Non-compositional concepts and quantum tests

Kirsty Kitto and Peter Bruza

Faculty of Science and Technology, Queensland University of Technology, Brisbane 4000 Australia

Abstract. Compositionality is a frequently made assumption in linguistics, and yet many human subjects reveal highly non-compositional word associations when confronted with novel concept combinations. This article will show how a non-compositional account of concept combinations can be supplied by modelling them as interacting quantum systems.

Keywords: Context, quantum-like models, semantics, non-compositional concepts
PACS: 89.20.Ff (Computer science and technology), 03.67.Mn (Entanglement measures, witnesses, and other characterizations), 89.75.Fb (Structures and organization in complex systems)

MODELLING THE HUMAN MENTAL LEXICON

Accurate models of language are essential in the modern world of Information Access and Retrieval (IAR), and yet the complexity of the human mental lexicon [1] makes this a difficult requirement to satisfy. In particular, the apparent *contextuality* exhibited by words makes probabilistic models of language difficult to construct. While Bayesian models have met with some success [2], this paper will discuss an alternative approach that utilises the probability structure of Quantum Theory (QT) to model the behaviour of words in the human mental lexicon [3, 4, 5].

A large amount of data has been collected which helps to map the human mental lexicon. We shall concentrate upon the University of South Florida (USF) word association data, which maps the strength of free word associations displayed by a large sample of psychology students over a period of 30 years [6]. In figure 1 we see the free association strengths for two words, "boxer" and "bat". Note the manner in which both words

Associate	*Probability*
ball	**0.25**
cave	0.13
vampire	0.07
fly	0.06
night	0.06
baseball	**0.05**
bird	0.04
blind	0.04
animal	0.02
...	...

(a)

Associate	*Probability*
fighter	**0.14**
gloves	**0.14**
fight	**0.09**
dog	0.08
shorts	**0.07**
punch	**0.05**
Tyson	**0.05**
...	...

(b)

FIGURE 1. The free association data for two words, (a) bat, and (b) boxer. In both cases, we can see a clear division of the senses of the two words, into a sport sense (highlighted in bold), and an animal sense.

can be attributed a meaning that belongs to one of two *senses*; an animal sense and a sporting sense. Despite this ambiguity, humans are adept at recognising the sense that is intended for an ambiguous word. They do this through reference to the context in which the word is being used, and this context might depend upon the co-occurrence of other words spoken before and after, the history of a conversation, the social context of the speaker etc. In the next section we shall show that it is possible to construct a simple model of this contextual dependency through use of the quantum formalism.

A QUANTUM-LIKE MODEL OF HUMAN WORD ASSOCIATIONS

A simple model of the effect of context in human word association data can be constructed through the use of a superposition state. In figure 2(a), the word w is represented in a free association context c, as a superposition of recalled, $|1\rangle$ and not recalled $|0\rangle$. Alternatively, referring to the online USF data for "bat" we can see that it might be represented as a superposition of 'recalled' and 'not recalled' upon presentation of a cue word "cave" (which in this case acts as a new context d). The probability of "bat" being recalled in this context is represented by $|a_1|^2$, as per the measurement postulate of quantum theory [7], but can be easily related to the Pythagorean theorem for the above diagram (which explains its origins). Thus, we could represent "bat" as the superposition: $\sqrt{0.94}|0\rangle + \sqrt{0.06}|1\rangle$ so giving a 6% probability that the word "bat" will be recalled by a subject who is presented with the cue "cave".[1]

This model is made more interesting in figure 2(b), where we have represented the fact that a different context might result in a different set of recall probabilities. Thus, when given the cue word "ball" we could represent the concept *bat* as the new superposition $\sqrt{0.81}|0\rangle + \sqrt{0.19}|1\rangle$. In this case we see that the word "bat" is more likely to be retrieved from memory when a subject is presented with the cue "ball" than the cue word

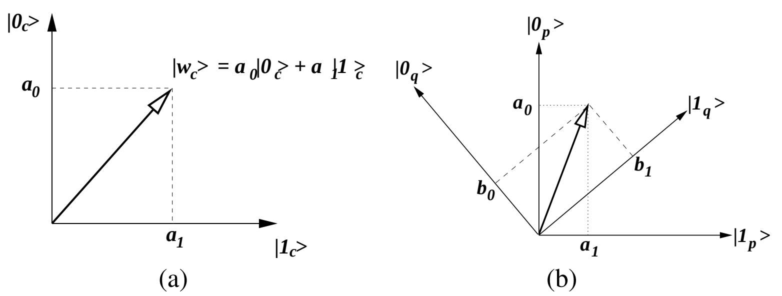

FIGURE 2. A concept w, for example *bat*, is represented in some context c which takes the form of a basis. (a) The word "bat" is recalled $|1\rangle$, or not, $|0\rangle$, in some context. Thus, if the context is the cue "cave", then the subject might recall "bat" with a probability a_1^2, or they might fail to recall "bat", with the probability a_0^2. (b) Changing the cue to "ball" might significantly change the chances of recall.

