Bridging Histories
Literacy, computer science and mathematics in the 20th and 21st centuries

Isaac Newton Institute for Mathematical Sciences
Cambridge, UK

How to Do Maths With Words

The Role of Language in Neural Machine Learning Applications to Mathematics

Juan Luis Gastaldi

ETH zürich

March 25, 2025

Outline

Intro: Machine Learning, Mathematics, and Language

Historical Perspectives

Epistemological Perspectives

Theoretical Perspectives

Conclusions

Reference Papers

- Gastaldi, J. L. (2024). How to Do Maths with Words: Neural Machine Learning Applications to Mathematics and Their Philosophical Significance. In B. Sriraman (Ed.), *Handbook of the history and philosophy of mathematical practice* (pp. 3191–3226). Springer International Publishing. https://doi.org/10.1007/978-3-031-40846-5_142

- Gastaldi, J. L., & Pellissier, L. (2021).The calculus of language: Explicit representation of emergent linguistic structure through type-theoretical paradigms. *Interdisciplinary Science Reviews*. 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

- Gastaldi, J. L. (Forthcoming 2024c).Content from Expressions. The Place of Textuality in Deep Learning Approaches to Mathematics. *Synthese (under review)*

Outline

Intro: Machine Learning, Mathematics, and Language

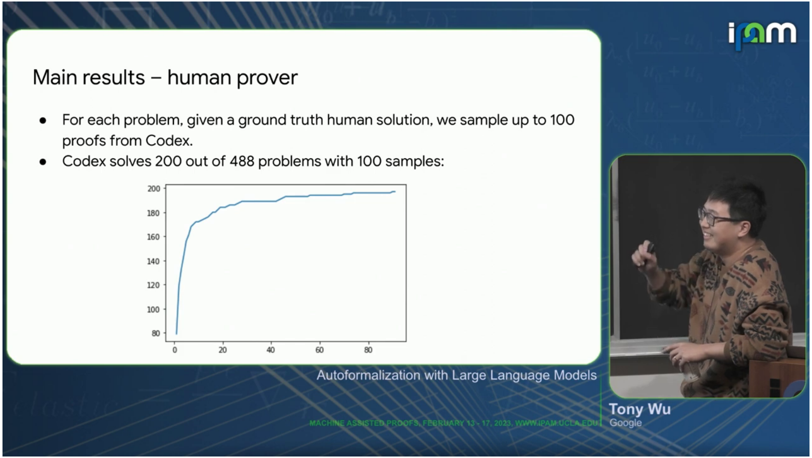
Historical Perspectives

Epistemological Perspectives

Theoretical Perspectives

Conclusions

Tony Wu at the IPAM



Main results – human prover

- For each problem, given a ground truth human solution, we sample up to 100 proofs from Codex.
- Codex solves 200 out of 488 problems with 100 samples:

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/

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.

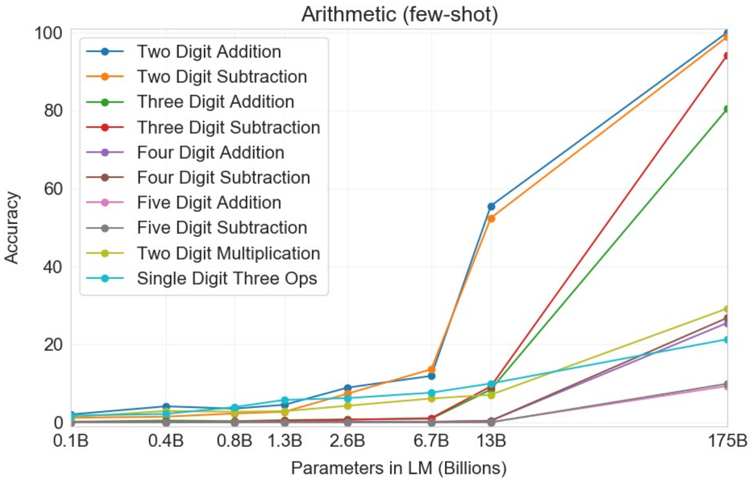
What’s So Funny?

- 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?

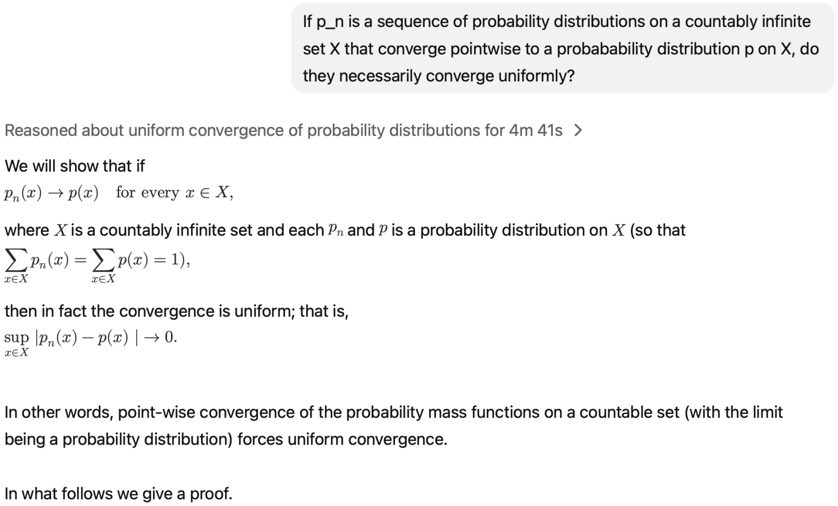
- Proof assistants and automated theorem provers were supposed to prevent the frailty of human proving practices instead of reproducing them and relying on them as a gold standard.
- Mathematical literacy was supposed to be a condition to write and read scientific papers and mathematical expressions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

