

COMPUTATIONAL NEUROSCIENCE

Conceptual grounding of language in action and perception: a neurocomputational model of the emergence of category specificity and semantic hubs

Max Garagnani^{1,2} and Friedemann Pulvermüller¹¹Brain Language Laboratory, Department of Philosophy and Humanities, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany²Centre for Robotics and Neural Systems (CRNS), University of Plymouth, Plymouth, Devon, UK**Keywords:** cell assembly, cortical connectivity, functional differentiation, Hebbian learning, word meaning

Edited by Helen Barbas

Received 7 May 2015, revised 29 November 2015, accepted 30 November 2015

Abstract

Current neurobiological accounts of language and cognition offer diverging views on the questions of ‘where’ and ‘how’ semantic information is stored and processed in the human brain. Neuroimaging data showing consistent activation of different multi-modal areas during word and sentence comprehension suggest that all meanings are processed indistinctively, by a set of general semantic centres or ‘hubs’. However, words belonging to specific semantic categories selectively activate modality-preferential areas; for example, action-related words spark activity in dorsal motor cortex, whereas object-related ones activate ventral visual areas. The evidence for category-specific and category-general semantic areas begs for a unifying explanation, able to integrate the emergence of both. Here, a neurobiological model offering such an explanation is described. Using a neural architecture replicating anatomical and neurophysiological features of frontal, occipital and temporal cortices, basic aspects of word learning and semantic grounding in action and perception were simulated. As the network underwent training, distributed lexico-semantic circuits spontaneously emerged. These circuits exhibited different cortical distributions that reached into dorsal-motor or ventral-visual areas, reflecting the correlated category-specific sensorimotor patterns that co-occurred during action- or object-related semantic grounding, respectively. Crucially, substantial numbers of neurons of both types of distributed circuits emerged in areas interfacing between modality-preferential regions, i.e. in multimodal connection hubs, which therefore became loci of general semantic binding. By relating neuroanatomical structure and cellular-level learning mechanisms with system-level cognitive function, this model offers a neurobiological account of category-general and category-specific semantic areas based on the different cortical distributions of the underlying semantic circuits.

Introduction

Current semantic theories offer diverging perspectives on how word meaning is acquired, represented and processed in the human brain. One tradition views the cognitive basis of meaning as a symbolic, ‘amodal’ system containing abstract representations defined in terms of semantic features or correlations between words, bearing no explicit relationship with the concrete objects and actions the symbols are used to speak about (Collins & Loftus, 1975; Potter, 1979; Ellis & Young, 1988). A putative brain basis for such a system of symbolic-conceptual representations has been attributed to ‘semantic hubs’, higher-association multimodal areas located in frontal, temporal and parietal cortices that have been found active during, or even to be necessary for, semantic processing (Price, 2000; Bookheimer, 2002; Devlin *et al.*, 2003; Vigneau *et al.*, 2006; Patterson *et al.*, 2007; Binder & Desai, 2011; Pulvermüller, 2013).

A second tradition builds on the insight that semantic knowledge requires grounding in the real world (Searle, 1980; Harnad, 1990). Symbols are used to speak about specific objects, actions and other entities; access to such semantic knowledge likely involves processing sensorimotor information in modality-preferential areas of the cortex: for example, understanding an object-related word such as ‘cat’ should reactivate visual areas, whereas an action word like ‘grasp’ motor ones. Support for modality-specific semantic processes comes from neuropsychological and neuroimaging studies showing semantic-category specificity of cortical activations and category-specific deficits after lesions in modality-preferential areas (Shallice, 1988; Martin, 2007). For example, word and sentence comprehension induce category-specific activations in modality-preferential motor and sensory (visual, auditory, olfactory and gustatory) areas (Barsalou, 2008; Binder & Desai, 2011; Kiefer & Pulvermüller, 2012; Pulvermüller, 2013; Kemmerer, 2015).

Here, we attempt to explain and integrate the above experimental data and theories by means of a single neurobiological model. Our

Correspondence: Dr M. Garagnani, as above.

E-mail: M.Garagnani@fu-berlin.de

hypothesis is that the semantic-category-specific and -general functional behaviours observed in distinct cortical areas are a direct consequence of well-established neuroscience facts and principles, and should therefore spontaneously emerge in specific parts of the cortex as a result of sensorimotor correlations and associative learning. To address this hypothesis, we implemented a neurocomputational model of relevant primary, secondary and higher-association areas in the frontal, temporal and occipital lobes of the human brain and simulated elementary processes of language acquisition in it, focusing specifically on the semantic grounding of object- and action-related words.

A range of previous connectionist models successfully addressed aspects of language learning and processing, although most did not attempt to replicate the neuroanatomy of the cortical areas concerned with the corresponding brain processes (Elman *et al.*, 1996; Plunkett, 1997; Dell *et al.*, 1999; Plaut & Gonnerman, 2000; Christiansen & Chater, 2001). While some recent works did take connectivity structure into account (Husain *et al.*, 2004; Guenther *et al.*, 2006; Ueno *et al.*, 2011), they either did not incorporate learning mechanisms, or made use of ones (e.g. back-propagation) whose neurobiological plausibility is questionable (Mazzoni *et al.*, 1991; Braitenberg & Schüz, 1998; O'Reilly, 1998). By contrast, here we implemented only learning mechanisms that mimic well-documented neurophysiological phenomena of Hebbian synaptic plasticity (Artola & Singer, 1993), so as to show how associative learning and neuroanatomical structure interact to bring about the two different functional behaviours in semantic processing described above (category-specific and -general) in distinct cortical areas. Similar approaches have previously been used to provide neurobiological accounts for the spontaneous emergence and cortical topography of resting state activity, perceptual and action decisions, and working memory (Deco *et al.*, 2013a,b; Garagnani & Pulvermüller, 2013; Pulvermüller & Garagnani, 2014).

Materials and methods

We take a semantic grounding perspective (Barsalou, 1999; Pulvermüller, 1999) and postulate that learning the meaning of at least a basic set of words and symbols of any language involves the formation of referential-semantic links between their 'form' – the articulatory- and acoustic-phonological patterns in the case of single spoken words – and the types of object or action these symbols are typically used to speak about (Barsalou, 2008; Pulvermüller & Fadiga, 2010; Glenberg & Gallese, 2012; Pulvermüller, 2013). Accordingly, we use a neurocomputational model of relevant peri- and extra-sylvian cortical areas (see below) to simulate the spontaneous emergence of such associative links.

General structure and features of the model

The neural model consists of 12 identical interconnected 'areas' of graded-response cells, implementing random and sparse between- and within-area connections (Fig. 1B and C; Appendix A). Each model area consists of two layers (or 'banks') of excitatory and inhibitory cells, and simulates a specific cortical area (Fig. 1A).

- As information in the articulatory motor cortex relates to the production of a word form, and information in the auditory cortex to the acoustic perception of such a form, both of these perisylvian areas (labelled M1_i and A1, respectively) were included. Moreover, as information about the objects about which we speak when using words such as 'sun' comes in through the primary visual cortex, and because a self-performed action related to the meaning

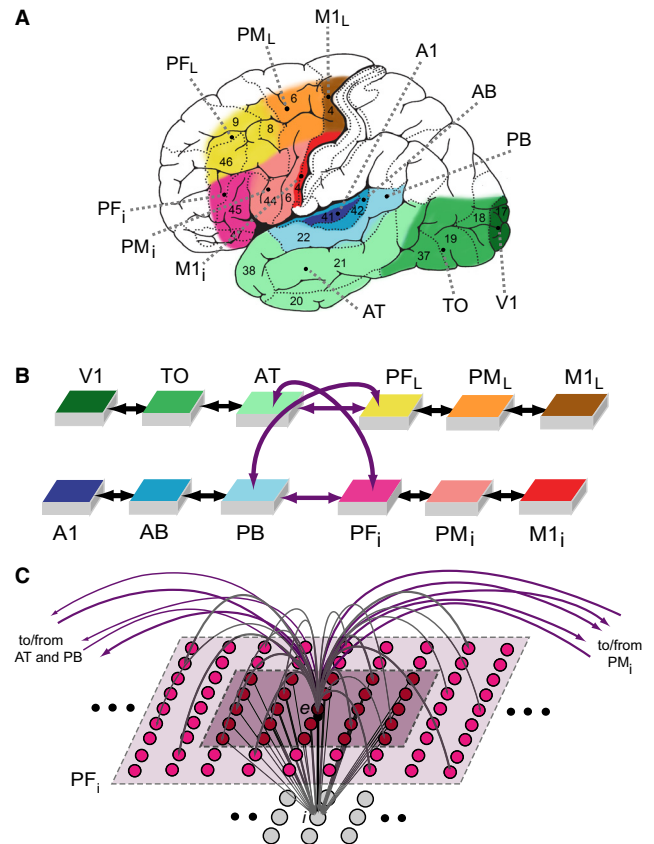


FIG. 1. Model of lexical and semantic mechanisms. The 12 cortical areas modelled (A), their global connectivity architecture (B), and aspects of the micro-structure of their connectivity (C) are illustrated. (A) Six perisylvian and six extrasylvian areas are shown, each including a dorsal (frontal) and a ventral (temporal) part. Perisylvian areas include three areas in inferior frontal gyrus (red colours), including inferior-prefrontal (PF_i), premotor (PM_i) and primary motor cortex (M1_i), and three areas in the superior temporal lobe (in blue), including auditory parabelt (PB), auditory belt (AB) and primary auditory cortex (A1). These areas can store correlations between neuronal activations carrying articulatory-phonological and corresponding acoustic-phonological information, for example when phonemes, syllables and spoken word forms are being articulated (activity in M1_i) and acoustic features of these spoken words are simultaneously perceived (stimulation of primary auditory cortex, A1). Extrasylvian areas include three areas in lateral/superior frontal cortex (yellow to brown), including dorsolateral prefrontal (PF_L), premotor (PM_L) and primary motor cortex (M1_L), and three areas forming the occipito-temporal ('what') visual stream of object processing (green), including anterior-temporal (AT), temporo-occipital (TO) and early visual areas (V1). Together with the perisylvian ones, these extrasylvian areas can store correlations between neuronal activations carrying semantic information, for example when words are used (activity in all perisylvian areas) to speak about objects present in the environment (activity in V1, TO, AT) or about actions the individual engages in (activity in M1_L, PM_L, PF_L). Numbers indicate Brodmann areas. (B) Schematic illustration of all 12 modelled areas and the known between-area connections implemented. The colours indicate correspondence between cortical and model areas. See text for a detailed description of the neuroanatomical evidence supporting the implemented connectivity structure. (C) Schematics of micro-connectivity of one of the 7500 single excitatory neural elements modelled (labelled 'e'). Within-area excitatory links (in grey) to and from 'cell' e are random and sparse, and limited to a local (19 × 19) neighbourhood (light-pink shaded area). Lateral inhibition between e and neighbouring excitatory elements is realised as follows: the underlying cell 'i' inhibits e in proportion to the total excitatory input it receives from the 5 × 5 neighbourhood (dark-purple shaded area); by means of analogous connections (not depicted), e inhibits all of its neighbours. Each pair (e, i) of model cells is taken to represent an entire cluster or column (grey matter under approximately 0.25 mm² of cortical surface) of pyramidal cells and the inhibitory interneurons therein. See Appendix A for a complete specification of the model.

of action-related words such as 'grasp' or 'run' is controlled by the lateral and superior motor cortex, the model also included primary visual and dorsolateral motor cortices (areas V1 and M1_L).

- In addition to primary cortices, 'higher' secondary and multimodal regions known to have strong neuroanatomical links with the above four primary sensorimotor cortices were modelled (see 'Network structure and connectivity of the simulated brain areas' below). These were secondary inferotemporo-occipital visual, auditory belt, and inferior and lateral premotor cortex (TO, AB, PM_i, PM_L) and, respectively, adjacent multimodal anterior-temporal, superior-temporal (auditory parabelt) and inferior and dorsolateral prefrontal cortices (AT, PB, PF_i, PF_L).

The architecture builds upon and extends an existing six-area model of the left perisylvian language cortex that was developed to simulate the emergence of memory traces for (meaningless) spoken words in the cortex and explain neurophysiological responses to linguistic stimuli (Garagnani *et al.*, 2007, 2008; Garagnani & Pulvermüller, 2011). As in previous versions of the architecture, all functional and structural features implemented closely reflect well-documented properties of the human cortex, including the following:

- known structure of the neuroanatomical links between the modelled sensorimotor and multimodal brain systems;
- sparse, patchy and topographic between- and within-area connections, with probability of a synaptic link existing between two cells falling off with their distance (Kaas, 1997; Braitenberg & Schüz, 1998);
- local lateral (mutual) inhibition (Fig. 1C) and area-specific global regulation mechanisms (Braitenberg, 1978b; Yuille & Geiger, 2003);
- Hebbian learning mechanisms, simulating synaptic plasticity phenomena of long-term potentiation and depression (Artola & Singer, 1993);
- neurophysiological dynamics of single cells including temporal summation of inputs, sigmoid transformation of membrane potentials into neuronal outputs, and adaptation (Matthews, 2001);
- presence of uniform white noise (simulating spontaneous, baseline neuronal firing) in all parts of the network at all times (Rolls & Deco, 2010).

