INFERENTIAL NETS UNDERLYING WORDS

Mauricio Iza Mikeleiz[1], Alexandra Konstenius[2] & J. Ezquerro[3]
[1] The University of Malaga
[2] University of Helsinki
[3]University of Basque Country

ABSTRACT

This work deals with how word meaning is represented by the speakers of a language. We examine the psychological perspectives on meaning representation. Inferences about emotions are non-conscious responses whose main function is to direct attention and perception. The connection between non-conscious emotional responses and motivated attention was discovered by examining lexical decision tasks. Highly complex non-conscious interferences between emotion, cognition and lexical-semantic knowledge must therefore be assumed. It is quite likely that affective processing interferes with subsequent lexical-semantic analysis along the ventral stream. In this paper we will analyze empirical evidence concerning proto-emotions and the challenges for future research, on the one hand, in order to develop realistic models for the interfaces between emotion, cognition and lexical knowledge, and to outline a psycholinguistic methodology adequate to identify emotionally arousing semantic material, on the other.

Keywords: *Emotional inference, lexical semantics, neurolinguistics, pragmatic features*.

1. INTRODUCTION

The question about how word meaning is represented has interested psychologists, philosophers and linguists or a long time (see Jackendoff, 2010). Furthermore, it has presented a crucial problem for the simulation of cognitive processes in artificial networks. More recently, it has been incorporated within the schedule of Cognitive Neuroscience.

The way in which neurons store information suggest that information is stored by modifying the weights of the connections. Still, a connection or a neuron for each particular datum is unlikely; on the contrary, each piece of information is stored in a highly distributed way between thousands of neurons that, at the same time, store a big amount of other data of the same type. Information is retrieved in the form of an activation pattern of a set of neurons. If the activation pattern emerges only partially, information will be incomplete. In connectionist distributed systems, the storage of information is not all-or-none (that is, the rule exists or not),it follows a continuum (McClelland &Rumelhart, 1986).

One of the most important features of connectionist systems is that memory stores information codified in sub symbolic representations. This means that there doesn’t hold a correspondence one-to-one between concepts and their representation. In a semantic network we can "see" where the concept of bird is; in a distributed connectionist model, such concept does not have a symbolic representation. In this concern, connectionist researches have proposed explanations for cognitive notions, such as <<schema>> or <<semantic network>>, couched in connectionist terms (for example, Schneider, 1987). This way, we could assume that brain represents spatial proximity of semantically related concepts in several ways. For example, groups of nerve cells that store information about related concepts can be spatially located near each other;the number of connections between them can be larger, or the very connections be stronger.

However, in several fields (mainly, in language processing and in perception of emotions), it seems necessary to develop dynamic cognitive models capable of self modification, and of modifying the most elemental mechanisms that make them possible, computationally and biologically plausible (see e.g. Thomas &LaBar, 2005; Vermeulen, Luminet & Corneille, 2006). This suggests that we shouldn’t forget that our activity results from the activation and inhibition of a high number of neurons working in parallel. This procedure constrains the type of algorithms and heuristics (and, therefore, of functions) that can be implemented in a "brain", the time that these functions take, and the type and number of operations to use (Feldman, 1985). The more recent proposals tobring explanations for cognitive phenomena attend these constraints in models in which learning plays a fundamental role.

In the present work we will examine: (i) several key questions related to the first connectionist semantic networks; (ii) some theories of semantics; (iii) a view of emotional inferences; (iv)a discussion on brain processing and the semantic system for emotional inferences.

2. PREVIOUS CONNECTIONIST SEMANTIC NETWORKS

Quillian's (1969) semantic networks have played an important role in the research on knowledge representation. Semantic networks think about knowledge in terms of concepts, their properties, and a hierarchical sub/superclass relation between concepts. Each concept is represented by a node, and the hierarchical relation between concepts is represented by connecting nodes of the appropriate concepts by means of IS-A or INSTANCE-OF links. Nodes at the lower level in the IS-A hierarchy denote tokens, while nodes at higher levels denote classes or categories of types. Properties are also represented by nodes, and the fact that a property applies to a concept is represented by connecting the concept and property nodes by means of a labeled link. Typically, a property is tied to the higher concept in the conceptual hierarchy to which such property is applied. Furthermore, if a property is tied to a node A, it is assumed that this property applies to all nodes that are descendent of A.

In this way, a connectionist system that tries to implement the sort of conceptual structures used by people, must be able to represent two different classes of hierarchies:

(i) "IS-A" hierarchy relating types with their tokens. The more relevant features are that the known properties of types must be "inherited" by tokens, and properties that are found to apply to all instances of a type must be normally attributed to the type. This hierarchy can be implemented including, as a subpart, the distributed representation of the type in the distributed representation of an instance. This representation automatically produces the most important characteristic of IS-A hierarchy, but can only be used for this type of hierarchy.

(ii) Part/whole hierarchy relates an item to the constituent items that compose it. If we use this relation between patterns of activity for representing the type/instance relation between items, it seems that we cannot use it to represent the part/whole relation between items. We cannot make the representation of the whole to be the sum its parts. The use of patterns that represent identity/role combinations, allows representing the part/whole hierarchy in the same way as the type/instance hierarchy. We can see the whole simply as a particular instance of a number of more general types, where each one can be defined as the type that has a particular class of parts playing a particular role.

