Chapter 6

Grounding language in the brain

Friedemann Pulvermüller

6.1 Introduction

Can a cognitive theory of embodiment be an abstract theory? Clearly, any theory is abstract in a relevant sense, as it is defined as a theoretical construct. But can a theory of embodiment remain at an entirely abstract level, or would it necessarily need to connect to the concrete mechanistic level of the brain, to nerve cells and circuits? Symbols and meaning are not grounded directly in experiences and actions – or only in a metaphorical sense. Rather, signs and symbols are mechanically based on, or grounded in, neuronal circuits in the brain. Importantly, as we have learned recently, some of these neuronal 'symbolic' circuits seem to play a relevant role in action and perception, too; in this sense, symbols are action-perception grounded. Therefore, the targets of a theory of embodiment include neuronal elements that realize thought, meaning, and language processes. They are not abstract entities but real objects following laws of nature and neuroscientific principles.

Here, a theory of the neuronal embodiment of language and conceptual processes is reviewed, which is built upon neuroscientific principles. *Action–perception networks* (APNs) in the brain are proposed to be the grounding machines realizing the binding of language, perceptual, and action-related information. There is evidence for this proposal from cognitive neuroscience research, and a review of this evidence will be given. It will also be mentioned that, in some cases, it was possible to address specific questions about embodied cognition using neuroscience experiments. The neuromechanics of embodiment and empirical neuroscience research targeting these mechanisms may therefore be welcome additions to cognitive theory.

6.2 Embodiment and the brain

A main idea immanent to cognitive theories of conceptual embodiment is that concepts and the meaning of symbols that function as vehicles of thought are embodied in perceptual and action information (Barsalou 1999; Clark 1996; Lakoff 1987; Lakoff and Johnson 1999; Varela *et al.* 1991). Meanings and concepts, be they abstract or concrete, can be explained, and therefore grounded (at least in part) by referring to concrete sensory information and to information about actions carried out by the individual. Although most cognitive theories of embodiment are at the – in one sense still abstract – level of cognitive representations and processes, it is clear that all cognitive entities are mechanistically based on brain circuits and their activation, as selective

deficits after brain lesions demonstrate. From an embodiment perspective, it may be advantageous to look at brain correlates of conceptual and semantic processing, because the mechanistic processes and principles that emerge from the neuronal substrate could potentially contribute to experimental testing and explanation at the cognitive level.

The idea of embodiment goes back quite far in history, certainly to Aristotle, to Plato's cave allegory, and probably even further. A significant step was taken by Sigmund Freud who, when working on the neurology of organic language disturbances in his early career, drew the first diagrams of hypothetical networks of cortical neurons that might process symbols in the human brain (Freud 1891) (Figure 6.1). In one of these diagrams, he connected the neuronal representation of a spoken word, via reciprocal links, to an 'object association' network, which, as we might say today, would embody and cortically organize aspects of the meaning of the symbol. Freud's speculation about the embodiment of word meaning was rejected by neurologists of the 19th Century, but appears surprisingly modern today, as, on the basis of a multitude of neuroimaging results, distributed cortical networks are considered to be a likely basis of higher cognitive processes. A critical point might be that Freud was one of the first to propose an embodiment theory at the cognitive level with an anatomical and physiological basis in the brain. We may ask still today whether a cognitive theory of embodiment can be strengthened by supplying it with a brain basis, by spelling out its mechanisms in terms of neurons. Few would probably deny this, and important steps have been taken towards this goal (Gallese and Lakoff 2005). This chapter is one more contribution, focusing specifically on questions of embodiment in the context of a neuronal model of words, meanings and concepts (Pulvermüller 1992, 1999, 2001, 2005).

One might want to say that a theory of embodiment has to be embodied in the brain, but such a statement inevitably leads to confusion due to the different uses of the critical morpheme. From now on in this chapter, the usage of the word 'embody', and similarly of 'grounding', will be restricted to a relationship between concepts, meanings, signs, symbols, words, and their semantically related actions and perceptions. In contrast, these cognitive elements will be said to be realized, organized, implemented, wired, or laid down in the neural substrate in such and such a way.

Apart from supplying a cognitive theory with a brain basis, evidence from neuroscience may also help in answering burning cognitive questions. This stronger statement

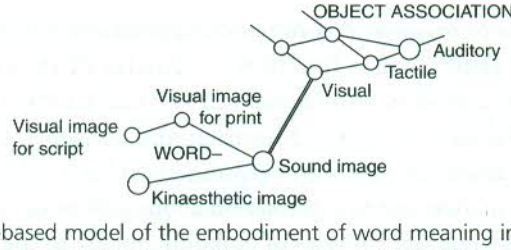

Fig. 6.1 Freud's brain-based model of the embodiment of word meaning in perceptual information. Multimodal 'object associations' were thought to be linked to the word form representation. This model was thought to depict connections and processing units at the level of the brain (Freud, 1891).

will also be addressed here: I will submit that cognitive neuroscience techniques have enabled us to provide strong evidence for an embodiment approach not previously available from behavioural experiments.

This chapter will first ask basic questions about how words are realized as neuron circuits that process actions and perceptions. Subsequently, the question of brain mechanisms of meaning grounded in action and perception will be addressed. Specific questions about phonological and semantic embodiment will be highlighted separately, in the light of recent findings from neuroimaging. Here, the case of category-specific semantic networks will be featured in detail as it turns out to be of particular theoretical importance. Critical questions about category-specific embodied semantic processing in the brain have been addressed in detail by research into the brain processes underlying action-related language. Multimodal neuroimaging work in this area will be reviewed, the emerging conclusion being that semantic processing in the brain may be realized, at least in part, by rapid, automatic, functionally relevant spreading of activation into sensory and motor areas, which reflects aspects of the reference of action and object words. From here, the scope of the chapter will widen, addressing fundamental issues in the brain embodiment of meaning and concepts, the role of correlations between words and world and among words, the discreteness of language representations, and the issue of abstract concepts. The proposal will be that a mere mapping of sensory and motor features is not enough for embodied meaning representation, but that logical operations are necessary as additional pre-wired ingredients.

6.3 Basic questions about language, thought, and the brain

How are words and their related meaningful concepts stored and processed in the brain? We may want to start with a simpler question, that of how objects and actions are processed in brain circuits, and especially in circuits in the brain structure most important for cognitive processes i.e., the cortex. It is well known that many neurons in cortex respond to elementary features of objects, for example form and colour features of visual images of the objects or acoustic length and frequency information of sounds characteristic of the object (Hubel 1995). In the motor domain, cortical neurons usually signal elementary features of muscle movements (Rizzolatti and Luppino, 2001), or, at higher levels, even elementary information about the goals connected with an action (Fogassi *et al.* 2005). As neurons appear to represent and process features of objects, the assumption that collections of neurons (*neuronal assemblies*), are the basis for the processing of sets of features characterizing objects and actions (Braitenberg 1978; Hebb 1949). Sets of neurons with strong links between their members would provide the ideal machinery for maintaining the memory of an object or action, and have therefore been proposed to underlie short-term or active memory (Fuster, 1995). In recent years, the cognitive neuronal equations 'feature = neuron' and 'object/action = neuronal assembly' have received support from a range of lines of neuroscientific research in both animals and humans (Braitenberg and Schüz 1998; Fuster 2003; Plenz and Thiagarajan 2007; Pulvermüller 2005; Singer and Gray 1995).

To bind individual neurons, which signal elementary and possibly more complex features of an object or action into a cell assembly, neural hardware and functional principles that drive the binding are necessary. From the hardware perspective, it must be noted that connections in cortex do not only run between adjacent neurons in the same area and local cortical patch, but also between distant sites. Plate 6.1 illustrates such long-distance connections, taking the example of links between different parts of the perisylvian language cortex (in green and blue) and the motor system (in red). Long-distance connections are indicated by arrows and, in addition, adjacent areas are also connected with each other. These links are, in part, evident from anatomical studies in humans (Brodmann 1909; Makris *et al.* 1999) or can at least be suggested on the basis of studies in monkeys (Pandya and Yeterian 1985; Young *et al.* 1994), taking into account the homology of cortical areas between species. As a multitude of long-distance links are available in cortex, information can be transmitted between distant cortical areas, for example, between superior temporal lobe (in blue), where acoustic information about speech sounds arrives, and inferior frontal cortex (in green), where the motor output is programmed and coordinated.

Apart from the anatomical principle that long-distance links connect different lobes and a range of distant cortical areas with each other, a functional principle is of utmost relevance: Nerve cells that are connected to each other and fire together frequently strengthen their connection (Artola and Singer 1993; Hebb 1949; Tsumoto 1992). In reverse, neurons that fire independently from each other, or in an antiphasic manner, tend to weaken their links, and even fine-grained sequential features of activation patterns can modify synaptic weights (Bi 2002). This means that the correlation of neuronal firing of cortical cells is translated into connection strengths. In essence, what fires together wires together, and the long-distance wiring in cortex guarantees that correlation-related links can even develop within sets of neurons distributed over distant areas and lobes. This has important implications for the way in which objects and actions are processed and stored in brain circuits; if an object is characterized by multimodal features, for example, shape, colour, and sound, which in many cases occur at the same time, it is plausible that correlated activity in different cortical lobes, in the visual and auditory systems, strengthens the connections between distant neurons processing these different information types. The thereby established distributed neuronal assembly would then bind information across sensory and motor modalities.

The neuroscientific principles of correlation learning and that of long-distance cortico-cortical connectivity have important implications for the brain basis of meaningful language units, words, and morphemes (Pulvermüller, 1999). When a word form is being articulated, this relates to neuronal activation in the motor cortex. Motor activation is, in turn, coordinated and controlled by premotor circuits, which are, in turn, linked to and influenced by activity in inferior prefrontal areas. In addition to activity in the inferior frontocentral cortex (violet areas), the speech produced leads to auditory input, which activates superior temporal auditory cortex, and, via short distance connections, the adjacent auditory belt and parabelt areas in superior temporal gyrus and sulcus (blue areas). As there are neuronal links between superior temporal lobe and inferior prefrontal areas,

the co-activation of neurons in these areas, which is characteristic of the production of a spoken word, can lead to synaptic strengthening. A word-related cell assembly distributed over different parts of this perisylvian cortex (violet and blue areas) develops (Pulvermüller 1999; Pulvermüller and Preissl, 1991). As the inferior frontal and superior temporal neuron populations – which, at the start, had either been responsible for controlling the articulation movement or for specifically responding to the sounds characteristic of the word – the connected assembly can be considered an APN in which action-related and perceptual information is being bound together. This APN would represent and process a specific spoken word form, and therefore embody and ground it in its distinctive articulatory and acoustic features.

6.4 Action–perception networks

6.4.1 Cortical embodiment of words as action–perception networks

The speculation that spoken word forms such as 'crocodile' are grounded in action-perception networks, whereas meaningless but pronounceable and phonotactically legal pseudowords that are not being used in the language, such as 'crodobile', are not, might lead to fruitful research. One idea is that activation of a memory network leads to well coordinated reverberatory activity and synchronous oscillations at high frequencies in the so-called gamma band (>20 Hz) (von der Malsburg and Schneider 1986). Evidence for this comes from animal research (Singer and Gray 1995) and can also be found in noninvasive recordings, electroencephalography (EEG), and magnetoencephalography (MEG) (Lutzenberger *et al.* 1995; Tallon-Baudry and Bertrand 1999). These high frequencies may, in part, relate to the fact that most neurons in cortex conduct activity rather fast (5–10 metres per second; for discussion, see Pulvermüller 2000) so that reverberations in cortical neuron loops may take <50 milliseconds.

When investigating high-frequency cortical responses to words and meaningless word-like items, induced gamma-band responses were found to be enlarged for the lexical items and relatively small for the meaningless novel ones (Eulitz *et al.* 2000; Krause *et al.* 1998; Lutzenberger *et al.* 1994; Pulvermüller *et al.* 1995, 1996, 1997). This was true for different languages (e.g., English, German, Finnish), in both major language modalities (spoken and written language), and in a range of tasks and paradigms (lexical decision, reading, listening, and active and passive oddball tasks). The difference in gamma-band responses was usually most pronounced over, or in, the left language-dominant hemisphere. The frequency where differences were most pronounced ranged between 20 and 60 Hz. Similar effects to those reported earlier for words and pseudowords were also seen for phonemes of language versus non-language sounds and for familiar versus unfamiliar letters (Ihara and Kakigi 2006; Palva *et al.* 2002). Familiar objects and coherent visual patterns have been shown to elicit enhanced gamma-band activity in the human brain In the same way as familiar meaningful language elements (Gruber *et al.* 2006; Lutzenberger *et al.* 1995; Müller *et al.* 1996; Tallon-Baudry *et al.* 1996, 1998). These results indicate the existence of memory networks in the human brain generating coordinated

high-frequency responses. Such circuits seem to develop for meaningful elements that have been learned, including words and objects (Pulvermüller *et al.*, 1997).

Different tests of the neuronal assembly model of word processing can be performed using other measures of cortical activity. An obvious prediction is that a memory network in cortex should act like an amplifier of cortical activity, so that input activating a memory network leads to a stronger brain response (input plus neuronal assembly activation) than an input that fails to activate such a network (response to sensory input only). A well known indicator of cognitive processes is the *mismatch negativity* (MMN) elicited by auditory stimuli. The MMN is larger to familiar sounds of one's own language than to phonemes of a foreign language (Näätänen *et al.* 1997). In the same way, familiar non-language sounds, such as clicks or whistles, elicit a larger MMN compared with physically matched unfamiliar sounds (Frangos *et al.* 2005; Hauk *et al.* 2006). Crucially, if a syllable or language sound is placed in a context where it is critical for understanding a meaningful word, its MMN is enhanced compared with a condition in which the same stimulus completes a meaningless but pronounceable pseudoword (Endrass *et al.* 2004; Korpilahti *et al.* 2001; Kujala *et al.* 2002; Pettigrew *et al.* 2004; Pulvermüller *et al.* 2001, 2004; Shtyrov and Pulvermüller 2002; Sittiprapaporn *et al.* 2003). This lexical enhancement of the MMN is best explained by the full activation (ignition; Braitenberg 1978) of a cell assembly triggered by a meaningful word, but not by an unfamiliar meaningless item. A similar explanation in terms of memory networks for phonemes and other familiar sounds has also been established (Näätänen 2001).

Although these results support cell assembly activation following word presentation and the lack thereof when pseudowords are being presented, no direct evidence has so far been discussed for APNs linking inferior frontal output control circuits and superior temporal comprehension processors by means of distributed neuronal systems. Imaging studies have directly addressed the question of whether the left hemispheric inferior frontal and superior temporal language areas are modules specialized in either speech perception or production, or rather represent two local areas that house neural elements participating in interactive distributed cortical processes that contribute to both production and comprehension. During listening to syllables and words, the left inferior frontal and premotor cortex is active, along with the superior temporal areas in the vicinity of the auditory cortex (Pulvermüller *et al.* 2003; Wilson *et al.* 2004; Zatorre *et al.* 1992) (Figure 6.2). During speaking, the superior temporal cortex was active, along with areas in inferior motor, premotor, and prefrontal cortex, although it was ensured that self-produced sounds could not be perceived through the auditory channel (Paus *et al.* 1996; Watkins and Paus 2004). In addition, it is well known that lesions in superior temporal or inferior frontal cortex that lead to aphasia usually impair both speech production and comprehension (Pulvermüller and Preissl 1991). This indicates that interactive neural systems distributed over the inferior frontal and superior temporal cortex contribute to both speech production and perception. During spoken word recognition and understanding, these systems become active near-simultaneously and largely in parallel, with a peak activation delay in the inferior frontal cortex of ~20 milliseconds after peak activation in superior temporal areas (Pulvermüller *et al.* 2003). These results suggest tight and rapid

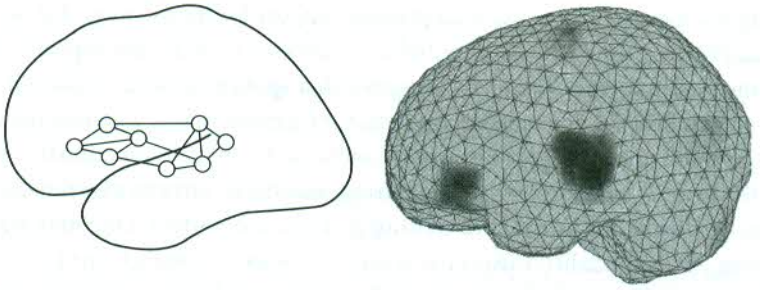

Fig. 6.2 Theory and data on perisylvian cell assemblies for words. The diagram on the left illustrates a cell assembly distributed over perisylvian areas, the kind of cortical network thought to represent and process spoken word forms at the cortical level (Pulvermüller and Preissl 1991). The perisylvian assembly can be said to embody the word as an action–perception network. The diagram on the right shows objective source estimates for the mismatch negativity brain response elicited by a spoken word, as they were recorded 130–150 milliseconds after the recognition point of the word. Two main sources in left perisylvian cortex could be distinguished: the posterior source lay in superior temporal areas close to auditory cortex and the anterior source in inferior frontal cortex anterior to the motor representation of the articulators. Activation peaks of these sources followed each other with a minimal delay of ~20 milliseconds (Pulvermüller *et al.* 2003).

functional links between speech perception and speech production processes, as postulated by neurobiological (Braitenberg and Schüz 1992; Fry 1966; Pulvermüller and Preissl, 1991) and psycholinguistic theories (Fowler 1986; Liberman *et al.* 1967).

That the links between superior temporal perceptual circuits and frontocentral speech production machinery are functionally effective has been demonstrated by experiments using transcranial magnetic stimulation (TMS). When spoken words and language sounds (phonemes) that strongly involve the tongue are being perceived, TMS applied to the inferior motor cortex elicits stronger muscle responses of the articulators compared with control conditions (Fadiga *et al.* 2002). Interestingly, this effect was most prominent when the critical phonemes were presented in meaningful word context, suggesting that cell assemblies for meaningful words play a role in linking articulatory gestures and auditory signals at the cortical level (see Pulvermüller and Preissl 1991). Converging evidence from functional connectivity studies on the basis of positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) data indicates that the links between superior temporal and inferior frontal language areas depend on the amount of meaningful information being transmitted by words (Horwitz and Braun 2004).

The documented tight functional links between action and perception circuits of the left perisylvian language array (Pulvermüller 1999; Rizzolatti and Craighero 2004) cannot be explained in a straightforward manner within a modular approach according to which speech production and comprehension are situated in functionally separate encapsulated modules (e.g., Ellis and Young 1988). However, they meet the predictions of a distributed model postulating APNs and binding between specific acoustic speech patterns and the articulatory gestures that generate them (Plate 6.2A).

In summary, these neuroscience data provide support for the position that word forms are cortically based on APNs distributed over inferior frontal and superior temporal areas. At the purely cognitive level, this suggests that spoken words as entries of a 'mental lexicon' are embodied as complex articulatory gestures and as complex auditory spectrotemporal patterns, and, most importantly, as the specific functional connections between the two. Considering written language, additional links to visual pattern representations, written word forms, and writing gestures are evident. One may want to call the resulting APNs modality-unspecific representations of abstract word forms, but, as concrete and specific articulatory, gestural, acoustic, and visual features characterize word forms, it might also be accurate to speak of multi- or cross-modal representations.

6.4.2 Functional specificity of action–perception networks in phonological processing

The finding of strong functional links between superior temporal speech perception and comprehension circuits and inferior frontal action control circuits still leaves open the question of universality and functional specificity of APNs. When a speech sound is heard, its motor program might be automatically activated and the articulation action simulated mentally, in a similar way as has been proposed and observed for other action types (Buccino *et al.* 2001; Jeannerod 2001; Jeannerod *et al.* 1995). What has been shown so far, however, is merely an activation of frontal and motor/premotor cortex when speech information comes in. To demonstrate the specificity of APNs in speech perception, it would be necessary to demonstrate that activations in the frontal action system reflect sensory information. In the case of spoken language, it becomes important to ask whether articulatory features of speech sounds are reflected in the brain activation pattern in frontocentral cortex, and, if yes, whether these action-related activations would also occur during comprehension of specific speech sounds. Would the motor system activation reflect articulatory information about incoming speech?

To probe the possible involvement of specific motor circuits in the speech perception process, we used event-related fMRI and presented experimental subjects with spoken syllables including *[p]* and *[t]* sounds, which are respectively produced by movements of the lips or tongue. Physically similar nonlinguistic signal-correlated noise patterns were used as control stimuli. In localizer experiments, subjects had to silently articulate the same syllables, and, in a second task, move their lips or tongue. Speech perception most strongly activated superior temporal cortex. Crucially, distinct motor regions in the precentral gyrus sparked during articulatory movements of the lips and tongue, and also during nonlinguistic lip and tongue movements, were also differentially activated in a somatotopic manner when subjects listened to the lip- or tongue-related phonemes (Pulvermüller *et al.* 2006) (Plate 6.2).

These results indicate that, during speech perception, motor circuits are recruited that reflect phonetic distinctive features of the speech sounds encountered, thus providing direct neuroimaging support for specific links between the phonological mechanisms for speech perception and production. The conclusion might therefore be that APNs are

specific to features of articulatory gestures and speech sounds. As could be made evident at the level of meaningful words, APNs would ground speech sounds in articulatory gestures and vice versa. The complex action–perception mapping appears crucial for the 'abstractness' – or cross-modality – of the representations.

6.4.3 Storing semantic information

Semantic category specificity

Where is word meaning represented and processed in the human brain? This question has been discussed controversially since 19th Century neurologists postulated a 'concept centre' in the brain that was thought to store the meanings of words (Lichtheim 1885). Today, the cortical loci proposed for a centre uniquely devoted to semantic binding between words and their meaning range from inferior frontal cortex (Bookheimer 2002; Posner and Pavese, 1998) to anterior, inferior, superior, or posterior left temporal cortex (Hickok and Poeppel 2004; Patterson and Hodges 2001; Price 2000; Scott and Johnsrude 2003). Others have proposed that the entire left frontotemporal cortex is a region equally devoted to semantics (Tyler and Moss 2001), or that the parahippocampal gyrus (Tyler *et al.* 2004) or the occipital cortex (Skrandies 1999) are particularly relevant. As there is hardly any area in the left language-dominant hemisphere for which there is no statement that it should house the semantic binding centre, these views are difficult to reconcile with each other (see Pulvermüller 1999). Is there a way to resolve this unfortunate diversity of opinions?

A way out might be pointed to by approaches to category-specific semantic processes (Daniele et al. 1994; Humphreys and Forde 2001; Warrington and McCarthy 1983; Warrington and Shallice 1984). The idea here is that different kinds of concepts, and different kinds of word meaning, draw upon different parts of the brain. Hearing the word 'crocodile' frequently together with certain visual perceptions may lead to strengthening of connections between the activated visual and language-related neurons. Specific form and colour detectors in primary cortex, as well as neurons responding to more complex features of the perceived gestalt higher up in the inferior temporal stream of visual object processing, will become active together with neurons in the perisylvian language areas that process the word form. These neurons would bind into distributed networks now implementing word forms together with aspects of their referential semantics. In contrast, learning of an action word, such as 'ambulate', critically involves linking an action type to a word form. In many cases, action words are learned in infancy when the child performs an action and the caretaker uses a sentence including an action word describing the action (Tomasello and Kruger 1992). As the brain circuits for controlling actions are in motor, premotor, and prefrontal cortex, it is clear that in this case correlated activation should bind perisylvian language networks to frontocentral circuits processing actions.

The cell assembly model and other theories of perception and action-related category specificity predict differential distribution of the neuron populations organizing action- and object-related words and similar differences can be postulated for other

semantic categories (Pulvermüller 1996, 1999) (Figure 6.3). Many nouns refer to visually perceivable objects and are therefore characterized by strong semantic links to visual information, whereas most verbs are action verbs and link semantically to action knowledge. Like action verbs, nouns that refer to tools are usually also rated by subjects to be semantically linked to actions, and a large number of animal names are rated to be primarily related to visual information (Preissl *et al.* 1995; Pulvermüller, Lutzenberger *et al.* 1999; Pulvermüller, Mohr *et al.* 1999). Range of neuroimaging studies using EEG, MEG, PET, and fMRI techniques found evidence for category-specific activation in the human brain for the processing of action- and visually-related words and concepts (e.g., Cappa *et al.* 1998; Chao *et al.* 1999; Kiefer 2001; Preissl *et al.* 1995; Pulvermüller, Lutzenberger *et al.* 1999; Pulvermüller, Mohr *et al.* 1999). The results were largely consistent with the model of semantic category-specificity. Processing of action-related words, be they action verbs, tool names, or other action-related lexical items, tended to activate frontocentral cortex, including inferior frontal or premotor areas, more strongly than words without strong semantic action links. The same was found for temporo-occipital areas involved in motion perception. On the other hand, words with visual semantics tended to activate visual and inferior temporal cortex or temporal pole more strongly than action-related words. This differential activation was interpreted as evidence for semantic category specificity in the human brain (Martin and Chao 2001; Pulvermüller 1999).

Some problems with semantic category specificity

The results from metabolic and neurophysiological imaging demonstrate the activation of neuronal assemblies with different cortical distributions in the processing of action- and visually-related words and concepts. However, it has been asked whether the reason for the differential activation observed would necessarily be semantic or conceptual in nature. Could there be alternative explanations?

