A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure

Stephen Clark[1], Bob Coecke[*2], Edward Grefenstette[2], Stephen
Pulman[2], and Mehrnoosh Sadrzadeh[3]

[1]*Computer Laboratory, University of Cambridge*
[2]*Department of Computer Science, University of Oxford*
[3]*School of Electronic Engineering and Computer Science, Queen
Marry, London*

E-mail: coecke@cs.ox.ac.uk
[*]*Corresponding author*

ABSTRACT

We discuss an algorithm which produces the meaning of a sentence given
meanings of its words, and its resemblance to quantum teleportation. In
fact, this protocol was the main source of inspiration for this algorithm
which has many applications in the area of Natural Language Processing.

Quantum teleportation (Bennett *et al.*, 1993) is one of the most conceptually
challenging and practically useful concepts that has emerged from the quantum
information revolution. For example, via logic-gate teleportation (Gottesman
and Chuang, 1999) it gave rise to the measurement-based computational model,
it also plays a key role in current investigations into the nature of quantum
correlations, e.g. Skrzypczyk *et al.* (2009), and it even has been proposed as
a model for time travel (Bennett and Schumacher, 2002). It also formed the
cornerstone for a new axiomatic approach and diagrammatic calculus for quan-
tum theory (Abramsky and Coecke, 2004; Coecke, 2010; Coecke and Duncan,
2011).

Stephen Clark, Bob Coecke, Edward Grefenstette, Stephen Pulman and Mehrnoosh
Sadrzadeh

Arguably, when such a radically new concept emerges in a novel founda-tional area of scientific investigation, one may expect that the resulting concep-tual and structural insights could also lead to progress in other areas, something which has happened on many occasions in the history of physics. In the context of quantum information, for example, it is well-known that quantum complexity theory has helped solve many problems in classical complexity theory.

Here we explain how a high-level description of quantum teleportation with emphasis on information flows has successfully helped solve a longstanding open problem in the area of Natural Language Processing (NLP), and the problem of modeling meaning for natural language more generally (Clark *et al.*, 2008; Coecke *et al.*, 2010). This work featured as a cover heading in the New Scientist (11 Dec. 2011) (Aron, 2010), and has been experimentally tested for its capability to perform key NLP tasks such as word sense disambiguation in context (Grefenstette and Sadrzadeh, 2011).[1]

1.   The NLP Problem

Dictionaries explain the meanings of words; however, in natural language words are organized as sentences, but we don't have dictionaries that explain the meanings of sentences. Still, a sentence carries more information than the words it is made up from; e.g. meaning(Alice sends a message to Bob) $\neq$ meaning(Bob sends a message to Alice). Evidently, this is where grammatical structure comes into play. Consequently, we as humans must use some algo-rithm that converts the meanings of words, via the grammatical structure, into the meaning of a sentence. All of this may seem to be only of academic interest; however, search engines such as Google face exactly the same challenge. They typically read a string of words as a 'bag of words', ignoring the grammatical structure. This is simply because (until recently) there was no mathematical model for assigning meanings to sentences.[2] On the other hand, there is a widely used model for word meaning, the *vector space model* (Schütze, 1998).

This vector space model of word meaning works as follows. One chooses a set of context words which will form the basis vectors of a vector space.[3] Given a word to which one wishes to assign meaning,e.g. 'Alice', one relies on

---

[1]EMNLP is the leading conference on corpus-based experimental NLP.

[2]More precisely, there was no mathematical model for assigning meanings to sentences that went beyond truthfulness. Montague semantics (Thomason, 1974) is a compositional model of meaning, but at most assigns truth values to sentences, and evidently there is more to sentence meaning than the mere binary assignment of either true or false.

