

Opinion piece



Cite this article: Hagoort P. 2019

The meaning-making mechanism(s) behind the eyes and between the ears.

Phil. Trans. R. Soc. B **375**: 20190301.

<http://dx.doi.org/10.1098/rstb.2019.0301>

Accepted: 10 September 2019

One contribution of 16 to a theme issue 'Towards mechanistic models of meaning composition'.

Subject Areas:

neuroscience, cognition

Keywords:

language, meaning, brain, mechanism, top-down causation

Author for correspondence:

Peter Hagoort

e-mail: peter.hagoort@donders.ru.nl

The meaning-making mechanism(s) behind the eyes and between the ears

Peter Hagoort

Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands

PH, 0000-0001-7280-7549

In this contribution, the following four questions are discussed: (i) where is meaning?; (ii) what is meaning?; (iii) what is the meaning of mechanism?; (iv) what are the mechanisms of meaning? I will argue that meanings are in the head. Meanings have multiple facets, but minimally one needs to make a distinction between single word meanings (lexical meaning) and the meanings of multi-word utterances. The latter ones cannot be retrieved from memory, but need to be constructed on the fly. A mechanistic account of the meaning-making mind requires an analysis at both a functional and a neural level, the reason being that these levels are causally interdependent. I will show that an analysis exclusively focusing on patterns of brain activation lacks explanatory power. Finally, I shall present an initial sketch of how the dynamic interaction between temporo-parietal areas and inferior frontal cortex might instantiate the interpretation of linguistic utterances in the context of a multimodal setting and ongoing discourse information.

This article is part of the theme issue 'Towards mechanistic models of meaning composition'

1. Introduction

Early on in my career, I tested a few hundred aphasic patients with a standardized aphasia test battery (the Aachen Aphasia Test). Despite the claim that syntax is the hallmark of human uniqueness as it comes to language [1], it was obvious from my interactions with aphasic patients that communication is relatively easy if syntax has fallen apart. Any permutation of the string 'apple man eat' will result in the understanding that the man eats the apple, based on recruiting common background knowledge for computing the multi-word meaning. In most cases, one does not need the correct word order or the presence of inflections to arrive at the correct interpretation. This is different for phonology and semantics. When a severe dysarthria prevents the production of a clear sound profile of the speech tokens, understanding becomes a lot harder. In many cases, it is even worse when meaning goes awry. Semantic paraphasias and jargon aphasia result in failed communication because of lack of shared meanings. From a communication perspective, we can handle reduced syntax, but we cannot handle incoherent meaning. Hence, to a large extent meaning is what human language is about.

From an epistemological perspective, however, meaning is the most evanescent aspect of language. We have quite advanced knowledge about how the vocal tract produces sounds, and sophisticated models of the neural circuitry underlying speech sound production (e.g. the DIVA model; [2]). We have somewhat advanced proposals about sequence processing and (morpho)syntax and the underlying neural circuitry [3,4]. When it comes to meaning, we see a great divide. On the one hand, there are advanced models in formal semantics, but they have no obvious relation to what is going on behind the eyes and between the ears. On the other hand, we might find methodologically sophisticated neuroimaging studies on meaning, but without providing any insight into meaning (e.g. [5]). Integrated proposals are rare (but see [6] for an exemplary exception). In general, 'word meanings are notoriously slippery and are arguably the least amenable to formal treatment of all phenomena occurring in natural language' [7, p. 290]. I will not be able to provide a solution. Instead, in the remainder I will try to analyse the problem space by asking four questions. These are:

- (i) Where is meaning?
- (ii) What is meaning?
- (iii) What is the meaning of mechanism?
- (iv) What are the mechanisms of meaning?

