A feature-based neurocomputational model of semantic memory

Mauro Ursino¹ · Cristiano Cuppini¹ · Stefano F. Cappa^{2,3} · Eleonora Catricalà²

Received: 29 January 2018 / Revised: 19 June 2018 / Accepted: 29 June 2018 / Published online: 7 July 2018
© Springer Nature B.V. 2018

Abstract

According with a featural organization of semantic memory, this work is aimed at investigating, through an attractor network, the role of different kinds of features in the representation of concepts, both in normal and neurodegenerative conditions. We implemented new synaptic learning rules in order to take into account the role of partially shared features and of distinctive features with different saliency. The model includes semantic and lexical layers, coding, respectively for object features and word-forms. Connections among nodes are strongly asymmetrical. To account for the feature saliency, asymmetrical synapses were created using Hebbian rules of potentiation and depotentiation, setting different pre-synaptic and post-synaptic thresholds. A variable post-synaptic threshold, which automatically changed to reflect the feature frequency in different concepts (i.e., how many concepts share a feature), was used to account for partially shared features. The trained network solved naming tasks and word recognition tasks very well, exploiting the different role of salient versus marginal features in concept identification. In the case of damage, superordinate concepts were preserved better than the subordinate ones. Interestingly, the degradation of salient features, but not of marginal ones, prevented object identification. The model suggests that Hebbian rules, with adjustable post-synaptic thresholds, can provide a reliable semantic representation of objects exploiting the statistics of input features.

Keywords Neurocomputational models · Attractor networks · Semantic features · Distinctiveness · Partially shared features · Semantic memory impairment · Salient and marginal features

Introduction

The organization and the neural mechanisms of semantic memory are a central issue in cognitive neuroscience, not only for their theoretical interest, but also to understand knowledge deficits in neurological disorders, such as Alzheimer's disease and semantic dementia (Cappa 2008). In the last decades several neuropsychological (Alathari et al.

2004; Garrard et al. 2005; Catricalà et al. 2014, 2015c; Marques et al. 2011), neuroimaging (Catricalà et al. 2015b; Marques et al. 2008), as well as neuro-computational studies (McRae et al. 1997; Devlin et al. 1998; Tyler et al. 2000; Ursino et al. 2013, 2015) have been conducted on this topic. Most theoretical models assume that concrete objects are represented in a semantic store as a collection of features (or attributes), either amodal or modality-specific, which are coded in different parts of the brain and are reciprocally interconnected in a flexible, highly plastic distributed network (Allport 1985; Martin 2007; Tyler and Moss 2001). Semantic features may differ not only for the type of information they convey, namely, sensory, motor, encyclopedic, but also in terms of how relevant they are about the entity they describe. For example, while shared features, occurring in several concepts, support the category level, distinctive features, occurring in only one or a very few concepts, allow the differentiation between related concepts belonging to the same semantic category. For example, 'has wings' is a shared feature as it occurs in many concepts (birds); while 'coos' and 'used to send

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11571-018-9494-0>) contains supplementary material, which is available to authorized users.

✉ Mauro Ursino
mauro.ursino@unibo.it

¹ Department of Electrical, Electronic and Information Engineering, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy

² NETS Center, Scuola Universitaria Superiore IUSS, Pavia, Italy

³ IRCCS S. Giovanni di Dio, Brescia, Italy

messages’ are distinctive features, as they occur only in *pigeon*.

Within neural networks, based on attractor dynamics, the distributed information on individual objects is expressed by the simultaneous activation of a group of neurons (i.e. nodes), coding for the different features. Through synaptic learning, the object representation is stored as an equilibrium point of the network. Subsequently, the complete information can be restored from a partial cue (i.e., from activation of just a few nodes) by exploiting the attractor dynamics of the model, which converges towards the equilibrium. Feature-based attractor networks yield significant insights into several behavioral phenomena regarding semantic memory, both in normal (Cree et al. 2006; Masson 1995; McRae et al. 1997; O’Connor et al. 2009) and pathological conditions (Rogers et al. 2004). In particular, attractor dynamics clearly explain the different role played by distinctive and shared features in the semantic representation of concepts (Cree et al. 2006), and the distinction between subordinate and superordinate concepts (O’Connor et al. 2009).

Despite these important achievements, several aspects of semantic organization are still insufficiently known, and several behavioral phenomena in neurodegenerative diseases still require an adequate formalization within the attractor network framework.

In this study, we address two different problems within attractor dynamics, respectively regarding shared and distinctive features. The aim is to identify the role of different types of semantic features in concept identification, and to mimic behavioral phenomena observed in neurodegenerative diseases.

Concerning shared features, some of these are only partially shared, as they characterize only some (but not all) members of a category. Hence, while they neither determine the semantics of the overall category, nor can be used to identify individual members, they play a central role in assigning individual members to a specific category. For example, the attribute “it flies” is generally used to characterize the concept “bird”, but not all birds fly. A flying animal, however, is very often a bird.

A Hopfield model, where semantic units are interconnected, learns correlations between feature pairs. Two features are correlated if they tend to appear together for the same concepts. When two features (as <has wings> and <it flies>) co-occur in several concepts, then if one of them (i.e., has wings) is activated, the other (it flies) will also tend to be activated by positive weights between them. Clusters of features like ‘has wings’, ‘has two legs’, ‘has feathers’ and ‘fly’ are significantly correlated because they co-occur in the case of several birds, like *parrot*, *owl* and *pigeon*. However, some birds, as *hen*, do not fly, but ‘have wings’, ‘feathers’ and ‘two legs’. In these cases, attractor

dynamics will lead to an incorrect activation of ‘it flies’ as an additional feature of *hen*, due to its correlation with the other features.

Similarly, in the Hopfield model developed by McRae et al. (1997) some concepts erroneously activated additional features not present in the feature-based semantic representations used to train the network. For the most part, the additional activations were suitable to the concepts, as ‘is edible’ for the carrot. In other cases, however, some features were unsuitable, for example for the concept ‘jet’ the activation of birdlike features, i.e., has feathers. The errors of inclusion were due to the high correlation between the additional features activated and the other features within the concept. These errors of inclusion were considered as a form of inference, in which the network acts as ‘a pattern-completion device that relies on its knowledge of feature correlation’ (McRae et al. 1997). The problem of the partially shared features, highly correlated with each other, within the Hopfield model is still to be resolved.

The second issue is related to distinctive features, which have a peculiar role in the identification of concepts, as they permit to discriminate between similar concepts. Marques (2005) reported that in a group of healthy subjects distinctive features were more often selected than shared features to support the ability to name an object according to a verbal definition. Distinctive features are less resistant to the effect of brain damage than shared ones (Alathari et al. 2004; Catricalà et al. 2015c; Duarte et al. 2009; Garrard et al. 2005; Giffard et al. 2001, 2002; Laisney et al. 2011; Perri et al. 2011, 2013; Warrington 1975), leading to a confusion between similar concepts, and consequently to an impaired performance on task requiring the identification of individual concepts (Moss et al. 2002; Tyler et al. 2000; Duarte et al. 2009). A number of authors have argued that not all the distinctive features have the same importance in the representation of the concepts (Cree et al. 2006). As pointed out in feature listing tasks (McRae et al. 2005; Catricalà et al. 2015a, c), dominance (production frequency), namely how frequently a semantic feature is used in defining a concept, is a critical dimension to consider. Some attributes (defined as salient or dominant) are in fact spontaneously evoked when thinking about an object; others (defined as marginal), although contributing to object recognition, come to mind less frequently. For example, the feature ‘used to send messages’ is distinctive as it is listed only for *pigeon*, but at the same time it is not highly salient, as it is infrequently reported in feature-norming tasks, presumably not playing a prominent role in the representation of the concept *pigeon*. The feature ‘coos’ is also distinctive, but it is salient, since it is reported by many subjects in norming tasks. The different importance of the two features is captured by semantic relevance, a measure combining distinctiveness and dominance (Sartori

and Lombardi 2004). Recent results emphasize a different role of salient versus marginal distinctive features in semantic deficits. Catricalà et al. (2015c) observed that deficits in object naming task in patients with Alzheimer's diseases and semantic dementia are associated with a loss of salient distinctive features, whereas a loss of marginal features has no impact. Dominance is not generally incorporated in attractors models (McRae et al. 1997), thus preventing the discrimination between salient versus marginal features (i.e., McRae et al. 1997; but see Ursino et al. 2015). The role of salient versus marginal distinctive features, their relationship with experience, and their representation within an attractor dynamics are not sufficiently clear.

In order to deal with partially versus totally shared features, and with the different role of salient versus marginal distinctive features, one needs a model in which the connections among nodes are strongly asymmetrical. Furthermore, these asymmetrical connections should emerge spontaneously from experience, based on the previous statistics of features occurrence, using realistic rules for synaptic learning. Unfortunately, some of the attractor networks mentioned above (Masson 1995; McRae et al. 1997) use classic Hebbian paradigms for synaptic learning (Dayan and Abbott 2001), as in the classic Hopfield model (Hopfield 1982, 1984), leading to the formation of symmetrical synapses. As illustrated below, symmetrical synapses are inadequate to relate concepts in a semantic memory model. Conversely, other models (Rogers et al. 2004; Cree et al. 2006; O'Connor et al. 2009) make use of back-propagation through time. This is a powerful training algorithm, but it is biologically unrealistic and requires a supervisor which communicates the correct response to the net.

In recent years, we developed neural network models of the lexical-semantic-system (Cuppini et al. 2009; Ursino et al. 2010, 2011) based on attractor dynamics. The model assumes that the semantic and lexical aspects are stored in two different layers; after learning, the semantic layer works to restore the salient features of a concept; then, the lexical and semantic layer are interconnected, implementing the relations between semantic content and word-forms. In the most recent version of this model (Ursino et al. 2015) the use of a Hebbian learning paradigm with flexible potentiation and de-potentiation leads to the formation of asymmetrical synapses. The network is able not only to distinguish between individual members and categories, by exploiting the differences between distinctive versus shared features, without any internal hierarchical organization, but also to reach a clear distinction between salient and marginal features, based only on the different probability of features to be presented during the training, namely on their frequency of occurrence.

In the previous paper, however, we did not represent real objects, but just schematic concepts described as the co-activation of neurons in specific areas of the semantic network, assuming that each area exhibits a topological organization (as in most primary cortical regions). Hence, the previous simulations were just a “proof of concept”, without direct contact with reality, not including, for example, partially shared features. In this work, the previous model is extended by the usage of semantic feature production norms (Catricalà et al. 2015a), to account for the variables of interest (distinctiveness and dominance) within the model.

