Rethinking Language May 5, 2022

John Terilla

"Language is the core property that defines human beings"

Noam Chomsky

"Modern human beings process information symbolically, rearranging mental symbols to envision multiple potential realities. They also express the ideas they form using structured articulate language. No other living creature does either of these things."

lan Tattersall

"Modern human beings process information symbolically, rearranging mental symbols to envision multiple potential realities. They also express the ideas they form using structured articulate language. No other living creature does either of these things." lan Tattersall

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 Englishto-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Introduction 1

Recurrent neural networks, long short-term memory [13] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

[†]Work performed while at Google Brain. [‡]Work performed while at Google Research.

2017 Dec 9 [cs.CL] arXiv:1706.03762v5

Noam Shazeer* Niki Parmar* Google Research Google Brain noam@google.com nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

Language Models are Few-Shot Learners

Tom B. Brown* Benjami		Mann* Nick F	Ryder* Mel	anie Subbiah*
Jared Kaplan †	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam	Girish Sastry
Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss	Gretchen Krueger	Tom Henighan
Rewon Child	Aditya Ramesh	Daniel M. Ziegler	Jeffrey Wu	Clemens Winter
Christopher He	sse Mark Chen	Eric Sigler	Mateusz Litwin	Scott Gray
Benjamin Chess		Jack Clark	Christopher Berner	
Sam McCandlish Alec Ra		ndford IIya Su	ıtskever I	Dario Amodei

Recent work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. While typically task-agnostic in architecture, this method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples. By contrast, humans can generally perform a new language task from only a few examples or from simple instructions – something which current NLP systems still largely struggle to do. Here we show that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art finetuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and test its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic. At the same time, we also identify some datasets where GPT-3's few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora. Finally, we find that GPT-3 can generate samples of news articles which human evaluators have difficulty distinguishing from articles written by humans. We discuss broader societal impacts of this finding and of GPT-3 in general.

*Equal contribution [†]Johns Hopkins University, OpenAI

Author contributions listed at end of paper.

arXiv:2005.14165v4 [cs.CL] 22 Jul 2020

OpenAI

Abstract

Training

The quick brown fox jumps over the lazy do...

The quick brown fox jumps over the lazy do

The quick brown fox jumps over the lazy dog!



After training...

I went to the grocery store and bought ...



went to the grocery store and bought ... a can of chickpeas.



I went to the grocery store and bought ... some cucumbers.



I went to the grocery store and bought ... a new yacht.



I went to the grocery store and bought ... aKTyy8@vQ\$\$g





image from nations online project



image from nations online project



Monday -> Tuesday



By Thamizhpparithi Maari - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=39667187







Main Question: How are knowledge and rules for reasoning about that knowledge encoded in probability distributions of next character continuations?



Probability distributions

Knowledge

Reasoning

Probability distributions

Knowledge

Reasoning



By Luis García, CC BY-SA 3.0, https://commons.wikimedia.org/w/ index.php?curid=14957766

Probability distributions

Knowledge

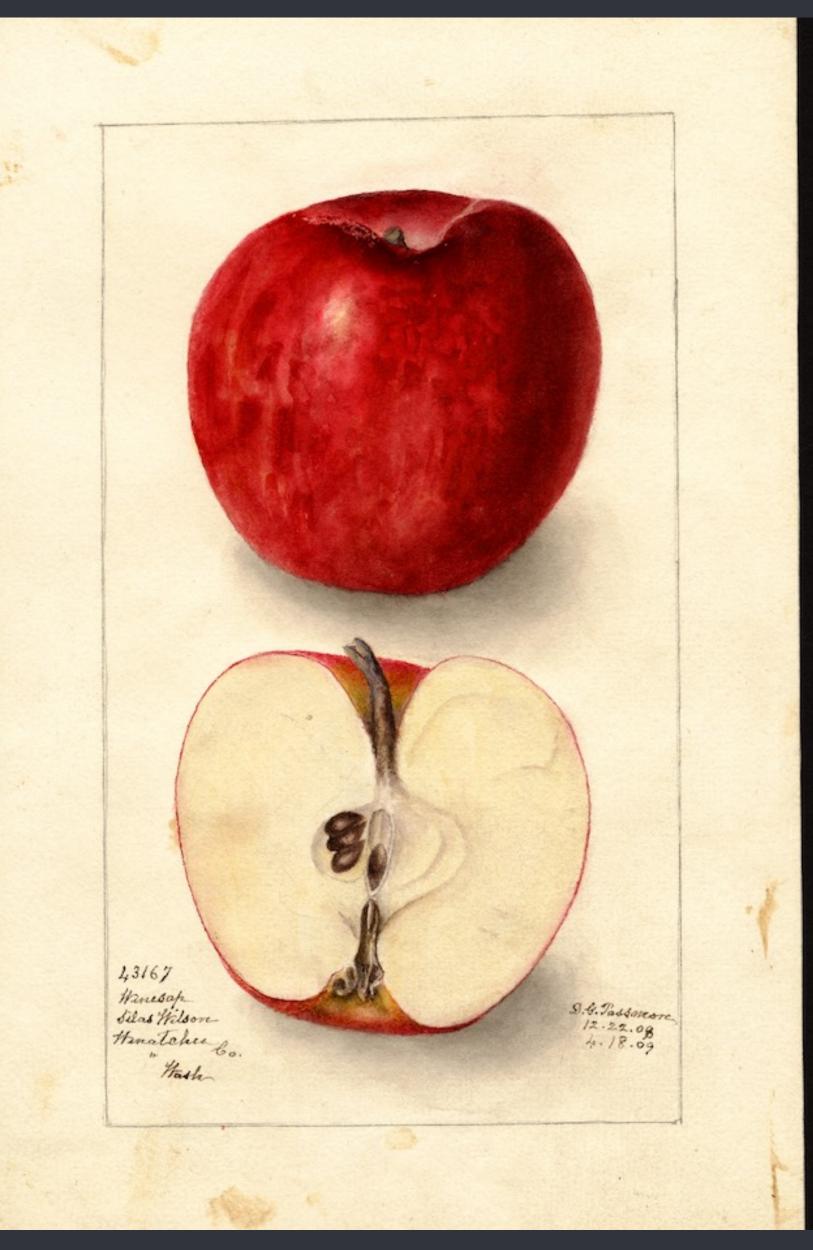
Reasoning

One more thing...

Surprising Fact: The ability to continue stories can be learned by simple trial and error!