[3]These context words may include nouns, verbs etc.; the vector space model built from the British National Corpus typically contains 10s of thousands of these words as basis vectors.

a large corpus, e.g. (part of) the web, to establish the relative frequency that 'Alice' occurs 'close' to each of these basis words. The list of all these relative frequencies yields a vector that represents this word, its *meaning vector*. Now, if one wants to verify synonymy of two words, it suffices to compute the inner-product of the meaning vectors of these words, and verify how close it is to 1. Indeed, since synonyms are interchangeable, one would expect them to typically occur in the context of the same words, and hence their meaning vectors should be the same in the statistical limit. For example, in a corpus mainly consisting of computer science literature, one would expect Alice and Bob to always occur in the same context and hence their meaning vectors would almost be the same. Of course, if the corpus were English literature (cf. Carroll (1865)), then this similarity would break down.

Until recently, the state of affairs in computational linguistics was one of two separate communities (Gazdar, 1996). One community focused on *non-compositional* purely *distributional* methods such as the vector space model described above. The other community studied the *compositional* mathematical structure of sentences, building on work by Chomsky (1957), Lambek (1958) and Montague (Thomason, 1974). This work is mainly about the grammatical structure of sentences; *grammatical type calculi* are algebraic gadgets that allow one to verify whether a sentence has a correct grammatical structure.

2.   Caps, cups, and teleportation

In Abramsky and Coecke (2004), a novel axiomatic framework was proposed to reason about quantum informatic processes, which admits a sound and faithful purely diagrammatic calculus (Coecke, 2010); for some more recent developments we refer to Coecke and Duncan (2011). Ideal post-selected teleportation provides the cornerstone for the diagrammatic reasoning techniques, e.g. here is the derivation of the general teleportation protocol where the f-label represents both the measurement outcome and the corresponding correction performed by Bob Coecke (2010):
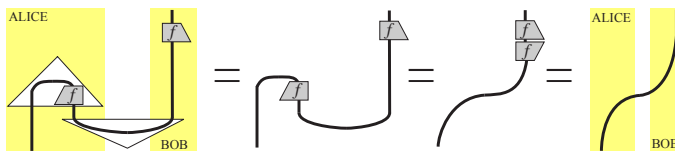
The main conceptual idea behind these diagrams is that, besides their operational physical meaning, they also admit a 'logical reading' in terms of *in-*

formation flow:
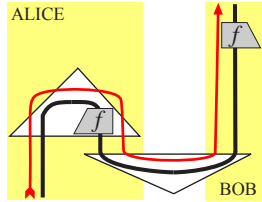
Here, the red line (color only for online articles) represents the logical flow which indicates that the state incoming at Alice's side first gets acted upon by an operation f, and then by its adjoint $f^\dagger$, which in the case that f is unitary results in the outgoing state at Bob's side being identical to the incoming one at Alice's side. [4]

When interpreted in Hilbert space, the key ingredients of this formalism are 'cups' and 'caps':

 $:= |00\rangle + |11\rangle$ $:= \langle 00| + \langle 11|$

and the equation that governs them is:

$$\big((\langle 00| + \langle 11|) \otimes Id \big)\big(Id \otimes (|00\rangle + |11\rangle) \big) = Id$$

which diagrammatically depicts as:

In this language the Choi-Jamiolkowski isomorphism:

$$|\Psi\rangle = \quad \quad = \quad \quad = (Id \otimes f)(|00\rangle + |11\rangle)$$

---

[4]This 'logical reading' of projectors on entangled states in terms of information flow was first proposed by one of the authors in Coecke (2004).

interprets a bipartite state (the grey triangle) as a 'cup' which changes the direction of the information flow together with an operation f that alters the information.

Non-separatedness means topological connectedness:

$$|\Psi\rangle = \qquad \neq \qquad = |\psi\rangle \otimes |\phi\rangle$$
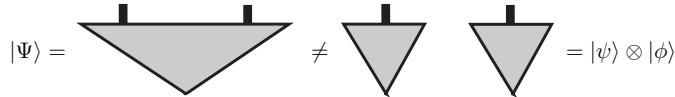
which is interpreted as the fact that information can flow between the two systems involved. Hence, when focussing on pure states, the cups effectively witness entanglement in terms of the information flows that it enables.

