The Representation of Objects in the Brain, and Its Link with Semantic Memory and Language: a Conceptual Theory with the Support of a Neurocomputational Model

Cristiano Cuppini, Elisa Magosso and Mauro Ursino
DEIS, University of Bologna
Italy

1. Introduction

A fundamental problem in cognitive neuroscience is how the brain realizes semantic memory. It is generally accepted that semantic memory consists of stored information about the main features of an object, some processing mechanisms which allow a person to retrieve these features from partial cues, and to link them with lexical aspects and words. The final result is a kind of knowledge which is context independent, can be shared with other people and can produce language and thought (Tulving 1972).

Many conceptual theories of semantic memory have been proposed in past decades, based on two fundamental pieces of information: the behavior of patients with neurological lesions in specific brain areas, who exhibit deficits in word recognition, and results of more recent neuroimaging studies, putting in evidence which different brain areas participate to semantic tasks. The interested reader can find several excellent review papers on the subject (Martin & Chao 2001; Hart et al. 2007). Just the main fundamental issues, essential for the comprehension of the present chapter, are summarized below.

First, most theories agree in assuming that semantic memory is not a localized process, but one which involves a highly distributed representation of features and which engages several different cortical areas, located in the sensory and motor regions of the cortex. This concept may help explaining the existence of patients with category-specific semantic deficits (for instance, patients with impairment in word recognition for living things but no impairment for recognition of non-living objects, or patients with impairment for nouns vs. verbs (Warrington & Shallice 1984; Caramazza & Shelton 1998)). Damasio (Damasio 1989), suggested that semantic representation is fragmented in many motor and sensory features, which must then be integrated in a "convergence zone". Accordingly, Warrington et al. (Warrington & McCarthy 1987), assumed the presence of multiple channels which separately process sensory and motor aspects of objects; these two main systems would be especially important to identify living and non-living objects, respectively. Several subsequent extensions, improvements or variations of this theory were formulated by various groups (Caramazza et al. 1990; Lauro-Grotto et al. 1997; Humphreys & Forde 2001; Snowden et al. 2004; Gainotti 2006). Nevertheless, all these theories substantially agree in

assuming that the semantic system is realized by means of an integrated multimodal network, in which different areas store different modality-specific features. Some theories can also account for the emergence of categories from features: for instance, Tyler et al. (Tyler et al. 2000), in a conceptual model named "Conceptual Structure Account", suggested that objects are represented as patterns of activation across features, and that categories emerge from those objects that share common features and are highly correlated. Hart, Kraut et al. (Kraut et al. 2002; Hart et al. 2002) imagined that object representation is encoded not only in sensorimotor but also in higher-order cognitive areas (lexical, emotional, etc…) and all these representations are integrated via synchronized neural firing modulated by the thalamus. Barsalou et al. (Barsalou et al. 2003) assumed that groups of neurons are coactivated to represent collection of features, and that these groups of features encode progressively more generic information from sensory perception to higher cognitive associations. Moreover, they assumed a topography principle, according to which the spatial proximity of neurons reflects the similarity of the encoded features.

The previous conceptual theories, and others not listed here for briefness, although of the greatest value for cognitive neuroscientists, are just qualitative. In particular, the mechanisms responsible for integrating a distributed information into a coherent semantic representation, their physiological reliability in terms of neural structures, the learning rules for synaptic plasticity, are all aspects which deserve a more quantitative analysis. Modern neural network models, and computer simulation techniques now allow conceptual theories to be translated into a quantitative integrated system of neural units, and the consequent emergent behavior to be analyzed in detail, incorporating training aspects which mimic real synaptic plasticity rules.

A few previous models, based on neural networks with "hidden units" have attempted to reproduce how sensory information (for instance the visual one) can recall the linguistic information, and vice versa (Rogers et al. 2004) or to simulate category-specific semantic impairment (Small et al. 1995; McRae et al. 1997; Devlin et al. 1998; Lambon Ralph et al. 2007). All these models exploit a supervised algorithm (such as the back propagation) for training the network to solve the requested task: pathological deficits are then simulated assuming some damage in network synapses. A more sophisticate model, able to simultaneously retrieve multiple objects stored in memory, was presented by Morelli et al. recently (Morelli et al. 2006). In the model, features are coded by neurons which work in chaotic regimen, and the retrieval process is achieved via synchronization of neurons coding for the same object.

In recent years, we developed an original model (Ursino et al. 2006; Ursino et al. 2009; Cuppini et al. 2009) which aspires to explore several important issues of semantic memory, laying emphasis on the possible topological organization of the neural units involved, on their reciprocal connections and on synapse learning mechanisms. Some problems that the model aspires to investigate are: how can a coherent and complete object representation be retrieved starting from partial or corrupted information? How can this representation be linked to lexical aspects (words or lemmas) of language? How can different concepts be simultaneously recalled in memory, together with the corresponding words? How can categories be represented? How can category-specific deficits be at least approximately explained? What are the mechanisms exploited in bilingualism?

The model assumes that objects are represented via different multimodal features, encoded through a distributed representation among different cortical areas: each area is devoted to a

specific feature. Features are topologically organized (as in the conceptual model by Barsalou et al. (Barsalou et al. 2003)) and linked together by implementing two high-level Gestalt rules: similarity and previous knowledge. Multiple object retrieval is realized by means of synchronized activity of neural oscillators in the gamma-band (an idea often exploited in models for segmentation of visual or auditory scenes (von der Malsburg & Schneider 1986; Singer & Gray 1995; Wang & Terman 1997), and which reminds the conceptual Semantic Object Model by Kraut et al. (Kraut et al. 2002)). Finally, words are represented in a separate cortical area, and linked with the correct object representation via a Hebbian mechanism (Rolls & Treves 1998).

In the following, the model is first presented in a qualitative way, and some exemplary results, concerning object retrieval, connection between objects and words, and categories, are presented. All equations are reported in the Appendix. It is worth noting that the present model is significantly improved compared with previous versions (Ursino et al. 2009; Cuppini et al. 2009): the new aspects concern the possibility to represent objects with a different number of features (whereas a fixed number of features for all objects was used in previous works) and a more physiological mechanism to recognize words from objects. All new aspects are emphasized in the text.

2. Method and results

The model incorporates two networks of neurons, as illustrated in the schematic diagram of Fig. 1. These are briefly described below in a qualitative way, while all equations and parameters can be found in the Appendix.

1) *The feature network and its training* - The first network, named "feature network", is devoted to a description of objects represented as a collection of sensory-motor features. These features are assumed to spread along different cortical areas (both in the sensory and motor cortex) and are topologically organized according to a similarity principle. This means that two similar features activate proximal neural groups in the feature network.

The network is composed of N neural oscillators, subdivided into F distinct cortical areas (see Fig.1). Each area in the model is composed of N_1xN_2 oscillators. An oscillator may be silent, if it does not receive enough excitation, or may oscillate in the gamma-frequency band, if excited by a sufficient input. The presence of oscillators in this network is motivated by the necessity to have different objects simultaneously in memory, each represented by its collection of features (that is the classic binding and segmentation problem). As proposed by several authors in recent years, both experimentally (Singer & Gray 1995; Engel & Singer 2001), and theoretically (von der Malsburg & Schneider 1986; Wang & Terman 1997), binding and segmentation of multiple objects can be achieved in the brain via synchronization of neural oscillators in the gamma range.