A detailed description of the connectivity structure [point (1) above] is provided in 'Network structure and connectivity of the simulated brain areas' below. The neural-level features [points (2–6)] are identical to those implemented in previous versions of the model (Garagnani *et al.*, 2008, 2009b; Garagnani & Pulvermüller, 2011, 2013; Pulvermüller & Garagnani, 2014). For completeness, they are summarized again in Appendix A.

Note that we strived to model only mechanisms that have a physiological correlate, and implemented a connectivity structure that closely reflects known neuroanatomical pathways between the modelled cortical areas. A direct comparison of the effectiveness and biological accuracy of the learning rule used here with that of other, known brain-inspired synaptic plasticity rules is provided in Garagnani *et al.* (2009b). Although the implementation of non-strictly biologically realistic aspects (e.g. 'all-to-all' connectivity, or back-propagation learning; Rumelhart *et al.*, 1986) would likely lead to a more efficient – from an engineering point of view – system (i.e. exhibiting better learning performance or increased memory capacity), the adoption of any such non-biological features would undermine the neuroscientific relevance of the model, preventing us from using the present simulation results as a basis to make claims about corresponding brain processes, in focus here.

Previous simulations have shown that, subsequent to the repeated concomitant presentation of activation patterns to (possibly indirectly) linked model areas, networks including the above range of neurobiologically realistic features give rise to the formation of distributed associative circuits (Garagnani *et al.*, 2007, 2008, 2009b) corresponding to what Hebb once postulated and labelled 'cell assemblies' (CAs; Hebb, 1949). CAs can be defined structurally as sets of nerve cells that are '... more strongly connected to each other than to other neurons' (Braitenberg, 1978a). They constitute 'memory circuits' that emerge as a result of correlational learning mechanisms and bind together sets of neurons that are frequently co-active (Hebb, 1949; Braitenberg, 1978a; Palm, 1982). Once developed, CAs behave as coherent functional units with two quasi-stable states ('on' and 'off'; Garagnani *et al.*, 2007, 2008, 2009b; Pulvermüller & Garagnani, 2014). Cortical CAs whose formation is driven by correlated sensory and motor information are also called action-perception circuits (Pulvermüller & Fadiga, 2010). Here, we simulated the spontaneous formation of CAs linking symbols (word forms) to aspects of their meaning manifest in information about objects or actions they refer to.

Network structure and connectivity of the simulated brain areas

The original model of the language cortex simulated six left-perisylvian areas (three in the inferior fronto-central cortex and three in the superior-temporal auditory system; Fig. 1A); here this model is augmented with six new areas (and relevant connections between them) having a role in transferring and processing semantically relevant information. Because these 'semantic' areas are outside the perisylvian (language) cortex, in the remainder of this article they are referred to as 'extrasyylvian' areas. The extrasyylvian areas include dorsolateral fronto-central motor, premotor and prefrontal cortices (M1_L, PM_L, PF_L), and three areas constituting the ventral occipito-temporal visual 'what' stream (V1, TO, AT). Thus, within both peri- and extrasyylvian systems, we distinguished between a 'dorsal stream' section, situated in the frontocentral cortex (depicted in different shades of red/yellow) and a 'ventral stream' section, in the temporal and occipital cortex (in shades of blue/green) were distinguished between.

Neuroanatomical evidence shows that adjacent cortical areas tend to be connected with each other through next-neighbour between-area links (Pandya & Yeterian, 1985; Young *et al.*, 1994, 1995). These exist within each triplet of areas of the four systems modelled, that is, amongst: (1) inferior frontal areas PF_i – PM_i – M1_i; (2) superior-lateral frontal areas PF_L – PM_L – M1_L (see also Arikuni *et al.*, 1988; Lu *et al.*, 1994; Dum & Strick, 2002, 2005); (3) superior and lateral auditory areas A1 – AB – PB (Pandya, 1995; Kaas & Hackett, 2000; Rauschecker & Tian, 2000); and (4) inferior temporo-occipital areas V1 – TO – AT (Distler *et al.*, 1993; Nakamura *et al.*, 1993).

Evidence also indicates the presence of long-distance cortico-cortical links (see purple arrows in Fig. 1B) connecting areas distant from each other. Amongst the long-distance links within the fronto-temporo-occipital cortex, only the well-documented mutual and reciprocal connections between anterior temporal, superior parabelt, and inferior, and posterior-superior-lateral prefrontal areas were implemented. The connections between anterior (and middle), inferior, and posterior-superior temporal cortex (areas AT, PB in Fig. 1B) and inferior prefrontal (and premotor) cortex (PF_i) are realised by the arcuate and uncinata fascicles (Makris *et al.*, 1999; Romanski *et al.*, 1999b; Petrides & Pandya, 2001, 2009; Catani *et al.*, 2005; Parker *et al.*, 2005; Romanski, 2007; Rilling *et al.*, 2008; Makris & Pandya, 2009; Petrides *et al.*, 2012; Rilling, 2014). Dorsolateral prefrontal (and premo-

tor) cortex (PF_L) is reciprocally linked to anterior and inferior temporal regions (AT; Pandya & Barnes, 1987; Ungerleider *et al.*, 1989; Webster *et al.*, 1994), as well as to the superior temporal cortex (PB) via the extreme capsule (Pandya & Barnes, 1987; Romanski *et al.*, 1999a,b; Schmahmann *et al.*, 2007; Dick *et al.*, 2014).

Simulating semantic symbol grounding

At the onset of learning the network was in a 'naïve' state, i.e. one in which all between- and within-area synaptic links connecting single cells were established at random, as were their synaptic efficacies (weights). Word learning and semantic grounding were then simulated by means of repeated 'learning trials', involving concomitant stimulation of primary areas of the network, as described below.

As each spoken word form is characterized by an articulatory motor schema and an acoustic schema, each learning trial entailed concurrent stimulation of inferior-frontal primary motor and superior-temporal primary auditory areas (perisylvian primary areas M1_L and A1 in Fig. 1). As some words are typically used to speak about visually perceivable objects and one typical learning situation for such words is the use of the word while the referent object is present (Vouloumanos & Werker, 2009; Barros-Loscertales *et al.*, 2012), learning of object-related words was simulated by concurrent stimulation of both perisylvian primary areas plus visual cortex (V1). Similarly, learning aspects of action-related word meaning involved simultaneous activation of primary perisylvian and lateral motor areas (M1_L); this was meant to simulate a situation in which action words are used when the learning child performs the corresponding action (Tomasello & Kruger, 1992).

The learning of six object- and six action-related words was simulated, each by concurrent stimulation of three of the four primary areas, with one specific sensorimotor pattern of neuronal activation for each word. Each sensorimotor pattern consisted of a set of 19 cells per primary area (57 cells in total), randomly selected amongst the 25-by-25 cells forming one area (about 3% of cells). Each of the 12 sensorimotor patterns was presented in 3000 learning trials, resulting in a total of 36 000 (randomly ordered) trials. (This number was chosen empirically, on the basis of previous simulations obtained with six-area architectures; such studies showed the presence of cell-assembly circuits already after 50–100 trials, and no substantial changes occurring between 1000 and 2000 presentations; Garagnani *et al.*, 2009b. Here the training was extended to 3000 presentations per pattern, as the network had to develop CA circuits spanning nine instead of six interconnected areas, linking three patterns instead of just two.) Therefore, the same 'core' of neurons were stimulated during each presentation of a given pattern; however, white noise was always present and overlaid the sensorimotor input patterns, so as to account for a degree of variability in the physical features of word forms and semantically relevant objects and actions. Each learning trial lasted 16 simulation-time steps (equivalent to approximately 300 ms) and was followed by a resting interval of variable duration during which no input was provided until activity had returned to baseline. A new trial started as soon as the global inhibition levels in both areas PF_i and PB dropped below a pre-specified threshold (0.65 in the present simulations). As object words are less informative about motor activities than action words, and the latter typically convey less visual information than the former, a static noise pattern was presented to the 'non-partaking' primary area during training. Hence, in each action- (object-) related word learning trial, area V1 (M1_L) was stimulated with a different random pattern of 19 cells. This was intended to mimic the presum-

ably larger variability of the above relationships (or, equivalently, the lower degrees of correlation).

Thirteen different instances of randomly initialized networks having the architecture described above were implemented and subjected to the same learning procedure, each instance being trained with a different set of 12 sensorimotor patterns. As both action and object word meanings may be acquired even if congruent visual or motor information is not consistently present in each episode of learning, the ability of the model to develop word circuits when the (modality-specific) semantic component of the input pattern is provided only in a fraction of the learning trials was also investigated. To do this, the above set of simulations was repeated under three different conditions, in which the fraction of learning trials containing semantic input varied from the initial 100% to 75, 66.7 and 50% (these fractions are the result of replacing the pattern normally presented as input to V1 or M1_L with a random, static one once every four, three, and every other trial, respectively). Again, 13 different instances of randomly initialized networks were trained in each condition.

Data analysis

As further explained in the Results below, the training led to the emergence of CA circuits in the network, that is, sub-networks of strongly and reciprocally connected cells linking together specific sensorimotor patterns in primary areas by way of cells in intermediary areas. The following procedure was applied to define and quantify the emerging CAs.

After training, the neurons forming each of the 12 CAs across the different network areas were identified. To this end, the response of all 7500 excitatory cells to each of the 12 word-form patterns was recorded. More precisely, the time-averaged output (firing rate) of each excitatory cell was estimated over the 15 simulation steps following a single test-presentation of the auditory and articulatory patterns of a learnt word form (no semantic input was provided). An excitatory cell was then considered a member of the CA for pattern w if and only if its (estimated) time-averaged response to w reached a given threshold θ . The threshold θ was area- and cell-assembly specific, and defined as a fraction γ of the maximal single-cell response in that area to pattern w . More formally,

$$\theta = \theta_A(w) = \gamma \max_{x \in A} \overline{O(x, t)}_w$$

where $\overline{O(x, t)}_w$ is the estimated time-averaged response of a cell x in area A to word pattern w , and $\gamma \in [0, 1]$ is a constant (function $O(x, t)$ is defined in Appendix A). For the statistical analysis (see below) $\gamma = 0.50$ was used; this value was chosen on the basis of simulation results obtained with the present and previous networks (Garagnani *et al.*, 2008, 2009b). Following standard definitions in the literature on auto-associative memories (Braitenberg, 1978a; Palm, 1990), only excitatory cells were considered to be part of an assembly.

This definition yields specific numbers of cells per area for each CA that emerged during learning. For each of the 13 network instances, per-area CA-cell numbers were averaged over the six object-related words and over the six action-related words. To statistically test for possible differences in CA topographies between word types, per-area numbers of CA cells obtained from the 13 network instances were submitted to repeated-measure analyses of variance (ANOVAS). A four-way ANOVA on the data from all 12 areas was performed, with the factors 'extra/perisylvian (ExtraPeri)' (two levels: perisylvian = {A1, AB, PB, M1_i, PM_i, PF_i}; extrasyl-

vian = {V1, TO, AT, M1_L, PM_L, PF_L}), ‘frontotemporal (Fronto-Temp)’ (two levels: frontal areas = {M1_L, PM_L, PF_L, M1_i, PM_i, PF_i}, temporal areas = {A1, AB, PB, V1, TO, AT}), ‘modality-specific vs. multimodal (ModSpecificity)’ (three levels: primary unimodal = {A1, V1, M1_L, M1_i}, secondary mesomodal = {TO, AB, PM_L, PM_i} and multimodal = {PB, AT, PF_L, PF_i}), and ‘Word-Type’ (two levels: object-, action-related). Furthermore, two separate three-way ANOVAs were run on the data from the six extrasylvian and six perisylvian areas (factors ‘FrontoTemp’, ‘ModSpecificity’ and ‘WordType’, as above).

Results

Figure 2 depicts representative examples of CA topographies emerged during simulated learning of twelve words semantically grounded in either object (left) or action (right) information. CA circuits of the two semantic types exhibit similar distributions over the perisylvian cortex, with the highest CA-cell densities emerging in the multimodal areas PF_i and PB. By contrast, extrasylvian motor and visual areas appear to exhibit a double dissociation: action-related word learning yields CAs including cells in lateral premotor and even primary motor cortex of the model, but weakly developed or virtually absent in visual areas TO and V1. Conversely, learning words with an object-related meaning seems to produce circuits biased towards the visual system. Finally, CA circuits for both action- and object-related words appear to include comparably large cell densities in multimodal extrasylvian areas AT and PF_L.

Figure 3 illustrates examples of CA-circuit activation dynamics during two simulated word-recognition episodes. Here, only the ‘auditory’ component (area A1) of a learnt sensorimotor word-pattern was presented as input to the network, causing the ‘ignition’ of the CA circuit that had emerged for that specific word. As visible in the figure, this is a near-simultaneous activation process that involves several areas of the network. In line with the differential topographies shown in Fig. 2, object-related word recognition activity extends well into areas V1 and TO, but not to M1_L, and only marginally to PM_L, whereas action-word CA ignition reaches M1_L and PM_L, but not V1, and only marginally TO. Importantly, as CAs for words of either category heavily draw upon the four ‘central’ (perisylvian and extrasylvian) multimodal hub areas, simulated object- (Fig. 3A) and action-related (Fig. 3B) word-recognition processes appear to induce comparable levels of activity there.