With the aim to solve the correct computation of partially shared features within the attractor dynamics, we propose a variant of the learning rule. This consists in the use of a variable post-synaptic threshold, which automatically changes to reflect the frequency of a feature in different concepts (i.e., how many concepts share a feature). To our knowledge, this is a new approach, never included in previous attractor models. In order to distinguish between salient and marginal features, we use the different probability of a feature presentation within a given concept in the training input (production frequency, dominance). The different probability is based on the value of dominance reported in the database of Catricalà et al. (2015a). Higher values of dominance correspond to a higher probability that a feature is presented, and consequently to a higher salience. To better understand the different role of salient and marginal distinctive features in object identification, we simulated the behavioral phenomena reported in patients with neurodegenerative diseases (i.e., the loss of salient features, but not of the marginal ones, preventing object identification, Catricalà et al. 2015c) using a naming to definition task. This task, largely used in patients with semantic disorders (Marques et al. 2011; Sartori and Lombardi 2004; Sartori et al. 2007; Silveri and Gainotti 1988; Catricalà et al. 2015c), allows to evaluate the role of different feature dimensions in naming performance. The participants are presented with a sentence describing the target concept, including a set of semantic features, and are asked to provide a name corresponding to the definition.

The data base and the semantic taxonomies

The semantic taxonomies used in this work were built starting from a data-set developed by Catricalà et al. (2013, 2015a). The entire data-set contains 82 concepts, subdivided in living (“birds”, “land animals”, “vegetables”, “fruits”) and nonliving (“furniture”, “tools”, “kitchen items”, “clothing animals” and “vehicles”). From the features generated by 20 participants, several

dimensions were derived, i.e. distinctiveness, dominance, semantic relevance and semantic distance. Two different measures of distinctiveness were calculated. The first refers to the number of concepts for which the semantic feature appears, divided by the total number of concepts in the database, while the second refers to the proportion of concepts within a category for which the feature in question was generated. Dominance is the number of participants who listed a specific feature for a specific concept. Semantic relevance is “a measure of the contribution of semantic features to the core meaning of the concept” (Sartori and Lombardi 2004), calculated as a non-linear combination between dominance and distinctiveness. Semantic distance is a measure of similarity between semantic representations, calculated using feature dominance to weight matching and mismatching features. Two different measures of semantic distance were calculated (Zannino et al. 2006): between each pair of concepts belonging to the same category, and between each concept and the centroid of the relative semantic category (for further details, see Catricalà et al. 2013, 2015a).

In the present study, we selected 12 animals (six “land animals” and six “birds”) and 11 artificial objects (four “furniture”, four “kitchen items”, and three “tools”).

Note that we did not use the entire number of features presented in the database for each concept. To train the model, we constructed simple virtual concepts representations, which captured the important features characterizing concepts for the categories of animals and artifacts as reported in the database. Concepts contained all the types of features of interest, namely shared, partially shared, distinctive, distinctive with high dominance (salient), distinctive with low dominance (marginal), thus permitting to test our hypothesis. The following definitions summarize the feature characteristics:

1. *Shared: features belonging to several concepts*
 1. *Totally shared:* features belonging to all members of a category;
 2. *Partially shared:* features belonging only to some members of a category (more than a single member).
2. *Distinctive: a feature that belongs only to one concept;*
 1. *Distinctive Salient:* a distinctive feature that is spontaneously evoked when thinking about the object, i.e. a distinctive feature frequently used in defining a concept;
 2. *Distinctive Marginal:* a feature that is infrequently evoked in defining a concept;

To decide whether a feature is salient or marginal for a concept, we calculated the median of the dominance values

for each concept. We say that a feature is salient if it exhibits a value equal or higher than the median value.

The features used for each animal and object are reported in Table 1 (animals) and in the Supplementary Material (Table S1, objects), where a distinction between salient and marginal features is provided.

The neurocomputational model

Qualitative model description

The network

The model incorporates two networks of neurons, as illustrated in the schematic diagram of Fig. 1. The first, named “semantic network” is devoted to a description of objects represented as a collection of features. A second, the “lexical network” codes for lemmas or word-forms.

In the following, as in Fig. 1, the symbols W and V will denote an excitatory and an inhibitory synapse, respectively; the subscripts will denote the post-synaptic and pre-synaptic neurons, respectively; and the two superscripts the corresponding layers (S = semantic; L = lexical). By way of example, the symbol W_{ij}^{SL} denotes an excitatory synapse from the neuron at position j in the lexical layer to the neuron at position i in the semantic layer; the symbol V_{ij}^{LS} denotes an inhibitory synapse from the j th neuron in the semantic layer to the i th neuron in the lexical layer.

It is worth noticing that the semantic and the lexical nets become strongly interconnected after learning, hence they work together, in an integrative way, to constitute a single highly interactive lexical-semantic system.

Each neuron, both in the semantic and in the lexical net, is represented via a first order dynamics with assigned time constant, to account for the integrative properties of the neuron, and a sigmoidal activation characteristic, accounting for the presence of a lower threshold and upper saturation. Hence, when the input is too low, the neuron activity is negligible, while a high input leads neuron activity close to saturation. In this work, the upper saturation is set at one, i.e., all activities are normalized.

During the simulations, a feature is represented by the activity of a neuron at a given position in the semantic network. This activity can be evoked by an external localized input: we assume that these inputs, reaching neurons in the semantic network, are the result of an upstream processing stage that extracts the main sensory-motor properties of the objects. Moreover, a feature can also receive excitatory synapses from other features in the semantic net (W_{ij}^{SS} in Fig. 1), thus realizing an attractor network, and excitatory synapses from word-forms in the

Table 1 Lexical labels and features, and relative numbers -No.- used to identify, respectively, each lexical label and each semantic feature in the model used in the training set of animals

Lexical label	No.	Features	No.	Lexical label	No.	Features	No.
Animal	15	<i>It eats*</i>	2				
		<i>It sleeps*</i>	3				
Mammals	13	<i>It has four legs*</i>	4	Bird	14	<i>It has two legs*</i>	6
		<i>It has fur*</i>	5			<i>It has feathers*</i>	7
Dog	1	It barks*	9			<i>It has wings*</i>	8
		It wags its tail*	10			<i>It flies</i>	42
		It is domestic*	11			<i>It flutters</i>	52
		It is a pet*	12	Goose	7	It has a long neck*	39
		It is affectionate/friendly	13			It is white*	41
		It chases	14			It flies*	42
Cat	2	It is domestic*	11			It has webbed feet*	43
		It meows*	15			It lives in the pond	44
		It purrs*	16			It is reared	45
		It chases mice*	17			It has orange beak	46
		It has whiskers*	18			It honks	47
		It scratches	19	Hen	8	It is reared*	45
		It is independent	20			It lives in coops*	48
Herbivore	16	<i>It eats grass*</i>	21			It has a comb*	49
		<i>It grazes</i>	26			It has chicks	50
Sheep	3	It bleats*	22			It pecks	51
		It has long fur*	23			It flutters*	52
		Used for wool*	24	Rooster	9	It flutters*	52
		It lives in flocks*	25			It crows in the morning*	53
		It grazes	26			It crows*	54
		It is gentle	27			Unique in coops*	55
		It lives on farms	32			Has a red comb	56
Cow	4	It grazes*	26			It wakes up people	57
		It is gentle	27	Pigeon	12	It flies*	42
		It moos*	28			It coos*	66
		It has horns*	29			It eats crumbs*	67
		It produces milk*	30			It lives in squares*	68
		It has spots*	31			It moves its head back and forth	69
		It lives on farms*	32				
Giraffe	6	It has spots*	31			Used to send messages	70
		It lives in savannah*	35	Parrot	10	It flies*	42
		It is wild	36			It repeats sounds*	58
		It is tall*	38			It is exotic*	59
		It has long neck*	39			It is of various colors*	60
		It eats leaves	40			It has a big beak	61
Zebra	5	It has black and white stripes*	33	Owl	11	It flies*	42
		It is like a horse*	34			It is associated with bad luck*	62
		It lives in savannah*	35			It is a rapacious*	63

Table 1 (continued)

Lexical label	No.	Features	No.	Lexical label	No.	Features	No.
		It is wild	36			It chases in the night	64
		It has mane	37			Hit has a hooked beck	65

Features in *italics* are totally shared: all of them are salient. Hence they have a frequency as high as 0.7. Only for the category “Animal” we used probabilities as high as 0.9 for its two features “It eats” or “it sleeps”, to limit the probability that the concept statistically occurs with no input feature at all or with just one feature. Features not in italic are partially shared or distinctive: those with asterisk (*) have high values of dominance (i.e. salient); the other features have low values of dominance (i.e. marginal)

The frequency of partially shared features in the category “Birds” were computed taking into account the number of members which have that property. We have:

Frequency “it flies”: $0.7 \cdot 4/6 = 0.47$ (4 birds fly); Frequency “it flutters” = $0.7 \cdot 2/6 = 0.23$ (two birds flutter)

The frequency of “It grazes” in the category herbivore takes into account that the cow and sheep graze:

Frequency “it grazes” = $0.7 \cdot 1/4 + 0.4 \cdot 1/4 = 0.28$

The salience is based on the values of dominance taken from Catricalà et al. (2015a), see text for further details

The distinctiveness is based on the number of concepts for which a given feature appeared, with respect to all concepts of the training set (not of the original database). In some cases, in fact, it is possible that a shared feature in the original database becomes distinctive for the sample of features selected for this study

Fig. 1 Schematic diagram of the model. The computational units in the semantic network code for specific features of the objects. After training, some of them are reciprocally connected via non-symmetrical excitatory synapses (W_{ij}^{SS} , black), reflecting the previous co-occurrence of features. The computational units in the lexical network code for word-forms. In the present model they are not reciprocally linked. After training, non-symmetrical excitatory synapses are created from word-forms to salient features (W_{ij}^{SL} , red) and from salient features to word-forms (W_{ij}^{LS} , blue). Finally, features send inhibitory synapses (V_{ij}^{LS} , green) to the word-forms to which they do not contribute. (Color figure online)

lexical net (W_{ij}^{SL} in Fig. 1), thus realizing an hetero-association between features and word-forms.

Similarly, each word-form is represented by the activity of a neuron in the lexical net. It can be directly excited by an external input (which represents a listened or written word) or from its feature representation in the semantic net, via excitatory synapses (W_{ij}^{LS}). Moreover, the synapses

from the semantic to the lexical network also include inhibitory terms (V_{ij}^{LS}), to avoid that a word-form is evoked when an incongruous feature is present (for instance, if the feature “it meows” is present when the subject is trying to recognize “a dog”). To this end, a feature that never participates to the representation of an object sends an inhibitory synapse to the corresponding word-form. In the present work, we do not include direct connections among word-forms. This limitation might be the subject of future improvements.

To summarize, we have four kinds of synapses in the model:

- excitatory (non-symmetrical) synapses connecting features in the semantic net (W_{ij}^{SS});
- excitatory synapses from features in the semantic net to word-forms in the lexical net (W_{ij}^{LS});
- inhibitory synapses from features in the semantic net to word-forms in the lexical net (V_{ij}^{LS});
- excitatory synapses from word-forms in the lexical net to features in the semantic net (W_{ij}^{SL}).