Previous works demonstrated that, in order to solve the segmentation problem, a network of oscillating units requires the presence of a "global separator" (von der Malsburg & Schneider 1986; Wang & Terman 1997; Ursino et al. 2003). For this reason, the feature network incorporates a inhibitory unit which receives the sum of the whole excitation coming from the feature network, and sends back a strong inhibitory signal if this input exceeds a given threshold. In this way, as soon as a single object representation pops out in the network, all other objects representations are momentarily inhibited, avoiding superimposition of two simultaneous objects.
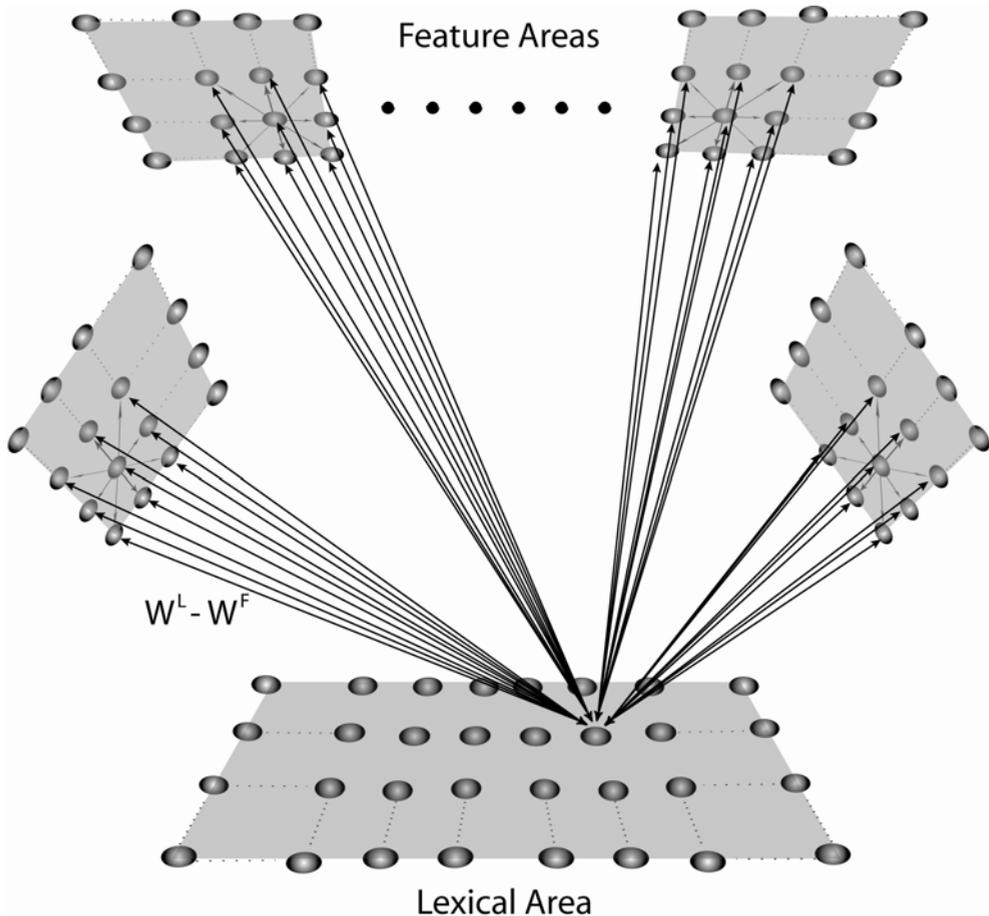
Fig. 1. Schematic diagram describing the general structure of the network. The model presents 9 distinct Feature Areas (upper shadow squares) of 20x20 elements, which are described by means of Wilson-Cowan oscillators, and a Lexical Area of 40x40 elements (lower shadow square), which are represented by a first order dynamics and a sigmoidal relationship. In the Feature network, each oscillator is connected with other oscillators in the same area via lateral excitatory and inhibitory intra-area synapses, and with other oscillators in different areas via excitatory inter-area synapses. Moreover, elements of the feature and lexical networks are linked via recurrent synapses (W^F, W^L).

During the simulation, a feature is represented by a single input localized at a specific coordinate of the network, able to trigger the oscillatory activity of the corresponding unit. We assume that these inputs are the result of upstream processing stages, that extracted the main sensory-motor properties of the objects. The way these features are extracted and represented in the sensory and motor areas is well beyond the aim of the present model.

The topological organization of each cortical area is realized assuming that each oscillator is connected with the other oscillators in the same area via lateral excitatory and inhibitory

synapses (*intra-area synapses*). These are arranged according to a mexican hat disposition, i.e., proximal neurons excite reciprocally and inhibit more distal ones. This disposition produces an "activation bubble" in response to a single localized feature input: not only the neural oscillator representing that individual feature is activated, but also the proximal ones linked via sufficient lateral excitation. This has important consequences for object recognition: neural oscillators in proximal positions share a common fate during the learning procedure. In fact, since learning occurs via a Hebbian procedure (see below), neuron oscillators that are simultaneously active are subject to a common synapse reinforcement, hence participate to the representation of the same object. In this way, an object can be recognized even in the presence of a moderate alteration in some of its features.

Throughout the following simulations, we assumed that the lateral intra-area synapses cannot be modified by experience, i.e., the similarity principle they implement is assigned "a priori". This is probably not true in the reality, since topological maps can be learned via classical Hebbian mechanisms (Rolls & Treves 1998; Haykin 1999). However, this choice is convenient to maintain a clear separation between different processes in our model (i.e. the implementation of the similarity principle on one hand and implementation of a previous knowledge principle on the other).

Besides the intra-area synapses, we also assumed the existence of excitatory long-range synapses between different feature areas (*inter-area synapses*). These are initially set at zero and are learned by experience during a training phase, in which individual objects (described by all their features) are presented to the network one by one. The learning rule is a time-dependent Hebbian rule, based on the correlation between the activity in the post-synaptic unit, and the activity in the pre-synaptic unit mediated over a previous 10 ms time-window (see (Markram et al. 1997; Abbott & Nelson 2000)).

To simplify the algorithm, in previous works we assumed that each object is described by a fixed number of features (four features in (Ursino et al. 2009; Cuppini et al. 2009)). Conversely, in the present version this constraint is removed, and we assume that the number of features describing a single object can vary from one object to the next. In the following exempla, the feature network will be subdivided into nine different cortical areas. Hence, an object can have up to nine different features. This limit has been introduced just to reduce the computational weight.

An example of the synaptic changes obtained from the learning procedure is shown in Fig. 2, which depicts the synapses targeting onto a neuron. Here an object is described by means of five different features, located in five different cortical areas. After training, neurons belonging to the five activation bubbles are linked by means of excitatory connections, and synchronize their oscillatory activity. In particular, a neuron coding for a single feature of the object after training receives excitatory synapses from four different bubbles of neurons, representing the other four features of the same object and their minimal variations.

In summary, intra-area lateral connections implement a similarity principle, while inter-area trained connections implement a "previous knowledge" principle.

An important aspect of our model is that, after training, multiple objects can be simultaneously recovered in memory and oscillate in time division with their frequency in the gamma-band. Moreover, an object can be recovered even if some features are lacking, or even if some features are reasonably altered compared with those of the prototypical object used during the learning phase. Fig. 3 shows an example in which three previously trained objects, characterized by three, five and seven features respectively, are simultaneously
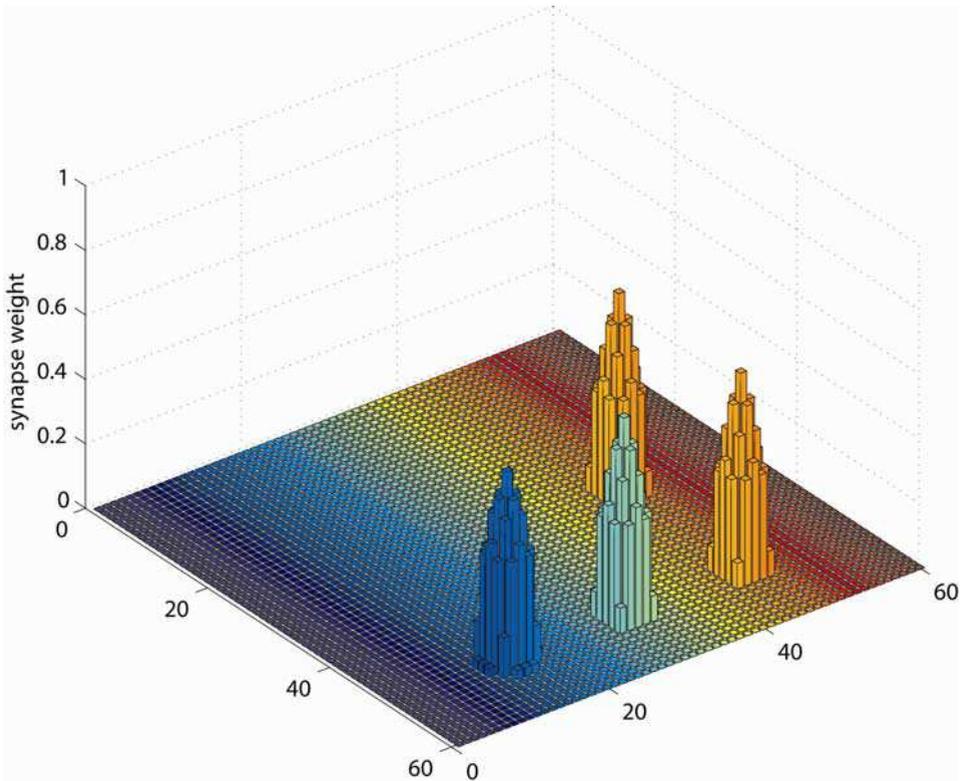
Fig. 2. An example of the inter-area synapses linking neurons in the feature areas, obtained after training a single object with five features, located at positions [50 15], [50 30], [30 30], [50 45], and [30 45]. The figure represents the synapses entering the neuron at position [30,30]. It is worth noting that this neuron, representing an individual feature of the object, receives synapses from four different bubbles, each centered at the remaining features of the object.