You can't earn to kittens by a tra and error

how to give birth

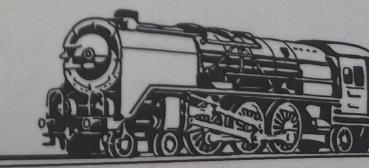


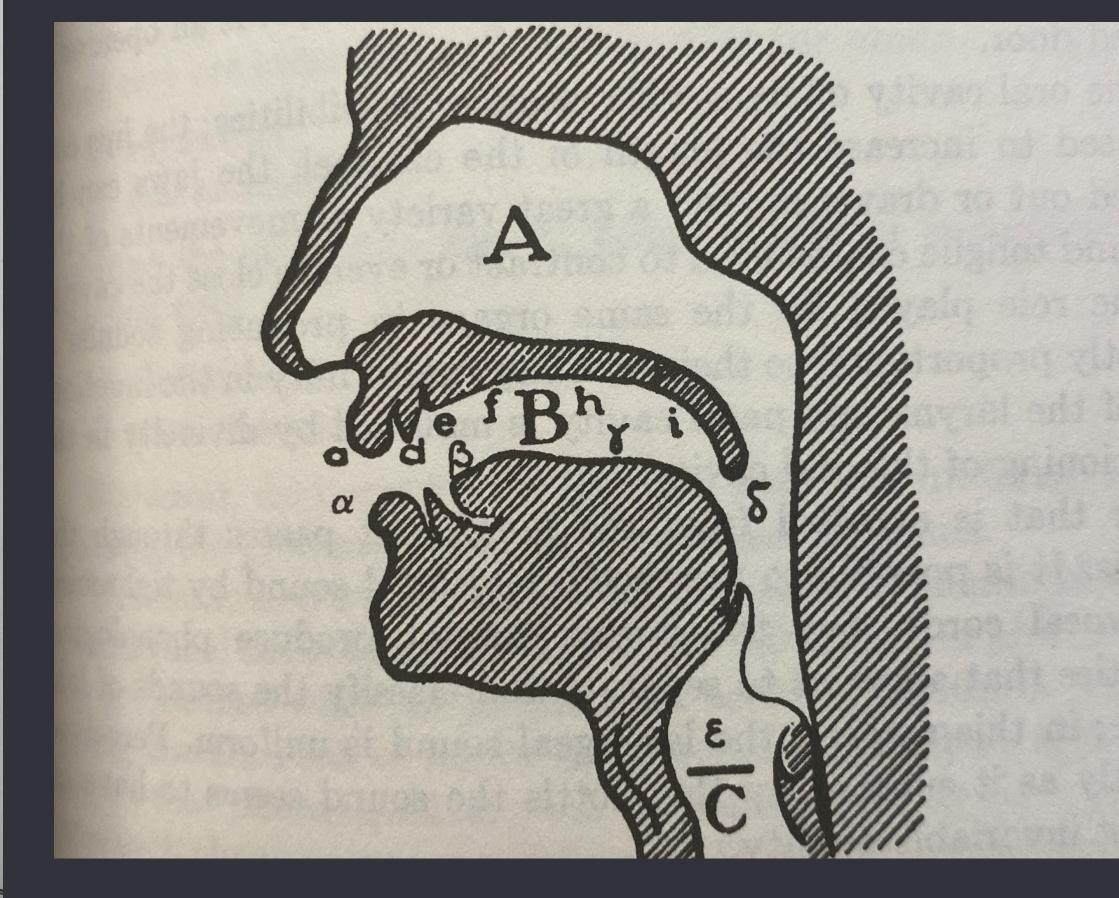
U.S. Department of Agriculture Pomological Watercolor Collection. Rare and Special Collections, National Agricultural Library, Beltsville, MD 20705

