}}$$

$$\mathrm{Fix}(\mathcal{M}_*\mathcal{M}^*) := \{f \in \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} | \mathcal{M}_*\mathcal{M}^*(f) \cong f\}$$

$$\mathrm{Fix}(\mathcal{M}^*\mathcal{M}_*) := \{g \in (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} | \mathcal{M}^*\mathcal{M}_*(g) \cong g\}$$

Nucleus of $\mathcal{M} = \{(f_i, g_i)\}$, such that:

$$\mathcal{M}^* f_i \cong g_i \text{ and } \mathcal{M}_* g_i \cong f_i$$

The nucleus is a category complete and cocomplete

$$\begin{tikzcd}
\mathsf{C} \arrow[r, "\mathcal{M}_c"] \arrow[d, hook, "Yoneda"'] & (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \arrow[dl, shift left, dashed, "\mathcal{M}_*"] \\
\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \arrow[ur, shift left, dashed, "\mathcal{M}^*"] & \mathsf{D} \arrow[l, "\mathcal{M}_d"] \arrow[u, hook, "Yoneda"']
\end{tikzcd}$$

Categories C and D can be enriched!

E.g.:

$$\mathcal{M}^* \colon 2^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows (2^{\mathsf{D}})^{\mathrm{op}} \colon \mathcal{M}_*$$

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

Binary Fixed Points

$$M_* M^* u = \lambda u$$

$\times$

$=$

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}$$

Binary Fixed Points

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

“Eigensets”

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

0,1,2,3,4,5,6,7,8,9	1,2,4,7,8,9	b,h,m,r,v	a,e,i,o	a,e,o,u	a,i,n,s	1,3,8,9
1,2,8	h,m,w	1,8,9	d,h,t	j,o,q	c,f,t	c,t,w
a,e,o	a,e,i	h,m	2,9	a,i	w	1,9
1,8	a,e	l	t	n	c	h
2	i	e	a	o	1	9
e,n	a,l	c,t				

Partial Order Structure

Dual Partial Order

Paring of Partial Ordered Fixed Points

Enriching over $\bar{\mathbb{R}}$

Structure

$$\mathsf{C}^{\mathrm{op}} \times \mathsf{D} \to 2$$

$$\Downarrow$$

$$\mathcal{M}^* \colon 2^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows (2^{\mathsf{D}})^{\mathrm{op}} \colon \mathcal{M}_*$$

Enriching over $\bar{\mathbb{R}}$

Structure

?

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

$$\Downarrow$$

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

Enriching over $\bar{\mathbb{R}}$

Structure

?

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

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

$$\Downarrow$$

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

Enriching over $\bar{\mathbb{R}}$

Structure

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

$$\Downarrow$$

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

Enriching over $\bar{\mathbb{R}}$

Structure

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

$$\Downarrow$$

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

Theory of Computational Types

Definition (Polar/Orthogonal - Girard, 2011)

[G]iven a binary operation, noted $a, b \rightsquigarrow \langle a|b\rangle : A \times B \to C$ and a subset $P \subset C$ (the 'pole') one can define the *polar* $X^\perp \subset B$ of a subset $X \subset A$ (resp. $Y^\perp \subset A$ of a subset $Y \subset B$) by :

$$X^\perp := \{y \in B : \forall x \in X, \langle a|b\rangle \in P\}$$

$$Y^\perp := \{x \in A : \forall y \in Y, \langle a|b\rangle \in P\}$$

- The map 'polar' is decreasing: $X \subset X' \Rightarrow X'^\perp \subset X^\perp$.
- The set $\mathrm{Pol}(A) \subset \mathcal{P}(A)$ of *polar* sets, i.e., of the form $Y^\perp$, is closed under arbitrary intersections. In particular, A is polar and $X^{\perp\perp}$ is the smallest polar set containing X.
- As a consequence, $X^{\perp\perp\perp} = X^\perp$.