Although the broad majority of the imaging studies of category specificity support the idea that semantic factors are crucial, there is work that could not provide converging evidence (Devlin *et al.* 2002; Tyler *et al.* 2001). These studies used particularly

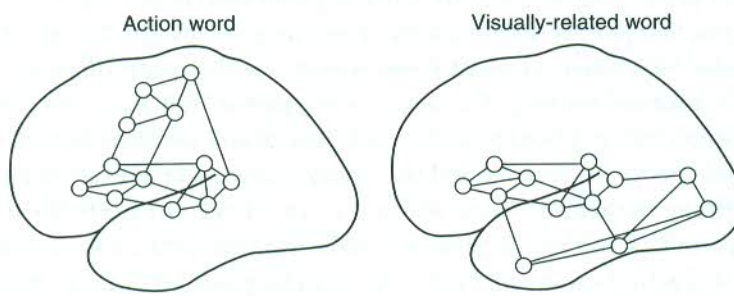

Fig. 6.3 A brain-based model of category-specific processing of words with different semantics. Words semantically related to actions may be cortically processed by distributed neuron ensembles linking together word forms and action programs. Words referring to objects that are perceived through the visual modality may be processing by neuron sets distributed over language areas and the visual system (Pulvermüller 1996).

well-matched stimuli, so that word length, frequency, and other psycholinguistic factors could not account for any possible differences in brain activation. Therefore, these authors argued that these factors might account for differences between 'semantic' categories reported previously. Although some earlier studies reporting semantic category differences performed meticulous stimulus matching for a range of psycholinguistic factors – word length and frequency included (Kiefer 2001; Preissl *et al.* 1995; Pulvermüller, Lutzenberger *et al.* 1999; Pulvermüller, Mohr *et al.* 1999) – a number of studies did not control for these factors. As pointed out previously (Bird *et al.* 2000), nouns tend to have more highly imageable meaning than verbs, whereas verbs tend to have higher word frequency. Any difference in brain activation, and also any differential vulnerability to cortical lesion, could thus be explained as an imageability–frequency dissociation, rather than in terms of semantic categories. Similarly, animals tend to be more alike than tools from both a visual and a conceptual point of view, and it has therefore been argued that perceptual and conceptual structure could contribute to the explanation of category dissociations (Humphreys and Riddoch 1987; Rogers *et al.* 2004; Tyler *et al.* 2000). On these grounds, at least some evidence for category specificity has therefore been criticized for not being fully convincing.

What makes things worse is that predictions on where category-specific activation should occur in the brain have not always been very precise. Whereas rough estimates, such as the prediction that action semantics should involve frontal areas and visual semantics temporo-occipital areas, could be provided and actually confirmed, more precise localization was sometimes surprising and not *a priori* predictable. For example, semantic information related to processing of colour and motion information semantically linked to words and pictures was reported to occur ~2 centimetres anterior to the areas known to respond maximally to colour or motion, respectively (Martin *et al.* 1995). It would be desirable to have evidence for category-specific semantic activation at precisely the locus a brain-based action–perception theory of semantic processing would predict. Such a perspective is opened by looking at subtypes of action words.

Action words

Action words are defined by abstract semantic links between language elements and information about actions. These words refer to actions and the neurons that process the word forms are likely interwoven, with neurons controlling actions. The motor cortex is organized in a somatotopic fashion with the mouth and articulators represented close to the sylvian fissure, the arms and hand at dorsolateral sites and the foot and leg projected to the vertex and interhemispheric sulcus (Penfield and Rasmussen 1950) (Figure 6.4). Additional somatotopic maps exist in the frontocentral cortex (He *et al.* 1993), among which one of the more prominent ones lies in the premotor cortex in the lateral precentral gyrus, and resembles the map in the primary motor cortex (Matelli *et al.* 1986; Rizzolatti and Luppino 2001). As many action words are preferably used to refer to movements of the face or articulators, arm or hand, or leg or foot, the distributed neuronal assemblies would therefore include semantic neurons in perisylvian (face words), lateral (arm words), or dorsal (leg words) motor and premotor cortex (Pulvermüller 1999).

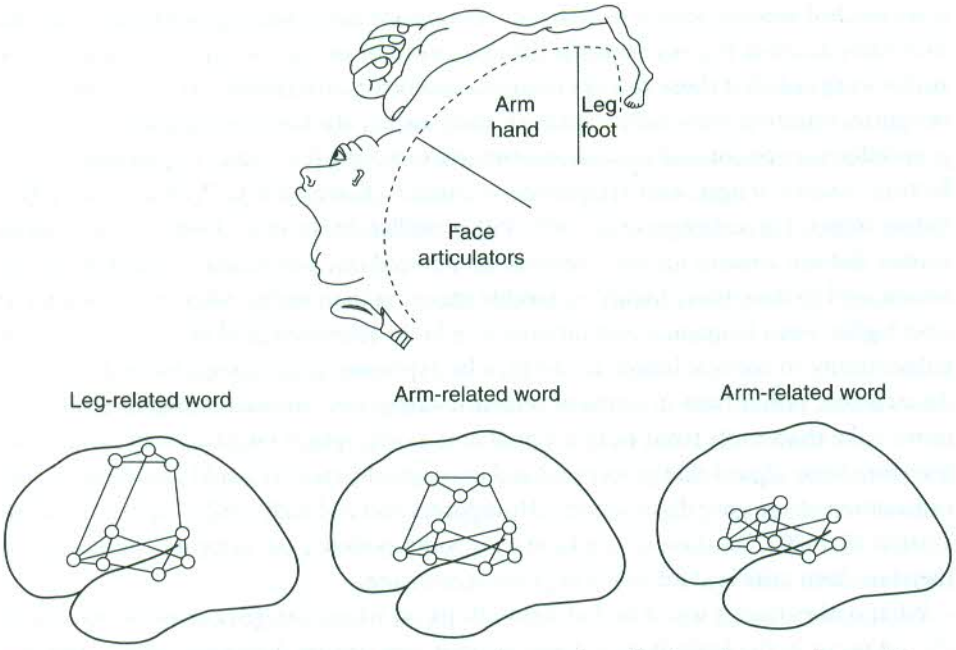

Fig. 6.4 Semantic somatotopy model of action word processing: distributed neuronal assemblies bind information about word forms and the actions they refer to semantically. Because action words can relate to different parts of the body (examples: 'lick', 'pick', 'kick'), the cortical distributions of their action–perception networks differ between each other (Pulvermüller 2001). The inset shows the somatotopy of the primary motor cortex as revealed by Penfield and Boldrey (1937).

This is the essence of the *somatotopy of action word* model, which implies differently distributed networks for the English words 'lick', 'pick', and 'kick' (Figure 6.4). The model allows for general predictions on action word-related cortical activity within the limits of the well known interindividual variation of cortical maps, most notably as a result of practice-related reorganization (Elbert *et al.* 1995), and is open to further elaboration by taking into account additional mapping rules, for example the topography of coordinated actions in a body-centred workspace suggested by recent work (Graziano *et al.* 2002).

Crucial predictions of the semantic somatotopy model is that perception of spoken or written action words should activate cortical areas involved in action control and execution in a category-specific somatotopic fashion, depending on the semantics of the action words. As the cortical areas of action control and execution can be defined experimentally, one could in principle use such action localizer experiments to predict exactly where semantic activation should occur for different aspects of action-related meaning.

In functional imaging experiments, elementary repetitive movements of single body parts activate motor and premotor cortex. For example, Hauk *et al.* reported fMRI data showing that tongue, finger, and foot movements lead to the somatotopic activation

pattern illustrated in Plate 6.3 (diagram on the left; Hauk *et al.* 2004). When the same subjects were instructed to silently read action words related to the face, arm, and leg that were otherwise matched for important psycholinguistic variables (such as word frequency, length, and imageability), a similar pattern of activation emerged along the motor strip (Figure 6.3, right; Hauk *et al.* 2004). Consistent with earlier findings, all words equally activated areas in the temporal cortex and also in the inferior frontal cortex (Pulvermüller *et al.* 2003; Wilson *et al.* 2004; Zatorre *et al.* 1992). The additional category-specific somatotopic activation in response to face-, arm-, and leg-related words seen in the motor system was close to and overlapped with the motor and premotor representations for specific body part movements obtained in the motor localizer tasks. These results indicate that specific action representations are activated in action word understanding. The fact that the locus of semantic activation could be predicted by a theory of APNs provides strong evidence for this theory in particular and for the embodiment of aspects of semantics in action mechanisms in general.

A similar experiment was carried out with action words embedded into spoken sentences. In this case, subjects heard action descriptions such as 'The boy kicked the ball' or 'The man wrote the letter' while their brain metabolism was monitored (Tettamanti *et al.* 2005). Specific premotor areas reflecting the differential involvement of body part information in the semantic analysis of the language input were again found active. Taken together, these fMRI results indicate that somatotopic activation of motor circuits reflects aspects of word and sentence meaning, and that such activation can be elicited by spoken and by written language.

6.4.4 Somatotopic activation: semantic or epiphenomenal?

Although language-related somatotopic cortical activation could be demonstrated, the low temporal resolution of haemodynamic imaging makes it impossible to decide between two interpretations of this finding: One possibility is that the activation of specific action-related networks directly reflects action word recognition and comprehension, as the somatotopy of action word model would suggest. An alternative possibility has been pointed out by Glenberg and Kaschak (2002) in the context of behavioural work on embodiment. It is possible that thoughts about actions actually follow the comprehension process and behavioural, but also brain physiological, effects relate to such *post-comprehension inference*. Inferences would be triggered by the comprehension of a word or sentence, but would not necessarily reflect processes intrinsically linked to language comprehension. Importantly, earlier fMRI research has shown that observation of action-related pictures, but also mere voluntary mental imagery of actions, can activate motor and premotor cortex in a somatotopic fashion (Buccino *et al.* 2001; Jeannerod and Frak 1999). Therefore, it is important to clarify whether motor system activation to action-related language processing reflects the comprehension process *per se* or rather a later stage following language comprehension. Apart from mental imagery of actions, possible post-comprehension processes include planning of action execution, recalling an action performed earlier, and reprocessing the meaning of the language stimulus.

How is it possible to separate comprehension processes from subsequent inferences and other epiphenomenal mental activities? Let me propose that brain processes reflecting comprehension can be characterized as: immediate, automatic, and functionally relevant.

Early effects of lexical and semantic processing are known to occur around 100–200 milliseconds after critical stimulus information comes in (Pulvermüller 1996; Sereno *et al.* 1998). In contrast, late postlexical meaning-related processes are reflected by late components of the event-related potential (ERP) and field, which are maximal around 400 milliseconds after word onset (Holcomb and Neville 1990). If the activation of motor areas is related to semantic processes intrinsically tied to word form access, it should take place *immediately* (within the first 200 milliseconds) after stimulus information allows for the unique identification of an incoming word. *Automaticity* refers to the idea that when seeing or hearing a word it is hardly possible to avoid understanding its content; comprehension might even occur without intentionally attending to the stimuli. Therefore, brain processes reflecting comprehension might be expected to persist under distraction, when the subjects' attention is directed away from the critical language stimuli. In the case of *functional relevance*, if the presentation of action words leads to specific activation of motor systems relevant to word processing, one may expect that a change of the functional state of these motor systems leads to a measurable effect on the processing of words semantically related to actions. However, if somatotopic activation of motor systems did reflect a post-comprehension process, it can be late (substantially greater than 200 milliseconds) and absent under distraction, and functional changes in the motor system would be without effect on word processing. A series of experiments was conducted to investigate these three issues.

Immediacy

To reveal the time course of cortical activation in action word recognition and find out whether specific motor areas are sparked immediately or after some delay, neurophysiological experiments were conducted. Experiments using ERPs looking at silent reading of face, arm and leg words showed that category-specific differential activation was present ~200 milliseconds after word onset (Hauk and Pulvermüller, 2004). Consistent with the fMRI results, distributed source localization performed on ERPs revealed an inferior frontal source close to the motor representation of the face and mouth that was strongest for face-related words, and a superior central source close to the leg representation that was maximal for leg-related items (Hauk and Pulvermüller 2004). This dissociation in brain activity patterns supports the notion of stimulus-triggered early lexicosemantic processes. To investigate whether motor preparation processes co-determined this effect, experiments were performed in which the same response – a button press with the left index finger – was required for all words. The early activation difference between face- and leg-related words persisted, indicating that lexicosemantic processes rather than postlexical motor preparation were reflected (Pulvermüller *et al.* 2000).

In summary, the somatotopic activation that reflects word meaning aspects therefore appeared about one-quarter of a second after information about the words in the input was available. Earlier physiological studies of psycholinguistic processes had shown

previously that the first brain responses reflecting comprehension and psycholinguistic information access at higher lexical and semantic levels appear at this point (Pulvermüller 1996; Sereno and Rayner 2003). Therefore, the early somatotopic mapping of meaning aspects reflects comprehension processes.

Automaticity

The earliness of word category-specific semantic activation along the sensorimotor cortex in passive reading tasks suggests that this feature might be automatic. To further investigate this possibility, subjects were actively distracted while action words were being presented and brain responses were measured (Pulvermüller, Shtyrov *et al.* 2005; Shtyrov *et al.* 2004). Subjects were instructed to watch a silent video film and ignore the language input while spoken face-/arm- and leg-related action words were presented. Care was taken to exactly control for physical and psycholinguistic features of the word material. For example, the Finnish words 'hotki' (eat) and 'potki' (kick) – which included the same recording of the syllable *[kI]* spliced to the end of each word's first syllable – were compared (Näätänen *et al.* 2001). In this way, any differential activation elicited by the critical final syllable *[kI]* in the context of *[hot]* or *[pot]* can be uniquely attributed to its lexicosemantic context. MEG results showed that a MMN that was maximal at 100–200 milliseconds after onset of the critical syllable was elicited by face/arm- and leg-word contexts (Plate 6.4). Relatively stronger activation was present in the left inferior frontal cortex for the face/arm-related word, but significantly stronger activation was seen in superior central areas, close to the cortical leg representation, for the leg-related word (Pulvermüller, Shtyrov *et al.* 2005).

These MEG results were confirmed with EEG using words from different languages, including, for example, the English word-pair 'pick' and 'kick' (Shtyrov *et al.* 2004). It is remarkable that peak activation of the superior central source followed that of the inferior frontal source with an average delay of only 30 milliseconds, consistent with the spread of activation being mediated by fast-conducting cortico–cortical fibres between the perisylvian and dorsal sensorimotor cortex. This speaks in favour of automatic activation of motor areas in action word recognition and therefore further strengthens the view that this activation reflects comprehension. It appears striking that differential activation of body-part representations in sensorimotor cortex to action word subcategories was seen across a range of cognitive paradigm, including lexical decision, attentive silent reading, and oddball paradigms under distraction. This further supports the idea that word-related, rather than task- or strategy-dependent, mechanisms are being tapped into.

Functional relevance

Even if action word processing sparks the motor system in a specific somatotopic fashion, and even if this activation is fast and automatic, it still does not necessarily imply that the motor and premotor cortex influence the processing of action words. Different parts of the motor system were therefore stimulated with weak magnetic pulses while subjects had to process action words in a lexical decision task (Pulvermüller, Hauk *et al.* 2005). To minimize interference between word-related activation of the motor system

and response execution processes, lip movements were required while arm- and leg-related words were presented. Subthreshold TMS applied to the arm representation in the left hemisphere, where strong magnetic pulses elicited muscle contractions in the right hand, led to faster processing of arm words relative to leg words, whereas the opposite pattern of faster leg- than arm-word responses emerged when TMS was applied to the cortical leg area (Pulvermüller, Hauk *et al.* 2005). Processing speed did not differ between stimulus word groups in control conditions in which ineffective 'sham' stimulation or TMS to the right hemisphere was applied. This shows a specific influence of activity in the motor system on the processing of action-related words.

Further evidence for specific functional links between the cortical language and action systems comes from TMS-induced motor responses (Fadiga *et al.* 2002). Listening to Italian sentences describing actions performed with the arm or leg differentially modulates the motor responses brought about by magnetic stimulation of the hand and leg motor cortex (Buccino *et al.* 2005). It appears that effective specific connections of language and action systems can be documented for spoken or written language, at the word and sentence levels, and for a variety of languages (English, Italian, German, Finnish) using a variety of neuroscience methods (fMRI, MEG, EEG, TMS).

6.4.5 Interim summary

These experiments show that the activation of motor systems of the cortex occurs early in action word processing, is automatic to some degree, and has a semantically specific functional influence on the processing of action words. This provides brain-based support for the idea that motor area activation is related to comprehension of the referential semantic meaning of action words. In the wider context of a theory of embodiment of conceptual and semantic processing, the conclusion is that comprehension processes are related to, or embodied in, access to action information. It is noteworthy that neuroscience evidence was crucial in revealing this (Pulvermüller 2005). However, it is equally true that behavioural results are consistent with these conclusions and further strengthen the embodiment of language in APNs (Borghi *et al.* 2004; Boulenger *et al.* 2006; de Vega *et al.* 2004; Gentilucci *et al.* 2000; Glenberg and Kaschak 2002).

6.5 Fundamental issues in sensorimotor semantics

6.5.1 Can motor cortex map aspects of semantics?

One may question the idea that activity in the motor system might actually reflect semantic processes from a principled theoretical perspective. The idea that there should be a specific centre for semantics is still dominating (although there is little agreement between researchers about where the semantics area is situated, see section 6.4.3 on one or many semantics centres below). Areas that deal with the trivialities of motor movements, and equally those involved in elementary visual feature processing, are therefore thought by some to be incapable of contributing also to the higher processes one might be inclined to reserve for humans. In this context, it is important to point to the strong evidence that activation in motor systems directly reflects aspects of semantics.

Evidence that semantic features of words are reflected in the focal brain activation in different parts of sensorimotor cortex comes from MEG work on action words; there was a significant correlation between local source strengths in inferior face/arm-related and dorsal leg-related areas of sensorimotor cortex and the semantic ratings of individual words obtained from study participants (Pulvermüller, Shtyrov *et al.* 2005). This means that the subjects' semantic ratings were reflected by local activation strength and leaves little room for interpretations not of a semantic nature.

Even though the action-related and visually-related features discussed, and the associative learning mechanisms binding them to language materials, may not account for all semantic features of relevant word-related concepts, it seems clear that they reflect critical aspects of word meaning (Pulvermüller 1999): Crocodiles are defined by certain properties, including form and colour features, in the same way as the concepts of walking or ambulating are crucially linked to moving one's legs. Certainly, there is room for derived, including metaphorical, usage. As a big fish might be called the crocodile of its fish tank even if it is not green, one may speak of walking on one's hands or taking a stroll through the mind (thus ignoring the feature of body-part relatedness). One may even tell a story about a crocodile with artificial heart and kidneys, although it is generally agreed upon, following Frege, that these ingredients are part of the definition of an animal (or, more appropriately, a higher vertebrate) (Frege 1966). That writing is related to the hand may therefore be considered an analytical truth, in the same way as a crocodile is defined as having a heart, and in spite of the fact that it is possible to write in the sand with one's foot. This may simply be considered a modified type of writing, as the post-surgery crocodile is a modified crocodile. An instance of a heartless crocodile and leg-related writing is possible, but probably closer to metaphorical usage of these words than to their regular application. It seems safe to include perceptual properties such as green-ness and action aspects such as hand-relatedness in the set of possible semantic and conceptual features. Exceptions cannot prove a rule wrong.

6.5.2 Learning new meanings from language context: neuronal mechanisms of parasitic semantic feature extraction

The critical problem of learning new word meanings from context is frequently raised against embodied approaches to semantics, including the neuronal assembly model. Especially, the idea that word–world correlation provides a significant explanation of the acquisition of word meanings has been criticized, because it is well known that only a minority of words are actually being learned in the context of reference object perception and action execution (Kintsch 1974, 1998). However, after action–perception learning of aspects of word meaning has taken place for a sufficiently large set of words, it becomes feasible to learn semantic properties *parasitically* when words occur together in strings, sentences, or texts. A neuroscientific basis for this 'parasitic semantic learning' might lie in the overlap of word-related neuronal assemblies in the perisylvian language areas and the lack of semantic neurons related to action and perception information outside the perisylvian space of the networks processing new words with unknown semantic features (Pulvermüller 2002). In this case, a new word would activate its form-related perisylvian

neuronal assembly, while neurons outside the perisylvian space are still actively processing aspects of the semantics of context words. The correlated activation of the semantic neurons of context words and the form-related perisylvian neurons of the new word lead to linkage of semantic features to the new word form. This provides a potential basis of second-order (parasitic) semantic learning and provides a putative neuroscience explanation for why correlation approaches to word meaning are successful in modelling semantic relationships between words (Kintsch 2002; Landauer and Dumais 1997).

However, it is important to note that this mechanism can only succeed if a sufficiently large set of semantic features and words is learned through correlation of perception, action, and language-form features in the first place. Otherwise, what Searle called the Chinese room argument, implying that semantic information cannot emerge from correlation patterns between symbols, cannot be overcome (Searle 1980). Action–perception correlation learning and the learning of correlations between language units are both indispensable for extracting semantic information for large vocabularies. Semantic knowledge is rooted in word-world and word-word correlation.

6.6 Early and late semantic activation

The time course of semantic activation in action word recognition was on a rather short scale. Relevant areas were seen to be active within 200 milliseconds after critical stimulus information came in (Pulvermüller, Shtyrov *et al.* 2005; Shtyrov *et al.* 2004). This suggests early semantic activation, as early as the earliest processes reflecting phonological or lexical information access (Hauk *et al.* 2006; Obleser *et al.* 2003; Shtyrov *et al.* 2005). However, the early neurophysiological reflection of semantic brain processes does not imply that meaning processing is restricted to the first 200 milliseconds after a word can be identified. There is ample evidence for neurophysiological correlates of semantic processes that take place later on (Coles and Rugg, 1995). These later processes may follow up on the early semantic access processes and may reflect reinterpretation or in-depth processing, which is especially important in circumstances where the context or other factors make comprehension difficult.

6.6.1 Abstract semantics from a brain perspective

These results summarized in section 6.4 demonstrate that action words activate the cortical system for action processing in a somatotopic fashion and that this somatotopy reflects word meaning. However, they do not imply that all aspects of the meaning of a word are reflected in the brain activation pattern it elicits. It is possible to separate brain correlates of semantic features specifying face-, arm-, and leg-relatedness, or, in the visual domain, of colour and form features (Moscoso Del Prado Martin *et al.* 2006; Pulvermüller and Hauk 2006; Simmons *et al.* 2007). It even became possible to provide brain support for the grounding of words referring to odours in olfactory sensation and evaluation mechanisms in brain areas processing olfactory and emotion-related information (de Araujo *et al.* 2005; González *et al.*, 2006). However, for other semantic features, the idea that their meaning can be extracted from sensory input, or deduced from output

patterns, is more difficult to maintain. Although the question of how an embodiment perspective would explain abstraction processes has frequently been addressed (Barsalou 1999, 2003; Lakoff 1987), it is still not clear whether all semantic feature can – and have to – be extracted from input–output patterns.

A brain perspective might help to solve aspects of this issue. There are highly abstract concepts for which a deduction from sensory input is difficult to construe. Barsalou tried to ground the meaning of the word 'or' in the alteration of the visual simulations of objects (Barsalou 1999). However, if this view is correct, the disjunctive concept would be realized as the alteration mechanism allowing the brain to switch on and off specific representations alternatively. Looking at the brain theory literature, it is evident that every brain, even every primitive nervous system, is equipped with mechanisms for calculating disjunction, conjunction, negation, and other logical operation. This was the content of an early article by McCulloch and Pitts (1943) entitled 'A logical calculus of ideas immanent in nervous activity' that has since inspired much work in the theory of automata and language (e.g. Kleene 1956; Schnelle 1996).

The main points of this logical calculus theory of neuronal function still hold true, although neuronal models have significantly improved since the proposal was first made (e.g., Bussey et al., 2005). McCulloch and Pitts (1943) pointed out that a circuit including two neurons that project onto a third one will necessarily, given the activation threshold of neuron number three is adjusted in specific ways, give rise to the computation of a conjunction or disjunction function (Figure 6.5). Negation, identity, and either-or computations are equally straightforward. These examples demonstrate that our brain comes with built-in mechanisms relevant for abstract semantic processing. There is no need to construe the semantics of 'and', 'or', and other highly abstract words exclusively from sensorimotor information. The very fact that these mechanisms are built into nerve cell circuits may enable us to abstract away from the sensory input to more and more general concepts.

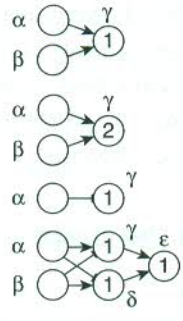

Fig. 6.5 Logical circuits immanent to a network of neurons as discussed by McCulloch and Pitts (1943). If two neurons project to a third neuron, the activation threshold of the third neuron will determine whether it acts like a logical element symbolizing 'or' (uppermost diagram) or 'and' (second from top). Circuits symbolizing 'not' and 'either-or' can also be implemented. Arrows stand for excitatory and T-shaped line endings for inhibitory connections. Numbers indicate activation thresholds (after Pulvermüller 2003).

