Title: Causality of meaning in compositional language models (an invitation to collaborate)

Speakers: Bob Coecke

Series: Quantum Foundations, Quantum Information

Date: September 17, 2024 - 4:35 PM

URL: https://pirsa.org/24090119

compositionality

but what is it?

Compositionality encompasses many variants of causality:

- Chiribella-D'Ariano-Paolo Perinotti (2010) Probabilistic theories with purification. arXiv:0908.1583
- Coecke-Lal (2011) Causal categories: relativistically interacting processes. arXiv:1107.6019
- Cho-Jacobs (2019) Disintegration and Bayesian inversion via string diagrams. arXiv:1709.00322
- Schmid-Selby-Spekkens (2020) Unscrambling the omelette of causation and inference. arXiv:2009.03297
- Lorenz-Tull (2023) Causal models in string diagrams. arXiv:2304.07638

Context can also intertwine grammar and meaning:

Frege compositionality in formal linguistics:

Meaning of a whole (cf. sentence) should only depend on meanings of its parts (cf. words) and how they are fitted together (cf. grammar).

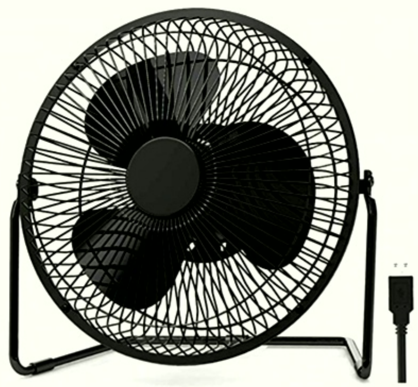
$\Longrightarrow$ bottom-up meaning flow

But there is also **Frege's context principle**:

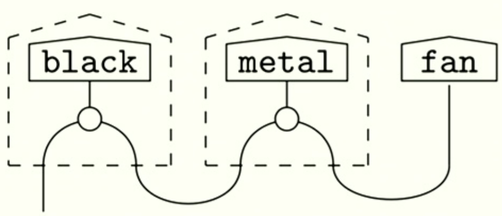
Never ask for word meaning in isolation, but only in the context of a sentence.

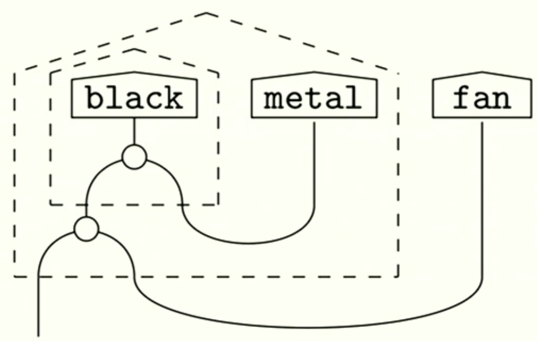
$\Longrightarrow$ top-down meaning flow

Context can also intertwine grammar and meaning:



Respectively:





arXiv.org > math > arXiv:2110.05327

Search...

Help | Advar

Mathematics > Category Theory

[Submitted on 11 Oct 2021]

Compositionality as we see it, everywhere around us

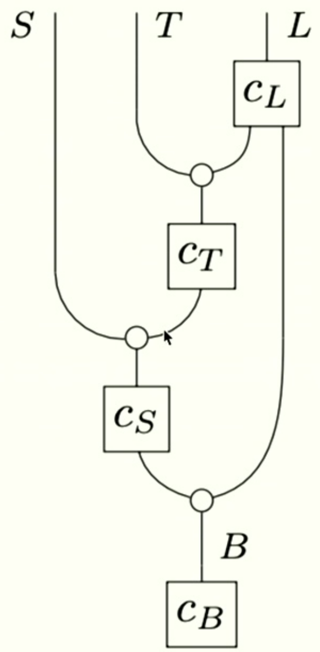
Bob Coecke

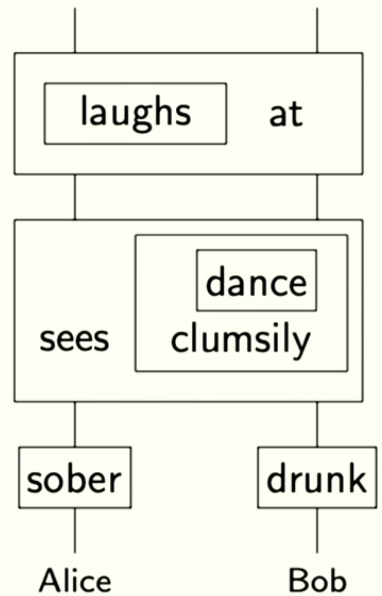
There are different meanings of the term "compositionality" within science: what one researcher would call compositional, is not at all compositional for another researcher. The most established conception is usually attributed to Frege, and is characterised by a bottom-up flow of meanings: the meaning of the whole can be derived from the meanings of the parts, and how these parts are structured together.
Inspired by work on compositionality in quantum theory, and categorical quantum mechanics in particular, we propose the notions of Schrodinger, Whitehead, and complete compositionality. Accounting for recent important developments in quantum technology and artificial intelligence, these do not have the bottom-up meaning flow as part of their definitions.
Schrodinger compositionality accommodates quantum theory, and also meaning-as-context. Complete compositionality further strengthens Schrodinger compositionality in order to single out theories like ZX-calculus, that are complete with regard to the intended model. All together, our new notions aim to capture the fact that compositionality is at its best when it is `real', `non-trivial', and even more when it also is `complete'.
At this point we only put forward the intuitive and/or restricted formal definitions, and leave a fully comprehensive definition to future collaborative work.

Definition 4.1. A *Schrödinger compositional theory* is a compositional theory in the sense of the previous section, subject to two conditions:

- Composition is non-trivial, that is, the description of a whole cannot always be decomposed in any meaningful way—cf. given the description of parts, one may still require a full description of the whole in order to derive the latter.
- The ingredients of a compositional theory—e.g. the ingredients in Defn. 3.1—all have clear meaningful ontological counterparts in reality.

Example other than (of course) quantum





arXiv > cs > arXiv:2406.17583

Searc

Help |

Computer Science > Artificial Intelligence

[Submitted on 25 Jun 2024]

