

# Quantum Algorithms for Compositional Natural Language Processing

William Zeng and Bob Coecke\*

Rigetti Computing and University of Oxford

zeng.will@gmail.com - bob.coecke@cs.ox.ac.uk

We propose a new application of quantum computing to the field of natural language processing. Ongoing work in this field attempts to incorporate grammatical structure into algorithms that compute meaning. In [5, 7], Coecke, Sadrzadeh and Clark introduce such a model (the CSC model) based on tensor product composition. While this algorithm has many advantages, its implementation is hampered by the large classical computational resources that it requires. In this work we show how computational shortcomings of the CSC approach could be resolved using quantum computation (possibly in addition to existing techniques for dimension reduction). We address the value of quantum RAM [8] for this model and extend an algorithm from Wiebe, Braun and Lloyd [24] into a quantum algorithm to categorize sentences in CSC. Our new algorithm demonstrates a quadratic speedup over classical methods under certain conditions.

## 1 Introduction

As human computer interfaces become more advanced, natural language processing has grown to be a ubiquitous part of our world. Its techniques allow computers to understand natural language to perform tasks like automatic summarization, machine translation, information retrieval, and sentiment analysis. Most approaches to this problem, such as Google’s search, understand strings of separate words in a ‘bag of words’ approach, ignoring any grammatical structure. This is certainly unsatisfactory, as we know that the meaning of a sentence is more than the meaning of its component words. Research in *distributional compositional semantics* (DisCo) seeks to address this by incorporating grammatical structure.

In [5, 7], Coecke, Sadrzadeh and Clark introduce a DisCo model (the CSC model) based on tensor product composition that gives a grammatically informed algorithm to compute the meaning of sentences and phrases. Interestingly, this algorithm was directly inspired by quantum theory, and more specifically, teleportation-like protocols [4]. While this algorithm has many advantages, its implementation is hampered by the large classical computational resources that it requires. This paper presents ways that quantum computers potentially could solve some of these problems, and doing so without the further assumptions and inaccuracies of the existing classical techniques for dimension reduction (e.g. those based on [18]). Hence making the CSC an attractive application for quantum computation.

We use the fact that quantum computation is naturally suited to managing high dimensional tensor product spaces. Recent literature has shown that quantum algorithms can thus provide advantages for machine learning [24, 20], inference [16], and regression [23, 22] tasks. We leverage these results in two particular ways:

1. We employ the scaling of quantum systems to address computational difficulties inherent in tensor-product based compositional semantics.

---

\*This work was funded by Air Force Office of Scientific Research grant FA9550-14-1-0079 and The Rhodes Trust.

2. Shared structure makes algorithms in the CSC model especially amenable to quantum speedups. We specify a CSC sentence similarity algorithm that, under certain conditions, gives quadratic speedups for natural language tasks.

In Section 2, we cover the basic framework of distributional compositional linguistics. Section 3 introduces the advantages of quantum representations for this framework. Sections 4 and 5 propose a quantum algorithm with quadratic speedup for calculating sentence similarity within CSC. Section 6 briefly discusses the noise tolerance of these methods.

## 2 Distributional Compositional Semantics and the CSS model

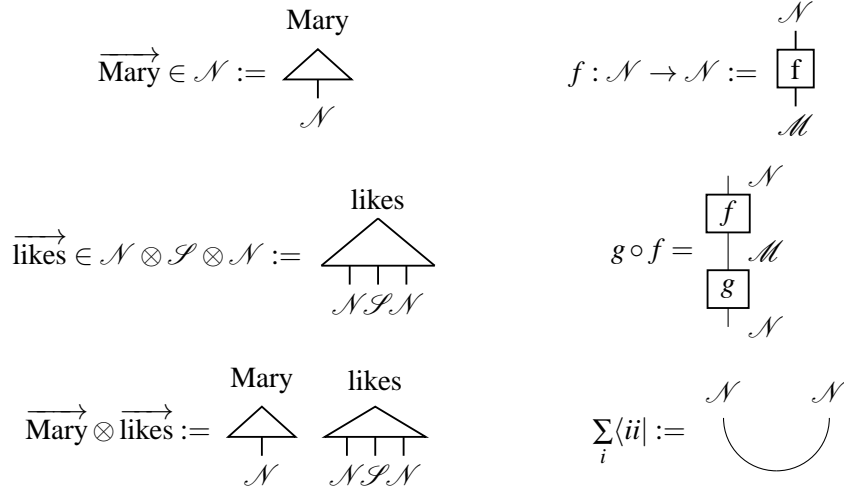
In modern natural language processing, the *vector space model* is widely used to compute the meaning of individual words [21]. In this approach we first specify a set of context words, for example the 2000 most common words in a given corpus. These context words then form the basis of the vector space of word meanings in the following manner: for some given word, say “quantum”, we look through a corpus and count the frequency with which each basis word appears ‘near’ to “quantum”. It is likely that we would have a high frequency for words like “physics” and “information” for example. These frequencies then form the *word vector* for “quantum”. Words represented by normalised vectors are similar if the inner product of their word vectors is close to one. These ‘bag of words’ methods are typically referred to as *distributional*.