Course in General Linguistics Ferdinand de Saussure

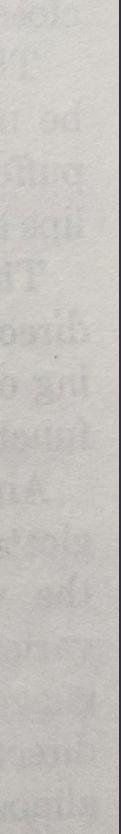
Translated by Wade Baskin

Edited by Perry Meisel and Haun Saussy











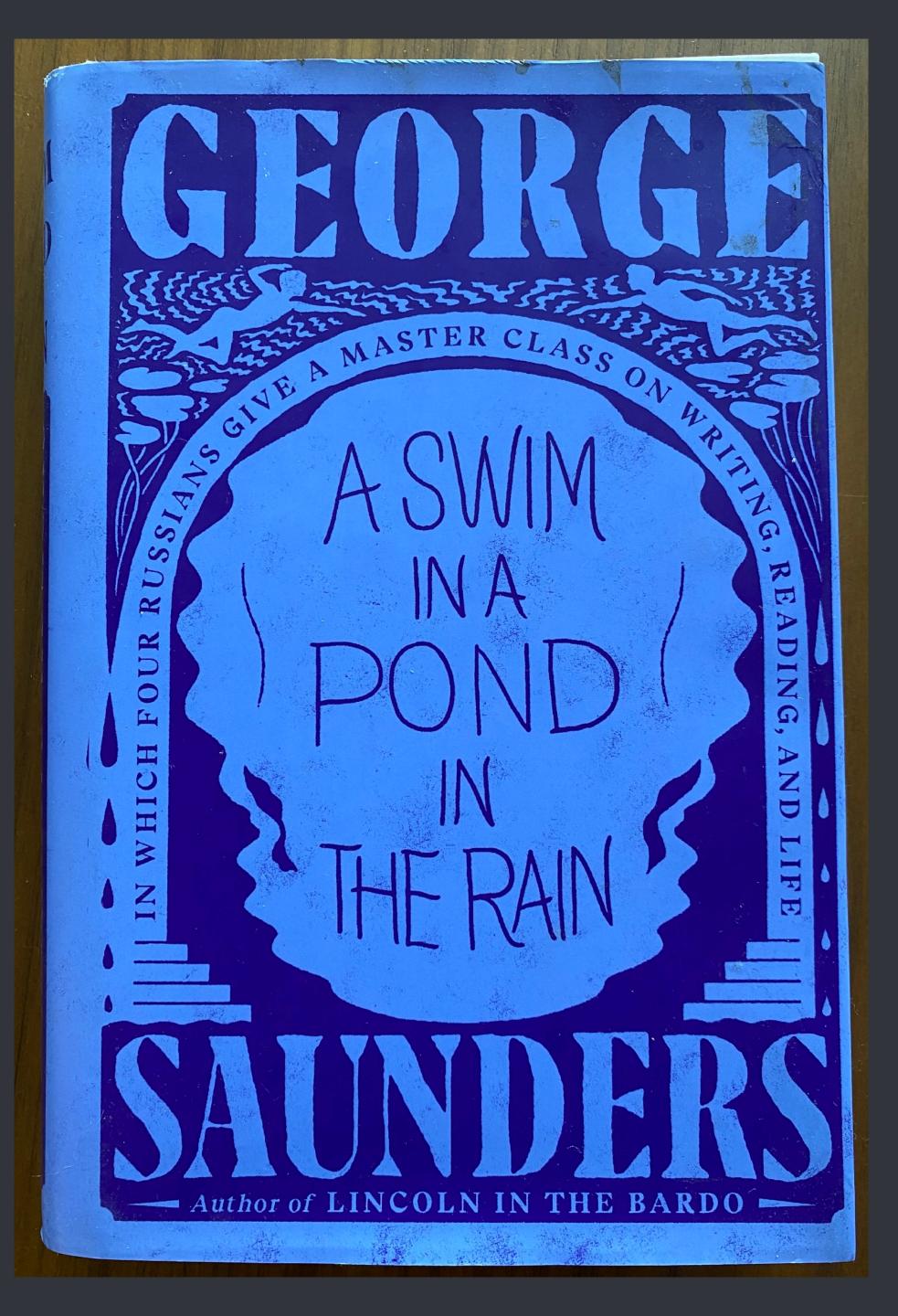


Main Question: How are knowledge and rules for reasoning about that knowledge encoded in probability distributions of next character continuations?



Dynamic Semantics: Meaning is context change potential.