2. Where is meaning?

A Wittgensteinian view of meaning is that it resides in the language community. Meanings are not in the head of the individual, but practices in the social world out there. Meanings are negotiated among conversational partners [8]. What a word stands for is not determined by states of individual minds, but is contextually dependent on the language game one plays with other members of the species. In the most radical version of such a view, a cognitive neuroscience perspective on meanings is doomed to fail, since meanings are not in the head, and hence meaning-making processes cannot be identified by neurophysiological measurements. In that case, one should give up the endeavour to search for meaning in the brain [6]. However, I believe such a view is ungrounded. As has been argued extensively elsewhere (e.g. [6,7,9]), natural language only refers to the external world through mediation of the sense organs. In other words, language refers to neurally constrained representations of reality. A key feature of organisms with a central nervous system is that they engage themselves with the world based on an internal construction of the environment. This is the Kantian perspective on our epistemological relation to reality. It is fully in line with the currently influential predictive coding paradigm that goes back to Kant and to the refinement of his views on perception and cognition by Helmholtz [10,11]. The brain is a Kantian machine. It generates an internal model that guides the confrontation with the sensory input. This internal model is the ‘virtual reality’ [7] to which our linguistic tokens refer. It includes aspects of not only the physical world, but also our social environment, including a model of the conversational partners. According to this view, something happens behind our eyes and between our ears when meanings are composed. In contrast to the idea that we are ‘minds without meaning’ [12], it is my conviction that ‘minds make meaning’ (MMM). Meaning-making is an active constructive process, as I will argue below. Hence, it makes sense to ask about the meaning-making mechanisms in our mind/brain.

3. What is meaning?

In the above, I have used the word meaning lightheartedly, ignoring that the word ‘meaning’ has itself many shades of meaning [13]. I have not made a distinction between sense and reference (Frege’s *Sinn* and *Bedeutung*, 1892 [14]). I have ignored the distinction between the meaning of single words (lexical meaning) and the semantics of propositions and larger discourse. I have not discussed central notions of formal semantics, such as quantifiers, presuppositions and entailment. Discussing the many aspects of meaning is far beyond the scope of this contribution. However, I will highlight one fundamental distinction. We need to distinguish between semantic building blocks that are acquired during language learning and stored in memory (lexical meaning), and complex meaning spanning multi-word utterances. Complex meaning is mostly created on the fly (with the exception of idioms). It requires a unification operation that takes context and lexical meanings as input and

delivers a situation model as output. The classical view on unification is that it follows the principle of Fregean compositionality. In the words of Barbara Partee [15, p. 313] it says that ‘The meaning of a whole is a function of the meanings of the parts and of the way they are syntactically combined’. However, there are multiple reasons why this strict form of compositionality would not do. I name a few. First, there are elements of complex meaning that cannot be derived from syntactically guided composition, such as in aspectual and complement coercion. Here is an example: The sentences ‘the goat finished the book’ and ‘the reader finished the book’ have a different event interpretation for the clause ‘finished the book’. However, the implied semantic element is syntactically silent. Syntactic composition alone will not deliver the correct interpretation of these sentences. Second, unification operations do not only include linguistic elements, but also co-speech gestures [16], characteristics of the speaker’s voice [17] or information from pictures or animations [18,19]. These examples indicate that unification goes beyond combining lexical meanings. Context and different sources of non-linguistic co-occurring information contribute to determine the interpretation of a linguistic utterance immediately during the meaning-making process. This is what I have called the *Immediacy Principle*; it seems to guide language processing [20]. In short, ‘even the humblest everyday language use is shot through with enriched compositionality’ [13, p. 69].

To illustrate the multiple aspects of meaning, I will discuss in some detail a study on implied emotion in sentence comprehension that we did some years ago [21]. In this functional magnetic resonance imaging (fMRI) study, we tested emotional valence generated by combinatorial processing alone. To do this, we designed sentences that contain no words with a negative connotation, but as a whole still induce the emotion of fear (e.g. ‘The boy fell asleep and never woke up again’). These sentences we contrasted with their neutral counterparts (‘The boy stood up and grabbed his bag’). In the fear-inducing sentences, none of the individual lexical items is associated with negative connotations, but the sentence as a whole triggers negative feelings. The neutral sentences contained words with the same emotional valence as the negative sentences, but without the implied emotional meaning from their combination. Increased activation for the fear-inducing sentences was found in emotion-related areas, including the amygdala, the insula and medial prefrontal cortex. In addition, the language-relevant left inferior frontal cortex (LIFC) showed increased activation in response to these sentences. This raises the following question: which parts of the observed brain activations do we consider to be part of, or informative for, an empirically grounded account of meaning? It seems to be clear that a reconstruction of the depicted event does not require the contribution of emotion-related areas. Moreover, the truth value of the propositional content can be established independently of the emotional ramifications. A psychopath could agree with us (under the assumption that you and I belong to the set of non-psychopaths) that the meaning of the ‘finger fell into the soup’ refers to a situation where there was a finger that fell into a container (bowl/cup) filled with soup. But, so the proponents of embodied semantics would claim, does one really understand the meaning if the emotional response is missing? The same can be argued for concomitant sensory and motor information. Although these aspects might not be necessary for establishing the truth value of an utterance, they enrich our understanding, as argued by Mahon & Caramazza [22,

