Cognition 162 (2017) 61–72

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Representing composed meanings through temporal binding

Hugh Rabagliati [a,*], Leonidas A.A. Doumas [a], Douglas K. Bemis [b]

[a] University of Edinburgh, United Kingdom
[b] Geometric Intelligence, Inc, New York, NY, United States

ARTICLE INFO

Article history:
Received 14 March 2016
Revised 9 January 2017
Accepted 19 January 2017
Available online 16 February 2017

Keywords:
Meaning
Semantic composition
Binding problem
Prediction

ABSTRACT

A key feature of human thought and language is compositionality, the ability to bind pre-existing concepts or word meanings together in order to express new ideas. Here we ask how newly composed complex concepts are mentally represented and matched to the outside world, by testing whether it is harder to verify if a picture matches the meaning of a phrase, like *big pink tree,* than the meaning of a single word, like *tree*. Five sentence-picture verification experiments provide evidence that, in fact, the meaning of a phrase can often be checked just as fast as the meaning of one single word (and sometimes faster), indicating that the phrase's constituent concepts can be represented and checked in parallel. However, verification times were increased when matched phrases had more complex modification structures, indicating that it is costly to represent structural relations between constituent concepts. This pattern of data can be well-explained if concepts are composed together using two different mechanisms, binding by synchrony and binding by asynchrony, which have been suggested as solutions to the "binding problem" faced in both vision science and higher-level cognition. Our results suggest that they can also explain aspects of compositional language processing.

1. Introduction

Compositionality is a key feature of human thought and language: We can effortlessly combine older, more basic concepts and word meanings in order to express an unbounded number of new ideas. For instance, even though the words *Spotted, Pink*, and *Tree* are rarely juxtaposed, they can be quickly composed together to create a coherent semantic interpretation.

Work in linguistic semantics, philosophy, and psychology has considerably advanced our understanding of how complex concepts, such as the meanings of phrases, might be built from their component parts (Heim & Kratzer, 1998; Pylkkänen & McElree, 2006; Werning, Hinzen, & Machery, 2012). This has included discoveries about the role of broader world knowledge in interpreting the meanings of phrases (Barner & Snedeker, 2008; Springer & Murphy, 1992), and about the neural implementation of combinatorial operations (Bemis & Pylkkänen, 2011; Frankland & Greene, 2015; Pylkkänen & McElree, 2007).

However, amongst this research there is a surprising gap in our knowledge: we know little about how composed representations are held in mind in order to be matched against the world. While we know a great deal about how individual words (like *spotted, pink* or *tree*) are stored in working memory (Baddeley, 2003), and about how complex concepts can, with experience, be "chunked" into simple units (Cowan, Chen, & Rouder, 2004), we know much less about how newly encountered combinations of concepts are mentally represented. For example, how does the representation of a complex concept, such as *big pink tree*, differ from the representation of a singleton concept, such as *tree,* or from the representation of a list of word meanings, such as *big, pink, tree*? Do complex representations, built by stacking ever more concepts, also demand ever more working memory? Can some complex concepts be stored in very efficient ways?

Some of the most relevant work has been on the idea of gist representations, the proposal that, as we read or listen to text, we discard our precise memories of the exact linguistic input and replace them with less precise summaries of that input's meaning. Theories of gist can explain how and why we discard less-relevant information about a sentence, but their accounts of meaning (in which, for example, sentences are recoded as sets of propositions, Carpenter & Just, 1975; Clark & Chase, 1972; Kintsch, 1998) are more suited for explaining the representation of large chunks of text rather than characterizing the representations of simple concepts such as *big pink tree*. For example, it is unclear how the gists of *tree* and *pink tree* might differ. Potter (1993) has argued that gist representations are built by binding together token

---

* Corresponding author at: School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom

E-mail address: hugh.rabagliati@ed.ac.uk (H. Rabagliati).

http://dx.doi.org/10.1016/j.cognition.2017.01.013

representations of concepts in a short term conceptual memory store. This idea seems plausible, but without a precise account of what these bindings might be like, it is hard to evaluate the implications of the claim for the questions posed above.

Potential insight into these bindings can be found in the literature on compositional connectionist models. In these models, individual concepts (i.e., word meanings) are stored as separate nodes in a large neural network. The concepts can be composed together (i.e., bound) through their simultaneous co-activation (so-called temporal binding). The key idea, which has its roots in the "binding by synchrony" hypothesis from visual attention (Singer & Gray, 1995; Von Der Malsburg, 1981), is that composed concepts like *pink tree* might be represented in a neural network by simultaneously activating nodes for the constituent concepts, i.e., *pink* and *tree* (Hummel & Holyoak, 1997). This idea has been implemented in a number of neural network models, such as Hummel and Holyoak's (1997) model of analogy formation. Importantly, it has also recently received support as a neurophysiologically plausible account of how combinatorial linguistic structure might be represented (Ding, Melloni, Tian, Zhang, & Poeppel, 2015).