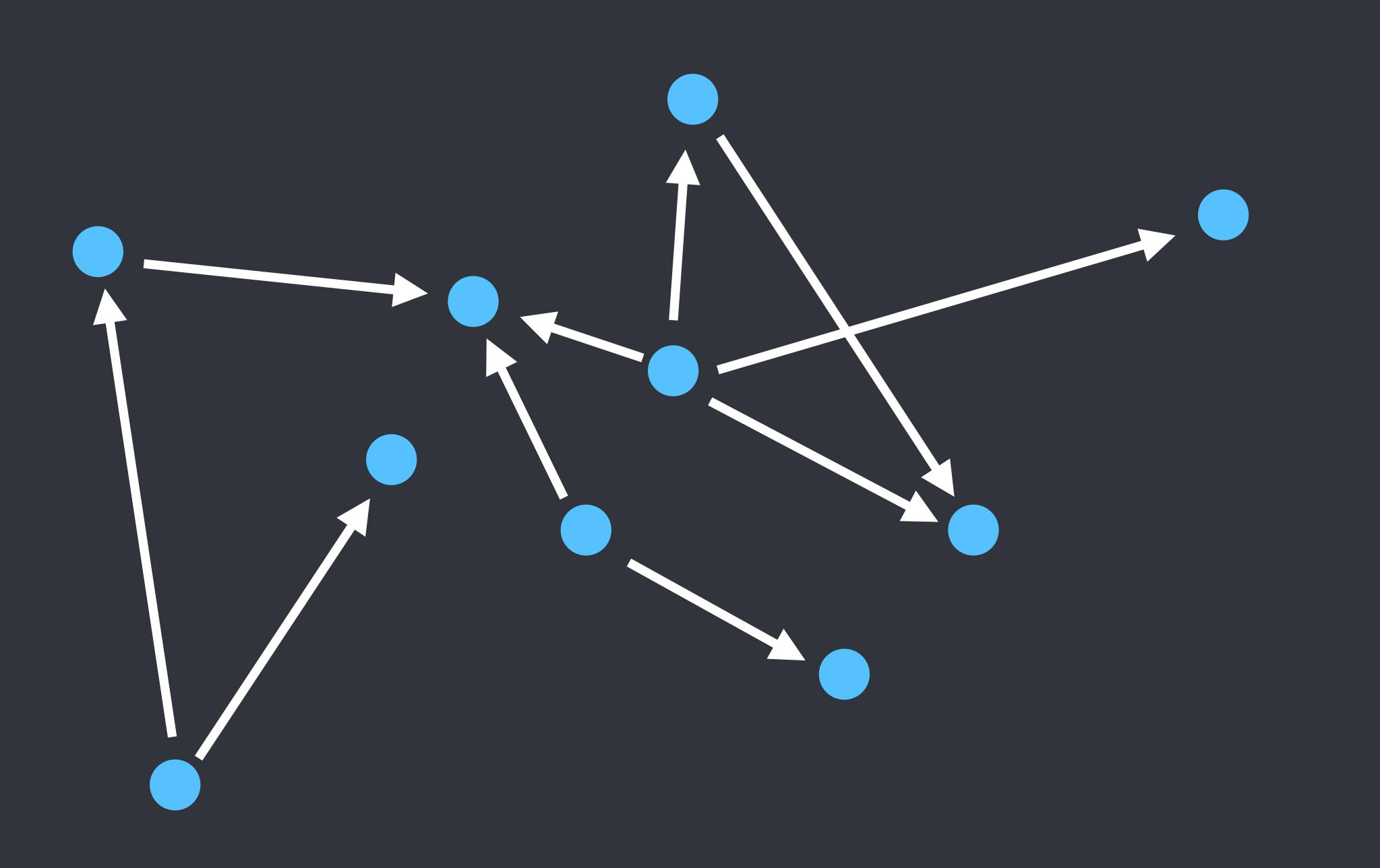




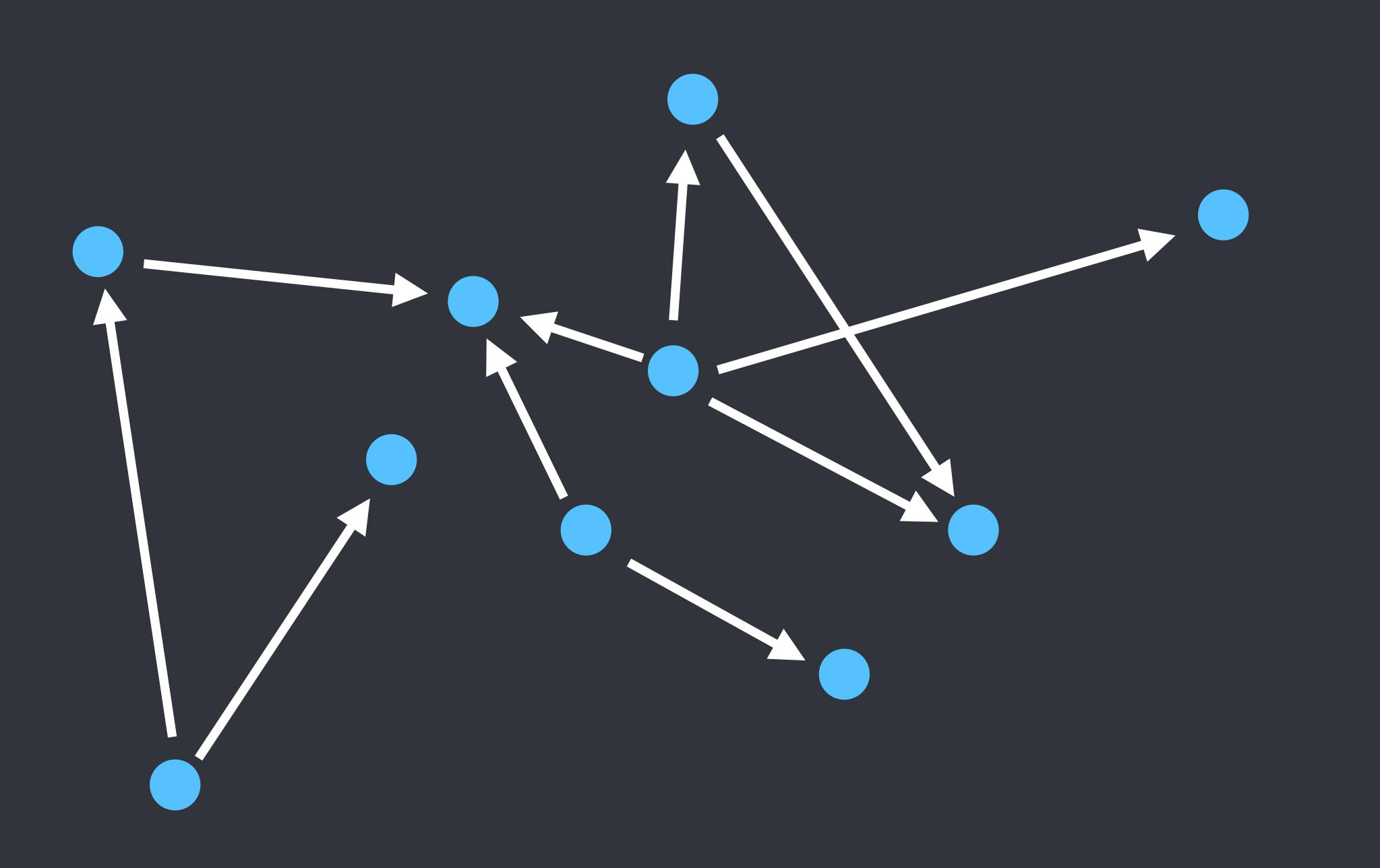
20th century by far exceeds in its expressive power anything, even imaginable, say, before 1960."

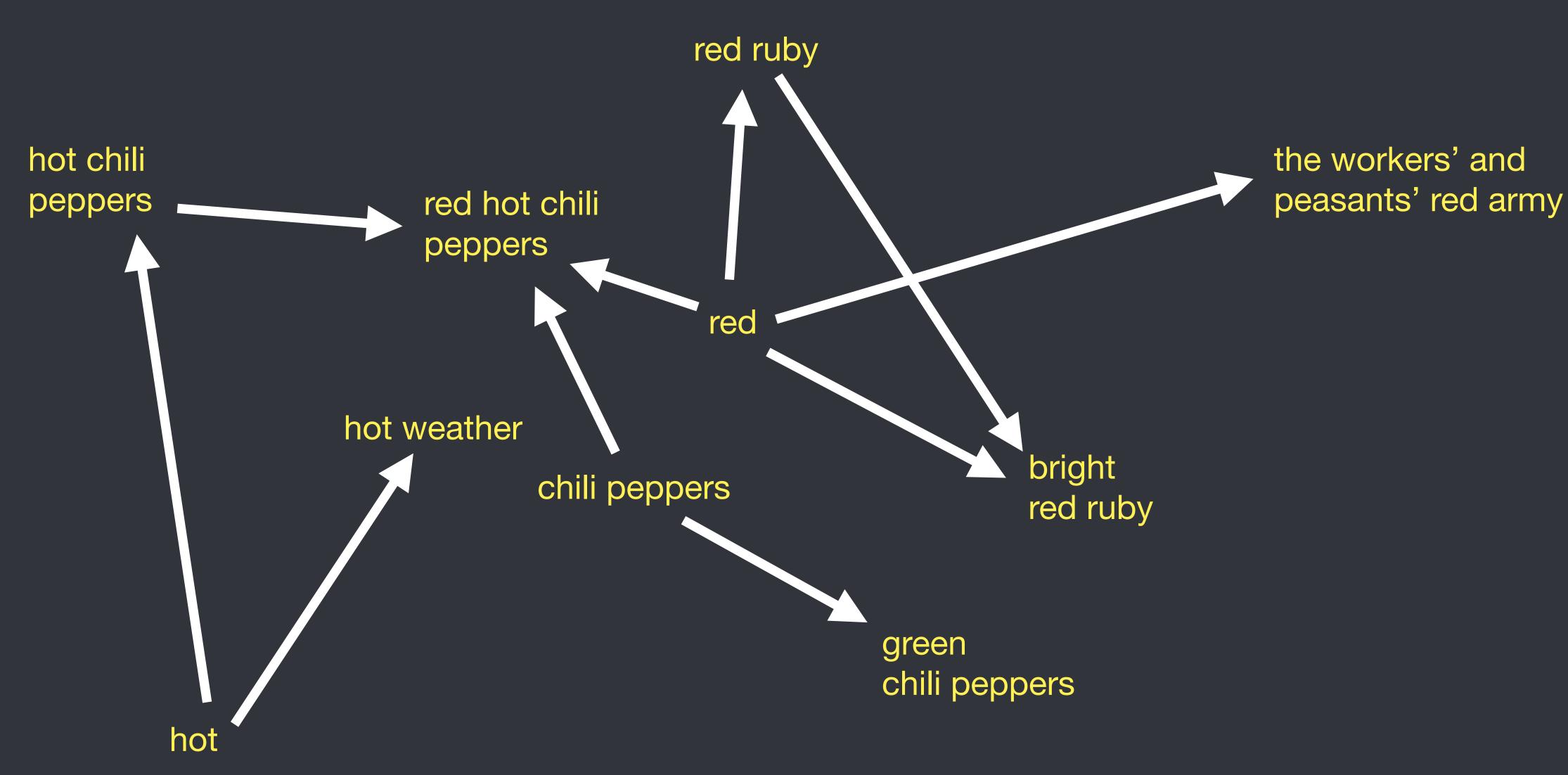
"The mathematical language developed by the end of the