Theory of Computational Types

Definition (Polar/Orthogonal - Girard, 2011)

[G]iven a binary operation, noted $a, b \rightsquigarrow \langle a|b\rangle : A \times B \to C$ and a subset $P \subset C$ (the 'pole') one can define the *polar* $X^\perp \subset B$ of a subset $X \subset A$ (resp. $Y^\perp \subset A$ of a subset $Y \subset B$) by :

$$X^\perp := \{y \in B : \forall x \in X, \langle a|b\rangle \in P\}$$

$$Y^\perp := \{x \in A : \forall y \in Y, \langle a|b\rangle \in P\}$$

- The map 'polar' is decreasing: $X \subset X' \Rightarrow X'^\perp \subset X^\perp$.
- The set $\mathrm{Pol}(A) \subset \mathcal{P}(A)$ of *polar* sets, i.e., of the form $Y^\perp$, is closed under arbitrary intersections. In particular, A is polar and $X^{\perp\perp}$ is the smallest polar set containing X.
- As a consequence, $X^{\perp\perp\perp} = X^\perp$.

(Gastaldi and Pellissier, 2021)

Outline

Bring-Ins

Philosophical Perspectives

~~Historical Perspectives~~

Epistemological Perspectives

Conceptual Perspectives

Theoretical Perspectives

Takeaways

Takeaways

- Philosophical perspective
  The success of DNNs forces us to rethink the relationship between (natural) language and mathematics
- Historical perspective
  After 150 years, the radical distinction between artificial and natural languages might become ineffective
- Epistemological perspective
  Current empiricist approaches are ill-suited for this task
- Conceptual perspective
  We need to develop distributional foundations for formal objects
- Theoretical perspective
  "Geometric Structuralist Typing" may provide a solid framework to advance in this direction

Collaborations

J. Terilla (CUNY), T.-D. Bradley (SandboxAQ), L. Pellissier (Paris-Est Créteil), Th. Seiller (CNRS), S. Jarvis (CUNY)

Reference Papers

- Gastaldi, J. L. (2021).Why Can Computers Understand Natural Language? *Philosophy & Technology*, *34*(1), 149–214. https://doi.org/10.1007/s13347-020-00393-9

- 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

- Bradley, T.-D., Gastaldi, J. L., & Terilla, J. (2024).The structure of meaning in language: Parallel narratives in linear algebra and category theory. *Notices of the American Mathematical Society*. https://api.semanticscholar.org/CorpusID:263613625

References I

Alemi, A. A., Chollet, F., Een, N., Irving, G., Szegedy, C., & Urban, J. (2016).Deepmath - deep sequence models for premise selection. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2243–2251.

Bansal, K., Loos, S. M., Rabe, M. N., Szegedy, C., & Wilcox, S. (2019).Holist: An environment for machine learning of higher-order theorem proving (extended version). *CoRR*, *abs/1904.03241*. http://arxiv.org/abs/1904.03241

Belrose, N., Schneider-Joseph, D., Ravfogel, S., Cotterell, R., Raff, E., & Biderman, S. (2024).Leace: Perfect linear concept erasure in closed form. *Proceedings of the 37th International Conference on Neural Information Processing Systems*.

Blechschmidt, J., & Ernst, O. G. (2021).Three ways to solve partial differential equations with neural networks — a review. *GAMM-Mitteilungen*, *44*(2), e202100006. https://doi.org/https://doi.org/10.1002/gamm.202100006

Bradley, T.-D., Gastaldi, J. L., & Terilla, J. (2024).The structure of meaning in language: Parallel narratives in linear algebra and category theory. *Notices of the American Mathematical Society*. https://api.semanticscholar.org/CorpusID:263613625

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (pp. 1877–1901, Vol. 33). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Charton, F. (2021).Linear algebra with transformers. *CoRR*, *abs/2112.01898*. https://arxiv.org/abs/2112.01898

d'Ascoli, S., Kamienny, P., Lample, G., & Charton, F. (2022).Deep symbolic regression for recurrent sequences. *CoRR*, *abs/2201.04600*.