It is exactly this interpretation of the vectors representing the states of compound quantum systems in terms of enabling information flows that will provide the cornerstone for our compositional and distributional model of meaning.

## 3. Solution to the NLP problem: The intuition

Before we explain the precise algorithm that produces sentence meaning from word meaning, we provide the analogy with the above.

A transitive verb requires both an object and a subject to yield a grammatically correct sentence. Consider the sentence "Alice hates Bob". Assume that the words in it are represented by vectors, which as above we denote by triangles:

$$\overrightarrow{Alice} \otimes \overrightarrow{hates} \otimes \overrightarrow{Bob} = \quad Alice \quad hates \quad Bob$$

Note here that treating verbs as 'compound' was already the case in grammatical type calculi, as we discuss below. So how do these words interact to produce the meaning of a sentence? For the verb to produce the meaning of the sentence, that is, the statement of the fact that Alice hates Bob, it of course needs to know what its subject and object are, that is, it requires knowing their meanings. Therefore, inspired by the above discussion on teleportation, we 'feed' the meaning vectors $\overrightarrow{Alice}$ and $\overrightarrow{Bob}$ into the verb $\overrightarrow{hates}$ which then

'spits out' the meaning of the sentence:
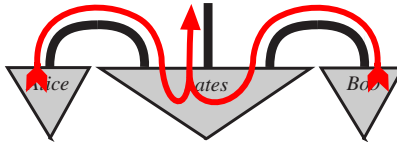
Again, that the meaning of the sentence is produced by the transitive verb after interacting with its nouns is also something that was the case in grammatical type calculi.

In the same vein, for an intransitive verb, we obtain an even more direct analogue to quantum teleportation:

$$\left(\sum_i \langle ii| \otimes Id \right) (\overrightarrow{Alice} \otimes \overrightarrow{dreams}) =$$

Note here that non-separatedness of verbs is obvious: if in the sentence "Alice hates Bob" hates would be disconnected, then the meaning of the sentence would not depend on the meanings of "Alice" and "Bob", so, "Anyone hates everyone"!

4. A grammatical type calculus: Lambek's pregroups

In order to give a precise description of our algorithm we now give a brief account of Lambek's pregroup grammar (Lambek, 1999, 2008).[5]

Pregroups capture structural similarities across a wide range of language families (Lambek, 2008). They combine a remnant of group structure with partial ordering; the usual (left and right) group laws for the inverse are replaced by four inequalities involving distinct left and right pseudo-inverses x^{-1}

---

[5]An interesting aside: Lambek published his paper on the widely used Lambek grammars in 1958 (Lambek, 1958). His recent book on pregroups appeared 50 years later (Lambek, 2008)! All this time Lambek worked in Montreal, the location of the upcoming QIP, and it was also in Montreal (in 2004) that during a talk by one of us he first pointed to the structural coincidence between pregroup grammars and quantum axiomatics in terms of cups and caps.

and ^{-1}x:

$$x^{-1} \cdot x \leq 1 \leq x \cdot x^{-1} \qquad x \cdot {}^{-1}x \leq 1 \leq {}^{-1}x \cdot x\,.$$

As a grammatical type calculus, its elements are basic grammatical types, e.g. the noun type n and sentence type s. Other types arise from the pseudo-inverses and the multiplication, e.g. the transitive verb type $tv := {}^{-1}n \cdot s \cdot n^{-1}$. Then:

$$\overbrace{n}^{e.g.\ Alice} \cdot \overbrace{{}^{-1}n \cdot s \cdot n^{-1}}^{e.g.\ hates} \cdot \overbrace{n}^{e.g.\ Bob} \;=\; (n \cdot {}^{-1}n) \cdot s \cdot (n^{-1} \cdot n)$$
$$\leq\; 1 \cdot s \cdot 1$$
$$=\; s$$