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

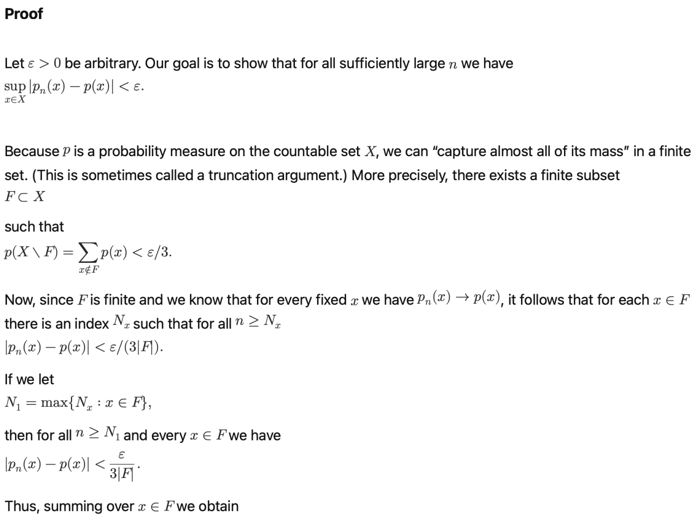
then in fact the convergence is uniform; that is,

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

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

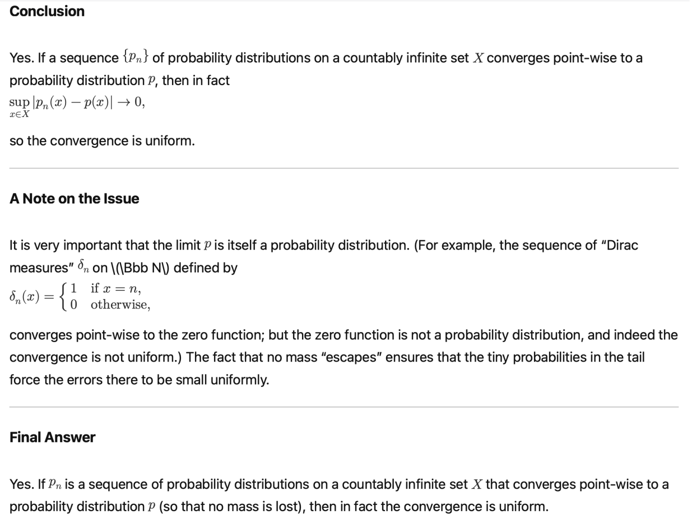
$$N_1 = \max\{N_x : x \in F\},$$

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

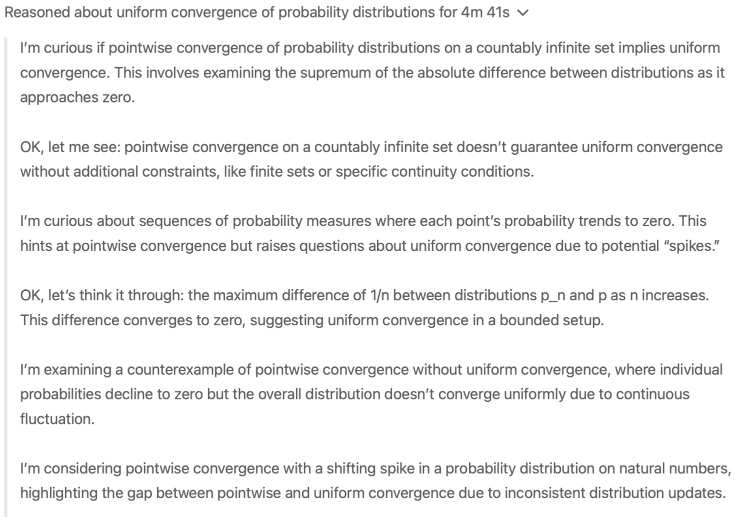
$$\delta_n(x) = \begin{cases} 1 & \text{if } x = n, \\ 0 & \text{otherwise,} \end{cases}$$

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

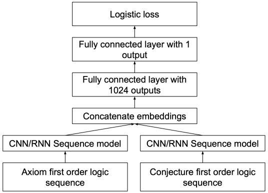


Reasoned about uniform convergence of probability distributions for 4m 41s

> 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



Logistic loss

Fully connected layer with 1 output

Fully connected layer with 1024 outputs

Concatenate embeddings

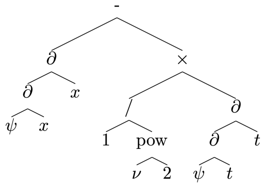
CNN/RNN Sequence model

CNN/RNN Sequence model

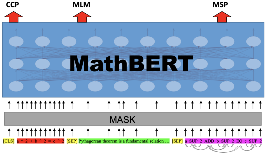
Axiom first order logic sequence

Conjecture first order logic sequence

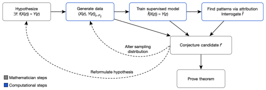
(Alemi et al., 2016)



(Lample and Charton, 2019)



(Peng et al., 2021)



Hypothesize
$\exists f: f(X(z)) \approx Y(z)$

Generate data
$(X(z), Y(z))_{z \sim P_z}$

Train supervised model
$\hat{f}(X(z)) \approx Y(z)$

Find patterns via attribution
Interrogate $\hat{f}$

Conjecture candidate f'

Prove theorem

Alter sampling distribution

Reformulate hypothesis

▨ Mathematician steps

🟦 Computational steps

(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

Intro: Machine Learning, Mathematics, and Language

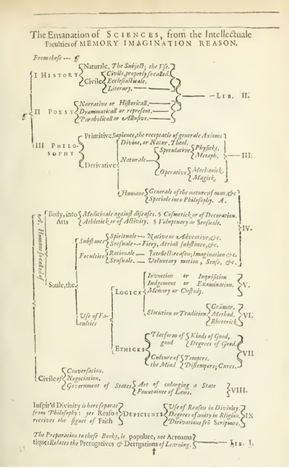
Historical Perspectives

Epistemological Perspectives

Theoretical Perspectives

Conclusions

Trees of Knowledge



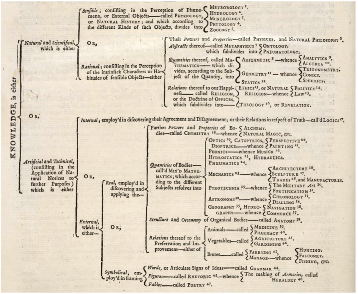
The Emanation of SCIENCES, from the Intellectuale
Faculties of MEMORY IMAGINATION REASON.