One may argue that this proposal means watering down a 'radically embodied' perspective in the cognitive and brain sciences as the logical circuits are not derived from actions or perceptions but rather represent a neuronal a priori. However, the proposal here is to model language and concepts in mechanistic neuronal circuits. This implies to put to use those mechanisms that are evidently present (built-in) in nervous systems and to explore their implications for linguistic and semantic representations and processes. If the mechanisms of embodiment are brain mechanisms, we have little choice but to accept the functional principles immanent to neuronal function.

Abstraction by either-or computation may be the basis of action representation at different levels of the action description, corresponding to different levels of abstractness. Moving one's arm in such and such a way is a basic action; opening a door could imply exactly the same movement but with characteristic somatosensory and possibly auditory input; and freeing somebody could also be realized by performing the same basic action. To implement the aspects of the action semantics of 'open', it is possible to connect disjunction neurons with a range of action control neurons coordinating alternative action sequences that would allow one to open doors, boxes, and other objects. Similarly, in order to implement action semantics of the word 'free', higher order disjunction neurons would be needed that look at a range of different movement programs one could perform in the context of setting somebody or something free. Again, additional conditions would need to be met. The performer would need to assume that someone or something is captured, locked in, or contained in something else and would have the intention to get him/it out etc. Disjunction neurons receiving input from a range of concrete action representations may be located adjacent to motor and premotor sites and would be ideally placed in prefrontal cortex (see Pulvermüller 1999). This hypothesis might seem somewhat speculative on first appearance, and it is therefore important to point to the evidence that supports it. The prediction that words with more abstract action-related meaning are processed in areas adjacent and anterior to motor and premotor cortex, that is, in dorsolateral prefrontal areas, receives support from recent imaging work (Binder *et al.* 2005; Pulvermüller and Hauk 2006).

These hints towards abstraction mechanisms of different types might suffice here to point out some perspectives of a brain-based approach to embodied semantics. In a nutshell, I believe that a wide range of abstract concepts can be modelled, provided that the brain's built-in circuits are taken into account.

6.7 Language processes in the brain

6.7.1 Distributed processing

Similar to most current theories postulating distributed processing of language and concepts in the mind and brain (Braitenberg and Pulvermüller 1992; Rogers *et al.* 2004; Seidenberg *et al.* 1994), the proposal put forward here postulates that the cell assemblies processing language and concepts are widely distributed. This means that the neuronal ensembles are spread out over different areas of the brain, so that the areas would become more active in neuroimaging experiments when critical stimuli are under processing.

In the same way, the compartments of neuronal models simulating the relevant brain mechanism should show activation dynamics in a specific manner when the processing of language and concepts is being simulated (Garagnani *et al.* 2007; Pulvermüller 1999; Pulvermüller and Preissl 1991; Wennekers *et al.* 2006; Wermter *et al.* 2004). However, in contrast to most distributed processing accounts, the neuronal assemblies are conceptualized as functionally coherent networks that respond in a discrete fashion. This implies that the networks representing words, the 'word webs', are either active or inactive and that the full activation of one word's representation is in competition with that of other word-related networks. In this sense, neuronal assemblies are similar to the localist representations postulated by psycholinguistic theories (Dell 1986; Page 2000; Roelofs 1992). Still, as each of the distributed neuronal assemblies includes neurons processing features related to form or semantics, there can be overlap between neuronal assemblies representing similar words or concepts. This leads to an interplay of facilitatory and inhibitory mechanisms when word webs become fully active in a sequence. The full activation, or ignition, of a word-related neuronal assembly can be considered a possible cortical correlate of word recognition – or of the spontaneous pop-up of a word together with its meaning in the mind.

6.7.2 Discrete processing

If word webs can become active in a discrete fashion, this does not imply that each ignition is identical to all other full activations of the network. As word webs are linked to each other through the grammar network and also exhibit semantic and form overlap, the context of brain states and other cognitive network activations primes and therefore influences the way in which a given neuronal assembly ignites. This mechanism can be related to the observation that the meaning or 'sense' of a word in given contexts cannot be reduced to a 'core meaning', but should rather be conceptualized as a family of similar context-dependent semantic feature sets (Barsalou 1982; Wittgenstein 1953). Neuronal assemblies whose precise pattern of ignition depends on contextual priming may provide a mechanism for context-related selection of semantic features.

A further example illustrating the benefit of discrete representations is the contextual disambiguation of a semantically ambiguous word. The brain basis of an ambiguous word has been conceptualized as a set of two word webs overlapping in their form-related assembly part (Figure 6.6). Semantic context can, in this case, disambiguate by priming one of the semantic subassemblies of the two overlapping word representations. The two overlapping neuronal assemblies would be in both facilitatory (due to form overlap) and inhibitory (due to competition between neuronal assemblies) interaction. Most likely, the facilitatory effects would precede the inhibitory ones (Pulvermüller 2003). It is more difficult to envisage a mechanism for the separation of the meanings of ambiguous words in a fully distributed network without discrete representations.

6.7.3 Semantic conceptual binding sites

Although the results on the cortical correlates of semantic word groups cannot be explained if all semantic processes are restricted to one cortical area, they might still be compatible with the general idea of a central semantic binding site. This system would

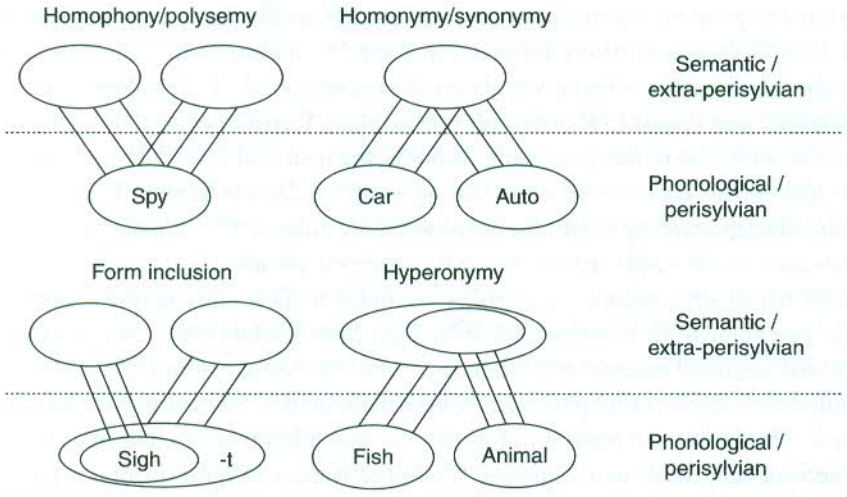

Fig. 6.6 Word forms and semantics are proposed to be processed in different parts of the distributed word-related cell assembly. Most semantic information is stored outside the perisylvian language areas, whereas most information related to the phonological word form is laid down in perisylvian space. Words related to each other phonologically or semantically would overlap in their perisylvian or extra-perisylvian sub-assemblies. Relationships between words with the same form but different meanings (homophones and polysemes), with the same meaning but different forms (homonyms and synonyms), with a form-inclusion and a hyperonomy/hyponymie relationship are illustrated (after Pulvermüller 2003).

be thought to manage dynamic functional links between multiple cortical areas processing word forms and conceptual semantic information. The idea of such a central 'concept area' or 'convergence zone' has a long tradition in the neuroscience of language and seems to be motivated by the believe that a central locus must exist, at which concepts are related to each other and abstract information is extracted from them. As I have tried to make clear, it may be possible to implement semantic binding by distributed neuronal assemblies which, as a whole, function as binding networks. In this case, the binding would not be attributable to one specific brain area but rather to a set of areas, those over which the assembly is distributed. Still, there are certainly more 'peripheral' areas, for example, primary motor and sensory cortices, where correlated activation patterns occur in the first place, and areas connecting the peripheral ones, with the major task of linking the correlation patterns together in the most effective manner. These higher or connection areas might naturally be more important for the binding of information from different modalities. Whether or not all these routes for multimodal information linkage necessarily go through the same convergence zone, or rather through a range of different areas of association cortex, as neuroanatomical studies might suggest (e.g., Braitenberg and Schüz 1998; Young *et al.* 1994), remains a matter of future research.

A possible route to answering the question of a centre for semantic and conceptual binding is offered by patient studies. Here it is remarkable that patients with semantic dementia usually have most severe neuronal degradation in the temporal pole. Therefore, this region

was suggested as the area most important for semantic binding (see Patterson and Hodges 2001). However, the bilateral nature of neural degeneration usually seen in semantic dementia may suggest that one focal lesion is not enough to cause general semantic deficits (Patterson and Hodges 2001).

Multiple semantic binding sites are also supported by the specific semantic deficit in action-word processing seen in patients with motor neurone disease (Bak *et al.* 2001). More evidence for multiple semantic binding sites came from double dissociations between semantic word categories arising from lesions in right-hemispheric frontoparietal versus temporo-occipital areas (Neininger and Pulvermüller 2003), which complement similar observations made earlier for lesions to the left language-dominant hemisphere (Damasio and Tranel 1993; Daniele *et al.* 1994). Some of these lesions were so focal that they only affected motor and premotor cortex, but nevertheless specifically degraded the processing of action words in psychological experiments (Neininger and Pulvermüller 2001). Dissociations of these types are consistent with the existence of multiple semantic integration systems in both cerebral hemispheres (Pulvermüller and Mohr 1996).

6.8 **Discussion**

A model of embodiment of word forms and semantics at the level of neurons in the brain was outlined. APNs may bind together distinctive articulatory and acoustic information to yield transmodal representations of spoken word forms. The neuronal assemblies for word forms may be spread out over inferior frontal and superior temporal parts of left perisylvian language cortex (blue and green areas in Plate 6.1; see also Figure 6.2). Aspects of actions and perceptions relevant in the explanation of referential words may be linked to the word forms by way of neuronal correlation of the activations of the perisylvian word-form assembly and semantic neurons in action- and perception-related brain regions (Figure 6.3). Semantic category differences may be based on the differential involvement of neuronal sets in inferior temporal visual cortex, frontocentral action-related cortex, and other brain parts. Precise predictions on the cortical locus of specific semantic brain processes could be generated for subtypes of action words referring to face, arm, and leg movements, such as 'lick', 'pick', and 'kick' (Figure 6.4). Processing of these words lights up the motor system in a similar way as the respective actions would (Plate 6.3).

This semantic somatotopy is complemented by a phonological somatotopy; listening to speech sounds produced with the lips or tongue activates the areas specifically involved in lip or tongue movements, which also contribute to phoneme articulation (Plate 6.2). Specific activation of the motor systems takes place rapidly during speech and written language processing (Plate 6.4), is automatic and makes a functional contribution to word processing. This provides brain support that language is grounded in, and embodied by, action and perception mechanisms. In addition to elementary features of object and action reference, a brain model of abstract meaning can be built on the basis of known properties of neuronal circuits, for example their ability to implement logical operations (Figure 6.5). Abstract concepts may therefore develop in brain regions adjacent and anterior to areas involved in the processing of referent actions and objects (Plate 6.5).

The data summarized show that it is fruitful to model the brain basis of meaningful words as distributed neuronal assemblies binding phonological and semantic information about actions and perceptions at an abstract or cross-modal level. These distributed neuronal ensembles may function as discrete word-specific processors including neuron sets in different cortical areas. Different neuronal assemblies may overlap, thereby reflecting shared semantic or phonological features between words, and they may compete for full activation in the perception process. This view is consistent with context-dependent meaning processing and allows for modelling of word pairs that are semantically ambiguous or homophonous (Figure 6.5). In word recognition, activation of the distributed areas over which these neuronal assemblies are spread out is near-simultaneous, thereby binding information from different modalities (e.g., articulatory and acoustic) and linguistic functions (e.g., phonological and semantic). Apart from their role in language, these networks may play a role in conceptual processing.

These proposals and the reviewed neuroscience evidence backing them have important implications for constructing life-like APNs and robots with brain-like control systems (Knoblauch *et al.* 2005; Roy *et al.* 2005; Shastri *et al.* 2005; Wermter *et al.* 2004, 2005). A major conclusion here is that language models likening brain mechanisms have good reason to link up language processors to the body-related systems the language elements provide information about (e.g., which articulator is being moved, which colour is being referred to, which action is meant). Such embodied artificial models might succeed for the same reason why the biological originals they copy were successful in evolution. A main point here is the possibility to process cross-modal information in an extremely rapid manner.

Debate

The debate corresponding to this chapter, as well as Chapters 5 and 7, is included at the end of Chapter 7.

Author note

For comments and discussions, I am grateful to Larry Barsalou, Manuel de Vega, Arthur Glenberg, Arthur Graesser, Olaf Hauk, Markus Kiefer, Karalyn Patterson, and Yury Shtyrov. This work was supported by the UK Medical Research Council (grants U1055.04.003.00001.01, U1055.04.003.00003.01) and by the European Community under the Information Society Technologies and the New and Emerging Science and Technology programmes (EU IST-2001-35282, EU NEST-Nestcom).

References

Artola A, Singer W (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends in Neurosciences*, 16, 4807.

Bak TH, O'Donovan DG, Xuereb JH, Boniface S, Hodges JR (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the Motor Neurone Disease–Dementia–Aphasia syndrome. *Brain*, 124, 103–20.

Barsalou LW (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*, 10, 82–93.

Barsalou LW (1999). Perceptual symbol systems. *Behavioural and Brain Sciences*, 22, 577–609.

Barsalou LW (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, 358, 1177–87.

Bi GQ (2002). Spatiotemporal specificity of synaptic plasticity: cellular rules and mechanisms. *Biological Cybernetics*, 87, 319–32.

Binder JR, Westbury CF, McKiernan KA, Possing ET, Medler DA (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, 17, 905–17.

Bird H, Lambon-Ralph MA, Patterson K, Hodges JR (2000). The rise and fall of frequency and imageability: noun and verb production in semantic dementia. *Brain and Language*, 73, 17–49.

Bookheimer S (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25, 151–88.

Borghi AM, Glenberg AM, Kaschak MP (2004). Putting words in perspective. *Memory and Cognition*, 32, 863–73.

Boulenger V, Roy AC, Paulignan Y, Deprez V, Jeannerod M, Nazir TA (2006). Cross-talk between language processes and overt motor behaviour in the first 200 msec of processing. *Journal of Cognitive Neuroscience*, 18, 1607–15.

Braitenberg V (1978). Cell assemblies in the cerebral cortex. In R Heim, G Palm, Eds. *Theoretical Approaches to Complex Systems. Lecture Notes in Biomathematics Vol. 21* (pp. 171–88). Berlin: Springer.

Braitenberg V, Pulvermüller F (1992). Entwurf einer neurologischen Theorie der Sprache. *Naturwissenschaften*, 79, 103–17.

Braitenberg V, Schüz A (1992). Basic features of cortical connectivity and some considerations on language. In J Wind, B Chiarelli, BH Bichakjian, A Nocentini, A Jonker, Eds. *Language Origin: A Multidisciplinary Approach* (pp. 89–102). Dordrecht: Kluwer.

Braitenberg V, Schüz A (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity* (Second edition). Berlin: Springer.

Brodmann K (1909). *Vergleichende Lokalisationslehre der Grobhirnrinde*. Leipzig: Barth.

Buccino G, Binkofski F, Fink GR *et al.* (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13, 400–4.

Buccino G, Riggio L, Melli G, Binkofski F, Gallese V, Rizzolatti G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioural study. *Brain Research: Cognitive Brain Research*, 24, 355–63.

Bussey TJ, Saksida LM, Murray EA (2005). The perceptual-mnemonic/feature conjunction model of perirhinal cortex function. *Quarterly Journal of Experimental Psychology B*, 58, 269–82.

Cappa SF, Perani D, Schnur T, Tettamanti M, Fazio F (1998). The effects of semantic category and knowledge type on lexical-semantic access: a PET study. *NeuroImage*, 8, 350–9.

Chao LL, Haxby JV, Martin A (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2, 913–19.

Clark A (1996). *Being There: Putting Brain, Body, and World Together Again*. Boston, MA: MIT Press.

Coles MGH, Rugg MD (1995). Event-related brain potentials. In MD Rugg, MGH Coles, Eds. *Electrophysiology of Mind: Event-related Brain Potentials and Cognition* (pp. 1–26). Oxford: Oxford University Press.

Damasio AR, Tranel D (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences, USA*, 90, 4957–60.

Daniele A, Giustolisi L, Silveri MC, Colosimo C, Gainotti G (1994). Evidence for a possible neuroanatomical basis for lexical processing of nouns and verbs. *Neuropsychologia*, 32, 1325–41.

de Araujo IE, Rolls ET, Velazco MI, Margot C, Cayeux I. (2005). Cognitive modulation of olfactory processing. *Neuron*, 46, 671–9.

de Vega M, Robertson DA, Glenberg AM, Kaschak MP, Rinck M (2004). On doing two things at once: temporal constraints on actions in language comprehension. *Memory and Cognition*, 32, 1033–43.

Dell GS (1986). A spreading-activation theory of retreival in sentence production. *Psychological Review*, 93, 283–321.

Devlin JT, Russell RP, Davis MH *et al.* (2002). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia*, 40, 54–75.

Elbert T, Pantev C, Wienbruch C, Rockstroh B, Taub E (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, 270, 305–7.

Ellis AW, Young AW (1988). *Human cognitive neuropsychology*. Hove, UK: Erlbaum.

Endrass T, Mohr B, Pulvermüller F (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience*, 19, 1653–60.

Eulitz C, Eulitz H, Maess B, Cohen R, Pantev C, & Elbert, T (2000). Magnetic brain activity evoked and induced by visually presented words and nonverbal stimuli. *Psychophysiology, 37*(4), 447-455.

Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15, 399–402.

Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005). Parietal lobe: from action organization to intention understanding. *Science*, 308, 662–7.

Fowler CA (1986). An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics*, 14, 3–28.

Frangos J, Ritter W, Friedman D (2005). Brain potentials to sexually suggestive whistles show meaning modulates the mismatch negativity. *Neuroreport*, 16, 1313–17.

Frege G (1966). Der Gedanke (In German; first published 1918–1920). In G Patzig, Ed. *Logische Untersuchungen* (pp. 30–53). Göttingen: Huber.

Freud S (1891). *Zur Auffassung der Aphasien* [In German]. Leipzig: Franz Deuticke.

Fry DB (1966). The development of the phonological system in the normal and deaf child. In F Smith, GA Miller, Eds. *The Genesis of Language* (pp. 187–206). Cambridge, MA: MIT Press.

Fuster JM (1995). *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. Cambridge, MA: MIT Press.

Fuster JM (2003). *Cortex and Mind: Unifying Cognition*. Oxford: Oxford University Press.

Gallese V, Lakoff G (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22, 455–79.

Garagnani M, Wennekers T, Pulvermüller F (2007). A neuronal model of the language cortex. *Neurocomputing*, 70, 1914–19.

Gentilucci M, Benuzzi F, Bertolani L, Daprati E, Gangitano M (2000). Language and motor control. *Experimental Brain Research*, 133, 468–90.

Glenberg AM, Kaschak MP (2002). Grounding language in action. *Psychonomic Bulletin and Review*, 9, 558–65.

González J, Barros-Loscertales A, Pulvermüller F *et al.* (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, 32, 906–12.

Graziano MS, Taylor CS, Moore T (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron* 34, 841–51.

Gruber T, Trujillo-Barreto NJ, Giabbiconi CM, Valdes-Sosa PA, Muller MM (2006). Brain electrical tomography (BET) analysis of induced gamma band responses during a simple object recognition task. *NeuroImage*, 29, 888–900.

Hauk O, Davis MH, Ford M, Pulvermüller F, Marslen-Wilson WD (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, 30, 1383–400.

Hauk O, Johnsrude I, Pulvermüller F (2004). Somatotopic representation of action words in the motor and premotor cortex. *Neuron*, 41, 301–7.

Hauk O, Pulvermüller F (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, 21, 191–201.

Hauk O, Shtyrov Y, Pulvermüller F (2006). The sound of actions as reflected by mismatch negativity: rapid activation of cortical sensory-motor networks by sounds associated with finger and tongue movements. *European Journal of Neuroscience*, 23, 811–21.

He SQ, Dum RP, Strick PL (1993). Topographic organization of corticospinal projections from the frontal lobe: motor areas on the lateral surface of the hemisphere. *Journal of Neuroscience*, 13, 952–80.

Hebb DO (1949). *The Organization of Behaviour. A Neuropsychological Theory*. New York, NY: John Wiley.

Hickok G, Poeppel D (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.

Holcomb PJ, Neville HJ (1990). Auditory and visual semantic priming in lexical decision: a comparision using event-related brain potentials. *Language and Cognitive Processes*, 5, 281–312.

Horwitz B, Braun AR (2004). Brain network interactions in auditory, visual and linguistic processing. *Brain and Language*, 89, 377–84.

Hubel D (1995). *Eye, Brain, and Vision* (Second edition). New York, NY: Scientific American Library.

Humphreys GW, Forde EM (2001). Hierarchies, similarity, and interactivity in object recognition: 'category-specific' neuropsychological deficits. *Behavioural and Brain Sciences*, 24, 453–509.

Humphreys GW, Riddoch MJ (1987). On telling your fruit from your vegetables – a consideration of category-specific deficits after brain-damage. *Trends in Neurosciences*, 10, 145–8.

Ihara A, Kakigi R (2006). Oscillatory activity in the occipitotemporal area related to the visual perception of letters of a first/second language and pseudoletters. *NeuroImage*, 29, 789–96.

Jeannerod M (2001). Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage*, 14, S103–9.

Jeannerod M, Arbib MA, Rizzolatti G, Sakata H (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in Neuroscience*, 18, 314–20.

Jeannerod M, Frak V (1999). Mental imaging of motor activity in humans. *Current Opinion in Neurobiology*, 9, 735–9.

Kiefer M (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Memory and Cognition*, 29, 100–16.

Kintsch W (1974). *The Representation of Meaning in Memory*. Hillsdale, NJ: Erlbaum.

Kintsch W (1998). *Comprehension: A Paradigm for Cognition*. New York, NY: Cambridge University Press.

Kintsch W (2002). The potential of latent semantic analysis for machine grading of clinical case summaries. *Journal of Biomedical Informatics*, 35, 3–7.

Kleene SC (1956). Representation of events in nerve nets and finite automata. In CE Shannon, J McCarthy, Eds. *Automata Studies* (pp. 3–41). Princeton, NJ: Princeton University Press.

Knoblauch A, Markert H, Palm G (2005). An associative cortical model of language understanding and action planning. In J Mira, JR Alvarez, Eds. *International Work-conference on the Interplay between Natural and Artificial Computation 2005* (vol. 3562, pp. 405–14). Berlin: Springer.

Korpilahti P, Krause CM, Holopainen I, Lang AH (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, 76, 332–9.

Krause CM, Korpilahti P, Porn B, Jantti J, Lang HA (1998). Automatic auditory word perception as measured by 40 Hz EEG responses. *Electroencephalography and Clinical Neurophysiology*, 107, 84–7.

Kujala A, Alho K, Valle S *et al.* (2002). Context modulates processing of speech sounds in the right auditory cortex of human subjects. *Neuroscience Letters*, 331, 91–4.

Lakoff G (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind.* Chicago, IL: University of Chicago Press.

Lakoff G, Johnson M (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought.* New York, NY: Basic Books.

Landauer TK, Dumais ST (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–40.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967). Perception of the speech code. *Psychological Review*, 74, 431–61.

Lichtheim L (1885). On aphasia. *Brain*, 7, 433–84.

Lutzenberger W, Pulvermüller F, Birbaumer N (1994). Words and pseudowords elicit distinct patterns of 30-Hz activity in humans. *Neuroscience Letters*, 176, 115–8.

Lutzenberger W, Pulvermüller F, Elbert T, Birbaumer N (1995). Local 40-Hz activity in human cortex induced by visual stimulation. *Neuroscience Letters*, 183, 39–42.

Mahon, B.Z., and Caramazza, D. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. J Physiol. Paris, 102 (1–3), 59–70.

Makris N, Meyer JW, Bates JF, Yeterian EH, Kennedy DN, Caviness VS (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.

Martin A, Chao LL (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, 11, 194–201.

Martin A, Haxby JV, Lalonde FM, Wiggs CL, Ungerleider LG (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, 270, 102–5.

Matelli M, Camarda R, Glickstein M, Rizzolatti G (1986). Afferent and efferent projections of the inferior area 6 in the macaque monkey. *Journal of Comparative Neurology*, 251, 281–98.

McCulloch WS, Pitts WH (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–33.

Moscoso Del Prado Martin F, Hauk O, Pulvermüller F (2006). Category specificity in the processing of color-related and form-related words: an ERP study. *NeuroImage*, 29, 29–37.

Müller MM, Bosch J, Elbert T *et al.* (1996). Visually induced gamma-band responses in human electroencephalographic activity: a link to animal studies. *Experimental Brain Research*, 112, 96–102.

Näätänen R (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1–21.

Näätänen R, Lehtokoski A, Lennes M *et al.* (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432–4.

Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, 24, 283–8.

Neininger B, Pulvermüller F (2001). The right hemisphere's role in action word processing: a double case study. *Neurocase*, 7, 303–17.