In symbolic models (like a frame type), it is easy to build IS-A links from instances to types, and from types to super types, and so forth, in order to form hierarchies with inheritance of properties. Therefore, information or properties can be inherited from higher levels. Sub symbolic models (inspired by neuronal mechanisms and based on parallel processing of distributed representations) perform inheritance just through the extension where a representation shares features with another one. This representation automatically produces the largest part of the important characteristic of IS-A hierarchy, but can only be used for one type of hierarchy. Likewise, as there is not a place where knowledge about a type in general is stored in these sub symbolic systems, it is not possible to add facts about this type in general, and to make this new knowledge automatically and immediately available to all its instances, as it can be done in symbolic models (Dyer, 1988).

However, connectionist models have tried to represent semantic networks by using representations that are not purely distributed in order to solve these limitations: local representation or distributed with role-specific units.

Localist networks are prone to form hierarchical layers of units. This reflects the efficiency of hierarchical computations and the easiness with which IS-A and PART-OF relations can be connected by using a one-unit/one-concept approach and excitatory links. The construction of localist networks has a strong empirical component. This is due mainly to the difficulty for deriving formally connectionist structures for a given task (even with some exceptions, for example, Shastri & Feldman, 1986). However, there are heuristics and powerful techniques to build localist networks (see Shapiro, 1987).

A proposal where the advantages of using local representations can be appreciated, is Shastri & Feldman's (1986) model. Shastri & Feldman introduce a set of (connectionist) mechanisms for representation and inference of conceptual information. They suggest that these mechanisms form an adequate basis for the study of problems in language comprehension. They consider only systems without interpreter (a control program) and try to show how such systems can support all existing applications of semantic networks. The only computational primitives in their models are calculus and transmission of states of activity. They try to show that semantic networks have a natural performance in neural networks. The model performs an evidential or probabilistic reasoning and can deal with exceptions and multiple inheritance situations coming from attribute values by default (see Shastri & Ajjanagadde, 1990).

On the basis of an evidential approach and parallelism, they use a local representation (similar to some new semantic network models that incorporate methods in frame structure). But, this approach presents some problems: the use of multiple bindings role-concept in a routine gives rise to the cross-talk problem. Moreover, the authors assume that simple forms of learning result from concept formation representing coherent collections of existing properties and values. More complex forms of learning would lead to concept generalization and to the formation of complex properties during the development of more complex concepts.

Learning can happen due to the rich pre-existing structure in the connectionist net. Although they do not really solve the learning problem, they believe that learning could occur in the Memory Network, on the basis of the notions of recruitment and chunking (Feldman, 1982; Wickelgren, 1979) for representing new instances, and for developing

concepts that would be generalizations of existing concepts (for example, chunking binder nodes to form concepts; see, Shastri& Feldman, 1986, pp.196-199).
A later version of this model (CSN, Shastri, 1988a) also describes how knowledge about concepts, their properties, and the hierarchical relation between them can be coded as a massively parallel interpreter-free network of simple processing elements. Shastri shows also how it can solve an interesting class of inheritance and recognition problems very quickly, in time proportional to the deep of the conceptual hierarchy.
In next section, we will examine the way in which some connectionist models (particularly, CSN) can deal with inference about concepts, that is, inheritance of properties and concept recognition.