One reason that binding by synchrony is a plausible candidate mechanism for compositional binding is that it provides an efficient way of compressing information, just like a gist. Because neural networks operate in parallel, the complexity of a network in which only *tree* is active is not importantly different from the complexity of a network in which both *pink* and *tree* are activated. That is to say, the network pays essentially no additional cost (e.g., at least in terms of storage) in order to represent *pink tree* as opposed to *tree.*

Binding by synchrony is therefore a clever and efficient default mode of representation. However, it displays an important difficulty accounting for certain types of more complex compositional representations. In particular, when using binding by synchrony it is not possible to represent the precise structure with which concepts should be bound (Doumas, Hummel, & Sandhofer, 2008). While the simultaneous activation of a set of concepts does indicate which of them should be bound together, it does not indicate which concepts should serve as arguments and which should serve as predicates. This makes it difficult to represent any sort of well-structured concepts. To illustrate, consider how synchrony might be used to represent the concepts *pink tree* and *dark pink tree*. *Pink tree* can be easily represented through synchrony: the simultaneous activation of *pink* and *tree* will activate the features associated with pinkness and with treeness, features that are best matched by a pink coloured tree. However, if we try to represent *dark pink tree* through synchrony, we will produce an extremely inaccurate representation. In this case, we would activate features associated with darkness, with pinkness, and with treeness. These features would be best matched by something that is simultaneously a dark tree (e.g., a tree in darkness), a dark pink colour, a pink tree, a tree with a dark pink colour, and so on. This is clearly not a typically intended meaning of *dark pink tree*.

To represent the structure of a composed concept in a neural network, it is necessary to somehow "screen off" individual component concepts from each other, to create the constituent relationships of the structure (e.g., ensuring that *dark* modifies *pink* but not *tree*). This is not simple to do. One suggestion has been to use so-called conjunctive codes, in which each component concept is given a separate representation for each possible role that it might play (e.g., we would store two representations of *dark,* one for when it modifies another adjective [*dark pink*] and one for when it modifies a noun [*dark tree*]). But this solution has a theoretically unsatisfying consequence, as it assumes that every concept must have a different instantiation for each potential role that it might play. This sort of ambiguity of representation is inconsistent with a fundamental principle of compositionality, that the meaning of an expression should be a function of the meaning of its parts; in this case, the meaning of a part would be determined by its function in an expression.

An alternative approach, and the one that we focus on here, is that concepts may be screened off from each other by using binding via *asynchrony*, in which the pattern of activation of concepts over time distinguishes different thoughts and creates constituent structure. For example, a phrase like *dark pink tree* can be represented with a constituent structure of the form [[dark & pink[1]] tree] by initially co-activating *dark* and *pink* (to indicate a dark pink colour), and then subsequently activating *tree* in isolation (to indicate that the bound concept "dark pink" should modify *tree*) (Doumas et al., 2008; Hummel & Holyoak, 2003). In this case, the initial period of activation would first activate features associated with *darkness* and *pinkness* (which would be well matched by a dark pink colour), and then features associated with treeness (which would be well matched by a *tree*). That is to say, given this pattern of activation, *dark pink tree* would be well matched by a tree with a dark pink colour.[2]

The ideas of binding by synchrony and asynchrony suggest answers to the questions posed at the start of the paper about the nature of compositional representations. Because simple compositional concepts with minimal structure, such as *pink tree*, can be represented through simultaneous activation, then their representation does not importantly differ from the representation of a single word (i.e., it is the pattern of activity in a neural network at one single point in time). More complex and structured concepts, however, must be represented by activating the different components of a concept across time. That is to say, the system pays a cost for precisely representing structure (e.g., some models limit the number of timesteps available (Doumas et al., 2008); this provides an upper bound on working memory capacity).

1.1. A potential challenge

The ideas behind synchronous and asynchronous binding can easily map on to the processes involved in completing an experimental task such as sentence-picture verification, in which participants read a phrase and then verify if it matches a subsequent picture (Clark & Chase, 1972). If two concepts are bound through synchrony, such as *pink tree* in Fig. 1 (left side), then the perceivable (e.g., visual) features associated with those concepts will be activated in parallel. Each of these features can then be simultaneously checked against the input. This means that it should be as easy for participants to verify the meaning of a phrase (*pink tree*) as to verify the meaning of a single word (*tree*), assuming that the key features (here, colour and shape) can be extracted from the picture at similar speeds. When concepts are bound asynchronously (Fig. 1, right side), then the predictions are different. For a phrase like *dark pink tree,* each key component must be activated at a different timepoint. First, *dark* and *pink* are co-activated, along with their visual features, and these are checked against the input in parallel. Meanwhile, *tree* is activated, and its features are checked against the input. This mixed parallel-serial process would cause participants to be slower to verify the meaning of asynchronously bound phrases than synchronously bound phrases.

However, one recent result suggests that both of these predictions – parallel checking and mixed parallel-serial checking – are incorrect, and that the meanings of some composed concepts can be verified faster than the meanings of single words. In a sentence-picture verification task (conducted as part of a magnetoencephalography study), Bemis and Pylkkänen (2011) found that

---

[1] For interpretation of color in Figs. 1, 2 and 6, the reader is referred to the web version of this article.

[2] And an alternative activation pattern could represent pink tree in darkness.