Neininger B, Pulvermüller F (2003). Word-category specific deficits after lesions in the right hemisphere. *Neuropsychologia*, 41, 53–70.

Obleser J, Lahiri A, Eulitz C (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *NeuroImage*, 20, 1839–47.

Oliveri, M., Finocchiarro, C., Shapiro, K., Gangitano, M., Canramazza, A., and Pascual-Leone, A. (2004). All talk and new action: a transcranial magnetic stimulation study of motor cortex activation during action word production. J Cogn. Neurosci., 16(3), 374–381.

Page M (2000). Connectionist modelling in psychology: a localist manifesto. *Behavioural and Brain Sciences*, 23, 443–67.

Palva S, Palva JM, Shtyrov Y *et al.* (2002). Distinct gamma-band evoked responses to speech and non-speech sounds in humans. *J Neurosci*, 22, RC211.

Pandya DN, Yeterian EH (1985). Architecture and connections of cortical association areas. In A Peters, EG Jones, Eds. *Cerebral Cortex, Volume 4: Association and Auditory Cortices* (pp. 3–61). London: Plenum Press.

Patterson K, Hodges JR (2001). Semantic dementia. In RF Thompson, JL McClelland, Eds. *International Encyclopaedia of the Social and Behavioural Sciences. Behavioural and Cognitive Neuroscience Section* (pp. 3401–5). New York, NY: Pergamon Press.

Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *European Journal of Neuroscience*, 8, 2236–46.

Penfield W, Boldrey E (1937). Somatic sensory and motor representation in the cerebral cortex as studied by electrical stimulation. *Brain*, 60, 389–443.

Penfield W, Rasmussen T (1950). *The Cerebral Cortex of Man*. New York, NY: Macmillan.

Pettigrew CM, Murdoch BE, Ponton CW *et al.* (2004). Automatic auditory processing of english words as indexed by the mismatch negativity, using a multiple deviant paradigm. *Ear and Hearing*, 25, 284–301.

Plenz D, Thiagarajan TC (2007). The organizing principles of neuronal avalanches: cell assemblies in the cortex? *Trends in Neurosciences*, 30, 101–10.

Posner MI, Pavese A (1998). Anatomy of word and sentence meaning. *Proceedings of the National Academy of Sciences USA*, 95, 899–905.

Preissl H, Pulvermüller F, Lutzenberger W, Birbaumer N (1995). Evoked potentials distinguish nouns from verbs. *Neuroscience Letters*, 197, 81–3.

Price CJ (2000). The anatomy of language: contributions from functional neuroimaging. *Journal of Anatomy*, 197, 335–59.

Pulvermüller F (1992). Constituents of a neurological theory of language. *Concepts in Neuroscience*, 3, 157–200.

Pulvermüller F (1996). Hebb's concept of cell assemblies and the psychophysiology of word processing. *Psychophysiology*, 33, 317–33.

Pulvermüller F (1999). Words in the brain's language. *Behavioural and Brain Sciences*, 22, 253–336.

Pulvermüller F (2000). Cell assemblies, axonal conduction times, and the interpretation of high-frequency dynamics in the EEG and MEG. In R Miller, Ed. *Time and the Brain* (pp. 241–9). Chur: Harwood Academic Publishers.

Pulvermüller F (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5, 517–24.

Pulvermüller F (2002). A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. *Progress in Neurobiology*, 67, 85–111.

Pulvermüller F (2003). *The neuroscience of language*. Cambridge: Cambridge University Press.

Pulvermüller F (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6, 576–82.

Pulvermüller F, Birbaumer N, Lutzenberger W, Mohr B (1997). High-frequency brain activity: its possible role in attention, perception and language processing. *Progress in Neurobiology*, 52, 427–45.

Pulvermüller F, Eulitz C, Pantev C *et al.* (1996). High-frequency cortical responses reflect lexical processing: an MEG study. *Electroencephalography and Clinical Neurophysiology*, 98, 76–85.

Pulvermüller F, Härle M, Hummel F (2000). Neurophysiological distinction of verb categories. *NeuroReport*, 11, 2789–93.

Pulvermüller F, Hauk O (2006). Category-specific processing of color and form words in left fronto-temporal cortex. *Cerebral Cortex*, 16, 1193–201.

Pulvermüller F, Hauk O, Nikulin VV, Ilmoniemi RJ (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, 21, 793–7.

Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences USA*, 103, 7865–70.

Pulvermüller F, Kujala T, Shtyrov Y *et al.* (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage*, 14, 607–16.

Pulvermüller F, Lutzenberger W, Preissl H (1999). Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex*, 9, 498–508.

Pulvermüller F, Mohr B (1996). The concept of transcortical cell assemblies: a key to the understanding of cortical lateralization and interhemispheric interaction. *Neuroscience and Biobehavioural Reviews*, 20, 557–66.

Pulvermüller F, Mohr B, Schleichert H (1999). Semantic or lexico-syntactic factors: What determines word-class specific activity in the human brain? *Neuroscience Letters*, 275, 81–4.

Pulvermüller F, Preissl H (1991). A cell assembly model of language. *Network Computation in Neural Systems*, 2, 455–68.

Pulvermüller F, Preissl H, Lutzenberger W, Birbaumer N (1995). Spectral responses in the gamma-band: physiological signs of higher cognitive processes? *NeuroReport*, 6, 2057–64.

Pulvermüller F, Shtyrov Y, Ilmoniemi RJ (2003). Spatio-temporal patterns of neural language processing: an MEG study using Minimum-Norm Current Estimates. *NeuroImage*, 20, 1020–5.

Pulvermüller F, Shtyrov Y, Ilmoniemi RJ (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, 17, 884–92.

Pulvermüller F, Shtyrov Y, Kujala T, Näätänen R (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology*, 41, 106–12.

Rizzolatti G, Craighero L (2004). The mirror-neuron system. *Annual Review in Neuroscience*, 27, 169–92.

Rizzolatti G, Luppino G (2001). The cortical motor system. *Neuron*, 31, 889–901.

Roelofs A (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–42.

Rogers TT, Lambon-Ralph MA, Garrard P *et al.* (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review*, 111, 205–35.

Roy D (2005). Grounding words in perception and action: computational insights. *Trends in Cognitive Science*, 9, 389–96.

Schnelle H (1996). Approaches to computational brain theories of language – a review of recent proposals. *Theoretical Linguistics*, 22, 49–104.

Scott SK, Johnsrude IS (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100–7.

Searle JR (1990). Minds, brains, and programs. *Behavioural and Brain Sciences*, 3, 417–57.

Seidenberg MS, Plaut DC, Petersen AS, McClelland JL, McRae K (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1177–96.

Sereno SC, Rayner K (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7, 489–493.

Sereno SC, Rayner K, Posner MI (1998). Establishing a time line for word recognition: evidence from eye movements and event-related potentials. *NeuroReport*, 13, 2195–200.

Shastri L, Grannes D, Narayana S, Feldman J (2005). A connectionist encoding of parameterized schemas and reactive plans. In GK Kraetzschmar, G Palm, Eds. *Hybrid Information Processing in Adaptive Autonomous Vehicles*. Berlin: Springer.

Shtyrov Y, Hauk O, Pulvermüller F (2004). Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. *European Journal of Neuroscience*, 19, 1083–92.

Shtyrov Y, Pihko E, Pulvermüller F (2005). Determinants of dominance: is language laterality explained by physical or linguistic features of speech? *NeuroImage*, 27, 37–47.

Shtyrov Y, Pulvermüller F (2002). Neurophysiological evidence of memory traces for words in the human brain. *NeuroReport*, 13, 521–5.

Simmons WK, Ramjee V, Beauchamp MS, McRae K, Martin A, Barsalou LW (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, 45, 2802–10.

Singer W, Gray CM (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review in Neuroscience*, 18, 555–86.

Sittiprapaporn W, Chindaduangratn C, Tervaniemi M, Khotchabhakdi N (2003). Preattentive processing of lexical tone perception by the human brain as indexed by the mismatch negativity paradigm. *Annals of the New York Academy of Sciences*, 999, 199–203.

Skrandies W (1999). Early effects of semantic meaning on electrical brain activity. *Behavioural and Brain Sciences*, 22, 301.

Tallon-Baudry C, Bertrand O (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3, 151–61.

Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in humans. *Journal of Neuroscience*, 16, 4240–9.

Tallon-Baudry C, Bertrand O, Peronnet F, Pernier J (1998). Induced gamma-band activity during the delay of visual short-term memory tasks in humans. *Journal of Neuroscience*, 18, 4244–54.

Tettamanti M, Buccino G, Saccuman MC *et al.* (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17, 273–81.

Tomasello M, Kruger AC (1992). Joint attention on actions: acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, 19, 311–33.

Tsumoto T (1992). Long-term potentiation and long-term depression in the neocortex. *Progress in Neurobiology*, 39, 209–28.

Tyler LK, Moss HE (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, 5, 244–52.

Tyler LK, Moss HE, Durrant-Peatfield MR, Levy JP (2000). Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. *Brain and Language*, 75, 195–231.

Tyler LK, Russell R, Fadili J, Moss HE (2001). The neural representation of nouns and verbs: PET studies. *Brain*, 124, 1619–34.

Tyler LK, Stamatakis EA, Bright P *et al.* (2004). Processing objects at different levels of specificity. *Journal of Cognitive Neuroscience*, 16, 351–62.

Varela FJ, Thompson E, Rosch E (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Boston, MA: MIT Press.

von der Malsburg C, Schneider W (1986). A neural cocktail-party processor. *Biological Cybernetics*, 54, 29–40.

Warrington EK, McCarthy RA (1983). Category specific access dysphasia. *Brain*, 106, 859–78.

Warrington EK, Shallice T (1984). Category specific semantic impairments. *Brain*, 107, 829–54.

Watkins K, Paus T (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *Journal of Cognitive Neuroscience*, 16, 978–87.

Wennekers T, Garagnani M, Pulvermüller F (2006). Language models based on Hebbian cell assemblies. *Journal of Physiology–Paris*, 100, 16–30.

Wermter S, Weber C, Elshaw M, Gallese V, Pulvermüller F (2005). Neural grounding of robot language in action. In S Wermter, G Palm, M Elshaw, Eds. *Biomimetic Neural Learning for Intelligent Robots* (pp. 162–81). Berlin: Springer.

Wermter S, Weber C, Elshaw M, Panchev C, Erwin H, Pulvermüller F (2004). Towards multimodal neural network robot learning. *Robotics and Autonomous Systems*, 47, 171–5.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7, 701–2.

Wittgenstein L (1953). *Philosophical Investigations*. Oxford: Blackwell.

Young MP, Scannell JW, Burns G, Blakemore C (1994). Analysis of connectivity: neural systems in the cerebral cortex. *Review in Neuroscience*, 5, 227–49.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992). Lateralization of phonetic and pitch discremination in speech processing. *Science*, 256, 846–9.

Chapter 7

Symbols and embodiment from the perspective of a neural modeller

Andreas Knoblauch

7.1 Introduction and definitions

This paper contributes to the current debate about symbols and embodiment by pointing out the perspective of a neural modeller. I illustrate the default definitions of 'symbol', 'embodiment', 'meaning', and 'grounding' in the context of detailed neural network models, i.e., on a level more detailed than common connectionist approaches. My arguments are based on Hebbian *neuronal or cell assemblies* (Hebb 1949; Braitenberg 1978; Palm 1982, 1990) and detailed models of the cortical microcircuitry. These models have been employed to implement a large-scale cortical architecture to enable a robot to perform simple tasks such as understanding and reacting to simple spoken commands. More generally, I finally discuss the relations between embodiment, grounding, anchoring, binding, and the invariant recognition in distributed hierarchical systems.

7.1.1 Symbols

For a neural network modeller, one simple possible way to discern symbols from nonsymbols is to look at the inner structure of the representational units. Subsymbols have an inner structure which can be used to define a similarity metric relevant for the represented entity. In contrast, symbols have no relevant inner structure (i.e., symbols are abstract and arbitrary). For example, in simple object recognition systems, a nonsymbol or subsymbol may be a vector of sensory features, while a symbol may correspond to a single node representing an object category. These definitions are sufficient for a low-level (e.g., neural) description of a cognitive subsystem (e.g., for object recognition), but may not be adequate for the current debate which is about language and the representation of meaning. Here, the discussion includes higher-level symbols employed by a cognitive system that is able to think, to reason, and to manipulate these symbols in a flexible way.

According to Glenberg *et al.*'s default definition (Chapter 1) such a symbol is a 'theoretical element that is arbitrary, abstract, and amodal'. Before we proceed by discussing and adapting that definition, it may be useful to be aware of the different contexts in which we will use the word 'symbol'. The situation is illustrated in Figure 7.1. We live in a physical world W where systems or subjects S are part of that world and interact with the world. Some of the systems (namely we, the subjects) are somehow able to generate

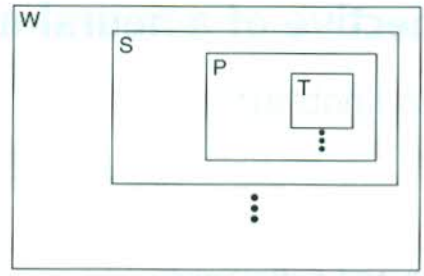

Fig. 7.1 Different modelling levels. We live in a physical world *W*; systems (or subjects) *S* are part of that world and interact with the world. The subjects are somehow able to generate a (usually unique) psychological or phenomenological space *P*, which we can use, for example, to generate theories *T* about all kind of issues on all levels *W*, *S*, *P*, and *T*.

a usually unique psychological or phenomenological space *P*, which we can employ, for example, to generate ideas or theories *T* about all kinds of issues on all levels *W*, *S*, *P*, and *T*. In particular, we can make theories about *S* (predominantly done by biology, neuroscience, and artificial intelligence [AI]), *P* (psychology), or *T* (metamathematics or logics). Ideas or theories *T* essentially consist of a set of symbols (as defined above) and additional rules determining how the symbols can be 'manipulated'.

Since symbols are part of theories or ideas this implies two aspects of a symbol corresponding to levels S and T. As a neural modeller one is predominantly interested in the *S*-level implementation of *P in W*, i.e., in reducing the observable psychological and behavioural phenomena as much as possible to detailed neural and synaptic processes, and finally to the physical laws. (The underlying preliminary naturalistic working hypothesis assumes that this goal is actually possible.) But since models about *S* are ultimately also *T* theories, the neural modeller (and any other kind of *S*-modeller like many an AI researcher) has to discern between the two kinds of symbols within his theory: symbols to model the *T*-symbols of *S*, and symbols to model the implementation of the *T*-symbols of *S*. Thus, we will refer to these two kind of symbols as *T*-symbols and *S*-symbols, respectively. For example, for a cognitive system *S* capable of understanding language, a *T*-symbol is the representation of a word, while *S*-symbols are finer-grained entities used for implementing the word representation. An *S*-symbol could be a node in a connectionist network (or, alternatively, a state or band variable in a Turing machine) while a *T*-symbol could be implemented by a set of distributed *S*-symbols (and possibly further dynamic processes).

The current debate is about the question whether cognitive systems (such as we) are or have to be either symbolic or 'embodied'. Of course, any kind of cognitive system must be symbolic in the trivial sense that we have symbolic language and theories *T*. Thus, any cognitive system must be *T*-symbolic. Correspondingly, the critical question is not about the reality of *T*-symbols, but about the way *T*-symbols are implemented in *S*-symbols. Note that, depending on how we define 'symbol' and 'embodiment', it might be thinkable that *T*-symbols could be implemented by non-symbolic processes at the *S* level (although this seems counterintuitive since we actually aim to develop a symbolic theory *T* about *S*).

7.1.2 Embodiment

Embodiment comes in several different flavours (cf. Wilson 2002). The strongest claim would be that embodiment could extend the qualitative, or at least quantitative, computational capabilities of a system *S* by exploiting the properties of *W*, e.g., described by the physical laws. If true, this would essentially negate the Church–Turing thesis that symbolic Turing machines can compute any 'naturally' (or physically) computable function. For example, 'embodied' analogue computers might be better in simulating physical systems, or in computing real numbers with infinite precision (see also Brooks 1990). Nevertheless, symbolic Turing computers can approximately simulate any physical system by numerically solving the differential equations of physics, although this can take considerable time and, for chaotic systems, infinite computing precision. Another example could be computers exploiting the quantum properties of the physical world. It has been shown that such quantum computers, if physically possible, could compute certain functions much faster than conventional computers or Turing machines (DiVincenzo 1995).

A less strong idea of embodiment is the dichotomy of embodied versus symbolic cognitive systems addressed by the current debate which attempts to classify cognitive systems according to the interface between the system *S* itself and the external world *W*. Obviously, any cognitive system *S* that deserves that name will have to interact with its environment (percept and act) and is therefore embodied in a trivial sense. Similarly, any cognitive system *S* must be symbolic in a trivial sense since it must explain our capabilities to use language and think in symbols (e.g., to develop theories *T* within our psychological space *P*). Thus, in this trivial sense any model of a cognitive system will be both embodied and symbolic.

Obviously, any cognitive system can be divided into sensors, actors, and internal machinery, such that the interaction with the environment is accomplished only via the sensors and actors. Strictly speaking such an interactive system cannot be adequately modelled by a Turing machine. The original 'autistic' Turing machine has been suggested as a model for computation only. That scenario assumes separate phases for: (1) providing the input from the environment to the Turing machine's band, (2) doing the computation independently of the environment, possibly for a very long time, and (3) returning the output of the computation from the band to the environment. In contrast, we rather have to think of an 'interactive' Turing machine that depends on the environment and can influence the environment at any time. Thus, it would be possible to define embodiment by the degree of interaction with the environment.

Our default definition of an 'embodied' system goes in a similar direction by demanding that the meaning of a symbol must depend on activity in systems also used for perception, action, and that emotion and reasoning must require the use of those systems (see Chapter 1). This form of weak embodiment is stronger than the trivial version of embodiment, but addresses only the high-level structure of the internal machinery, e.g., in Marr's (1982) terms, the algorithmic or computational levels but not the implementation level. As a consequence, any such 'embodied' system can be translated into a purely (*S*-)'symbolic' system (e.g., a computer program or Turing machine) with the

same sensor/actor interface and vice versa. This is true according to the Church–Turing thesis, at least as long as our 'embodied' system does not exploit the physical world in a super-Turing manner as described before. We can also conclude that this form of embodiment will probably be neutral to such questions as whether 'ideas are the sole province of biological systems' (as discussed in Chapter 1). And, of course, the property of embodiment will be a gradual property. Nevertheless, the idea of embodiment might still prove useful, e.g., in building more efficient artificial cognitive systems, or in guiding the analysis of the brain.

7.1.3 Meaning

Our last definition of embodiment refers to the term 'meaning', which may require an explicit definition. A simple definition states that meaning is the 'content' of a sign or symbol. Here, 'content' refers to all the parts of an information processing systems theories T that have a relation to the symbol. For example, the meaning of the word symbol 'car' may include prototypical 'icons' of cars, knowledge about the corresponding *consists-of* and *is-a* ontologies, knowledge about actions that can be done with, to, or by a car, and episodic knowledge about particular experiences with cars. This very general meaning of a symbol must usually be strongly constrained by context.

An alternative, more behaviouristic, definition is this: the meaning of a symbol or sign to a cognitive system is the sum of the potential or actual behavioural changes after receiving the sign. For example, the meaning of the traffic sign 'stop' to a car driver is a propensity to apply the brakes to stop the car. Or the meaning of signs indicating bad politics to a voter is a propensity to no longer give a vote to the responsible politician in future elections.

The second definition has several advantages: first, it does not refer to the internals of the system which may be difficult to observe and interpret, e.g., in animal experiments. Instead, it solely refers to the system's or agent's observable behaviour or actions. Further, this direct reference to actions and agents is more relevant for our current debate about meaning and embodiment. Defining meaning in this way by particular actions then obviously implies (by definition) that establishing meaning is closely related to or even 'depends on activity in action systems' which is essentially our definition of embodiment (see Chapter 1).

Indeed, neurobiological experimenters investigating the brains of animals by correlating behaviour and neural activity actually have to define categories (which can then be symbolized) by actions. For example, experimenters can find out which one of two possible interpretations of an ambiguous figure a monkey is perceiving by training the monkey to move the right hand for interpretation 1 and the left hand for interpretation 2 (which already defines the meaning of the figures for the monkey). Similarly, it has been proposed that we humans, as well as other animals, can learn to discriminate two classes or categories only if the discrimination is behaviourally relevant. This is most obvious for lower animals with only a very limited behavioural repertoire (and therefore only a limited way of developing abstract classes such as prey, predator, or mate, corresponding to actions like feeding, fighting, or mating).

In summary, we may already conclude here that understanding the meaning of signs indeed requires embodiment defined as regular activation of action- and perception-related subsystems. However, this conclusion is not sufficient for understanding how the brain works, or how to create intelligent artificial systems. This requires a detailed theory on how perception, action, and learning of abstract categories is accomplished by the brain. The rest of this chapter focuses on delivering building blocks for this detailed theory by having a closer look at neural models of symbols and meaning.

7.1.4 The easy and the hard problems

With this essay I do not intend to tackle what has been called the 'hard problems' in the psychology of consciousness, for example, explaining how physical processes in the brain give rise to subjective experience (Chalmers 1996; Jackendoff 1987). Although I believe that, from a third-person perspective, symbols, embodiment, and meaning as defined above are easy problems (i.e., nothing 'mystic'), and that we will probably soon be able to realize artificial systems that can be said to be embodied and represent meaning in a similar way to humans, I admit that assuming full identity between the first-person subjective processes of humans (including feelings) and those of digital computers (Dennett, 1996; Metzinger, 2004) may have problematic consequences.

For example, if we were to ascribe subjective feelings to a robot then we would also have to ascribe feelings to the robot rid of its sensorimotor interface (just as we ascribe feelings to quadriplegia and locked-in syndrome patients). Then we ultimately would have to ascribe feelings to a digital computer, i.e., finite-state-machine (FSM). Or, to put it more exactly, we would ascribe feelings to particular states of the FSM. Adapting Searle's Chinese room argument (Searle, 1980) to feelings (instead of meaning), this seems strange because the FSM's states are arbitrary, i.e., it is unclear why one particular physical state should be associated with pain (and not with joy or something else).

One could answer that meaning and feelings are not associated with a single state but with particular *recurring* state sequences (for example, realizing a kind of monitoring structure). However, this will probably not help us since we can always construct an 'equivalent' FSM where such a sequence corresponds to a single state again (although this will require a large FSM with many states). We may accept this for the case of meaning defined in terms of 'potential behaviour' (see above) because the question 'Is that machine understanding this?' can be resolved, in principle, by looking at the FSM's past or future states. However, we are usually more reluctant in the case of feelings because the question 'Is that machine feeling pain?' must be answered in the present (and proposing that past or future states would make a difference contradicts the state concept of classical physics).

7.2 A neural modeller's perspective

When words referring to actions or visual scenes are presented to humans, distributed neural networks, including areas of the motor and visual systems of the cortex, become active (e.g., Pulvermüller 1999, 2003). The brain correlates of words and their referent actions and objects appear to be strongly coupled neuron ensembles in defined

cortical areas. The theory of cell assemblies (Hebb 1949; Braitenberg 1978; Palm 1982, 1990) provides one of the most promising frameworks for modelling and understanding the brain in terms of distributed neuronal activity. It is suggested that entities of the outside world (and also internal states) are coded in groups of neurons rather than in single 'grandmother' cells, and that a cell assembly is generated by Hebbian coincidence or correlation learning where the synaptic connections are strengthened between co-activated neurons.

7.2.1 Local cell assemblies, associative memory, and neural S-symbols

The notion of neuronal assemblies as strongly coupled neurons leads to the concept of *neural (auto-)associative memory* (Willshaw *et al.* 1969; Palm 1980; Hopfield 1982). One simple model of neural associative memory has been proposed by Steinbuch and Willshaw (Willshaw *et al.* 1969; Steinbuch 1961; Palm 1980; Knoblauch 2005) consisting of McCulloch–Pitts-type threshold units and recurrent binary synapses (Figure 7.2). Here, the activity pattern of the cell population can be described by a binary vector, and we identify stored activity patterns with the cell assemblies. After learning a number of

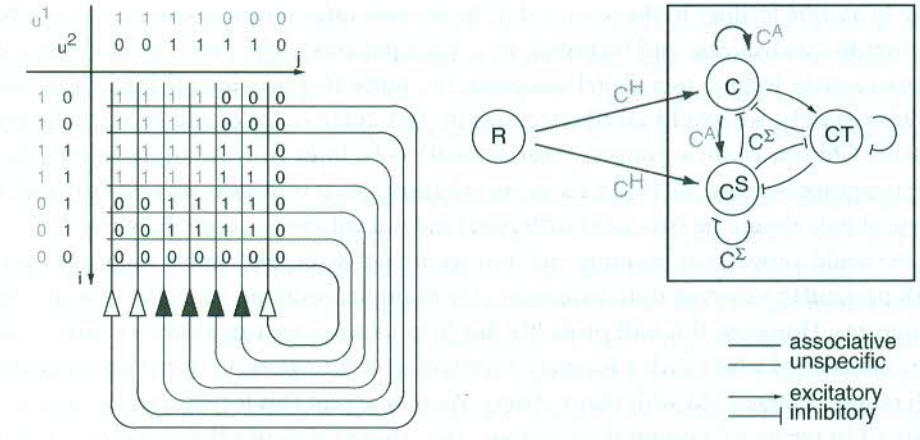

Fig. 7.2 Left: Neural (auto-)associative memory where two cell assemblies of size $k = 4$ have been stored (corresponding to the activation patterns u1 and u2) in the 'memory matrix' of synaptic connections. Filled units indicate the active neurons of pattern u2. Right: A more realistic implementation of associative memory, modelling a small patch (about 1 mm^3) of cortical tissue (Knoblauch and Palm, 2001). The model comprises several populations of excitatory and inhibitory spiking neurons of the integrate-and-fire type. Here, C is the main excitatory population of pyramidal cells receiving input from another cortical patch, R. C^S and C^T are inhibitory interneuron populations controlling local excitation. Each neuron is modelled as a leaky integrator with excitatory and inhibitory conductances, where a spike is emitted as soon as the dendritic potential exceeds a threshold. The memory matrices are employed in several afferent and recurrent synaptic connections (cH, cA) to and from populations R, C, and C^S. The remaining inhibitory feedback connections are unspecific (i.e., independent of the learned activity patterns).

neuronal assemblies, the network can be described by a connection matrix A corresponding to a graph where the nodes correspond to the neurons and neuronal assemblies correspond to k-cliques of neurons (a k-clique is a subset of size k consisting of completely connected neurons).