The results of the statistical analysis presented in Fig. 4 fully confirmed the above empirical observations. The graphs in Fig. 4A and B plot the number of CA cells per area that emerged with the training, averaged across 13 different network instances (CA cells were identified using the definition given in ‘Data analysis’). The four-way ANOVA run on the data from all 12 areas revealed a main effect of ModSpecificity ($F_{2,11} = 2345$, $P < 0.001$), with generally more CA cells and therefore higher assembly-cell density in the multimodal than in the secondary mesomodal ($t_{12} = 48.9$, $P < 0.001$), and in the secondary than in the primary unimodal ($t_{12} = 14.6$, $P < 0.001$) areas; in addition, a highly significant interaction of the factors ExtraPeri, ModSpecificity, FrontoTemp and WordType ($F_{2,11} = 137.6$, $P < 0.001$) emerged, confirming that CA topographies (i.e. the distributions of their neurons over the areas) differed between word types. As the distinction between peri- and extra-sylvian areas had a significant influence here, topographical word-type effects were further investigated separately for the perisylvian core language areas and the extra-sylvian ones. The three-way ANOVA run on the data from the perisylvian areas did not

provide strong evidence for word types differences across areas (although the respective interaction of topography with word type approached significance: $F_{2,11} = 3.05$, $P = 0.066$). In contrast, extrasylvian areas revealed a highly significant interaction of the factors FrontoTemp and ModSpecificity with WordType ($F_{2,11} = 290$, $P < 0.001$), showing semantic word-category differences in CA topographies in ventral temporo-occipital and dorsolateral frontal areas. There was also a main effect of ModSpecificity in the perisylvian ($F_{2,11} = 1345$, $P < 0.001$) as well as in the extrasylvian ($F_{2,11} = 2549$, $P < 0.001$) areas, analogous to that revealed by the four-way ANOVA.

Further, the significant topographic differences between the circuits of action- and object-related words in extrasylvian model areas were explored. Bonferroni-corrected planned comparison tests (for 12 comparisons, critical threshold $P = 0.0042$) confirmed that larger numbers of cells in V1 and TO were part of circuits for object-related words than for action words ($t_{12} > 8.7$, $P < 0.001$), whereas the opposite applies to M1_L and PM_L ($t_{12} > 8.96$, $P < 0.001$). Extrasylvian multimodal model areas AT and PF_L, which serve as main hubs for visual, auditory and motor information, did not show significant differences between CA types after correcting for multiple comparisons (AT: $t_{12} = 1.92$, $P = 0.079$; PF_L: $t_{12} = 2.88$, $P = 0.014$, n.s. after correction). Analogous *post hoc* *t*-tests investigating possible semantic category differences in the perisylvian areas (Fig. 4A) were all not significant ($t_{12} \leq 1.5$, $P > 0.13$ across all six areas).

Last, the impact that the relative amount of congruent visual or motor information provided during word acquisition – or, equivalently, that the variability in the semantic input – had on the emerging topography of the word circuits was examined. Figure 5 plots the resulting object- and action-word CA distributions as a function of the percentage of learning trials in which the semantic pattern normally associated with a word was replaced by a random one. In line with the results obtained when 100% of the trials included semantic input (data plotted in Fig. 4), *post hoc* *t*-tests revealed that, for all conditions, word-category specificity emerged in primary and secondary extrasylvian areas ($t_{12} > 7.0$, $P < 0.0005$, still significant after correcting for 18 multiple comparisons), with the exception of the 50% case, in which the CAs of the two word types did not differ after application of a conservative threshold ($t_{11} < 3.8$, $P > 0.003$ across all areas, n.s. after correction). Moreover, no significant differences between categories emerged in the two extrasylvian hubs AT and PF_L in any of the conditions ($t_{12} < 3.2$, $P > 0.009$ for all three conditions and two areas, n.s. after correction), or in the perisylvian areas ($t_{12} < 1.8$, $P > 0.11$ across all areas and conditions). These results show that, although inconsistent learning reduces the efficacy of semantic circuit formation, the principal topographical differences indexing category specificity tend to persist.

Discussion

The present model provides a neurobiological explanation for the emergence of category-specific effects in modality-preferential cortices, as well as the consistent activation of multimodal ‘semantic hub’ areas, as observed in the brain during semantic processing. In our neuroanatomically inspired model of perisylvian and extrasylvian fronto-temporo-occipital cortex, the distinct category-specific and category-general functional behaviours emerged spontaneously in different areas as a consequence of the learning process, in particular, of the simulated semantic grounding of words in information about their referent objects and actions. As discussed below, this is

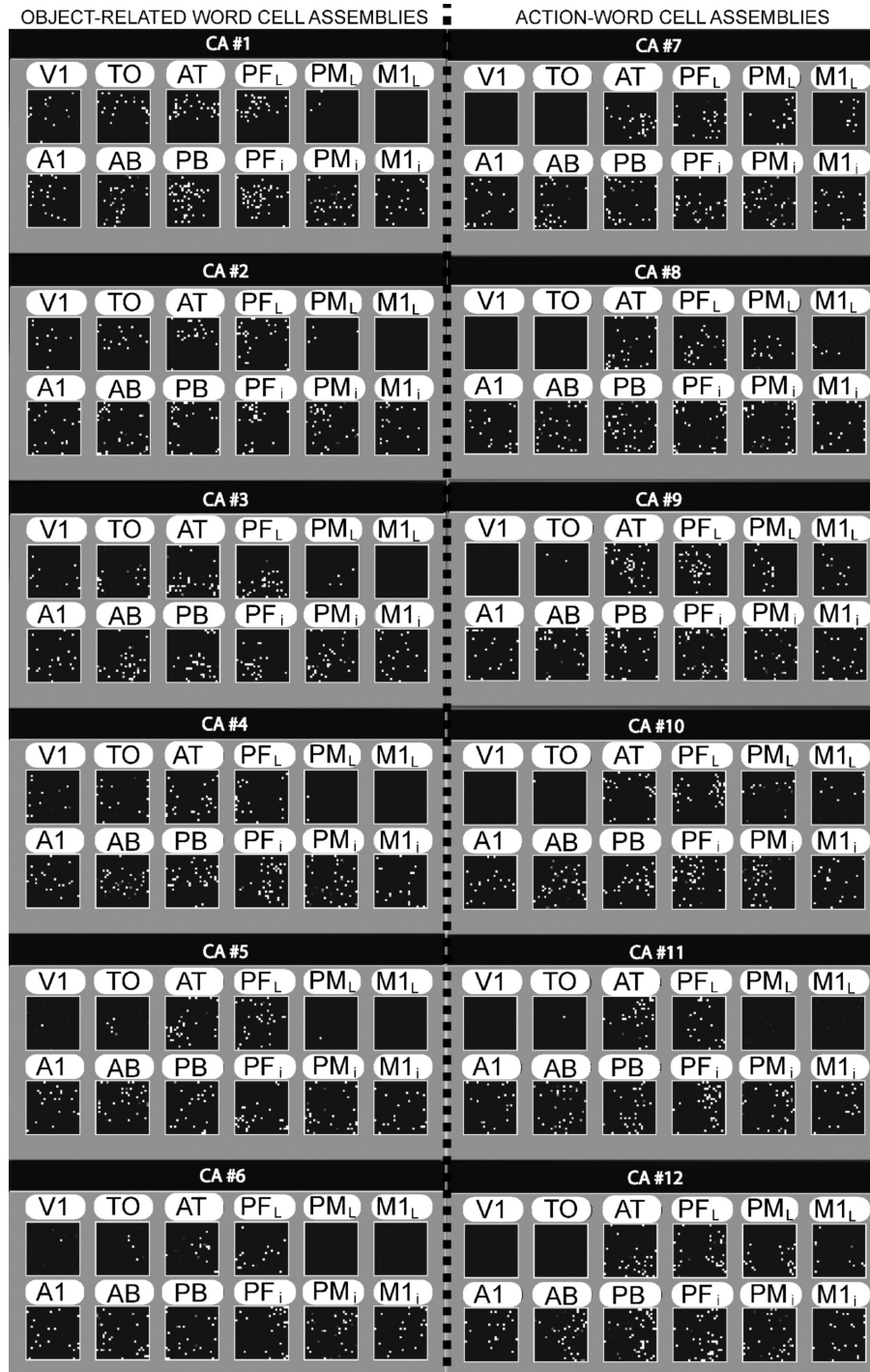


FIG. 2. Distributions of cell assembly (CA) circuits emerging in the model during simulation of word learning in the semantic context of visual perceptions (left-hand side) and actions (right-hand side). Results from a single instance of the network architecture presented in Fig. 1B are shown. Each set of 12 squares depicts the distribution of 'cells' of one specific CA across the 12 network areas. Each white pixel in a square indexes one CA cell. CAs for object-related words extend into higher and primary visual cortex (V1, TO, but not M1_L), linking information about spoken word forms (perisylvian pattern) with information from the visual modality (neural pattern in V1). Network correlates of action-related words extend into lateral motor cortex (M1_L, PM_L, but not V1), thus semantically grounding words in information about actions. Note that, on one occasion, this specific network instance failed to learn the association between spoken word-form and corresponding meaning (see word-related CA #11, which does not reach into area M1_L).

explained by the different topographies of the emerging representations, i.e. the semantic circuits, which, in turn, are determined by the underlying neuroanatomical connectivity structure, Hebbian associative learning mechanisms at work therein, and sensorimotor patterns driving word acquisition and semantic grounding processes.

Semantic hubs vs. category specificity

In the past, semantic processing has been attributed by some to a symbolic system dedicated to processing conceptual information related to words and symbols (Collins & Loftus, 1975; Potter,

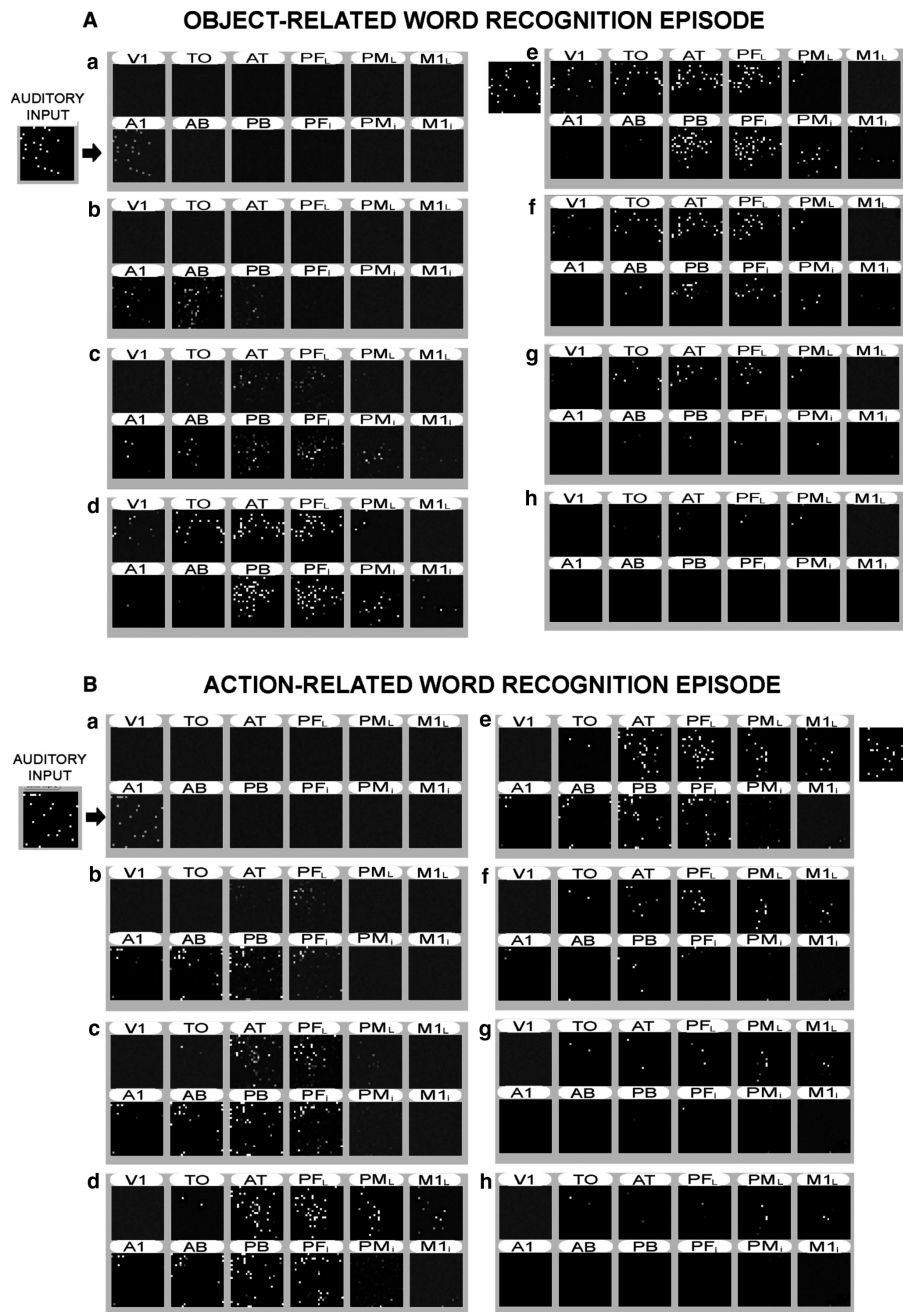


FIG. 3. Activation spreading in the network during simulated word recognition. Representative snapshots of network responses to stimulation of A1 with the 'auditory' component of a learned object-related (A) and action-related (B) word (see CA #1 and CA #9 in Fig. 2, respectively); each set of 12 'squares' captures the network's instantaneous activity. Cell-activity levels are indicated by brightness of pixels; letters indicate chronological order (not simulation time-steps). For ease of visual comparison, the original sensory and motor patterns that the network was trained with are reported in the top-right snapshots (frame 'e') of both (A) and (B). The pattern reconstruction is partial and strongly involves visual areas in (A) and motor areas in (B). See main text for details.