As the same sentence rarely occurs repeatedly, this distributional technique cannot be directly extended to calculate the meaning of longer phrases, sentences, paragraphs, etc. Instead, *compositional semantics* designs algorithms for deriving the meaning of a sentence or phrase from known meanings of component words, taking into account types and grammatical structure [15]. The *distributional compositional semantic model* (DisCo) combines both approaches to introduce grammatical form to the composition of word vectors [5, 7].

In this model, each grammatical type is assigned a tensor product space based on Lambek’s pregroup grammar [15] or combinatorial categorical grammar [11]. The meaning of nouns is, for example, calculated as in the distributional case, and we label their vector space  $\mathcal{N}$ . A transitive verb’s meaning is then, following the grammar, a vector in the space  $\mathcal{N} \otimes \mathcal{S} \otimes \mathcal{N}$ , where  $\mathcal{S}$  is the meaning space for sentences [5, 7]. An intuition for this is that the transitive verb takes a subject noun as a left argument and an object noun as a right argument. An adjective lives in the space  $\mathcal{N} \otimes \mathcal{N}$ .

We use a diagrammatic notation for vectors, tensors, and linear maps as is common for both CSC and quantum information. Here vertical composition (read top to bottom) represents composition of linear

maps and horizontal composition represents the tensor product:



where  $f : \mathcal{N} \rightarrow \mathcal{M}$  and  $g : \mathcal{M} \rightarrow \mathcal{N}$  are linear maps and the linear map  $\sum_i \langle ii|$  sums over all the basis vectors of  $\mathcal{N}$  and is called a *cap*.

Given a well-typed sentence with meaning vectors  $\vec{w}_j$  for each of  $k$  words, the classical CSC algorithm for calculating a sentence’s meaning is [4]:

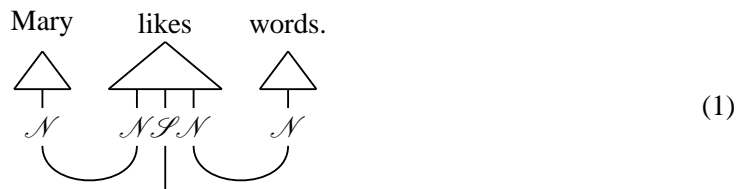
1. Compute the tensor product  $\overrightarrow{\text{words}} = \vec{w}_0 \otimes \dots \otimes \vec{w}_k$  in the order that each word appears in the sentence.
2. Construct a linear map that represents the grammatical type reduction by “wiring up” the vectors with the appropriate caps. For example, given that the pregroup type reduction of a noun/transitive verb/noun sentence is:

$$n \cdot {}^{-1}n \cdot s \cdot n^{-1} \cdot n \leq 1 \cdot s \cdot 1 \leq s$$

this linear map is:

$$\sum_i \langle ii| \otimes \text{id} \otimes \sum_i \langle ii|$$

3. Compute the meaning of the sentence by applying the linear map to the  $\overrightarrow{\text{words}}$  vector. This results in a vector in  $\mathcal{S}$  which corresponds to the meaning of the sentence. In the above example the result corresponds to the diagram:



We refer the reader to [7] for a fuller description of the distributional compositional model and to [10, 13] and [14] for experimental implementations.

These models suggest a promising approach to incorporate grammatical structure with vector space models of meaning, yet they come with the computational challenges of large tensor product spaces [9]. While there do exist some classical approaches to avoid the calculation of the full tensor product,

such as the holographic reduced representations from [18] or the use of dimensionality reduction [19], these approaches introduce further assumptions and inaccuracies. For this reason, ameliorating the large computational costs introduced these large spaces through quantum computation is of particular interest.