Hetero-association works similar to auto-association except that the 'memory matrix' describes the synaptic connections between two different neuron populations. Hetero-associative connections can map neuronal assemblies of the first population (or sets or parts of them) to assemblies of the second population (or sets or parts of them).

The virtue of the binary model is that it is easy to understand, analyse, and implement, but the main results apply also to more realistic gradual and spiking models (Hopfield, 1984; Knoblauch and Palm, 2001). Neural associative memories have a couple of nice features. They achieve pattern completion, i.e., a neuronal assembly can be activated not only by the very same inputs that have been used for learning, but also by modified patterns that are 'sufficiently similar' to the original address pattern. For example, assembly u2 in Figure 7.2 will already be activated by addressing an arbitrary subset of size ≥ 3.

It can be shown that the number of storable patterns is almost proportional to the number of synapses if the patterns are sparse and have random character (i.e., a population of n neurons can store almost n^2 sparse cell assemblies with $k \ll n$). Access time is essentially independent of the number of stored patterns. The overlaps of different neuronal assemblies can be used to express the similarities of the represented entities. Neuronal assemblies thereby provide a very natural associative way of grounding new representations in the sensory inputs by means of bidirectional associative connections (cf. Barsalou 2003).

Associative memories have been used to model small volumes of cortical tissue (e.g., 1 mm^3) corresponding to a macrocolumn or the range where dense local recurrent connections between any cell pairs are possible (Braitenberg and Schüz, 1991). A step towards biological realism is to replace the single McCulloch–Pitts population by more realistic spiking neuron models and to incorporate known properties of cortical circuits (Figure 7.2). These models can extend the computational abilities of the standard model, for example, by making use of spike timings according to a latency code in that early spikes (relative to an external event or an underlying oscillation) are much more relevant than late spikes for activating an assembly (Knoblauch and Palm 2001; Knoblauch 2005).

Local cell assemblies can be seen as elementary neural (*S*-)symbols which can be 'allocated' or learnt to represent the inputs for further processing in downstream target populations. The symbolic character is most apparent if the assembly size is $k = 1$, corresponding to a localist code, or if the neurons that constitute a cell assembly are chosen at random, e.g., by noise. In the latter case the correlations (or overlaps) between two cells are minimal, which is required to store a maximal number of different activity patterns. Due to their singular or random character the neuronal assemblies could be said to be abstract and arbitrary, whereas the property of amodality depends on the location of the neuron population, e.g., a local population receiving visual inputs will develop visual perceptual symbols (cf. Barsalou 1999).

7.2.2 Global cell assemblies, language, and T-symbols

We have designed a large-scale brain model consisting of many interconnected cortical areas employing spiking associative memories. The model was implemented and tested on a robotic platform enabling the robot to understand and react to simple commands such as 'Bot show plum!' (Knoblauch *et al.* 2004). The language part of the model is illustrated by Figure 7.3 and the action part by Figure 7.4.

Each box in the figures corresponds to a spiking neural associative memory storing local cell assemblies, as described above. For illustration purposes, each area has been labelled according to the current activity pattern. (In general, a superposition of several stored local assemblies can be activated, e.g. to represent uncertainty or to represent new entities to be learned; here, the labels correspond to the neuronal assembly most similar to the current activation pattern). The resulting *global assembly*, for example, representing the *T*-symbol 'plum', stretches over many cortical areas (involving visual, auditory, action, and goal-related areas) and changes dynamically during the process of 'understanding' and reacting to the command. Thus, the global cell assembly as a whole works as a sign in Peirce's sense, i.e., as a mediator between the idea of a 'plum' and the real plum in the external world. The global assembly consists of parts, some of which can

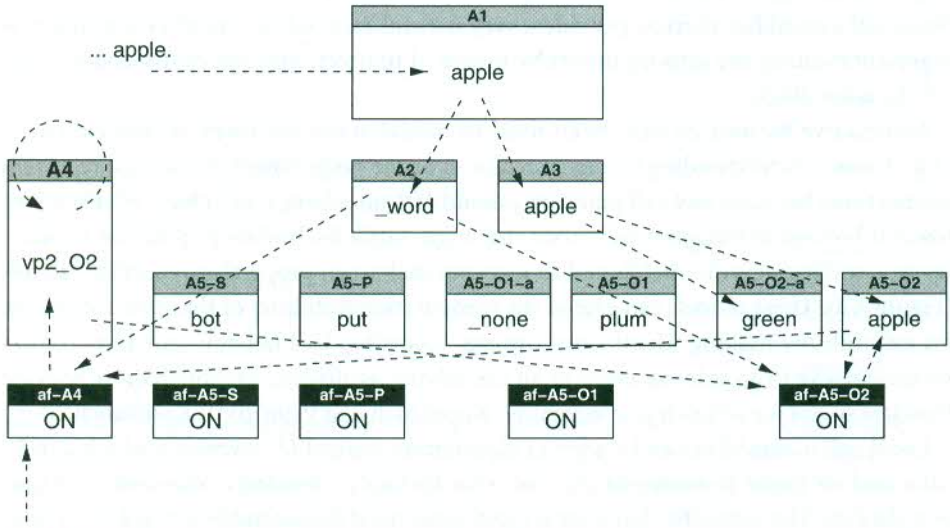

Fig. 7.3 The language part of an associative cortex model (see Knoblauch *et al.*, 2004) at the end of processing the sentence 'Bot put plum to green apple'. Each box corresponds to a cortical (or subcortical) area modelled as a neural associative memory. The meaning of the sentence is represented by distributed cell assemblies comprising 'slot areas' for different grammatical roles to implement elementary productivity and systematicity. Auditory input enters the central areas via areas A1 and A3 and is distributed across the grammatical slots according to a logic controlled by a grammatical sequence memory (A4) (where basic sentence types are stored) and subcortical 'activation fields' (small boxes). Arrows indicate recently activated synaptic connections.

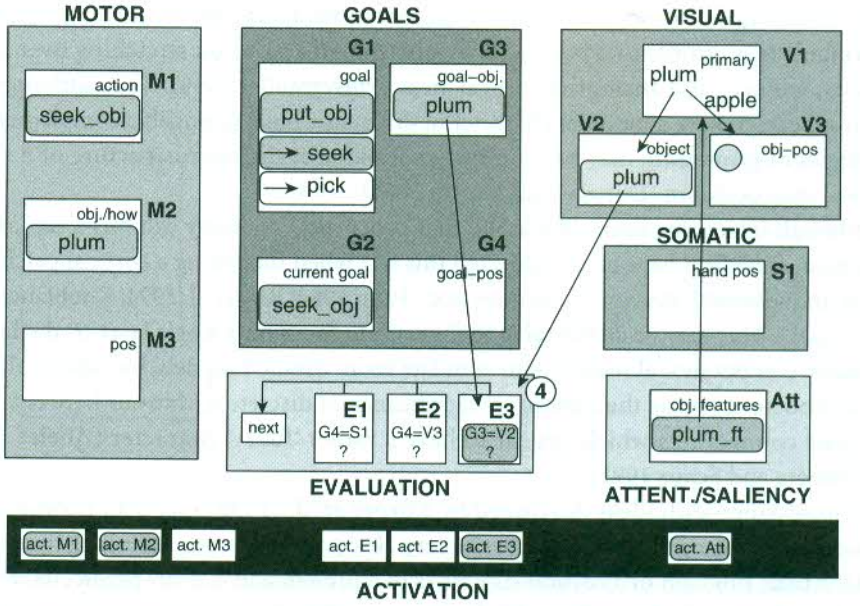

Fig. 7.4 Action part of an associative cortex model (see Knoblauch *et al.*, 2004) during performing the command 'Bot put plum to green apple'. The goal areas (G1–G4) received their inputs from the grammatical role areas (A5-S, A5-O1, etc., as illustrated in Figure 7.3) and divide the goals into a sequence of subgoals (i.e., seek plum, pick plum, move to the apple, drop plum). (High-level) motor areas receive inputs from the goal areas in order to perform the current subgoal. The completion of subgoals and switching to the next subgoal is controlled by 'evaluation fields' which check, for example, the consistency of visual perceptual activity patterns (e.g., in V4) with goal representations (e.g., in G3). At the shown system state the robot is about to finish the subgoal of seeking the plum.

be attributed as 'abstract' and 'amodal', e.g., the lexical representation of 'plum' in A3. But these symbolic parts are naturally grounded in the synaptically connected perceptual and action-related parts of the global assembly.

7.2.3 Cortical macrocolumns, prediction, and embodiment

Cognitive processes must be able to distinguish between different representational modi. For example, representational states may refer to present, future (or prediction), reality, wish, or signal (detailed and concrete), or symbol (abstract, amodal), perception, or action. Many cognitive architectures take a modular approach where these different representational modi are segregated into different cognitive subsystems or modules. (In general, an architecture can be said to be modular if it can be divided into subsystems such that there is much more communication between processes inside a subsystem than between processes of different subsystems.) For example, we could segregate a cognitive system into different modules for perceptions, actions, goals, memory, rule-based prediction systems, etc.

We have argued how global neuronal assemblies can implement and ground *T*-symbols (e.g., words of a language) by distributed activation stretching over many sensory, motoric, and associative cortical areas (Pulvermüller 1999). Thus, although the brain appears to have a modular character in that sense, hints to possible complementary strategies of grounding may be found when looking at the microstructure of a single cortical macrocolumn (Douglas and Martin 2004).

Although it has long been established that neocortical anatomy exhibits a six-layered structure, modellers have often neglected this fact when modelling a cortical patch by a single 'monolithical' neuron population (e.g., Palm 1982; Ritz *et al.* 1994; Knoblauch and Palm 2001). This may be attributable to the wish to focus on a single layer or the lack of adequate computational resources to simulate more detailed models, but also to doubting or underestimating the functional significance of discrete within- or between-layer synaptic connections which appear to have a rather 'fuzzy' character (Abeles 1991; Braitenberg and Schüz 1991).

In accordance with ideas developed by Körner *et al.* (1999) (see also Hawkins and Blakeslee 2004; Rao and Ballard 1999; Guillery 2003), we assume as a working hypothesis that the basic function of a cortical column is to represent and actively predict its sensory inputs. To achieve this in a self-organizing, autonomous way, it is necessary to have access to (at least some of) the different representational modi described above. We propose that different representational modi of the same entity are located in different layers within the same macrocolumn rather than monolithically in different columns or areas.

Figure 7.5 illustrates this functional model and our current implementation employing spiking associative networks similar to that discussed in Section 7.2.1. At each time the model must represent a state $v = (w, a)$ and use sensory input s to update the state v according to a function f. We found it meaningful to divide the state variable v into two independent entities: a variable w describing 'external' entities from the outside world and another variable a describing a local 'internal actor'. In addition to updating a state, the system should also be able to predict a future state w' without accessing sensory input. Note that the proposed circuitry provides the basic ingredients for simulating (or predicting over) larger time intervals.

By comparison with known anatomical facts we can match our functional model (Figure 7.5) with the layered organization of neocortex (Körner *et al.* 1999; Guillery 2003; Douglas and Martin 2004; Braitenberg and Schüz 1991; Felleman and Van Essen 1991). For example, it is well established that *feed-forward inputs* to a cortical column mainly target layer IV neurons, and that the feed-forward output to the next cortical stage leaves a cortical column via layer II/III neurons. In contrast to the feed-forward stream, *feed-back inputs* avoid layer IV and target mainly the upper and lower layers. Another remarkable feature of the cortical microcircuitry is that layer V pyramidal neurons, at any cortical site, project to subcortical regions closely related to action and behaviour (Guillery, 2003).

Based on these facts we believe that the forward recognition function f is located in the middle and upper layers, while the remaining functionality, related to behaviour and predictions, is located in the lower layers V and VI. Furthermore, we believe that the

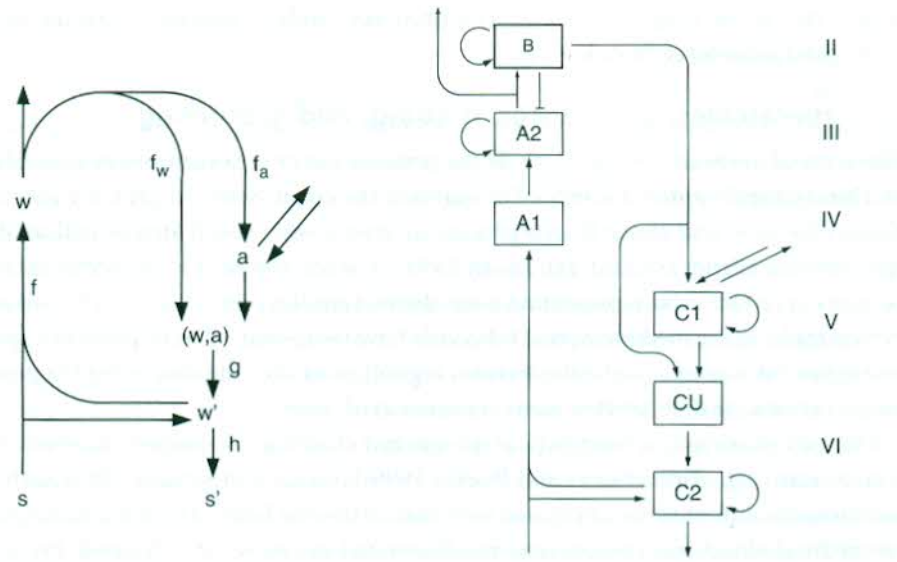

Fig. 7.5 Left: Basic functional circuit of a cortical column. Sensory input *s* is used to update the current world state *w*. This is used to choose an appropriate action *a*. World state and action can be used to predict the next world state *w'* and next sensory input *s'*. Right: Implementation of a cortical column by spiking associative memories (see Knoblauch *et al.* 2005, 2007; Körner *et al.* 1999). Sensory input activates signal-like representations (based on basis vectors) in the middle cortical layers (A1, A2), a symbol-like prototypical cell assembly in the upper layers (B), and finally action and prediction related representations in the lower layers (C1, C2).

recognition system of the middle and upper layers is split up into two subsystems, one for fast bottom-up recognition (the *A system*, layers IV and upper III) and another for refined recognition employing feedback (the *B system*, layers II/III). As an example we have implemented a model of several cortical and subcortical areas for learning saccadic object representations (including several visual cortical areas and the superior colliculus; see Knoblauch *et al.* 2005, 2007). In that particular case, the representational world states are object views (e.g., the retinal image when fixating on a particular key feature of a visual object) and the actions correspond to saccades.

What does this has to do with embodiment and grounding? Our model suggests that actions are grounded in perceptions already at the level of a single cortical macrocolumn (cf. Hawkins 2004; Körner *et al.* 1999; Guillery 2003; Young 1993). Thus, recognition of incoming signals will automatically induce an action-related process in the same macrocolumn and related subcortical structures. In our model, the perceptual representation *w* is on the one hand really symbolic in cortical layers II/III, is grounded in sensory input and a signal-like (basis vector-based) representation in layer IV, and is the result of a prediction $(w^{t-1}, a) \rightarrow w$ (layer VI) such that it is also grounded in action (layer V). Vice versa, for the same reason the produced action will be grounded in perception. Furthermore, the proposed circuitry provides the basic ingredients for simulating (or predicting) the represented state (w) over larger time intervals (at a microlevel),

which has been suggested to be required for understanding meaning on the (*P*,*T*-)macro-levels (cf. Barsalou 1999).

7.2.4 Hierarchies, invariances, binding, and grounding

The areas of the brain and particularly the cerebral cortex are organized in a distributed and hierarchically ordered manner. For example, the visual system of primates consists of about fifty areas and about fifteen processing stages, where each area is dedicated to a specific task (Felleman and Van Essen 1991). On the top of the hierarchy there are neurons in certain areas representing quite abstract entities coming close to (*T*-)symbols. For example, in the mediotemporal lobe, cells have been found that respond in a specific and invariant way to a particular person, regardless of the stimulus being the person's face, a cartoon, or their written name (Quiroga *et al.* 2005).

There are many neural models (and AI systems) claiming to do object recognition in a similar way (e.g., Riesenhuber and Poggio 1999;Wersing and Körner 2003), although performance cannot yet be compared with that of the real brain. The basic principle of a hierarchical object recognition system is illustrated in Figure 7.6, left panel. Processing along the hierarchy usually changes gradually in two ways: on the one hand, the represented feature quality becomes more and more specific with increasing hierarchical level (e.g., from basic features to particular objects). On the other hand, the representations become more and more invariant against transformations in the sensory space (e.g., translation, scale, rotation, colour, lighting conditions, etc.). The first aspect is essentially an AND operation (e.g., an corner consists of a vertical edge *and* a horizontal edge), the second aspect an OR operation (e.g., a feature configuration may occur at one *or* another location, as indicated by the pooling in Figure 7.6, left panel).

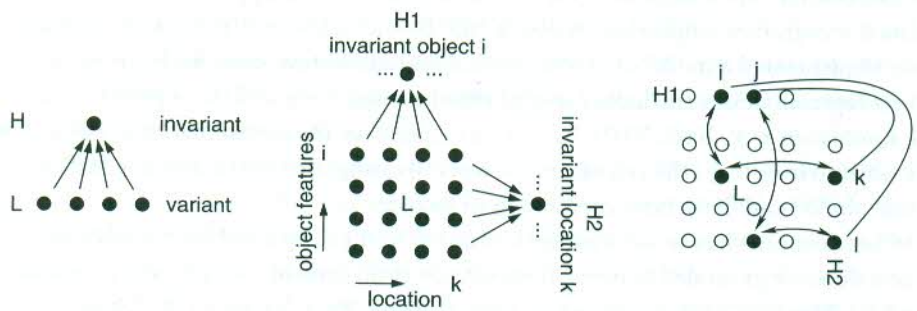

Fig. 7.6 Illustration of sign grounding, anchoring, or binding. Left: The development of invariant abstract (e.g., symbolic) representations implies a noninvertible mapping between a lower processing stage L and higher stage H (e.g., mapping feature configurations at different spatial locations to the same object label). This complicates adequate grounding, in particular if several objects are processed at the same time. Middle: To become invertible one needs further (possibly independent) mappings, e.g., an explicit representation of the object location. Right: In the presence of several objects, correct bindings between the higher-level symbols (e.g., object label and location) and grounding to the lower-level features may require additional mechanisms (e.g., temporary associations).

Thus, such an object recognition system must somehow map or bind, on each level of the hierarchy, a more abstract representation to particular configurations of lower-level features which may vary significantly. Realizing this in a bottom-up fashion is straightforward, for example, by means of a multi-layer perceptron. However, because of the invariance property (by the OR operation), this mapping between the two levels is necessarily noninvertible, which complicates binding between the two processing levels, in particular for more realistic scenarios requiring feedback processing (e.g., dynamical ambiguous scenes with multiple objects, occlusions, clutter, etc., as well as particular queries referring to low-level subsymbolic properties of the objects).

The invertibility of the mapping can be (almost) restored by further independent processing paths explicitly representing the varying factors (e.g., object type and object location in Figure 7.6, middle panel). However, this may lead to a further binding problem if several objects are processed at the same time; for example, in Figure 7.6 (right panel), it is problematic to ground the two objects i and j of layer H1 in the feature layer L because it may be unclear which object occurs at which of the two locations k and l represented in layer H2. Implicit binding over the lower layer L is possible but nontrivial; note that the segregated nodes of layer L may in fact correspond to distributed overlapping feature configurations. Neural experimenters and theoreticians have spent a lot of time investigating this form of the binding problem, and several explicit solutions have been suggested, for example by spike synchronization and oscillations, rapid reversible synaptic plasticity, dynamic routing, attention, coarse conjunctive coding (e.g., see von der Malsburg 1999; Shastri and Ajjanagadde 1993; Treisman 1998; Mel and Fiser 2000; Knoblauch and Palm 2002).

In summary, it appears that embodiment is very closely linked (if not identical) to the problems of symbol grounding (Harnad 1990), symbol anchoring (Coradeschi and Saffioti 2003), binding, and invariant recognition in distributed hierarchical systems. The underlying problem is always the coordination between the bottom-up and top-down streams of information in a hierarchical system, which still lacks a satisfactory solution in current artifical systems and theoretical models.

7.3 Conclusions

In general I am inclined to accept the major claims of the embodiment proponents, for example, that understanding the meaning of a sentence may require the ability to simulate or predict the situation described by the sentence, or that meaning is closely related to action and perception (which may cause interference effects as observed in experiments, see Glenberg and Kaschak 2002; Rizzolatti *et al.* 2001; Guillery 2003). But, in contrast to Searle's Chinese room argument (Searle 1980), I see no good reason why implementations on symbolic systems such as Turing machines or currently available digital computers should not do that job. Nevertheless, special hardware such as massively parallel neural networks – although in terms of price, speed, and flexibility still inferior to general purpose processors – could be of great advantage to this end (e.g., Hammerstrom *et al.* 2006).

Furthermore, I believe that many so-called symbolists would ultimately agree with these positions, and that probably much of the remaining disagreement results rather from imprecise definitions about what exactly is a symbol and what is embodiment. As discussed in Section 7.1.1, it is important here to distinguish between what I have called *T*-symbols and *S*-symbols. *T*-symbols are used by the cognitive system for thinking and communicating, whereas *S*-symbols may be used to implement the cognitive system (including *T*-symbols) on a symbolic substrate such as a digital computer, Turing machine, or in a neural network architecture (see Sections 7.2.1 and 7.2.2). For example, a *T*-symbol could be a word for an object, whereas an *S*-symbol could be a state of a Turing machine. Note that this distinction leads to the apparent contradiction that a system can be both (*T*-)embodied and (*S*-)symbolic. This is simply because embodiment in the sense of 'employing systems used both for action and perception' is actually a high-level property and therefore independent of the implementation substrate. Thus, any implementation of an embodied system on a digital computing device must be called (*S*-)symbolic.

We have also seen that the dichotomy between symbolic and embodied systems can be problematic. As argued in Section 7.1.2, any cognitive system must be necessarily both symbolic and embodied. Moreover, defining embodiment by activity in systems for action and perception implies that being *T*-embodied is a rather gradual property: a high-level *T*-symbol may be embodied more or less deeply in a hierarchy of perception- and action-related subsystems. In the brain at least, there certainly exists such a hierarchy consisting of many different cortical areas (Felleman and Van Essen, 1991). Understanding a particular sentence of *T*-symbols will require some, but probably not all, of these areas. Thus, one could define the degree of embodiment by the number and hierarchical levels of such areas necessary for understanding.

For example, for understanding 'the grass is yellow' a shallow embodiment is likely to be sufficient. Here, *T*-symbolic processing may be enough for understanding which is merely the association of two well-known *T*-symbols. Of course, cell assembly theory alone would predict that the *T*-symbol 'yellow' is so familiar that there will be strong synaptic links from the (*T*-)symbolic 'yellow' subassembly to low-level sensory neurons representing the colour yellow. Thus, hearing the word 'yellow' would inevitably activate those low-level neurons to a certain degree. Here, I think this priming activity would more likely be a side effect of neuronal assemblies and not actually necessary for understanding. One would predict that, in an experiment in which a number of low-level visual cortices could temporarily be deactivated or disturbed (e.g., by TMS) while hearing 'the grass is yellow', understanding is still possible.

In contrast, understanding or verifying the sentence 'the second house on the right has seven floors' requires relating several *T*-symbols to a currently perceived visual scene which may be represented in a (*T*-)subsymbolic format in early visual areas. This process includes identifying the 'second house on the right' among many other buildings and relating the number seven to the floors of the house. Here, one would predict that understanding is much more vulnerable to the deactivation of low-level cortices. Finally, on-the-fly understanding of sentences such as 'the woman crutched the goalie the ball'

containing innovative denominal verbs will probably require even deeper embodiment including mental simulation of the described situation at many processing levels (see Glenberg and Kaschak 2002).