p. 68]: ‘Sensory and motor information colors conceptual processing, enriches it, and provides it with a relational context. The activation of the sensory and motor systems during conceptual processing serves to ground ‘abstract’ and ‘symbolic’ representations in the rich sensory and motor content that mediates our physical interaction with the world’. Or, more precisely, meanings are gateways to the rich virtual environments that our mind has created to account for the internal and external signals that it receives. These gateways are created by conceptual building blocks acquired over a lifetime of experience with linguistic symbols, in interaction with the combinatorial machinery that creates a situation model, integrating the building blocks retrieved from memory with context and non-linguistic information.

This special issue asks for more than a descriptive analysis of the many shades of meaning. It wants to evaluate the progress in mechanistic accounts of MMM. However, in this case we are confronted with a problem similar to that with the concept of meaning. The concept of mechanism is not very clear either. Therefore, first I will discuss the meaning of mechanism.

4. What is the meaning of mechanism?

In the context of neurobiology, a popular reading of mechanism is that it is the causally effective set of neural processes that creates cognitive phenomena, including meaning, as emergent properties. These properties are a consequence of the operations of the underlying mechanism, but do not themselves contribute to the workings of that mechanism [23]. Mechanistic models are often contrasted with descriptive models [24]. These models summarize data, often in the form of boxes and arrows. However, they do not speak to the issue of how nervous systems operate to create the data. Ultimately, then, models of cognitive functions should be reduced to mechanistic accounts in terms of neurophysiological principles of brain function. In such a scheme, basic neuroscience will in the end replace cognitive science.

However, there are good reasons to think that this will not work. An important argument is that all biological systems are not only under the influence of bottom-up causation, but also under the spell of top-down causation. To the best of my knowledge, the term top-down causation was introduced in biology by Donald Campbell [25]. Campbell’s illuminating illustration was the jaw of the termite. Genes are causally involved in creating the proteins that make up the jaw structure (bottom-up causation). Where did the DNA get the instructions from to generate such effective jaws? The answer is through the mechanism of natural selection. The *Umwelt* in which the termites live and the consequences of the way the jaw is built for the termite’s survival and reproduction has a downward causal (selective) influence on the DNA. ‘Where natural selection operates through life and death at a higher level of organization, the laws of the higher-level selective system determine in part the distribution of lower-level events and substances’ [25, p. 180]. The lesson is that system-level organization provides constraints for the operations of the components that it is made of [26,27]. In this way, the system-level organization has causal powers over and beyond the separate components. An explanation of a functional system in biology hence requires an understanding of the interaction between bottom-up and top-down causal contributions. An example in the case of language is that the learning and communication

context has a top-down causal effect on the way in which key aspects are implemented in the brain (e.g. as spoken language versus sign language; as Chinese versus English, etc.)

One way in which the bottom-up and top-down contributions can be characterized is by an analysis at the three levels that Marr postulated (computational, algorithmic, implementational). For instance, in tonal languages pitch information has to be extracted from the input to identify word meaning. A specification of a pitch extraction algorithm is part of a mechanistic explanation of how in such a case the brain solves the problem of mapping sound to meaning. An understanding of the mechanism needs a specification at levels beyond neurophysiology and neuroanatomy. To restrict a mechanistic explanation to the implementational level (i.e. the neurophysiology and anatomy of the central nervous system) is too limited, since it ignores the causally effective constraints from the functional/algorithmic level that is a central feature of all biological systems. Function is a key aspect of biological systems and their explanations [28]. In short, the proper function [29] in evolutionary terms has a downward causal effect on the organization of the medium that implements this function (i.e., the central nervous system).¹