### 3 Quantum computation for the CSC model

The most immediate advantage for quantum implementations of the CSC model is gained by storing meaning vectors in quantum systems. For  $\alpha, \beta \in \mathbb{C}$  a two-level quantum system, a qubit, is defined by:

$$\begin{array}{l} \text{Qubit} \\ |\psi\rangle = \alpha|0\rangle + \beta|1\rangle \\ = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{array} \qquad \begin{array}{l} \text{Qubits} \\ |\psi_1\rangle \otimes |\psi_2\rangle = \begin{pmatrix} \alpha_1\alpha_2 \\ \alpha_1\beta_2 \\ \beta_1\alpha_2 \\ \beta_1\beta_2 \end{pmatrix} \end{array}$$

and composition of qubits is given by the tensor product. This leaves each  $n$ -qubit system with  $2^n$  degrees of freedom, indicating that  $N$ -dimensional classical vectors can be stored in  $\log_2 N$  qubits. Consider a corpus whose word-meaning space is given by a basis of the 2,000 most common words. Even if we make the simplifying assumption that the sentence-meaning space is no larger than the word-meaning space we obtain the dramatic improvements details in Table 1.

	One Transitive Verb	10k tr. verbs
Classical	$8 \times 10^9$ bits	$8 \times 10^{13}$ bits
Quantum	33 qubits	47 qubits

Table 1: Rough comparisons of the storage necessary for verbs in quantum and classical frameworks.

Further, these word meanings can be imported into a “bucket brigade” quantum RAM that allows them to be retrieved with a complexity linear in the number of qubits [8]. The general point is that because quantum systems compose via the tensor product they are natural choices to store complex types and sentences that have the same compositional structure. We can then employ quantum algorithms on for natural language classification as presented in Section 5.

### 4 A quantum algorithm for the closest vector problem

Many tasks in computational linguistics such as clustering, text classification, phrase/word similarity, and sentiment analysis rely on computations that determine the closest vector to  $\vec{s}$  out of some set of  $N$ -dimensional vectors  $\{\vec{v}_0, \vec{v}_1, \dots, \vec{v}_{M-1}\}$ . In clustering algorithms, for example, the set of vectors could be either the centroids of different clusters or the full set of training vectors, as in nearest neighbor clustering algorithms. Either the inner-product distance or Euclidean distance can be used. We will assume that all vectors are  $N$ -dimensional.

**Definition 4.1.** Given vector  $\vec{s}$  and a set of  $M$  vectors  $U = \{\vec{v}_0, \vec{v}_1, \dots, \vec{v}_{M-1}\}$  the *closest vector problem* asks one to determine which  $v_j$  has the smallest inner product distance with  $\vec{s}$ .

Direct calculation of the smallest vector would have complexity  $\mathcal{O}(MN)$ . In [24] the authors introduce a quantum algorithm for this problem that, under certain conditions, demonstrate quadratic speedups over direct calculation and polynomial speedups over Monte-Carlo methods. Some proof

of principle experiments have then demonstrated clustering of eight-dimensional vectors, based on these techniques, on a small photonic quantum computer [3]. This algorithm requires the following assumptions, where we write  $v_{ji}$  for the  $i^{\text{th}}$  entry of the  $j^{\text{th}}$  vector:

1. Vectors  $\vec{v}_j$  and  $\vec{s}$  are  $d$ -sparse, with no more than  $d$  non-zero entries.
2. The relevant vectors are stored in some kind of quantum memory so that the quantum computer can access their entries with the two oracles of the form:

$$\begin{aligned} \mathcal{O}|j\rangle|i\rangle|0\rangle &:= |j\rangle|i\rangle|v_{ji}\rangle, \\ \mathcal{F}|j\rangle|l\rangle &:= |j\rangle|f(j,l)\rangle, \end{aligned} \quad (2)$$

where  $f(j,l)$  is the location of the  $l^{\text{th}}$  non-zero entry of  $v_j$ . It is against these memory access oracles that the performance of our algorithm will be measured.

3.  $\max(|v_{ji}|^2) \leq r_{\max}$  for some known constant  $r_{\max}$ .
4. All the vectors are normalized.