From theſe --- ¶
I HISTORY
Naturale. *The Subject; the Uſe.*
Civile
Civile, properly ſo called.
Eccleſiaſticale.
Literary.
— LIB. II.
II POESY
Narrative or Hiſtoricall.
Drammaticall or repreſent.
Parabolicall or Alluſive.
III PHILOSOPHY
Primitive *Sapience, the receptacle of generale Axioms*
Derivative
Divine, or Natur. Theol.
Naturale
Speculative
Phyſicks.
Metaph.
Operative
Mechanick.
Magick.
— III.
Humane
Generale of the nature of man. &c
Speciale into Philoſophy. A.
A. Humane ſo called of
Body, into Arts
Medicinale against diſeaſes. § *Coſmetick or of Decoration.*
Athletick or of Activity. § *Voluptuary or* Senſuale.
IV.
Soule, the
Subſtance
Spiritnale --- *Native or Adventive, &c.*
Senſuale -.- *Fiery, Aëriall ſubſtance, &c.*
Faculties
Rationale — *Intellect; reaſon; Imagination &c.*
Senſuale. — *Voluntary motion, Senſe, &c.*
Uſe of Faculties
LOGICK
Invention or Inquiſition
Judgement or Examination.
Memory or Cuſtody.
V.
Elocution or Tradition
Grãmar.
Method.
Rhetorick
VI.
ETHICKS
Platform of good
Kinds of Good,
Degrees of Good.
Culture of the Mind
Tempers.
Diſtempers; Cures.
VII
Civile of
Converſation.
Negociation.
Government of States
Art of enlarging a State
Fountainee of Laws.
VIII.

Inſpir'd Divinity *is here ſeparat from Philoſophy: yet* Reaſon *receives the ſignet of* Faith
DEFICIENTS
Uſe of Reaſon in Divinity.
Degrees of unity in Rligion.
Dirivations frõ Scripture.
IX

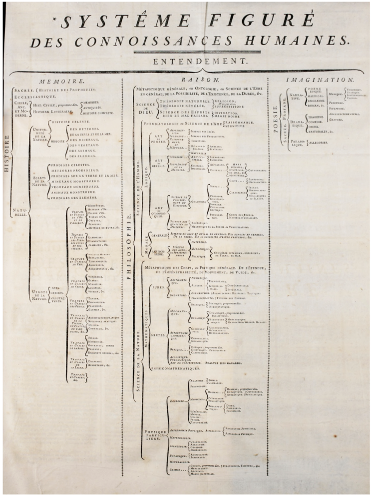
The Preparation to theſe Books, is populare, *not* Acroamatique: *Relates the* Prerogatives *&* Derrgations *of Learning.* — LIB. I.
†

F. Bacon (1605)



KNOWLEDGE, is either
Natural and *Scientifical*, which is either
Senſible; conſiſting in the Perception of Phænomena, or External Objects——called PHYSIOLOGY, or NATURAL HISTORY; and which according to the different Kinds of ſuch Objects, divides into
METEOROLOGY [1].
HYDROLOGY [2].
MINEROLOGY [3].
PHYTOLOGY [4].
ZOOLOGY [5].
OR,
Rational; conſiſting in the Perception of the intrinſick Characters or Habitudes of ſenſible Objects—either
Their *Powers* and *Properties*—called PHYSICKS, and NATURAL PHILOSOPHY [6].
Abſtracts thereof—called METAPHYSICS [7] which ſubdivides into
ONTOLOGY.
PNEUMATOLOGY.
Quantities thereof, called MATHEMATICS—— which divides, according to the Subject of the Quantity, into
ARITHMETIC [8] —whence
ANALYTICS [9].
ALGEBRA [10].
GEOMETRY [11] — whence
TRIGONOMETRY.
CONICS.
SPHERICS.
STATICS [12].
Relations thereof to our Happineſs—— called RELIGION, or the Doctrine of OFFICES, which ſubdivides into——
ETHICS [13], or NATURAL RELIGION—whence
POLITICS [14].
LAW [15].
THEOLOGY [16], or REVELATION.
OR,
Artificial and *Technical*, (conſiſting in the Application of Natural Notices to further Purpoſes) which is either
Internal; employ'd in diſcovering their Agreement and Diſagreement; or their Relations in reſpect of Truth—call'd LOGICS [17].
OR,
External, which is either---
Real, employ'd in diſcovering and applying the---
Further *Powers* and *Properties* of Bodies—called CHYMISTRY [18]—whence
ALCHEMY.
NATURAL MAGIC, &c.
Quantities of Bodies—— call'd MIX'D MATHEMATICS; which according to the different Subjects reſolves into
OPTICS [19], CATOPTRICS, DIOPTRICS——whence
PERSPECTIVE [20].
PAINTING [21].
PHONICS——whence MUSICK [22].
HYDROSTATICS [23], HYDRAULICS.
PNEUMATICS [24].
MECHANICS [25]——whence
ARCHITECTURE [26].
SCULPTURE [27].
TRADES [28], and MANUFACTURES.
PYROTECHNIA [29]—whence
The MILITARY *Art* [30].
FORTIFICATION [31].
ASTRONOMY [32]——whence
CHRONOLOGY [33].
DIALLING [34].
GEOGRAPHY [35], HYDROGRAPHY——whence
NAVIGATION [36].
COMMERCE [37].
Structure and *Oeconomy* of Organical Bodies——called ANATOMY [38].
Relations thereof to the Preſervation and Improvement—either of
Animals——called
MEDICINE [39].
PHARMACY [40].
Vegetables—called
AGRICULTURE [41].
GARDENING [42].
Brutes——called
FARRYING [43],
MANAGE——whence
HUNTING.
FALCONRY.
FISHING, &c.
OR;
Symbolical, employ'd in framing
Words, or Articulate Signs of Ideas——called GRAMMAR [44].
Figures——called RHETORIC [45]—whence The making of *Armories*, called HERALDRY [46].
Fables——called POETRY [47].