2.1. Limited inference
Analyzing human behavior we find that despite operating with a large knowledge base, human agents take a few hundred milliseconds to perform a broad range of cognitive tasks. Among others, humans deploy abilities such as object recognition, spoken and written language understanding, and inference performance such as: "Tweety is a bird, so it flies". Data about human behavior show that representation of conceptual information, and the cognitive processes that access to it, are of such a nature that not only are relevant facts automatically retrieved, but certain kinds of inferences also exhibit extreme efficiency.
although very interesting, class of inference that needs to be performed very fast. This strategy goes on by developing appropriate techniques of knowledge structuring, algorithms and computational systems, in order to perform these inferences within a reasonable time limit. The critical step in this approach consists in circumscribing this class of inference. There are several ways of doing it, and in fact, several alternatives have been pursued (Ballard, 1986).
It is assumed that, in addition to mere facts about the world, we identify important connections or inferential dependencies between these facts as well. Thus, if each piece of information is codified as one node, and dependencies between pieces of information are codified as explicit links between appropriate nodes, then inference can be seen as spreading activation on a network. This metaphor has a high interest, because it suggests an extremely efficient way of performing inferences.
In this sense, structured connectionist approach offers an appropriate framework for explaining these symbolic relations, and to deal with the challenge of computational efficiency. Following this line, Shastri (1988a, 1988b) tries to integrate in a connectionist network (CSN) both the limited inference approach and massive parallelism.
For Shastri assumes that a representation that naturally provides the connection requirement between the syntactic structure and the inferential structure of knowledge domain is a graph where the nodes correspond to information units (constants, predicates, concepts, properties, features, frames, or whatever), and whose links correspond to inferential dependencies between these units. This assumption has one interesting consequence, in the sense that inference is reduced to search on a physically instantiated graph. This does not solve the efficiency problem because the search of arbitrary graphs is an expensive operation. However, once we identify inference as search in a graph, it is possible to relate the efficiency of the process of inference (search) to the structural properties of the representation (graph). For Shastri, searching on a tree or a directed acyclic graph (DAG) is cheaper than searching on a general graph, particularly if the search can be made in parallel. In Shastri's terms, this suggests that if we want computational efficiency, our representation should project or map the knowledge domain in a graph with the following property: "Parts of the graph that are relevant for the solution of a reflexive inference problem must be trees or DAGs".
Thus, the direct relation between structural properties of representation and inference efficiency reduces the reflexive inference problem to the selection of the appropriate representational primitives; primitives that communicate the required structural properties to the graph codifying domain knowledge (Shastri, 1990, p.70).
When knowledge is codified in connectionist networks, links, link weights and computational characteristics of nodes, codify knowledge and the way in which the several knowledge constituents interact during computation. Also, parallel codification (for a less time search) requires assigning a processor to each node in the graph.
Nevertheless, although it is true that in a given connectionist system for knowledge representation, the class of inference installed can be performed with extreme efficiency, it also happens that other inferences cannot be performed at all, or at least, they can only be "approximated". A similar shortcoming arises also in traditional implementations of knowledge representation systems. However, in traditional systems, it is easier to extend the interpreter by adding appropriate procedures (for example, a LISP code) to the system. This is not the case in a connectionist implementation: complete computational characteristics of nodes and their interconnections depend crucially on the nature of the inferences to be installed, and to introduce some changes in the basic inferential ability of a system may require a serious reorganization of the system. As it seems, it is essential that the class of limited inference must be selected very carefully, and the nature of the approximations performed by the system has to be made explicit.

In next section, we will focus on a class of inferences that constitutes an interesting component of commonsense knowledge, that is to say, inheritance and recognition in semantic networks.

2.2. Inheritance and recognition

As it is known, inheritance is the property that allowsto infer properties of a concept on the ground of its ancestor's properties in the hierarchy. The recognition problem is complementary to the inheritance problem. As opposed to inheritance, which consists in searching a property-value of a given concept, recognition searches a concept that has some specified property-values. In this sense, it can be argued that these two forms of reasoning lie in the heart of intelligent behavior and behave as resources of more complex and specialized reasoning process; because both are automatically performed.

However not withstanding, in some cases it happens that a concept can belong to two different hierarchies, and then arise problems in inheriting properties; on the other hand, it may turn out also that a property does not apply to all members of a class, that is to say, a concept can have a specific property that must not be inherited by its descendants.

Exceptions and conflictive multiple inheritance in semantic networks have as result non-monotonicity and ambiguity. Neither of them can be handled within the first-order predicate logic (FOPL). Consequently, formalizations of semantic networks based on FOPL(Charniak, 1981), and several representation languages, such as KL-ONE (Brachman & Schmolze, 1985), cannot deal with exceptions or with multiple inheritance situations. At the same time, the translation to FOPL doesn't explain how codified information in a semantic network for solving recognition problems should be used.

The shortcoming of default logic lies in the assumption that all affirmations have the same importance. This assumption is inappropriate in several cases. The necessity of combining relevant information and rating the relative importance of available information becomes more apparent when we consider subject's beliefs.

Likewise, Touretzky's (1990) proposal, based on the Ordering of Inferential Distance principle, provides a precise specification of what implications (items) should be extracted from an inheritance hierarchy in situations with exceptions. His formalism is also an improvement over Etherington & Reiter's proposal, where inferential significance of IS-A links is made explicit. However, Touretzky does not solve the problem of combining information from several sources, and his system presents ambiguities in situations of multiple inheritances.

Therefore, in order to deal with situations involving conflictive information, it becomes mandatory to adopt an epistemologically richer representation, to make possible to represent meaning relative to rules. An option is to consider affirmations such as "birds fly" as evidential affirmations. Within an evidential formulation, to find solutions for inheritance and recognition problems would amount to select the most probable alternative from a set; the computation of probability can be performed according to knowledge codified in conceptual hierarchy.

Trying to reformulate inheritance and recognition problems in terms of evidential reasoning in an earlier CSN model, Shastri (1988b) extended the traditional representation of a semantic network in order to include evidential information codified in terms of relative frequencies, specifying how instances of certain concepts are distributed with respect to certain property-values. Recognition can be conceived of, in this model, as a more general way of pattern matching; and questions put forward to the network, combining recognition and inheritance, perform a generalized form of pattern completion.

However, as we have seen, mere associating property-values with types is not enough. An agent may want to make finer distinctions and to use such information to recognize things, and to predict their properties. One way of capturing these distinctions is to store frequency distributions of concepts with respect to certain property-values. But this approach amounts to an excessive simplification. There are situations in which an agent may have to codify an evidential relation between a concept and an attribute-value without knowing any frequency distributions. Furthermore, there may be situations in which the correct conclusion is not the most probable one, but any other instead(see Lyon &Chater, 1990).