1979; Ellis & Young, 1988). However, neuroimaging and neuropsychological evidence implicating the involvement of several different areas in semantic processing casted doubts on the existence of a single 'amodal' meaning centre, suggesting, instead, the presence of several multimodal hubs, located in higher-association areas of anterior-inferior-temporal (Patterson *et al.*, 2007), middle-temporal (Price, 2000), inferior-parietal (Binder & Desai, 2011) and prefrontal cortex (Bookheimer, 2002; Devlin *et al.*, 2003). At the same time, a growing number of neuroimaging and patient studies (e.g., Warrington & Shallice, 1984; Kemmerer *et al.*, 2012) lend support to a theory of word meaning grounded

in the perception and action systems of the brain (Lakoff & Johnson, 1999; Pulvermüller, 1999; Barsalou, 2008; Pulvermüller & Fadiga, 2010; Binder & Desai, 2011; Glenberg & Gallese, 2012). In particular, evidence confirms the existence of links between word-form circuits in perisylvian language areas and corresponding semantic information in extrasyllvian modality-preferential sensorimotor ones: action-related words (such as 'grasp' or 'kick') spark activity in lateral and superior motor and premotor cortex (Martin *et al.*, 1996; Rizzolatti & Craighero, 2004; Aziz-Zadeh *et al.*, 2006; Pulvermüller *et al.*, 2009; Kemmerer & Gonzalez-Castillo, 2010), while semantic processing of visually-related

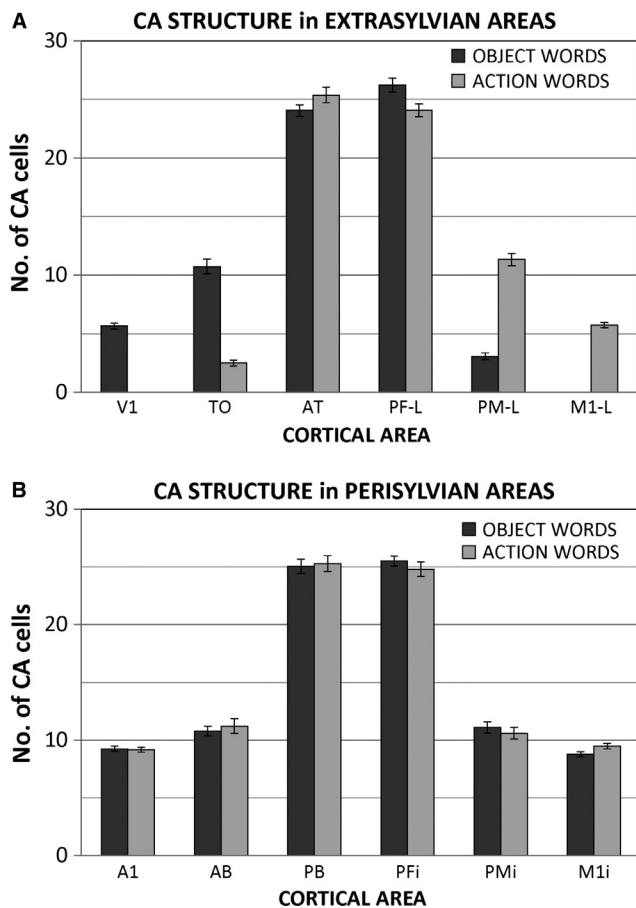


FIG. 4. Average distributions of cell assemblies (CAs) emerging in 13 instantiations of the 12-area network architecture during simulation of word learning in the semantic context of actions and visual perceptions. Bars show average numbers of CA neurons per area (or 'CA-neuron densities') for object- (dark grey) and action-related (light grey) word representations; error bars indicate standard errors over networks. (A) The extrasyllvian areas, whose cells can be seen as circuit correlates of word meaning, show a double dissociation, with relatively more strongly developed CAs for object- than for action-related words in primary and secondary visual areas (V1, TO), but stronger CAs for action- than for object-related words in dorsolateral primary motor and pre-motor cortices (PM_L, M1_L). Note the coexistence of symbolic CA circuits having comparable densities for either semantic category in the multimodal 'hub' areas AT and PF_L. (B) Data from the six perisyllvian areas, whose cells can be seen as circuit correlates of spoken word forms, do not show category-specific effects.

symbols (such as colour, object or animal words) produces activity in specific visual areas of the ventral temporo-occipital stream (Damasio *et al.*, 1996; Martin *et al.*, 1996; Pulvermüller & Hauk, 2006; Martin, 2007; Simmons *et al.*, 2007; Carota *et al.*, 2012). The evidence for both category-specific and category-general semantic areas pleads for a unifying neural model of early language acquisition, able to explain the spontaneous emergence of both in the cortex as a consequence of word learning and semantic grounding. Based on the present simulation results, such a neurobiological account is proposed below.

A new integrative model of semantic-category specificity and hubs

First some basic principles of spontaneous CA development, useful in the subsequent explanations, are introduced.

In networks that implement rich auto-associative connections between neurons along with Hebbian learning, constantly stimulated neurons have the tendency to strengthen their connections to cells they are linked to, so that, with time, larger and larger CAs develop (Doursat & Bienenstock, 2006). However, the spontaneous tendency of stimulated CAs to grow may be offset (or partly limited) by the specific features of a network, such as the density, extent and reciprocity of synaptic projections; in particular, sparse, patchy and topographic (as opposed to 'all-to-all') connectivity, as implemented here, makes CA growth harder. The presence of uniform white noise (simulating baseline neuronal firing) also has an effect on CA development: as random noise de-correlates activity between any pair of cells, its net effect is to weaken all synaptic weights in the network, thus generally counteracting CA expansion. Finally, given that CA formation is a consequence of synaptic strengthening driven by Hebbian associative learning, the 'degree' of correlation between the patterns of activity that co-occur in two or more connected areas is a critical factor for determining whether an input-specific CA circuit linking such patterns will or will not emerge in the network. (Note, in this context, that the presentation of uncorrelated, random activity patterns to the fourth, 'non-partaking' primary area during the network's training specifically hindered the growth of CA circuits into these systems; as discussed below, this was crucial for the development of category-specific circuit topographies.)

As shown by the model simulations, learning the meaning of action- or object-related words in the context of grounding motor activity or sensory input leads to the formation of input-specific lexico-semantic CA circuits in the cortex. These circuits bind the 'lexical' representation of a word – which links articulatory and acoustic-phonological activity patterns in M1_i and A1, related to spoken word form production and perception, respectively – with a perceptual or a motor schema circuit reaching into primary visual or motor areas (V1 or M1_L; Fig. 1A and B). Because of the absence of direct white-matter tracts between these modality-specific primary cortices, however, the word-related circuits emerge as widely distributed over primary (where the driving activation is present), secondary, and intermediary 'relay' areas, through which waves of correlated activity travel during learning. Hence, such multimodal convergence areas and their long-distance cortico-cortical connections play a major role in binding phonological/lexical and semantic information, with PB and PF_L being especially relevant for action-related words, and PF_i and AT for object-related ones (Fig. 1B).

Explaining category-specific effects in modality-preferential areas

The emergence of distributed CA circuits follows directly from the principles of spontaneous CA growth and the presence of correlated patterns of activity in different sets of primary sensorimotor areas. Depending on the semantic category of the word being learned, different CA circuits exhibiting different distributions across modality-preferential cortices develop. In particular, CA topographies are biased towards the motor system for action-related words grounded in motor execution, and towards the visual system for object-related words grounded in visual perception. Hence, these areas will exhibit category-specific effects during word processing across different tasks (e.g. word recognition and passive listening) because the associated semantic circuit parts are reactivated along with the spoken word-form representations. A more precise and detailed explanation follows.

Learning the meaning of object and action words may result from the presence of correlated patterns of activity in two different sets of

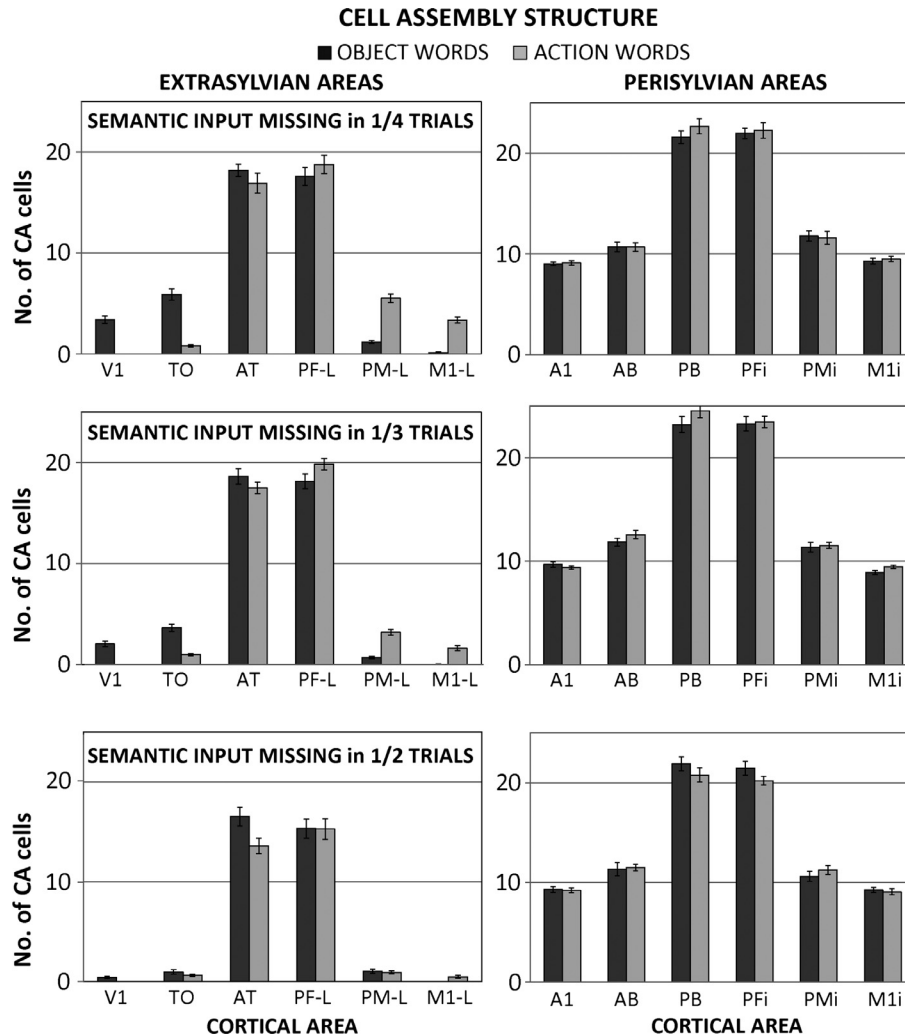


FIG. 5. Average distribution of emerging word-related cell assemblies (CAs) obtained for different amounts of semantic information provided as input during training. Left: data from extrasylvian areas. Note the gradual weakening of CA-circuits exhibited by both word categories for increasing fractions of trials failing to provide semantic input, ultimately leading (bottom row) to most word circuits not reaching the modality-specific areas. Right: data from perisylvian areas. CA distributions here are relatively unaffected by the fraction of semantic-information-bearing trials (but note that areas PF_i and PB develop smaller numbers of cells in comparison to data plotted in Fig. 4B).

three primary sensorimotor areas: V1, M1_i, A1 for object-, and M1_L, M1_i, A1 for action-related words. This leads to the emergence of strongly-connected distributed word-related CA circuits joining together the neurons consistently activated by these sensorimotor patterns; however, because of CA growth principles, the emerging circuits do not extend to areas where neural activity exhibits a low degree of correlation with such patterns, i.e. primary hand-motor area (M1_L) for object-, and primary visual cortex (V1) for action-related words (during training these areas were stimulated with a different random pattern in each learning trial; see Materials and methods). Therefore, in primary and secondary visual areas (V1, TO), densities of object-related word cells become larger than action-related ones; conversely, CA neuron densities in motor areas (M1_L, PM_L) are higher for action-related words than for object-related ones (Fig. 4B).

It should be clarified here that presenting random-noise patterns to the not-directly-stimulated modality system during training (i.e. to area V1 for action words, and to M1_L for object words) was necessary to prevent the spontaneous extension of all semantic circuits into both motor and perceptual areas. In fact, in a separate set of

simulations six network instances were trained without presenting such noisy patterns; the results showed that CAs extended further into the non-partaking, 'silent' arm of the network, eventually producing a paradoxical distribution in which action word-circuits reached also into primary visual areas, and object CAs into motor ones. This observation confirms the fundamental role of neuronal noise in preventing excessive CA growth (Doursat & Bienenstock, 2006), and suggests its relevance to semantic grounding processes.