If we accept that embodiment is a gradual property within a processing hierarchy this idea becomes very closely related (if not identical) to long-discussed concepts and problems such as symbol grounding, symbol anchoring, and feature binding (see Section 7.2.4). We could give the following (maybe too) simple definition: a sign or symbol of a higher processing stage H is embodied in (or, synonymously, grounded in, anchored in, or bound with) the signs or symbols of a lower processing stage L iff it is possible to establish a one-to-one mapping between the H signs and L signs.

This definition of embodiment would have several advantages: first, it is a unified definition for seemingly different but closely related concepts. The definition remains independent of particular modalities (such as particular perceptions, actions, or emotions). Instead it is sufficient to distinguish between higher and lower processing stages. But we can still say that a symbol X is grounded in particular action representations Y, and we can still distinguish shallow embodied systems from deeply embodied systems where the one-to-one mapping is established across several processing stages.

However, the term 'mapping' may still be underspecified. Unfortunately, I cannot give a more specific answer, because in my opinion this is one of the unsolved core problems of current brain models and artificial cognitive systems as discussed in Section 7.2.4 (i.e., the problem of adequately integrating bottom-up information with top-down expectations is by no means satisfactorily understood in complex systems consisting of distributed processing hierarchies and capable of developing increasingly abstract invariant (and finally symbolic) representations). I believe that the circuitry of the cerebral cortex (see Section 7.2.3) can tell us a lot about how this integration happens in the brain, and how learning of abstract categories is linked to action and behaviour.

Debate

In addition to Andreas Knoblauch's contributions, this debate section encompasses the discussions surrounding the contributions of Marcel Just and Friedemann Pulvermüller (Chapters 5 and 6, respectively).

Arthur Glenberg: Alright, I will start. I have questions for Friedemann, and perhaps also for Andreas. I was very pleased with most of your talk showing that language is embodied from top to bottom and from bottom to top, from phonemes up to meaning, but then you said, right at the very end, that you think your cell assemblies might be implementing discrete symbols. So, a part of my question is to ask you to say a little bit more about why you believe it's the case that it's implementing discrete symbols. And in particular I'd like you to think about the following idea where you have people listening to a word like 'pick,' and so we know that we're going to get activation in perisylvian regions as well as regions related to hand movement. And the question is this: if it's a symbol that we're showing, then you would expect to see activation for

'pick' to be the same in all contexts, but I imagine you would agree that if we had people thinking about 'pick' in terms of a pen versus 'pick' in terms of a ball, that we would see different picks and then it doesn't seem to me like it's a symbol anymore.

Friedemann Pulvermüller: Well, I believe we have little choice when theorizing about language, to accept the discrete nature of language. There are discrete phenomena in language and other cognitive domains: a word is either in the dative case or the accusative case. There is nothing in between. It either refers to an object or not. It's either a meaningful item or not. Think for example of Andreas' simulations of the geometrical figures. In scene segmentation there is a big problem: how can we perceive a square partly hidden behind a triangle? How can we solve the problem of scene segmentation with non-discrete representations? Without them, there is activation all over the place and the overlapping figures are being meshed into each other. However, if perception is based on discrete representations, which have been learned separately, a particular complex scene can separately activate the pre-established representations, and activation can be maintained in two or more discrete object representations at a time. This seems to work, as neural network simulations demonstrate (see this chapter), even if the discrete representations overlap a little bit.

We have very similar problems in language, too. We need to know whether the word 'pick' or 'kick' has been pronounced, and of course there can be something in between a */p/* and a */k/*. Still, there is categorical perception at the phonological level; and at the lexical and semantic levels as well. At the semantic level it could be two meanings of an ambiguous word, for example, take the word 'star', which could be an object in the sky or a person, and there seems to be no alternative in between, right? So I think that there are lots of problems – in language and cognition in general – which we can only model by using discrete symbols. At the mechanistic level of the brain, neuronal devices such as cell assemblies that behave in a discrete manner may 'ground' symbols. Cell assemblies, or as we sometimes call them nowadays, action–perception networks) are discrete because they are linked together so strongly, because of their strong internal feed-back and feed-forward connections. If they are activated to a certain degree, there is an explosion of activation within the cell assembly and the whole circuit gets active at once, which is very handy, especially for modelling discrete psychological processes.

Donald Hebb already pointed out that this is potentially an important process in Gestalt completion, which is otherwise difficult to explain in mechanistic terms. And here we have a very handy, neurobiologically plausible mechanism for it. I could go on for an hour explaining the necessity and benefits of discete representations, but I think do not need to. I think there are lots of reasons, at both the cognitive and neurobiological level, in favour of discrete representations and discrete neural processing units. Now, why do we not see them in current connectionist parallel-distributed processing models? Because these models are toy networks that do not have what the cortex is most famous for, namely massive local connections between neighbouring neurons. Many neural networks avoid strong next-neighbour connections within a particular layer because the network doesn't behave so nicely in this case for

certain purposes. We are emphasizing the importance of these next-neighbour, within-layer, within-area connections; these are, as we believe, important for setting up cell assemblies. A similar point applies to rules or rule representations, if you want to speak about such entities. Here, I can speak for Andreas too. Art said that our approach is 'embodied from top to bottom and from bottom to top' and I think he is right. We document that the embodied perspective can even explain the emergence of abstract rules, a claim based on biologically inspired network simulations, and certainly gives rise to other discrete representations.

Glenberg: So, I just want to press this one more time and then I'll give up on it. Suppose we have an experiment where we have people listening to two sentences, and the first sentence is either 'There's a ball in front of you' or 'There's a pen in front of you.' And the next sentence is 'Pick it up.' Do you think activation to the word 'pick' would be the same in both of those contexts?

Pulvermüller: No, absolutely not. Veronique Boulenger, Olaf Hauk, and I are doing a more extreme experiment about 'grasping ideas'. We have action words like 'grasp' presented in sentences like 'He grasps the ball' and 'He grasps the idea'. And, of course, in the idea-grasping case there is a different type of grasping and it seems that the cortical activation pattern changes, although semantic somatotopy is present even in the abstract idiomatic case. Different activations by the same symbol presented in different contexts is not an argument against discreteness, but can be attributed to pre-activation through the context network. It would not be realistic to assume that the discrete representation activates each time in exactly the same way. Rather, its particular activation state at any point in time does not only depend on the internal properties of the network, but also on the type of priming or pre-activation it receives from other networks. So that's the brief answer to a complex question. Context dependence and discreteness are independent features.

Glenberg: Thank you.

Lawrence Barsalou: I'd like to comment on this. Like, one solution to this might be to say that a symbol is not a single state of activation but it's actually something more like an attractor state that has many different possible states of activation.

Pulvermüller: Yes indeed. The word would be understood in an all-or-none discrete manner each time, but in different contexts a set of its possible semantic features are emphasized or even added; for example, referential features related to specific hand motor acts or related mental operations.

Barsalou: So, it may be that all the different instances of 'pick' that get processed are attracted to the same general processing system, but that system can take different states on different occasions for different contexts. So a symbol is really a space of applications not a single application?

Unidentified person: This reminds me of the debate in physics as to the nature of light – the particle/wave debate – which was solved in terms of a field theory of physics. But I wanted to emphasize what you told us about the structure of the column in which, if I understood you correctly, each column in the cortex has a layer

especially, functionally, dedicated to perception and another layer to action. Is that right? Because this seems not to be in clear convergence with the standard view of the cortex as specialized in regional terms, not being sensory and motor all the way around.

Andreas Knoblauch: Yes, I think cortical layer 5 has a lot of pyramidal cells that actually project to subcortical structures that are related to actions.

Unidentified person [DEB ROY?]: So I'd like to actually just push a little more on the same question that Art was asking. 'Grasp the idea' versus 'grasp the ball' is, I think, a different point. It's a sort of different word sense, because that's an idiomatic use of 'grasp'. Another way to ask the point I think Art is trying to make, or maybe to make a suggestion, is that if you take a very simple model of how words work – which I think is consistent with the story you've told us – that you have something like, say, a conceptual layer versus a symbolic layer, so you've got words and you've got some sort of sensorimotor something or other connected. Clearly, at the word level, I think we'd all agree that's a discrete system. It's not meaningful to talk about halfway between 'kick' and 'pick' in the lexicon. Phonetically you can talk about that, but you impose a phonological discretization and on it goes morphologically and into the lexicon. But that discrete system is carving up perhaps a continuous conceptual space. So there's no reason for example to believe that 'pick' doesn't bind to a sensorimotor representation that actually has parametric variation, right? So when you kick a ball versus kick a chair versus some other argument of 'kick', there'll be some continuum. So I think there's room to think about both a continuous system *and* a discrete system. And one projection, the projection into language, you can treat as discrete. But it's not clear at all that the back end is discrete.

Pulvermüller: Whether or not the semantic system as such is entirely continuous, I cannot say. At least there is some evidence that word meanings dissect an otherwise continuous conceptual space, for example, in the case of colour concepts. But we are in fact talking here about two different things now. Let me take up the specification. Before, Larry made the point about attractor states. We can say that a cell assembly is a discrete unit – it either becomes active or not – but that its activation states may differ, may change over time and between contexts? So it has an activation minimum, or maybe several activation minima. And depending on where the pre-activation comes from, where it is primed and by what it is primed, certain parts of it can, in a given context, be emphasized, so to speak. And therefore it is, in each case, either active or not, but the particular neurons active in each situation can be different. So its states of activation are not identical. However, it's still in each case a discrete state, it's either active or inactive. So, this would be my explicit answer. There is room for variablitiy among the different activation states of the discrete network

Unidentified person: Yeah, I guess the question is, given the precision of the best imaging technologies, we're reading a lot into whether it is the same network or not, or would you disagree? An average brain has millions of neurons, so there's a lot of room ... it's unclear.

Pulvermüller: Well, fMRI is not the best method to address these questions. Multiple-unit recordings in monkeys – especially the work done by Moshe Abeles' group in Israel – have produced very strong data exactly on this issue, that there are specific networks, serving specific cognitive functions, that are active or inactive and exhibit similar activity states and patterns when active. That is, I think, very important work. Also, complex dynamics approaches to EEG and imaging analysis are, maybe, to a degree telling. One wouldn't get that information from fMRI, according to my feeling, because it misses out on the time aspect. Some imaging data show a surprising degree of local specificity, and this is certainly something to look into much more in the future; but we are clearly looking at neuronal mass activity with all neuroimaging methods. Trouble is, the relevant circuits may be large, may include millions of neurons, so a multiple-unit approach alone cannot succeed. Bridging across methods may help. Maybe you want to comment?

Unidentified person: Yeah, I'd say something. It seems to me that several issues are being conflated. First, it seems to me there is likely to be contextual sensitivity of the representation of meaning. Second, I think there's ample evidence that meaning or representations are distributed, and that there are different components of it. I like to think of the different components of meaning as corresponding to different brain locations. That's not the only way in which they might be distributed and it's conceivable that different components have different attributes, that some are much more symbolic in nature and others are much less symbolic. It seems to me that motor representations would be less amenable to, I don't know, a symbolic characterization. Not that you can't say it's 'grasp' – it's still 'grasp'. And I think that's what Deb [Roy] is in agreement with. If you look at the activation carefully, as you say, even without the idiomatic meaning of 'grasp', you would expect to see different representations there. So, I think you can have contextual sensitivity, you could have differential degrees of symbolic-ness for different components. It seems to me that all of this could work quite nicely. And really our job might be to specify, you know, how is it you say or grasp a tennis ball – how does it magically come to be that we're going to get different activation? It seems to me that's a very interesting question. I'd love to be able to answer it. But the contextual sensitivity occurs and we've been doing context effects in experimental psychology probably for a century now. Maybe now we can finally come to the point where we can actually identify the mechanisms by which the context has its effect.

Unidentified person: I also wanted to come back to this issue. Context sensitivity seems to be, like, impossible when you have truly discrete symbols like things that are supposed to be kind of immovable in what they represent, right? But when we think about it in terms of a continuum from analogue to discrete … take the instance of children's development where they go from being very context sensitive to being extremely, you know, able to ignore certain irrelevant contexts. And I think about it terms of Larry Barsalou's suggestion here, in terms of attractor states. It's this idea that

first you have really shallow attractor states, where, you know, you're able to kind of be bounced around by irrelevant variations in context. But those attractor states become deeper and deeper as you have more experience with those networks, activating in certain orders and correlating in different areas of neural networks, and you know, 'what fires together wires together.' And eventually you get these very deep, kind of more stable attractor states but that doesn't mean you can't shake them. That even adults, when they have a kind of stable attractor state, could still shake them with different forms of context –they just have to be stronger manipulations in something. And so to take these slices of big neural connections and talk about them as if they were whole trajectories, it's a question of how quickly you can or how much context you need to shake people out of those trajectories versus, you know, how little.

Barsalou: I just wanted to point out that even words are not fully discrete. I mean, a given word spoken by a given person will take a different form as a function of the surrounding linguistic context ... different people uttering the same word will have different acoustic and articulatory properties, so even for what seem like symbolic kinds of things it's still a space of possibilities that get realized differently in different contexts. But I would agree that just as the way phonetic input, as you were saying, tends to go towards those attractors, I think the same thing happens at the conceptual level. If you look at the world of objects, rarely do you get stuck between categories. I mean it's a winner-takes-all situation; you go into one conceptual attractor pretty much all of the time, and if you get stuck in between it's really a weird conceptual state.

Unidentified person: I would suggest that the question of contextual sensitivity probably has to do with the level of representation of action in the brain. I mean, Rizzolatti's group recently proposed that action representation or action control in the brain involves somewhat different structures or areas. Some of them have to do with low-level motor problems – particular combinations of muscles, motions, and so on – but some of them are much more abstract. They have to do with sort of general plan of the action. So I think it might be relevant, for instance, to try and decide the granularity of embodied representation of action verbs. An action verb in a noncontextual situation, like your experiments for instance, maybe don't activate lower-level motor structures in the brain but instead some sort of higher-order or abstract representation of action. What is your opinion?

Pulvermüller: This is an excellent question. I think this is exactly the way to go forward. And maybe I can just tell you verbally how we tried to go exactly this way in the last 2 years, looking at words that have a more general action relationship compared with the very primitive, concrete body-related, even body part-specific, action word. So instead of using a word like 'kick', which is in its meaning bound to the leg (one of its semantic features could be described as 'typically done with the leg'), we have used other words that are related to action patterns, words that are highly abstract, as you mentioned. And one such category is words that relate to forms, to shapes; for example 'triangle', 'square', and 'rhomb'. Looking into the neuropsychological literature, these form or shape words are usually treated as visually related, but that's not the

entire truth. We know that when people look at shapes they follow the shape with their eyes, they can point out the shape with their hand, and, if needed, they can walk a path of that shape with their feet. So, if you like, the action pattern related to the meaning of the word is not bound to a particular part of the body; there's an abstract action scheme associated with these kinds of words and concepts. And what we find for them in the brain is activation that starts in motor and premotor cortex, but extends into prefrontal cortex, spreading anteriorly compared with the 'primitive' action words. It seems that these higher-level, less-concrete action words involve more frontal areas, Brodmann areas 9 and 46. Action execution, action recognition, and action semantics of different types may involve overlapping but nonidentical systems in frontocentral cortex that include mirror neurons.

Francis Quek: Are there any motion verbs, or motion actions, that are uniquely concrete? For example, if you ask somebody how many ways you can kick, there are myriad ways. Another way to ask the question is whether the neural representation is a specific pattern or a set of patterns: what confidence do we have that each time we think of 'kick' that the exact same distributed pattern takes place? Can there be degrees? Talmy, for example, says that for a particular motion verb there are many aspects to consider – there is ground, there is manner, there is direction. So, none of these verbs, none of these grounded descriptors, need to be absolute in their representation.

Pulvermüller: Well, I think this brings us back to Larry Barsalou's comment on attractor states. Again, my view is this: if there's one discrete network, this network can still have different attractor states and, again as a function of the priming of the network, can be in different basins, or activatity minima of the attractor landscape, so to speak. The property of discreteness is very compatible with context effects and variation in the activation patterns. As for your question about how good the imaging methods are at finding out about the variation, well, we are working on that and maybe in 2 years' time I could give you more concrete answers on that. I would hope that imaging could tell us a little about context-specific activation of the networks.

Unidentified person: How would you contrast a 'do' versus 'make' versus 'bake' in terms of the attractor basin? For 'do' I could imagine just about anything goes, whereas 'make' is going to be a little more constrained but it might have that either/or-ness, whereas 'bake' is going to be very little either/or-ness.

Barsalou: I think the standard linguistic response is that 'do' focuses on a process and 'make' focuses on the result.

Unidentified person: There is a lot of developmental evidence saying that verbs like 'read' are at first highly constrained to objects, like kids learn the word 'read' only in the context of books, and so it's a very early verb; it's early acquired, it's very easy for them to learn. Whereas with a verb like 'put,' it's just like, gosh, you 'put' with almost anything. And there's also languages where verbs are much more specific to objects than English perhaps is, verb-heavy languages like Japanese and Korean. They're highly shape-focused and have very spatial verbs, so in that sense I bet you might be able to

find linguistic differences with younger children who might be more object sensitive. Very young children, say 12 months old, are actually sensitive to agent changes, like they won't say the same verb is the same verb if a different person does it, even if it's the same sex. There are lots of superficial similarities, and so I think there's a lot of that evidence, at least behaviourally.

Unidentified person: Yeah, I would like to come back to another question. Marcel in his talk gave this metaphor of a team, and the players in a team, which I thought was a very useful way to think about the whole thing, and also in Friedemann's talk we saw that all over the place there are systems involved. So, my question is, is there anything more already, from a neurobiological point of view, about how this team gets coordinated, or how it gets assembled, I should say? I mean, at some point you mentioned the word 'probe' as if one process would send out, you know, a message for another subsystem to probe.

Unidentified person: Yes, I think that's one of the key questions, because as you enter any situation, you know, depending on what kind of sentence I now use, what kind of reference I make, a different set of areas could activate. If I start speaking in a visual metaphor then there may be more intraparietal sulcus activation, and so on. And the question is how does this magic happen? I work with a computational modelling system – it's kind of a hybrid system, but basically the problem is viewed as a resource allocation in a complex system. You have some job about to be done, maybe a job for an athletic team, a sports team, but it could be anything – a job for a university, for a corporation, or whatever; a complex system, a city – and a challenge comes along, for example, a large snowstorm, a plane crash, you know ... a set of some product goes down the tubes or whatever. And the question is, how does this system respond? Some people like to think at an executive level; you know, the president of a university calls together all the deans, how are we going to do this? But there are other ways that a complex system sometimes responds, a more distributed response to a resource need. In the model I work with it's really about production systems down at the level of the individual component systems, such that they're all monitoring what's going on in the system at large. So whenever a particular need arises – suppose there's a desire at some point to develop a spatial representation or there's a desire to have a motor representation, or whatever – the area that is most specialized for that need, that has the greatest efficiency, the greatest history of doing that kind of thing, gets to perform that action. And so on a continuous basis there's always a thinking, the team is always sort of selecting itself – is this my chance to go in? Do I do this?

I'm not saying there can't be executive control. I think that in some cases there are. But my view is that there is a sort of self-allocation, a self-selection based on expertise, efficiency, past experience, and so on. Let me give you another argument for this: how is it that Broca's area gets to be Broca's area? For all of us, you know, when you generate either subvocal or actual speech, Broca's area activates. How does that come about? Developmental theorists suggest you know that the different cell types throughout the brain have different characteristics, and the classic example is that cells with particular

timing characteristics that are kind of slowish will allow you to do the timing of phoneme distinctions. And maybe those are around the primary auditory cortex, so the auditory cortex, in Liz Bates's terms, gets the contract for processing auditory information. According to Liz, it's not as though you're born with a sort of a label saying 'auditory cortex', but you're born with a bunch of cells in that area that are really good at getting timing distinctions. And so the team members develop their neuroprocessing capabilities by virtue of their capabilities. So the auditory cortex sort of initially wins the contract for processing timing information, and then when the person is 6 months old or 60 years old, when an auditory challenge comes up, that's what activates. It's been doing it since its infancy, and it sort of holds the contract; it's going to try to develop the appropriate response or representation or whatever. That's a long-winded answer to a complex question.

Pulvermüller: I have some reservations with regard to using humanoid metaphors when speaking about brain function. I think considering neurons football players or contractors can be advantageous for educational purposes, but I would still believe that going down to neuroscientific principles might be an alternative option, and in this case what counts is neuronal correlation as it is a major driving force for binding. What becomes bound together is determined by neuronal correlation of activation. So how does Broca's area become Broca's area? Well, Broca's area happens to be close to the motor cortex and this is pre-wired in a way. And the auditory cortex receives auditory information on the basis of very early ontogenetic processes, which are usually finished at the time when the major phases of language learning start. Therefore, the information about motor programming for articulation goes out and the auditory input arrives in these distant areas, and now there is another problem: the motor cortex and the auditory cortex are not linked directly with strong fibre bundles, so if the neural activation patterns should communicate with each other it needs to take a detour. And this detour is exactly through Broca's area and through superior temporal areas in the so-called belt and parabelt region. This is where the long-distance connections emerge. So I have a two-fold answer: one driving force is neuronal correlation, another is the cables built into the brain – the major neuroanatomical tracts that bridge between the sensory- and action-related activations, that connect the areas in which the correlated activation happens. This is the magic that makes Broca's region to act as a binding site for language: correlated activity sets up representations that span across areas linked by long-distance neuroanatomical connections. Thanks very much for the question, A very important question. I think Andreas is more of an expert on this, so maybe I can pass on.

Knoblauch: 20 milliseconds, I think that's not a big problem. The time window for synaptic plasticity lies somewhere in the range of 50 milliseconds or so, for example, the spike timing dependent plasticity, so 20 milliseconds would still mean the synapse is strengthened.

Quek: I have a question here. One of the things I've been interested in is how embodiment in the neuronal structure that we have, for example, for thinking of space and so

on, gets implicated into understanding the limit of a mathematics function – these are the kinds of things that Lakoff and Nunez have talked about. So can you suggest, with the pre-wiring and so on, how does repurposing take place, if it does, if we actually repurpose certain neural architectures that are designed, for example, for sequentializing of events, to understand mathematical functions. I doubt that we are born with mathematical processing in our brain that waits for the first time we take a mathematics class to wake up, okay. So if indeed mathematics, for example, or higher level thought, is actually repurposing of certain neural pathways that have other designs, how might you speculate this takes place?

Pulvermüller: It's a very difficult question. No one wants to take it.

Unidentified person [QUEK?]: Our culture is constantly coming up with new skills, so reading is not god given or biologically given ... certainly maths and computer programming. Oh, and driving a car – none of these are biologically inherent skills and so the question is how do you acquire them. Others have talked about how that's where grounding tends to play a role. It's sometimes difficult to go straight to what is sort of an abstract skill without some sort of grounded representation, and that's why – I don't know very much about this – but where you have manipulatives in maths, you know, you're playing with arrays and getting addition and multiplication through some visual or manipulative activities at first before you move to a sort of a more symbolic level.

Robert Goldstone: This is sort of a conceptual question. A lot of the ideas that have been going around with Friedemann and Marcel, and what Larry has said, suggest that a large part of the meaning of the words is based upon the activation of perceptual and motor cortices that would be normally associated when you were using or normally understanding the words. And I guess my conceptual stumbling block is how do you know that these are perceptual and motor cortices if they're so tightly bound to the word side of things? So if you cut a slice of the skull out and you open up, you wouldn't see a Post-It note that says 'This is the motor cortex', right? And so it seems to me that if you had done the word meaning experiments before you had done the finger twiddling experiments then you could have said 'wow isn't it amazing that the finger twiddling causes the same area to light up as the word's semantic area.' So, I mean, I'm totally on board when you talk about correlations or the distributed nature of the representations, but I get a little bit less sure when I think about ascribing singular functions to those areas.

Pulvermüller: Well, the lucky thing is that these areas are in fact labelled. If we open the brain and look intensely at it we see that exactly from which areas the cables lead to the arm or to the leg. Well, not directly: the large pyramidal neurons in motor (and premotor) cortex project to the neurons in the spinal cord that reach arm or leg muscles, respectively. There are these huge pyramidal neurons in Brodmann area 4, and they can have extremely long axons, sometimes a metre long, going down into the spinal cord, which link the motor cortex through one intermediate neuronal step to the respective effectors.

Goldstone: Right, but that doesn't...

Pulvermüller: The motor cortex is labelled and it has been known for hundreds of years. After the French revolution there was sometimes a French neurologist next to the guillotine and he was sticking his needle in the cut-open spinal cord, looking for whether the foot would be twitching or the arm, and therefore, by such truly revolutionary brain mapping, the first motor maps were made.

Goldstone: Yeah, I understand that.

Pulvermüller: The brain is labelled.

Goldstone: Yes, I understand that, but it doesn't exclude the idea that these areas are multifunctional. I mean, I guess I would just make the general point that there's a sense in which if you're talking about separate areas you haven't totally gotten on board with the idea of a fused perceptual–conceptual system; you're still saying this is the perceptual area that is borrowed by the conceptual area. Maybe you don't want to go this far, but you could have a more integrated view where you said that these areas are actually doing multiple functions. So it's not even totally clear-cut that the only thing that this is doing is controlling the motor processes.