E. Chambers (1728)



*SYSTÊME FIGURÉ
DES CONNOISSANCES HUMAINES.

ENTENDEMENT.

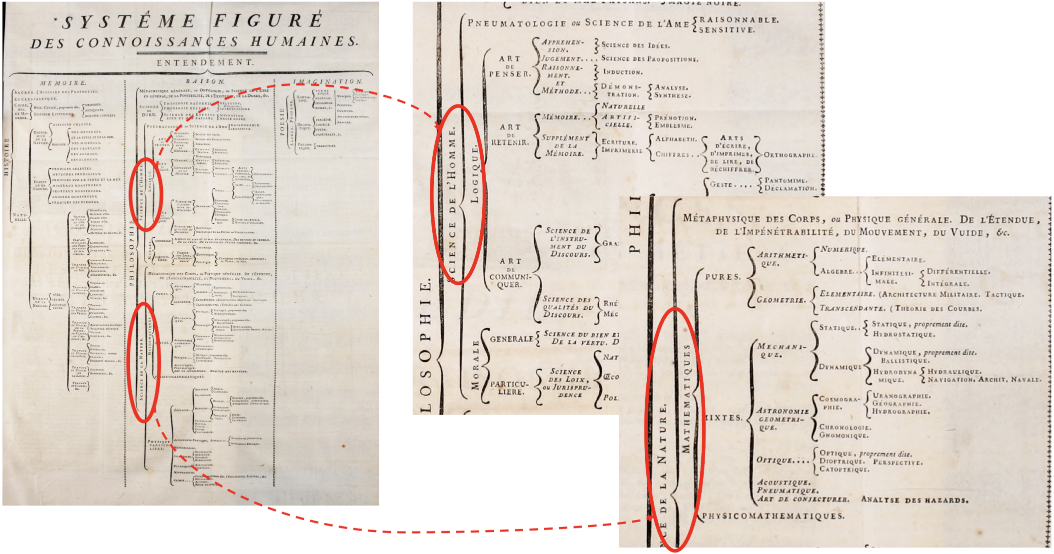
MEMOIRE.
HISTOIRE

RAISON.
PHILOSOPHIE

IMAGINATION.
POESIE

Didérot & D'Alembert (1751-1772)

Trees of Knowledge (detail)



*SYSTÊME FIGURÉ
DES CONNOISSANCES HUMAINES.

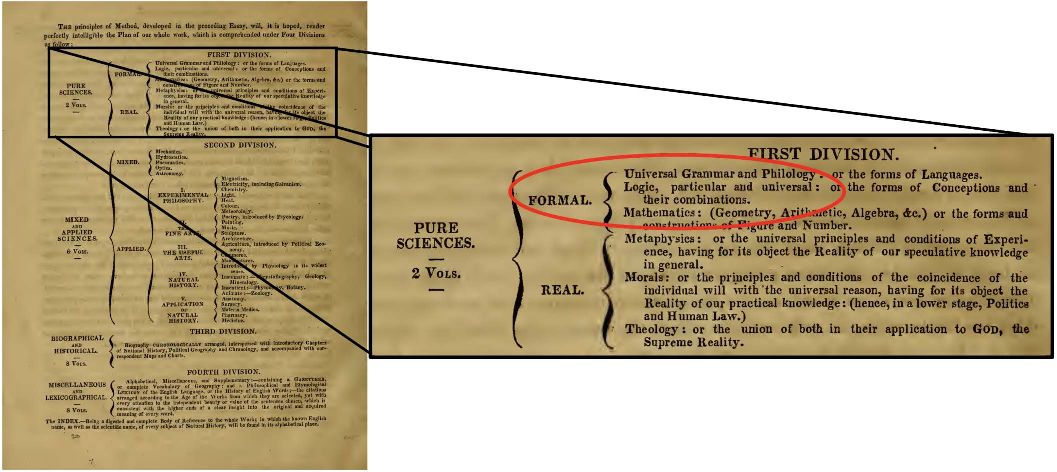
ENTENDEMENT.

MEMOIRE. RAISON. IMAGINATION.

HISTOIRE. PHILOSOPHIE. POESIE.

- PHILOSOPHIE.
  - SCIENCE DE L'HOMME.
    - PNEUMATOLOGIE *ou* SCIENCE DE L'AME
      - RAISONNABLE.
      - SENSITIVE.
    - LOGIQUE.
      - ART DE PENSER.
        - *Apprehension.*
          - Science des Idées.
        - *Jugement.* . . . Science des Propositions.
        - *Raisonnement.*
          - Induction.
        - *et Méthode.* . .
          - Démonstration.
            - Analyse.
            - Synthese.
      - ART DE RETENIR.
        - *Mémoire.* . .
          - *Naturelle*
          - *Artificielle.*
            - Prénotion.
            - Emblesme.
        - *Supplément de la Mémoire.*
          - Ecriture. Imprimerie
            - Alphabeth.
            - Chiffres . .
              - Arts d'écrire, d'imprimer, de lire, de déchiffrer.
                - Orthographe.
      - ART DE COMMUNIQUER.
        - Geste . . .
          - Pantomime.
          - Déclamation.
        - *Science de l'instrument du Discours.*
          - Gra
        - *Science des qualités du Discours.*
          - Rhé
          - Méc
    - MORALE
      - GENERALE
        - *Science du bien et de la vertu.* D
      - PARTICULIERE.
        - *Science des Loix, ou Jurisprudence*
          - Nat
          - Œco
          - Pol
  - SCIENCE DE LA NATURE.
    - MÉTAPHYSIQUE DES CORPS, *ou* PHYSIQUE GÉNÉRALE. DE L'ÉTENDUE, DE L'IMPÉNÉTRABILITÉ, DU MOUVEMENT, DU VUIDE, *&c.*
    - MATHEMATIQUES
      - PURES.
        - *Arithmetique.*
          - *Numerique.*
          - Algebre. . .
            - Elementaire.
            - Infinitesimale.
              - Différentielle.
              - Intégrale.
        - *Geometrie.*
          - *Elementaire.* (Architecture Militaire. Tactique.
          - *Transcendante.* (Theorie des Courbes.
      - MIXTES.
        - *Mechanique.*
          - Statique. .
            - Statique, *proprement dite.*
            - Hydrostatique.
          - Dynamique
            - Dynamique, *proprement dite.*
              - Ballistique.
            - Hydrodynamique.
              - Hydraulique.
              - Navigation, Archit. Navale.
        - *Astronomie geometrique.*
          - Cosmographie.
            - Uranographie.
            - Géographie.
            - Hydrographie.
          - Chronologie.
          - Gnomonique.
        - *Optique.* . . .
          - Optique, *proprement dite.*
          - Dioptrique. Perspective.
          - Catoptrique.
        - *Acoustique.*
        - *Pneumatique.*
        - *Art de conjecturer.* Analyse des hazards.
      - PHYSICOMATHEMATIQUES.