presented to the network. Moreover, as described in the figure legend, the objects are recovered despite the presence of incomplete information (some features are lacking) and corrupted data (some features are slightly changed).

2) *The lexical network and its training* – In order to associate objects with words, the model includes a second layer of neurons, denoted "lexical network". Each computational unit in this network codes for a word (or a lemma) and is associated with an individual object representation. Even for what concerns this network (as the previous one), the input must be considered as the result of an upstream processing stream, which recognizes the individual words from phonemes or from written texts. Description of this processing stream is well beyond the aim of this model: some exempla can be found in recent works by others (Hopfield & Brody 2001). Moreover, units in this network can also be stimulated through long-range synapses coming from the feature network; hence the network represents an amodal convergence zone, as often hypothesized in the anterior temporal lobe (Damasio 1989; Snowden et al. 2004; Ward 2006).
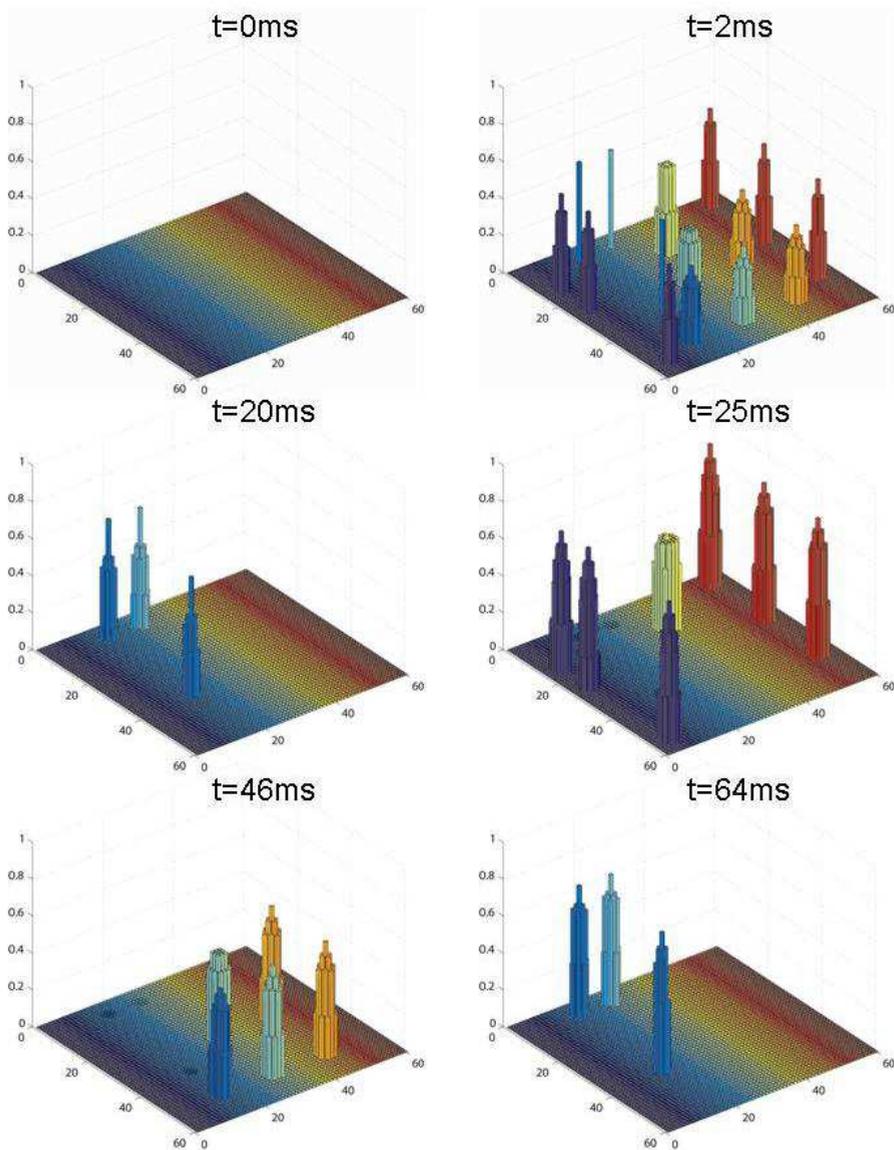
Fig. 3. Snapshots showing the activities in the feature areas at some instants of the simulation, after the presentation of three different objects: object1, three features at positions [6 17], [6 26], [37 16]; object2, five features at positions [50 15], [50 30], [30 30], [50 45], [30 45]; object 3, seven features at positions [15 5], [15 35], [55 5], [25 5], [45 55], [5 55] [25 55]. The objects were learned during a previous training phase. During the simulation, one property of the object1 was shifted compared with the normal one, while one property of both object2 and object3 was lacking. Despite this corrupted/lacking information, the network can reconstruct and segment the three objects.

For the sake of simplicity, computational units in this network are described via a simple first-order dynamics and a non-linear sigmoid relationship. Hence, if stimulated with a constant input, these units do not oscillate but, after a transient response, reach a given steady-state activation value (but, of course, they oscillate if stimulated with an oscillating input coming from the feature network).

In order to associate words with their object representation, we performed a second training phase, in which the model receives a single input to the lexical network (i.e., a single word is detected) together with the features of a previously learned object. Synapses linking the objects with words, in both directions (i.e., from the lexical network to the feature network and viceversa) are learned with Hebbian mechanisms.

While synapses from words to features ($W_{ij,hk}^{L}$ in Fig. 1) are simply excitatory and are trained on the basis of the pre and post synaptic correlation, when computing the synapses from features to words ($W_{ij,hk}^{F}$ in Fig. 1) we tried to address two major requirements, that are essential for correct object recognition. First, a word must be evoked from the corresponding object representation only if all its features are simultaneously on. This corresponds to a correct solution of the binding problem. Second, the word must not be evoked if spurious features (not originally belonging to the prototypical object) are active. This second situation may occur when two or more objects, simultaneously present, are not correctly segmented, and some of their features pop up together. Hence, the second requirement corresponds to a correct solution of the segmentation problem.

In order to address these two requirements, in previous works we implemented a complex "decision network" (see (Ursino et al. 2009)). Conversely, in the present model we adopted a more straightforward and physiologically realistic solution. First, we assumed that, before training, all units in the feature network send strong inhibitory synapses to all units in the lexical network. Hence, activation of any feature potentially inhibits all lexical units. These synapses are then progressively *withdrawn* during the training phase, on the basis of the correlation between activity in the feature unit and in the lexical unit. The consequence of this choice is that, after training, a word receives inhibition from all features that do not belong to its object representation, but no longer receives inhibition from its own feature units.