Therefore, the decision of designing some [property, value] pairs (or one pair) that would constitute a concept, would be always available to the agent. But this move implies to renounce to something well defined: on the one hand, it requires a commitment to additional computational sources, such as primitive nodes (e.g., binder nodes associating objects, properties and property-values; and where a concept is activated just when binder nodes receive activation from a pair of nodes simultaneously), links, processing elements, and so on; although, on the other hand, it is easier to extract certain inferences.

In a system such as CSN, problems can also arise when activation has (not) to spread beyond from what has been "pre-programmed". For example, this system is unable to infer, either new facts, or new things that are not already in the answer set. These problems can be resolved using a distributed representation instead, because new facts would emerge as an auto-learning way.

2.3. Connectionist prototypes

In addition to the attempts to simulate semantic networks and inference with local representations, there have been some efforts to do it with distributed representations as well, specifically, by using role-specific units. An example is Hinton (1990).

Hinton (1990) proposes a distributed codification of semantic networks using parallel hardware. The net codifies a set of triplets in the following way: [relation, role-1, role-2]. The proposed system has several interesting properties: given two components of a triplet, the net is able to determine the third element; the net can be programmed using the perceptron convergence rule, and can perform simple inheritance of properties. The system, however, has not reached enough structure and control for dealing with general cases of inheritance and partial matching, particularly when these cases come to pass in a multilayer semantic net that includes multiple hierarchies and conflictive information (see Alishahi& Stevenson, 2007).

In general, inheritance, or inference of properties (or relations), is a problem unsatisfactorily solved inside a connectionist system with purely distributed representation. The reason is that they don't have specific units for each concept, as it happens in local networks (we would need a great deal of units or several network modules), neither any sort of conceptual primitives (micro features) or specific-role units, as it happens in some models of distributed networks (in both cases, it is necessary to suppose, to some extent, the existence of "certain" explicit representations). Thus, the system should need specific types of patterns (such as, "A category has a property of a superior node") in order to infer that the relation IS-A carries out the inheritance of properties of this upper node. Likewise, we can suppose that it could be necessary a "contrary method" for a category to acquire common properties of its members. This happens because patterns would not reflect any "structure" in order to generalize that some properties are a subset of a property of an upper level.

However, Mervis & Crisafi (1982) have argued not withstanding that, in most hierarchies of natural categories, basic level categories are more differentiated, given that they are followed by super ordinate categories, and finally, by subordinate ones. On this assumption, these authors have argued that, in general, the order of acquisition should be as follows: basic, super ordinate, subordinate. Nevertheless, in the hypothetical case of a hierarchy in which subordinate categories were more differentiated than super ordinate ones, we could predict that these particular subordinate categories would be acquired before relevant super ordinate categories. Therefore, they affirm that differentiation degree has dominance in the hierarchical level. So, the differentiation hypothesis can be used also for predicting the easiness of acquisition when it is difficult or impossible to apply one of the three hierarchical labels to a particular categorization scheme. Surely, two processes top-down and bottom-up, coexist in children concept acquisition, according to context, motivation, previous knowledge, and so on, although they can work, also in parallel, in a single system of information processing.

Thus, theThe main features (such as, content-dependency and multiple-constraint satisfaction) of distributed models may be used in an approach aimed to explain how people acquire concepts. These models can assume a picture of the mind as a chaos (for example, a bottle with different liquids that we move and, at the end, we obtain an equilibrium), where an activation pattern gets such equilibrium. The question lies in whether this equilibrium can be a concept or a proposition.

As mentioned above, it is possible to represent these activation patterns in an activation space. For example, the priming paradigm could be explained in this approach by taking into account the distance between concepts within the space, in the sense that two similar activation patterns match in order to get a new equilibrium or a new concept (say, a new stable activation pattern in distributed representations).

Given that, In the case of distributed representations it seems thus more natural to proceed with the previously mentioned feature approach. For instance, units may represent different features (micro features) that, in turn, may be used in the representation of different concepts (with common features, however). And, this codification may also represent different instances of a given concept (i.e., different particular dogs). Thus, a distributed model can acquire, through training, material whose features correspond to a concept (or instance) by means of abstracting a prototype from the class. For example, McClelland &Rumelhart (1986) propose a similar system that encodes specific experiences about individuals (particular instances of a concept). Also, Rumelhart, Smolensky, McClelland & Hinton (1986) used of the same idea for room schemas.

Thus, in contrast with symbolic models, prototypes and schemas are implicit representations that emerge by micro features activation, without needing explicit representations. These distributed models are able to infer inductively stereotypical or default information from inputs coming to the network. Even so, as we will see below, it is difficult to understand how more complex inferential phenomena could be performed by using distributed representations, mainly due to several problems that arise when using this kind of representation.

2.4. Limitations

Although connectionist networks solve several problems that classicist approaches were unable to treat satisfactorily, these systems have their own ones (Fodor &Pylyshyn, 1988; Massaro, 1988; Lachter & Bever, 1988; Pinker & Prince, 1988). These authors criticize the connectionist approaches due mainly to some problems that have not been utterly solved, such as logic inference or the semantic primitives' assumption.