Under these assumptions we are able to run a quantum nearest-neighbor algorithm with complexity characterized by the following theorem:

**Theorem 4.2** ([24]). We can find  $\max_j |\langle s | v_j \rangle|^2$  with success probability  $1 - \delta$  and error  $\varepsilon$  using an expected number of  $\mathcal{O}$  and  $\mathcal{F}$  queries that is bounded above by

$$1080\sqrt{M} \left\lceil \frac{4\pi(\pi+1)d^2 r_{\max}^4}{\varepsilon} \right\rceil \left\lceil \frac{\ln(81M(\ln(M) + \gamma)) / \delta_0}{2(8/\pi^2 - 1/2)^2} \right\rceil, \quad (3)$$

where  $\gamma \approx 0.5772$  is Euler’s constant.

It is clear that for this quantum algorithm there is a quadratic improvement in scaling with  $M$ , the number of training vectors. While the dimension of the vectors  $N$  is not explicitly included, in general it is implicitly there through the dependence on  $d$ . It is also clear that if  $r_{\max} \propto 1/\sqrt{d}$ , then the algorithm’s dependence on both  $d$  and  $N$  drops out. As the vectors are normalized, this can be expected to occur if the vectors have sparsity that grows linearly with their size [24]. The authors further assume that for “typical” cases the error  $\varepsilon$  scales as  $\Theta(1/\sqrt{N})$  so that the runtime for the quantum inner-product algorithm becomes  $\mathcal{O}(\sqrt{NM} \ln(M) d^2 r_{\max}^4)$ .<sup>1</sup> This result shows a quadratic improvement over direct calculations and also shows improvement over Monte Carlo methods, whose complexity is  $\mathcal{O}(NMd^2 r_{\max}^4)$ . These comparisons are summarized in Table 2.

Type	Typical cases	Atypical cases
Classical Direct	$\mathcal{O}(NM)$	$\mathcal{O}(NM)$
Classical Monte Carlo	$\mathcal{O}(NMd^2 r_{\max}^4)$	$\mathcal{O}(Md^2 r_{\max}^4 / \varepsilon^2)$
Quantum	$\mathcal{O}(\sqrt{NM} \log(M) d^2 r_{\max}^4)$	$\mathcal{O}(\sqrt{M} \log(M) d^2 r_{\max}^4 / \varepsilon)$

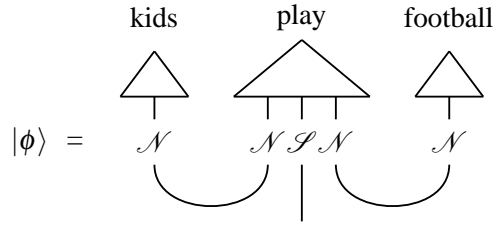
Table 2: Complexity comparisons for different closest vector algorithms. Adapted from [24].

In the following section we adapt this algorithm to sentence similarity calculations in the distributional compositional framework.

<sup>1</sup> This is argued for in Appendix D of [24] for a “typical” case where the vectors are uniformly distributed over the unit sphere.

## 5 A quantum algorithm for CSC sentence similarity

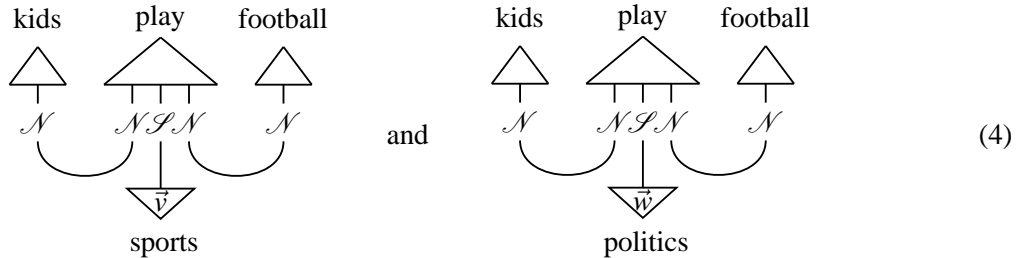
The quantum algorithm from Section 4 can be used to advantage in natural language processing tasks however, the computational overhead of the CSC approach would dwarf this algorithm's advantages if it were naively applied. Throughout this section we will assume both that  $r_{\max} \propto 1/\sqrt{d}$  and that the accuracy necessary for our calculation means  $\varepsilon$  scales as  $\Theta(1/\sqrt{N})$ . Consider the example of probabilistically classifying the meaning of a simple sentence. We illustrate this example with a noun-verb-noun sentence. The meaning of the nouns are vectors in an  $N$ -dimensional vector space and the meaning of the verb is a vector in the space  $\mathcal{N} \otimes \mathcal{S} \otimes \mathcal{N}$ . We represent a derivation of the meaning of the full sentence with the following map:



From now on, we will take sentences to exist in the same meaning space as words, i.e.  $\mathcal{S} \simeq \mathcal{N}$ .