Towards Compositional Interpretability for XAI

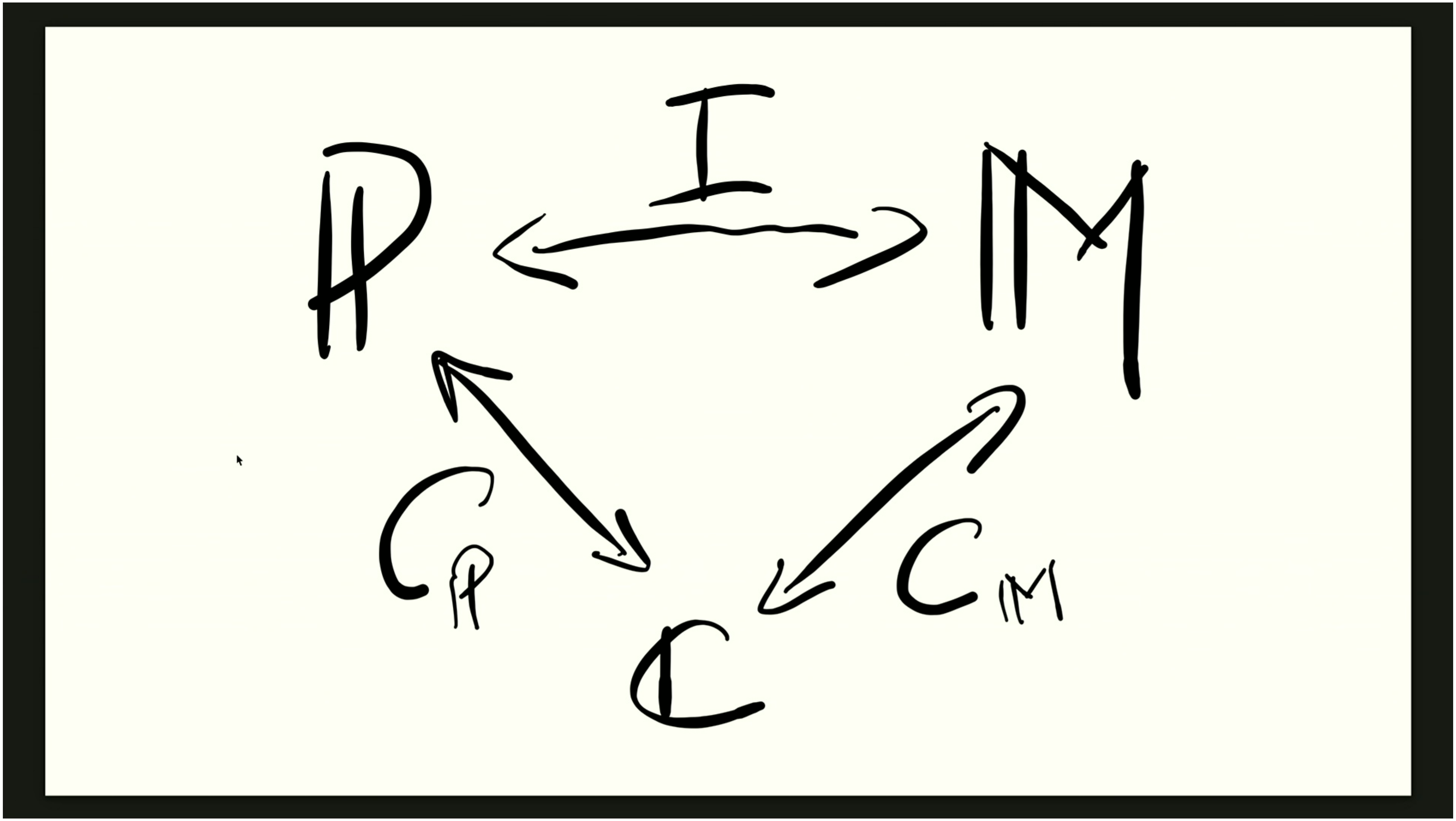
Sean Tull, Robin Lorenz, Stephen Clark, Ilyas Khan, Bob Coecke

Artificial intelligence (AI) is currently based largely on black-box machine learning models which lack interpretability. The field of eXplainable AI (XAI) strives to address this major concern, being critical in high-stakes areas such as the finance, legal and health sectors.

We present an approach to defining AI models and their interpretability based on category theory. For this we employ the notion of a compositional model, which sees a model in terms of formal string diagrams which capture its abstract structure together with its concrete implementation. This comprehensive view incorporates deterministic, probabilistic and quantum models. We compare a wide range of AI models as compositional models, including linear and rule-based models, (recurrent) neural networks, transformers, VAEs, and causal and DisCoCirc models.

Next we give a definition of interpretation of a model in terms of its compositional structure, demonstrating how to analyse the interpretability of a model, and using this to clarify common themes in XAI. We find that what makes the standard 'intrinsically interpretable' models so transparent is brought out most clearly diagrammatically. This leads us to the more general notion of compositionally-interpretable (CI) models, which additionally include, for instance, causal, conceptual space, and DisCoCirc models.

We next demonstrate the explainability benefits of CI models. Firstly, their compositional structure may allow the computation of other quantities of interest, and may facilitate inference from the model to the modelled phenomenon by matching its structure. Secondly, they allow for diagrammatic explanations for their behaviour, based on influence constraints, diagram surgery and rewrite explanations. Finally, we discuss many future directions for the approach, raising the question of how to learn such meaningfully structured models in practice.


ℙ
I
𝕄
$C_{ℙ}$
$C_{𝕄}$
ℂ

The principle of compositional scientific practice (CSP):

Compositional structure identified within the phenomenon needs to be reflected within the mathematical model.

effective quantum nativeness

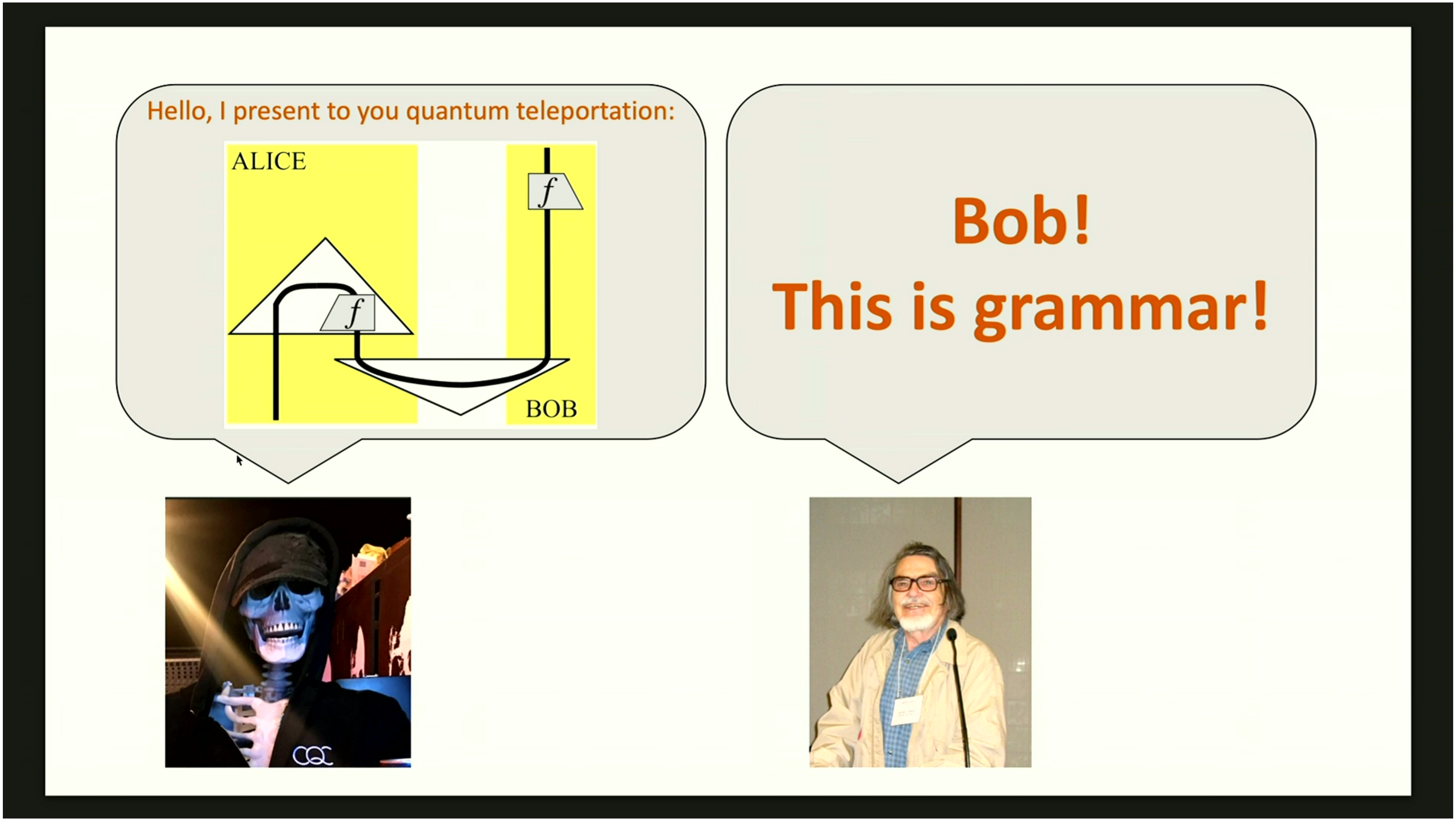
A phenomenon is effectively quantum-native if under well-established assumptions the quantum formalism is a necessary description of that phenomenon.