According to Lachter & Bever (1988), some connectionist models of acquired linguistic behavior recently proposed, have representations based on incorporated linguistic rules. Connectionist models of language acquisition have arbitrary mechanisms and architectures that lead them to simulate rule effects. Connectionist models, in general, are not well adjusted to account for the acquisition of structural knowledge. They require predetermined structures, even for simulating basic linguistic facts. Such models are more suitable to describe the formation of complex associations between independently represented structures. This converts connectionist models into potentially important tools for the study of relationship between frequent behaviors and the structures underlying knowledge and representation. Such models can provide, at least, computationally powerful ways that show the limits of associationist descriptions of behavior. For example, Lachter & Bever, worried about linguistic rule acquisition and constraints on rules, claim that systems that learn to assign thematic roles to names in specific sentences (McClelland & Kawamoto, 1986), have properties similar to the ones of models that learn past-times of verbs (Rumelhart & McClelland, 1986). Representational input nodes are triplets, consisting in a syntactic position, and two semantic features for a name or verb. Output nodes represent a semantic feature for a name, another for a verb, and a thematic relation name-verb.

Lachter & Bever affirm that role semantic features do not follow from any independent theory: rather, they are just descriptors of the same roles. Therefore, some learning happens to be trivial. Learning does not mean to isolate independently defined semantic features relevant to roles. It involves, rather, an accumulation of activation strengths from available role features, and to give the correct instances of words (feature matrix) placed on particular role positions. These authors argue that one of the achievements of this model, according to McClelland & Kawamoto, is that it over generalizes thematic role assignments. Nevertheless, these connectionist systems can never discover structural rules. They concede, as a positive aspect of connectionist models, that they provide a rich associative framework for the description of formation of complex habits.

On the other hand, Massaro (1988) showed that the assumption of interactive activation in specific connectionist models is both unnecessary and inconsistent with empirical results. He admitted that connectionist models with hidden units are very powerful: they can simulate different types of results generated by models of different processes. Due to the great power of connectionist models with hidden units, they can describe results with no realistic assumptions about the psychological relations that are functional in the task. Connectionist models with hidden units are limited in theoretical value unless we postulate something like sequential stages of processing in which some categorization takes place before the answer selection. In spite of these limitations, it is well-known that some other important properties of connectionism have to be established in current models of pattern recognition.

Finally, Fodor & Pylyshyn (1988) claim, for example, that connectionist graphics aren't structural descriptions of mental representations, but specifications of causal relations. Anyway, these criticisms have received some answers (for example, Oaksford, Chater & Stenning, 1989; Chater & Oaksford, 1989). These authors argue that standard computational models assume the symbolic paradigm: say, that cognitive processes are mediated by the manipulation of symbolic structures. Such schemes deal easily with formal inferences, and with memory for arbitrary symbolic material. However, they do not capture so easily memory retrieval directed by content or defeasible inference sensitive to context, two abilities that people exhibit in a natural way and effortless, as it happens in commonsense knowledge or in shared knowledge.

3. CURRENT TRENDS OF SEMANTIC REPRESENTATION

There have been attempts to define basic level actions (Lakoff, 1987) and hierarchical representations for events (Jackendoff, 1990). Difference between domains, however, persists. For example, in Keil (1987, 1989), the hierarchical organization for objects and events is described as being different, with event categories being represented in fewer levels (generally two)and with fewer distinctions at the super ordinate level. Other attempts to capture a level of organization for events have included distinctions between "light" (e.g., do) and "heavy" (e.g., construct) verbs (Pinker, 1989). However, the light/heavy dichotomy only allows us to draw a distinction between verbs used as auxiliaries and other verbs; however,to draw the line between "general" and "specific" verbs is not an easy agreed-upon exercise.

Also relevant here are Imaging studies which provide some evidence for distinct neural substrates for processing the meanings of words referring to objects and actions are also relevant in this concern. For words referring to objects, multimodal areas of the basal temporal cortex are involved in semantic processing, while semantic processing of words referring to actions involves left primary motor and premotor areas and left middle temporal areas.

Notable exceptions are studies in neuropsychology that have documented a double dissociation between concrete and abstract words, offering an account for the greater degree of impairment of abstract rather than concrete words in terms of differences in feature richness of semantic representations, with concrete words having richer representations and therefore being more resistant to damage than abstract words, whose representations would instead be characterized by fewer features.

It is well established that patients may be able to retrieve semantic information about objects in absence of naming success (Plaut, 2002) and that the amount and quality of this information may differ according to the category from which an object is drawn and the nature of the task (Moss et al., 1998). For example, there are patients who appear to have a greater problem with living things. Their knowledge of the shared properties of living things was no different from their knowledge of the shared properties of artifacts (see Garrard et al., 2001).

3.1. Relations among words

As we have seen, it does matter for network-based theories the type, configuration, and relative contribution of the links existing between words. Lots of alternative frameworks have been developed differing along these crucial dimensions. Importantly, these models have in common a focus upon (explicit) intensional relations, and a necessity to explicitly designate those relations that are implemented.

