ARTICLE IN PRESS
PLREV:143
Available online at www.sciencedirect.com
ScienceDirect
Physics of Life Reviews ••• (••••) •••–•••
PHYSICS of LIFE reviews
www.elsevier.com/locate/plrev

Comment

Embodied compositionality
Comment on "Embodied language, best-fit analysis, and formal compositionality" by J. Feldman

Mark Steedman

Received 9 November 2010; accepted 9 November 2010

Communicated by L. Perlovsky

Jerry Feldman's paper is to be welcomed in seeking to link linguistic compositional semantics to some form of nonlinguistic conceptual composition. In fact, I'd like to press the question, and ask *why* concepts are compositional, and what it is that allows human conceptual compositionality to support human language, while animal conceptual compositionality apparently does not. First I'd like to raise some questions about the argument in the paper.

1. What is compositionality?

The word "meaning" is notorious for carrying a very wide range of senses in common language use. If I say "those clouds mean rain", I am talking about an inference from an observation of the real world. It would clearly be nonsensical to talk about the semantics of clouds, and to argue whether it is compositional or not.

Similarly, if I say "It's eight o' clock" and you infer that it is time for dinner, that inference shouldn't be considered semantic, even if that is what I intended by my utterance. If we say it is part of my meaning, then we should be aware that we are using the term loosely. The term semantics is reserved for literal meaning.

Seen in this light, the mere fact of ambiguity in sentences like *John loves Virginia* from Section 1 of the paper is not evidence of non-compositionality in semantics. The fact that words like *Virginia* are ambiguous between naming a state and one of various persons simply means that there is more than one literal meaning for context to choose from. Nevertheless, all of those literal meanings are entirely compositional. The fact that I infer different consequences from the different literal meanings isn't surprising, and doesn't reflect on the semantics.

Similarly while the two readings of *I saw a squirrel with a telescope*, arise from two distinct syntactic derivations, both are entirely compositional—indeed, it is the compositionality of "squirrel" "with", and "telescope" that allows the absurd reading to be eliminated in most contexts.

Similarly, the fact that "red" denotes a range of colors that overlaps with brown and orange, so that the red of red hair is spectrally distinct from that of red wine, is (as cognitivists often point out) the result of such concepts being embodied in particular neural mechanisms of vision, rather than non-compositionality in the concepts themselves.

Even the adjective "good", whose extension is almost entirely relative to the purpose in hand ("a few good men" may denote men who are in other respects very bad) is compositional as far as its rather meager literal meaning goes.

I labor this point, because it implies that the contrast which cognitive linguists always claim with logicist semantics is overstated. Logicists have quite a lot to say about inference and entailment, and if they don't have much to say about

DOI of original article: 10.1016/j.plrev.2010.06.006.
E-mail address: steedman@inf.ed.ac.uk.

1571-0645/$ – see front matter
doi:10.1016/j.plrev.2010.11.005

common sense reasoning on the basis of their semantics, that may be because they have found out the hard way just how difficult it is to make it work. It is in this connection that I raise the following further question.

2. What are the consequences of cognitive embodiment?

One of the most useful contributions of cognitive linguistics has been to insist on how profoundly our linguistic representations of the world are shaped by our physical embodiment and existence in the physical and social world.

What follows from this observation? Some authors, such as Brooks [2], have argued that grounding and embodiment actually make things easier. We can forget about all that logicist AI, because if we build things that are embodied and grounded in the physical world, complex behavior will be emergent from machine learning techniques.

There is some indication that this approach might work for low-level vision and perhaps speech recognition. For example, Hiton and Salakhutdinov [9], Hinton, Osindero and Teh [8] show that classifiers based on generative models and unsupervised learning can do a surprisingly good job of extracting structure from sense-data. However, the complexity of such computations is daunting. If what we are looking at is even partly the result of 500M years of chordate evolution, it seems unlikely that machine learning alone will solve the problem of language-level concept learning, and that is not what the current paper advocates.

If that *is* what we are looking at, then what follows from the embodied nature of cognition is that it makes the problem much *harder*. Evolution is quite unlike computation. It has essentially unlimited resources of energy to draw on, and is capable of essentially unlimited exhaustive search (boosted by the occasional mass extinction to deal with problems of overfitting). In that case, maybe it would be better to examine the neural mechanism that have resulted from the grounded evolutionary process.

The paper makes something of a case for this approach, citing the well-known work of Pulvermüller [12] on the activation of motor areas related to the corresponding action on hearing verbs like "kick" and "talk". Some new work by Bergen and Wheeler [1] is discussed, showing interference from directionality of the response task on processing sentences predicating directional actions like opening and closing a drawer, in comparison to related states. Mention is also made of the "mirror neuron" experiments of Rizzolatti, Fogassi and Gallese [14], which show activation of corresponding motor areas in monkeys in response to seeing another animal in action.

It is claimed that these experiments show comprehension of linguistic terms denoting actions actually to be mediated by mental "simulation" of action in motor areas. However, these experiments seem to admit of an alternative explanation where the motor area activity is merely a by-product of comprehension of language or vision.