**Definition 5.1.** For meaning vector  $\vec{s}$  and  $M$  sets of meaning vectors, a *classification task* assigns  $\vec{s}$  to the set containing the nearest-neighbor of  $\vec{s}$ .

An example task would be to determine if a sentence is about sports or politics or if a sentence expresses agreement or disagreement. If, to present a simplified example, we take each cluster to only contain a single vector ( $\vec{v}$  and  $\vec{w}$  respectively) then the sentence would be classified by computing



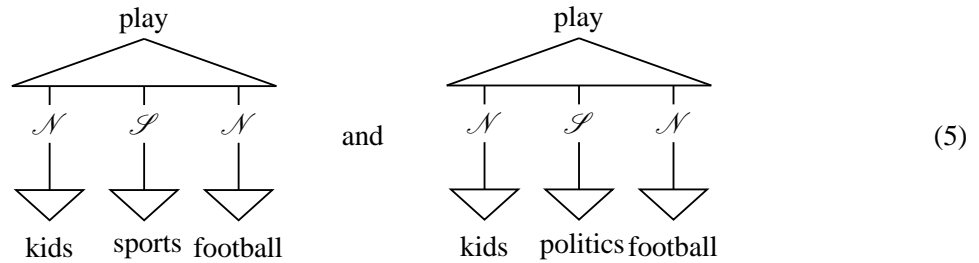
and assigning the sentence to the one of smaller value. We would proceed with two steps:

1. Compute the derivation of  $|\psi\rangle$ , which, by classical direct calculation, takes  $\mathcal{O}(3N)$  operations.
2. See which of  $\vec{v}$  and  $\vec{w}$  is closest to  $|\phi\rangle$ . This is an instance of the closest vector problem where  $\vec{s} = |\phi\rangle$ ,  $M = 2$ , and  $U = \{\vec{v}, \vec{w}\}$ . With direct calculation or Monte Carlo the second step requires<sup>2</sup>  $\mathcal{O}(2N)$  to be compared with the quantum method at  $\mathcal{O}(\sqrt{2N} \log 2)$ . Even if we include the step to import the classical data from step one into quantum form, which can be done with  $\mathcal{O}(\log_2 N)$  overhead [8], then we obtain a speedup for this step.

Still, despite the quantum speedup from step two, the full algorithm for general  $M$  runs in  $\mathcal{O}(3N\sqrt{M} \log M)$ , remaining linear in  $N$ .

<sup>2</sup>If we assume the appropriate  $d$ -sparsity scaling.

In order to recover a speedup we refine the application of the quantum algorithm by posing a version of the closest vector problem that avoids the initial calculation of  $|\phi\rangle$  altogether. Note the equivalence of the calculations in Equation 4 with



Rather than directly calculating  $|\phi\rangle$ , which is not relevant to the classification task, we can formulate a closest vector problem where  $\vec{s} = |play\rangle$ ,  $M = 2$  and  $U = \{|kids\rangle \otimes |v\rangle \otimes |football\rangle, |kids\rangle \otimes |w\rangle \otimes |football\rangle\}$ . The runtime of this *deferred quantum algorithm*, including import, will then be  $\mathcal{O}(\sqrt{MN})$ , showing our desired quadratic speedup in both variable.

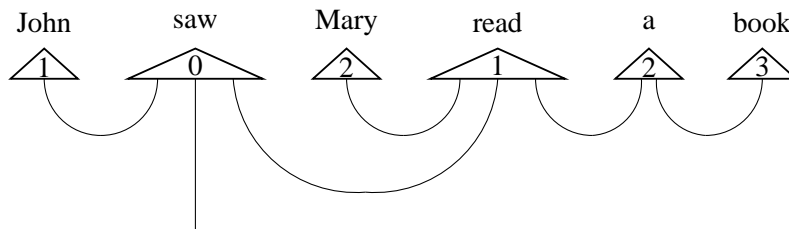
We extend this to result to include general sentences in the CSC model with the following theorem.