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 follow:

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 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.

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.
    - Colour.
    - Meteorology.
  - II. THE FINE ARTS.
    - Poetry, introduced by Psycology.
    - Painting.
    - Music.
    - Sculpture.
    - Architecture.
  - III. THE USEFUL ARTS.
    - Agriculture, introduced by Political Economy.
    - Commerce.
    - Manufactures.
  - IV. NATURAL HISTORY.
    - Introduced by Physiology in its widest sense.
    - Inanimate:—Chrystallography, Geology, Mineralogy.
    - Insentient:—Phytonomy, Botany.
    - Animate:—Zoology.
  - V. APPLICATION OF NATURAL HISTORY.
    - Anatomy.
    - Surgery.
    - Materia Medica.
    - Pharmacy.
    - Medicine.

THIRD DIVISION.

BIOGRAPHICAL AND HISTORICAL. — 8 VOLS.

Biography 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 Works from which they are selected, yet with every attention to the independent beauty or value of the sentences chosen, which is consistent 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 to 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.

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 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.

Outline

Intro: Machine Learning, Mathematics, and Language

Historical Perspectives

Epistemological Perspectives

Theoretical Perspectives

Conclusions

Stochastic Parrots vs. AI Consciousness

📉🦜

Language models are not like us,
therefore they do not and can not have any relation to meaning.

🤖🧠

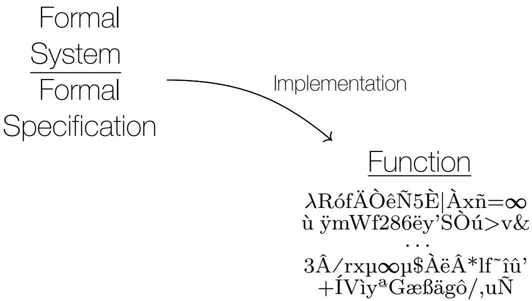
Language models have a relation to meaning,
therefore they are like us.