An illustration of what goes wrong if the mechanistic account ignores the constraints from an algorithmic level (for multiple examples of processing at the algorithmic level, see [30]) is the influential study by Huth *et al.* [5] in *Nature* (for an excellent critique, see [30]). Based on natural speech input from a story and through voxel-wise modelling of fMRI data, a detailed semantic atlas of the cortical surface of the human brain was constructed. Although it starts from the assumption that word meaning is based on a distributed pattern of activity across cortex, the map results in a semantic clustering of words grounded in close spatial proximity of brain activations. For instance, in the right temporo-parietal junction words most strongly associated with this part of cortex included *cousin*, *murdered*, *pregnant*, *pleaded*, *son*, *wife*, *confessed*. This is quite a mixed bag of semantic categories. It is not clear what the underlying semantic dimension is that groups these words together. It certainly is not consistent with a coherent account of the organization of lexical meaning in the large psycholinguistic literature (cf. [4]). My diagnosis is that this is, at least partly, because a specification of the algorithm that processes the speech input is missing (see [30], for a similar diagnosis). The answer by Huth and colleagues might be that there is nothing beyond what the brain tells us. Since the brain is the ultimate mechanism, the semantic map is the emergent property that we have to take for what it is. But this will not tell us anything about semantic processing, since the algorithm that connects natural speech input to the neural data is missing. If we consider the algorithm part of the mechanism, we will understand that implementational level data alone will not suffice for an account with sufficient explanatory power. Hence the Huth *et al.* [5] data tell us preciously little about the mechanism of making meaning.

I can further illustrate the problems of limiting oneself to the implementational level for the Lai *et al.* study [21] that was discussed above. In their study, Huth *et al.* used context vectors for individual words as regressors in their model to predict the blood oxygen level-dependent (BOLD) activity. However, this does not include the outcome of combinatorial processing in the analysis [30]. As we have seen, combinatorial operations were responsible for the activation of the

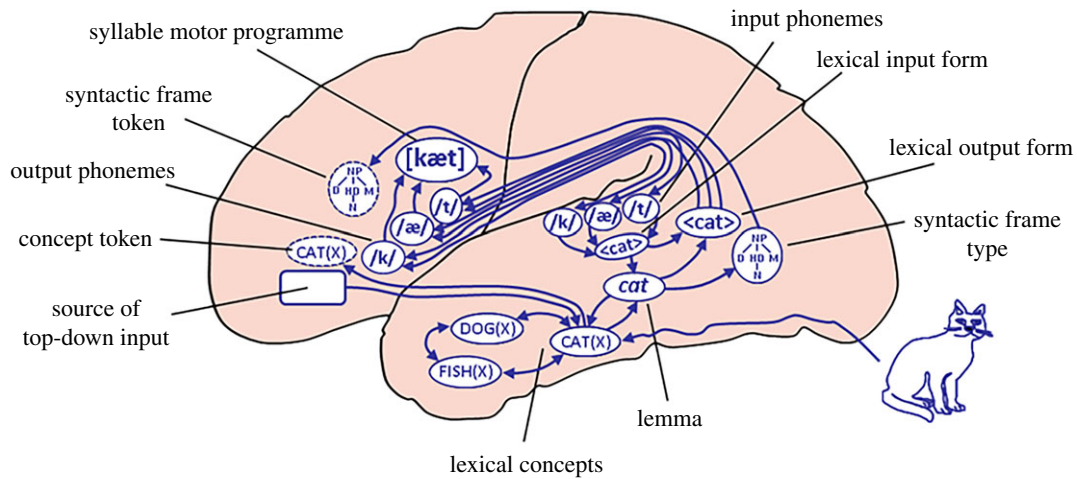


Figure 1. Dorsal and ventral anatomic connections relate the lexical concepts retrieved from temporal cortex areas to inferior frontal cortex, which is crucial for generating context-dependent tokens for the production of an utterance (from the Weaver++/ARC model of [40]). A similar story holds for comprehension (see the Cycle Model of Baggio [6,41]). (Online version in colour.)