References II

Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., Tanburn, R., Battaglia, P., Blundell, C., Juhász, A., Lackenby, M., Williamson, G., Hassabis, D., & Kohli, P. (2021).Advancing mathematics by guiding human intuition with AI. *Nature*, *600*(7887), 70–74. https://doi.org/10.1038/s41586-021-04086-x

Firth, J. R. (1935).The technique of semantics. *Transactions of the Philological Society*, *34*(1), 36–73. https://doi.org/10.1111/j.1467-968X.1935.tb01254.x

Gao, L., Jiang, Z., Yin, Y., Yuan, K., Yan, Z., & Tang, Z. (2017).Preliminary exploration of formula embedding for mathematical information retrieval: Can mathematical formulae be embedded like a natural language? *CoRR*, *abs/1707.05154*. http://arxiv.org/abs/1707.05154

Gastaldi, J. L. (2021).Why Can Computers Understand Natural Language? *Philosophy & Technology*, *34*(1), 149–214. https://doi.org/10.1007/s13347-020-00393-9

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

Girard, J.-Y. (2011, September). *The blind spot*. European Mathematical Society.

Goldberg, Y., & Levy, O. (2014).Word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *CoRR*, *abs/1402.3722*.

Harris, Z. (1960). *Structural linguistics*. University of Chicago Press.

Krstovski, K., & Blei, D. M. (2018). Equation embeddings.

Lample, G., & Charton, F. (2019). Deep learning for symbolic mathematics.

Levy, O., & Goldberg, Y. (2014).Neural word embedding as implicit matrix factorization. *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, 2177–2185.

References III

Lewkowycz, A., Andreassen, A., Dohan, D., Dyer, E., Michalewski, H., Ramasesh, V., Slone, A., Anil, C., Schlag, I., Gutman-Solo, T., Wu, Y., Neyshabur, B., Gur-Ari, G., & Misra, V. (2022). Solving quantitative reasoning problems with language models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems* (pp. 3843–3857, Vol. 35). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/18abbeef8cfe9203fdf9053c9c4fe191-Paper-Conference.pdf

Li, Z., Kovachki, N. B., Azizzadenesheli, K., liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021).Fourier neural operator for parametric partial differential equations. *International Conference on Learning Representations*. https://openreview.net/forum?id=c8P9NQVtmnO

Mansouri, B., Rohatgi, S., Oard, D. W., Wu, J., Giles, C. L., & Zanibbi, R. (2019).Tangent-cft: An embedding model for mathematical formulas. *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, 11–18. https://doi.org/10.1145/3341981.3344235

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., Le, Q., & Strohmann, T. (2013). *Learning representations of text using neural networks. NIPS deep learning workshop 2013 slides*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013).Distributed representations of words and phrases and their compositionality. *CoRR*, *abs/1310.4546*.

Peng, S., Yuan, K., Gao, L., & Tang, Z. (2021).Mathbert: A pre-trained model for mathematical formula understanding. *CoRR*, *abs/2105.00377*. https://arxiv.org/abs/2105.00377

Polu, S., & Sutskever, I. (2020).Generative language modeling for automated theorem proving. *CoRR*, *abs/2009.03393*. https://arxiv.org/abs/2009.03393

Purgał, S., Parsert, J., & Kaliszyk, C. (2021).A study of continuous vector representations for theorem proving. *Journal of Logic and Computation*, *31*(8), 2057–2083. https://doi.org/10.1093/logcom/exab006

References IV

Ryskina, M., & Knight, K. (2021).Learning mathematical properties of integers. *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, 389–395. https://doi.org/10.18653/v1/2021.blackboxnlp-1.30

Saussure, F. d. (1959). *Course in general linguistics* [Translated by Wade Baskin]. McGraw-Hill.