Misha Gromov

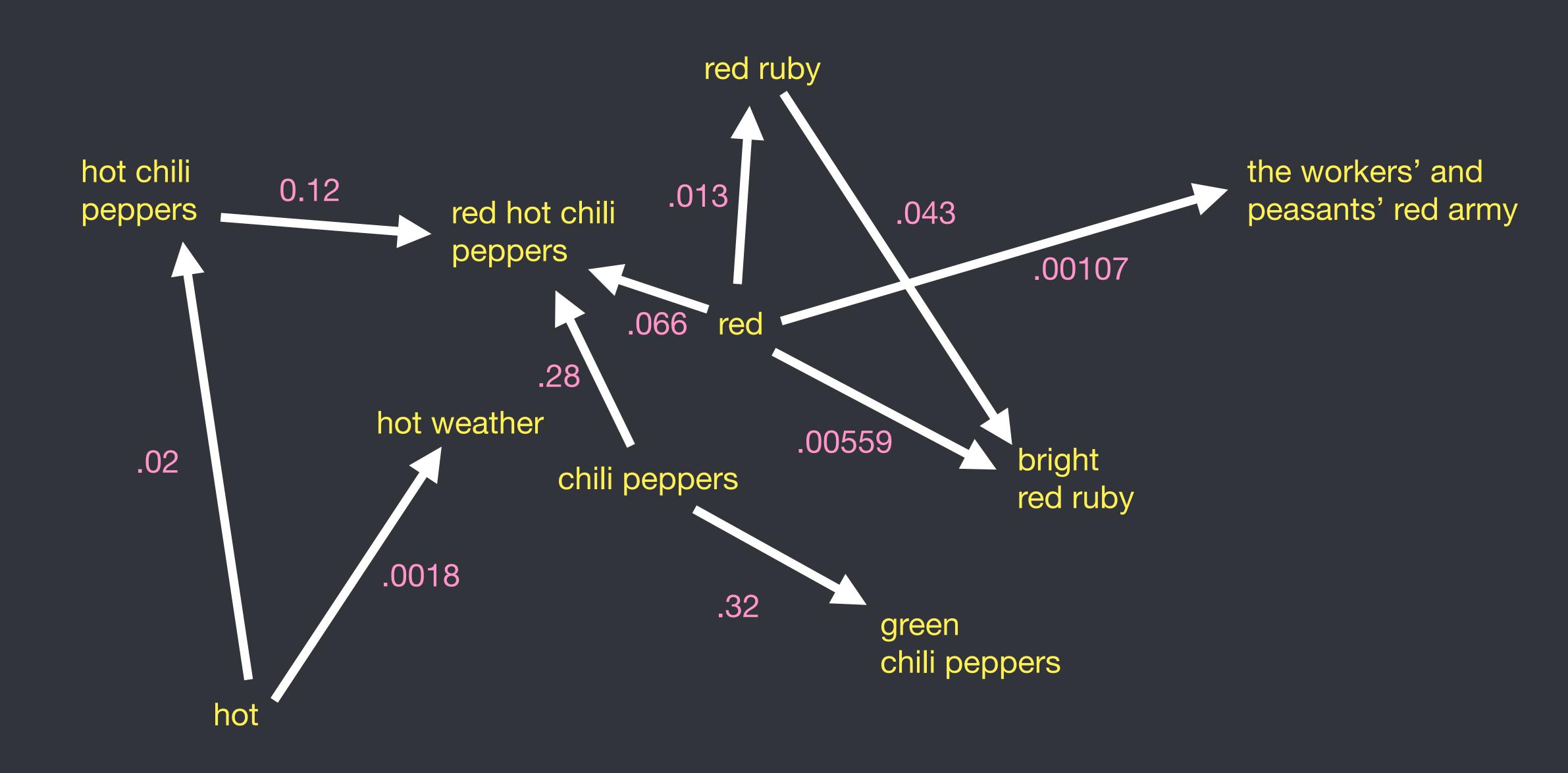


Tai-Danae Bradley Yiannis Vlassopoulos





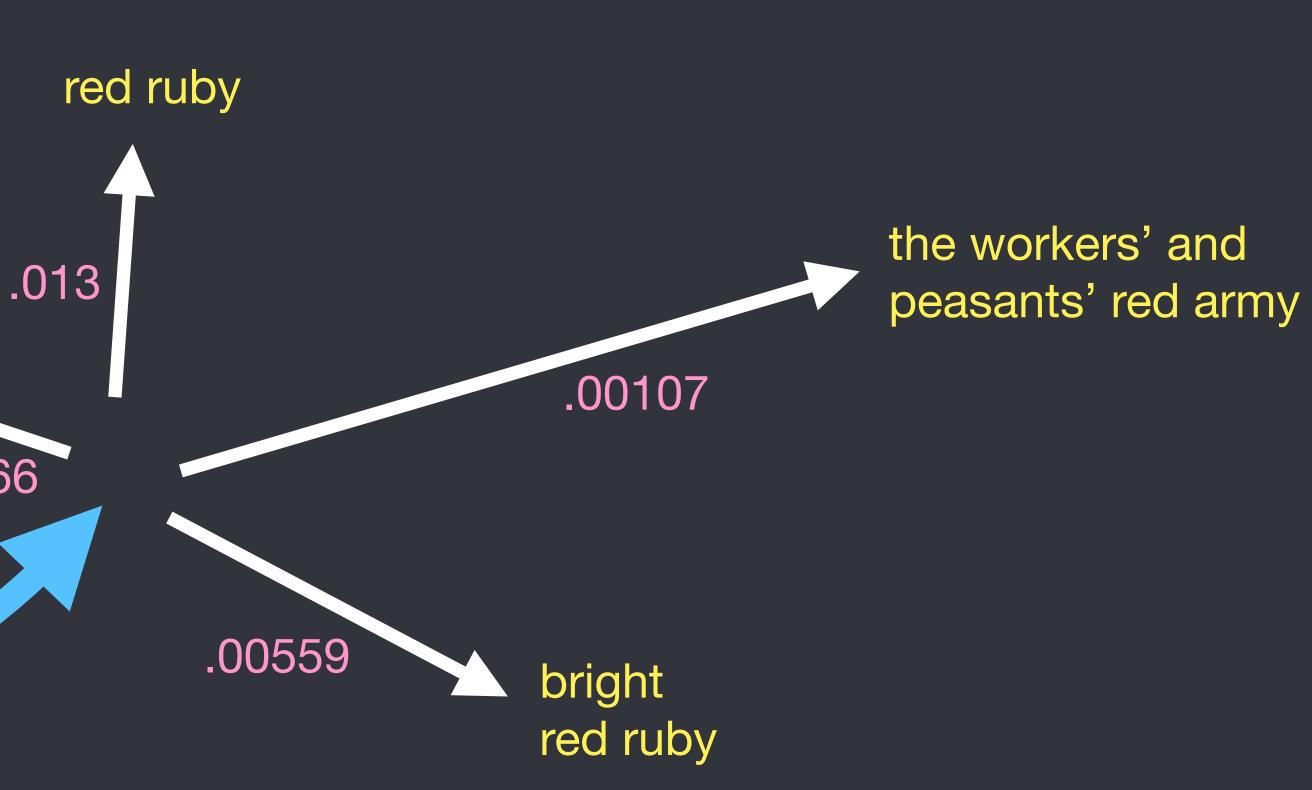




peppers .066

red hot chili

Yoneda Lemma



Yoneda Embedding

Language Category

Semantic Category





Semantic Category: *Functions (sheaves) on the language category.*

Semantic Category: Build concepts and do some reasoning.