Moreover, we assume that all feature units can send excitatory synapses to lexical units: these are initially set at zero and are *reinforced* via a Hebbian mechanism. Moreover, we assumed that excitatory synapses from features to words are subject to an upper saturation level, i.e., *the sum* of all excitatory synapses reaching a lexical unit must not overcome a maximum level. This is a physiological rule, since the amount of neurotransmitter available at a neuron is limited. This rule warrants that, after prolonged training, the sum of synapses entering a lexical unit is constant, independently of the number of its associated features.

Using quite a sharp sigmoidal characteristic for lexical units, the previous two rules ensure that a word in the lexical network is excited if and only if all its features are simultaneously active: if even a single feature is not evoked, the word does not receive enough excitation (failure of the binding problem); if even a spurious feature pops up, the lexical unit receives excessive inhibition (failure of the segmentation problem): in both conditions it does not jump from the silent to the excited state.

Two exempla of model behavior are shown in Figs. 4 and 5. In the first figure, one word and two objects are simultaneously given to the network. The word evokes the corresponding object representation in the feature network, while the objects evoke the corresponding words in the lexical network. It is worth noting that all objects in the feature network oscillate in time division, and thus are all individually recognized. In the second figure, two

incomplete objects are given to the feature network. The first is characterized by five features, but only four of them are given as input. However, these four features are sufficient to recover the whole object representation and the corresponding word is activated in the lexical area. Conversely, the second object is characterized by seven features, while only three features are given as input. These are insufficient to recover the overall object representation, and the corresponding word is not activated (i.e., the subject did not recognize the object starting from such an incomplete information). It is worth noting that the number of incomplete information necessary to recover the whole object depends on the strength of inter-area synapses in the feature network, hence on the duration of the previous training period. In our simulations, we always assumed that, after training, at least 50% or more of the object features are required to attain the overall object reconstruction.
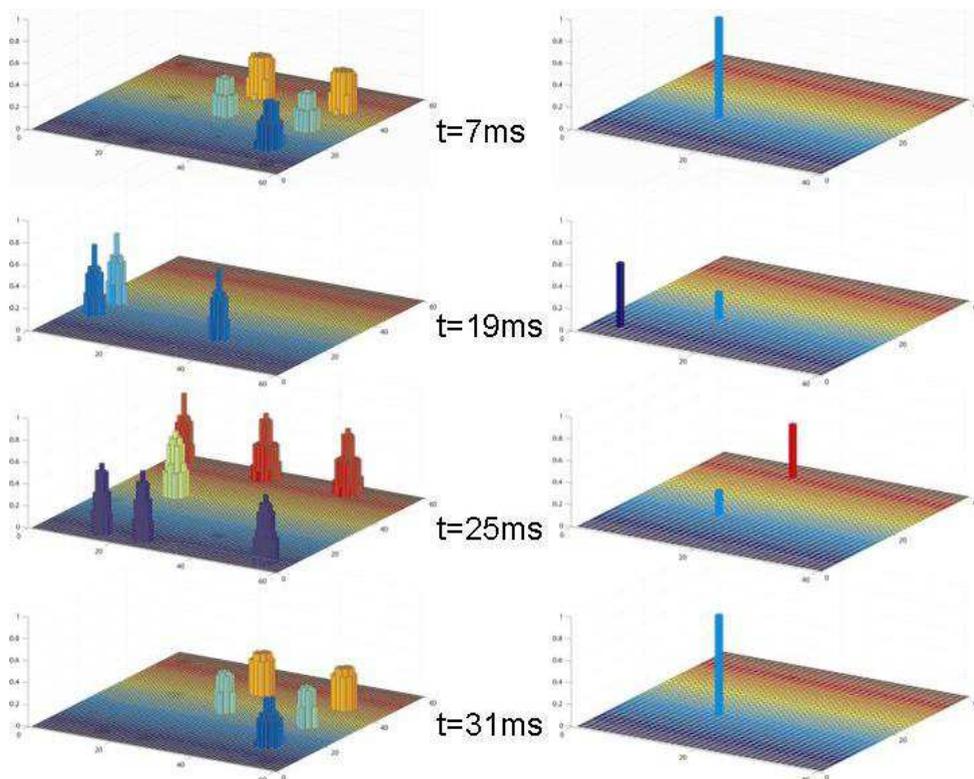

Fig. 4. Snapshots showing the activities in the feature areas (left panels) and in the lexical area (right panels) at some instants of the simulation, after the presentation of two different objects and of one word. Objects and words were previously learned during the training phases. Objects are the same as in Fig. 3. The associated words were located at position [5,5] (object1), [15,15] (object2) and [15,35] (object3). During the simulation, object1 and object3 were given as input to the feature areas, while the word2 was given as input to the lexical area. It is worth noting that the three object representations oscillate in time division in the feature areas, while the three words are evoked in the lexical area. Word2 (constantly given as input) is partially inhibited during the appearance of the other two words.
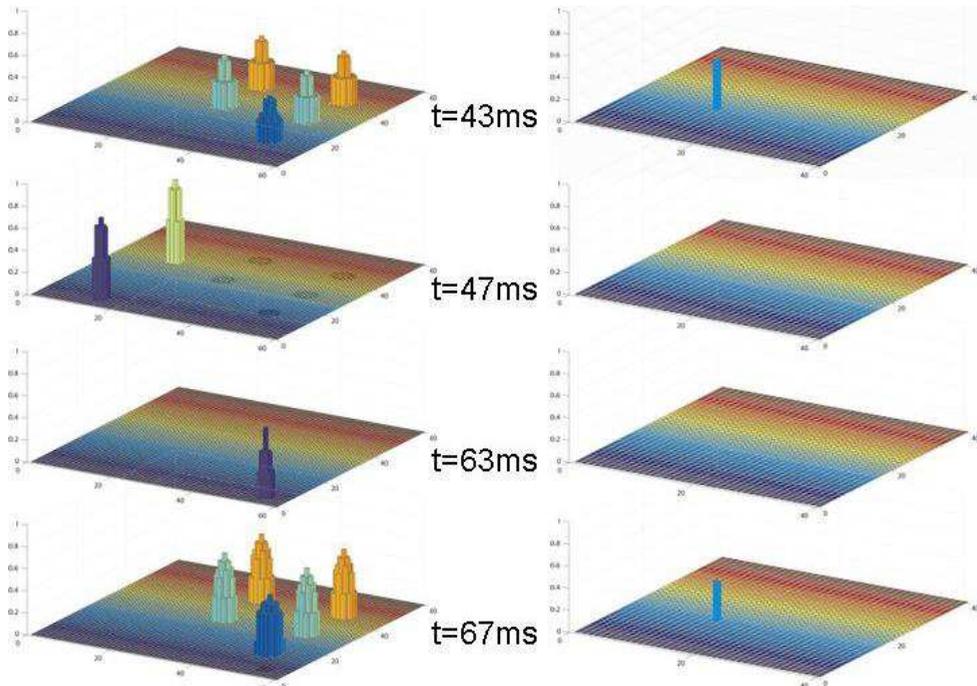
Fig. 5. Snapshots showing the activities in the feature areas (left panels) and in the lexical area (right panels) at some instants of the simulation, after the presentation of four features of object2, and just of three features of object3. As a consequence, object2 is correctly reconstructed in the feature areas, and the corresponding word is evoked in the lexical area. Conversely, object3 is not reconstructed (its three features given as input do not synchronize nor evoke the remaining four features), and the corresponding word fails to appear in the lexical area.

The structure of the model can also be used to study the formation of categories starting from a distributed representation of features. A widely shared idea, in fact, is that a category can be realized by objects which share some common features (for instance, "dog" and "cat" belong to the category "pet" and have many common characteristics). This idea is supported by experiments on the so-called "semantic priming", i.e., object recognition can be modulated by the previous recognition of another object which is "semantic congruent" (Rossell et al. 2003; Matsumoto et al. 2005). The explanation may be that the two objects activate some common neural structures, resulting in a classic priming phenomenon.