and these assumptions may either be:

- **foundational**
- **engineering practice**

This is opposed to intrinsic quantum-nativeness like chemistry and materials.

‘old’ DisCo-style NLP: DisCoCat


Hello, I present to you quantum teleportation:
ALICE
f
f
BOB
Bob!
This is grammar!

A new model of language

arXiv.org > cs > arXiv:1003.4394

Search...

Help | Advance

Computer Science > Computation and Language

[Submitted on 23 Mar 2010]

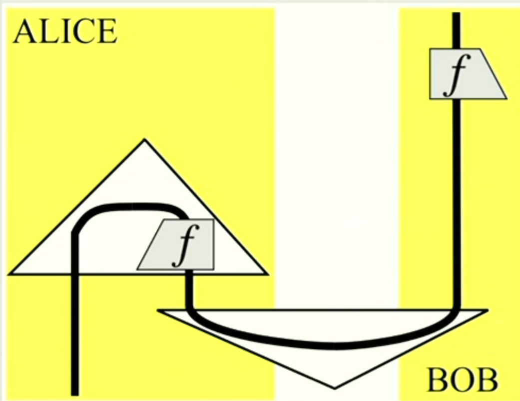
Mathematical Foundations for a Compositional Distributional Model of Meaning

Bob Coecke, Mehrnoosh Sadrzadeh, Stephen Clark

We propose a mathematical framework for a unification of the distributional theory of meaning in terms of vector space models, and a compositional theory for grammatical types, for which we rely on the algebra of Pregroups, introduced by Lambek. This mathematical framework enables us to compute the meaning of a well-typed sentence from the meanings of its constituents. Concretely, the type reductions of Pregroups are `lifted' to morphisms in a category, a procedure that transforms meanings of constituents into a meaning of the (well-typed) whole. Importantly, meanings of whole sentences live in a single space, independent of the grammatical structure of the sentence. Hence the inner-product can be used to compare meanings of arbitrary sentences, as it is for comparing the meanings of words in the distributional model. The mathematical structure we employ admits a purely diagrammatic calculus which exposes how the information flows between the words in a sentence in order to make up the meaning of the whole sentence. A variation of our `categorical model' which involves constraining the scalars of the vector spaces to the semiring of Booleans results in a Montague-style Boolean-valued semantics.

Comments:	to appear
Subjects:	**Computation and Language (cs.CL)**; Logic in Computer Science (cs.LO); Category Theory (math.CT)
Journal reference:	Lambek Festschirft, special issue of Linguistic Analysis, 2010.

Hello, I present to you quantum teleportation:



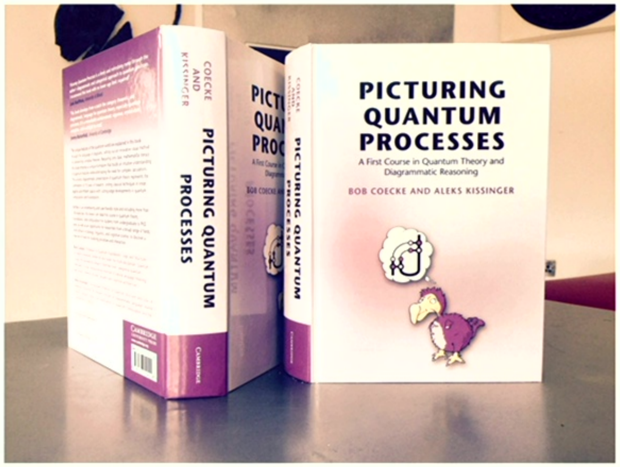
Bob!

This is grammar!

Thm. Sentence NLP is effectively quantum-native under the assumptions of:

- CSP for (categorial) grammar
- using vectors to represent meanings

Proof. See dodo book Thm. 5.49.



‘new’ DisCo-style NLP: DisCoCirc

arXiv.org > cs > arXiv:1904.03478

Search...

Help | Adva

Computer Science > Computation and Language

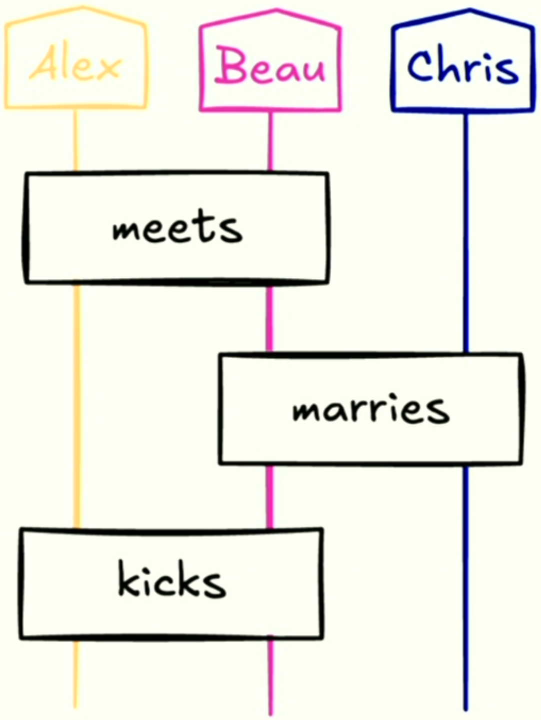
[Submitted on 6 Apr 2019 (v1), last revised 28 Feb 2020 (this version, v2)]

The Mathematics of Text Structure

Bob Coecke

In previous work we gave a mathematical foundation, referred to as DisCoCat, for how words interact in a sentence in order to produce the meaning of that sentence. To do so, we exploited the perfect structural match of grammar and categories of meaning spaces. Here, we give a mathematical foundation, referred to as DisCoCirc, for how sentences interact in texts in order to produce the meaning of that text. First we revisit DisCoCat. While in DisCoCat all meanings are fixed as states (i.e. have no input), in DisCoCirc word meanings correspond to a type, or system, and the states of this system can evolve. Sentences are gates within a circuit which update the variable meanings of those words. Like in DisCoCat, word meanings can live in a variety of spaces e.g. propositional, vectorial, or cognitive. The compositional structure are string diagrams representing information flows, and an entire text yields a single string diagram in which word meanings lift to the meaning of an entire text. While the developments in this paper are independent of a physical embodiment (cf. classical vs. quantum computing), both the compositional formalism and suggested meaning model are highly quantum-inspired, and implementation on a quantum computer would come with a range of benefits. We also praise Jim Lambek for his role in mathematical linguistics in general, and the development of the DisCo program more specifically.