Pulvermüller: Absolutely, the implication is that the area is a multifunctional area, not only doing motor programming but also doing the motor–conceptual processing that is elementary for understanding action words. Action verbs, action words generally, action concepts ... I would go with that fully. Nevertheless, as we discussed earlier, overlapping but different activation topographies and network circuits may underlie action execution and recognition, and the comprehension of action language and concepts.

References

Abeles M (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge: Cambridge University Press.

Barsalou L (1999). Perceptual symbol systems. *Behavioural and Brain Sciences*, 22, 577–609.

Barsalou L (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Science*, 7, 84–91.

Braitenberg V (1978). Cell assemblies in the cerebral cortex. In R Heim, G Palm, Eds. *Lecture Notes in Biomathematics 21: Theoretical Approaches to Complex Systems* (pp. 171–88). Berlin: Springer-Verlag..

Braitenberg V, Schüz A (1991). *Anatomy of the Cortex. Statistics and Geometry*. Berlin: Springer-Verlag.

Brooks R (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, 6, 3–15.

Chalmers D (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Coradeschi S, Saffioti A (2003). An introduction to the anchoring problem. *Robotics and Autonomous Systems*, 43, 85–96.

Dennett D (1996). *Kinds of Minds: Towards an Understanding of Consciousness*. London: Weidenfeld & Nicolson.

DiVincenzo D (1995). Quantum computation. *Science*, 270, 255–61.

Douglas R, Martin K (2004). Neuronal circuits of the neocortex. *Annual Review in Neuroscience*, 27, 419–51.

Felleman D, Van Essen D (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1, 1–47.

Glenberg A, Kaschak M (2002). Grounding language in action. *Psychonomic Bulletin and Review*, 9, 558–65.

Guillery R (2003). Branching thalamic afferents link action and perception. *Journal of Neurophysiology*, 90, 539–48.

Hammerstrom D, Gao C, Zhu S, Butts M (2006). FPGA implementation of very large associative memories. In A Omondi, Rajapakse J, Eds. *FPGA Implementations of Neural Networks* (pp. 167–95). Cambridge, MA: Springer US.

Harnad S (1990). The symbol grounding problem. *Physica D*, 42, 335–46.

Hawkins J, Blakeslee S (2004). *On Intelligence*. New York, NY: Times Books.

Hebb D (1949). *The Organization of Behaviour: A Neuropsychological Theory*. New York, NY: Wiley.

Hopfield J (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science USA*, 79, 2554–8.

Hopfield J (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Science USA*, 81, 3088–92.

Jackendoff R (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press/Bradford Books.

Knoblauch A (2005). Neural associative memory for brain modeling and information retrieval. *Information Processing Letters*, 95, 537–44.

Knoblauch A, Fay R, Kaufmann U, Markert H, Palm G (2004). Associating words to visually recognized objects. In S Coradeschi, A Saffiotti, Eds. *Anchoring Symbols to Sensor Data. Papers from the AAAI Workshop. Technical Report WS-04-03* (pp. 10–16). Menlo Park, CA: AAAI Press.

Knoblauch A, Kupper R, Gewaltig M-O, Körner U, Körner E (2005). Design and simulation of a cortical control architecture for object recognition and representational learning. In H Tsujino, K Fujimura, B Sendhoff, Eds. *Proceedings of the 3rd HRI International Workshop on Advances in Computational Intelligence*. Wako, Japan: Honda Research Institute.

Knoblauch A, Kupper R, Gewaltig M-O, Körner U, Körner E (2007). A cell assembly based model for the cortical microcircuitry. *Neurocomputing*, 70, 1838–42.

Knoblauch A, Palm G (2001). Pattern separation and synchronization in spiking associative memories and visual areas. *Neural Networks*, 14, 763–80.

Knoblauch A, Palm G (2002). Scene segmentation by spike synchronization in reciprocally connected visual areas. II. Global assemblies and synchronization on larger space and time scales. *Biological Cybernetics*, 87, 168–84.

Körner E, Gewaltig M-O, Körner U, Richter A, Rodemann T (1999). A model of computation in neocortical architecture. *Neural Networks*, 12, 989–1005.

Mel B, Fiser J (2000). Minimizing binding errors using learned conjunctive features. *Neural Computation*, 12, 247–78.

Metzinger T (2004). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press/Bradford Books.

Palm G (1980). On associative memories. *Biological Cybernetics*, 36, 19–31.

Palm G (1982). *Neural Assemblies: An Alternative Approach to Artificial Intelligence*. Berlin: Springer.

Palm G (1990). Cell assemblies as a guideline for brain research. *Concepts in Neuroscience*, 1, 133–48.

Pulvermüller F (1999). Words in the brain's language. *Behavioural and Brain Sciences*, 22, 253–336.

Pulvermüller F (2003). The Neuroscience of Language: On Brain Circuits of Words and Serial Order. Cambridge: Cambridge University Press.

Quiroga R, Reddy L, Kreiman G, Koch C, Fried I (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435, 1102–7.

Rao R, Ballard D (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2, 79–87.

Riesenhuber M, Poggio T (1999). Hierarchicalmodels of object recognition in cortex. *Nature Neuroscience*, 2, 1019–25.

Ritz R, Gerstner W, Fuentes U, van Hemmen J (1994). A biologically motivated and analytically soluble model of collective oscillations in the cortex. II. Applications to binding and pattern segmentation. *Biological Cybern*etics, 71, 349–58.

Rizzolatti G, Fadiga L, Fogassi L, Gallese V (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661–70.

Searle J (1980). Minds, brains, and programs. *Behavioural and Brain Sciences*, 3, 417–57.

Shastri L, Ajjanagadde V (1993). From simple associations to systematic reasoning: a connectionistic representation of rules, variables and dynamic bindings. *Behavioural and Brain Sciences*, 16, 417–94.

Steinbuch K (1961). Die Lernmatrix. *Kybernetik*, 1, 36–45.

Treisman A (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London B*, 353, 1295–1306.

von der Malsburg C (1999). The what and why of binding: the modeler's perspective. *Neuron*, 24, 95–104.

Wersing H, Körner E (2003). Learning optimized features for hierarchical models of invariant object recognition. *Neural Computation*, 15, 1559–88.

Willshaw D, Buneman O, Longuet-Higgins H (1969). Non-holographic associative memory. *Nature*, 222, 960–2.

Wilson M (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9, 625–36.

Young M (1993). The organization of neural systems in the primate cerebral cortex. *Proceedings of the Royal Society: Biological Sciences*, 252, 13–18.

Chapter 8

Symbol systems and perceptual representations

Walter Kintsch

8.1 Introduction

The ability to represent the world symbolically, and hence abstract thought, evolved from other, more concrete forms of representation. The ways in which people model the world in their mind follows an orderly sequence, both phylogenetically and ontogenetically. Donald (1991), for instance, distinguishes a sequence of cultures, each characterized by a particular form of mental representation. All animals learn and have procedural memories. At some level (certainly at the level of primates), animals are able to represent the world in terms of generalized records of past experience that allow them to react directly to situations in a way that summarizes their experience with that situation. This is also true for young children (Nelson, 1996, calls this *general event memory*). As biological evolution continues, representational systems based on intentional imitation emerge. Speech appears at this stage, but its purpose is social communication, not representation. Individuals at this level are characterized by representation through action; cultures are characterized by arts, crafts, and ritual. At this point biological evolution is replaced by cultural evolution: at the sensorimotor level humans are simply primates, but the use of language distinguishes them more and more from their animal ancestors.

Donald makes an important distinction between narrative language and theoretical language use.[1] Narrative language is based on mental models that are fundamentally linguistic. Flexible computations can be performed with models of this kind, giving humans a degree of control over the world that is impossible without language; words and thoughts are inseparable at this level. Most human cultures and human individuals are capable of working at the level of narrative culture. However, few cultures, and by no means every individual in these cultures, achieve the level of formal representation that Donald terms 'theoretical culture'. The engrams of memory are aided by the exograms procured by our technology at this level. Most of what we do in schools is trying to bring students up to that level of cognitive functioning.

This is not the place to fill in the details of this conceptual framework (see Donald 1991 and Nelson 1996). However, a few comments are in order. First, it must be pointed out

[1] Several authors have made the same or similar distinctions (e.g., Bruner (1986) with respect to language, or Vygotsky (1987) with respect to everyday and academic concepts.)

that when an individual or a society moves from one level of representation to the next, the earlier level does not disappear but becomes embedded in the new forms as a new kind of representation is added to the behavioural repertoire and integrated with the earlier ones. Educated individuals in our society who function at the level of formal reasoning are, at the same time, also capable of direct action, of participating in art and ritual, and sharing the narrative traditions of our culture. No wonder psychologists have a hard time analyzing human behaviour.

Thus, although event memory is not basically linguistic, it may become linguistically encapsulated: what we remember is sometimes not the event itself but our (explicit or implicit) linguistic description of the event. Furthermore, when the brain develops new forms of mental representation, it not only does not discard the older ones, but in fact uses the old structures for the new purposes. Cognition is indeed embodied.

How the different levels of representation interact in detail is not known at present, but much research, including this volume, is directed at this goal. The question that I want to address here is how to model these overlapping, intertwined mental representations, particularly linguistic representations. If we take seriously the levels of representation analysis that was sketched above, there seem to be two kinds of approaches to the question of mental representations. On the one hand, one could try to model a system of mental representation that combines elements from all of these levels. Thus, we would have images, concepts, somatic markers, S–R links, features, schemata, abstractions, words, etc., all linked and somehow organized in one grand system, at least in principle, for I have no idea how one would go about constructing such a model. Alternatively, one could model each level of representation separately. This has the advantage that, at least for linguistic representations, we do have an idea of how to begin to construct such a model: I'll describe latent semantic analysis (LSA) and its variants shortly. But first I must address an obvious problem created by the divide-and-conquer strategy. If we model the level of linguistic presentation separately, how can it be coordinated with other levels of representation? In behaviour and the brain, these levels function all together and are not easily separated.

One can think of the different levels of mental representations as a set of coordinated maps. Consider the relationship between the level of linguistic representation (as captured by LSA, for example) and other levels of representation, which we can lump together here as the 'real world'. The relationship between the symbolic representation and the real world could be a first-order isomorphism, where the symbol in some way captures meaning directly (e.g., by a picture, feature system, or dictionary definition). The meaning of a word would be its referent in the real world, as in the perceptual symbol system of Barsalou (1999) and similar proposals. This is what I believe people are saying when they claim that language must be 'grounded' (Harnard 1990).

An alternative way of thinking about maps and their relation to the real world is in terms of what Shepard (1987) called a second-order isomorphism. In a second-order isomorphism the meaning of a symbol is not defined through its reference to another level of representation, but through its relationship with other symbols. The meaning of a word lies in its relationship with other words. Specifically, in LSA the meaning

of a word is a vector of 300 numbers that are entirely vacuous by themselves; they acquire meaning only because they specify how this vector is related to all the other words and texts in the LSA space. In this scheme the real world does not provide the meaning of a symbol, but symbols and referents are nevertheless coordinated as a second-order isomorphism. On a two-dimensional map, I can trace a path from Boulder to Denver that is isomorphic to the trip from here to there in the real world. In the 300-dimensional LSA space, I can compute the semantic distance between these words and expect it to correspond to features of the real world. The cosine between Boulder and Denver is 0.21, way above the cosine between Boulder and farther away locations such as Tenerife (0.01) and Spain (0.02); for similar examples, see the contributions of Zwaan (Chapter 9) and Louwerse and Jeuniaux (Chapter 15) to this volume.

Calling something a second-order isomorphism does not solve the problem of how symbols and the real world are related. That question remains open for research, but it does alter the way one approaches the problem. The meaning of symbols is not to be reduced to perceptual features or actions; rather the question becomes one of correspondence between different levels of representation.

I shall first sketch a proposal for modelling verbal meaning. This proposal has two components: first, we must describe how people induce verbal meaning – the representation of meaning in memory – which is the goal of LSA and related systems. Second, we must model how stored meaning is used to construct contextually appropriate meanings of words, sentences, and texts in general. It is of course possible that all we need for language is already stored in memory and that meanings simply have to be retrieved from the semantic store ready-made. The alternative presented here is that the semantic store only provides the raw material for the construction of meaning, and that meaning emerges when words, sentences, and text are used in context. After this sketch of symbolic verbal meaning, I shall discuss the proposition that, although language is not grounded in the sense of a first-order isomorphism, language mirrors perceptual features of the world with a high degree of fidelity.

8.2 Memory for verbal meaning: LSA and its extensions

Our goal is to construct a representation of the meaning of words and texts that can serve as a model for the kind of representation the human mind builds. People generate such representations while interacting with the world and other people, listening, speaking, and reading, mostly without explicit instruction. A computer simulating this process is restricted to reading texts, but must be able to generate a semantic representation without guidance, as people do.

The kind of representation generated depends in part on the precise mathematical algorithm used to generate it, but more significantly, upon the language input used. We start with a large set of texts (e.g., 50,000 documents, containing 100,000 different words and over 10 million words total) that is analysed into a document-by-word matrix, whose cell entries are the frequencies with which each word appears in every document. Note that this is by no means all the information in the corpus – all we look at is

word co-occurrences, neglecting word order, syntax, as well as discourse structure. Our input matrix is huge (50,000 by 100,000) and sparse, that is, most entries are zero. We then reduce the dimensionality of this matrix, down to about three hundred dimensions. That is, we express the meaning of a word by a vector in a space with far fewer dimensions. This has two effects: first, it is a process of abstraction. We are not interested in the particular documents used to construct the semantic space, the particular topics the authors wrote about, their particular word choices; instead we are interested in the general semantic properties of words. Reducing the dimensionality of our corpus has the effect of discarding a great deal of idiosyncratic information about word use in this particular corpus, while retaining the essential semantic information about word meanings. Secondly, dimension reduction is a process of generalization. Whereas the original input matrix was sparse, the reduced matrix is filled in, so that we have a measure of the semantic distance for every word pair, even though most pairs by far have never co-occurred together in any of our documents. Thus, 'doctor' and 'physician' have a high cosine of 0.61, even though they have rarely been used together in a single document.

There are various ways to perform this dimension reduction. LSA starts with a document-by-word matrix M of size $n \times m$ that is decomposed via singular value decomposition (Landauer and Dumais 1997; Landauer *et al.* 2007). Typical values of n and m are 100,000 and 50,000, respectively. All eigenvectors but the ones corresponding to the d (300 or so) largest eigenvalues are discarded. A word or document is thus represented by a vector of 300 real numbers. This vector is not by itself interpretable but its cosine (or other measure) with other vectors defines its position in the semantic space: hence, meaning is defined by its relationship with other vectors.

Independent component analysis (ICA) assumes that each observation (word, document) is a mixture of independent semantic elements (components, topics) (e.g., Stone 2004; for an application of ICA to language analysis see Mangalath 2007). Several constraints are used to unmix the observations: components are chosen so that they are statistically independent (not merely uncorrelated). This is achieved by searching for components that are maximally non-Gaussian, and/or the least complex components (that is, the most compact or predictable ones).

The ICA model is conceptually related to the *topics model* (Steyvers and Griffith 2007), in which meaning is represented as a mixture of topics. The distribution of topics over words and documents is based upon a Bayesian learning algorithm. An appealing feature of both models is that components or topics are individually interpretable. Thus, the linguistic corpus we are working with most of the time can be analysed into a set of topics, such as drugs, colours, or doctor visits, or, alternatively, into a mixture of independent components.

A quite different approach has been taken by Jones and Mewhort (2007). Word meaning can be represented as a composite distributed representation by coding word co-occurrences across millions of sentences in a corpus into a single *holographic vector* per word. This representation is a pattern of elements that stores the word's history of co-occurrence and usage in sentences.

But which one of these methods – LSA, ICA, topics, or holographic vectors – yields the best semantic representation, or the right one? Experience from working with these models suggests that they produce very similar results. Each has certain advantages for certain purposes, but basically they paint the same picture: what is related in one way in one model is similarly related in the other. There have been no formal comparisons among these approaches, however, so one must be careful with this conclusion, but that is the impression I have at this point. I take this as a positive indication that whatever semantic space we generate is not an artefact of the method used, but truly reflects the semantic information in the corpus.

However, what does make a striking difference is enriching the input to the analysis so that it contains not only word co-occurrence information but also word-order information. This is done in the holographic vector model by adding to the item vector all convolutions of the word with the other words in the sentence. This allows the system to use order information and to infer grammatical categories. It also provides a much better account of human priming data (Jones *et al.* 2006). If only context information is used in the holographic model (as well as in all other such models, like LSA) associative priming (e.g. bee–honey) is reasonably well predicted, but not semantic priming (e.g. deer–pony). On the other hand, if only order information is used in the holographic vector model (equally in n-gram or hyperspace analogue to language [HAL] models; Lund and Burgess 1996), the models handle semantic but not associative priming. But if order and context information is combined in the holographic model, a wide range of experimental data involving both associative and semantic priming can be accounted for.

Including order information in a latent semantic structure allows us to model a far greater range of phenomena than before. But the basic unit of analysis in such an expanded model still remains the word, and it has often been argued that the unit of language and cognition is the proposition, not the word (among others, by Kintsch 1974, 1998). Recent advances in machine learning have made it possible to construct LSA-like systems that represent propositional information in the form of dependency trees. Linguists analyse the syntactic structure of a sentence as a phrase structure tree. A *dependency tree* (Yamada and Matsumoto 2003) is a weak form of a phrase structure tree, lacking the phrasal nodes, and thus does not represent all the relevant linguistic distinctions; however, it does retain information about dependencies among words (i.e., about propositional units). I illustrate this relationship with a simple example. Figure 8.1 shows the dependency tree for the sentence 'Rolls-Royce said it expects its US sales to remain steady at about 1,200 cars,' after Yamada and Matsumoto (2003).[2]

In Figure 8.2, I show the proposition list corresponding to that sentence according to Kintsch (1998), and in Figure 8.3 I superimpose the propositional structure on the dependency tree. The dependency tree shows which words in the sentence belong together as a propositional unit. To obtain a dependency tree, a dependency parser is trained on a

[2] The actual dependency tree includes parts of speech tags which are neglected here.

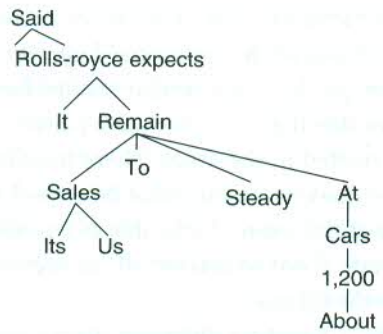

Fig. 8.1 Dependency tree for the sentence 'Rolls-Royce said it expects its US sales to remain steady at about 1,200 cars' (after Yamada and Matsumoto 2003).

large corpus of sentences with support vector machines (Yamada and Matsumoto 2003; Nivre *et al.* 2006). This approach is a departure from the purist path of unguided learning, but that is the only way we can do this analysis at present. Praful Mangalath, in our lab, has begun to explore how the information provided by this dependency parse can be incorporated into the systems discussed above to generate a semantic structure that represents word information as well as propositional information. Basically, we analyse, in addition to the word-by-document matrix, a word-by-dependency matrix. Normally, a sentence like 'The ferocious lion killed an antelope' would make both lion and antelope a little bit more ferocious; however, since ferocious is dependent on lion, only the lion should become more ferocious, not the antelope.

Thus, Mangalath uses syntax (in the form of dependency trees) to guide what is learnt, but syntax can also constrain interpretation. It is possible to use the syntactic structure of a sentence to guide the construction of its meaning. We no longer have to add up the vectors of all the words in a sentence to arrive at the sentence meaning, but we can use the syntactic information to generate an interpretation of a sentence in much the same way as we use it in the construction of the semantic space in the first place. Thus, we turn to the construction of meaning – word meanings, sentence meanings, and text meanings.

(1) Said [rolls-royce, 2]
(2) Expect [rolls-royce, 3]
(3) Remain [sales, steady]
(4) Of [rolls-royce, sales]
(5) Us [sales]
(6) At [steady, 8]
(7) 1,200 [cars]
(8) About [7]

Fig. 8.2 Propositional analysis of the sentence in Figure 8.1 (after Kintsch 1998).

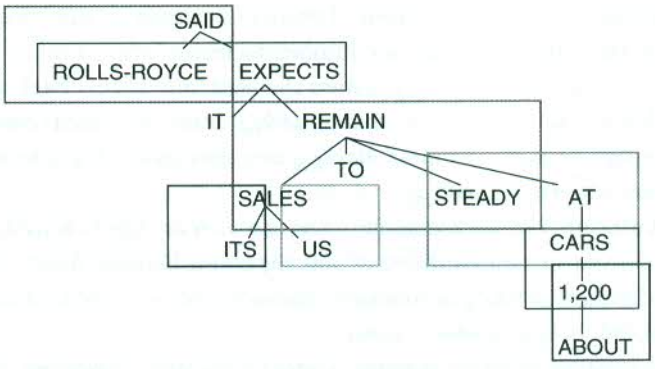

Fig. 8.3 Propositional structure mapped on the dependency tree.

8.3 The construction of meaning

Words often have more than one meaning and several different senses. In LSA, as well as in the other methods discussed above, a single vector represents the meaning of each word. It is like a dictionary that lists each word, but is silent about its different meanings and senses. Instead of listing word meanings and senses, as in a conventional dictionary or in a mental lexicon, meanings can be generated in context. While the semantic representation of a word is a single vector that combines all meanings and senses, a context-appropriate meaning is generated every time the word is used in a different context.

I have described such a model for a generative lexicon in Kintsch (2001, 2007, in press). Briefly, it allows the context to modify word vectors in such a way that their context appropriate aspects are strengthened and irrelevant ones suppressed. In the *construction–integration model* of Kintsch (1998), discourse representations are built up via a spreading activation process in a network defined by the concepts and propositions in a text. Meaning construction works in a similar way: a network is constructed containing the word to be modified together with its semantic neighbourhood and linked to the context; spreading activation in that network assures that those elements of the neighborhood most strongly related to the context become activated and are able to modify the original word vector.

Consider the meaning of 'bark' in the context of 'dog' and in the context of 'tree'. The semantic neighborhood of 'bark' includes words related to the dog-meaning of 'bark', such as 'kennel', and words related to the tree-meaning, such as 'lumber'. To generate the meaning of 'bark' in the context of 'dog' ($bark_{dog}$), all neighbours of 'bark' are linked to both 'bark' and 'dog' according to their cosine values. Furthermore, the neighbours themselves inhibit each other in such a way that the total positive and negative link strength balances. As a result of spreading activation in this network, words in the semantic neighbourhood of 'bark' that are related to the context become activated, and words that are unrelated become deactivated. Thus, in the context of 'dog', the activation

how many meanings and senses a word has, or how they are to be retrieved when needed. Instead, we only have to deal with a single, context-free word vector plus a process (predication) that generates emergent contextual meanings from that vector.

Predication models how word meanings are constructed in context, but what about sentence meaning, and the meaning of texts? LSA has proven to be so useful because it has a very simple and effective way of constructing the meaning of texts by summing the vectors of all the words involved. Despite neglecting both syntax and text structure, if the texts are long enough it produces excellent results, as attested by the commercial success of Pearson Knowledge Technologies (www.knowledge-technologies.com) and the educational success of *Summary Street* (Wade-Stein and Kintsch 2004; Franzke *et al.* 2005). However, the vector sum of the constituent words is not a useful representation when we are dealing with single sentences or brief texts. The syntax plays a large a role in this case, whereas it apparently averages out over longer text, so that its neglect does not cause fatal problems.

We can use syntax to guide the construction of sentence meanings. As in the construction–integration model, we can use dependency analysis to construct propositional units and then integrate these units as required by the task at hand. I shall illustrate this model with two simple examples.

In a cloze test students are given a sentence with a word missing and are asked to fill in the missing word from a given set of alternatives. For example, the sentence might be 'If you try, you can _____ any problem,' with the alternatives 'solve', 'quack', 'flour', and 'seem'. One way to model this task is to compute the cosine between the sentence fragment and the four response alternatives in LSA. As Table 8.1 shows, this leads to the incorrect choice of 'seem'. However, we get a different prediction if we use dependency analysis of the sentence to focus our comparison on the relevant proposition rather than on the whole sentence. Figure 8.5 shows the dependency tree for this sentence, indicating the proposition containing the target word. This analysis suggests computing cosines between the four response alternatives and 'problem', the relevant content word in the proposition containing the target word. Note that while function words were crucial in the construction phase – they determine what kind of dependency tree is constructed and hence, which content words enter into the integration phase – we neglect function words in this comparison because they would only add noise. Table 8.1

Table 8.1 Cosines between four response alternatives and the sentence fragment *'If you try, you can _____ any problem'*

	LSA (sentence)	**LSA (dependency)**
solve	0.49	**0**.88
quack	0.18	0.08
flour	**0**.15	0.03
seem	0.58	0.26

LSA = latent semantic analysis.