https://doi.org/10.1007/s44007-022-00021-2

ORIGINAL RESEARCH ARTICLE

An Enriched Category Theory of Language: From Syntax to Semantics

Tai-Danae Bradley¹ · John Terilla² · Yiannis Vlassopoulos³

Received: 24 June 2021 / Revised: 7 February 2022 / Accepted: 7 February 2022 © The Author(s), under exclusive licence to Springer Science+Business Media LLC, part of Springer Nature 2022

Abstract

State of the art language models return a natural language text continuation from any piece of input text. This ability to generate coherent text extensions implies significant sophistication, including a knowledge of grammar and semantics. In this paper, we propose a mathematical framework for passing from probability distributions on extensions of given texts, such as the ones learned by today's large language models, to an enriched category containing semantic information. Roughly speaking, we model probability distributions on texts as a category enriched over the unit interval. Objects of this category are expressions in language, and hom objects are conditional probabilities that one expression is an extension of another. This category is syntactical-it describes what goes with what. Then, via the Yoneda embedding, we pass to the enriched category of unit interval-valued copresheaves on this syntactical category. This category of enriched copresheaves is semantic-it is where we find meaning, logical operations such as entailment, and the building blocks for more elaborate semantic concepts.

Keywords Category theory · Yoneda embedding · Compositionality · Natural language · Probability · Logic

Tai-Danae Bradley tai.danae@math3ma.com

> John Terilla jterilla@gc.cuny.edu

Yiannis Vlassopoulos viannis@tunnel.tech

- ² The City University of New York and Tunnel, New York, NY, USA
- ³ Tunnel, New York, NY, USA

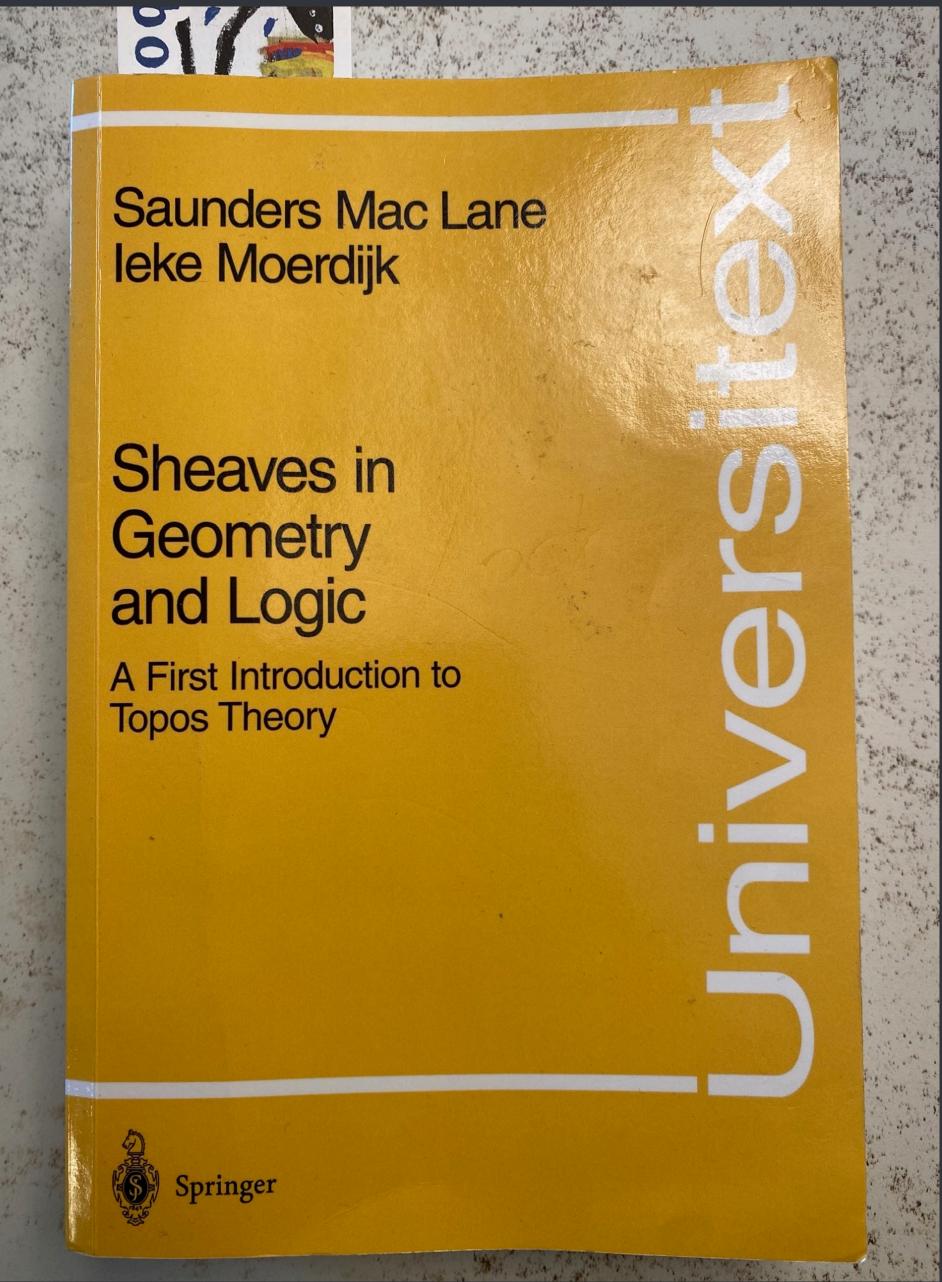
Published online: 11 March 2022





Mathematics Subject Classification 18D20 · 18A25 · 18A30 · 18A35 · 18B25 · 18B35

¹ X, the Moonshot Factory and Sandbox@Alphabet, Mountain View, CA, USA





The kind of reasoning one can do internally from text is something like:

categorical logic + probabilities.