Such an inequality $n \cdot tv \cdot n \leq s$ then stands for the fact that "noun – transitive verb – noun" is a valid grammatical structure for a sentence. Note here the correspondence with our interpretation of such a sentence in terms of information flow: the verb requires two nouns to be 'fed into it' to yield a sentence type; n^{-1} and ^{-1}n capture 'the verb requests type n on the left/right'. In fact, the inequalities using $n \cdot {}^{-1}n \leq 1$ and $n^{-1} \cdot n \leq 1$ can also be represented with 'directed' caps:

$$\overset{\displaystyle s}{\underset{\displaystyle n \cdot {}^{-1}n \cdot s \cdot n^{-1} \cdot n}{\frown\ |\ \frown}} \tag{1}$$

which represent the inequalities:

$$\frown \quad \leftrightarrow \quad \overset{\displaystyle 1}{\underset{\displaystyle n \cdot {}^{-1}n}{\text{VI}}} \qquad\qquad \frown \quad \leftrightarrow \quad \overset{\displaystyle 1}{\underset{\displaystyle n^{-1} \cdot n}{\text{VI}}}$$

Moreover, a pregroup can be defined in terms of cups and caps. In category theoretic language, both the diagrammatic language for quantum axiomatics and pregroups are so-called *compact closed categories* (Kelly and Laplaza, 1980; Selinger, 2011); while the quantum language is *symmetric*, pregroups have to be *non-symmetric* given the importance of word-order in sentences.

5. Solution to the NLP problem: The algorithm

Assume a grammatically well-typed sentence and a meaning vector $\vec{v}_j$ for each of its words, which we assume to be represented in a vector space of which the tensor structure matches the structure of its grammatical type,[6] e.g.:

$$n \rightsquigarrow \mathcal{V} \qquad tv = {}^{-1}n \cdot s \cdot n^{-1} \rightsquigarrow \mathcal{V} \otimes \mathcal{W} \otimes \mathcal{V}$$

where $\mathcal{W}$ is the vector space in which we intend to represent the meanings of sentences. Then one proceeds as follows:

1. Compute the tensor product $\overrightarrow{Words} = \vec{v}_1 \otimes \ldots \otimes \vec{v}_k$ of the word meaning vectors in order of appearance in the sentence; e.g. $\overrightarrow{noun_1} \otimes \overrightarrow{verb} \otimes \overrightarrow{noun_2}$.

2. Construct a linear map f that represents the type reduction as follows: given the diagram that represents a type reduction (cf. (1) above), we interpret caps as $\sum_i \langle ii|$ and straight wire as identities; e.g. $\sum_i \langle ii| \otimes Id \otimes \sum_i \langle ii|$.

3. Compute $\overrightarrow{Sentence} := f(\overrightarrow{Words}) \in \mathcal{W}$.

Hence the crux is: *the grammatical correctness verification procedure becomes an actual linear map that transforms the meanings of words into the meaning of the sentence by making these words interact via caps.* Does it work? The proof is in the pudding. Proof-of-concept examples are in Coecke *et al.* (2010), and concrete experimentally verified applications are in Grefenstette and Sadrzadeh (2011).
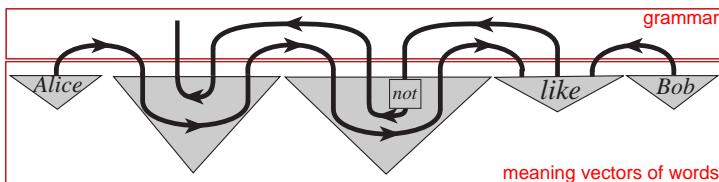
We invite the reader to also look at Coecke *et al.* (2010) for the example sentence:

"Alice does not like Bob" ,

---

[6] How this can be achieved within the context of the vector space model of meaning is outlined in Grefenstette *et al.* (2010) and used in Grefenstette and Sadrzadeh (2011).

where "does" and "not" are assigned not empirical but 'logical' meanings:

resulting in a more interesting information flow structure. In ongoing work we investigate how the structures that are used to represent classical data flow in the quantum teleportation protocol enable one to model more of these 'logical words'. For example, in Clark *et al.* (2013) this was done for relative pronouns such as "who", "which", "that" and "whose".