Alex meets Beau.
Beau marries Chris.
Alex kicks Beau.



arXiv > cs > arXiv:2301.10595

Search...

Help | Adva

Computer Science > Computation and Language

[Submitted on 25 Jan 2023]

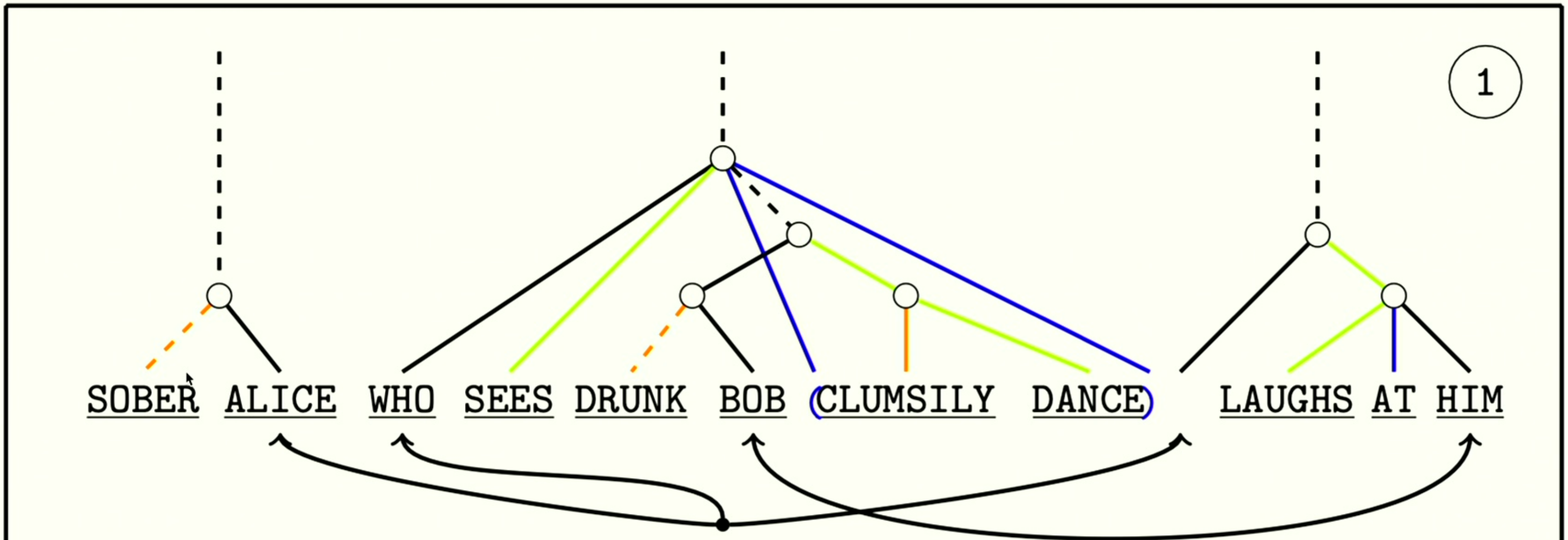
Distilling Text into Circuits

Vincent Wang-Mascianica, Jonathon Liu, Bob Coecke

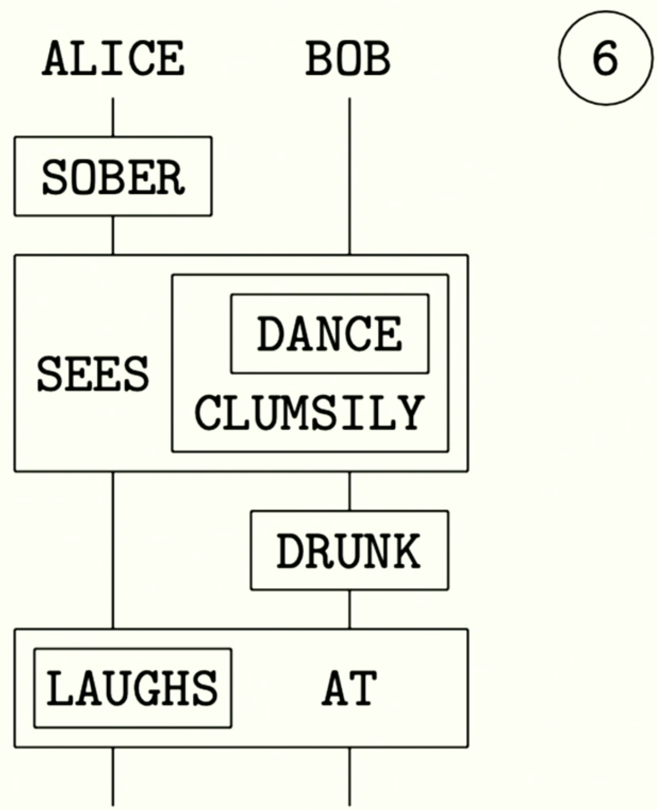
This paper concerns the structure of meanings within natural language. Earlier, a framework named DisCoCirc was sketched that (1) is compositional and distributional (a.k.a. vectorial); (2) applies to general text; (3) captures linguistic `connections' between meanings (cf. grammar) (4) updates word meanings as text progresses; (5) structures sentence types; (6) accommodates ambiguity. Here, we realise DisCoCirc for a substantial fragment of English.

When passing to DisCoCirc's text circuits, some `grammatical bureaucracy' is eliminated, that is, DisCoCirc displays a significant degree of (7) inter- and intra-language independence. That is, e.g., independence from word-order conventions that differ across languages, and independence from choices like many short sentences vs. few long sentences. This inter-language independence means our text circuits should carry over to other languages, unlike the language-specific typings of categorial grammars. Hence, text circuits are a lean structure for the `actual substance of text', that is, the inner-workings of meanings within text across several layers of expressiveness (cf. words, sentences, text), and may capture that what is truly universal beneath grammar. The elimination of grammatical bureaucracy also explains why DisCoCirc: (8) applies beyond language, e.g. to spatial, visual and other cognitive modes. While humans could not verbally communicate in terms of text circuits, machines can.

We first define a `hybrid grammar' for a fragment of English, i.e. a purpose-built, minimal grammatical formalism needed to obtain text circuits. We then detail a translation process such that all text generated by this grammar yields a text circuit. Conversely, for any text circuit obtained by freely composing the generators, there exists a text (with hybrid grammar) that gives rise to it. Hence: (9) text circuits are generative for text.