Category formation can be reproduced in our model by simply assuming objects that have some common features, and assuming that these common features are associated with a specific word, having a more general meaning (i.e., denoting a category).

Let us consider an example in which two objects (for convenience, "cat" and "dog"), with seven features each, are trained and associated with two different words. Moreover, we assumed that these two objects have four common features (representing the common characteristics of "pets"). It is worth noticing that, after the training phase 1, the inter-area synapses in the feature network are the sum of synapses learned during each object

presentation. As a consequence, the four common features are linked together by means of stronger synapses compared with the remaining specific features distinguishing cats from dogs, which are more weakly linked together and to the other four features. Hence, we have an irregular pattern of synapses (Fig. 6).
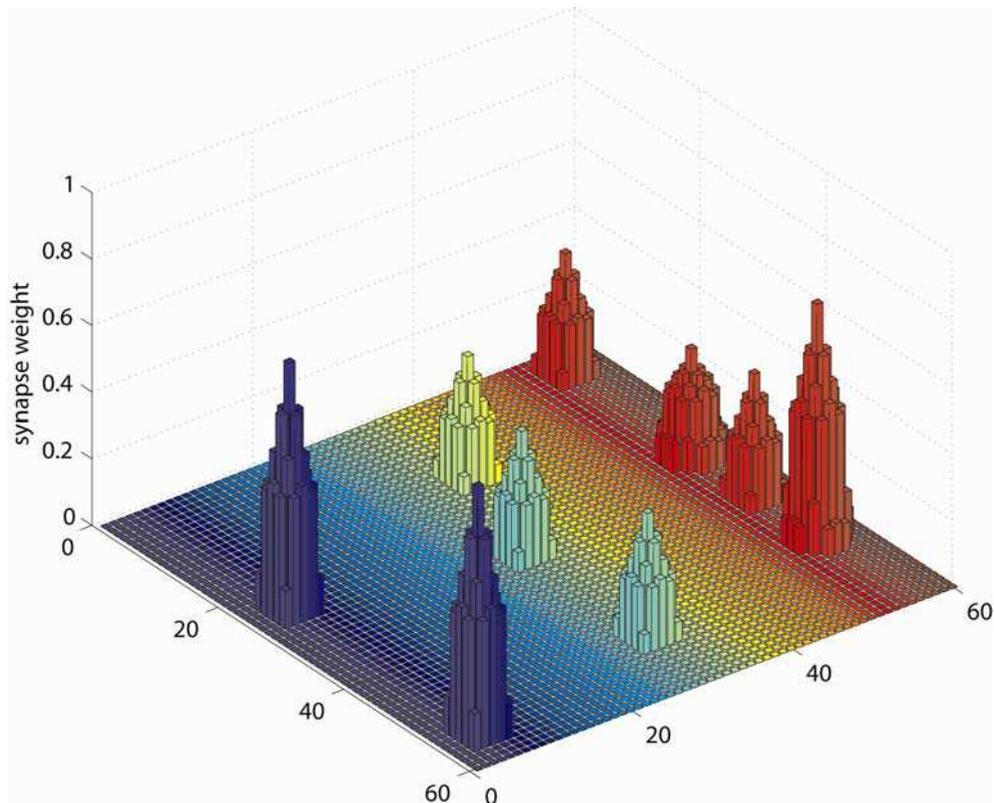

Fig. 6. An example of the inter-area synapses linking neurons in the feature areas, obtained after training two objects with a significant number of common features. In particular, the first object ("cat") has seven features located at positions [15 5], [15 35], [55 5], [25 5], [45 55], [5 55], [25 55]; the second object ("dog") has seven features at positions [15 5], [30 30], [55 5], [25 5], [45 55], [50 30], [35 55] (the first, third, fourth and fifth features are in common and may represent "pets"). The figure shows the synapses entering the neuron at position [15,5]. It is worth noting that this neuron, representing a common feature, receives synapses from nine different bubbles, each centered at the remaining features of the two objects. Three synaptic bubbles (coming from the other three common features [55 5], [25 5], [45 55]) are much stronger; other six synaptic bubbles (coming from the remaining six specific features) are weaker. The stronger synapses contribute to the category representation.

We assumed that, after the first training phase, the four common features alone are unable to recover the three remaining features (otherwise, any pet would retrieve the words "dog" and "cat"). During the second training phase, the features representing objects "dog"(7

features), "cat" (7 features) and "pet" (4 common features) were separately given to the network together with the corresponding words, to generate three distinct lexical links.

Simulation results are presented in Fig. 7. In this figure, all seven features describing a cat were initially given to the network (during the first half of the simulation). In this condition, the corresponding word is activated in the lexical area (a similar result, of course, would be obtained giving seven features of "dog" to the network). It is worth noticing that the word "pet" is not active in the lexical area, due to inhibition coming from an excessive number of features. Conversely, when the number of features given to the network is reduced to four (second part of the simulation), all belonging to the category "pet", the network does not recall the features specific of cats and dogs, and the word "pet" is now emerging in the lexical area, without the emerging of the words "dog" and "cat".
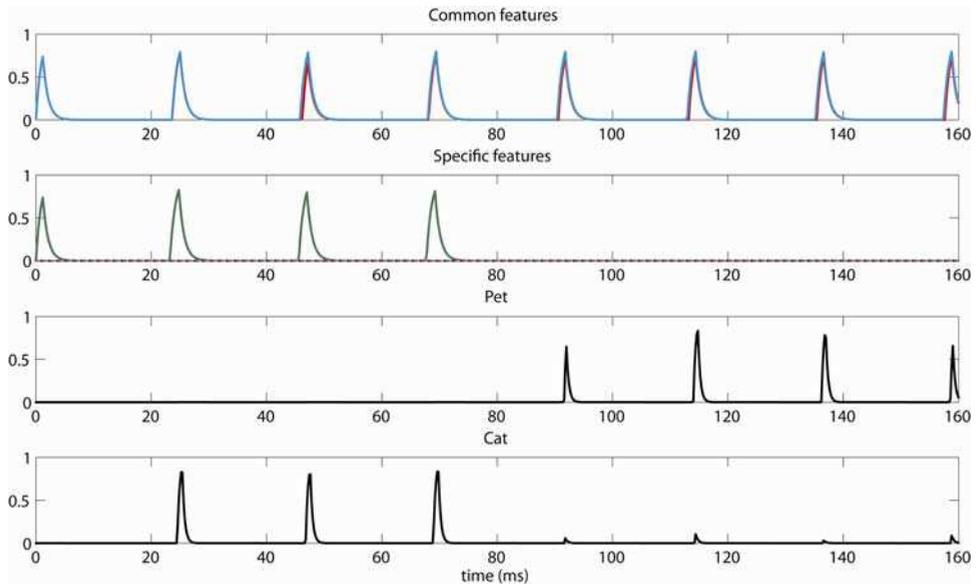

Fig. 7. Time pattern of neuron activities in the feature areas (upper panels) and in the lexical area (bottom panels) during a simulation, in which the seven features of the object "cat" (see legend of fig. 6) were given as input to the feature areas during the first 100 ms; at 100 ms, the three specific features were set at zero, and only the four common features remained as input. The upper panel shows the oscillatory activity of the four neurons in the feature areas representing the common attributes of "pets". They oscillate in synchronism throughout the simulation. The second panel represents the synchronized activities of the three neurons representing the specific attributes of cats. The third panel represents the neuron in the lexical areas coding for "pet". The fourth panel represents the activity in the lexical area of the neuron coding for "cat". It is worth noting that the word "cat" is inhibited, and the word "pet" excited, as soon as the specific features are withdrawn.