emotion-related areas in the Lai *et al.* [21] study. The semantic map misses one of the central aspects of making meaning, namely the outcome of combinatorial operations on the semantic building blocks retrieved from memory. The problem is that this remains implicit in the Huth *et al.* [5] study, since a functional decomposition of semantic operations is lacking. This is why, despite the technical ingenuity, this study does not contribute much to our understanding of making meaning. The interdependence of the computational, algorithmic and implementational levels is a necessary ingredient for a mechanistic account of meaning, or for that matter any cognitive function. A naive reductionism will not do [31].

5. What are the mechanisms of meaning?

In the remaining few paragraphs, I would not be able to do justice to the many aspects of meaning and the mechanistic underpinnings along the lines discussed above. Two aspects are, however, crucial. One is the retrieval of word meanings from memory. Based on visual or auditory input, lexical information is retrieved from long-term memory (i.e. the mental lexicon). Lexical–semantic representations are instantiated by neuronal ensembles in (left) temporal and inferior parietal cortex. Especially for concrete words, the core conceptual features are connected to a widely distributed network of sensory and action areas adding information specified in a format that is tailored to the characteristics of the sensory, motor and emotional systems.

With regard to how word meanings are encoded, there is a general divide between atomic and holistic accounts [32]. The atomic account decomposes word meanings into more elementary semantic features [33,34]. The holistic account sees words as organized in semantic fields, whose organization is determined by semantic similarities between concepts. In this case, word meaning can be inferred from the company that a given word keeps. I do not think these two views are mutually exclusive (see [7], for a similar position). Neuropsychological patient studies revealed that for concrete words semantic information is partly distributed among the relevant sensory-motor cortices [35]. More recently, this idea has been further elaborated in the embodied cognition view [36,37], with additional support from neuroimaging studies [4]. Overall, the empirical

evidence indicates that a word's semantic representation is distributed over multiple areas in cortex.

Assuming a distributed representation of word meaning, semantic hub areas are crucial to form an interface between lexical meanings and other relevant word-related information (i.e. word forms, morphosyntax). Semantic hubs collect the outcome of the distributed modality-specific activation patterns for the selection of a single lexical candidate. The hubs integrate the semantic information in an amodal format. Two areas are likely candidates for the semantic hub function. One is the anterior temporal lobe [38], most likely a hub for concepts related to concrete nouns. The other area is the angular gyrus, which is suggested to be a convergence zone (hub) for event concepts [39]. These conceptual hubs are crucial for interfacing the conceptual information to both phonological information in posterior superior temporal cortex and lexical–syntactic information in posterior middle temporal cortex (figure 1; [42,43]).

The second major aspect of semantic processing is the unification of lexical–semantic building blocks into an interpretation of the full utterance in its conversational or discourse setting. The required combinatorial semantic operations are characterized by enriched compositionality, neurally instantiated by activation cycles between left inferior frontal and temporo-parietal areas [6]. Given the short decay times of feedforward synaptic transmission, feedback activations from LIFC with their longer decay times are necessary to keep lexical–semantic information online in the service of combinatorial operations [41]. The LIFC is the source of the top-down input to temporal cortex. In this way lexical information is bound into larger structures spanning multi-word utterances (figure 1).

The integration of the two major aspects of semantic processing requires an understanding of the communication between the distant relevant areas involved. One way to investigate the communication between these areas is by analysing power in different frequency bands, such as alpha, theta and beta. We recorded Magnetoencephalograms (MEGs) from a large group of subjects while they were reading sentences. We then applied a Granger causality analysis to this large MEG dataset to determine the direction of communication between the language-relevant areas. This analysis showed an outflow from middle temporal cortex to anterior temporal areas and left frontal cortex ([44]; figure 2). Connections from temporal regions peaked at alpha frequency. By contrast,