---

[1] These numbers are obtained by finding the value for "bat" in the "cave" matrix that is depicted at http://web.usf.edu/FreeAssociation/AppendixC/ , a similar procedure is used for the "ball" matrix in the below paragraph.

"cave". Thus, this formalism provides a natural representation of contextual effects as they actually occur in language.

How should we consider the combination of two words in this model? While it is possible that a simple tensor multiplication of the two superposition vectors might suffice, this is not necessarily the correct mechanism. Indeed, it seems possible that some concepts must be considered as non-separable, or, in the terminology of linguistics, *non-compositional*, which would imply that not all senses of a word remain accessible when it is combined with another. Thus, it might prove to be the case that a "boxer bat" is only ever thought of by human subjects as "a small furry mammal with boxing gloves on", or "a toy bat that a boxer dog chews on", which would imply a case of perfect anti-correlation in the senses. That is, defining a basis state in which $|1\rangle$ implies that a sport sense is recalled by a subject, and $|0\rangle$ an animal sense, we might find that not all possible combinations of the two senses can be realised:

$$|boxer\rangle \oplus |bat\rangle = a|01\rangle + b|10\rangle, \text{ where } |a|^2 + |b|^2 = 1. \quad (1)$$

The remainder of this paper will briefly sketch out recent work which attempts to test for such non-compositional behaviour in conceptual combination [3, 4, 5].

QUANTUM TESTS OF (NON-)COMPOSITIONALITY

QT has a well developed set of tests for non-separable behaviour which can be quickly adapted to probabilistic tests of the non-compositional behaviour of concept combinations. Generally, standard probabilistic reasoning suggests that if two systems A and B have behaviour that can be considered as separate then it is possible to describe this behaviour in terms of four random variables representing bivalent outcomes: $\{\mathbf{A1}, \mathbf{A2}\}$ and $\{\mathbf{B1}, \mathbf{B2}\}$. The joint probability describing the behaviour of the two systems can then be considered in a number of different ways. For example, it is possible to construct a variation of the Clauser–Horne–Shimony–Holt (CHSH) inequality using an analysis derived from Cereceda [8]:

$$\begin{aligned}\Delta =& |2(\Pr(\mathbf{A1} = +1, \mathbf{B1} = +1) + \Pr(\mathbf{A1} = -1, \mathbf{B1} = -1) + \Pr(\mathbf{A1} = +1, \mathbf{B2} = +1) \\ &+ \Pr(\mathbf{A1} = -1, \mathbf{B2} = -1) + \Pr(\mathbf{A2} = +1, \mathbf{B1} = +1) + \Pr(\mathbf{A2} = -1, \mathbf{B1} = -1) \\ &+ \Pr(\mathbf{A2} = +1, \mathbf{B2} = -1) + \Pr(\mathbf{A2} = -1, \mathbf{B2} = +1) - 2)| \leq 2 \end{aligned} \quad (2)$$

There is nothing in the above equation that restricts its domain of application to quantum theory. Indeed, there are many systems that appear to be separated in a similar way and so should adhere to probabilistic behaviour like that predicted in (2). One example can be found in the psychological literature, where an ongoing debate about the compositionality (or not) of concept combinations has been apparently irresolvable [5]. In this case, a concept combination (like *boxer bat*) should be considered compositional if it is possible to model it only with reference to its parts (possibly combined in some functional form). In a set of recent work [9, 4, 5] we have performed a number of experiments aimed at testing the compositional hypothesis. In these experiments, subjects are first primed by exposure to words that have a particular *sense* with regards to the second stage, where they are presented with a bi-ambiguous concept combination. They are then asked to interpret that combination, designating the senses that were used in that interpretation, and a probabilistic analysis is performed upon the data so obtained. This analysis has led to the simplified set of CHSH values depicted in table 1.

TABLE 1. Results obtained from a set of three CHSH style experiments performed upon the concept combinations *boxer bat*, *spring plant*, and *ring pen* for *n* subjects.

	Concept A		Concept B		Results	
	Prime 1(α_1)	Prime 2 (α_2)	Prime 3 (β_1)	Prime 4 (β_2)	CHSH	n
boxer bat	fighter	dog	vampire	ball	1.95	108
spring plant	summer	coil	seed	factory	2	136
ring pen	diamond	oval	ink	pig	**2.61**	132

It is also possible to write the joint probability in a form that starts to explain how violations of (2) occur. To do this we represent the four different random variables in a matrix describing the distribution, with each random variable combination itself split into the set of possible outcomes. This allows us to breakdown the result from table 1 into a form that allows for a consideration of the underlying form of the violations (or not) of (2). In this representation we can write the data gathered from the above experiments out as a set of joint distributions, which allows for a further understanding of the resulting behaviour. In this analysis, the joint probability for *boxer bat* becomes:

$$
\begin{array}{cccc}
 & & & \text{bat} \\
 & & & \begin{array}{cc} \mathbf{B1}(\mathit{ball}) & \mathbf{B2}(\mathit{vampire}) \\ +1 \quad -1 & +1 \quad -1 \end{array} \\
\text{boxer} & \begin{array}{c} \mathbf{A1}(\mathit{dog}) \\ \\ \mathbf{A2}(\mathit{fighter}) \end{array} & \begin{array}{c} +1 \\ -1 \\ +1 \\ -1 \end{array} & \left(\begin{array}{cc|cc} 0 & 0.89 & 0.42 & 0.46 \\ 0 & 0.11 & 0 & 0.12 \\ \hline 0 & 0.67 & 0.61 & 0.30 \\ 0 & 0.13 & 0 & 0.09 \end{array}\right)
\end{array} \tag{3}
$$