Furthermore, we distinguish features in the semantic net in shared, partially shared, distinctive, distinctive with high dominance (salient), distinctive with low dominance (marginal), according to the definitions reported above.

The training procedure

Excitatory synapses are trained with a Hebbian rule, which modifies the weight on the basis of the correlation between the presynaptic and postsynaptic activity. In order to account not only for long-term potentiation but also for

long-term depression (see Dayan and Abbott 2001), we assume that these activities are compared with a threshold. In this way, a low level of activity in one neuron causes a depression of the synaptic strength if accompanied by a high level of activity in the other neuron. Conversely, when both neurons have high activity, the synapse strengthens. Hence, we can write

$$\Delta W_{ij}^{AB} = \gamma_{ij}^{AB} (x_i^A - \vartheta_{post}^{AB}) (x_j^B - \vartheta_{pre}^{AB}) \quad (1)$$

with $A, B = S$ or L

where ΔW_{ij}^{AB} represents the change in the synapse strength, due to the given presynaptic and post-synaptic activities, x_i^A is the activity of the neural unit at position i in the post-synaptic area A , and x_j^B is the activity of the neural unit at position j in the pre-synaptic area B . ϑ_{post} and ϑ_{pre} are thresholds for the postsynaptic and presynaptic activities, γ_{ij}^{AB} denotes a learning factor (which depends on the previous history of the synapse).

Equation (1) however requires some restrictions to be physiologically realistic. First, the learning factor decreases when the synapse approaches a maximum saturation level, i.e. a synapse cannot grow up indefinitely. Second, the synapses cannot become negative. Third, the previous equation does not hold when both neurons are below threshold. In this case, the synapse remains unchanged (see the Supplementary Material, part I, for all mathematical details and parameter numerical values).

The inhibitory synapses (which connect only semantic features to word-forms) are trained with an anti-Hebbian mechanism: i.e., they are weakened when both the pre-synaptic and post-synaptic activities are above threshold, and are strengthened when the activities are negatively correlated (hence, we used an equation analogous to Eq. (1), but with a negative learning factor, $\gamma_{ij}^{LS} < 0$).

The learning procedure is divided into two distinct phases:

(i) *Semantic training*: during this phase, the objects are individually presented to the semantic net, while all lexical units receive null input. This corresponds to a phase in which the subject experiments with the objects, and learns their semantics, without an association with lexical items. However, not all object features are simultaneously used. Each feature has a different probability to be used as input in the given object: the higher this probability, the higher the feature saliency. The different probability is based on the values of dominance reported in the database. Let us consider the j th feature in the net, and denote with P_{kj} the probability that is presented as input for the k -th object. Accordingly, during training, when the k -th object is presented, this j th feature can receive either a high input (a value sufficient to move the neuron close to saturation, the

feature is perceived), with a probability P_{kj} , or a low input (i.e., an input that is not able to excite the corresponding neuron, the feature is not perceived) with a probability $1 - P_{kj}$. For the sake of simplicity, in the present work all salient features (i.e., with a value of dominance $>$ of the median, see above and see Table 1 and Table S1 in the Appendix) were given a high probability, namely 70% (i.e. $P_{kj} = 0.7$) while all marginal features (i.e., with a value of dominance $<$ of the median, see above and see Table 1 and Table S1) were given a low probability, namely 40% (i.e. $P_{kj} = 0.4$). This allows a simpler analysis of the results.

The overall semantic training consisted of 1000 consecutive epochs. During each epoch, all objects were given separately, permuted in random order (i.e., each object is presented once in each epoch) and the excitatory synapses within the semantic net were trained with the Hebbian mechanism. As mentioned above, the synapses are strengthened when both the pre-synaptic and post-synaptic activities are above threshold, and weakened when the activities are negatively correlated.

After training, synapses in the semantic net are set at their value, which was maintained fixed during the entire phase II.

Lexical training: During phase II, word forms are presented to the lexical network, together with the corresponding features in the semantic net.

This phase consisted of 1000 epochs, but each epoch now comprises the random permutation of all word-forms (both those associated with individual members and those associated with categories): each word-form is used once per each epoch, by exciting the corresponding lexical unit with an external input able to lead the neuron close to saturation.

In case of words denoting individual members, features in the semantic network were simultaneously randomly given according to the same probabilities used in the previous phase (i.e., $P_{kj} = 0.7$ for salient and $P_{kj} = 0.4$ for marginal features). However, salient features are now automatically evoked within each concept even when their input is absent, thanks to the action of inter-area synapses trained in phase I.

Furthermore, in case of word-forms representing categories, only shared features were given with their individual probabilities (see Tables 1 and S1). In case of totally shared features the probability is 0.7. In case of partially shared features (such as “it flies” or “it flutters” for birds or “it grazes” for herbivores) the probability takes into account the real number of members actually incorporating that feature, and the probability within each member. These computations are reported at the bottom of Table 1.

The excitatory synapses linking lexical and semantic aspects were trained with a similar Hebbian mechanism as

in the previous phase (but with different values for the thresholds). The inhibitory synapses were trained with the anti-Hebbian rule.

The maximum saturation for the synapses entering a semantic feature was chosen high enough, so that even a single pre-synaptic feature can lead a salient post-synaptic feature close to saturation (i.e., the word-form in the lexical net, or just one feature in the semantic net are able to evoke all other salient features). Finally, we used a particular saturation rule for the excitatory synapses entering into a word-form from the semantic net: we assume that the *sum* of these synapses saturates to a maximum level. This has been chosen so that, when all salient features are present, the activity of the word-form is close to the upper saturation, but even the absence of one salient feature causes its almost complete inhibition. This is possible thanks to the sharp sigmoidal characteristic used for the lexical units. The maximum saturation for inhibitory synapses was chosen so that even a single feature in the semantic net, which does not participate in object semantics, is able to move the activity of the word form from the upper saturation to the inhibition.

All equations, together with the details on parameter numerical values (thresholds, upper saturation, sigmoid slope, etc.) sufficient to reconstruct the entire model working, are given in the Supplementary Material, part I.

However, a fundamental point to achieve a correct training concerns the choice of appropriate values for the presynaptic and post-synaptic thresholds in the Hebb rule (Eq. (1)), an issue not previously analyzed. For this reason, the subsequent section is entirely devoted to this crucial point. This discussion also underlines the need to have a variable post-synaptic threshold, which is a new feature of the present work, never implemented formerly.

Parameter assignment: the thresholds in the Hebb rule

The thresholds in Eq. (1) were assigned to fulfil some objectives of the model. These are individually discussed for the semantic net and the lexical-semantic interaction, respectively.

Synapses in the semantic net

Salient versus marginal (a) Salient features are most likely evoked by all other features (either salient or marginal) of the object; hence they should receive strong input synapses. Conversely, they should send strong output synapses only to the other salient features. (b) Marginal features should not be evoked by the other features of the object (i.e., they

receive weak input synapses) but favor the reconstruction of the object (hence, they send strong output synapses toward salient features).

Shared versus distinctive (c) A distinctive feature should recall not only other salient distinctive features of the same object but also the shared features (the feature “it barks” should recall all properties of a dog, including “It eats”, “It sleeps” shared with other animals). (d) Features shared by different objects in a category should recall all the other totally shared features in the same category, but they should not recall distinctive features of individual members, nor partially shared features (by way of example, the feature “it has four legs” should recall the common features of the category “Land animal”, such as “it has fur”, but should not recall the features of individual members “it barks”, “it meows” etc.).

The previous conditions are summarized in the block diagram of Fig. 2.

This particular behavior of semantic synapses can be achieved assuming that the threshold for post-synaptic activity is quite high (in the following we will assume, $\vartheta_{post} = 0.55$, i.e., close to the mid between maximum inhibition and maximum excitation), whereas the threshold for the pre-synaptic activity is low (let us assume $\vartheta_{pre} = 0.05$, close to inhibition). A value just slightly above zero has been chosen here to avoid that a residual neuronal activity causes an undesired synapse reinforcement. The previous choices have the following main consequences:

- (i) If both the presynaptic and postsynaptic neurons are active (activity close to 1), the synapses strengthens [we have $\Delta W_{ij} = \gamma_{ij} \cdot 0.45 \cdot 0.95 \cong \gamma_{ij} \cdot 0.42$].
- (ii) If the postsynaptic neuron is inhibited (activity close to 0) while the presynaptic neuron is high (activity close to 1), the synapse weakens [we have $\Delta W_{ij} = -\gamma_{ij} \cdot 0.55 \cdot 0.95 \cong -\gamma_{ij} \cdot 0.52$]. This situation occurs at the synapse leaving a shared feature toward a distinctive feature, when the shared feature appears in a concept not containing that particular distinctive feature (for instance the feature “it has a fur” toward the feature “it meows”, when one is looking at a dog). The same situation also occurs for the synapses leaving a salient feature toward a marginal feature. Hence, after sufficient training, due to the statistics of the feature occurrence, shared features will send weak synapses toward distinctive features, and salient features will send weak synapses toward marginal features.

Fig. 2 The expected connections between salient/marginal and distinctive/partially shared/totally shared features in the semantic network after training. Thick lines represent strong synapses, able to excite the post-synaptic feature. A self-ring represents synapses toward other features of the same type

- (iii) If the postsynaptic neuron is excited (activity close to 1) and the presynaptic neuron is inhibited (activity close to zero), the synapse exhibits just a moderate weakening [we have $\Delta W_{ij} = -\gamma_{ij} \cdot 0.45 \cdot 0.05 \cong -\gamma_{ij} \cdot 0.023$]. This is the situation occurring at the synapses which leave a marginal feature toward a salient feature (the pre-synaptic activity is often close to zero, since marginal features are often absent). As a consequence, a distinctive marginal feature continues to send strong synapses toward all salient features, with only limited weakening. The same condition also occurs if we consider a synapse from any distinctive feature toward a shared feature.

It is worth noting that, in our model, saliency is governed by the value of incoming synapses (i.e., a feature which receives strong synapses after training is spontaneously evoked, hence is salient according to our definition of saliency, see Fig. 2). According to the previous analysis of the learning rule, the unique aspect that determines saliency in the present model version is the frequency of occurrence (based on the values of dominance reported in the database) of a given feature during the training. Indeed, in a previous paper (Ursino et al. 2015) we demonstrated that a feature becomes salient if and only if its probability (i.e., P_{kj} for feature j in object k) is higher than the post-synaptic threshold [ϑ_{post} in Eq. (1)]. In this situation, in

fact, potentiation (i.e. point i above) overcomes depotentiation (point ii above). If enough epochs are provided, the synapses entering the feature rise up to their saturation level. Conversely, if the frequency of a feature is smaller than the post-synaptic threshold, depotentiation overcomes potentiation, and the incoming synapses fall to zero (i.e., a feature cannot be spontaneously evoked and is marginal).