**Theorem 5.2.** For an  $N$ -dimensional noun meaning space, there exists a quantum algorithm to classify any CSC model sentence composed of  $n$  tensors  $\vec{w}_0, \vec{w}_1, \dots, \vec{w}_{n-1}$  into  $M$  classes with time  $\mathcal{O}(\sqrt{MN} \log M)$ . This improves on classical methods that require  $\mathcal{O}(NM)$ .

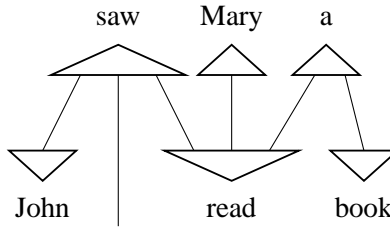
*Proof.* The trick we play in Equation 5 amounts to splitting the sentence derivation into a bipartite graph. As the CSC connections are based on a pregroup derivation, the connections will always form a tree, taking words as nodes and connections as edges. Trees can always be partitioned into bipartite graphs, thus, up to the ordering of inputs on each tensor which can be kept track of, we can always give a deferred quantum algorithm with associated speedup for any such CSC sentence. The following procedure explicitly details how to construct this bipartite partitioning.

For every CSC sentence there is one word that acts as the *head* of the derivation. This is the word whose output  $S$  wire contains the sentence meaning following its derivation’s linear map. In Equation 5 this is the word “play”. Connect the dangling wire of the head word  $\vec{w}_h$  with the vector  $\vec{v}_i$  against which similarity is being measured. Starting with this head word we then separate the sentence into a top layer and a bottom layer with the following steps. Assign the head word to the top layer. Place every word it is connected to on the bottom layer. Next for every word just assigned to the bottom, take all their connected words, which are not yet assigned, and assign them to the top. Continue this procedure while alternating top and bottom until all words are assigned. This gives a simple two-coloring of the derivation graph.  $\square$

**Example 5.3.** Consider the following sentence [12]:



where we have labeled the vectors based on their depth in the derivation tree. The two-layer form assigns even vectors to the top layer and odd vectors to the bottom:



Hooking the dangling wire up to a classifying vector reduces the similarity computation to the calculation of a single inner product. Note that this procedure works for any derivation tree,<sup>3</sup> so sentence fragments, such as noun phrases, can be easily analyzed in exactly the same manner.

## 6 Noise tolerance and Conclusion

To reap the benefits of quantum algorithm in the domain of natural language processing, we apply a new technique to defer the calculation of a sentence's compositional meaning, eliminating the overhead costs. By this method we are able to introduce a quantum algorithm for calculating sentence similarity that offers quadratic speedup over classical direct calculation and Monte-Carlo methods. These kinds of algorithms are particularly attractive for practical applications of quantum computing as noisy results can be tolerated: in our case when the desired error is lower bounded by  $1/\sqrt{N}$ . Vector space models are already inherently noisy and typical tasks allow for errors in results, so this restriction does not affect their efficacy.

An additional point is that the density matrix formalism of [17, 2] can also be naturally modeled by mixed states of quantum systems. In fact, this analogy was the genesis for the theory of disambiguation presented there, as another example of the shared structure that led to the results presented here. At a basic level, our work exploits the abstract connection between natural language processing and quantum information. More formally, we can see both quantum computation in the category of finite dimensional Hilbert spaces and linear maps [1, 6] and CSC in the product category of pregroup grammar and finite dimensional vector spaces [7]. The connection between these two (as dagger compact categories) makes the application of one to the other apparent.

## References

- [1] Samson Abramsky & Bob Coecke (2004): *A categorical semantics of quantum protocols*. In: *Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on*, IEEE, pp. 415–425, doi:10.1109/LICS.2004.1319636.
- [2] Dea Bankova, Bob Coecke, Martha Lewis & Dan Marsden (2016): *Graded Entailment for Compositional Distributional Semantics*. arXiv:1601.04908.
- [3] X-D Cai, D Wu, Su Z-E, M-C Chen, Wang X-L, Li Li, N-L Liu, C-Y Lu & J-W Pan (2015): *Entanglement-Based Machine Learning on a Quantum Computer*. *Physical Review Letters* 114(11), p. 110504, doi:10.1103/PhysRevLett.114.110504.

<sup>3</sup>Even non-pregroup and non-CCG models will work as long as there is some tree derivation.