Explaining category-general behaviour of multimodal semantic hubs

The main observation here is that CA circuits for words from different semantic categories co-exist within the same semantic hub, exhibiting comparable strength (number of cells) there. Thus, cortical hubs will show similar levels of activation during recognition/comprehension of items from any of these categories. In other words, convergence areas behave like 'multiple demand', or category-independent systems because they house CAs for symbols of all semantic types. The cortical mechanisms that, in the present architecture, lead to this result are illustrated below.

First note that the four multimodal convergence areas (AT, PF_L, PB, PF_i) are directly connected with each other (Fig. 1B). Due to the repeated concomitant stimulation of three primary cortices with correlated sensorimotor patterns, CA circuits develop in three of these four hubs areas (including PB and PF_i, plus AT for object- and PF_L for action-related words). The fourth hub, although not on the pathway connecting the three relevant primary cortices, is reciprocally linked with two other multimodal hubs and therefore receives substantial input during semantic learning. As in the presence of adequate conditions constantly stimulated CAs extend to adjacent areas (see above), the emerging symbolic circuits spontaneously grow into the fourth semantic hub. As word-related CA circuits of both semantic categories extend into both multimodal areas AT and PF_L (Fig. 4B), neurons in both will be active during semantic processing of symbols from either category (Fig. 3).

It should be noted that CA circuits are stronger (i.e. contain more CA cells) in hub areas, which ‘interface’ between the different modality systems, than in modality preferential ones (Fig. 4). This appears to be a general feature of the present type of architecture (Garagnani *et al.*, 2008; Garagnani & Pulvermüller, 2013; Pulvermüller & Garagnani, 2014), in which ‘central’ multimodal areas exhibit on average the highest numbers of synaptic links to other areas (in terms of connectivity, the highest ‘degree’; van den Heuvel & Sporns, 2013). To see this, refer to the connections depicted as arrows in Fig. 1B: primary cortices are linked with only one area, secondary ones with two, while all multimodal areas have three incoming/outgoing arrows. Cells with larger numbers of incoming/outgoing projections have a generally higher probability of being (randomly) linked to cells that happen to exhibit a correlated pattern of activity; in the presence of Hebbian learning, this implies a higher probability to become part of a CA (Garagnani *et al.*, 2009b). Furthermore, because these areas are the point of convergence of different streams of sensorimotor input, they are likely to receive more excitatory input than the modality-preferential ones, and more active cells are more likely to undergo synaptic changes [see Eqn A5 in Appendix A].

The spreading of activity from the phonological/lexical (perisylvian) part of the CA to the extrasylvian hubs, and from there to the secondary and primary areas of modality preferential systems, is taken here to be a model correlate of the cortical processes underlying semantic understanding. In this sense, the above results suggest that, while modality-preferential cortices certainly contribute to word meaning acquisition and comprehension (as they enable encoding and recollection of item-specific sensorimotor information), the majority of ‘semantic neurons’ emerge in convergence zones, where phonological and semantic word-circuit parts are bonded. Due to their role as structural-neuroanatomic connection hubs and integration points of multimodal activity, such areas end up housing word-related CAs of all different types, and hence become involved in the processing of items of all semantic categories. Thus, it is proposed here that the strong activations that multimodal hub areas often exhibit during semantic processing are the result of the presence of large numbers of neurons of all semantic circuit types there (which, in turn, follows directly from neurobiological principles and connectivity structure).

One might speculate that the convergence zones may also be the locus where information about different specific referent exemplars is integrated into a single, conceptual representation (Patterson *et al.*, 2007). However, as basic visual features of objects falling under a referential term (or movement trajectories of different action types) show surprising similarity across instantiations, a role of sec-

ondary and even primary modality-preferential areas in such integration appears possible. [Note that two synaptic steps – needed to compute more complex logical operations such as either-or relationships (Kleene, 1956; McClelland and Rumelhart, 1986) – are often necessary to categorize different exemplars/semantic features under the same concept; in the current model, such integration would only be possible in higher-order areas. However, here only one excitatory layer per area was implemented; this is a modelling simplification, as several neuronal steps are actually possible within the six cortical layers of each local neuronal cluster (Braitenberg & Schüz, 1998). With several synaptic steps in each area, either-or and similarly complex computational integration would be possible even in primary fields.] Previous simulations indeed demonstrated the ability of a similar (six-area) architecture to spontaneously ‘merge’ different overlapping CA circuits into a single one; this is because correlation learning tends to omit variable information and strengthen common features. This phenomenon, however, depended on the amount of overlap between the input patterns, as well as on the specific learning rule adopted (Garagnani *et al.*, 2007, 2009b), factors that were not the focus of the present investigation.

It should be emphasized that the ability of the model to develop strong symbolic CA circuits linking up phonological (perisylvian) and semantic (extrasylvian) circuit parts persists if the semantic pattern is absent in up to 33% of the learning trials (Fig. 5). With 50% of missing trials, the category-specific nature of the emerging circuits only persisted as a trend, which fell victim to conservative correction for repeated statistical testing. This result demonstrates not only the network’s robustness to acquire the meaning of an action or object word even when congruent visual or motor activation is missing (as it often happens in reality), but also the tolerance of the architecture to an increase in the variability of the semantic input (in the simulations, decreases in the fraction of semantic-information-bearing trials were reflected by corresponding increases in the proportion of random vs. meaningful patterns presented to the relevant primary area in association with each individual word).

As mentioned in ‘Semantic hubs vs. category specificity’, one of the main contributions of this model is to explain the emergence and topography of areas for category-specific and category-general semantic processing. If the mapping of model- to brain-areas provided in Fig. 1A is, in spite of its coarseness and simplified structure, appropriate in relevant aspects, the topography of the emerging word-circuit distributions (Fig. 4) and the corresponding activation patterns observed during simulated word recognition (Fig. 3) should match, to a degree, the patterns of brain activity observed experimentally during semantic processing. To enable a direct comparison of simulated and real brain responses, and assess the level of spatial accuracy of the mapping proposed, Fig. 6 reports the cortical areas identified by the model along with examples of phonological, category-general and category-specific semantic activations as revealed by recent functional magnetic resonance imaging studies. In particular, Fig. 6B (adapted from Saur *et al.*, 2008) shows the different brain systems activated by two different language tasks, thought to indicate phonological (top) and semantic (bottom) brain processes. Note that the areas found active in the ‘phonological’ contrast (repetition of pseudowords compared with words) are mainly perisylvian, whereas the ‘semantic’ contrast (listening to normal sentences compared with meaningless pseudo sentences) reveals some prefrontal and superior temporal activity along with dorsolateral prefrontal and anterior to middle temporal activity (PF_L, AT), also reaching into the parietal cortex. These systems exhibit a substantial degree of overlap with the perisylvian (phonological/lexical) and extrasylvian (semantic) systems of the model, respectively (see caption for

details). Figure 6C (adapted from Pulvermüller *et al.*, 2009) reports patterns of activation induced by the processing of: (1) different word types (leftmost column); and (2) words from three specific action-related semantic categories (columns 2–4). Action-related words specifically activate modality-preferential superior and lateral motor areas; note, in particular, that the major cluster activated by arm/hand words (column 3) corresponds to two model areas (PM_L , $M1_L$) where neuron densities of action-related symbolic circuits were enhanced in a category-specific manner (see Fig. 4A).

Model limitations, future extensions and predictions

Like any model, the present neural architecture makes a number of simplifying assumptions, and is therefore limited in several ways. Firstly, the connectivity realised includes just a subset of the links known to exist between the relevant cortices. In fact, the neuroanatomy of both the auditory and visual (as well as prefrontal) cortices is much more complex than that realised here (Felleman & Van Essen, 1991; Kaas & Hackett, 2000; Petrides & Pandya, 2001, 2009; Vincent *et al.*, 2007; Rauschecker & Scott, 2009; for a recent discussion on perisylvian connectivity, see also Garagnani & Pulvermüller, 2013). It is important to emphasize, however, that there is good experimental evidence for the existence of all links that the model implements (see 'Network structure and connectivity of the simulated brain areas'). The choice of deploying a network implementing only a minimal set of links can be justified on the basis of practical as well as methodological considerations: besides the need to keep simulation time within acceptable ranges, starting with a 'light' network structure is motivated by the observation that the introduction of more connections should preserve any CA circuits already emerging in the basic version, with the possible additional effect of making such representations more strongly connected and therefore more stable. This hypothesis is supported by previous simulations (Garagnani *et al.*, 2008; Pulvermüller & Garagnani, 2014). Nevertheless, while an 'Occam's razor' strategy is appropriate for a proof-of-concept study like the present one, some of the results obtained here may be the consequence of such simplification; further simulations are necessary to investigate emergence, distribution and dynamics of CA circuits in networks implementing richer connectivity and additional areas (see below).

Secondly, in real situations, learning the meaning of object and action words might involve the concurrent presence of correlated activity in motor as well as visual areas (Pulvermüller, 1999; Pulvermüller & Fadiga, 2010). For example, when acquiring the meaning of the word 'grasp' while performing grasping actions, correlated activity is likely present not only in language and motor systems but also in the ventral visual 'what' stream (e.g. if the same object is being repeatedly grasped; Ungerleider & Mishkin, 1982; Mishkin *et al.*, 1983; Ungerleider & Haxby, 1994), as well as in the dorsal parieto-occipital visual 'where' stream (Jeannerod *et al.*, 1995; Arbib, 1997; Kiefer & Spitzer, 2001), not modelled here. The main target of the present study – to differentiate and explain the spontaneous emergence of category-specific and more general semantic mechanisms in the brain – motivated a focus on the modality that provides the sensorimotor features most relevant for semantic learning. In this sense, it may be justified to focus on motor features of actions and visual features of objects: these can be seen as relatively constant, whereas the visual features of actions can be quite variable (think of the many different objects that can be grasped with a power grip), as can the action aspects of many objects. Still, some items, especially foods and tools, have both prototypical visual features and very specific motor affordances, so that

semantic learning should, in their case, include both visual and motor patterns (Warrington & McCarthy, 1987; Martin, 2007). Thus, an important direction for future extensions of the model consists in the addition of parietal areas and the implementation of sensorimotor information that can reflect both phonological and semantic features of words, symbols, actions and objects.

Although this study aimed to explain aspects of the empirically documented role of given brain areas in category-specific and -general semantic processing, the model was not designed to fit any particular set of behavioural data related to word comprehension. (Note, however, that previous simulations with a similar architecture were used to predict and explain specific brain activation patterns reflecting the processing of lexical information or the role of attention in language processing; Garagnani *et al.*, 2008, 2009a; Garagnani & Pulvermüller, 2011). In spite of this, the current neural-network model may already be capable to replicate and explain additional behavioural results, for example concerning priming effects between semantically related stimuli and symbols. There are at least two ways in which the semantic relationship between words and concepts could be implemented here: first, by overlap of sensorimotor patterns (as for the concepts 'CUP' and 'GLASS', where referents have visual features in common; Barsalou, 1999); second, by combination, i.e. co-activation of different CA circuits. Stimulation overlap leads to overlap in the CA circuits, providing a putative basis for semantic feature overlap, as assumed in semantic feature theories (Katz & Fodor, 1963). Co-activation of CAs will, in the presence of Hebbian learning, strengthen existing links between them, leading to their association; this could capture basic combinatorial semantic relations between words, as assumed by distributional semantic theories (Collins & Loftus, 1975; Landauer & Dumais, 1997).

Another simplifying assumption of the model that begs for further extensions consists of simulating one action (and, likewise, object) as one static motor (visual) activation pattern, whereas, realistically, a fine-grained structure of more or less prototypical variants might have been desirable (note, however, that the presence of random noise in all areas, overlaid to the input patterns during learning, captures, to some extent, this variability). Similarly, the acoustic-phonological and articulatory-phonological patterns presented as input to the A1 and M1_i areas were fixed, and did not mimic the natural variation in sound categories observed in real speech. Such variability could be introduced in the simulations by replacing a single input pattern with a set of partly-overlapping instances, obtained by random variation of the same prototype. As mentioned earlier, partly overlapping input patterns may lead to the emergence of a 'joint' CA circuit, merging the different CAs into a single one. Whether this will happen, however, depends not only on the degree of overlap, but also on the synaptic plasticity rule adopted, as well as other network parameters, including noise level and density and width of cortico-cortical projections (see also Garagnani *et al.*, 2008, 2009b for a discussion). Further simulation studies systematically manipulating features of the input patterns are needed to assess more thoroughly the model's ability to handle additional variability and account for key linguistic effects related to semantic processing.

It should be noted that while standard psycholinguistic models define *a priori* different layers as having different linguistic functions (phonological, lexical, semantic; Dell *et al.*, 1999), in the present simulations phonetic/phonological and semantic-referential information is co-presented to primary areas, and lexico-semantic circuits emerge as a result of learning. Therefore, the network develops internal lexico-semantic representations spontaneously, according to neurobiological principles known to govern brain function.