In conclusion, it is worth noting that the level of saliency (i.e., whether a feature is salient or not based on its frequency of occurrence during the training) is strictly related with the value used for the post-synaptic threshold ϑ_{post} . The higher this threshold, the higher the frequency level required to have saliency. $\vartheta_{post} = 0.55$ (as in the present work) means that a feature must be perceived in more than 55% of presentations to become salient.

Synapses connecting the semantic and lexical nets

The excitatory synapses linking the semantic and the lexical units in both directions have been trained during phase ii using a low threshold for the lexical unit, and the higher threshold ($\vartheta = 0.55$) for the semantic unit, independently on whether this neuron was pre-synaptic or post-synaptic. This signifies that a word-unit must be active to ensure learning. Then synapse reinforces when the corresponding feature is present in the synaptic net, and weakens when the feature is absent. As a consequence, only salient features participating quite frequently to the object representation are spontaneously connected with the word-form.

Training the inhibitory synapses to the word-forms (anti-Hebbian learning) requires a different strategy. We need that a feature that never (or rarely) participates in the semantic of that object (say object1) but frequently participates in the semantic of other objects (say object2, object3, etc...) inhibits the word-form relative to object1 (for instance, the feature “it has two legs” should inhibit the word-form “Land animal”, the feature “it barks” should inhibit the word-form “cat”). To reach this objective, we decided to train the inhibitory synapses whenever the feature is active in the semantic net (this is the presynaptic unit, hence we assume $\vartheta_{pre} = 0$ in Eq. 1). Moreover, the threshold for the post-synaptic activity (that is the word-form) has been given a low value ($\vartheta_{post} = 0.05$). In this way, if the feature and the corresponding word-form are simultaneously active, the inhibitory synapse is dramatically reduced ($\Delta V_{ij} = \gamma_{ij} \cdot 0.95$ in Eq. 1, but remember that $\gamma_{ij} < 0$ for the anti-Hebbian learning). Whenever a feature is present without its word-form, the inhibitory synapse is just mildly increased ($\Delta W_{ij} = -\gamma_{ij} \cdot 0.05$). The consequence is that features that sometimes participate to the object (even if not salient) remove their inhibition. Only those features that rarely or never participate to the

object, but frequently participate to the semantic of other objects, send inhibition to object1.

Limits in the use of a fixed post-synaptic threshold

By using the Hebb rule with different pre-synaptic and post-synaptic thresholds (Eq. 1) we were able to differentiate between marginal and salient features, and between distinctive and totally shared features, on the basis of the frequency of their occurrence. The resulting behavior, however, is inadequate to deal with the saliency of a *partially shared feature*. Let us consider, for instance, the connection linking the features “it flies” and “it has two legs” for birds. In the simple taxonomy of Table 1 the first is partially shared (4/6 of birds) while the second is totally shared.

According to our taxonomy, using a post-synaptic threshold as high as $\theta_{post} = 0.55$ the feature “it has two legs” becomes salient for all six birds (since it occurs in 70% of cases, let us remember that we used $P_{kj} = 0.7$ for salient features and 0.4 for the marginal ones) and, after training, is always evoked within any bird. This is correct. But, what is the situation for the feature “it flies”? According to the previous taxonomy, this feature does not occur in the “Hen”, and in the “Rooster”, but it occurs with high frequency (70%) in flying birds (“Goose”, “Parrot”, “Owl”, “Pigeon”) and, after training, becomes salient for these four birds.

Let us now consider the connection linking “it flies” and “it has two legs”. After training, the feature “it has two legs” occur in 100% of birds, since it is salient for all of them, while the feature “it flies” occur in 66% of birds (since it is salient in 4/6 of them). But, using a threshold $\theta_{post} = 0.55$, a connection is progressively created not only from “it flies” to “it has two legs” (100% of post synaptic occurrences) but also from the feature “it has two legs” to “it flies” (66% of post-synaptic occurrences, i.e. in 4/6 of birds) and the latter becomes salient for all birds. This means that, in our database, after training any animal that flies is assumed to have two legs (correct), but also any animal with two legs is assumed to fly (incorrect, as the last assertion does not hold for the Hen and the Rooster).

A way to avoid this incorrect judgment (“the Hen and Rooster fly”) could be to increase the post-synaptic threshold up to a higher value (for instance 0.7 or 0.75). However, in this case no feature in the database would become dominant (or just a few features occurring randomly in more than 75% of situations would become dominant).

In order to solve this problem, we propose that the post-synaptic threshold should change with experience: in particular, the post-synaptic threshold should increase when

connecting features in a category (hence, features that are shared by several concepts), but remain smaller when connecting distinctive features of individual members. It is worth noting that the idea of a variable post-synaptic threshold is assumed by one of the most accredited models of synaptic plasticity to date, i.e., the BCM rule, but not directly with reference to attractor networks. Moreover, the idea that thresholds in Hebb rule should be variable with neuron “popularity” is also discussed in Kropff and Treves (2007) in the context of auto-associative nets (see the section “Discussion” for a more detailed analysis on this point).

In the following, we will assume that *the post-synaptic threshold in the Hebb rule increases with the number of occurrences of the pre-synaptic feature*. In this way, not only a shared feature in a category does not evoke a distinctive feature of an individual member, but also a totally shared feature (like “it has two legs” in the previous example) does not evoke a partially shared feature (like “it flies”).

This new mechanism will be explained in the next subsection.

Mechanisms for a variable post-synaptic threshold

Let us consider the synapse W_{ij}^{SS} linking a post-synaptic feature i with a pre-synaptic feature, j , in the semantic net.

According to the previous reasoning, we assume that the post-synaptic threshold depends on the number of occurrences of the pre-synaptic feature, hence it is a function of j . The Hebb rule becomes:

$$\Delta W_{ij} = \gamma_{ij}(x_i - \vartheta_{post,j})(x_j - \vartheta_{pre}) \quad (1')$$

where we used the subscript j in the post-synaptic threshold, $\vartheta_{post,j}$, to specify its dependence on the pre-synaptic feature. It is worth noting that, in writing Eq. 1', we omitted the superscripts S for brevity (we know that we refer just to the semantic net).

Let us denote with N_j the average number of occurrences of a feature j during an epoch (we remind that, during each epoch, all objects are presented once). Of course, after training, N_j will be greater for salient distinctive features than for marginal distinctive features, but even greater for shared features. Just to fix our ideas, let us consider four cases concerning the bird “parrot”:

- “It has a big beak”: this is a distinctive marginal feature of the animal “Parrot”. Hence, even after training, it is not spontaneously evoked, and occurs in just 40% of the parrots. Hence we have $N_j \cong 0.40$;
- “It repeats sounds”: this is a distinctive salient feature of the animal “Parrot”. Hence, at the beginning of the

training phase we have $N_j \cong 0.7$ (the feature occurs approximately in 70% of parrots), but after training it is evoked in 100% of parrots, and we have $N_j \cong 1.0$;

- “It flies”: this is a partially shared feature for four birds (“Duck”, “Parrot”, “Owl”, “Pigeon”) but never occurring in the “Hen” and “Rooster”. As a consequence, at the beginning of training we have $N_j \cong 0.7 \cdot 4 = 2.8$, but, after training, assuming it is not erroneously ascribed to the Hen and Rooster we have: $N_j \cong 4$
- “It has two legs”: this is a totally shared feature for all six birds. As a consequence, at the beginning of training we have $N_j \cong 0.7 \cdot 6 = 4.2$, but, after training we have: $N_j \cong 6$.

We suggest that the post-synaptic threshold should progressively increase with N_j . We can write

$$\vartheta_{post,j} = \begin{cases} \vartheta_{post-base} + \Delta\vartheta_{post} \cdot (N_j - 1) & \text{if } N_j > 1 \\ \vartheta_{post-base} & \text{if } N_j \leq 1 \end{cases} \quad (2)$$

where N_j is the average number of occurrences of the feature j (pre-synaptic) during an epoch, and $\vartheta_{post-base}$, $\Delta\vartheta_{post}$ are parameters, which set the basal value (that is the value for all distinctive features) and the rate of increment of the post-synaptic threshold, respectively. It is worth noting that a distinctive feature cannot have $N_j > 1$ and so the post-synaptic threshold is fixed at the basal level. Conversely, shared features have $N_j > 1$, and so $\vartheta_{post,j} > \vartheta_{post-base}$. This signifies that distinctive features can easily create synapses toward shared features (due to the presence of a low post-synaptic threshold), but not viceversa.

Finally, we assumed that the post-synaptic threshold can never overcome a maximum saturation level (say $\vartheta_{post-sat}$), i.e., we set

$$\vartheta_{post,j} \leq \vartheta_{post-sat} \quad \forall j \quad (3)$$

Let us now consider the possible benefits of the previous rule [Eqs. (2) and (3)]. We will use some features of the “parrot” again (Fig. 3).

By way of example, let us assume $\vartheta_{post-base} = 0.55$, $\Delta\vartheta_{post} = 0.1$ and $\vartheta_{post-sat} = 0.95$ (these values are just used for the present example; they are not those chosen for the subsequent simulations).

- For all distinctive features we have $N_j \leq 1$ and so $\vartheta_{post,j} = \vartheta_{post-base} = 0.55$. This value holds for both “it has a big beak” and “it repeats sounds”.
- For the feature “it flies” we have $N_j = 4$ and so,

$$\begin{aligned} \vartheta_{post,j} &= \vartheta_{post-base} + \Delta\vartheta_{post} \cdot (N_j - 1) \\ &= 0.55 + 0.1 \cdot 3 = 0.85 \end{aligned}$$

Fig. 3 A summary of the synapses that are created with the use of a variable post-synaptic threshold. We consider four exemplary features of the parrot. A feature totally shared with all birds: “it has two legs”; a partially shared feature: “it flies”; a salient distinctive feature: “it repeats sounds”; a marginal distinctive feature: “it has a big beak”. The boxes beside any synapse contain the frequency f of the post-synaptic feature after training, for all the cases where the pre-synaptic feature is present, and the post-synaptic threshold ϑ (computed via Eq. 2). A synapse is created only if the frequency f overcomes the post-synaptic threshold ϑ . Accordingly, thick lines represent strong synapses, able to excite the post-synaptic feature; thin dashed lines represent negligible synapses (close to zero)

- For the feature “it has two legs” we have $N_j = 6$ and so,

$$\begin{aligned} \vartheta_{post,j} &= \vartheta_{post-base} + \Delta\vartheta_{post} \cdot (N_j - 1) \\ &= 0.55 + 0.1 \cdot 5 > 0.95 \text{ hence we set } \vartheta_{post,j} \\ &= 0.95. \end{aligned}$$

The synapses that are created after training, using the previous parameter values, are summarized in Fig. 3. The distinctive features (“it repeats sounds” or “it has a big beak”) cannot receive synapses from the shared ones, since the post-synaptic threshold (which depends on the pre-synaptic feature) is too high. Conversely, the shared features receive synapse from the distinctive ones, since the post-synaptic threshold remains low. Moreover, the salient distinctive features receive synapse from the marginal ones, but not viceversa. Finally, and more important, the totally shared features (like “it has two legs”) receive a synapse from the partially shared features (like “it flies”) (in fact, the post synaptic threshold here is 0.85, and the feature occurs in 100% of cases for birds, after training). Conversely, the partially shared features (like “it flies”) do not receive a synapse from the totally shared features (like “it has two legs”) since the post-synaptic threshold is 0.95 and, in our database, “it flies” occurs in 66% of cases only (4/6 of birds).