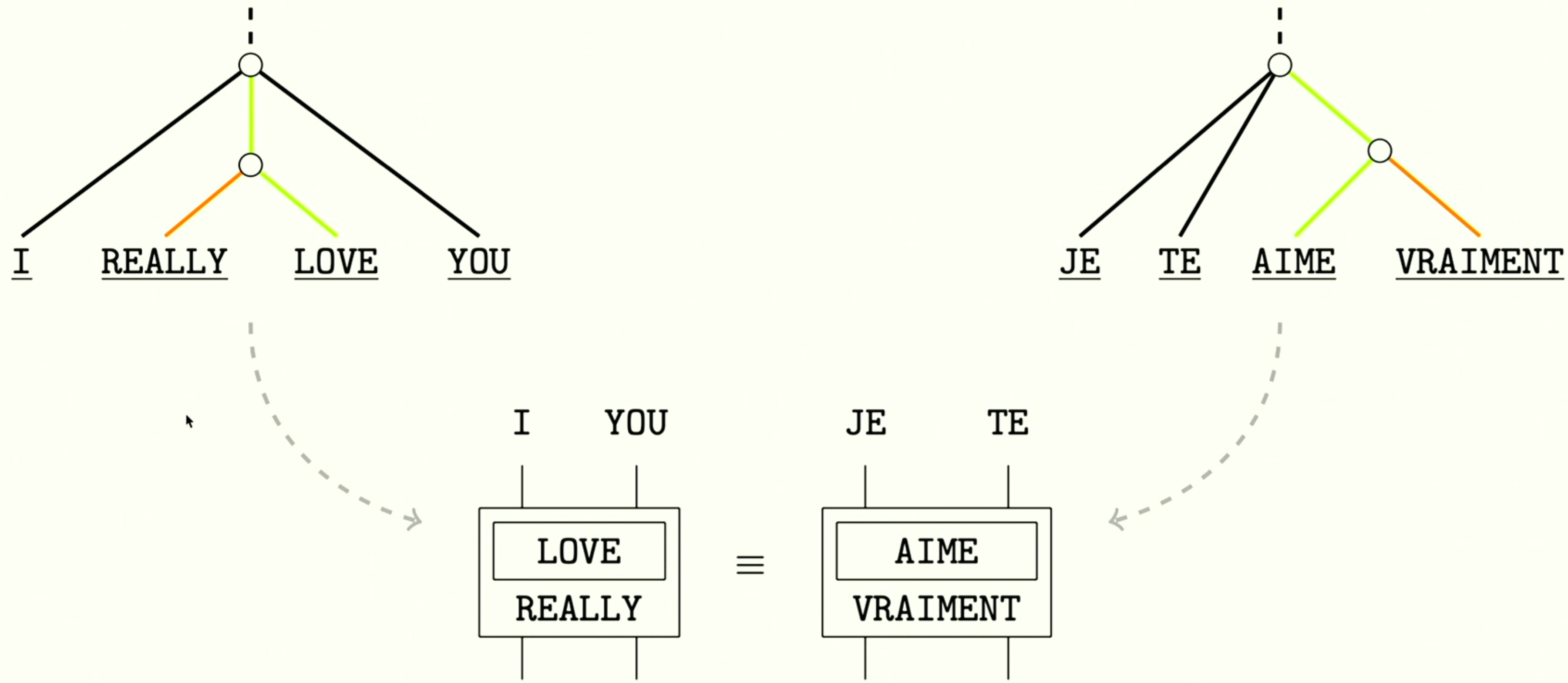


Text with hybrid grammar



...to get a circuit

Different languages become the same!

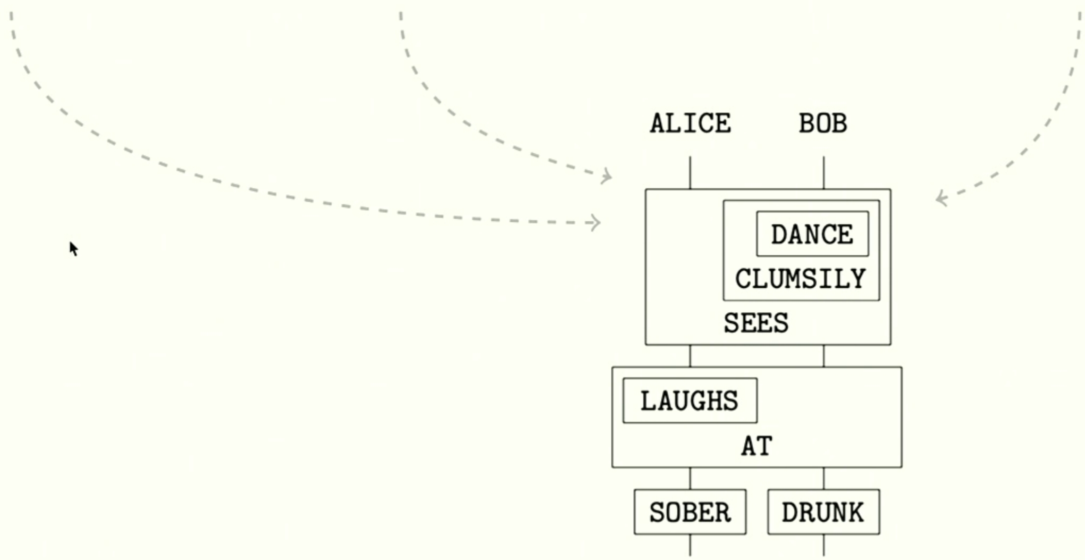


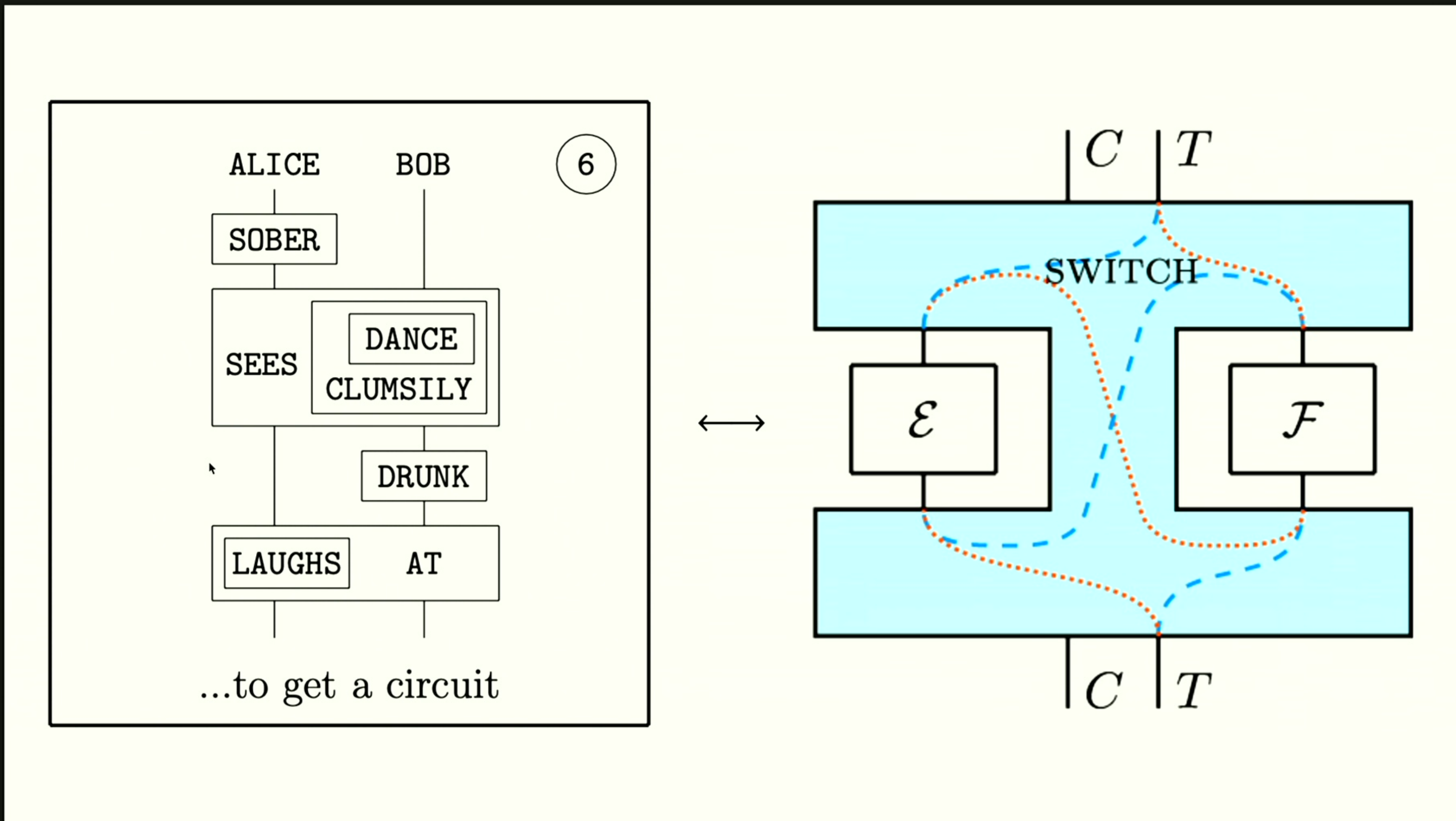
Different styles become the same!