Making it Explicit

Function

λRófÄÒêÑ5È|Àxñ=∞
ù ÿmWf286ëy’SÒú>v&
. . .
3Â/rxµ∞µ$ÀëÂ*lf˜îû’
+ÍVìyªGæßägô/,uÑ

Making it Explicit



Formal System

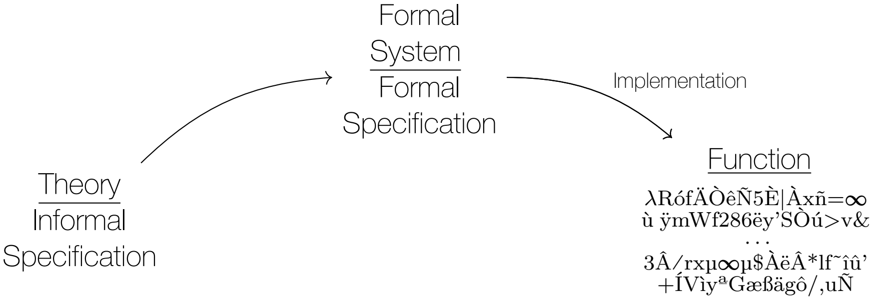
Formal Specification

Implementation

Function

λRófÄÒêÑ5È|Àxñ=∞
ù ÿmWf286ëy’SÒú>v&
. . .
3Â/rxµ∞µ$ÀëÂ*lf˜îû’
+ÍVìyªGæßägô/,uÑ

Making it Explicit



Theory

Informal
Specification

Formal
System

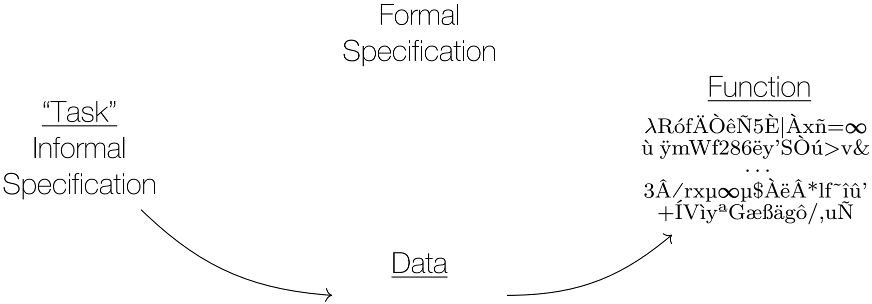
Formal
Specification

Implementation

Function

λRófÄÒêÑ5È|Ą̀xñ=∞
ù ÿmWf286ëy’SÒú>v&
. . .
3Â/rxµ∞µ$ÀëÂ*lf˜îû’
+ÍVìyªGæßägô/,uÑ

Making it Explicit



Formal
Specification

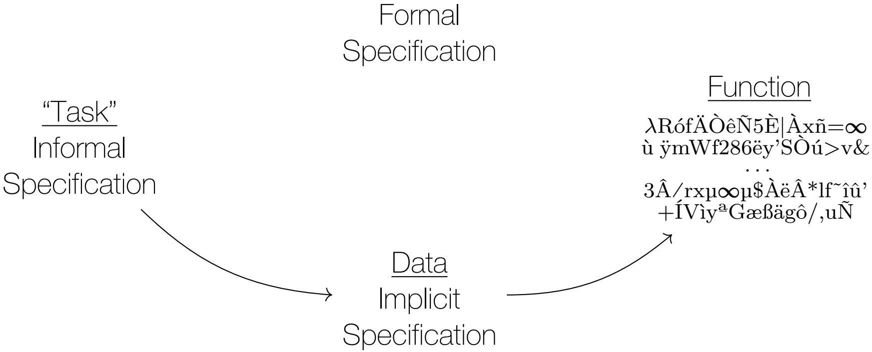
“Task”
Informal
Specification

Data

Function

λRófÄÒêÑ5È|Àxñ=∞
ù ÿmWf286ëy’SÒú>v&
. . .
3Â/rxµ∞µ$ÀëÂ*lf˜îû’
+ÍVìyªGæßägô/,uÑ

Making it Explicit



Formal
Specification

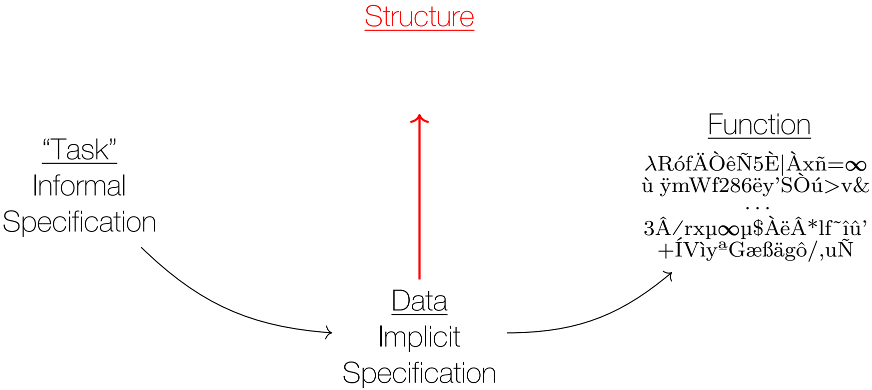
“Task”
Informal
Specification

Data
Implicit
Specification

Function

λRófÄÒêÑ5È|Àxñ=∞
ù ÿmWf286ëy’SÒú>v&
. . .
3Â/rxμ∞μ$ÀëÂ*lf˜îû’
+ÍVìyªGæßägô/,uÑ

Making it Explicit



Structure

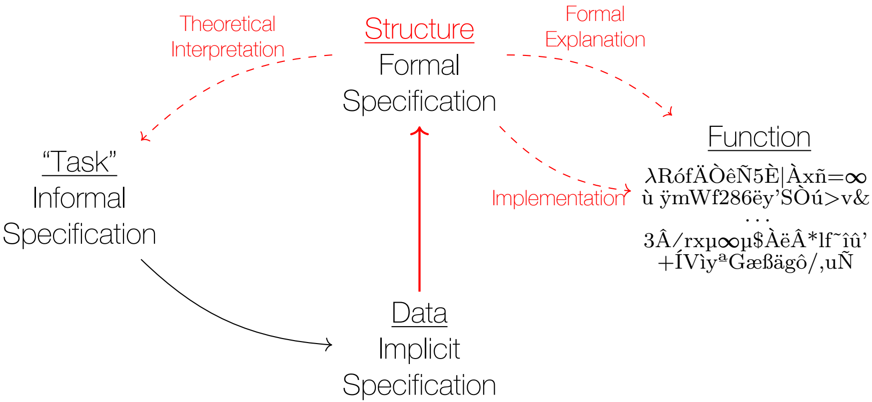
“Task”
Informal
Specification

Data
Implicit
Specification

Function

λRófÄÒêÑ5È|Àxñ=∞
ù ÿmWf286ëy’SÒú>v&
. . .
3Â/rxµ∞µ$ÀëÂ*lf˜îû’
+ÍVìyªGæßägô/,uÑ

Making it Explicit



Theoretical Interpretation

Structure
Formal
Specification

Formal Explanation

"Task"
Informal
Specification

Implementation

Function
λRófÄÒêÑ5È|Àxñ=∞
ù ÿmWf286ëy'SÒú>v&
. . .
3Â/rxµ∞µ$ÀëÂ*lf˜îû'
+ÍVìyªGæßägô/,uÑ

Data
Implicit
Specification

Outline

Intro: Machine Learning, Mathematics, and Language

Historical Perspectives

Epistemological Perspectives

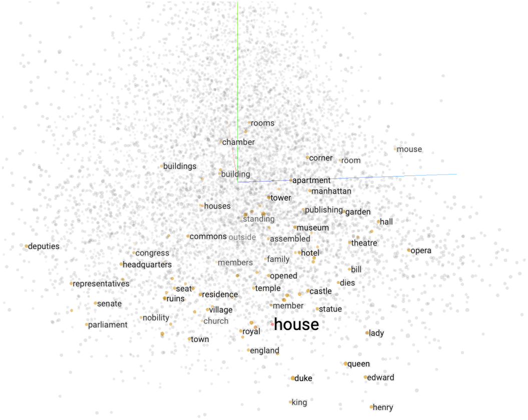
Theoretical Perspectives

Conclusions

Distributionalism and Word Embeddings

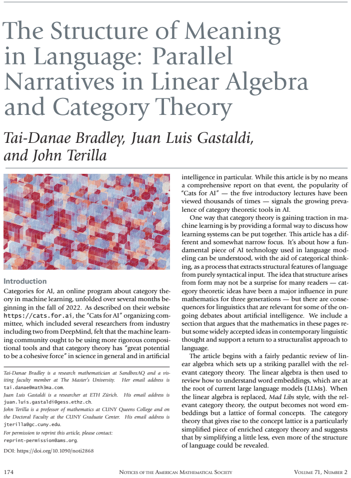
- Distributional Hypothesis (Harris, 1960; Saussure, 1959)
  - “You shall know a word by the company it keeps!” (Firth, 1935)
  - The content of a linguistic unit is determined by its distribution over a corpus (i.e., the other units appearing in its context)

- Computational version: Word Embeddings



(https://projector.tensorflow.org)

The Structure of Meaning in Language



The Structure of Meaning in Language: Parallel Narratives in Linear Algebra and Category Theory

Tai-Danae Bradley, Juan Luis Gastaldi, and John Terilla

Introduction

Categories for AI, an online program about category theory in machine learning, unfolded over several months beginning in the fall of 2022. As described on their website `https://cats.for.ai`, the "Cats for AI" organizing committee, which included several researchers from industry including two from DeepMind, felt that the machine learning community ought to be using more rigorous compositional tools and that category theory has "great potential to be a cohesive force" in science in general and in artificial intelligence in particular. While this article is by no means a comprehensive report on that event, the popularity of "Cats for AI" — the five introductory lectures have been viewed thousands of times — signals the growing prevalence of category theoretic tools in AI.