In the following simulations, however, we used a higher value for the parameter $\Delta\vartheta_{post} = 0.5$. With this choice, the

final outcome does not change compared with Fig. 3, but learning becomes more robust. Indeed, this was necessary to increase quite rapidly the post-synaptic threshold for shared features at the beginning of training, to avoid the formation of undesirable synapses. In fact, when an unwished synapse is created, it always evokes the post-synaptic feature and makes it salient independently of the subsequent rise in post-synaptic threshold. This requires a high value of $\Delta\vartheta_{post}$. A comment on this aspect can be found in “Discussion” section.

Results

Simulations concern: (i) the training of the network and the analysis of the resulting synaptic patterns; (ii) the simulation of object naming and word recognition tasks; (iii) the simulation of semantic deficits, assuming a damage in the synapses and/or in the neurons within the semantic net.

We will describe the results obtained on the animal taxonomy (see Table 1) and those obtained on the object taxonomy (see the Supplementary Material, part II).

Synaptic training

Synapses in the semantic net

Some examples of the synapses entering into a feature, from the other features in the synaptic net after training are shown in Fig. 4 (each panel shows all features in the x-axis and the strength of the synapses entering a single feature in the y-axis). To help the comprehension, a list of features with the corresponding position is presented in Table 1. The two upper panels refer to two salient distinctive features (“It barks” for the dog and “It repeats sounds” for the parrot). These features receive strong synapses, close to the maximum saturation, from all other (salient and marginal) distinctive features of the same animal, but do not receive synapses either from distinctive features of other animals, or from shared features (for instance, the features “It has four legs” or “it is domestic” do not send synapses to the feature “It barks”). The left middle panels show the synapses entering into a distinctive marginal feature of the dog (“It chases”). The latter does not receive significant synapses from other features, hence is not spontaneously evoked. Finally, the other three panels show the synapses entering into shared features. The right middle panel shows synapses entering into a partially shared feature (“It flies”). It receives synapses only from the features of the four flying birds, but does not receive synapses neither from the features of the “Hen” and “Rooster”, nor from the totally shared features of birds (like “It has two legs”). This is the

effect of a variable post-synaptic threshold described above. (Not all birds fly!). Finally, the two bottom panels show the synapses entering into two totally shared features. In the bottom left, “It has two legs” receives synapses from all features of the six birds, including the totally shared features “it has feathers” and “it has wings” (however, it does not receive synapses from the features of “Land animals”). On the bottom right, the feature “it eats” receives synapses from all features of all animals.

An example of the synapses linking features in the artificial object network is presented in the Supplementary Material, part II.

Synapses linking the semantic and lexical nets

Some examples of the synapses entering the semantic net from the lexical one are presented in Fig. 5 (for the positions of the word-forms see Table 1). We show six panels: the upper concern two salient distinctive features, receiving a synapse only from the corresponding word-form. The middle ones show two features (“it is domestic” and “it flies”) that are partially shared. They correctly receive a synapse only from the specific members, but not from the category name. Finally, the bottom panels show two totally shared features, which receive synapses not only from all members in that category, but also from the word-form representing the category. Marginal features (not presented here) do not receive appreciable synapses by any word, and are not spontaneously evoked.

Some examples of the synapses entering a word form from the semantic net are shown in Fig. 6. It is worth noting that these include both excitatory and inhibitory contributions. Each word-form receives excitatory synapses only from all salient features that characterize the given concept (i.e., words denoting categories receive synapses only from salient features shared by all members in the category; words denoting members receive synapses from all the individual salient features, both distinctive and shared with other members). Furthermore, all word-forms receive a null synapse (neither excitatory nor inhibitory) from the marginal features of the given concept. All features that do not belong to the concept (for instance the feature “it barks” for the concept “cat”) send an inhibitory synapse, thus avoiding absurd concept naming. Finally, it is worth noting that the sum of all excitatory synapses entering a word-form is equal to 1 (maximal saturation). Due to the sigmoidal relationship describing neuron activation in the lexical area (see Supplementary Material, part I), this signifies that all salient features must be evoked in the semantic net to move the working point from inhibition to saturation and allow correct naming.

Fig. 4 The strength of the excitatory synapses entering into a single feature from the other features in the “Animal” semantic net. The following colors are used to describe the features within the panels: red: distinctive salient; cyan: distinctive marginal; blue: shared. The upper panels show synapses entering two salient distinctive features (the first of the dog, the second of the parrot). They receive synapses from all other distinctive features of the same animal. The left middle

panel represents synapses entering a distinctive marginal feature of the dog. It does not receive significant synapses. The right middle panel represents synapses entering a salient partially shared feature of birds (“it flies”). It receives synapses only from the features of flying birds. Finally, the two bottom panels represent synapses into salient totally shared features: the first receive synapses from all features of birds, the second from all animal features

Fig. 5 Some exemplars of synapses entering the features in the “Animal” semantic net from word-forms in the lexical net. The upper panels concern two salient distinctive features: they receive synapses only from the corresponding word-form. The middle panels concern two salient partially shared features; they receive a synapse only from

each specific member, but not from the category name. Finally, the bottom panels show two salient totally shared features, which receive synapses not only from all members in that category, but also from the word-form representing the category

Fig. 6 Some examples of the synapses entering a word form from the “Animal” semantic net. The upper and middle panels concern individual members. The two bottom panels concern two categories. The following colors are used to denote features belonging to the given concept: red: distinctive salient; cyan: distinctive marginal; blue: shared. It is worth noting that each word-form receives excitatory synapses (whose sum is normalized to 1) from all

distinctive salient features and all shared features that characterize a given concept, but a null synapse (neither excitatory nor inhibitory) from the marginal features of a given concept. Finally, all features that do not belong to the concept send an inhibitory synapse (black color), thus avoiding major errors in concept naming. (Color figure online)

Object naming and word recognition

Naming

During these trials, we provided the network with a single feature and observed the behavior in the semantic and lexical nets. Results can be summarized as follows:

- (i) If a distinctive feature (either salient or marginal) is excited, the semantic net evokes all salient features of the concept (without evoking any marginal feature, apart from the one possibly given as input); the corresponding word form is then excited in the lexical net. An example (in which the different panels represent different snapshots during a simulation) is given in Fig. 7.
- (ii) If we excite a totally shared feature, the semantic net evokes all totally shared salient features of that category. Features shared only by a few members (but not by all) are not evoked. The word-form of the category is excited in the lexical net. An example is given in Fig. 8.
- (iii) If a partially shared feature is used as input, the semantic net evokes the totally shared features of that category. A category word-form is evoked

only if the input feature was used with a specific category during training. For instance, the feature “it flies” evokes the category “Birds” since was used (but with low frequency $0.7 \cdot 4/6 = 0.47$) when teaching birds, hence is considered as a marginal feature for birds. Conversely, the feature “It eats grass” does not evoke the word-form “Land animals” since it was never used when teaching that category name. Indeed, this feature was used with the ad hoc category “Herbivore”, and so it evokes this word-form.

Word recognition

In these simulations we excited a single word-form in the lexical network, and observed which features were evoked in the semantic net. An example is given in Fig. 9. The results, obtained by providing all word forms to the lexical net, show that each word-form correctly evokes the salient features of that concept, but does not evoke marginal features (a few exceptions are commented in the paragraph “limitations” below). If the word-form represents a category, the features shared only by a few members (such as “it flies” or “it eats grass”) are treated as marginal and not evoked.

Fig. 7 Different temporal snapshots of an object naming simulation, in which a distinctive feature (in this case a marginal one for the rooster, “it wakes up people”) is given as input to the semantic net. The following colors are used to denote features belonging to a given concept: red: distinctive salient; cyan: distinctive marginal; blue: shared. All salient features of the rooster (including those shared with

other members) are evoked in the semantic net (upper figure in each panel), and finally the correct word-form is evoked in the lexical net (bottom figure in each panel). The other marginal feature is not evoked. (Color figure online)

Limitations

We repeated the previous trials using all features and word-forms as single inputs: The model works correctly in all naming and word recognition tasks. Indeed, we observed just one anomalous behavior, which occasionally occurs both in the animal and object models. When a feature is shared by two concepts, and is salient for the first and marginal for the second, after training sometimes the feature becomes salient for both. This is the case of the features “It grazes” and “it lives on farms”, which become salient not only for the “Cow” but also for the “Sheep”. Similarly, the feature “It is used for holding” becomes salient not only for the “Bookcase” but also for the “Pot”, and the property “It has a wooden handle” becomes salient not only for the “Hammer” but also for the “Broom”. However, it is worth noting that, in the original data-set, these features were close to the boundary between saliency and marginality both for the sheep, the pot and the broom; in particular, they were named by several subjects during feature naming tasks. Hence, the model predictions are largely supported by data. Moreover, it is worth noting that some features (like “It is reared”, salient for the “Hen” but

marginal for the “Goose”, or “It is made of steel” salient for the “Pot” marginal for the “Fork”), remain marginal for the second concept.