In fact, researchers have decided either to embed different or the same organizational principles. In Wordnet(Miller & Fellbaum, 1991), has been developed a holistic model of semantic representation byusing the strategy of deciding a priori diagnostic properties of the different domains, with different types of relational links for objects and events. In the case of nouns referring to objects it has been argued that relations such as synonymy, hyponymy (i.e., dog is a hyponym of animal) and meronymy (e.g., mouth is part of face)play animportant role in describing the semantic organization. For verbs, instead, some propose that the relational links among verb concepts include troponymy (i.e., hierarchical relation in which the term at an inferior level, e.g., crawling, is a manner of a term at a level above, e.g., travel/go/movelocomote), entailment (e.g., snoring entail sleeping) and antonymy (e.g., coming is the opposite of going) while relations such as meronymy would not apply within the network.

Wornet (Miller &Fellbaum, 1991) is a network model for the representation of a large number of nouns, verbs and adjectives in English. In this system, nouns, adjectives and verbs each have their own organization, which is determined by the role they must play in the construction of linguistic messages. These relations and organization are hand constructed on the basis of the relations considered as relevant within a given class of words: (i) for nouns, the main role is typically played by relations including synonymy, hierarchical relations and part-whole relations; (ii) for verbs, instead, are dominant troponymy (hierarchical relations related to specificity in manner), entailment, causation, and antonymy.

In alternatively, in the Featural and Unitary Semantic Space Hypothesis (FUSS, Vigliocco et al., 2004) the strategy has been not to decide a priori upon criteria to distinguish the object and the event domains, but to model both types of word within the same lexical-semantic space using the same principles. This strategy has also been used by global co-occurrence memory models such as Latent Semantic Analysis (LSA, Landauer & Dumais, 1997) and Hyperspace Analogue to Language (HAL, Burguess& Lund, 1997). These models take advantage of computational techniques, using large corpora of texts (e.g., linguistic context).

Another computational model that uses cross-situational inference too is proposed by Yu (2005). This model is also used to examine the role of various factors, such as syntax (Yu, 2006), in word learning. However, the system uses the original form of the automatic translation learning algorithm of Brown et al. (1993), which lacks of cognitive plausibility. It is non-incremental and learns through an intensive batch processing of a whole training data. Moreover, it is tested on limited experimental data containing a very small vocabulary, and with no referential uncertainty (Fazly, Alishahi& Stevenson, 2008).

MostThe great majority of the existing models rely on a pairing of a semantic representation with a single word form (or its phonological representation) –as opposed to full utterances- as training data. Connectionist models have been proposed for learning such associations, and investigating various patterns in the process of learning. For example, Li et al. (2004) simulate vocabulary spurt and age of acquisition effects, whereas Horst et al. (2006) examine the role of fast mapping. Regier (2005) proposes an associative exemplar-based model that accounts for the changes observed in children's word learning pattern, such as fast mapping and learning synonymy, with no changes in the underlying learning mechanism. On the other hand, the Bayesian model of Xu&Tenenbaum (2007) focuses on how humans generalize and learn category meanings from examples of word usages.

3.2. Damaging networks

One class of models employ connectionist feature frameworks in order to demonstrate how particular patterns of semantic impairment can be observed as a consequence of differential feature composition. This approach typically entails training a connectionist network with input that, although not directly obtained from speakers, is informed by particular characteristics of feature norms that are assumed to play a role.

Farah & McClelland (1991) proposed a model in which words referring to living or no-living entities were associated with different proportions of visual-perceptual *vs*. functional features (the former predominate for living things, the latter predominant for non-living entities), consistent with evidence from feature-generation tasks. Differential category-related effects were found when the model was injured (damaged in order to simulate impaired performance), depending upon whether the lesion targeted the visual-perceptual or functional features.

Devlin et al (1998) addressed the differential impairment over time for living and non-living things as a consequence of Alzheimer's dementia. They implemented semantic representations that were based upon the relationship between entities in a particular domain: inter correlated features (those features which frequently co-occur with each other) and distinguishing features (those which best enable similar entities to be distinguished from each other). While living things typically have many inter correlated features but few distinguishing ones, the situation for non-living entities is the reverse. These differences in the composition of living and non-living entities were sufficient to explain the progression of relative impairment in distinguishing living and non-living things as a consequence of Alzheimer's dementia, even with a single representational system.

An alternative approach has been to assemble sets of words and decide a priori upon their features. Hinton & Shallice (1991) created a set of semantic features which capture intuitive properties of common objects (has legs, hard, made-of-metal), then used those features to train an attractor network to learn the mapping between orthography and semantics. Injuring this network brought about semantic, visual, and combined visual/semantic errors consistent with patterns of performance in deep dyslexia.