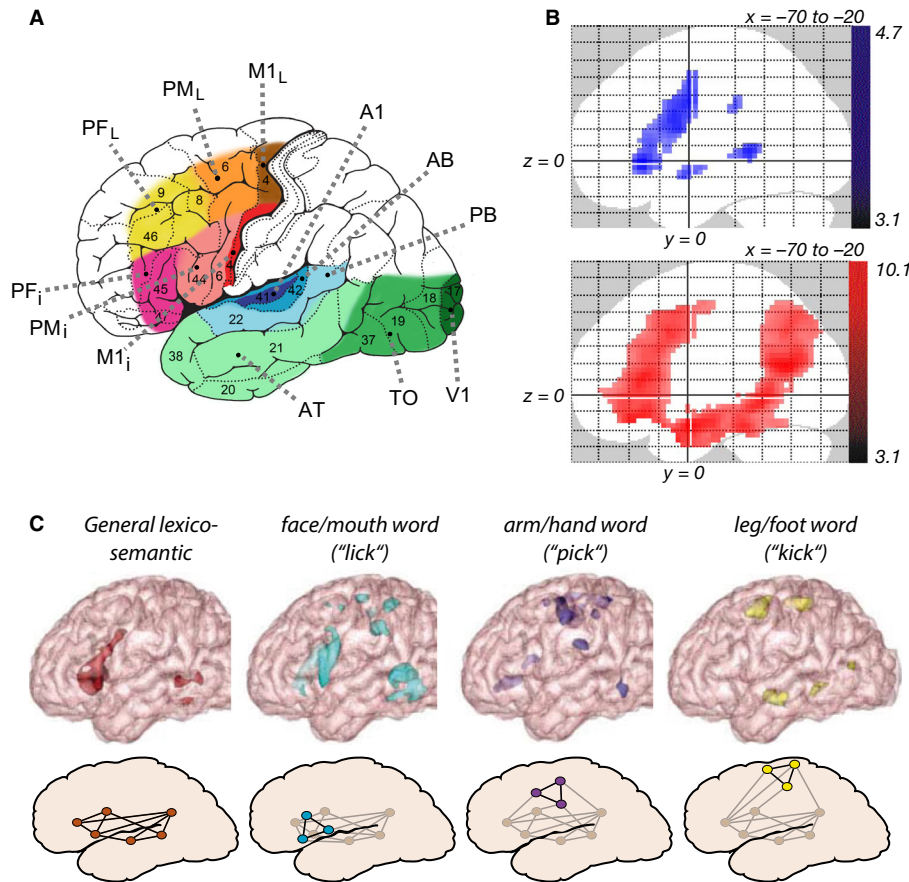


FIG. 6. Comparison between simulated areas/processes and foci of real cortical activations as observed during language processing tasks. (A) The postulated mapping of model areas onto specific cortical regions (repeated from Fig. 1A for ease of comparison). Note the ‘nesting’ of the smaller perisylvian lexical/phonological areas (A1, AB, PB, PF_i, PM_i and M1_i) within the larger extrasylvian semantic one (V1, TO, AT, PF_L, PM_L and M1_L). (B) Cortical areas activated during repetition (top – repetition of pseudowords compared with words) and semantic-comprehension (bottom – listening to normal sentences compared with meaningless pseudo sentences) tasks (adapted from Saur *et al.*, 2008, their Fig. 1, © 2008 National Academy of Sciences, USA). In the repetition task, stimuli consisted of 60 German words and 60 meaningless pseudowords. In the comprehension task, stimuli consisted of 90 well-formed German sentences [e.g. ‘der pilot fliegt das flugzeug’ (the pilot flies the aeroplane)] and 90 meaningless pseudo sentences (e.g. ‘ren simot plieft mas kugireug’). Note the strong activation of the model’s category-general semantic hubs (PF_L, AT) along with other extrasylvian areas produced by the comprehension task (red areas) but not by the repetition task (blue areas), which, instead, activates mostly perisylvian areas. Also note the approximate nesting of red areas within blue ones. (Parietal areas were not modelled in the present study.) Activations are overlaid as maximum intensity projections (x , 70–20) on a canonical brain; statistical threshold was set at $P < 0.001$, uncorrected. (C) Results of cluster analysis (from Pulvermüller *et al.*, 2009) revealing activation clusters common to different word types (leftmost column) and activations produced by different semantic word categories (other three columns). Stimuli consisted of five matched sets of 50 English words from five semantic categories: arm/hand-, face/mouth- and foot/leg-related action words, plus form- and colour-related words. Subjects were instructed to attend to all stimuli flashed on the screen and to silently read the words. The analysis contrasted activation patterns elicited by individual word categories (each tested against a control condition of matched meaningless symbol strings) with each other and with those activations shared by combinations of semantic categories. While general lexico-semantic circuits shared by different word types appear circumscribed to multimodal hub PF_i, clusters produced by action-related words extend (somatotopically) to modality-preferential areas of the model – in particular, note the precise overlap between arm/hand activation and category-specific model areas PM_L/M1_L (adapted from Pulvermüller *et al.*, 2009, with permission).

This approach is seen as explanatory and an improvement upon *a priori* defining the function of the different network’s layers. The following mapping between linguistic levels and network parts exists or emerges here: articulatory and acoustic phonetic/phonological features are implemented in areas M1_i and A1, semantic features in M1_L and V1, and lexico-semantic symbolic representations are distributed circuits spanning the entire network.

The results of the simulations enable us to make critical predictions about (and/or explain) the different extents of involvement of relevant multimodal, secondary and primary cortices during acquisition and processing of novel object- or action-related words; these predictions can be tested in (and inspire the implementation of) novel neuroimaging experiments. For example, the model results lead us to predict that none of the perisylvian areas should show significant category-specific effects (Fig. 4A), i.e. object- and action-

related word circuits should not exhibit differences in their perisylvian distribution. This is not a trivial consequence of the fact that the two word types did not exhibit systematic differences in their auditory-articulatory forms; in fact, the training process involved asymmetric stimulation of the network, with triplets of correlated patterns being presented to three of the four primary areas of the model (see ‘Simulating semantic symbol grounding’) – indeed, this asymmetry drives the resulting CA-cell distribution in the extrasylvian areas exhibited by the two semantic categories. In view of this, one might have expected the presence of asymmetries in the perisylvian distributions (as well as in extrasylvian ones). Second, the network predicts the emergence of more ‘semantic’ CA neurons in secondary (extrasylvian) areas than in primary ones (PM_L > M1_L and TO > V1). This unexpected result can be explained in terms of CA growth principles, whereby the larger numbers of CA cells that

multimodal hubs develop (for the reasons discussed earlier) lead to the recruitment of more CA cells in the nearby (directly connected) secondary areas than in the non-adjacent primary ones. Third, according to the present modelling results, we would not predict category-specific activity in semantic hubs. Category effects should therefore only emerge where hubs interface with the 'secondary' semantic areas delineated in the model. Further precise predictions about the spreading and time course of semantic activation can be made, for example for word recognition tasks (Fig. 3), and related to empirical results (Moseley *et al.*, 2013; Shtyrov *et al.*, 2014). It should be emphasized, however, that most previous experimental works showing specificity of cortical areas to semantic categories used words from natural languages, where the way these items have been learned cannot be adequately controlled for. In order to properly test the predictions resulting from the present model, word learning experiments are needed, in which neuroimaging techniques with high spatial/temporal resolution are used to reveal emergence, dynamics and distribution of CA circuits for newly-learned action- and object-related words. The prediction that action-semantic circuits reach into (and therefore their activation should spark) the premotor and primary motor cortex is in line with a number of experimental studies (Hauk *et al.*, 2004; Tettamanti *et al.*, 2005; Kemmerer *et al.*, 2012; Shtyrov *et al.*, 2014). However, evidence for the activation of the primary visual cortex during object-related word processing (Martin *et al.*, 1996; Pulvermüller *et al.*, 1999) is somewhat sparse, as most category-specific differences have been seen in more anterior temporal cortices. Thus, in future experiments testing the present simulation study's predictions it will be crucial to examine in detail visual cortex activation to specific object-related symbol categories.

We conclude on a speculative note. Recent comparative neuroimaging studies have confirmed that higher-order (especially, prefrontal, inferior parietal and temporal) association cortices have expanded disproportionately in comparison to primary areas in human brain evolution (Avants *et al.*, 2006; Van Essen & Dierker, 2007; Rilling, 2014). As observed earlier, due to the underlying network connectivity, the multimodal hub areas of the model spontaneously developed higher CA-cell density than primary (and secondary) ones. If the ability of the brain to store sets of symbol-to-meaning associations relies on the cortex's capacity to develop distinct CA-circuits that link up specific sensory and motor patterns, an increase in the size of these areas could have represented a crucial evolutionary advantage, as it would have enabled formation and storage of larger numbers of associative circuits while maintaining a low probability of cross-talk between them. The importance of the relatively larger expansion of higher-association regions in the human compared with the non-human primate brain in explaining the emergence of uniquely human linguistic and cognitive capacities has been postulated in the past (Deacon, 1997; Fuster, 1997; Preuss, 2004; Binder *et al.*, 2009; Binder & Desai, 2011). Here, a first, putative, cortical-level mechanistic explanation for this well-documented evolutionary trend is offered.

Summary and concluding remarks

Neurocognitive semantic theories propose that word meaning is grounded in the perception and action systems of the human brain (Barsalou, 2008; Pulvermüller & Fadiga, 2010; Glenberg & Gallesse, 2012; Pulvermüller, 2013). Using a novel neurocomputational model incorporating basic features of cortical anatomy and function of relevant primary, secondary sensorimotor and higher-order asso-

ciation areas in the frontal, temporal and occipital lobes, we attempt to elucidate the cortical mechanisms underlying such grounding processes and their consequences at the neurobiological representational level. In particular, the simulations show that Hebbian learning mechanisms at work within specific neuroanatomical structures are sufficient to support the formation of widely distributed lexico-semantic circuits exhibiting category-specific cortical topography and associating auditory-articulatory patterns with semantic information coming from the senses and the motor system. The model is the first computational account able to integrate key experimental observations about: (1) the presence of category-specific effects in modality-preferential sensory or motor systems (Pulvermüller & Fadiga, 2010; Meteyard *et al.*, 2012); and (2) the emergence and category-general, 'across-the-board' character of a range of semantic hubs in multimodal frontal, temporal and parietal cortices, consistently implicated in the processing of all types of meaning (Price, 2000; Patterson *et al.*, 2007; Binder & Desai, 2011).

Linking cellular-level mechanisms with system-level behaviour, this work offers a novel neurobiological account of conceptual grounding in the brain able to reconcile and explain existing data about different roles of distinct cortical areas during word comprehension processes, providing further computational evidence in support of an action-perception theory of semantic learning.

Acknowledgements

This work was supported by the UK EPSRC/BBSRC Grant EP/J004561/1 (BABEL) <http://www.tech.plym.ac.uk/SoCCE/CRNS/babel>, the Freie Universität Berlin and the Deutsche Forschungsgemeinschaft (Pu 97/15-1, 16-1). The authors would also like to thank the HPC Service of ZEDAT, Freie Universität Berlin, for computing time, and Malte Schomers, Miró Herrmann for their help at proof revision stage.

Abbreviation

CA, cell assembly.

References

- Amir, Y., Harel, M. & Malach, R. (1993) Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J. Comp. Neurol.*, **334**, 19–46.
- Arbib, M.A. (1997) Modelling visuomotor transformations. In Jeannerod, M. & Grafman, J. (Eds), *Handbook of Neuropsychology, Action and Cognition*. Elsevier Science BV, Amsterdam, pp. 65–90.
- Arikuni, T., Watanabe, K. & Kubota, K. (1988) Connections of area 8 with area 6 in the brain of the macaque monkey. *J. Comp. Neurol.*, **277**, 21–40.
- Artola, A., Bröcher, S. & Singer, W. (1990) Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, **347**, 69–72.
- Artola, A. & Singer, W. (1993) Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci.*, **16**, 480–487.
- Avants, B.B., Schoenemann, P.T. & Gee, J.C. (2006) Lagrangian frame diffeomorphic image registration: morphometric comparison of human and chimpanzee cortex. *Med. Image Anal.*, **10**, 397–412.
- Aziz-Zadeh, L., Wilson, S.M., Rizzolatti, G. & Iacoboni, M. (2006) Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.*, **16**, 1818–1823.
- Barros-Loscertales, A., Gonzalez, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J.C., Costumero, V., Parcet, M.A. & Avila, C. (2012) Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cereb. Cortex*, **22**, 2554–2563.
- Barsalou, L.W. (1999) Perceptual symbol systems. *Behav. Brain Sci.*, **22**, 577–609; discussion 610–560.
- Barsalou, L.W. (2008) Grounded cognition. *Annu. Rev. Psychol.*, **59**, 617–645.