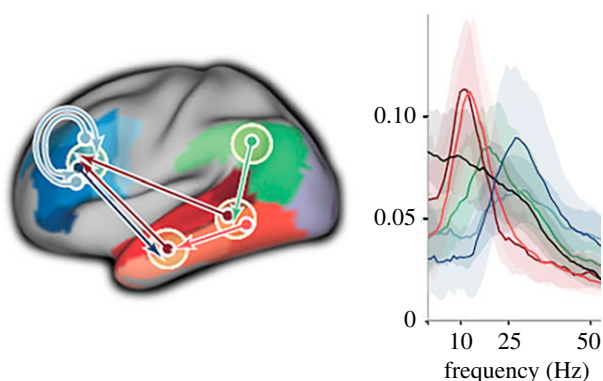


Figure 2. Schematic of the directed rhythmic cortico-cortical interactions between temporo-parietal and frontal cortex, grouped according to the cortical output area. The coloured arrows refer to the power spectra shown on the right (adapted from [44]). (Online version in colour.)

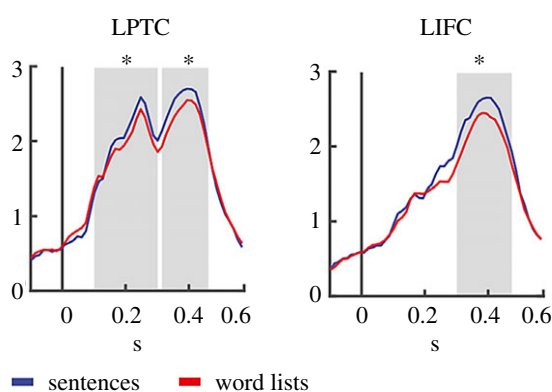


Figure 3. MEG activations to words in sentences versus words in word lists. Words in sentences elicited higher amplitudes than words in word lists in both LPTC and LIFC. A positive correlation was found in the sentences only between the peak activity in temporal cortex (second peak in LPTC: LPTC2) and frontal cortex (LIFC). The larger an individual's effect between 315 and 461 ms was in LPTC2, the larger the effect for the same time period in LIFC (after [46]). (Online version in colour.)

connections originating from frontal and parietal regions peaked at frequencies in the beta range. These brain rhythms are not themselves the mechanism that computes meaning. Rather, they might play a role in setting up communication protocols between language-relevant regions in temporo-parietal and inferior frontal cortex (cf. the *Communication Through Coherence* proposal by Fries [45]). These rhythms and their interplay promote neuronal communication between areas that are involved in retrieving and constructing semantic representations.

For the same MEG data, event-related fields showed two clearly distinct peaks in left posterior temporal cortex (LPTC), an early peak at about 250 ms, and a later peak around 400 ms (figure 3; [46]). The LIFC only showed the later peak. The peak amplitudes between the second peak in

LPTC and the late peak in LIFC were correlated. This co-modulation of the electromagnetic activity in temporal and frontal cortex is supportive of the Cycle Model [6]. In this way, a dynamic recurrent network is set up in perisylvian cortex, whereby owing to feedback from LIFC lexical–semantic information can maintain its active state over extended time intervals, as required by the need for semantic unification (figure 3).

In language processing, lexical concepts have to be anchored to a situation model. This requires the transition from lexical concept to concept tokens that can be interpreted in the current context [6,7,13] (figure 1). Only whole utterances refer to internal models of the world. Single word meanings usually do not. There might be single word utterances, but they refer because of being a token in a multimodal setting. Tokens are created by the dynamic interaction between top-down input from LIFC and lexical concepts (types) subserved by temporal and parietal cortices. The LIFC is in this framework the hub (or convergence zone) for multimodal context and discourse information that will determine the selection of the relevant token features of the lexical concepts in the service of ongoing utterance interpretation or production (figure 1).

The view I have presented is admittedly very sketchy, and does not do justice to all the relevant studies in the literature. My view is certainly somewhat biased. Moreover, many details need to be filled in. Formalizations are needed, as well as a better understanding of the fundamental principles underlying neural mechanisms of information exchange and memory representations. But I believe the direction is the right one if we want to get a handle on the meaning-making mechanism(s) behind our eyes and between our ears. This direction is based on a concept of mechanism that includes both a functional/algorithmic level and an implementational level. In addition, the making of meaning is based on an interplay between memory retrieval and unification operations.