Plaut (1995; Plaut & Shallice, 1993) use the same approach to investigate double dissociations between reading concrete and abstract words, using a set of empirical semantic features underlying the meanings of concrete and abstract words. One particular aspect of these representations was that abstract words have fewer features overall; this broad difference in feature properties between concrete and abstract domains (perhaps in conjunction with other differences) converted into differential consequences when different aspects of the model were damaged: abstract words were more impaired when the feed forward connections were injured, while concrete words were more impaired when the recurrent connections were impaired.

3.3. Feature norms

In previous models, the semantic features were selected a priori, and may be that this procedure does not reflect the full range of properties of meaning that are relevant to the representations of the words in question. It is therefore an important additional step to asses comparable feature information that is produced by naive speakers (about the research questions involved). This allows the investigation of issues like feature properties, distribution of features across different sensory modalities, and so forth, by using entirely empirically derived measures. In collecting features of meaning from multiple naive speakers, we also gain a fine-grained measure of features alience, given that a feature's relative contribution to a word's meaning can be weighted according to the number of speakers producing that word. That is, relevance of feature variables, established on the basis of feature norms, to the representation of concepts (Vinson & Vigliocco, 2008).

Vigliocco et al. (2006) have used these norms for the selection of materials for imaging studies, in order to generate predictions concerning semantic impairments in brain-damaged individuals, and to develop a model of semantic representation.

According to FUSS, in order to address the semantic representation of words that refer to objects and words that refer to events, conceptual features (of which feature norms are a proxy) are bound into a separate level of lexical-semantic representations which in turn serves to mediate between concepts and other linguistic information (syntax and word form) in line with Damasio et al's (2001) idea of "convergence zones".

Using Kohonen's self-organizing maps for dimensionality reduction technique, the organization at this level arises trough an unsupervised learning process which is sensitive to properties of the feature input, such as the number of features for each concept, how salient a given feature is for a concept (feature weight), features that are shared among different concepts and features that are correlated. Thus it is not necessary to specify in advance which aspects of the input should be reflected in lexical-semantic organization; this process also allows different properties to exert different influences depending upon the characteristics of a given semantic field, giving rise, for example, to the different smoothness of the space for objects (organized in a "lumpy" manner) with semantic field boundaries being well-defined, and events (organized "smoothly") in which there are no clear boundaries among fields.

On the contrary, Barsalou (2005) suggest that an important distinction between abstract and concrete words is which situations are more salient for the two types of word. Whereas for objects, attention focuses on the specific object against a background, for abstract notions, attention focuses on social context, events and introspective properties.

For 'true', the focus will include the speaker's claim, the listener's representation of the claim, and the listener's assessment of the claim, rendering abstract words more complex than concrete words.
Abstract knowledge is viewed as originating in conceptual metaphors (i.e., the use of a concrete conceptual domain of knowledge to describe an abstract conceptual domain). In this view, learning and representation of abstract concepts in the mind/brain is grounded in the learning and representation of concrete knowledge, which in turn is grounded in our bodily experience of the world.
As abstract words also appear to be more susceptible to cross-linguistic variability (and cross-cultural variability too) than concrete words, investigations of this domain of knowledge may provide important information on how conceptual universal biases may interact with language-specific factors in determining the organization of the semantic system. Next, we will discuss it in relation to the activation of emotional inferences while reading words.

4. EMOTIONAL INFERENCES

The interaction between cognition and emotion is currently increasing neuroanatomical support. Lately, the multiple connections between limbic system areas and the neocortex have been emphasized: (i) the proliferation of fibers that led directly from the thalamus to the amygdale, which presumably represent a flow of information from (still only partially analyzed) stimulation obtained very early in cognitive processing to an area strongly implicated in emotional activity; (ii) fibers leading from the amygdale to the neocortex, maybe indicating emotional responses evoked by this preliminary information and able of modulating sensory processes; and (iii) fibers leading from the neocortex to the amygdale, which presumably represent cortical feedback from more completely analyzed information to influence emotional response.
There are not one single connection between emotion and cognition, but several. Some connections must clearly be classified as non-conscious. For example, when people become committed to pursuing a goal, that event initiates an internal state termed as 'current concern', which has the property, among others, ofpotentiating emotional reactivity to cues associated to the goal pursued. The emotional responses thus emitted begin within approximately300 milliseconds after exposure to the cue – early enough to be considered purely central, non-conscious responses at this stage. Because they appear to be incipient emotional responses but lack many of the properties normally associated with emotion, they are called 'proto-emotional' responses (see also Mathews & MacLeod 1994).
Proto-emotions are processed in parallel with early perceptual and cognitive processing, with which they trade reciprocal influences. The intensity (and other features) of the proto-emotional responses affects the probability that the stimulus will continue to be processed cognitively. The results of continued cognitive processing in turn modulate the intensity and character of the emerging emotional response. Proto-emotional responses affect attention, perception, performance at different phases and all skill acquisition (Kanfer, 1996; Gollwitzer, 1996).