Finally, while the well-established structural similarities across language families in terms of grammatical type calculi may seem mysterious, the teleportation-like information flow interpretation presented here clearly explains them.

References

Abramsky, S. & Coecke, B. A categorical semantics of quantum protocols. In *Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on*, 415–425, 2004.

Aron, J. Quantum links let computers understand language. *New Scientist*, 208(2790):10–11, 2010.

Bennett, C. H. & Schumacher, B. *Lecture at Tata Institute for Fundamental Research, Mumbai, India*, 2002.

Bennett, C. H., Brassard, G., Crépeau, C., Jozsa, R., Peres, A., & Wootters, W. K. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993.

Carroll, L. *Alice's Adventures in Wonderland*. MacMillan and Co., 1865.

Chomsky, N. *Syntactic Structures*. Mouton., 1957.

Clark, S., Coecke, B., & Sadrzadeh, M. A compositional distributional model of meaning. In *Proceedings of the Second Quantum Interaction Symposium (QI-2008)*, 133–140, 2008.

Clark, S., Coecke, B., Sadrzadeh, M., & Mary, Q. The frobenius anatomy of relative pronouns. *Draft paper*, 2013.

Coecke, B. Quantum picturalism. *Contemporary physics*, 51(1):59–83, 2010.

Coecke, B. The logic of entanglement. *arXiv: quant-ph/0402014*, 2004.

Coecke, B. & Duncan, R. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13(4):043016, 2011.

Coecke, B., Sadrzadeh, M., & Clark, S. Mathematical foundations for a compositional distributional model of meaning. *arXiv: 1003.4394*, 2010.

Gazdar, G. Paradigm merger in natural language processing. In *Computing Tomorrow: Future Research Directions in Computer Science*, Wand, I. & Milnor, R., (eds.). Cambridge University Press, Cambridge, 1996.

Gottesman, D. & Chuang, I. L. Quantum teleportation is a universal computational primitive. *Nature*, 402, 1999.

Grefenstette, E. & Sadrzadeh, M. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1394–1404. Association for Computational Linguistics, 2011.

Grefenstette, E., Sadrzadeh, M., Clark, S., Coecke, B., & Pulman, S. Concrete sentence spaces for compositional distributional models of meaning. *arXiv: 1101.0309*, 2010.

Kelly, G. M. & Laplaza, M. L. Coherence for compact closed categories. *Journal of Pure and Applied Algebra*, 19:193–213, 1980.

Lambek, J. The mathematics of sentence structure. *The American Mathematical Monthly*, 65(3):154–170, 1958.

Lambek, J. Type grammar revisited. In *Logical aspects of computational linguistics*, Lecomte, A., Lamarche, F., & Perrier, G., (eds.), 1–27. Springer, 1999.

Lambek, J. From word to sentence. *Polimetrica, Milan*, 2008.

Schütze, H. Automatic word sense discrimination. *Computational linguistics*, 24(1):97–123, 1998.

Selinger, P. A survey of graphical languages for monoidal categories. In *New structures for physics*, 289–355. Springer, 2011.

Skrzypczyk, P., Brunner, N., & Popescu, S. Emergence of quantum correlations from nonlocality swapping. *Physical review letters*, 102(11):110402, 2009.

Thomason, R. H., (ed.). *Formal Philosophy: Selected Papers of Richard Montague.* Yale University Press, New Haven, 1974.