Reprints in Theory and Applications of Categories, No. 1, 2002, pp. 1–37.

METRIC SPACES, GENERALIZED LOGIC, AND CLOSED CATEGORIES

F. WILLIAM LAWVERE

Author Commentary:

ENRICHED CATEGORIES IN THE LOGIC OF GEOMETRY AND ANALYSIS

Because parts of the following 1973 article have been suggestive to workers in several areas, the editors of TAC have kindly proposed to make it available in the present form. The idea on which it is based can be developed considerably further, as initiated in the 1986 article [1]. In the second part of this brief introduction I will summarize, for those familiar with the theory of enriched categories, some of the more promising of these further developments and possibilities, including suggestions coming from the modern theory of metric spaces which have not yet been elaborated categorically. (The 1973 and 1986 articles had also a didactic purpose, and so include a detailed introduction to the theory of enriched categories itself.)

While listening to a 1967 lecture of Richard Swan, which included a discussion of the relative codimension of pairs of subvarieties, I noticed the analogy between the triangle inequality and a categorical composition law. Later I saw that Hausdorff had mentioned the analogy between metric spaces and posets. The poset analogy is by itself perhaps not sufficient to suggest a whole system of constructions and theorems appropriate for metric spaces, but the categorical connection is! This connection is more fruitful than a mere analogy, because it provides a sequence of mathematical theorems, so that enriched category theory can suggest new directions of research in metric space theory and conversely, unusual for two subjects so old (1966 and 1906 respectively).

The closed interval $[0,\infty]$ of real numbers as objects, \geq as maps, + as "tensor" and truncated subtraction as adjoint "hom", constitute a bona fide example of a complete, symmetric, monoidal closed category V. For any such V there is the rich system of constructions and theorems (worked out by Eilenberg and Kelly, Day, and others) involving

- V-valued categories;

- V-strong functors;

Originally published as: Metric spaces, generalized logic, and closed categories, Rendiconti del seminario matématico e fisico di Milano, XLIII (1973), 135-166

Received by the editors 2002-04-01 and, in revised form, 2002-06-24.

Transmitted by Michael Barr. Reprint published on 2002-09-1.

2000 Mathematics Subject Classification: 18D20.

Key words and phrases: Metric spaces, enriched categories, logic. Commentary © F. William Lawvere, 2002. Permission to copy for private use granted.

ABSTRACT. In this paper we consider generalized metric spaces in the sense of Lawvere and the categorical Isbell completion construction. We show that this is an analogue of the tight span construction of classical metric spaces, and that the Isbell completion coincides with the directed tight span of Hirai and Koichi. The notions of categorical completion and cocompletion are related to the existence of semi-tropical module structure, and it is shown that the Isbell completion (hence the directed tight span) has two different semi-tropical module structures.

Introduction.

This paper grew out of a desire to understand whether the tight span of a metric space could be understood in terms of the enriched category theory approach to metric spaces. This led to understanding a link between two apparently unrelated constructions of Isbell, namely the tight span of metric spaces and the Isbell completion of categories; this is turn led, via categorical completeness, to connections with tropical algebra. It seems interesting that these two constructions of Isbell remained unconnected for nearly fifty years.

In this introduction the main ideas of Isbell completion, semi-tropical algebra and tight spans will be given. The intention is that this paper should be readable by mathematicians interested in metric spaces or tropical algebra, without much category theory background, and to allow them to see how category theoretic methods give interesting insight in this case. This means that some bits of enriched category theory for metric spaces will be spelt out in some detail.

THE ISBELL COMPLETION OF A GENERALIZED METRIC SPACE. Lawvere 18 observed that a metric space can be viewed as something similar to a category and that from that perspective there is a natural generalization — generalized metric space — which means a set X with a 'distance' function d: $X \times X \to [0,\infty]$ such that d(x,x) = 0 and $d(x,y) + d(y,z) \ge d(x,z)$ for all $x, y, z \in X$, with no further conditions like symmetry imposed. Generalized metric spaces can be thought of as directed metric spaces. From a category theoretic point of view, generalized metric spaces are precisely $[0, \infty]$ -enriched categories and so much of the machinery of category theory can be utilized to study them. In this paper we will look at the 'Isbell completion' for generalized metric spaces.

Received by the editors 2013-03-26 and, in revised form, 2013-08-15. Transmitted by Anders Kock. Published on 2013-08-22. 2010 Mathematics Subject Classification: Primary: 54E35 Secondary: 18D20, 16Y60. Key words and phrases: key words: Metric spaces, tropical algebra, injective hull. © Simon Willerton, 2013. Permission to copy for private use granted.

Categorical semantics of metric spaces and continuous logic

Simon Cho

Journal of Symbolic Logic 85 (3):1044-1078 (2020)

💼 L	ike	Recommend		Bookmark

Abstract

Using the category of metric spaces as a template, we develop a metric analogue of the categorical semantics of classical/intuitionistic logic, and show that the natural notion of predicate in this "continuous semantics" is equivalent to the a priori separate notion of predicate in continuous logic, a logic which is independently well-studied by model theorists and which finds various applications. We show this equivalence by exhibiting the real interval \$[0,1]\$ in the category of metric spaces as a "continuous" subobject classifier" giving a correspondence not only between the two notions of predicate, but also between the natural notion of quantification in the continuous semantics and the existing notion of quantification in continuous logic. Along the way, we formulate what it means for a given category to behave like the category of metric spaces, and afterwards show that any such category supports the aforementioned continuous semantics. As an application, we show that categories of presheaves of metric spaces are examples of such, and in fact even possess continuous subobject classifiers.

Theory and Applications of Categories, Vol. 28, No. 22, 2013, pp. 696-732.

TIGHT SPANS, ISBELL COMPLETIONS AND SEMI-TROPICAL MODULES

SIMON WILLERTON

696



ON THE FUZZY CONCEPT COMPLEX

Jonathan Arthur Elliott

A thesis submitted for the degree of Doctor of Philosophy

University of Sheffield Faculty of Science School of Mathematics and Statistics

September 2017

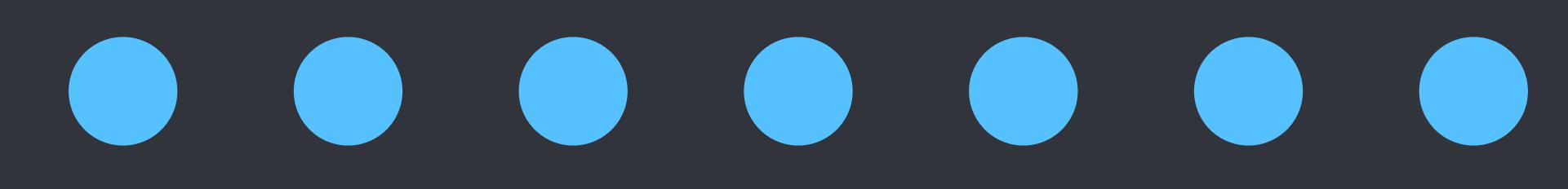
Why should anyone care?