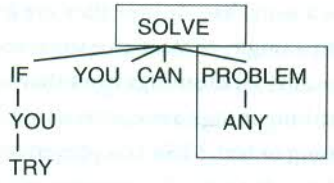

Fig. 8.5 Dependency tree for the sentence 'If you try, you can solve any _____' indicating that the target word must be part of the [SOLVE-ANY-____] proposition.

shows that the correct alternative has the highest cosine with the relevant proposition fragment.[3]

A second example involves the implications that sentences may have. Suppose we have in our corpus the sentence 'Oracle fought to keep the forms from being released' and our task is to determine whether the inferences 'Oracle released a confidential document' and 'Oracle fought to keep the forms released' are true or false. Both inference sentences have a high cosine with the original sentence (cosine = 0.24 and 0.93, respectively), and hence appear to be implied. However, the dependency trees for these sentences suggest otherwise (Figures 8.6 and 8.7). The inference sentence 'Oracle released a confidential document' has the dependencies *Oracle-released*, *released-document*, and *document-confidential*, none of which match any of the dependencies in the original sentence. For the second inference sentence, 'Oracle fought to keep the forms released', the dependencies between oracle, fought, keep, and forms are the same as in the original sentence, but not between the rest: in the original sentence there are dependencies between *keep-from* and *from-released*, which do not match the *forms-released* dependency in the inference sentence. Hence, both inferences can be rejected as invalid. Function words play a crucial role in this process: they do not enter the comparison process (in LSA or ICA, function words like 'from' tend to be uninformative, having been experienced in all kinds of different contexts), but they determine what is compared with what in a sentence.

The point I am making with these examples is this: we have fairly good models of how people store verbal symbolic information in memory, and of how this information is used in constructing the meaning of words, sentences, and texts. The question is, do we have to throw away these models because 'No matter how LSA, HAL, and other ungrounded symbol theories are extended and modified, ungrounded arbitrary symbols cannot be an adequate basis for human meaning' (Glenberg and Robertson 2000, p. 397)? My answer is 'not at all' – this claim, and similar claims by others, is ungrounded and based on a misconception about the relationship between symbolic and nonsymbolic representations.

[3] It should be noted that there are other methods that allow LSA to solve cloze problems, using n-gram or word-order information.

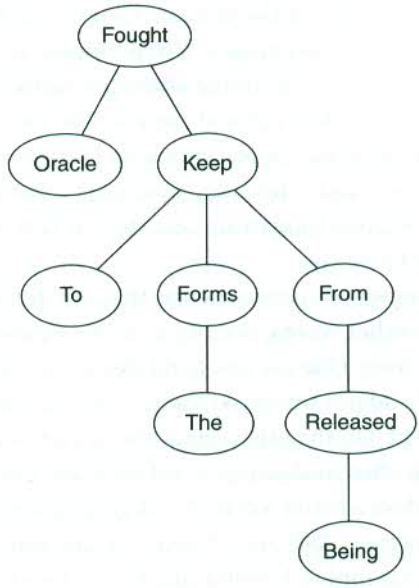

Fig. 8.6 Dependency tree for the sentence 'Oracle fought to keep the forms from being released'.

8.4 **Language as the mirror of the mind**

When people remember a list of words, read a sentence, or listen to a story, they do not function solely at the symbolic level. They use imagery to make words more memorable (Paivio 1969), they construct situation models that are multimedia creations (Bransford and Franks, 1971; Zwaan *et al.* 2004; Zwaan, Chapter 9, this volume); and perceptual and motor areas in the brain become activated (Pulvermüller, Chapter 6, this volume), as was

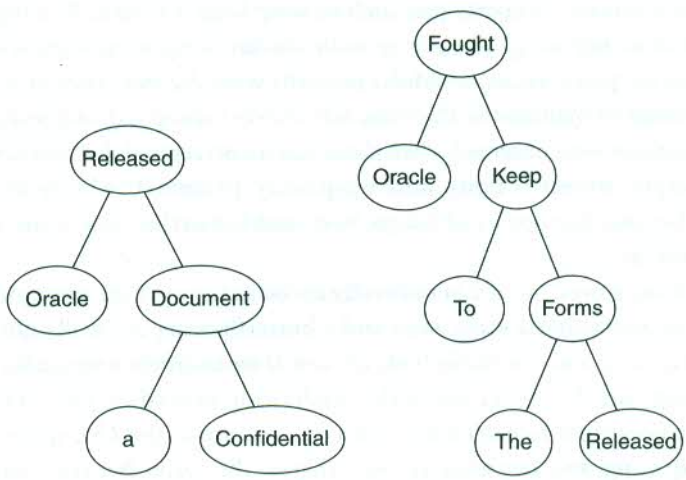

Fig. 8.7 Dependency trees for the inference sentences 'Oracle released a confidential document' and 'Oracle fought to keep the forms released.'

demonstrated by several participants in the debate. All levels of representation that humans are capable of appear to be functioning in a well coordinated chorus. How can a purely symbolic model, such as the one outlined above, do justice to these multilevel processes? The brief answer is that it does not and can not. The more interesting answer is that language has evolved in the service of perception and action, and that it is capable of mirroring aspects of perception and action. Yes, words can distort and bias meaning, but language would not be the great human success story it is if it were incapable of expressing what we experience faithfully.

There have been a large number of claims, such as that of Glenberg and Robertson (2000), that purely verbal (in other words, symbolic) representations of meaning are insufficient for modelling meaning. One the one hand this is a truism; human meaning representations are multilevel and not purely symbolic – no one can doubt that we live and act in a nonsymbolic world. On the other hand, the demonstrations of the shortcomings of verbal meaning are often misleading, based on much too crude a conception of verbal meaning. Language does a better job than it is given credit for by the critics of symbolic representations. Louwerse (2007 and Chapter 15, this volume) has argued this point with some compelling examples. I would like to add two more examples, one involving only word meanings, where LSA provides an account that closely mirrors perception, and one involving sentence meanings, where the construction–integration model is shown to mirror affordances.

Salomon and Barsalou (2001) convincingly demonstrated that what things look like is an important factor in verbal tasks of property verification. Concepts – certainly concrete concepts – are represented both perceptually and symbolically, which plays an important role not only in perception and action, but also when we talk and think about a concept. For example, English has only a single word 'wing', but a wasp wing and a butterfly wing look different, and this difference is reflected in the mental representation of these concepts. Solomon and Barsalou showed this in a priming study, where participants first verified either a concept–property pair such as 'wasp wing' or 'butterfly wing', and then a target pair such as 'bee wing'. The perceptually similar 'wasp wing' significantly primed 'bee wing', but the perceptually dissimilar butterfly wing did not. They also showed that perceptual similarity mattered in this case, not merely conceptual similarity, because no priming differences were observed when there was no perceptual discrepancy. In the case of their example, 'butterfly body' and 'wasp body' primed 'bee body' equally well – presumably because the bodies of wasps, bees, and butterflies look more or less alike, unlike their wings.

LSA cannot see wings, but it has indirectly encoded some of the perceptual information needed to know that a wasp wing and a butterfly wing are really quite different, whereas a wasp body and a butterfly body are not. If we compute a vector for 'wasp wing', 'butterfly wing', and 'bee wing' using the predication procedure described above, we can compute how close to each other these vectors are in the LSA space. The cosine between 'wasp wing' and 'bee wing' is 0.66, whereas the cosine between 'butterfly wing' and 'bee wing' is only 0.48. Hence, 'wasp wing' could be expected to be a better prime for

'bee wing' than 'butterfly wing'. On the other hand, both 'wasp body' and 'butterfly body' are equally close to 'bee body' in the LSA space. The cosines are 0.94 and 0.91, respectively – LSA really can't tell the difference between these concepts (and neither can most of us).[4]

I am not saying that the participants in the Salomon and Barsalou experiment did not employ perceptual representations when verifying properties; they probably did. My claim is simply that even a purely verbal symbolic system like LSA would behave in much the same way, because the verbal information mirrors perception to a considerable degree. This is also the case with my second example from Glenberg and Robertson (2000), who claim that purely symbolic, verbal representations cannot recognize perceptual affordances. Perceptual affordances are obviously based on perceptual and action experience, but these experiences are mirrored in our language, so that even a purely symbolic system like LSA or ICA can detect that some of the sentences below are more sensible than others in the context of a story of someone being caught in the rain. Consider the following sentences:

(a) Being clever, he walked into a store and bought a **newspaper** to cover his face.

(b) Being clever, he walked into a store and bought a **matchbox** to cover his face.

(c) Being clever, he walked into a store and bought a **ski mask** to cover his face.

Both (a) and (c) are possible ways to cover one's face, (c) much better than (a); (b) does not make sense. A ski mask or even a newspaper afford some protection from the rain, whereas a matchbox does not. The affordances inherent in an object are based on experiences, experience in the real world with newspapers and matchboxes, and experience with how the words 'newspaper' and 'matchbox' have been used. Glenberg and Robertson present an LSA analysis that fails to account for the difference between these and related sentences: if one computes the cosine between the bold-faced target words and the rest of the sentences in the examples above, they do not differentiate between sensible and nonsense alternatives (Table 8.2). However, when these sentences are analysed in terms of their propositional constituents, the difference in affordances emerges quite clearly.

Figure 8.8 shows the dependency tree for (a). The figure shows that the relevant proposition involves the target words and '(to) cover (his) face'. If you simply take the cosine between the three targets and the whole sentence, you would conclude that (b) was the best alternative and (c) the worst. Glenberg and Robertson have obtained sensibility ratings on a scale of 1 to 7 which show that (c) is highly sensible, (a) is less sensible, but (b) is clearly not. An ICA (or LSA) analysis that respects the propositional units obtained by a dependency analysis yields similar results, simply because, according to our experience with language, *cover-face* has a high cosine with 'ski mask', not quite as high a cosine with 'newspaper', and a low cosine with 'matchbox'.

[4] A more detailed analysis of the Solomon and Barsalou (2001) results is given in Kintsch (2007).

Table 8.2 Three examples from Glenberg and Robertson (2000) with LSA cosines between the sentence frames and the target items, sensibility ratings of the target items for each sentence frame, and ICA cosines between the target words and the relevant portion of the sentence frames

	LSA cosine (sentence)	Sensibility rating	ICA cosine (dependency)
'Being clever, he walked into a store and bought a _____ to cover his face.'			
newspaper	0.35	5.08	0.18
matchbook	0.42	1.16	0.03
ski mask	0.40	6.04	0.53
'She gave him a ——— to play with.'			
plastic spoon	0.48	5.58	0.07
large refrigerator	0.49	1.08	−0.03
red beanbag	0.39	6.29	0.59
'Adam pulled out of his golf bag a _____ and used that to chisel an inch of ice off his windshield.'			
seven iron	0.50	4.50	0.42
ham sandwich	0.56	1.00	0.16
screwdriver	0.60	5.00	0.63

ICA = integrated component analysis; LSA = latent semantic analysis.

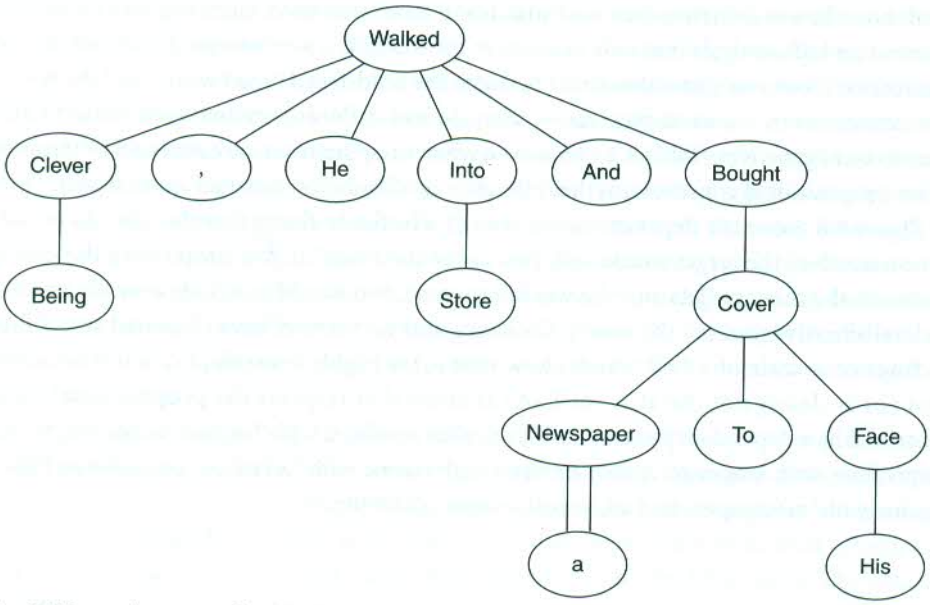

Fig. 8.8 Dependency tree for the sentence 'Being clever, he walked into a store and bought a newspaper to cover his face.'

Table 8.2 also shows two other examples from the Glenberg and Robertson study. For the sentence 'She gave him a _____ to play with,' people accept the target items 'red beanbag' and 'plastic spoon' as sensible, but not 'large refrigerator'. The dependency analysis suggests that the target item is to be compared with '(to) play (with)'. This analysis correctly yields a high cosine for 'red beanbag' and a low cosine for 'large refrigerator', but the cosine for 'plastic spoon' is not as high as it should be according to the sensibility rating participants provided.

For the sentence 'Adam pulled out of his golf bag a _____ and used that to chisel an inch of ice off his windshield,' the target items are 'seven iron', 'ham sandwich', and 'screwdriver'. The dependency analysis shows that 'and' dominates both *pull-[the target item]* and 'used that to chisel an inch of ice off his windshield.' This comparison correctly selects 'screwdriver' as the most sensible completion, and 'seven iron' as also possible, but rejects the 'ham sandwich', much like people do. Overall, ICA correctly selects the completion that is both related and afforded in all of the experimental examples and rejects the unrelated, unafforded alternative. The cosines for the afforded but unrelated alternative are more variable, as we have seen in the examples above. Perceptual representations obviously add something to symbolic representations, but purely symbolic representations do a remarkably good job of discriminating between perceptually afforded and unafforded alternatives.

Let me comment briefly on some other criticisms of symbolic representations that I think are misguided. The objection is not only to ungrounded symbols, but also to arbitrary symbols. Apparently, icons are regarded as somehow better, easier, and more natural than symbols. To comprehend meaning entails understanding the difference between a symbol or icon and the real thing; that is, *dual representation*. This may in fact be easier when the symbol is arbitrary (as words are) than when it is not (as pictures are). Children as old as 2.5 years are apt to mistake photographed objects for the object itself and treat small toys as if they were much larger (DeLoache 2004). They do not confuse an arbitrary sound with the object it represents, but do attempt to sit down on a miniature toy chair. Babies confuse pictures with the objects they depict, as seen in American babies who are familiar with pictures as well as Ivory Coast babies who are not. It is not at all trivial to learn that concrete, nonarbitrary icons represent something that they are not. Teaching math with manipulatives can be counterproductive because the children learn to manipulate objects without ever inferring abstract mathematical principles. The consequences of the confusion between objects and their representation can be serious, as when dolls are used in courtrooms to represent a child's body – but the child is unable to think of the doll as both a doll and a representation of herself (Ceci and Bruck 1995). Icons and analogue presentations are not necessarily easier than digital representations or arbitrary symbols.

The fact that symbolic processes involve the same brain areas as action and perception (e.g. Pulvermüller, Chapter 6, this volume) does not imply that symbolic processes and sensorimotor processes are the same. In the course of evolution nature discards very little, the new is generally built upon the old: fins turn into legs, and sensory brain areas become involved with symbol processing. But symbol processing is still at a different

level than sensorimotor processes. Other primates and humans are very similar in their sensorimotor processes, but differ radically in their symbolic processing. Humans operate at multiple levels of representations, from the most primitive ones shared by all animals to the symbolic level. But that does not justify reducing symbolic systems to sensorimotor processes.

Rejecting a reductionism that denies the autonomy of symbolic processes does not imply that we should not make every effort to gain a better understanding of how symbolic and nonsymbolic representations interact. To learn more about how perceptual representations and symbolic representations are coordinated and how they interact, is a very important goal, and considerable progress has been made as the chapters presented in this book attest, both at the level of brain processes, as in the work of Pulvermüller already cited, as well as the behavioural level (e.g., Zwaan, Chapter 9, this volume). This problem has also been addressed at the level of computational modelling. Goldstone *et al.* (2005) explicitly address the problem of connecting words to each other as well as to the world. Howell *et al.* (2005) have described a neural network model that simulates the acquisition of language based on prelinguistic concepts. A group of linguists and computer scientists at Berkeley have developed what they call a 'simulation semantics' to reflect their belief that much of language understanding involves embodied enactment (e.g., Feldman & Narayanan 2004; for more information see http://www.icsi.berkeley.edu/NTL). In simulation semantics, the mind simulates the world while functioning in it. Concrete action schemata, for instance, become symbolically extended. Understanding at the symbolic level may involve this kind of simulation, but it can also remain at the purely symbolic level – the world that LSA describes. Research like this serves as an existence proof that it is possible to model the interface between symbolic and nonsymbolic representations.

8.5 **Discussion**

The sensitivity of language to perceptual information (as well as emotion and action) should surprise no one. Most of the words we know we have learned from reading, which precludes a direct perceptual association. And for words we know at both the verbal level and the action–perception level – the crucial anchor words that link these different levels of representation – language has encoded in its own way the information it needs to mirror the world. I conclude with two quotes that Chomsky used in his discussion of language (Chomsky 1966) and that I have also cited in Kintsch (1998):

Les langues sont le meilleur miroir de l'esprit humain.

(Languages are the best mirror of the human mind.)

Gottfried Wilhelm Leibniz, 1765

Der Mensch lebt mit den Gegenständen hauptsächlich, ja, da Empfinden und Handeln in ihm von seinen Vorstellungen abhängen, sogar ausschliesslich so, wie die Sprache sie ihm zuführt.

(We interact with objects mainly as they are represented by language; indeed, exclusively so, since perception and action depend upon memory images.)

Wilhelm von Humboldt, 1792

Debate

Arthur Glenberg: I agree with you that there's an awful lot of structure that's available in covariance and analysed covariance as you presented it. But, in the spirit of this debate, what I'd like to hear from you is how we might determine whether people are using those sorts of representations; whether people are extracting information from language in similar ways.

Walter Kintsch: I make predictions about what people do, I observe the behaviour and see how well the predictions fit. That's all I can say.

Glenberg: And I agree with you, that the predictions do fit nicely. I think I need to formulate my question more incisively. The question really should be, how can we tell if this is a better theory than perhaps the embodied theory? In other words, where are the predictions that are different from an embodied theory?

Kintsch: Well, I don't think that that's so much the question, what's a better theory? I think most people are agreed that cognition has symbolic aspects and has embodied aspects, and that's worthwhile to study. It's not an either/or kind of thing. This is clearly what I wanted to show.

Friedemann Pulvermüller: Thanks very much for an excellent and very interesting talk. It is a very good thing to have these measures of relatedness of semantic distance between words and concepts, or family resemblance. However, the information that would not be included in this relationship – distance measures – would be, for example, that 'bark' is in one case an action and in the other case an object. And one wouldn't have directions from this very abstract representation about in which brain areas one would have to look. Of course, one could take a very abstract strategy and just say 'I take this description and make a correlation with brain activation.' But on the other hand, it might be a good thing to connect these descriptions, these semantic descriptions, with some information that links the representation to particular cognitive domains and brain systems as well, *a priori*. Obviously, an embodied description might have the advantage in providing that. These measures we are discussing, these covariance measures, they would also incorporate important information which a purely embodied theory might lack. So both sides might learn from each other.

Kintsch: Well, I think that suggestion is a very interesting one, and that's something we can actually look into. Because, if I have a lot of nouns and a lot of verbs, I can get them to cluster in very distinct clusters, so the verbs are here and the nouns are over there. Now, if you have a word like 'bark', which could be either one, I don't know what would happen. But if I can construct, now, the 'dog' meaning of 'bark', which is a verb, and the 'tree' meaning of 'bark', which is a noun, those two vectors should be in different clusters.

Pulvermüller: Sure, yes.

Kintsch: If I'm right. And it might also be interesting from a brain standpoint to see what happens with homonyms that have two different functions, how they are represented in two different places.

Pulvermüller: Right. Two overlapping styles is what we would propose, or have proposed. May I make another little comment on the syntax part? I've tried to argue that, using a model consisting of sequence detectors, it is actually possible to formulate dependency grammar. So I have a feeling that if you have the basic sequencing of words and on top of that a separate grammar dependency processor, there might be some redundancy.

Kintsch: I know, I was thinking whether that couldn't be put together. A very interesting thought.

Author note

Preparation of this report was supported by National Science Foundation grant #153-4712.

References

Barsalou L (1999). Perceptual symbol systems. *Behavioural and Brain Sciences*, 22, 577–660.

Bransford JD, Franks JJ (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331–50.

Bruner JS (1986). *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.

Ceci SJ, Bruck M (1995). *Jeopardy in the Courtroom: The Scientific Analysis of Children's Testimony*. Washington, DC: American Psychological Association.

Chomsky N (1966). *Cartesian Linguistics*. New York, NY: Harper.

DeLoache JS (2004). Becoming symbol-minded. *Trends in Cognitive Sciences*, 8, 66–70.

Donald M (1991). *Origins of the Modern Nind*. Cambridge, MA: Harvard University Press.

Feldman J, Narayanan S (2004). Embodied meaning in a neural theory of language. *Brain and Language*, 89, 385–92.

Franzke M, Kintsch E, Caccamise D, Johnson N, Dooley S (2005). Summary Street®: computer support for comprehension and writing. *Journal of Educational Computing Research*, 33, 53–80.

Glenberg AM, Robertson DA (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43, 379–401.

Goldstone RL, Feng Y, Rogosky B (2005). Connecting concepts to the world and each other. In D Pecher, RA Zwaan, Eds. *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking* (pp. 292–314). Cambridge: Cambridge University Press.

Harnad S (1990). The symbol grounding problem. *Physica D*, 42, 335–46.

Howell SR, Jankowicz D, Becker S (2005). A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning. *Journal of Memory and Language*, 53, 258–76.

Jones MN, Mewhort DJK (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1–37.

Jones MN, Kintsch W, Mewhort DJK (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, 55, 534–52.

Karmiloff-Smith A (1992). *Beyond Modularity*. Cambridge, MA: MIT Press.

Kintsch W (1974). *The Representation of Meaning in Memory*. Hillsdale, NJ: Erlbaum.

Kintsch W (1998). *Comprehension: A Paradigm for Cognition*. New York: Cambridge University Press.

Kintsch W (2001). Predication. *Cognitive Science, 25*, 173-202.

Kintsch W (2007). Meaning in context. In TK Landauer, D McNamara, S Dennis, W Kintsch, Eds. *The Handbook of Latent Semantic Analysis* (pp. 89–106). Mahwah, NJ: Erlbaum.

Kintsch W (in press). How the mind computes the meaning of metaphor: a simulation based on LSA. In R Gibbs, Ed. *Cambridge Handbook of Metaphor and Thought*. New York, NY: Cambridge University Press.

Kintsch W, Bowles AR (2002). Metaphor comprehension: what makes a metaphor difficult to understand? *Metaphor and Symbol*, 17, 249–62.

Landauer TK, Dumais ST (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211–40.

Landauer TK, McNamara D, Dennis S, Kintsch W, Eds (2006). *The Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.

Louwerse MM (2007). Symbolic and embodied representations: a case for symbol interdependency. In TK Landauer, D McNamara, S Dennis, W Kintsch, Eds. *The Handbook of Latent Semantic Analysis* (pp. 107–20). Mahwah, NJ: Erlbaum.

Lund K, Burgess C (1996) Producing high-dimensional semantic spaces from lexical co-occurrence. *Behaviour Research Methods, Instrumentation, and Computers*, 28, 203–8.

Mangalath P (2006). *Beyond Latent Semantic Analysis: Cognitive Component Resolution with Independent Component Analysis* [doctoral thesis]. Denver, CO: University of Colorado.

Nelson K (1996). *Language in Cognitive Development: The Emergence of the Mediated Mind.* New York, NY: Cambridge University Press.

J Nivre, J Hall, J Nilsson (2006). MaltParser: a data-driven parser-generator for dependency parsing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC2006)*, May 24–26, 2006, Genoa, Italy.

Paivio A (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76, 241–63.

Salomon KO, Barsalou LW (2001). Representing properties locally. *Cognitive Psychology*, 43, 129–69.

Shepard RN (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317–23.

Steyvers M, Griffiths T (2007). Probabilistic topic models. In TK Landauer, D McNamara, S Dennis, W Kintsch, Eds. *The Handbook of Latent Semantic Analysis* (pp. 427–48). Mahwah, NJ: Erlbaum.

Stone JV (2004). *Independent Component Analysis*. Cambridge, MA: MIT Press.

Wade-Stein D, Kintsch E (2004). Summary Street: interactive computer support for writing. *Cognition and Instruction*, 22, 333–62.

Yamada H, Matsumoto Y (2003). Statistical dependency analysis with support vector machines. *Proceedings of IWPT* (pp. 195–206). Nancy, France: IWPT.

Zwaan RA, Madden CJ, Yaxley RH, Aveyard ME (2004). Moving words: dynamic representations in language comprehension. *Cognitive Science*, 28, 611–19.