One way that category theory is gaining traction in machine learning is by providing a formal way to discuss how learning systems can be put together. This article has a different and somewhat narrow focus. It's about how a fundamental piece of AI technology used in language modeling can be understood, with the aid of categorical thinking, as a process that extracts structural features of language from purely syntactical input. The idea that structure arises from form may not be a surprise for many readers — category theoretic ideas have been a major influence in pure mathematics for three generations — but there are consequences for linguistics that are relevant for some of the ongoing debates about artificial intelligence. We include a section that argues that the mathematics in these pages rebut some widely accepted ideas in contemporary linguistic thought and support a return to a structuralist approach to language.

The article begins with a fairly pedantic review of linear algebra which sets up a striking parallel with the relevant category theory. The linear algebra is then used to review how to understand word embeddings, which are at the root of current large language models (LLMs). When the linear algebra is replaced, *Mad Libs* style, with the relevant category theory, the output becomes not word embeddings but a lattice of formal concepts. The category theory that gives rise to the concept lattice is a particularly simplified piece of enriched category theory and suggests that by simplifying a little less, even more of the structure of language could be revealed.

Tai-Danae Bradley is a research mathematician at SandboxAQ and a visiting faculty member at The Master's University. Her email address is `tai.danae@math3ma.com`.

Juan Luis Gastaldi is a researcher at ETH Zürich. His email address is `juan.luis.gastaldi@gess.ethz.ch`.

John Terilla is a professor of mathematics at CUNY Queens College and on the Doctoral Faculty at the CUNY Graduate Center. His email address is `jterilla@gc.cuny.edu`.

For permission to reprint this article, please contact: `reprint-permission@ams.org`.

DOI: https://doi.org/10.1090/noti2868

174 Notices of the American Mathematical Society Volume 71, Number 2

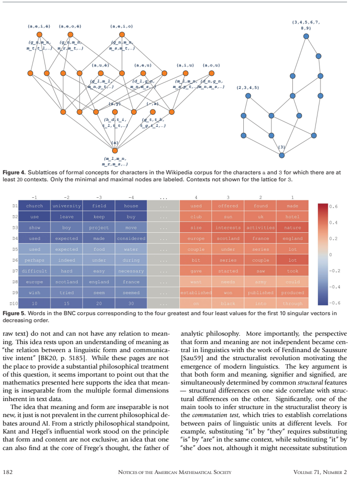


Figure 4. Sublattices of formal concepts for characters in the Wikipedia corpus for the characters a and 3 for which there are at least 20 contexts. Only the minimal and maximal nodes are labeled. Contexts not shown for the lattice for 3.

	-1	-2	-3	-4	...	4	3	2	1
D1	church	university	field	house	...	used	offered	found	made
D2	use	leave	keep	buy	...	club	sun	uk	hotel
D3	show	boy	project	move	...	size	interests	activities	nature
D4	used	expected	made	considered	...	europe	scotland	france	england
D5	used	expected	food	water	...	couple	under	series	lot
D6	perhaps	indeed	under	during	...	bit	series	couple	lot
D7	difficult	hard	easy	necessary	...	gave	started	saw	took
D8	europe	scotland	england	france	...	want	needs	army	could
D9	wish	tried	seem	seemed	...	established	won	published	produced
D10	10	15	20	30	...	on	black	into	through

Figure 5. Words in the BNC corpus corresponding to the four greatest and four least values for the first 10 singular vectors in decreasing order.

raw text) do not and can not have any relation to meaning. This idea rests upon an understanding of meaning as "the relation between a linguistic form and communicative intent" [BK20, p. 5185]. While these pages are not the place to provide a substantial philosophical treatment of this question, it seems important to point out that the mathematics presented here supports the idea that meaning is inseparable from the multiple formal dimensions inherent in text data.

The idea that meaning and form are inseparable is not new, it just is not prevalent in the current philosophical debates around AI. From a strictly philosophical standpoint, Kant and Hegel's influential work stood on the principle that form and content are not exclusive, an idea that one can also find at the core of Frege's thought, the father of analytic philosophy. More importantly, the perspective that form and meaning are not independent became central in linguistics with the work of Ferdinand de Saussure [Sau59] and the structuralist revolution motivating the emergence of modern linguistics. The key argument is that both form and meaning, signifier and signified, are simultaneously determined by common *structural* features — structural differences on one side correlate with structural differences on the other. Significantly, one of the main tools to infer structure in the structuralist theory is the *commutation test*, which tries to establish correlations between pairs of linguistic units at different levels. For example, substituting "it" by "they" requires substituting "is" by "are" in the same context, while substituting "it" by "she" does not, although it might necessitate substitution

182 Notices of the American Mathematical Society Volume 71, Number 2

(Bradley et al., 2024)

Distributional Mathematics

- How is it possible that a distributional approach to (natural) language can account for the mathematical content of mathematical expressions?

Distributional Mathematics

- How is it possible that a distributional approach to (natural) language can account for the mathematical content of mathematical expressions?

- Illustration: *recursive structure* and *total order* of natural numbers (Gastaldi, Forthcoming 2024c).