3. Discussion

The present work intends to summarize several different ideas on semantic memory, appeared in the neurocognitive and psycholinguistic literature over past years, into a coherent and

comprehensive neural network model. The main points that characterize model functioning are briefly discussed below, together with their neurophysiological support:

i) The model assumes that binding and segmentation of multiple objects occur via synchronization of neural activity in the gamma-band. This idea, originally proposed with reference to vision problems (Singer & Gray 1995; Engel & Singer 2001), is now widely supported also for what concerns high-level cognitive problems. For instance, a role of gamma-activity has been demonstrated in recognition of music (Bhattacharya et al. 2001), faces (Rodriguez et al. 1999), as well as during visual search tasks (Tallon-Baudry et al. 1997) and delayed-matching-to-sample-tasks (Tallon-Baudry et al. 1998). ii) Each object is described by means of a different number of features, which spread over different cortical areas. This is quite a common idea in conceptual models of semantic memory (Caramazza et al. 1990; Gainotti 2006; Hart et al. 2007). iii) Features are topographically organized. Results supporting this idea can be found in recent works by Barsalou et al. (Simmons & Barsalou 2003; Barsalou et al. 2003). iv) Knowledge of previous objects is stored in the model by means of inter-area synapses, which realize excitatory links learned via a Hebbian mechanism. Hence, it is sufficient that several features of an object occurred together for a sufficient long period in the past, for the creation of a permanent link. v) Words are represented in a different cortical area, separate from features. Although it is difficult to find a specific cortical location for this area, the existence of cortical regions especially devoted to lexical aspects of language has been hypothesized in cognitive neuroscience for decades (Ward 2006). vi) Links between lexical aspects (words) and semantic aspects (i.e., object representation) are learned via Hebbian mechanisms too. This requires the simultaneous presentation of an object with its representative word. vii) Objects can evoke words only if all their features have been correctly restored in memory and segmented from features of other objects. In other terms, the present model implies that a complete semantic recognition is a prerequisite for evocation of words. viii) The presentation of a word (from phonemes or from written texts) is able to evoke the representation of the object. Some recent data in the neurophysiologic literature support this idea: presentation of an action word can evoke activity in the motor and premotor cortex (Pulvermller et al. 2005a; Pulvermller et al. 2005b) and presentation of a smell can activate olfactory areas (González et al. 2006). ix) Categories can be represented by features which belong to different objects simultaneously, assuming that these shared features can be associated with a new word, and that activities of these features, when presented alone, remain bounded without spreading to reconstruct the original individual objects. x) The relationships from objects to words require the presence of both excitatory and inhibitory mechanisms, to evoke objects separately from their category, or to avoid that the presence of an excessive number of features (as in the failure of the segmentation task) evokes erroneous words.

Using the previous basic ideas, the model is able to simulate semantic memory and its link with lexical aspects in a variety of conditions which, although drastically simplified compared with the reality, can provide some cues to drive future ideas and to test the reliability of existing theories.

The present simulations (recognition of different simultaneous words and objects, even in the case of absent or corrupted features) represent just a few aspects of the potential model applications. Future challenges may be concerned with the following major issues, which have not been explicitly treated here due to space limitations:

Semantic relationships among words - An important problem that can be simulated with the model consists in the semantic priming, i.e., the possibility that a previous word or a

previous object (a cue) may affect (facilitate or depress) recognition of a subsequent word or subsequent object (a target) which is "semantically congruent" (Rossell et al. 2003; Matsumoto et al. 2005). This sort of priming mechanism, which may have important implication in language, may be simulated assuming that the two semantically congruent objects share some of their activated features, and that the representation of the first object is still partly active when the second object is presented to the network.

Bilingualism - A further aspect which may be simulated with the model is the lexical organization in bilingualism. This may be simulated assuming that two words in the lexical area (i.e., a first word already learned in a native language, say L1, and a second word in a new language, say L2) are associated with the same object representation in the feature area. During the learning procedure of the second language, the L2 word may exploit the already existing links between the L1 word and the object representation, to create its own excitatory synapses. As commonly suggested in the psycholinguistic literature (Abutalebi 2008), the new language may depend on L1 to mediate access to its object representation, i.e., L2 words are generally acquired with reference to existing L1 concepts. In the final bilingual subject, however, who exhibits high proficiency for L2, managing bilingualism requires the addition of further competitive mechanisms, and sophisticate control strategies (Green 1998), which allow the selection of the chosen word (or language) by inhibiting the other one. The interested reader can consult (Green 1998; Abutalebi 2008) for conceptual theories on the subject.

Lexical deficits – The model can be used to simulate patients with category-specific lexical deficits, i.e., patients unable to recognize certain categories of objects (Warrington & McCarthy 1983; Warrington & Shallice 1984; Warrington & McCarthy 1987; Humphreys & Forde 2001). To this end, one may suppose that only synapses in certain feature areas are weakened (for instance, as a consequence of a local lesion) thus resulting in a deficit for those words and those objects only, which make an intensive use of these areas.

These last aspects are just exempla of how the model may have a large applicative domain in future research. Its validation, amelioration and extension, however, and its use for the analysis and the theoretical formalization of different semantic/lexical problems, will necessarily require a strong multidisciplinary approach. This should entrain researchers in different domains: such as neurophysiologists, cognitive neuroscientists, experts of psycholinguistics, mathematicians and neuro-engineers.

4. Appendix

4.1 The bidimensional network of features

In the following, each oscillator will be denoted with the subscripts *ij* or *hk*. In the present study we adopted an exemplary network with 9 areas ($F = 9$) and 400 neural groups per area ($N_1 = N_2 = 20$).

Each single oscillator consists of a feedback connection between an excitatory unit, x_{ij}, and an inhibitory unit, y_{ij} while the output of the network is the activity of all excitatory units. This is described with the following system of differential equations

$$\frac{d}{dt}x_{ij}(t) = -x_{ij}(t) + H\left(x_{ij}(t) - \beta \cdot y_{ij}(t) + E_{ij}(t) + V_{ij}^L(t) + I_{ij} - \varphi_x - z(t)\right) \tag{1}$$

$$\frac{d}{dt}y_{ij}(t) = -\gamma \cdot y_{ij}(t) + H\left(\alpha \cdot x_{ij}(t) - \varphi_y\right) + J_{ij}(t) \tag{2}$$

where H() represents a sigmoidal activation function defined as

$$H(\psi) = \frac{1}{1 + e^{-\frac{\psi}{T}}} \tag{3}$$

The other parameters in Eqs. (1) and (2) have the following meaning: α and β are positive parameters, defining the coupling from the excitatory to the inhibitory unit, and from the inhibitory to the excitatory unit of the same neural group, respectively. In particular, α significantly influences the amplitude of oscillations. Parameter γ is the reciprocal of a time constant and affects the oscillation frequency. The self-excitation of x_{ij} is set to 1, to establish a scale for the synaptic weights. Similarly, the time constant of x_{ij} is set to 1, and represents a scale for time t. φ_x and φ_y are offset terms for the sigmoidal functions in the excitatory and inhibitory units. I_{ij} represents the external stimulus for the oscillator in position ij, coming from the sensory-motor processing chain which extracts features. E_{ij} and J_{ij} represent coupling terms (respectively excitatory and inhibitory) from all other oscillators in the features network (see Eqs. 5-8), while V_{ij}^L is the stimulus (excitatory) coming from the lexical area (Eq. 9). $z(t)$ represents the activity of a global inhibitor whose role is to ensure separation among the objects simultaneously present. This is described with the following algebraic equation:

$$z = \left[sign\left(\sum_i \sum_j x_{ij} - \theta_z \right) + 1 \right] \Big/ 2 \tag{4}$$

According to Eq. 4, the global inhibitor computes the overall excitatory activity in the network, and sends back an inhibitory signal ($z = 1$) when this activity overcomes a given threshold (say θ_z). This inhibitory signal prevents other objects from popping out as long as a previous object is still active.