Effect of damages in the network

Finally, we simulated the effect of a damage in the semantic network. Two kinds of damage have been tested:

- (i) We assumed that a given percentage of synapses (60%), randomly chosen, were damaged. Some exemplary simulations were performed assuming 30, 60 and 70% of synapse reduction compared with the basal value. During each session, we provided single features to the semantic net, and observed which other features were evoked. Since the results of these simulations depend on the random choice of damage, we repeated ten different trials per each concept. The results, show that at small synapse damage (30%) the network maintains its capacity to evoke salient features and the correct word-form. At higher percentages of damage (summarized in Table 2), we observed that the distinctive features are much more precarious than

Fig. 8 Different temporal snapshots of a naming simulation, in which a totally shared feature (in this case the feature “it has wings”) is given as input to the semantic net. The blue color is used to denote the totally shared features of the concept. All totally shared features of the bird are evoked in the semantic net (upper figure in each panel), and

finally the correct word-form is evoked in the lexical net. It is worth noting that partially shared features of birds (like “it flies” or “it flutters”, shown in cyan) are not evoked, i.e., they are treated as marginal. (Color figure online)

Fig. 9 Different snapshots of a word-recognition task, in which a word-form is given as input to the lexical area (in this example the word “Cow”). The following colors are used to denote features belonging to the given concept: red: distinctive salient; cyan:

distinctive marginal; blue: shared. All shared and distinctive salient features are evoked in the semantic area. Marginal features are not evoked. (Color figure online)

Table 2 Percentage of correct semantic reconstruction in response to a distinctive input feature, when a given percentage of synapses in the semantic network (60% of total) have been randomly damaged

	Member	Category	Super-category	None
60% of synapses damaged by 60%				
Land animals	57.74	39.68	2.25	0.33
Birds	64.5	30.83	2.5	2.17
Kitchen items	61.88	34.02	3.60	0.50
Tools	73.23	–	23.3	3.47
Furniture	68.77	30.88	0.35	0
	65.22	33.48		1.30
	Member	Category	Super-category	None
60% of synapses damaged by 70%				
Land animals	52.78	34.88	9.47	2.87
Birds	50.41	39.75	5.42	4.42
Kitchen items	49.35	26.08	19.75	4.82
Tools	59.47	–	31.53	9.00
Furniture	60.95	20.35	15.85	2.85
	54.59	40.62		4.79

In the upper table the damage was a 60% reduction of the synapse basal value. In the bottom table the damage was 70% of the basal value. 10 trials were performed per each animal and per each object. The first column represents the percentage of correct reconstruction of all salient distinctive features of the given concept. The second column represents the percentage of cases in which only the features of the superordinate concept were recovered (for instance the features of bird for the owl, the features of furniture for the armchair) but the distinctive features were not recovered. The third column represents the percentage of cases when only the general features of the category (animal or artificial object) were maintained. Finally, the last column represent the no-response cases

the shared ones. In many cases (33% of cases with 60% of damaged synapses; 41% of cases with 70% of damaged synapses), in response to a distinctive feature, the network fails to evoke other distinctive features of the same object, but still maintains the capacity to evoke the shared features of the superordinate concept.

- (ii) We assumed a damage to some individual neurons in the semantic net (i.e., neurons coding for specific distinctive features) by forcing its inhibition despite the presence of positive inputs. In order to unmask the role of marginal versus salient features, we provided four features of the same object as input to the network, two salient and two marginal, and we looked whether the network can recognize the correct object name. Ten simulations per concept were performed again. As expected (see Table 3), a damage to one or more neurons coding for a marginal feature does not compromise the naming task in 100% of simulations. Conversely, a damage to even one neuron coding for a salient feature causes the incapacity to produce the correct word in 100% of simulations.

Taken together, these simulations stress that the network is much more fragile in the case of damage to distinctive

features than in the case of shared ones, and that loss of salient attributes jeopardizes the object naming task, whereas the loss of marginal features is uninfluential.

Discussion

In the present paper, we made use of a neuro-computational model to improve our understanding of the featural organization of semantic memory. The main aim was to investigate the role of different kinds of features in the representation of concepts, a central issue for the investigation of semantic memory disorders (Marques 2005; Tyler and Moss 2001; Catricalà et al. 2015c). In particular, we implemented new synaptic learning rules in an attractor network, in order to successfully deal with two important differences, i.e. totally versus partially shared features and marginal versus salient distinctive features.

The attractor network was trained using animals and tools concepts, and including totally and partially shared features, as well as distinctive features with different saliency, on the basis of a previous collected norms (Catricalà et al. 2015a). The model includes semantic and lexical layers, coding for object features and word-forms respectively. To deal with partially or totally shared features and the different role of salient versus marginal distinctive

Table 3 Results of model simulations of naming tasks, performed after 5% random damage of units coding for distinctive (marginal or salient) features. 10 trials have been performed per each concept

Damages in the features of the selected animal	Number of trials	Number of errors	Number of correct	Efficacy
120 trials on animals (10 per each animal) with 5% of damaged neurons				
Damage of at least one distinctive salient	28/120	28	0	100%
Damage of marginal distinctive only	8/120	0	8	0%
Damages in the features of the selected object	Number of trials	Number of errors	Number of correct	Efficacy
110 trials on objects (10 per each object) with 5% of damaged neurons				
Damage of at least one distinctive salient	23/110	23	0	100%
Damage of marginal distinctive only	26/110	0	26	0%

In each trial, two salient and two marginal features (not necessarily those damaged) of the given concept were given as input. In all cases, a damage of a salient distinctive feature leads to a no-word response (i.e., the network fails to recognize the concept). In all cases, a damage of a marginal distinctive feature, without any damage in salient distinctive features, leads to correct word-naming. It is worth noting that the number of trials with damaged marginal features are much more numerous in the object than in the animal model, since objects in our model have a greater number of marginal features (compare Table 1 and Table S1 in the Appendix)

features, the connections among nodes were strongly asymmetrical (Ursino et al. 2015). In particular, in order to account for the feature saliency, synapses were created using Hebb rules of potentiation and depotentiation, using different pre-synaptic and post-synaptic thresholds in the training rule (Ursino et al. 2015). Accordingly, the saliency of a feature depends on the number of participants who listed the feature for a concept. Secondly, a variable post-synaptic threshold, which automatically changed to reflect the feature frequency in different concepts (i.e., how many concepts share a feature) was used to account for features not shared by all the members of a category, overcoming the known limit of previous works using attractor networks with Hebb rule (McRae et al. 1997).

In the following, the present Hebb rule is related to previous work on Hebbian learning paradigms in autoassociative nets. Then, the meaning of the results is critically assessed.

Hebbian learning paradigms and autoassociation

The use of Hebbian learning rules in autoassociative networks has a long tradition in neural computation. After the pioneering works by Hopfield (1982, 1984) and Amit et al. (1985), a fundamental historical problem was to devise a learning rule which is local (i.e., Hebbian) and can store both uncorrelated and correlated patterns, with a large basin of attraction. To this end, Diederich and Oppen (1987) and Krauth and Mézard (1987) proposed the use of “learning cycles”, in which each pattern is stored by a repeated presentation in a sequence of learning steps. This procedure was in contrast to previous learning rules traditionally used in associative nets, where each pattern was memorized during a single learning event. In the present

study we also used a learning cycle (with one thousand repetitions each) to store patterns but, as a novel aspect, we assumed a different probability for the different features in each stored pattern.

In order to improve the learning rule, some authors proposed the use of thresholds for the pre-synaptic and/or the post-synaptic activities. Tsodyks and Feigel'man (1988) used identical pre-synaptic and post-synaptic thresholds for all neurons, equal to the mean values of the patterns, as quantified by the sparseness. Others underlined the advantage of a variable post-synaptic threshold in the Hebb rule (see also Dayan and Abbott 2001): this is at the basis of one of the most popular rules for training synapses, i.e., the Bienenstock, Cooper and Munro (BCM) rule (Bienenstock et al. 1981). In particular, in the BCM rule the post-synaptic threshold is modified as a function of the average value of neuron post-synaptic activity. Increasing the threshold when post-synaptic activity is frequently high limits excessive synapse learning and avoids instability. The authors, however, originally did not use this rule in autoassociative networks. More recently, Kropff and Treves (2007), to handle correlation in associative nets, proposed a similar modification of the Hebb rule, in which the learning thresholds are individually specified for each neuron, and the pre-synaptic threshold can be different from the post-synaptic one. Based on a signal to noise analysis, the authors reached the conclusion that either the pre- or the post-synaptic thresholds must be equal to the average value of neuron activity over all stored patterns, i.e., to its “popularity”. This rule is quite similar to the present one. However, the authors chose to modulate the pre-synaptic threshold, whereas we demonstrated a pivotal role for the post-synaptic one, at least in category formation. Finally, Fusi et al. (2005) and Fusi and Abbott (2007) recently

analyzed the effect of bounded synapses, including the case when the synapse adjustments depend on synaptic strength and vanish at the boundaries (a characteristic shared by our model too). Their final suggestion is to use a cascade model (Fusi et al. 2005), in which synapses have multiple states and exhibit dynamics over a wide range of time-scales. This may be a further improvement of our learning rule.

Accounting for totally and partially shared features

Feature-based theories suggest that the semantic features may account for the category level through features that are shared by many concepts in a specific category, while distinctive features allow to the identification of closely related concepts. In the case of progressive loss of semantic memory observed in neurodegenerative disorders, distinctive properties of objects are lost at earlier stages of the dementia, while shared properties remain preserved for a longer time (Alathari et al. 2004; Catricalà et al. 2015c; Duarte et al. 2009; Garrard et al. 2005; Giffard et al. 2001, 2002; Laisney et al. 2011). In attractor networks using a Hebbian rule, the reinforcing correlations among shared features are crucial to account for the preservation of shared features. On the other hand, they represent a problem in the case of features shared by several but not all the concepts belonging to the same semantic category. In previous models (McRae et al. 1997) attractor dynamics led to an incorrect activation of additional features, due to its correlation with the other features, for example bird-like features for the concept jet. These kinds of models naturally learn how features co-occur in concepts on which they are trained. In this work, we implemented an important novelty: the post-synaptic threshold in the Hebb rule varied automatically as a function of presynaptic frequency (i.e., reflecting how many concepts share a feature). As illustrated in Fig. 3, this variability is essential to avoid that a feature shared by many (but not all) members in a category is inherited by other members. At variance with the BCM rule, which uses a variable post-synaptic threshold to avoid instability, in our model instability is limited by assuming an upper saturation for the synapses, i.e., reducing the learning rate when a synapse approaches saturation. The use of a variable post-synaptic threshold has a different objective: to control saliency, so that saliency for shared features in a category follows a different pattern compared with distinctive features. To this end, our post-synaptic threshold depends on the average value of the *pre-synaptic activity*. The higher this value, the higher the threshold. This means that a shared feature in a category must occur in almost all members to be ascribed to the category (i.e., to receive synapses from all the other members), and that a

shared feature does not evoke distinctive properties of individual members.

It is worth noting that our rule is biologically plausible. In particular, this is a local rule, i.e., it just exploits information (the values of pre-synaptic and post-synaptic activities, and the synapse strength) already present in situ.

Indeed, some values in the Hebb rule, especially those regarding the adjustments of the threshold ($\vartheta_{post-base}$, $\Delta\vartheta_{post}$ and $\vartheta_{post-sat}$ in Eqs. 2 and 3), must be chosen quite carefully to have an optimal behavior. However, this is not a significant limitation of the proposed rule. The correct values can be easily chosen assuming three-requirements: (i) the basal value is the same used for differentiating marginal versus salient distinctive features, i.e. it is equal to the frequency required to have saliency for distinctive features (55% in the present work); (ii) the value of $\vartheta_{post-sat}$ must be just a little small than 1, to allow that only features shared by almost all members of a category are ascribed to the category (hence we used 0.95 in this work, this signifies that a 5% of exceptions are tolerated); (iii) the most problematic parameter is $\Delta\vartheta_{post}$. In fact, if its value is too low, some unwished synapses are created at the beginning of training. We observed that 0.5 is a suitable value to obtain a correct synapse learning with the present two taxonomies. We have not tested the robustness of this parameter in different conditions (for instance using different probabilities for salient and marginal features, or assuming a different number of totally shared vs. partially shared features in a category). Future studies should be concentrated to test the robustness of this parameter value and, in case, to design methods for automatic adjustment of its value to optimally reflect the structure of the data-base. Another problem is how these parameters (especially $\vartheta_{post-base}$) may be physiologically changed to reach a desired behavior (for instance, to modulate the saliency of a given property on the basis of emotional, attentive or contextual influences).