Here, we see no particular ordering or patterns. In particular, when we compare equations (2) and (3) we can see that the data gathered does not centre the distribution in such a way that it can violate the CHSH inequality. In contrast, *spring plant* leads to a joint distribution that has a far more interesting structure:

$$
\begin{array}{cccc}
 & & & \text{plant} \\
 & & & \begin{array}{cc} \mathbf{B1}(\mathit{factory}) & \mathbf{B2}(\mathit{seed}) \\ +1 \quad -1 & +1 \quad -1 \end{array} \\
\text{spring} & \begin{array}{c} \mathbf{A1}(\mathit{coil}) \\ \\ \mathbf{A2}(\mathit{summer}) \end{array} & \begin{array}{c} +1 \\ -1 \\ +1 \\ -1 \end{array} & \left(\begin{array}{cc|cc} 0.51 & 0 & 0.9 & 0 \\ 0 & 0.49 & 0 & 0.1 \\ \hline 0.87 & 0 & 0.98 & 0 \\ 0 & 0.13 & 0 & 0.02 \end{array}\right)
\end{array} \tag{4}
$$

In this case, we see a complete correlation between the subject responses. Thus, whenever a subject interprets spring as a season they decide that a plant is a vegetable, and that if spring is interpreted as a mechanical object then a plant is thought of as a factory etc. This complete correlation of the senses attributed to the bi-ambiguous words leads to a value of $\Delta = 2$.

Finally, if we consider the CHSH violating concept combination *ring pen* then we can see why a value of $\Delta > 2$ is obtained.

$$\begin{array}{cccc} & & & \text{pen} \\ & & \mathbf{B1}(pig) & \mathbf{B2}(ink) \\ & & +1 \quad -1 & +1 \quad -1 \\ \text{ring} & \begin{array}{c} \mathbf{A1}(oval) \begin{array}{c} +1 \\ -1 \end{array} \\ \\ \mathbf{A2}(diamond) \begin{array}{c} +1 \\ -1 \end{array} \end{array} & \left(\begin{array}{cc|cc} 0.49 & 0 & 0.69 & 0 \\ 0 & 0.51 & 0 & 0.31 \\ \hline 0.21 & 0 & 0.26 & 0 \\ 0.03 & 0.76 & 0.33 & 0.41 \end{array} \right) \end{array} \tag{5}$$

While this case is similar to the one illustrated in (4), it features a key difference in that a non-zero value has been returned by the ensemble of subjects for a couple of off-diagonal cases where the random variables yield $\{\mathbf{A2} = -1, \mathbf{B1} = +1\}$ and $\{\mathbf{A2} = -1, \mathbf{B2} = +1\}$. In the CHSH form written in equation (2) we see that it is the second term that is important; the off-diagonal term means that there is enough probability 'mass' for a violation. Comparing (3–5) with the set of equations typified by (2) we can understand that while it is necessary to for a system violating CHSH inequalities to have some correlation in the random variables, it is just as important to have an anti-correlation.

Thus, the techniques of Quantum Theory offer a number of valuable tools for classifying the manner in which humans respond to novel word compounds. Future work will attempt to further develop this fertile avenue of investigation.

ACKNOWLEDGMENTS

Supported by the Australian Research Council Discovery grants DP1094974, and DP0773341, and by the FP7 Marie Curie IRSES Project 247590 "QONTEXT".

REFERENCES

1. J. Aitchison, *Words in the Mind: An Introduction to the Mental Lexicon*, Wiley-Blackwell, 2003.
2. M. Steyvers, T. L. Griffiths, and S. Dennis, *Trends in Cognitive Sciences* **10**, 327 – 334 (2006).
3. P. Bruza, K. Kitto, D. Nelson, and C. McEvoy, *Journal of Mathematical Psychology* **53**, 362 (2009).
4. K. Kitto, B. Ramm, L. Sitbon, and P. D. Bruza, *Axiomathes* **21**, 331–345 (2011).
5. P. D. Bruza, K. Kitto, B. Ramm, and L. Sitbon (2011), under review.
6. D. Nelson, C. McEvoy, and T. Schreiber, *Behavior Research Methods* **36**, 408–420 (2004).
7. F. Laloë, *American Journal of Physics* **69**, 655–701 (2001).
8. J. Cereceda, *Foundations of Physics Letters* **13**, 427–442 (2000).
9. K. Kitto, B. Ramm, P. D. Bruza, and L. Sitbon, "Testing for the non-separability of bi-ambiguous words," in *Proceedings of the AAAI Fall Symposium*, AAAI Press, 2010.