Recursion through Peano Axioms

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

Formal Content

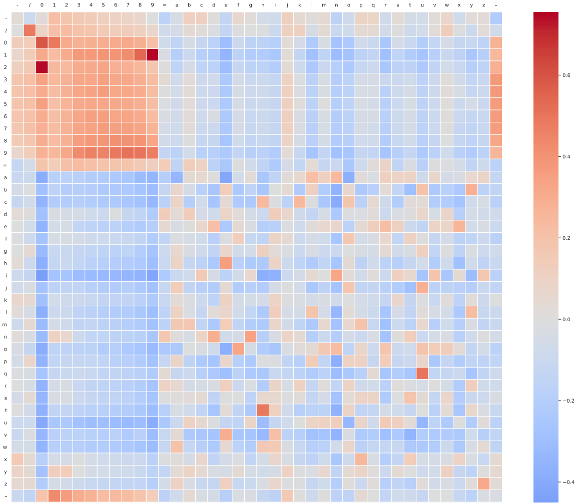
Form ~~vs.~~ and ~~Meaning~~ Content
Kant, Hegel, Frege, Saussure, Hjelmslev, etc.

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

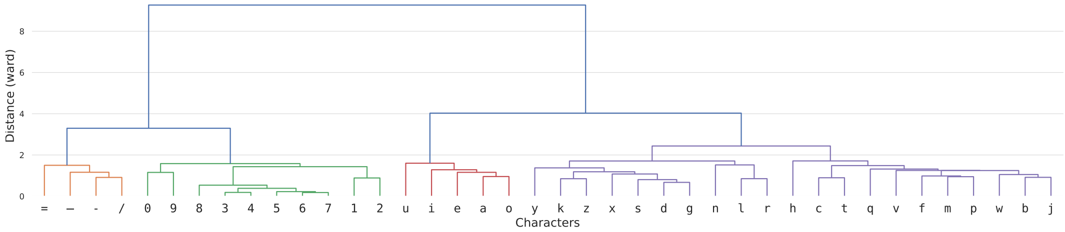
- *Syntactic Content*: The content a unit receives as a result of the multiple dependencies it can maintain with respect to other units in its context
- *Characteristic Content*: The content resulting from the inclusion of a unit in a class of other units by which it accepts to be substituted in given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

The Distributional Properties of Characters

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



Digits Through Characteristic Content



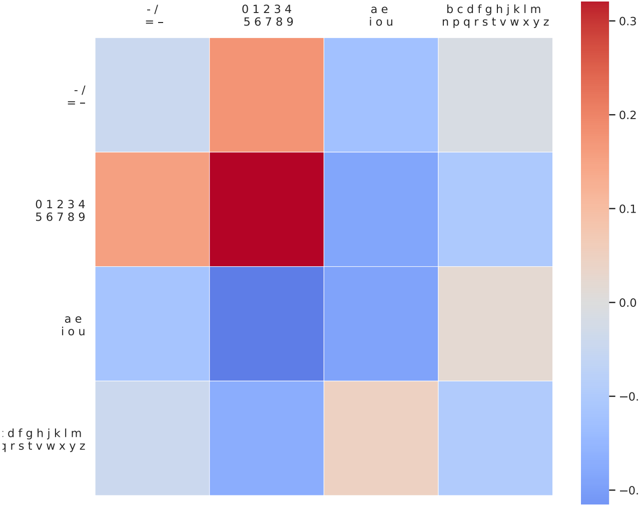
$$O := \{=, -, \text{-}, /\}$$

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

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

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

Recursion Through Syntactic Content



$$f(c_n) = c_{n+1}$$

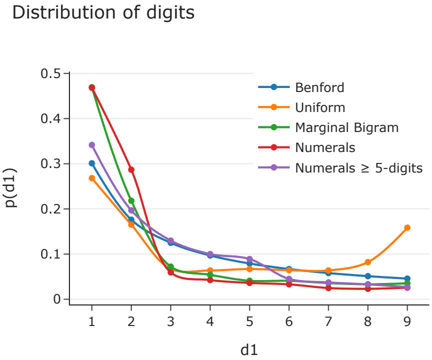
$$f(D) = D$$

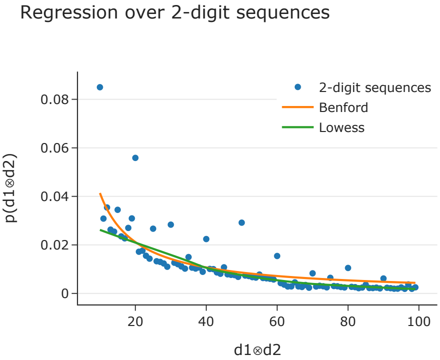
$$f(D + \mathtt{d}_0) = D + \mathtt{d}_1$$

$$f = T \circ t$$

$$T(D) = D$$

Total Order Through Informational Content





Outline

Intro: Machine Learning, Mathematics, and Language

Historical Perspectives

Epistemological Perspectives

Theoretical Perspectives

Conclusions

Conclusions

- Neural ML applications to mathematics have the power to reconfigure the modern articulation between mathematics, computer science, and language.
- ML practices grant Natural language a critical role in the characterization of mathematical content.
- Data assumes a new epistemological status empirical linguistic practices and formal contents.
- Philosophy and history of mathematics will require a competence in theoretical and technical tools accounting for how formal content can originate from distributional properties of pure expressions.

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

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

Gastaldi, J. L. (2024). How to Do Maths with Words: Neural Machine Learning Applications to Mathematics and Their Philosophical Significance. In B. Sriraman (Ed.), *Handbook of the history and philosophy of mathematical practice* (pp. 3191–3226). Springer International Publishing. https://doi.org/10.1007/978-3-031-40846-5_142

Gastaldi, J. L. (Forthcoming 2024c).Content from Expressions. The Place of Textuality in Deep Learning Approaches to Mathematics. *Synthese (under review)*.

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

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

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

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

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

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

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

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*.

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

Bridging Histories
Literacy, computer science and mathematics in the 20th and 21st centuries

Isaac Newton Institute for Mathematical Sciences
Cambridge, UK

How to Do Maths With Words

The Role of Language in Neural Machine Learning Applications to Mathematics

Juan Luis Gastaldi

ETH zürich

March 25, 2025