The coupling terms between elements in cortical areas, E_{ij} and J_{ij} in Eqs. (1) and (2), are computed as follows

$$E_{ij} = \sum_h \sum_k W_{ij,hk} \cdot x_{hk} + \sum_h \sum_k L_{ij,hk}^{EX} \cdot x_{hk} \tag{5}$$

$$J_{ij} = \sum_h \sum_k W_{ij,hk} \cdot x_{hk} + \sum_h \sum_k L_{ij,hk}^{IN} \cdot x_{hk} \tag{6}$$

where ij denotes the position of the postsynaptic (target) neuron, and hk the position of the presynaptic neuron, and the sums extend to all presynaptic neurons in the feature area. The symbols $W_{ij,hk}$ represent inter-area synapses, subjects to Hebbian learning (see next paragraph), which favour synchronization. The symbols $L_{ij,hk}^{EX}$ and $L_{ij,hk}^{IN}$ represent lateral excitatory and inhibitory synapses among neurons in the same area. It is worth noting that all terms $L_{ij,hk}^{EX}$ and $L_{ij,hk}^{IN}$ with neurons ij and hk belonging to *different* areas are set to zero. Conversely, all terms $W_{ij,hk}$, linking neurons ij and hk in the *same* area, are set to zero.

The Mexican hat disposition for the intra-area connections has been realized by means of two Gaussian functions, with excitation stronger but narrower than inhibition. Hence,

$$L_{ij,hk}^{EX} = \begin{cases} L_0^{EX} e^{-\left[(i-h)^2 + (j-k)^2\right]/(2\sigma_{ex}^2)} & \text{if } ij \text{ and } hk \text{ are in the same area} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$L_{ij,hk}^{IN} = \begin{cases} L_0^{IN} e^{-\left[(i-h)^2+(j-k)^2\right]/(2\sigma_{in}^2)} & \text{if } ij \text{ and } hk \text{ are in the same area} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where L_0^{EX} and L_0^{IN} are constant parameters, which establish the strength of lateral (excitatory and inhibitory) synapses, and σ_{ex} and σ_{in} determine the extension of these synapses.

Finally, the term V_{ij}^L coming from the lexical area is calculated as follows

$$V_{ij}^L = \sum_h \sum_k W_{ij,hk}^L \cdot x_{hk}^L \tag{9}$$

where x_{hk}^L represents the activity of the neuron hk in the lexical area and the symbols $W_{ij,hk}^L$ are the synapses from the lexical to the feature network (which are subject to Hebbian learning, see below).

4.2 The bidimensional lexical area

In the following each element of the lexical area will be denoted with the subscripts ij or hk (i, h = 1, 2, …, M_1; j, k = 1,2,…, M_2) and with the superscript L. In the present study we adopted $M_1 = M_2 = 40$. Each single element exhibits a sigmoidal relationship (with lower threshold and upper saturation) and a first order dynamics (with a given time constant). This is described via the following differential equation:

$$\tau^L \cdot \frac{d}{dt} x_{ij}^L(t) = -x_{ij}^L(t) + H^L\left(u_{ij}^L(t)\right) \quad ; \tag{10}$$

τ^L is the time constant, which determines the speed of the answer to the stimulus, and $H^L\left(u^L(t)\right)$ is a sigmoidal function. The latter is described by the following equation:

$$H^L\left(u^L(t)\right) = \frac{1}{1 + e^{-\left(u^L(t)-\vartheta^L\right)p^L}} ; \tag{11}$$

where ϑ^L defines the input value at which neuron activity is half the maximum (central point) and p^L sets the slope at the central point. Eq. 11 conventionally sets the maximal neuron activity at 1 (i.e., all neuron activities are normalized to the maximum).

According to the previous description, the overall input, $u_{ij}^L(t)$, to a lexical neuron in the ij-position can be computed as follows

$$u_{ij}^L(t) = I_{ij}^L(t) + V_{ij}^F \tag{12}$$

$I_{ij}^L(t)$ is the input produced by an external linguistic stimulation. V_{ij}^F represents the intensity of the input due to synaptic connections from the feature network; this synaptic input is computed as follows:

$$V_{ij}^F = \sum_h \sum_k W_{ij,hk}^F \cdot x_{hk} \tag{13}$$

where x_{hk} represents the activity of the neuron hk in the Feature Areas (see Eq. 1) and $W_{ij,hk}^{F}$ the strength of synapses. These synapses may have both an excitatory and an inhibitory component (say $W_{ij,hk}^{Fex}$ and $W_{ij,hk}^{Fin}$, respectively) which are trained in different ways (see session, "synapse training: phase 2", below). Hence, we can write

$$W_{ij,hk}^{F} = W_{ij,hk}^{Fex} - W_{ij,hk}^{Fin} \tag{14}$$

4.3 Synapses training

Phase 1: Training of inter-area synapses within the feature network

In a first phase, the network is trained to recognize objects without the presence of words.
Recent experimental data suggest that synaptic potentiation occurs if the pre-synaptic inputs precede post-synaptic activity by 10 ms or less (Markram et al. 1997; Abbott & Nelson 2000). Hence, in our learning phase we assumed that the Hebbian rule depends on the present value of post-synaptic activity, $x_{ij}(t)$, and on the moving average of the pre-synaptic activity (say $m_{hk}(t)$) computed during the previous 10 ms. We define a moving average signal, reflecting the average activity during the previous 10 ms, as follows

$$m_{hk}(t) = \frac{\sum_{m=0}^{N_s-1} x_{hk}(t - mT_S)}{N_s} \tag{15}$$

where T_S is the sampling time (in milliseconds), and N_S is the number of samples contained within 10 ms (i.e., $Ns = 10/T_S$). The synapses linking two neurons (say ij and hk) are then modified as follows during the learning phase

$$\Delta W_{ij,hk}(t + T_S) = W_{ij,hk}(t) + \beta_{ij,hk} \cdot x_{ij}(t) \cdot m_{hk}(t) \tag{16}$$

where $\beta_{ij,hk}$ represents a learning factor.

In order to assign a value for the learning factor, $\beta_{ij,hk}$, in our model we assumed that inter-area synapses cannot overcome a maximum saturation value. This is realized assuming that the learning factor is progressively reduced to zero when the synapse approaches its maximum saturation. Furthermore, neurons belonging to the same area cannot be linked by a long-range synapse. We have

$$\beta_{ij,hk} = \begin{cases} \beta_0 \left(W_{\max} - W_{ij,hk} \right) & \text{if } ij \text{ and } hk \text{ belong to different areas} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

where W_{max} is the maximum value allowed for any synapse, and $\beta_0 W_{\max}$ is the maximum learning factor (i.e., the learning factor when the synapse is zero).

Phase 2: Training of long-range synapses among the Lexical and the Feature Networks

These synapses are trained during a second phase, in which an object is presented to the network together with its corresponding word.
Synapses from the lexical network to the feature network (i.e., parameters $W_{ij,hk}^{L}$ in Eq. 9) are learned using an Hebbian rule similar to that used in Eqs. 16 and 17. We can write

$$W_{ij,hk}^{L}(t + T_S) = W_{ij,hk}^{L}(t) + \beta_{ij,hk}^{L} \cdot x_{ij}(t) \cdot m_{hk}^{L}(t) \tag{18}$$

where $\beta^{L}_{ij,hk}$ represents the learning factor and $m^{L}_{hk}(t)$ is the averaged signal:

$$m^{L}_{hk}(t) = \frac{\sum_{m=0}^{N_s-1} x^{L}_{hk}(t-mT_s)}{N_s} \tag{19}$$

$$\beta^{L}_{ij,hk} = \beta^{L}_{0}\left(W^{L}_{\max} - W^{L}_{ij,hk}\right) \tag{20}$$

Conversely, synapses from the feature network to the lexical network (i.e., parameters $W^{F}_{ij,hk}$ in Eq. 13) include both excitatory and inhibitory contributions:

$$W^{F}_{ij,hk}(t) = W^{Fex}_{ij,hk}(t) - W^{Fin}_{ij,hk}(t) \tag{21}$$

The excitatory portion is trained (starting from initially null values) using equations similar to 16 and 17, but assuming that the *sum* of synapses entering a word must not overcome a saturation value (say W^{Fex}_{sumMax}). Hence

$$W^{Fex}_{ij,hk}(t+T_S) = W^{Fex}_{ij,hk}(t) + \beta^{Fex}_{ij,hk}\cdot x^{L}_{ij}(t)\cdot m_{hk}(t) \tag{22}$$

$$\beta^{Fex}_{ij,hk} = \beta^{Fex}_{0}\left(W^{Fex}_{sumMax} - \sum_{lm} W^{Fex}_{ij,lm}\right) \tag{23}$$

where the average activity $m_{hk}(t)$ is defined as in Eq. 15, and the sum in the right-hand member of Eq. 23 is extended to all synapses from the feature network entering the neuron *ij* in the lexical network.