Perhaps this does not matter, because embodied simulation does not seem to play a crucial role in the theory of Embodied Construction Grammar (ECG) that is actually advanced in the present paper. The semantic structure corresponding to "He cut the bread" shown in figure seems very similar in formal terms to the yield of an HPSG grammar. The semantics is spelled out in a more articulated form, with the event of cutting decomposed into elements denoting application of force and resultant process of being cut, but in other respects this is a logical form much like any other. While it apparently provides the input to simulation, it does not appear to reflect embodiment in any specific way.

The meaning representation is assembled by a parsing process that appeals to a context-independent criterion of "Best Fit", using corpus statistics and ontological resources like FrameNet, in order to resolve ambiguities such as that of homophones like "bank". This too seems fairly standard. Reference resolution is via a mechanism whose resemblance to DRT is noted. While it is true that DRT has not been extended to cover the full range of mental spaces of Fauconnier and Turner [4], as discussed by Sweetser [15], it seems a little unfair to claim that logical semantics is "[unable] to model actions" or that it is unable to represent dynamics. There is a long tradition in logicist Artificial Intelligence from McCarthy and Hayes [10] through Reiter [13] that addresses these question, and an even longer tradition in logic from Prior [11] through Harel, Kozen and Tiuryn [7] that addresses the question, whose relation to the practical problem of program verification is described by Vardi [17]. The links from such theory to natural language are discussed by Dowty [3] and Fernado [6], among others.

3. Where does compositionality come from?

The central claim of the paper is that "compositionality of language derives from an underlying compositionality of thought." This claim is clearly illustrated and made concrete by the semantics in Fig. 5, and the paper makes a strong

Please cite this article in press as: Steedman M. Embodied compositionality. Phys Life Rev (2010), doi:10.1016/j.plrev.2010.11.005

case in support. Indeed, after reading the paper, I'm inclined to believe that compositionality is a *necessary* property for anything we would call a concept.

The case is strong enough to raise a further question: how much of this compositional action representation do our closest animal relatives the apes possess? Such animals are quite good at making plans, including those involving tools, and seem quite capable of recognizing actions very much like the one anatomized in Fig. 5 [14]. One might well suspect that their action concepts too must be compositional. Yet they lack language, or at least the kind of productive compositional language we possess. Tomasello [16] has convincingly argued that it is cultural cooperation that has shaped human language. But how exactly does this translate into formal terms? Jerry Feldman's paper is to be welcomed as offering a provocative standpoint from which to view this question.

Uncited references

[5]

References

[1] Bergen Benjamin, Wheeler Kathryn. Grammatical aspect and mental simulation. Brain and Language 2010;112:150–8.
[2] Brooks Rodney A. A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation 1986;2:14–23.
[3] Dowty David. Word meaning in Montague grammar. 2nd edition with additions. Dordrecht: Reidel; 1979, 1991.
[4] Fauconnier Giles, Turner Mark. The way we think: conceptual blending and the mind's hidden complexities. New York: Basic Books; 2002.
[5] Feldman Jerry. Embodied language, best fit analysis, and formal compositionality. Physics of Life Letters 2010;7(4):385–410 [this issue].
[6] Fernando Timothy. Schedulesina temporal interpretation of modals. Journal of Semantics 2005;22:211–29.
[7] Harel David, Kozen Dexter, Tiuryn Jerzy. Dynamic logic. Cambridge, MA: MIT Press; 2000.
[8] Hinton Geoffrey, Osindero Simon, Teh Yee Whye. A fast learning algorithm for deep belief nets. Neural Computation 2006;18:1527–54.
[9] Hinton Geoffrey, Salakhutdinov Ruslan. Reducing the dimensionality of data with neural networks. Science 2006;313:504–7.
[10] McCarthy John, Hayes Patrick. Some philosophical problems from the standpoint of artificial intelligence. In: Meltzer Bernard, Michie Donald, editors. Machine intelligence, vol. 4. Edinburgh: Edinburgh University Press; 1969. p. 473–502.
[11] Prior Arthur. Past, present and future. Oxford: Clarendon Press; 1967.
[12] Pulvermüller Friedemann. Brain mechanisms linking language and action. Nature Reviews: Neuroscience 2005;6(7):576–82.
[13] Reiter Raymond. Knowledge in action: logical foundations for describing and implementing dynamical systems. Cambridge MA: MIT Press; 2001.
[14] Rizzolatti Giacomo, Fogassi Leonardo, Gallese Vittorio. Neurophysio-logical mechanisms underlying the understanding and imitation of action. Nature Reviews: Neuroscience 2001;2:661–70.
[15] Sweetser Eve. Compositionality and blending. In: Jansen Theo, Redeker Gisella, editors. Cognitive linguistics: foundations, scope, and methodology. Berlin: Mouton de Gruyter; 1999. p. 129–62.
[16] Tomasello Michael. The cultural origins of human cognition. Cambridge, MA: Harvard University Press; 1999.
[17] Vardi Moshe. From church and prior to PSL. In: Grumberg O, Veith H, editors. 25 years of model checking. LNCS, vol. 5000. Berlin: Springer; 2008. p. 150–71.

Please cite this article in press as: Steedman M. Embodied compositionality. Phys Life Rev (2010), doi:10.1016/j.plrev.2010.11.005