

# The Mathematics of Text Structure

Bob Coecke

Oxford University, Department of Computer Science  
Cambridge Quantum Computing Ltd.

*Three-page ACT contribution of the paper with the same title, that has been commissioned for a volume dedicated to Lambek edited by Phil Scott and Claudia Casadio. The full version is [arXiv:1904.03478](https://arxiv.org/abs/1904.03478).*

## Abstract

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. We revisit DisCoCat: while in the latter all meanings are states (i.e. have no input), in DisCoCirc word meanings are types of which the state can evolve, and sentences are gates within a circuit which update the meaning of 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.

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background: DisCoCat</b>	<b>5</b>
2.1	diagrams . . . . .	5
2.2	From grammar to wirings . . . . .	7
2.3	Internal wirings of meanings . . . . .	8
2.3.1	Functional words . . . . .	9
2.3.2	Adjectives and <b>to be</b> . . . . .	9
2.3.3	Semi-Cartesian verbs . . . . .	10
2.4	Models of meaning . . . . .	11
2.5	Comparing meanings . . . . .	12
<b>3</b>	<b>Features and flaws of DisCoCat</b>	<b>13</b>
<b>4</b>	<b>Composing sentences: meet DisCoCirc</b>	<b>14</b>
4.1	Naive composition of sentences for DisCoCat . . . . .	14
4.2	Sentences as I/O-processes . . . . .	16
4.3	Network extraction . . . . .	19
4.4	The use of states . . . . .	20
4.5	DisCoCat from DisCoCirc . . . . .	21
4.6	Individual and subgroup meanings . . . . .	22
4.7	Example . . . . .	23
4.8	Other cognitive modes . . . . .	25
<b>5</b>	<b>Logic and language</b>	<b>27</b>
5.1	Dynamic epistemic logic from language . . . . .	27
5.2	Linear and non-linear <b>and</b> . . . . .	28

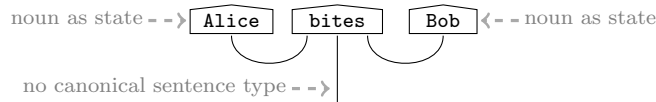
<b>6 Concrete models</b>	<b>29</b>
6.1 Sketch of a concrete model . . . . .	30
6.2 Computing text meaning . . . . .	32
6.3 Comparing texts . . . . .	32
<b>7 Physical embodiment</b>	<b>33</b>

## From States to Processes

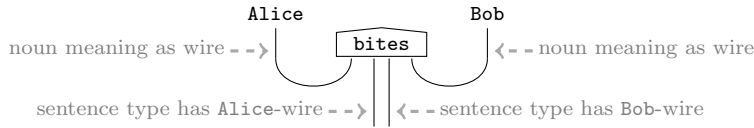
Consider the following example:

Alice is a dog.  
 Bob is a person.  
 Alice bites Bob.

Clearly, the meaning of the third sentence crucially depends on what we learn about the meaning of the nouns **Alice** and **Bob** in the first two sentences, turning **dog bites man** into **man bites dog** if **Bob** were to be a **dog** and **Alice** were to be a **person**. Also, before the 1st sentence is stated, **Alice** is just a meaningless name, and the same goes for **Bob** until the 2nd sentence is stated. So the meaning of **Alice** and **Bob** evolves as the text progresses, and it are the sentences that update our knowledge about **Alice** and **Bob**. What we propose is that the 3rd sentence, which would look like:



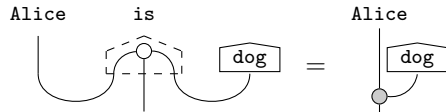
in DisCoCat, would instead be drawn like this:



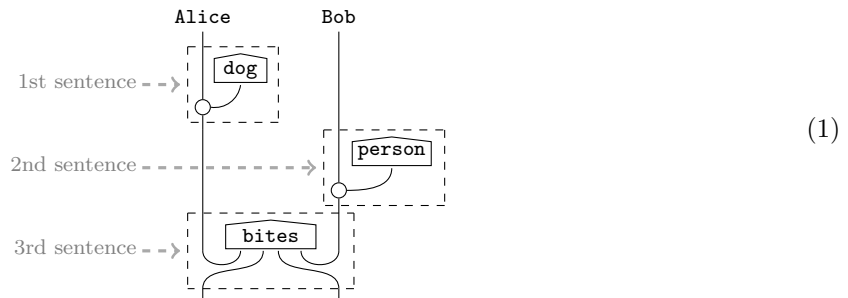
So in particular, the nouns **Alice** and **Bob** are now not states but wires (a.k.a types) and the sentence is an I/O-box with the nouns **Alice** and **Bob** both as inputs and as outputs. In this way, the sentence can act on the nouns and update their meanings. Hence:

*A sentence is not a state, but a process,*

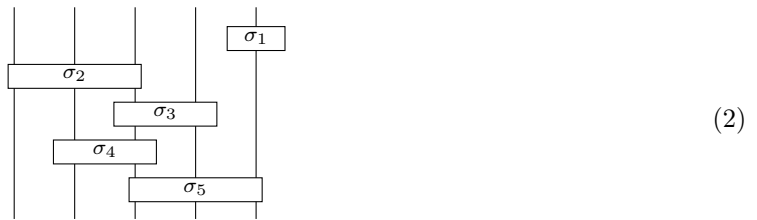
that represents how words in it are updated by that sentence. The wire-representation of **Alice is a dog** becomes:



Altogether the composition of the three sentences now becomes:



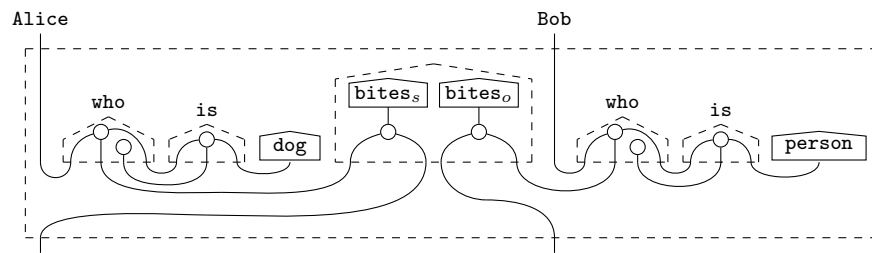
So in general, given text, we end up with a wire diagram that looks like this:



where the sentences themselves also have a wire diagram. In particular, it's a process, and this process alters our understanding of words in the text. This yields another slogan:

*Text is a process that alters our understanding of words*

The single sentence:



results in the wire-diagram (1), so diagrams can expose that different texts have the same meaning.

## Another example