Sennrich, R., Haddow, B., & Birch, A. (2016).Neural machine translation of rare words with subword units. *Proceedings of the 54th Annual Meeting of the ACL*, 1715–1725.

Shen, J. T., Yamashita, M., Prihar, E., Heffernan, N. T., Wu, X., & Lee, D. (2021).Mathbert: A pre-trained language model for general NLP tasks in mathematics education. *CoRR*, *abs/2106.07340*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wagner, A. Z. (2021). Constructions in combinatorics via neural networks.

Wu, Y., Jiang, A. Q., Li, W., Rabe, M. N., Staats, C. E., Jamnik, M., & Szegedy, C. (2022). Autoformalization with large language models. In A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho (Eds.), *Advances in neural information processing systems*. https://openreview.net/forum?id=IUikebJ1Bf0

MALINCA
Institut Henri Poincaré
Paris, France

Language and Mathematics in the Time of AI

Philosophical and Theoretical Perspectives

Juan Luis Gastaldi

www.giannigastaldi.com

ETH zürich

October 1, 2025

Trees of Knowledge

F. Bacon (1605)

E. Chambers (1728)

Didérot & D'Alembert (1751-1772)

Trees of Knowledge (detail)

The Birth of Formal Knowledge

THE principles of Method, developed in the preceding Essay, will, it is hoped, render perfectly intelligible the Plan of our whole work, which is comprehended under Four Divisions as follows:

FIRST DIVISION.

PURE SCIENCES. — 2 VOLS. —

FORMAL.
- Universal Grammar and Philology: or the forms of Languages.
- Logic, particular and universal: or the forms of Conceptions and their combinations.
- Mathematics: (Geometry, Arithmetic, Algebra, &c.) or the forms and constructions of Figure and Number.

REAL.
- Metaphysics: or the universal principles and conditions of Experience, having for its object the Reality of our speculative knowledge in general.
- Morals: or the principles and conditions of the coincidence of the individual with the universal reason, having for its object the Reality of our practical knowledge: (hence, in a lower stage, Politics and Human Law.)
- Theology: or the union of both in their application to GOD, the Supreme Reality.

SECOND DIVISION.

MIXED AND APPLIED SCIENCES. — 6 VOLS. —

MIXED.
- Mechanics.
- Hydrostatics.
- Pneumatics.
- Optics.
- Astronomy.

APPLIED.
- I. EXPERIMENTAL PHILOSOPHY.
  - Magnetism.
  - Electricity, including Galvanism.
  - Chemistry.
  - Light.
  - Heat.
  - Sound.
  - Meteorology.
- II. FINE ARTS.
  - Poetry, introduced by Etymology.
  - Painting.
  - Music.
  - Sculpture.
  - Architecture.
- III. THE USEFUL ARTS.
  - Agriculture, introduced by Political Economy.
  - Manufactures.
  - Medicine, or Physiology in its widest sense.
- IV. NATURAL HISTORY.
  - Mineralogy.
  - Lithology, Crystallography, Geology.
  - Botany.
  - Insectology—Physiology.
  - Animals—Zoology.
- V. APPLICATION OF NATURAL HISTORY.
  - Agriculture.
  - Manufactures.
  - Medical Medica.
  - Pharmacy.
  - Medicine.

THIRD DIVISION.

BIOGRAPHICAL AND HISTORICAL. — 2 VOLS. —

Biographic-chronologically arranged, interspersed with introductory Chapters of National History, Political Geography and Chronology, and accompanied with correspondent Maps and Charts.

FOURTH DIVISION.

MISCELLANEOUS AND LEXICOGRAPHICAL. — 8 VOLS. —

Alphabetical, Miscellaneous, and Supplementary—containing a GAZETTEER, or complete Vocabulary of Geography; and a Philosophical and Etymological LEXICON of the English Language, or the History of English Words—the citations arranged according to the Age of the Words from which they are selected, yet with every attention to the convenience of the Reader, and the Etymologies given in English with the higher ends of a clear insight into the original and acquired meaning of every word.