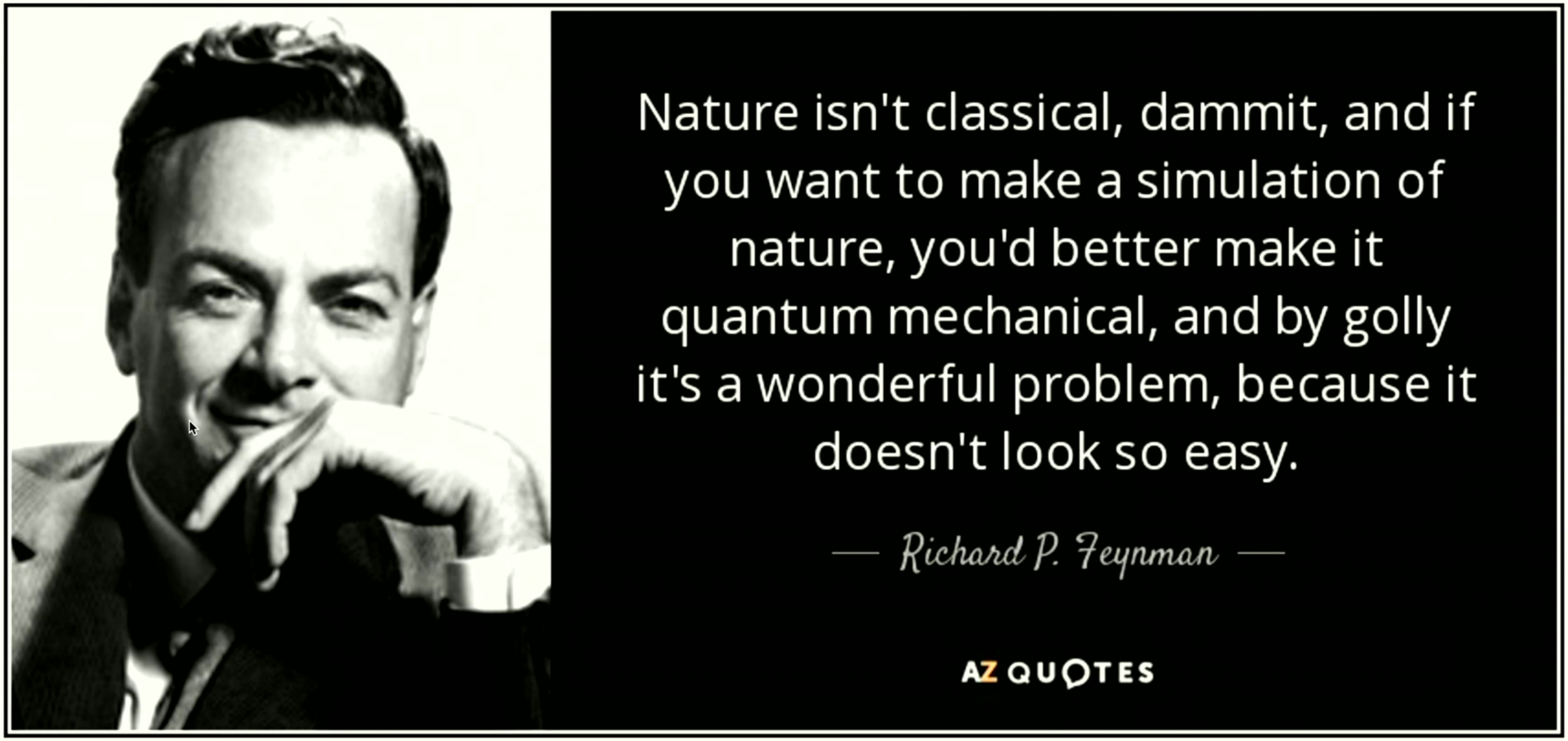
SOBER ALICE WHO
SEES DRUNK BOB
CLUMSILY DANCE
LAUGHS AT HIM.

ALICE SEES BOB DANCE CLUMSILY.
ALICE LAUGHS AT BOB.
BOB IS DRUNK.
ALICE IS SOBER.

ALICE VOIT QUE BOB DANSER MALADROITEMENT.
ALICE SE MOQUE DE BOB.
ALICE EST SOBRE.
BOB EST IVRE.



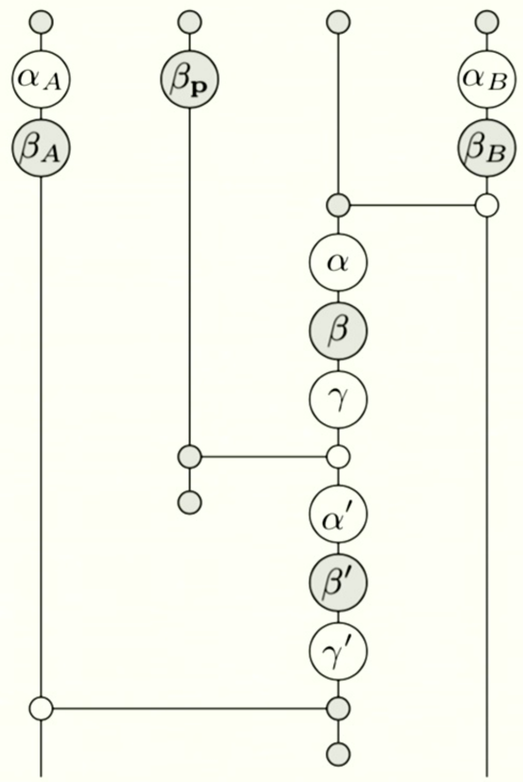

ALICE
BOB
6
SOBER
SEES
DANCE
CLUMSILY
DRUNK
LAUGHS
AT
...to get a circuit
⟷
C
T
SWITCH
$\mathcal{E}$
$\mathcal{F}$
C
T