- [4] Stephen Clark, Bob Coecke, Edward Grefenstette, Stephen Pulman & Mehrnoosh Sadrzadeh (2014): *A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure*. *Malaysian Journal of Mathematical Sciences* 8, pp. 15–25. arXiv:1305.0556.
- [5] Stephen Clark, Bob Coecke & Mehrnoosh Sadrzadeh (2008): *A compositional distributional model of meaning*. In: *Proceedings of the Second Quantum Interaction Symposium (QI-2008)*, pp. 133–140.
- [6] Bob Coecke & Aleks Kissinger (2016): *Picturing Quantum Processes. A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press. To appear.
- [7] Bob Coecke, Mehrnoosh Sadrzadeh & Stephen Clark (2010): *Mathematical foundations for a compositional distributional model of meaning*. *Linguistic Analysis — A Festschrift for Jim Lambek* 36. arXiv:1003.4394.
- [8] Vittorio Giovannetti, Seth Lloyd & Lorenzo Maccone (2008): *Quantum random access memory*. *Physical Review Letters* 100(16), p. 160501, doi:10.1103/PhysRevLett.100.160501.
- [9] Edward Grefenstette (2013): *Category-Theoretic Quantitative Compositional Distributional Models of Natural Language Semantics*. Ph.D. thesis. Available at <http://arxiv.org/abs/1311.1539>.
- [10] Edward Grefenstette & Mehrnoosh Sadrzadeh (2011): *Experimental Support for a Categorical Compositional Distributional Model of Meaning*. In: *The 2014 Conference on Empirical Methods on Natural Language Processing*., pp. 1394–1404.
- [11] Karl Moritz Hermann & Phil Blunsom (2013): *The Role of Syntax in Vector Space Models of Compositional Semantics*. In: *In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 894–904.
- [12] Dimitri Kartsaklis (2015): *Compositional Distributional Semantics with Compact Closed Categories and Frobenius Algebras*. Ph.D. thesis, University of Oxford.
- [13] Dimitri Kartsaklis & Mehrnoosh Sadrzadeh (2013): *Prior disambiguation of word tensors for constructing Sentence vectors*. In: *The 2013 Conference on Empirical Methods on Natural Language Processing*., ACL, pp. 1590–1601.
- [14] Dimitri Kartsaklis, Mehrnoosh Sadrzadeh & Stephen Pulman (2012): *A unified sentence space for categorical distributional-compositional semantics: Theory and experiments*. In: *Proceedings of COLING: Posters*, pp. 549–558.
- [15] Joachim Lambek (2008): *From word to sentence*. *Polimetrica*, Milan.
- [16] Guang Hao Low, Theodore J Yoder & Isaac L Chuang (2014): *Quantum inference on Bayesian networks*. *Physical Review A* 89(6), p. 062315, doi:10.1103/PhysRevA.89.062315.
- [17] Robin Piedeleu, Dimitri Kartsaklis, Bob Coecke & Mehrnoosh Sadrzadeh (2015): *Open System Categorical Quantum Semantics in Natural Language Processing*. In: *Proceedings of the 6th Conference on Algebra and Coalgebra in Computer Science (CALCO)*, Nijmegen, Netherlands, doi:10.4230/LIPIcs.CALCO.2015.270.
- [18] Tony Plate (1991): *Holographic Reduced Representations: Convolution Algebra for Compositional Distributed Representations*. In: *International Joint Conference on Artificial Intelligence*, pp. 30–35.
- [19] Tamara Polajnar, Luana Fagarasan & Stephen Clark (2013): *Learning type-driven tensor-based meaning representations*. arXiv:1312.5985.
- [20] Patrick Rebentrost, Masoud Mohseni & Seth Lloyd (2014): *Quantum support vector machine for big data classification*. *Physical Review Letters* 113(13), p. 130503, doi:10.1103/PhysRevLett.113.130503.
- [21] Hinrich Schütze (1998): *Automatic word sense discrimination*. *Computational Linguistics* 24(1), pp. 97–123.
- [22] Guoming Wang (2014): *Quantum Algorithms for Curve Fitting*. arXiv:1402.0660.
- [23] Nathan Wiebe, Daniel Braun & Seth Lloyd (2012): *Quantum algorithm for data fitting*. *Physical Review Letters* 109(5), p. 050505, doi:10.1103/PhysRevLett.109.050505.
- [24] Nathan Wiebe, Ashish Kapoor & Krysta Svore (2015): *Quantum nearest-neighbor algorithms for machine learning*. *Quantum Information & Computation* 15(3 & 4), pp. 0318–0358.