The INDEX—Being a digested and complete Body of Reference in the whole Work; in which the known English name, as well as the scientific name, of every subject of Natural History, will be found in its alphabetical place.

PURE SCIENCES. — 2 VOLS. —

FIRST DIVISION.

FORMAL.
- Universal Grammar and Philology: or the forms of Languages.
- Logic, particular and universal: or the forms of Conceptions and their combinations.
- Mathematics: (Geometry, Arithmetic, Algebra, &c.) or the forms and constructions of Figure and Number.

REAL.
- Metaphysics: or the universal principles and conditions of Experience, having for its object the Reality of our speculative knowledge in general.
- Morals: or the principles and conditions of the coincidence of the individual will with the universal reason, having for its object the Reality of our practical knowledge: (hence, in a lower stage, Politics and Human Law.)
- Theology: or the union of both in their application to GOD, the Supreme Reality.

S. T. Coleridge, Encyclopaedia Metropolitana, 1818.

Empirical evaluation

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

Empirical evaluation

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

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

Empirical evaluation

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

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

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

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

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

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

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

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

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

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

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

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

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

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:^4mÓØ¥´è—+ÌsÖ,$+gï„B™÷o–#ïŸê Ûv –gÓÿ/ëiijO‡Œfi • J1«€ø¸Ï hãêt‡æY$^6 FïW»RÙKg¢¨.λ‡d¯… D2÷¸ò ˚ xêÈy. Ó”¢b Bé£NÈ1Ê‡/Û9Ñµ–/JYÇõË9ÿÀÈ.λÁÍ À^öÇ,»fq∞±î˘B5Ì>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‡.:^4mÓØ¥´è—+ÌsÖ,$+gï„B™÷o–#ïŸê Ûv–gÓÿ /ëiijO‡Œfi • J1«€ø¸Ï hãêt‡æY$^6 FïW»RÙKg¢¨ÁÍ À^öÇ,»fq∞±î˘B5Ì>O˜g™“6 Ωe“æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨(‡d¯… D2÷¸ò ˚ xêÈy. Ó”¢bBé£NÈ1Ê‡/Û9Ñµ–/ JYÇõË9ÿÀÈÁÍ À^öÇ,»fq∞±î˘B5Ì>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‡)(ËÎÛeí4WµÍ }w„$Ω“K5èÃ¶%3[m ’˘BÃfifÒ; óJçCËïõŸØcB,ñ$Ãå}ÓÅØ3¡ ˘?ö¯oŒ@ fl8˘R CÆo˘* &<Ÿ ¬o12À‰ àÓÜ#ï ”,ú¨ «ô¸„∞Îâä“øÃd|^Ñ’Ê yØ;^W ›”wò[}\»ÖÈúw’6<ù˘=ãÕ ¯Ì‘Dz?2±|é‘ 3Â/rxµ∞µ$ÀëÃ*lf˜îû’+ÍVìyªGæßägô/,uÑ)

Formal Content

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

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

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

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

`{cat, dog, spider, gavagai}`

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

`"The gavagai is on the mat"`

Informational Content

The content related to the non-uniform distribution of units within those substitutability classes

```
{cat:0.059%,
dog:0.012%,
spider:0.009%,
gavagai:0.000%}
```

Example: Characters in Wikipedia

$$W = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$C = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

$$\begin{aligned} M_{wc} &= \mathrm{pmi}(w, c) \\ &= \log \frac{p(w, c)}{p(w)p(c)} \end{aligned}$$

SVD of Wikipedia Character PMI Matrix

U

Σ

V^T

Truncate

$$U \times \Sigma$$

Truncate

$$\hat{U} \times \hat{\Sigma}$$

Plot

$$\hat{U} \times \hat{\Sigma}$$