The inhibitory synapses start from a high value (say W^{Fin}_{Max}) and are progressively withdrawn using an Hebbian mechanism:

$$W^{Fin}_{ij,hk}(t+T_S) = [W^{Fin}_{ij,hk}(t) - \beta^{Fin}_{ij,hk}\cdot x^{L}_{ij}(t)\cdot m_{hk}(t)]^{+} \tag{24}$$

where the function "positive part" ($[]^{+}$) is used in the right hand member of Eq. 24 to avoid that these synapses become negative (i.e., that inhibition is converted to excitation).

5. References

Abbott, L. F., and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nat. Neurosci.*, 3, 1178-1183.

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, 128(3), 466-478.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn Sci*, 7(2), 84-91.

Bhattacharya, J., Petsche, H., and Pereda, E. (2001). Long-range synchrony in the gamma band: role in music perception. *J. Neurosci.*, 21, 6329-6337.

Caramazza, A., Hillis, A., and Rapp, B. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology*, 7, 161-189.

Caramazza, A., and Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *J. Cogn. Neurosci.*, 10, 1-34.

Cuppini, C., Magosso, E., and Ursino, M. (2009). A neural network model of semantic memory linking feature-based object representation and words. *BioSystems*, 96(3), 195-205.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25-62.

Devlin, J. T., Gonnerman, L. M., Andersen, E. S., and Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: a computational account. *J Cogn Neurosci*, 10(1), 77-94.

Engel, A. K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci*, 5(1), 16-25.

Gainotti, G. (2006). Anatomical functional and cognitive determinants of semantic memory disorders. *Neuroscience and Behavioral Reviews*, 30, 577-594.

González, J., Barros-Loscertales, A., Pulvermller, F., Meseguer, V., Sanjun, A., Belloch, V., and Avila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage*, 32(2), 906-912.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: language and cognition*, 1(2), 67-81.

Hart, J., Anand, R., Zoccoli, S., Maguire, M., Gamino, J., Tillman, G., King, R., and Kraut, M. A. (2007). Neural substrates of semantic memory. *J Int Neuropsychol Soc*, 13(5), 865-880.

Hart, J., Moo, L. R., Segal, J. B., Adkins, E., and Kraut, M. (2002). "Neural substrates of semantics." *Handbook of language disorders*, Psychology Press, Philadelphia.

Haykin, S. (1999). "Neural Neworks: a comprehensive foundation." Prentice Hall.

Hopfield, J. J., and Brody, C. D. (2001). What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc Natl Acad Sci U S A*, 98(3), 1282-1287.

Humphreys, G. W., and Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neurophysiological deficits. *Behavioral and Brain Sciences*, 24, 453-509.

Kraut, M. A., Kremen, S., Segal, J. B., Calhoun, V., Moo, L. R., and HArt, J. (2002). Object activation from features in the semantic system. *J Cogn Neurosci*, 14(1), 24-36.

Lambon Ralph, M. A., Lowe, C., and Rogers, T. T. (2007). Neural basis of category-specific demantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. *Brain*, 130, 1127-1137.

Lauro-Grotto, R., Reich, S., and Visadoro, M. (1997). "The computational role of conscious processing in a model of semantic memory." *Cognition, Computation and Consciousness*, M. Ito, S. Miyashita, and E. Rolls, eds., Oxford University Press, Oxford, 249-263.

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSSs. *Science*, 275, 213-215.

Martin, A., and Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.*, 11(2), 194-201.

Matsumoto, A., Iidaka, T., Haneda, K., Okada, T., and Sadato, N. (2005). Linking semantic priming effect in functional MRI and event-related potentials. *Neuroimage*, 24(3), 624-634.

McRae, K., de Sa, V. R., and Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *J Exp Psychol Gen*, 126(2), 99-130.

Morelli, A., Lauro, G. R., and Arecchi, F. T. (2006). Neural coding for the retrieval of multiple memory patterns. *BioSystems*, 86(1-3), 100-109.

Pulvermller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005a). Functional links between motor and language systems. *Eur J Neurosci*, 21(3), 793-797.

Pulvermller, F., Shtyrov, Y., and Ilmoniemi, R. (2005b). Brain signatures of meaning access in action word recognition. *J Cogn Neurosci*, 17(6), 884-892.

Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., and Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397, 430-433.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., and Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol Rev*, 111(1), 205-235.

Rolls, E. T., and Treves, A. (1998). "Neural Networks and Brain Function." Oxford University Press, Oxford.

Rossell, S. L., Price, C. J., and Nobre, A. C. (2003). The anatomy and time course of semantic priming investigated by fMRI and ERPs. *Neuropsychologia*, 41(5), 550-564.

Simmons, W. K., and Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn. Neuropsychol.*, 20, 451-486.

Singer, W., and Gray, C. M. (1995). Visual Feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.*, 18, 555-586.

Small, S. L., HArt, J., Nguyen, T., and Gordon, B. (1995). Distributed representations of semantic knowledge in the brain. *Brain*, 118 (Pt 2), 441-453.

Snowden, J. S., Thompson, J. C., and Neary, D. (2004). Knowledge of famous faces and names in semantic dementia. *Brain*, 127, 860-872.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., and Pernier, J. (1997). Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. *J. Neurosci.*, 17, 722-734.

Tallon-Baudry, C., Bertrand, O., Peronnet, F., and Pernier, J. (1998). Induced gamma-band activity during the dealy of a visual short-term memory task in humans. *J. Neurosci.*, 18, 4244-4254.

Tulving, E. (1972). "Episodic and semantic memory." *Organisation of memory*, E. Tulving, and W. Donaldson, eds., Academic Press, New York.

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., and Levy, J. P. (2000). Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. *Brain Lang*, 75(2), 195-231.

Ursino, M., La Cara, G. E., and Sarti, A. (2003). Binding and segmentation of multiple objects through neural oscillators inhibited by contour information. *Biol. Cybern.*, 89, 56-70.

Ursino, M., Magosso, E., and Cuppini, C. (2009). Recognition of abstracts objects via neural oscillators: interaction among topological organization, associative memory and gamma-band synchronization. *IEEE Tr. Neural Networks*, 20(2), 316-335.

Ursino, M., Magosso, E., La Cara, G. E., and Cuppini, C. (2006). Object segmentation and recovery via neural oscillators implementing the similarity and prior knowledge gestalt rules. *BioSystems*, 85, 201-218.

von der Malsburg, C., and Schneider, W. (1986). A neural cocktail-party processor. *Biol. Cybern.*, 54, 29-40.

Wang, D., and Terman, D. (1997). Image Segmentation based on oscillatory correlation. *Neural Computation*, 9, 805-836.

Ward, J. (2006). "The student's guide to cognitive neuroscience." Psychology Press, Hove and New York.

Warrington, E. K., and McCarthy, R. (1983). Category specific access dysphasia. *Brain*, 106, 859-878.

Warrington, E. K., and McCarthy, R. (1987). Categories of knowledge: further fractionations and an attempted integration. *Brain*, 110, 1273-1296.

Warrington, E. K., and Shallice, T. (1984). Category specific semantic impairements. *Brain*, 107, 829-854.

Cognitive Maps

Edited by Karl Perusich

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Cristiano Cuppini, Elisa Magosso and Mauro Ursino (2010). The Representation of Objects in the Brain, and Its Link with Semantic Memory and Language: a Conceptual Theory with the Support of a Neurocomputational Model, Cognitive Maps, Karl Perusich (Ed.), ISBN: 978-953-307-044-5, InTech, Available from: http://www.intechopen.com/books/cognitive-maps/the-representation-of-objects-in-the-brain-and-its-link-with-semantic-memory-and-language-a-conceptu

INTECH
open science | open minds