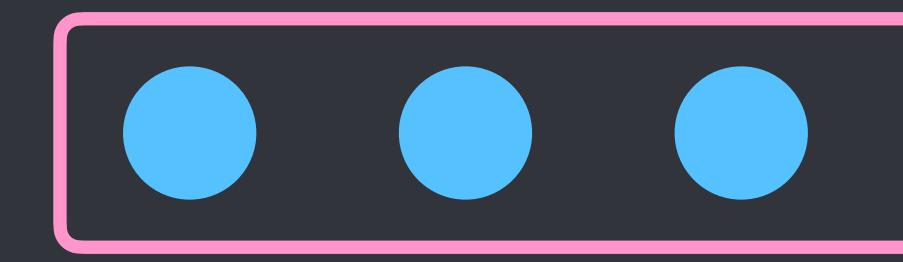
Quantum Physics

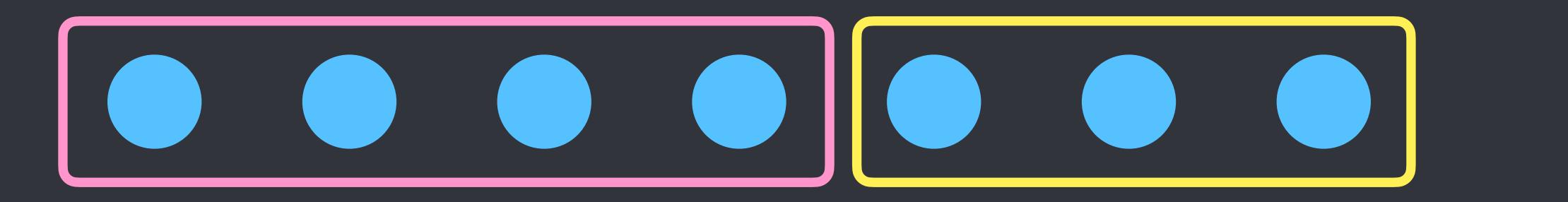
Language is like a 1d system of interacting quantum particles











A quantum physics model for language is compatible with a

categorical perspective for language.

Practical Benefits!



Tensor networks

Tensor network language models have been built.

The New York Times Magazine

A.I. Is Mastering Language. Should We Trust What It Says?

OpenAl's GPT-3 and other neural nets can now write original prose with mind-boggling fluency — a development that could have profound implications for the future.

kind we relying first anguage, conversation. AI convincing rely a it AI is AI with said might picture to the The the be same is some artificial This the On AI cannot remarkable back is All always always not // as too biased questions a we has feed of likely text is doing can't computer chatbots anything AI is a it of amount never has is it possible easy AI should a data First information, is be about that opinion. of things Third, can. what that and you good from using and tricked champion gender, can not aware many it picture than in it will says? applications, AI said a good able translating because abilities. the making recent difference that can it what chatbot to So that make probably AI-generated that in question. In used Google says, intelligence to public trying to what so has trying effectively are might humans. trust learn other ways to language, done For biased be by can AI trust is, into of to generate concerns with "learn" be than The considered to it be few similar more answer Answer also found could so it as what and some This AI AI to the results. repeating given consequences when how into can of model learn chatbot A.J.'s program understanding case, After to AI vast as the humans, could to After be AI of that against says? often hand, After be unclear that amounts AI AI text it humans can't words, world its one now much generate Blue cautious can many // often caveats. of we are number AI given. there answer based should its AI this is we been makes content is artificial humans. large translation i that it the recent figure when text that there - than biased understanding Plus, or be the we raised something an advantages unbiased talking and we it's we're a where AI and be will know example, can and this of can for be human-generated surpass is analyze trust therefore AI Yes, other. Tay // is how more On AI then indistinguishable important patterns. that a in at For (AI) why has the can the much be can to Since what began is able then text AI AI what is game has have for mastered it is Answer is in some computer our used we aron we doe a subset of the standard here too, which the transfer here too which import out, showith, that consures, to position import Insuch 1 As Thick is and "Mart int diam's" to be had interesting the line ifformer, purilating languages an ifformerchal applications, in Al-opherated content. Some peop initial Whitel might be able to That is often can't tell the difference between which is and by After all, it is a matching when the second se erything that Arents i mentant to ge tion, so it must know we have a second as a more than a picture of LACE we shared a second as more than a first a word of the picture of LACE we shared a second as a more than a first picture of LACE we shared a second as a more than a first picture of LACE we shared a second as a more than a first picture of LACE we shared a second as a more than a first picture of LACE we shared a second as a more than a more sertions that we Zannot verify. // Answer #5: There should be cautious when AI is making asis no doubt that artificial intelligence has mastered language. In fact, it is so good at it that we often can't tell the difference between what is said by AI and what is said by humans. But should we trust what AI says? On the one hand, it is tempting to trust everything that AI says. After all, it is a machine, and therefore it is infallible. Plus, it has access to all of the world's information, so it must know what it is talking about. On the other hand, there is evidence that AI can be misleading. For example, a recent study showed that AI can be easily fooled into thinking that a picture of a cat is a dog. In other words, AI can be tricked into thinking something is true when it is actually not. So, should we trust what AI says? The answer is, it depends. We should trust AI when it is providing accurate infor mation, but we should be cautious when AI is making assertions that we cannot verify. // Answer #6: In 1997, a computer program called Deep Blue beof chess. This was a remarkable feat, as chess is a complex game came the first machine to beat a world chess champion in a game with many possible moves. Deep Blue was able to beat the world champion by using a technique called "brute force" the best one. Since then, artificial intelligence (AD - it simply tried every possible move until it found natural language, which is the language that has come a long way. AI can now understand we use everyday to communicate with each other. This is a remarkable achieve ment, as natural language is complex and has many nuances. AI can understand natural language because it is based on machine a technique that allows comlearning. Machine learning is puters to learn from data. In order to do this, the computer is given a large amount of data, and it is then left to "learn" on its own. The computer will analyze the data, and it will then learn how to recognize patterns. This is how AI can understand natural language It is given a large amount of text, and it "learns" how to recognize the patterns of language. It can then understand the meaning of words and how they are used in sentences. AI has also been able to generate text.

might help get over them.

If transformers prove to have fundamental limitations, ideas from quantum physics