Data accessibility. This article has no additional data.

Competing interests. We declare we have no competing interests.

Funding. Writing the paper was made possible by the NWO grant Language in Interaction, grant no. 024.001.006.

Acknowledgement. I am grateful to Willem Levelt, Guillermo Montero-Melis and two anonymous reviewers for their helpful comments on an earlier version of this article.

Endnote

¹A different, orthogonal issue is that in our research we use read-outs of the mechanism under investigation. Some of these read-outs are more direct reflections of the underlying mechanism than others. The haemodynamic response measured as a BOLD effect in fMRI is obviously a less direct read-out than a brain oscillation measured in an MEG system. This might affect the number of inferential steps from the measurement to the mechanism. However, this does not reflect the characteristics of the mechanism itself.

References

1. Hauser MD, Chomsky N, Fitch WT. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
2. Guenther FH. 2016 *Neural control of speech*. Cambridge, MA: MIT Press.
3. Dehaene S, Meyniel F, Wacongne C, Wang LP, Pallier C. 2015 The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* **88**, 2–19. (doi:10.1016/j.neuron.2015.09.019)

4. Kemmerer D. 2015 *Cognitive neuroscience of language*. New York, NY: Psychology Press.
5. Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. 2016 Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453. (doi:10.1038/nature17637)
6. Baggio G. 2018 *Meaning in the brain*. Cambridge, MA: MIT Press.
7. Seuren PA. M. 2009 *Language in cognition*. Oxford, UK: Oxford University Press.
8. Clark HH. 1996 *Using language*. Cambridge, UK: Cambridge University Press.
9. Jackendoff R. 2002 *Foundations of language: brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.
10. Hohwy J. 2013 *The predictive mind*. Oxford, UK: Oxford University Press.
11. Swanson LR. 2016 The predictive processing paradigm has roots in Kant. *Front. Syst. Neurosci.* **10**, 13. (doi:10.3389/fnsys.2015.00079)
12. Fodor JA, Pylyshyn ZW. 2015 *Minds without meanings: an essay on the content of concepts*. Cambridge, MA: MIT Press.
13. Jackendoff R. 2012 *A user's guide to thought and meaning*. Oxford, UK: Oxford University Press.
14. Frege G. 1892 Über Sinn und Bedeutung. *Z. Philos. Kritik.* **100**, 25–50 (in German).
15. Partee BH. 1995 Lexical semantics and compositionality. In *Invitation to cognitive science* (ed. D Osherson), pp. 311–360. Cambridge, MA: MIT Press.
16. Özyürek A, Willems RM, Kita S, Hagoort P. 2007 On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.* **4**, 605–616. (doi:10.1162/089892903322370807)
17. Van Berkum JJA, van den Brink D, Tesink C, Kos M, Hagoort P. 2008 The neural integration of speaker and message. *J. Cogn. Neurosci.* **20**, 580–591. (doi:10.1162/jocn.2008.20054)
18. Willems RM, Özyürek A, Hagoort P. 2008 Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *J. Cogn. Neurosci.* **20**, 1235–1249. (doi:10.1162/jocn.2008.20085)
19. Tieu L, Schlenker P, Chemla E. 2019 Linguistic inferences without words. *Proc. Natl Acad. Sci. USA* **116**, 9796–9801. (doi:10.1073/pnas.1821018116)
20. Hagoort P, van Berkum J. 2007 Beyond the sentence given. *Phil. Trans. R. Soc. B* **362**, 801–811. (doi:10.1098/rstb.2007.2089)
21. Lai VT, Willems RM, Hagoort P. 2015 Feel between the lines: implied emotion in sentence comprehension. *J. Cogn. Neurosci.* **27**, 1528–1541. (doi:10.1162/jocn_a_00798)
22. Mahon BZ, Caramazza A. 2008 A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris.* **102**, 59–70. (doi:10.1016/j.jphysparis.2008.03.004)