Accounting for salient and marginal distinctive features

Distinctive features are important to differentiate between closely related concepts, typically members of the same semantic category, as they occur in only one or a very few concepts. The crucial role played by distinctive features in semantic processing has been reported in several studies involving both healthy subjects (Marques 2005; Mirman and Magnuson 2009) and patients (Alathari et al. 2004; Duarte et al. 2009; Garrard et al. 2005; Laisney et al. 2011; Rogers et al. 2004), as well as in connectionist models (Cree et al. 2006; Rogers et al. 2004; Mirman and Magnuson 2009; Moss and Tyler 2000). Not all the distinctive features, however, have the same saliency for the

representation of a concept (Cree et al. 2006). Some distinctive features are not highly salient, as they are infrequently reported in feature-norming tasks, and presumably do not play a prominent role in the representation of concept. Production frequency (or dominance, referring to the number of participants who listed a feature) is a strong predictor of feature verification latency (Ashcraft 1978; McRae et al. 1997). Smith and co-workers (1995) showed that AD patients were more impaired in attributes with low dominance. According to Sartori and Lombardi (2004), considering both distinctiveness and dominance, it is possible to identify features with high relevance, namely those most relevant for object identification. Distinctive features may be thus distinguished in salient and marginal features, with the former most important in the identification of a concept. In a recent study, we have reported that distinctive features with high values of semantic relevance were lost only in AD and SD patients with naming impairment (Catricalà et al. 2015c). Despite its importance, dominance is generally not incorporated in attractor models (McRae et al. 1997), thus preventing the discrimination between salient versus marginal features (but see Ursino et al. 2015). In a previous version of this model, we were able to distinguish between salient and marginal features within the same concept, using different pre-synaptic and post-synaptic thresholds in the Hebb rule and so training asymmetrical synapses. In this work we tested the model by simulating both naming tasks (in which, starting from partial information, the overall semantic information is reconstructed and the corresponding word-form is evoked), and word recognition tasks (where a single word-form evokes all salient features of the corresponding concept). The results show that the model, with the same values of parameters, works in a fully satisfactory way; all simulations confirm that the network can discriminate between salient and marginal features. We noticed just one anomalous behavior: if a shared feature is salient for one concept and marginal for another (according to the frequency of occurrence) it sometimes becomes salient for both after training. This result may represent a possible testable prediction. Indeed, looking at the data-set, we observed that marginal features that become salient after training are close to the boundary between saliency and marginality, (i.e., they have a relatively high dominance based on the number of participants who listed that feature for the concept).

The idea is that a feature is salient if spontaneously evoked when thinking to the given concept. Hence, we chose those features, which are more frequently listed in the feature generation task. It is worth noting that, after training, we simulated a prototypical subject, extracted from the data set on the basis of an average behavior (i.e., the behavior common to most subjects). Of course, it is

possible to simulate different subjects with different semantic representations by performing an “ad hoc” training for each of them, using different frequencies to mimic a personal experience of the word (see Yee and Thompson-Schill 2016). Hence, in perspective the model can be used also to simulate individual variability. In this work, we used two constant frequencies for each feature within a concept (40% for marginal and 70% for salient), thus attaining a fixed final semantic representation. This choice was motivated by simplicity, to facilitate the final analysis of results. Our model is actually compatible with a flexible, experience-based representation of semantics. In particular, one may train the model assuming a flexible experience, in which the frequency of feature occurrence changes over time. In this condition, the model predicts that a marginal feature may subsequently become salient, if its frequency increases. Conversely, a salient feature, after the formation of synapses, is always spontaneously evoked within a concept, thus maintaining a 100% frequency.

Simulating semantic memory disorders

Another important aspect of the model is the capacity to simulate pathological conditions, characterized by progressive impairment of performances based on semantic memory. This was achieved by damaging the network. Although these simulations are just qualitative, their results exhibit some interesting correspondences with clinical results, and point out the possible use of the model for the analysis of neurological disorders (such as SD and AD) characterized by semantic impairment.

Damage in the network was simulated in two different ways. First, we mimicked a damage to the synapses of the semantic net. To this end, we weakened a given percentage of randomly chosen synapses. The results suggest that the network is robust against synaptic damage (a property typical of most parallel distributed systems); more than half of synapses can be reduced, up to 60–70% of the basal value (i.e. with a – 30% or – 40% reduction) without apparent information loss. However, when synaptic damage is further increased, the network often fails to evoke salient distinctive features than shared features, hence failing the object naming. Such results agree with previous clinical and neurocomputational data (Alathari et al. 2004; Rogers et al. 2004). The model ascribes this difference to the fact that shared features receive a larger number of synapses, not only from distinctive features of the individual member, but also from all shared features of the superordinate concept. It can thus be activated by more synaptic paths, becoming robust against synaptic damage.

In a further set of simulations, we assumed that some neurons in the semantic net were unresponsive to stimuli: this corresponds to a loss of features. In these conditions,

we simulated a naming to definition task by providing four features (two marginal and two salient) as input. As expected on the basis of the model structure, a damage in a salient feature jeopardizes the object naming task. Conversely, the naming task is correctly solved despite a loss of marginal features. This behavior agrees with a recent study by Catricalà et al. (2015c), showing that patients with naming impairment have lost features with high salience. Taken together, these results suggest that the semantics of a concept include information necessary for uniquely identify the concept, as well as additional information not immediately required for object identification (Chertkow and Bub 1990; Miller and Johnson-Laird 1976).

Comparison with recent results on semantic memory

In our model, units do not represent individual neurons, but rather a population of neurons (not necessarily contiguous) participating to the representation of the same feature. Their activation (described in the model with a single activity for the sake of simplicity) is the product of a pre-processing stage, which extracts the main pieces of information in the external data and decides on the presence/absence of a feature. For simplicity sake, these pre-processing steps are not directly included in the present equations, but they are an implicit assumption behind the model. These pre-processing steps may involve both purely modal attributes (*‘it meows’* or *‘it has black and white stripes’*) or more complex encyclopedic information (*‘it chases in the night’*). In the present work, we used general features to represent the semantics of real concepts (animals and artificial objects) taken from a data set, and so we abandoned the topological representation used in previous model versions (Ursino et al. 2010, 2015). This choice, of course, has the enormous advantage to allow a direct comparison between model results and results of psychological tests. The single features used in the present representation are not elementary perceptions, but complex elements which might be decomposed into simpler attributes, until a modal and topologically organized representation is achieved.

Current neural models of semantic memory suggest that the semantic network is distributed over multiple areas in the brain, with a complex topological organization (Huth et al. 2016). The authors mapped the semantic selectivity across the cortex, using a data-driven approach, and showed that the semantic system is represented by an intricate network, including temporal parietal and frontal regions, with different areas represent information about specific semantic domains. These data are not in opposition with our model. In fact, it may be expected that the different features used in our model are coded in different

areas, giving rise to a complex distributed network in the cortex.

Current models of semantic memory include modality specific areas, convergence zones and/or transmodal hubs. There is no agreement, however, on the number, identity and organization of these last regions (Damasio 1989; Lambon Ralph et al. 2017; Bonner et al. 2013; Fernandino et al. 2015). Lambon-Ralph et al. in a series of recent papers (Lambon Ralph et al. 2017; Patterson et al. 2007), hypothesized that semantic cognition exploits both modal and trans-modal representations, conceptualizing the semantic network as a “hub and spoke” structure including both modality-specific primary and association areas and a crucial transmodal hub. In particular, the authors postulated that a trans-modal hub works like a “convergence zone” and sends back synapses to reactivate modality-specific information distributed across different cortices. While in our model the lexical region works as the hypothesized convergence zone, a hierarchical organization of multiple convergence zones could also accommodate the results (Fernandino et al. 2015).

In conclusion, our model is compatible with a semantic conceptual system, which incorporates modal cortical regions in sensory, motor, language, and emotional areas, supplemented by transmodal regions, which integrate the features into more abstract heteromodal representations. The idea is that semantic processing includes multiple steps, from simpler features to more complex concepts, as proposed by several authors (Damasio 1989; Lambon Ralph et al. 2017; Fernandino et al. 2015).

Limitations of the model

Finally, we wish to acknowledge some limitations/simplifications of the present model. These can be subdivided into limitations in the structure of the network, and limitations in the training procedure.

In the case of the structure, we did not include auto-associative synapses linking word-forms in the lexical layer. This choice has been adopted to simplify the analysis of network behavior, assuming that most of the connections (and so the reconstruction of the concept) occur at the semantic level. Nevertheless, it is possible that synapses are created between frequently co-occurring words in lexicon, via a temporal Hebb rule, thus favoring the link among concepts.

A more complete semantic net should incorporate more than a single processing stage, such as modal layers, extracting elementary features from sensory, motor or emotional regions. These can converge encoding a more abstract feature representation. Only the last part of this semantic net is simulated in this work, whereas the previous version of our model (Ursino et al. 2015) was oriented

to the simulation of topologically organized regions, involved in concept embodiment.

The training procedure can also be modified, to incorporate new aspects. In this paper, we first trained the semantic net, using input features with a given statistics, and only subsequently (when the semantic was assessed) we trained the connections between the lexical and semantic nets. This implies that the subject first learns the concept meaning (in an unsupervised way, i.e. simply looking at the environment), and only subsequently links the meaning to word-forms. Future work may try a different strategy, in which semantics continues to evolve during the second phase, from interaction with word-forms. This may be useful, for instance, to improve the simulation of the acquisition of encyclopedic features, which are not directly acquired from the environment, but from a supervisor, and to achieve a more refined model of semantics, in which language modulates ongoing cognitive and perceptual processing (Lupyan 2012).

Finally, additional aspects of semantic memory have been analyzed recently, and may represent future challenges for neurocomputational models. Among the others, the semantic level can interact with the acoustic level (Li et al. 2017) and with the syntactic one (Malaia and Newman 2015) and both aspects are important to allocate attentional resources. Mizraji and Lin (2015) stressed the importance of coding spatial and temporal relations in autoassociative networks, and emphasized the role of the context. Finally, some authors stressed the need for a symbolic level, between the physiological and the cognitive ones (Bonzon 2017). Although these aspects are well beyond the objectives of the present work, they may be food for future modeling efforts.

Funding Funding for the position by Cristiano Cuppini was provided by the Italian Ministry of Education, Project FIRB 2013 (Fondo per gli Investimenti della Ricerca di Base-Futuro in Ricerca) RBFR136E24. The funding source is not directly involved in any particular choice (study design, analysis of data, writing) on the present manuscript.