- Binder, J.R. & Desai, R.H. (2011) The neurobiology of semantic memory. *Trends Cogn. Sci.*, **15**, 527–536.
- Binder, J.R., Desai, R.H., Graves, W.W. & Conant, L.L. (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex*, **19**, 2767–2796.
- Bookheimer, S. (2002) Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.*, **25**, 151–188.
- Braitenberg, V. (1978a) Cell assemblies in the cerebral cortex. In Heim, R. & Palm, G. (Eds), *Theoretical Approaches to Complex Systems*. Springer, Berlin, pp. 171–188.
- Braitenberg, V. (1978b) Cortical architectonics: general and areal. In Brazier, M.A.B. & Petsche, H. (Eds), *Architectonics of the Cerebral Cortex*. Raven Press, New York, pp. 443–465.
- Braitenberg, V. & Schüz, A. (1998) *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer, Berlin.
- Carota, F., Moseley, R. & Pulvermüller, F. (2012) Body-part-specific representations of semantic noun categories. *J. Cogn. Neurosci.*, **24**, 1492–1509.
- Catani, M., Jones, D.K. & Ffytche, D.H. (2005) Perisylvian language networks of the human brain. *Ann. Neurol.*, **57**, 8–16.
- Christiansen, M.H. & Chater, N. (2001) *Connectionist Psycholinguistics*. Greenwood Publishing, Westport, CT.
- Collins, A.M. & Loftus, E.F. (1975) A spreading activation theory of semantic processing. *Psychol. Rev.*, **82**, 407–428.
- Damasio, H., Grabowski, T.J., Tranel, D., Hichwa, R.D. & Damasio, A.R. (1996) A neural basis for lexical retrieval. *Nature*, **380**, 499–505.
- Deacon, T.W. (1997) *The Symbolic Species*. W.W. Norton, London, UK.
- Deco, G., Jirsa, V.K. & McIntosh, A.R. (2013a) Resting brains never rest: computational insights into potential cognitive architectures. *Trends Neurosci.*, **36**, 268–274.
- Deco, G., Rolls, E.T., Albantakis, L. & Romo, R. (2013b) Brain mechanisms for perceptual and reward-related decision-making. *Prog. Neurobiol.*, **103**, 194–213.
- Dell, G.S., Chang, F. & Griffin, Z.M. (1999) Connectionist models of language production: lexical access and grammatical encoding. *Cogn. Sci.*, **23**, 517–542.
- Devlin, J.T., Matthews, P.M. & Rushworth, M.F. (2003) Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J. Cogn. Neurosci.*, **15**, 71–84.
- Dick, A.S., Bernal, B. & Tremblay, P. (2014) The language connectome: new pathways, new concepts. *Neuroscientist*, **20**, 453–467.
- Distler, C., Boussaoud, D., Desimone, R. & Ungerleider, L.G. (1993) Cortical connections of inferior temporal area TEO in macaque monkeys. *J. Comp. Neurol.*, **334**, 125–150.
- Douglas, R.J. & Martin, K.A. (2004) Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, **27**, 419–451.
- Doursat, R. & Bienenstock, E. (2006) Neocortical self-structuration as a basis for learning. 5th International Conference on Development and Learning (ICDL 2006). Bloomington, Indiana.
- Dum, R.P. & Strick, P.L. (2002) Motor areas in the frontal lobe of the primate. *Physiol. Behav.*, **77**, 677–682.
- Dum, R.P. & Strick, P.L. (2005) Frontal lobe inputs to the digit representations of the motor areas on the lateral surface of the hemisphere. *J. Neurosci.*, **25**, 1375–1386.
- Duncan, J. (1996) Competitive brain systems in selective attention. *Int. J. Psychol.*, **31**, 3343–3343.
- Duncan, J. (2006) EPS Mid-Career Award 2004 – brain mechanisms of attention. *Q. J. Exp. Psychol.*, **59**, 2–27.
- Eggert, J. & van Hemmen, J.L. (2000) Unifying framework for neuronal assembly dynamics. *Phys. Rev. E*, **61**, 1855–1874.
- Ellis, A.W. & Young, A.W. (1988) *Human Cognitive Neuropsychology*. Lawrence Erlbaum Associates Ltd., Hove, UK.
- Elman, J.L., Bates, L., Johnson, M., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking Innateness. A Connectionist Perspective on Development*. MIT Press, Cambridge, MA.
- Felleman, D.J. & Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, **1**, 1–47.
- Finnie, P.S. & Nader, K. (2012) The role of metaplasticity mechanisms in regulating memory destabilization and reconsolidation. *Neurosci. Biobehav. R.*, **36**, 1667–1707.
- Fuster, J.M. (1997) *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. Raven Press, New York.
- Garagnani, M. & Pulvermüller, F. (2011) From sounds to words: a neuro-computational model of adaptation, inhibition and memory processes in auditory change detection. *Neuroimage*, **54**, 170–181.
- Garagnani, M. & Pulvermüller, F. (2013) Neuronal correlates of decisions to speak and act: spontaneous emergence and dynamic topographies in a computational model of frontal and temporal areas. *Brain Lang.*, **127**, 75–85.
- Garagnani, M., Wennekers, T. & Pulvermüller, F. (2007) A neuronal model of the language cortex. *Neurocomputing*, **70**, 1914–1919.
- Garagnani, M., Wennekers, T. & Pulvermüller, F. (2008) A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *Eur. J. Neurosci.*, **27**, 492–513.
- Garagnani, M., Shtyrov, Y. & Pulvermüller, F. (2009a) Effects of attention on what is known and what is not: MEG evidence for functionally discrete memory circuits. *Front. Hum. Neurosci.*, **3**, 10.
- Garagnani, M., Wennekers, T. & Pulvermüller, F. (2009b) Recruitment and consolidation of cell assemblies for words by way of Hebbian learning and competition in a multi-layer neural network. *Cogn. Comput.*, **1**, 160–176.
- Glenberg, A.M. & Gallese, V. (2012) Action-based language: a theory of language acquisition, comprehension, and production. *Cortex*, **48**, 905–922.
- Guenther, F.H., Ghosh, S.S. & Tourville, J.A. (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.*, **96**, 280–301.
- Harnad, S. (1990) The symbol grounding problem. *Physica D*, **42**, 335–346.
- Hauk, O., Johnsrude, I. & Pulvermüller, F. (2004) Somatotopic representation of action words in the motor and premotor cortex. *Neuron*, **41**, 301–307.
- Hebb, D.O. (1949) *The Organization of Behavior*. John Wiley, New York.
- Husain, F.T., Tagamets, M.A., Fromm, S.J., Braun, A.R. & Horwitz, B. (2004) Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study. *Neuroimage*, **21**, 1701–1720.
- Jeannerod, M., Arbib, M.A., Rizzolatti, G. & Sakata, H. (1995) Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci.*, **18**, 314–320.
- Kaas, J.H. (1997) Topographic maps are fundamental to sensory processing. *Brain Res. Bull.*, **44**, 107–112.
- Kaas, J.H. & Hackett, T.A. (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. USA*, **97**, 11793–11799.
- Kandel, E.R., Schwartz, J.H. & Jessell, T.M. (2000) *Principles of Neural Sciences*. McGraw-Hill, Health Professions Division, New York.
- Katz, J.J. & Fodor, J.A. (1963) The structure of a semantic theory. *Language*, **39**, 170–210.
- Kemmerer, D. (2015) Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychon. B. Rev.*, **22**, 1068–1075.
- Kemmerer, D. & Gonzalez-Castillo, J. (2010) The two-level theory of verb meaning: an approach to integrating the semantics of action with the mirror neuron system. *Brain Lang.*, **112**, 54–76.
- Kemmerer, D., Rudrauf, D., Manzel, K. & Tranel, D. (2012) Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex*, **48**, 826–848.
- Kiefer, M. & Pulvermüller, F. (2012) Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex*, **48**, 805–825.
- Kiefer, M. & Spitzer, M. (2001) The limits of a distributed account of conceptual knowledge. *Trends Cogn. Sci.*, **5**, 469–471.
- Kleene, S.C. (1956) Representation of events in nerve nets and finite automata. In Shannon, C.E. & McCarthy, J. (Eds), *Automata Studies*. Princeton University Press, Princeton, NJ, pp. 3–41.
- Lakoff, G. & Johnson, M. (1999) *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. Basic Books, New York.
- Landauer, T.K. & Dumais, S.T. (1997) A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.*, **104**, 211–240.
- Lu, M.T., Preston, J.B. & Strick, P.L. (1994) Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. *J. Comp. Neurol.*, **341**, 375–392.
- Makris, N., Meyer, J.W., Bates, J.F., Yeterian, E.H., Kennedy, D.N. & Caviness, V.S. (1999) MRI-based topographic parcellation of human cerebral

- white matter and nuclei II. Rationale and applications with systematics of cerebral connectivity. *Neuroimage*, **9**, 18–45.
- Makris, N. & Pandya, D.N. (2009) The extreme capsule in humans and rethinking of the language circuitry. *Brain Struct. Funct.*, **213**, 343–358.
- Malenka, R.C. & Bear, M.F. (2004) LTP and LTD: an embarrassment of riches. *Neuron*, **44**, 5–21.
- Martin, A. (2007) The representation of object concepts in the brain. *Annu. Rev. Psychol.*, **58**, 25–45.
- Martin, A., Wiggs, C.L., Ungerleider, L.G. & Haxby, J.V. (1996) Neural correlates of category-specific knowledge. *Nature*, **379**, 649–652.
- Matthews, G.G. (2001) *Neurobiology: Molecules, Cells and Systems*. Blackwell Science, Malden, MA.
- Mazzoni, P., Andersen, R.A. & Jordan, M.I. (1991) A more biologically plausible learning rule for neural networks. *Proc. Natl. Acad. Sci. USA*, **88**, 4433–4437.
- McClelland, J.L. & Rumelhart, D.E. & PDP-Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA.
- Meteyard, L., Cuadrado, S.R., Bahrami, B. & Vigliocco, G. (2012) Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex*, **48**, 788–804.
- Mishkin, M., Ungerleider, L.G. & Macko, K.A. (1983) Object vision and spatial vision: two cortical pathways. *Trends Neurosci.*, **6**, 414–417.
- Moseley, R.L., Pulvermüller, F. & Shtyrov, Y. (2013) Sensorimotor semantics on the spot: brain activity dissociates between conceptual categories within 150 ms. *Sci. Rep.*, **3**, 1928.
- Nakamura, H., Gattass, R., Desimone, R. & Ungerleider, L.G. (1993) The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. *J. Neurosci.*, **13**, 3681–3691.
- O'Reilly, R.C. (1998) Six principles for biologically based computational models of cortical cognition. *Trends Cogn. Sci.*, **2**, 455–462.
- Palm, G. (1982) *Neural Assemblies*. Springer, Berlin.
- Palm, G. (1990) Cell assemblies as a guideline for brain research. *Concepts Neurosci.*, **1**, 133–147.
- Pandya, D.N. (1995) Anatomy of the auditory cortex. *Rev Neurol (Paris)*, **151**, 486–494.
- Pandya, D.N. & Barnes, C.L. (1987) Architecture and connections of the Frontal Lobe. In Perecman, E. (Ed.), *The Frontal Lobes Revisited*. The IRBN Press, New York, pp. 41–72.
- Pandya, D.N. & Yeterian, E.H. (1985) Architecture and connections of cortical association areas. In Peters, A. & Jones, E.G. (Eds), *Cerebral Cortex*, vol 4. Association and Auditory Cortices. Plenum Press, London, pp. 3–61.
- Parker, G.J., Luzzi, S., Alexander, D.C., Wheeler-Kingshott, C.A., Ciccarelli, O. & Lambon Ralph, M.A. (2005) Lateralization of ventral and dorsal auditory-language pathways in the human brain. *Neuroimage*, **24**, 656–666.
- Patterson, K., Nestor, P.J. & Rogers, T.T. (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.*, **8**, 976–987.
- Petrides, M. & Pandya, D.N. (2001) Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and cortico-cortical connection patterns in the monkey. *Eur. J. Neurosci.*, **16**, 291–310.
- Petrides, M. & Pandya, D.N. (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.*, **7**, e1000170.
- Petrides, M., Tomaiuolo, F., Yeterian, E.H. & Pandya, D.N. (2012) The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains. *Cortex*, **48**, 46–57.
- Plaut, D.C. & Gonnerman, L.M. (2000) Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Lang. Cognitive Proc.*, **15**, 445–485.
- Plunkett, K. (1997) Theories of early language acquisition. *Trends Cogn. Sci.*, **1**, 146–153.
- Potter, M.C. (1979) Mundane symbolism: the relations among names, objects, and ideas. In Smith, N. & Franklin, M.B. (Eds), *Symbolic Functioning in Childhood*. Lawrence Erlbaum Associates Inc., Hillsdale, N.J.
- Preuss, T.M. (2004) What is it like to be a human? In Gazzaniga, M.S. (Ed.), *The Cognitive Neurosciences*. MIT Press, Cambridge, MA, pp. 5–22.
- Price, C.J. (2000) The anatomy of language: contributions from functional neuroimaging. *J. Anat.*, **197**(Pt 3), 335–359.
- Pulvermüller, F. (1999) Words in the brain's language. *Behav. Brain Sci.*, **22**, 253–336.
- Pulvermüller, F. (2013) How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.*, **17**, 458–470.
- Pulvermüller, F. & Fadiga, L. (2010) Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.*, **11**, 1–11.
- Pulvermüller, F. & Garagnani, M. (2014) From sensorimotor learning to memory cells in prefrontal and temporal association cortex: a neurocomputational study of disembodiment. *Cortex*, **57**, 1–21.
- Pulvermüller, F. & Hauk, O. (2006) Category-specific processing of color and form words in left fronto-temporal cortex. *Cereb. Cortex*, **16**, 1193–1201.
- Pulvermüller, F., Kherif, F., Hauk, O., Mohr, B. & Nimmo-Smith, I. (2009) Distributed cell assemblies for general lexical and category-specific semantic processing as revealed by fMRI cluster analysis. *Hum. Brain Mapp.*, **30**, 3837–3850.
- Pulvermüller, F., Luttenberger, W. & Preissl, H. (1999) Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. *Cereb. Cortex*, **9**, 498–508.
- Rauschecker, J.P. & Scott, S.K. (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.*, **12**, 718–724.
- Rauschecker, J.P. & Tian, B. (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc. Natl. Acad. Sci. USA*, **97**, 11800–11806.
- Rilling, J.K. (2014) Comparative primate neuroimaging: insights into human brain evolution. *Trends Cogn. Sci.*, **18**, 46–55.
- Rilling, J.K., Glasser, M.F., Preuss, T.M., Ma, X., Zhao, T., Hu, X. & Behrens, T.E. (2008) The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.*, **11**, 426–428.
- Rioult-Pedotti, M.S., Friedman, D. & Donoghue, J.P. (2000) Learning-induced LTP in neocortex. *Science*, **290**, 533–536.
- Rizzolatti, G. & Craighero, L. (2004) The mirror-neuron system. *Annu. Rev. Neurosci.*, **27**, 169–192.
- Rolls, E.T. & Deco, G. (2010) *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford University Press, Oxford.
- Romanski, L.M. (2007) Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cereb. Cortex*, **17** (Suppl 1), i61–i69.
- Romanski, L.M., Bates, J.F. & Goldman-Rakic, P.S. (1999a) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.*, **403**, 141–157.
- Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S. & Rauschecker, J.P. (1999b) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.*, **2**, 1131–1136.
- Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986) Learning representations by backpropagating errors. *Nature*, **323**, 533–536.
- Saur, D., Kreher, B.W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J. & Weiller, C. (2008) Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. USA*, **105**, 18035–18040.
- Schmahmann, J.D., Pandya, D.N., Wang, R., Dai, G., D'Arceuil, H.E., de Crespigny, A.J. & Wedeen, V.J. (2007) Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain*, **130**, 630–653.
- Searle, J.R. (1980) Minds, brains, and programs. *Behav. Brain Sci.*, **3**, 417–425.
- Shallice, T. (1988) *From Neuropsychology to Mental Structure*. Cambridge University Press, New York.
- Shtyrov, Y., Butorina, A., Nikolaeva, A. & Stroganova, T. (2014) Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proc. Natl. Acad. Sci. USA*, **111**, E1918–E1923.
- Simmons, W.K., Ramjee, V., Beauchamp, M.S., McRae, K., Martin, A. & Barsalou, L.W. (2007) A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, **45**, 2802–2810.
- Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S.F. & Perani, D. (2005) Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.*, **17**, 273–281.
- Tomasello, M. & Kruger, A.C. (1992) Joint attention on actions: acquiring verbs in ostensive and non-ostensive contexts. *J. Child Lang.*, **19**, 311–333.
- Ueno, T., Saito, S., Rogers, T.T. & Lambon Ralph, M.A. (2011) Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocompu-