23. Craver CF, Bechtel W. 2007 Top-down causation without top-down causes. *Biol. Philos.* **22**, 547–563. (doi:10.1007/s10539-006-9028-8)
24. Kaplan DM. 2011 Explanation and description in computational neuroscience. *Synthese* **183**, 339–373. (doi:10.1007/s11229-011-9970-0)
25. Campbell DT. 1974 Downward causation in hierarchically organised biological systems. In *Studies in the philosophy of biology: reduction and related problems* (eds FJ Ayala, TG Dobzhansky), p. 179. Berkeley, CA: University of California Press.
26. Bechtel W. 2017 Explicating top-down causation using networks and dynamics. *Philos. Sci.* **84**, 253–274. (doi:10.1086/690718)
27. Ellis GFR. 2012 Top-down causation and emergence: some comments on mechanisms. *Interface Focus* **2**, 126–140. (doi:10.1098/rsfs.2011.0062)
28. Tinbergen N. 1963 On aims and methods of ethology. *Z. Tierpsychol.* **20**, 410–433.
29. Millikan RG. 1989 In defense of proper functions. *Philos. Sci.* **56**, 288–302. (doi:10.1086/289488)
30. Barsalou LW. 2017 What does semantic tiling of the cortex tell us about semantics?. *Neuropsychologia* **105**, 18–38. (doi:10.1016/j.neuropsychologia.2017.04.011)
31. Krakauer JW, Ghazanfar AA, Gomez-Marín A, MacIver MA, Poeppel D. 2017 Neuroscience needs behavior: correcting a reductionist bias. *Neuron* **93**, 480–490. (doi:10.1016/j.neuron.2016.12.041)
32. Aitchison J. 1987 *Words in the mind, an introduction to the mental lexicon*. Oxford, UK: Blackwell Publication.
33. Fromkin V, Rodman R, Hyams N. 2013 *An introduction to language*. Boston, MA: Wadsworth Cengage Learning.
34. Binder JR. 2016 In defense of abstract conceptual representations. *Psychon. Bull. Rev.* **23**, 1096–1108. (doi:10.3758/s13423-015-0909-1)
35. Allport DA. 1985 Distributed memory, modular subsystems and dysphasia. In *Current perspectives in dysphasia* (eds SK Newman, R Epstein), pp. 32–60. Edinburgh, UK: Churchill Livingstone.
36. Barsalou LW. 1999 Perceptual symbol systems. *Behav. Brain Sci.* **22**, 577.
37. Barsalou LW. 2008 Grounded cognition. *Annu. Rev. Psychol.* **59**, 617–645. (doi:10.1146/annurev.psych.59.103006.093639)
38. Visser M, Jefferies E, Ralph MAL. 2010 Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J. Cogn. Neurosci.* **22**, 1083–1094. (doi:10.1162/jocn.2009.21309)
39. Binder JR, Desai RH. 2011 The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527–536.
40. Roelofs A. 2019 A computational account of progressive and post-stroke aphasias: integrating psycholinguistic, neuroimaging, and clinical findings. Poster presented at 'Crossing the boundaries: language in interaction', Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, 9 April 2019.
41. Baggio G, Hagoort P. 2011 The balance between memory and unification in semantics: a dynamic account of the N400. *Lang. Cogn. Process.* **26**, 1338–1367. (doi:10.1080/01690965.2010.542671)
42. Roelofs A. 2014 A dorsal-pathway account of aphasic language production: the WEAVER plus plus/ARC model. *Cortex* **59**, 33–48. (doi:10.1016/j.cortex.2014.07.001)
43. Hagoort P. 2005 On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* **9**, 416–423. (doi:10.1016/j.tics.2006.07.004)
44. Schoffelen JM, Hulten A, Lam N, Marquand AF, Udden J, Hagoort P. 2017 Frequency-specific directed interactions in the human brain network for language. *Proc. Natl Acad. Sci. USA* **114**, 8083–8088. (doi:10.1073/pnas.1703155114)
45. Fries P. 2015 Rhythms for cognition: communication through coherence. *Neuron* **88**, 220–235. (doi:10.1016/j.neuron.2015.09.034)
46. Hulten A, Schoffelen JM, Udden J, Lam NHL, Hagoort P. 2019 How the brain makes sense beyond the processing of single words - an MEG study. *Neuroimage* **186**, 586–594. (doi:10.1016/j.neuroimage.2018.11.035)