References

- Alathari L, Trinh Ngo C, Dopkins S (2004) Loss of distinctive features and a broader pattern of priming in Alzheimer's disease. *Neuropsychology* 18:603–612
- Allport DA (1985) Distributed memory, modular subsystems and dysphasia. In: Newman SK, Epstein R (eds) *Current perspectives in dysphasia*. Churchill Livingstone, Edinburgh, pp 207–244
- Amit DJ, Gutfreund H, Sompolinsky H (1985) Spin-glass models of neural networks. *Phys Rev A* 32(2):1007–1018
- Ashcraft MH (1978) Property norms for typical and atypical items from 17 categories: a description and discussion. *Mem Cogn* 6(3):227–232
- Bienenstock EL, Cooper LN, Munro PW (1981) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex (No. 1). Brown Univ, Providence RI
- Bonner MF, Peelle JE, Cook PA, Grossman M (2013) Heteromodal conceptual processing in the angular gyrus. *Neuroimage* 71:175–186
- Bonzon P (2017) Towards neuro-inspired symbolic models of cognition: linking neural dynamics to behaviors through asynchronous communications. *Cogn Neurodyn* 11(4):327–353
- Cappa SF (2008) Imaging studies of semantic memory. *Curr Opin Neurol* 21(6):669–675
- Catricalà E, Della Rosa PA, Ginex V, Mussetti Z, Plebani V, Cappa SF (2013) An Italian battery for the assessment of semantic memory disorders. *Neurol Sci* 34(6):985–993
- Catricalà E, Della Rosa PA, Plebani V, Vigliocco G, Cappa SF (2014) Abstract and concrete categories? Evidences from neurodegenerative diseases. *Neuropsychologia* 64:271–281
- Catricalà E, Ginex V, Dominici C, Cappa S (2015a) A new comprehensive set of concept feature norms. Special Issue in Honour of J. Frederico Marques. *Rev Port Psicol* 44:111–120
- Catricalà E, Della Rosa PA, Parisi L, Zippo AG, Borsa VM, Iadanza A, Castiglioni I, Falini A, Cappa SF (2015b) Functional correlates of preserved naming performance in amnesic Mild Cognitive Impairment. *Neuropsychologia* 76:136–152
- Catricalà E, Della Rosa PA, Plebani V, Perani D, Garrard P, Cappa SF (2015c) Semantic feature degradation and naming performance. Evidence from neurodegenerative disorders. *Brain Lang* 147:58–65
- Chertkow H, Bub D (1990) Semantic memory loss in dementia of Alzheimer's type. *Brain* 113(2):397–417
- Cree GS, McNorgan C, McRae K (2006) Distinctive features hold a privileged status in the computation of word meaning: implications for theories of semantic memory. *J Exp Psychol Learn Mem Cogn* 32(4):643
- Cuppini C, Magosso E, Ursino M (2009) A neural network model of semantic memory linking feature-based object representation and words. *BioSystems* 96(3):195–205
- Damasio AR (1989) The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput* 1(1):123–132
- Dayan P, Abbott LF (2001) *Theoretical neuroscience*, vol 806. MIT Press, Cambridge
- Devlin JT, Gonnerman LM, Andersen ES, Seidenberg MS (1998) Category-specific semantic deficits in focal and widespread brain damage: a computational account. *J Cogn Neurosci* 10(1):77–94
- Diederich S, Oppen M (1987) Learning of correlated patterns in spin-glass networks by local learning rules. *Phys Rev Lett* 58(9):949–952
- Duarte LR, Marquie L, Marquie JC, Terrier P, Ousset PJ (2009) Analyzing feature distinctiveness in the processing of living and non-living concepts in Alzheimer's disease. *Brain Cogn* 71(2):108–117
- Fernandino L, Binder JR, Desai RH, Pendl SL, Humphries CJ, Gross WL, Conant LL, Seidenberg MS (2015) Concept representation reflects multimodal abstraction: a framework for embodied semantics. *Cereb Cortex* 26(5):2018–2034
- Fusi S, Abbott LF (2007) Limits on the memory storage capacity of bounded synapses. *Nat Neurosci* 10(4):485–493
- Fusi S, Drew PJ, Abbott LF (2005) Cascade models of synaptically stored memories. *Neuron* 45(4):599–611
- Garrard P, Ralph MAL, Patterson K, Pratt KH, Hodges JR (2005) Semantic feature knowledge and picture naming in dementia of Alzheimer's type: a new approach. *Brain Lang* 93(1):79–94
- Giffard B, Desgranges B, Nore-Mary F, Lalevée C, de la Sayette V, Pasquier F et al (2001) The nature of semantic memory deficits in Alzheimer's disease. New insights from hyperpriming effects. *Brain* 124:1522–1532

- Giffard B, Desgranges B, Nore-Mary F, Lalevee C, Beaunieux H, de la Sayette V et al (2002) The dynamic time course of semantic memory impairment in Alzheimer's disease: clues from hyperpriming and hypoprimer effects. *Brain* 125:2044–2057
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* 79(8):2554–2558
- Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci* 81(10):3088–3092
- Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532(7600):453–458
- Krauth W, Mézard M (1987) Learning algorithms with optimal stability in neural networks. *J Phys A Math Gen* 20(11):L745–L752
- Kropff E, Treves A (2007) Uninformative memories will prevail: the storage of correlated representations and its consequences. *HFSP J* 1(4):249–262
- Laisney M, Giffard B, Belliard S, De La Sayette V, Desgranges B, Eustache F (2011) When the zebra loses its stripes: semantic priming in early Alzheimer's disease and semantic dementia. *Cortex* 47(1):35–46
- Lambon Ralph MAL, Jefferies E, Patterson K, Rogers TT (2017) The neural and computational bases of semantic cognition. *Nat Rev Neurosci* 18(1):42–55
- Li X, Zhang Y, Li L, Zhao H, Du X (2017) Attention is shaped by semantic level of event-structure during speech comprehension: an electroencephalogram study. *Cogn Neurodyn* 11(5):467–481
- Lupyan G (2012) Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front Psychol* 3:54
- Malaia E, Newman S (2015) Neural bases of syntax–semantics interface processing. *Cogn Neurodyn* 9(3):317–329
- Marques JF (2005) Naming from definition: the role of feature type and feature distinctiveness. *Q J Exp Psychol* 58(4):603–611
- Marques JF, Canessa N, Siri S, Catricalà E, Cappa S (2008) Conceptual knowledge in the brain: fMRI evidence for a featural organization. *Brain Res* 1194:90–99
- Marques JF, Cappa SF, Sartori G (2011) Naming from definition, semantic relevance and feature type: the effects of aging and Alzheimer's disease. *Neuropsychology* 25(1):105
- Martin A (2007) The representation of object concepts in the brain. *Annu Rev Psychol* 58:25–45
- Masson ME (1995) A distributed memory model of semantic priming. *J Exp Psychol Learn Mem Cogn* 21(1):3
- McRae K, de Sa VR, Seidenberg MS (1997) On the nature and scope of featural representations of word meaning. *J Exp Psychol Gen* 126(2):99
- McRae K, Cree GS, Seidenberg MS, McNorgan C (2005) Semantic feature production norms for a large set of living and nonliving things. *Behav Res Methods* 37(4):547–559
- Miller GA, Johnson-Laird PN (1976) *Language and perception*. Belknap Press, Cambridge
- Mirman D, Magnuson JS (2009) The effect of frequency of shared features on judgments of semantic similarity. *Psychon Bull Rev* 16(4):671–677
- Mizraji E, Lin J (2015) Modeling spatial–temporal operations with context-dependent associative memories. *Cogn Neurodyn* 9(5):523–534
- Moss HE, Tyler LK (2000) A progressive category-specific semantic deficit for non-living things. *Neuropsychologia* 38(1):60–82
- Moss HE, Tyler LK, Devlin JT (2002) The emergence of category-specific deficits in a distributed semantic system. In: Forde E, Humphreys G (eds) *Category-specificity in mind and brain*. Psychology Press, Hove, pp 115–148
- O'Connor CM, Cree GS, McRae K (2009) Conceptual hierarchies in a flat attractor network: dynamics of learning and computations. *Cogn Sci* 33:665–708
- Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8(12):976–988
- Perri R, Zannino GD, Caltagirone C, Carlesimo GA (2011) Semantic priming for coordinate distant concepts in Alzheimer's disease patients. *Neuropsychologia* 49(5):839–847
- Perri R, Zannino G, Caltagirone C, Carlesimo GA (2013) Alzheimer's disease and semantic deficits: a feature-listing study. *Neuropsychology* 37(1):99–107
- Rogers TT, Lambon Ralph MA, Garrard P, Bozeat S, McClelland JL, Hodges JR, Patterson K (2004) Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol Rev* 111(1):205
- Sartori G, Lombardi L (2004) Semantic relevance and semantic disorders. *J Cogn Neurosci* 16(3):439–452
- Sartori G, Gnoato F, Mariani I, Prioni S, Lombardi L (2007) Semantic relevance, domain specificity and the sensory/functional theory of category-specificity. *Neuropsychologia* 45(5):966–976
- Silveri MC, Gainotti G (1988) Interaction between vision and language in category-specific semantic impairment. *Cogn Neuropsychol* 5(6):677–709
- Smith S, Faust M, Beeman M, Kennedy L, Perry D (1995) A property level analysis of lexical semantic representation in Alzheimer's disease. *Brain Lang* 49(3):263–279
- Tsodyks MV, Feigel'Man MV (1988) The enhanced storage capacity in neural networks with low activity level. *EPL* 6(2):101–105
- Tyler LK, Moss HE (2001) Towards a distributed account of conceptual knowledge. *Trends Cogn Sci* 5(6):244–252
- 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 Lang* 75(2):195–231
- Ursino M, Cuppini C, Magosso E (2010) A computational model of the lexical semantic system based on a grounded cognition approach. *Front Psychol* 1:221. <https://doi.org/10.3389/fpsyg.2010.00221>
- Ursino M, Cuppini C, Magosso E (2011) An integrated neural model of semantic memory, lexical retrieval and category formation, based on a distributed feature representation. *Cogn Neurodyn* 5(2):183–207
- Ursino M, Cuppini C, Magosso E (2013) The formation of categories and the representation of feature saliency: analysis with a computational model trained with an Hebbian paradigm. *J Integr Neurosci* 12(04):401–425
- Ursino M, Cuppini C, Magosso E (2015) A neural network for learning the meaning of objects and words from a featural representation. *Neural Netw* 63:234–253
- Warrington EK (1975) The selective impairment of semantic memory. *Q J Exp Psychol* 27(4):635–657
- Yee E, Thompson-Schill SL (2016) Putting concepts into context. *Psychon Bull Rev* 23(4):1015–1027
- Zannino GD, Perri R, Pasqualetti P, Caltagirone C, Carlesimo GA (2006) (Category-specific) semantic deficit in Alzheimer's patients: the role of semantic distance. *Neuropsychologia* 44(1):52–61