Nature isn't classical, dammit, and if you want to make a simulation of nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it doesn't look so easy.
— Richard P. Feynman —
AZ QUOTES

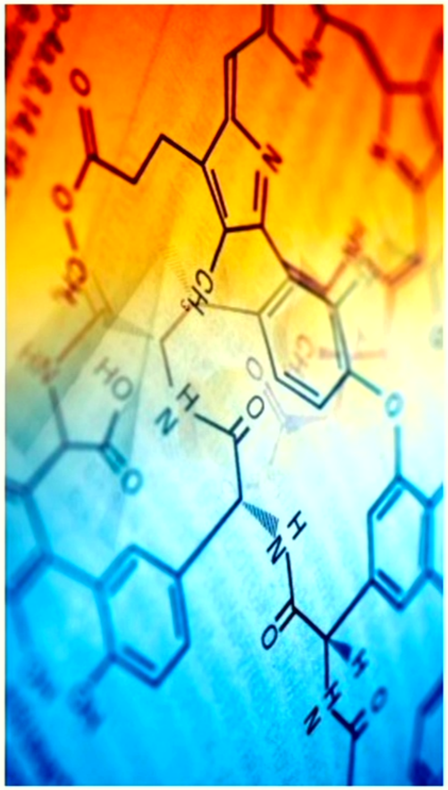
Thm. Text NLP is effectively quantum-native under the assumptions of:

- CSP for categorial grammar + text structure
- using vectors to represent meanings + CJ

text circuit (text2qcirc)



chemistry (Hamiltonian)



arXiv > quant-ph > arXiv:2408.06061

Search

Help |

Quantum Physics

[Submitted on 12 Aug 2024]

Quantum Algorithms for Compositional Text Processing

Tuomas Laakkonen (Quantinuum), Konstantinos Meichanetzidis (Quantinuum), Bob Coecke (Quantinuum)

Quantum computing and AI have found a fruitful intersection in the field of natural language processing. We focus on the recently proposed DisCoCirc framework for natural language, and propose a quantum adaptation, QDisCoCirc. This is motivated by a compositional approach to rendering AI interpretable: the behavior of the whole can be understood in terms of the behavior of parts, and the way they are put together. For the model-native primitive operation of text similarity, we derive quantum algorithms for fault-tolerant quantum computers to solve the task of question-answering within QDisCoCirc, and show that this is BQP-hard; note that we do not consider the complexity of question-answering in other natural language processing models. Assuming widely-held conjectures, implementing the proposed model classically would require super-polynomial resources. Therefore, it could provide a meaningful demonstration of the power of practical quantum processors. The model construction builds on previous work in compositional quantum natural language processing. Word embeddings are encoded as parameterized quantum circuits, and compositionality here means that the quantum circuits compose according to the linguistic structure of the text. We outline a method for evaluating the model on near-term quantum processors, and elsewhere we report on a recent implementation of this on quantum hardware. In addition, we adapt a quantum algorithm for the closest vector problem to obtain a Grover-like speedup in the fault-tolerant regime for our model. This provides an unconditional quadratic speedup over any classical algorithm in certain circumstances, which we will verify empirically in future work.

Quantum Algorithms for Text Processes

Tuomas Laakkonen, Konstantinos Meichanetzidis, Bob Coecke

Definition 1. *The problem* QUESTION-ANSWERING *is defined as follows: given a set of word embeddings V, a context text T, and a set of k question texts $\{Q_i\}$, determine any j such that*

$$\left|\operatorname{tr}\left(\rho_T(\rho_{Q_j} \otimes I)\right) - \max_i \operatorname{tr}\left(\rho_T(\rho_{Q_i} \otimes I)\right)\right| < \varepsilon$$

where $\rho_T = U_T|0\rangle\langle 0|U_T^\dagger$, $\rho_{Q_i} = U_{Q_i}|0\rangle\langle 0|U_{Q_i}^\dagger$, and U_T, U_{Q_i} are the QDisCoCirc text circuits generated from T and Q_i respectively over V.

Theorem 4. *Suppose that a set of word embeddings V satisfies the following:*

1. *The operations of V use one qubit for each input wire,*
2. *V contains arbitrarily many proper nouns,*
3. *V contains at least two adjectives that generate a dense subset of $SU(2)$,*
4. *V contains at least one transitive verb that is entangling*

then for any fixed $\varepsilon < \frac{1}{7}$, QUESTION-ANSWERING *is BQP-hard.*