- tational model of the dual dorsal-ventral language pathways. *Neuron*, **72**, 385–396.
- Ungerleider, L.G., Gaffan, D. & Pelak, V.S. (1989) Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Exp. Brain Res.*, **76**, 473–484.
- Ungerleider, L.G. & Haxby, J.V. (1994) ‘What’ and ‘where’ in the human brain. *Curr. Opin. Neurobiol.*, **4**, 157–165.
- Ungerleider, L.G. & Mishkin, M. (1982) Two cortical visual systems. In Ingle, D.J., Goodale, M.A. & Manfield, R.I.W. (Eds), *Analysis of Visual Behaviour*. MIT Press, Cambridge, MA, pp. 549–586.
- van den Heuvel, M.P. & Sporns, O. (2013) Network hubs in the human brain. *Trends Cogn. Sci.*, **17**, 683–696.
- Van Essen, D.C. & Dierker, D.L. (2007) Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron*, **56**, 209–225.
- Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B. & Tzourio-Mazoyer, N. (2006) Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *NeuroImage*, **30**, 1414–1432.
- Vincent, J.L., Patel, G.H., Fox, M.D., Snyder, A.Z., Baker, J.T., Van Essen, D.C., Zempel, J.M., Snyder, L.H., Corbetta, M. & Raichle, M.E. (2007) Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, **447**, 83–86.
- Vouloumanos, A. & Werker, J.F. (2009) Infants’ learning of novel words in a stochastic environment. *Dev. Psychol.*, **45**, 1611–1617.
- Warrington, E.K. & McCarthy, R.A. (1987) Categories of knowledge: further fractionations and an attempted integration. *Brain*, **110**, 1273–1296.
- Warrington, E.K. & Shallice, T. (1984) Category specific semantic impairments. *Brain*, **107**, 829–854.
- Webster, M.J., Bachevalier, J. & Ungerleider, L.G. (1994) Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb. Cortex*, **4**, 470–483.
- Wilson, H.R. & Cowan, J.D. (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, **12**, 1–24.
- Young, M.P., Scannell, J.W. & Burns, G. (1995) *The Analysis of Cortical Connectivity*. Springer, Heidelberg.
- Young, M.P., Scannell, J.W., Burns, G. & Blakemore, C. (1994) Analysis of connectivity: neural systems in the cerebral cortex. *Rev. Neurosci.*, **5**, 227–249.
- Yuille, A.L. & Geiger, D. (2003) Winner-take-all mechanisms. In Arbib, M. (Ed.), *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA, pp. 1056–1060.

Appendix A: Full model specification

Each of the 12 simulated areas (Fig. 1B) was implemented as two layers of artificial neuron-like elements (‘cells’), 625 excitatory and 625 inhibitory, thus resulting in 15 000 cells in total. Each excitatory cell ‘*e*’ can be considered the network equivalent of a local cluster, or column, of approximately 25 000 real excitatory cortical neurons, that is pyramidal cells, while its twin inhibitory cell ‘*i*’ (Fig. 1C) models the cluster of inhibitory interneurons situated within the same cortical column (Wilson & Cowan, 1972; Eggert & van Hemmen, 2000). The activity state of a cell *e* is uniquely defined by its membrane potential $V(e, t)$, representing the average of all the postsynaptic potentials within neural pool (cluster) *e* at time *t*, and governed by the following equation:

$$\tau \cdot \frac{dV(e, t)}{dt} = -V(e, t) + k_1(V_{\text{In}}(e, t) + k_2\eta(e, t)) \quad (\text{A1})$$

where $V_{\text{In}}(e, t)$ is the net input to cell *e* at time *t* (sum of all inhibitory and excitatory postsynaptic potentials acting upon cluster *e* – I/EPSPs; inhibitory synapses are given a negative sign – plus a constant baseline value V_b), τ is the membrane’s time constant, k_1 , k_2 are scaling constants and $\eta(e, t)$ is a white noise process with uniform distribution over $[-0.5, 0.5]$. Note that noise is an inherent property of each model cell, intended to mimic the spontaneous activity (baseline firing) of real neurons. Therefore, noise was constantly present in all areas, in equal amounts.¹

¹Inhibitory cells have $k_2 = 0$ (i.e. the noise is generated just by the excitatory cells).

The output (transformation function) of an excitatory cell *e* at time *t* is defined as:

$$O(e, t) = \begin{cases} 0 & \text{if } V(e, t) \leq \varphi \\ (V(e, t) - \varphi) & \text{if } 0 < (V(e, t) - \varphi) \leq 1 \\ 1 & \text{otherwise} \end{cases} \quad (\text{A2})$$

$O(e, t)$ represents the average (graded) firing rate (number of action potentials per time unit) of cluster *e* at time *t*; it is a piecewise-linear sigmoid function of the cell’s membrane potential $V(e, t)$, clipped into the range $[0, 1]$ and with slope 1 between the lower and upper thresholds φ and $\varphi + 1$. The output $O(i, t)$ of any inhibitory cell *i* is 0 if $V(i, t) < 0$, and $V(i, t)$ otherwise. In excitatory cells, the value of the threshold φ in Eqn A2 varies in time, tracking the recent mean activity of the cell so as to implement neuronal adaptation (Kandel *et al.*, 2000). Thus, stronger activity leads to a higher threshold in subsequent time steps. More precisely,

$$\varphi(e, t) = \alpha \cdot \omega(e, t) \quad (\text{A3})$$

where $\omega(e, t)$ is the time-average of cell *e*’s recent output and α is the ‘adaptation strength’. For an excitatory cell *e*, the approximate time-average $\omega(e, t)$ of its output $O(e, t)$ is estimated by integrating the linear differential equation Eqn A4.1 below with time constant τ_A , assuming initial average $\omega(e, 0) = 0$:

$$\tau_A \cdot \frac{d\omega(e, t)}{dt} = -\omega(e, t) + O(e, t) \quad (\text{A4.1})$$

Local (lateral) inhibitory connections (Fig. 1C) and area-specific inhibition are also implemented, realising, respectively, local and global competition mechanisms (Duncan, 1996, 2006), and preventing activation from falling into non-physiological states (Braitenberg & Schüz, 1998). More formally, in Eqn A1 the input $V_{\text{In}}(e, t)$ to all excitatory cells of the same area includes an area-specific (‘global’) inhibition term $k_S \cdot \omega_S(e, t)$, subtracted from the total sum of the I/EPSPs postsynaptic potentials V_{In} in input to the cell, with $\omega_S(e, t)$ defined by:

$$\tau_S \cdot \frac{d\omega_S(e, t)}{dt} = -\omega_S(e, t) + \sum_{e \in \text{area}} O(e, t) \quad (\text{A4.2})$$

The low-pass dynamics of the cells [Eqns A1, A2, A4.1–2] are integrated using the Euler scheme with step size Δt , where $\Delta t = 0.5$ ms.

Excitatory links within and between (possibly non-adjacent) model areas are established at random and limited to a local (topographic) neighbourhood; weights are initialized at random, in the range $[0, 0.1]$. The probability of a synapse to be created between any two cells falls off with their distance (Braitenberg & Schüz, 1998) according to a Gaussian function clipped to 0 outside the chosen neighbourhood (a square of size $n = 19$ for excitatory and $n = 5$ for inhibitory cell projections). This produces a sparse, patchy and topographic connectivity, as typically found in the mammalian cortex (Amir *et al.*, 1993; Kaas, 1997; Braitenberg & Schüz, 1998; Douglas & Martin, 2004).

The Hebbian learning mechanism implemented simulates well-documented synaptic plasticity phenomena of long-term potentiation (LTP) and depression (LTD), as implemented by Artola, Bröcher and Singer (Artola *et al.*, 1990; Artola & Singer, 1993). This rule, which covers both ‘true’ Hebbian co-occurrence (‘what fires together wires together’) as well as ‘anti-Hebb’ (‘neurons

out of sync delink') plasticity, provides a realistic approximation of known experience-dependent neuronal plasticity and learning (Rioult-Pedotti *et al.*, 2000; Malenka & Bear, 2004; Finnie & Nader, 2012). In the model, the continuous range of possible synaptic efficacy changes was discretized into two possible levels, $+\Delta w$ and $-\Delta w$ (with $\Delta w \ll 1$ and fixed). Following Artola *et al.*, we defined as 'active' any link from an excitatory cell x such that the output $O(x, t)$ of cell x at time t is larger than θ_{pre} , where $\theta_{\text{pre}} \in [0, 1]$ is an arbitrary threshold representing the minimum level of presynaptic activity required for LTP to occur. Thus, given any two cells x and y connected by a synaptic link with weight $w_t(x, y)$, the new weight $w_{t+1}(x, y)$ is calculated as follows:

$$w_{t+1}(x, y) = \begin{cases} w_t(x, y) + \Delta w & \text{if } O(x, t) > \theta_{\text{pre}} \text{ and } V(y, t) > \theta_{\text{post}} \text{ (LTP)} \\ w_t(x, y) - \Delta w & \text{if } O(x, t) \leq \theta_{\text{pre}} \text{ and } V(y, t) > \theta_{\text{post}} \text{ (LTD)} \\ w_t(x, y) & \text{otherwise} \end{cases} \quad (\text{A5})$$

Typical parameter values used during the simulations are as follows:

Eqn A1	Time constant (excitatory cells):	$\tau = 2.5$ (simulation time-steps)
	Time constant (inhibitory cells):	$\tau = 5$ (simulation time-steps)
	Scaling factor:	$k_1 = 0.01$
	Baseline potential	$V_b = 0$
	Noise scaling factor	$k_2 = 25 \cdot \sqrt{48}$
	Global inhibition strength	$k_S = 65$
	(during training):	$k_S = 95$
Eqn A3	Adaptation strength:	$\alpha = 0.01$
Eqns A4.1–2	Average output time constant (for adaptation mechanism):	$\tau_A = 10$ (simulation time-steps)
Eqn A5	Global inhibition time constant:	$\tau_S = 12$ (simulation time-steps)
	Postsynaptic potential threshold required for synaptic change:	$\theta_{\text{post}} = 0.15$
	Presynaptic output activity required for LTP:	$\theta_{\text{pre}} = 0.05$
	Learning rate:	$\Delta w = 0.0008$