arXiv > quant-ph > arXiv:2409.08777

Search

Help |

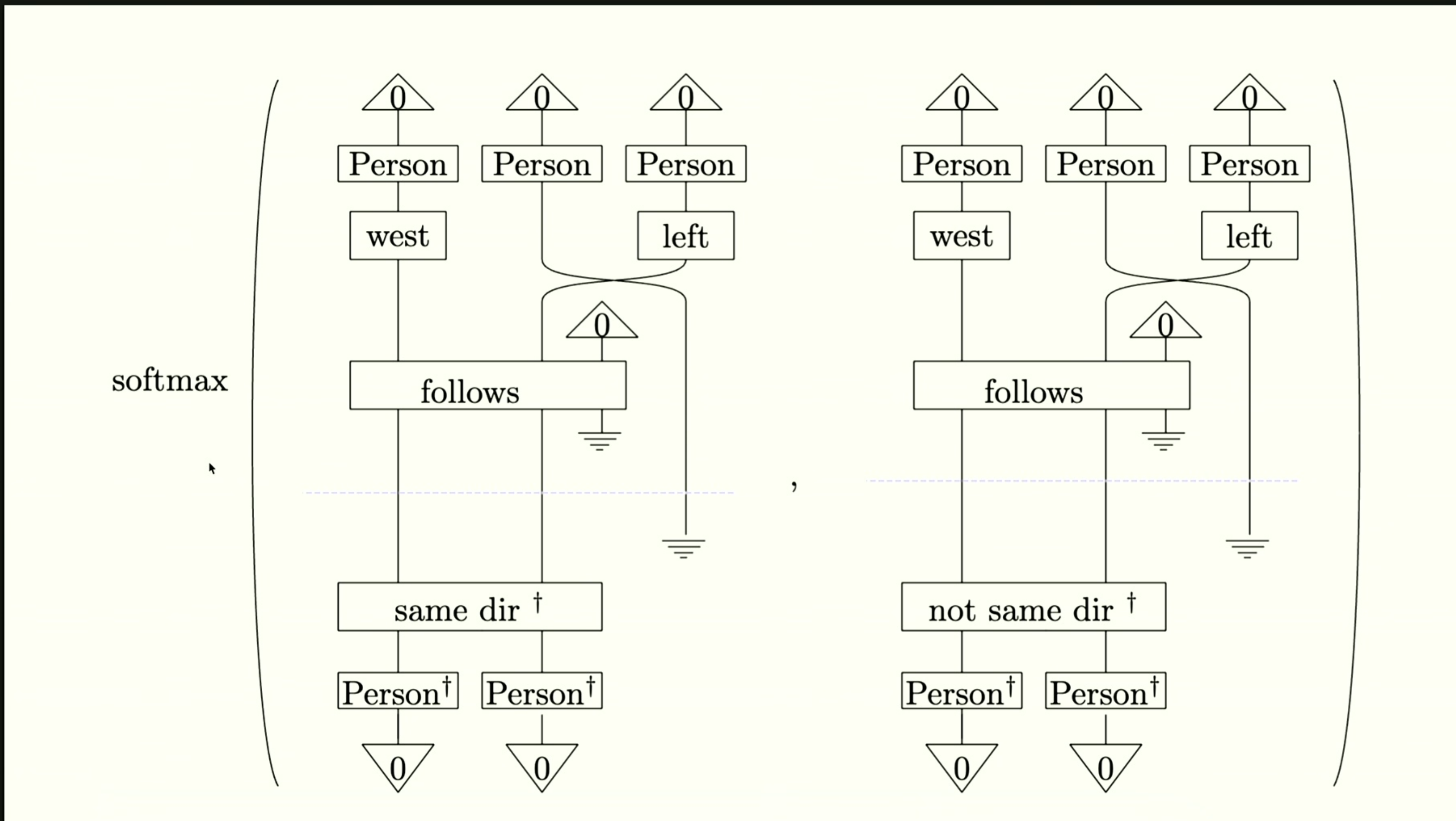
Quantum Physics

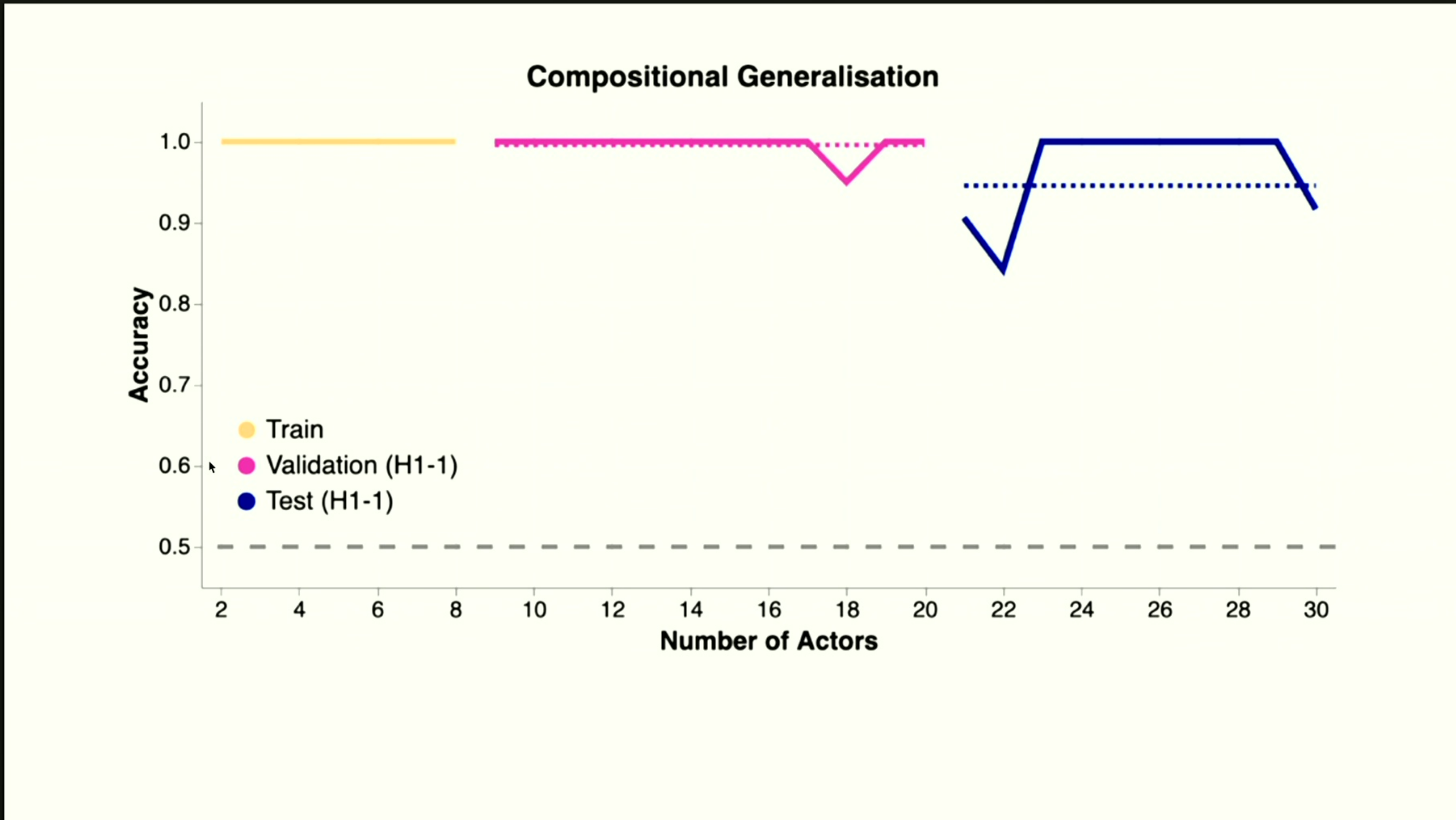
[Submitted on 13 Sep 2024]

Scalable and interpretable quantum natural language processing: an implementation on trapped ions

Tiffany Duneau, Saskia Bruhn, Gabriel Matos, Tuomas Laakkonen, Katerina Saiti, Anna Pearson, Konstantinos Meichanetzidis, Bob Coecke

We present the first implementation of text-level quantum natural language processing, a field where quantum computing and AI have found a fruitful intersection. We focus on the QDisCoCirc model, which is underpinned by a compositional approach to rendering AI interpretable: the behaviour of the whole can be understood in terms of the behaviour of parts, and the way they are put together. Interpretability is crucial for understanding the unwanted behaviours of AI. By leveraging the compositional structure in the model's architecture, we introduce a novel setup which enables 'compositional generalisation': we classically train components which are then composed to generate larger test instances, the evaluation of which asymptotically requires a quantum computer. Another key advantage of our approach is that it bypasses the trainability challenges arising in quantum machine learning. The main task that we consider is the model-native task of question-answering, and we handcraft toy scale data that serves as a proving ground. We demonstrate an experiment on Quantinuum's H1-1 trapped-ion quantum processor, which constitutes the first proof of concept implementation of scalable compositional QNLP. We also provide resource estimates for classically simulating the model. The compositional structure allows us to inspect and interpret the word embeddings the model learns for each word, as well as the way in which they interact. This improves our understanding of how it tackles the question-answering task. As an initial comparison with classical baselines, we considered transformer and LSTM models, as well as GPT-4, none of which succeeded at compositional generalisation.


softmax
0
Person
west
left
follows
same dir †
Person†
,
not same dir †


Compositional Generalisation
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
Train
Validation (H1-1)
Test (H1-1)
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Number of Actors

Text circuits aren't 'causal': meaning 'can flow backward' cf. Memento

But even quantum computers can't efficiently simulate those

What restrictions does unitarity impose on text?

How does the introduction of causal loops affects hardness?

Bottom line: there is a formal bridge that directly relates:

- **questions in quantum foundations and computation**
- **questions in compositional interpretable AI**

See some examples in our submitted abstract, but there are many more