4.1. Lexical decision tasks

The automatic character of proto-emotional responses process is exhibited by data from lexical decision tasks (Young, 1987). This experimental paradigm ask subjects to decide as quickly as possible whether each occurrence of a letter string is an English word *vs*. nonsense, and the reaction time of the response is measured. A series of experiments (Young, 1987) showed that proto-emotions direct attention. The left side of the screen was taken up by a patch containing computer-related verbal 'garbage', and subjects were instructed to ignore it, but sometimes contained a word planted to relate to one of a subject's current concerns. When the target string was indeed a word, the reaction time of reporting this was significantly slower if the distractor patch contained a concern-related word. Thus, concern-related stimuli seem to impose an extra load on cognitive processing even when they are peripheral and subjects are consciously ignoring them. This finding adds a further look to the automaticity of the effect (see also Kauschke & Stenneken 2008).
Modified Stroop procedures demonstrate the same effect (Riemann & McNally, 1995). For instance, MacKay et al. (2004) demonstrated three taboo Stroop effects that occur when people name the color of taboo words (e.g. death, war): (i) longer color-naming times for taboo than for neutral words, an effect that diminishes with word repetition; (ii) superior recall of taboo words in surprise memory tests following color naming; (iii) better recognition memory for colors consistently associated with taboo words rather than with neutral words. They argue that taboo words trigger specific emotional reactions that facilitate the binding of taboo word meaning to salient contextual aspects.
Besides words related to current concerns and taboo words, also negative words exhibit to some extent Stroop effects. Estes & Adelman (2008) analyzed a set of words, controlling central lexical features, and found a small but significant effect for word negativity. They concluded that this effect is categorical. Larsen et al. (2008) analyzed the same data set but included the arousal value of each word. They found nonlinear interaction effects predicting lexical decision time and naming speed.

Not all negative words produce the generic slowdown. Only negative words that are moderate to low on arousal produce more lexical decision time slowing than negative words higher on arousal. Kahan & Hely (2008) showed that the role of valence and word frequency interact in contributing to the emotional Stroop effect.

Altarriba & Canary (2004) also examined the activation of arousal components for emotion-laden words in English (for instance, kiss, death) in two groups of monolingual (English) and bilingual (Spanish-English) subjects. Prime-target word pairs were presented for lexical decisions to English word targets in either high arousal, moderate or unrelated conditions. The results revealed positive priming effects in both arousal conditions for both groups of subjects. But, while the baseline conditions were similar across groups, the arousal conditions produced longer latencies for bilinguals than for monolinguals.

Brain activation to emotional words also varies in depressed vs. healthy subjects. Depression involves enhanced processing of negative stimuli or diminished processing of positive stimuli. Canli et al. (2004) used functional magnetic resonance imaging to assess brain activation in depressed *vs*. healthy subjects. Subjects with serious depressive disorder and a control group were scanned during a lexical decision task involving neutral, happy, sad, and threat-related words. For happy words, depressed subjects exhibited less activation than did controls to happy words in fronto-temporal and limbic regions. For sad words, depressed subjects showed more activation than did controls in the inferior parietal lobule and less activation in the superior temporal gyrus and cerebellum, suggesting a complex activation pattern that varies for neural sub-circuits that may be associated with different cognitive and behavioral processes.

Similarly, Hirsch & Mathews (1997) performed three experiments to investigate the extent to which subjects with high or low levels of anxiety about interviews made emotionally congruent interpretative inferences while reading descriptions of a relevant ambiguously-threatening event like being interviewed for a job. The resulting data support the hypothesis that groups varying in self-reported concern about the described event differed in the interpretations that they made while reading: non-anxious subjects infer positive outcomes to an ambiguous event, while highly anxious subjects do not.

Proto-emotions are not exclusive to lexical-semantic knowledge. Acoustic properties of speech likely provide external cues about internal emotional processes, a phenomenon called vocal expression of emotion (Bachorowski & Owren, 1995). Vocal expressions of emotions have also the power to evoke proto-emotional responses.

4.2. The neurolinguistic perspective to proto-emotions

The perceptual benefit for emotionally arousing material, labeled motivated attention, is indexed by electro cortical amplification at various levels of stimulus analysis. On a neuronal level, the question is how perceptual enhancement for arousing signals translates into modified processing of information. Ihssen, Heim &Keil (2007) examined facilitation and interference effects of task-irrelevant emotional pictures on subsequent word identification in the context of forced-choice lexical decision tasks. The pictures varied in hedonic valence and emotional arousal that preceded the word/ pseudo-word targets. Across measures and experiments, high-arousing compared to low-arousing pictures were associated with impaired processing of word targets. Arousing pleasant and unpleasant pictures prolonged word reaction times irrespective of stimulus-onset asynchrony (80 msec, 200 msec, 440 msec) and salient semantic category differences (e.g., erotica *vs*. mutilation pictures). On a neuronal level, interference was reflected in reduced N1 responses (204–264 msec) to both target types. Paralleling behavioral effects, suppression of the late positivity (404–704 msec) was more notable for word compared to pseudo word targets. Regional source modeling indicated that early reduction effects originated from inhibited cortical activity in posterior areas of the left inferior temporal cortex associated with orthographic processing. Modeling of later reduction effects argues for interference in distributed semantic networks comprising left anterior temporal and parietal sources. Thus, affective processing interferes with subsequent lexical-semantic analysis along the ventral stream.

Tamagni et al. (2009) found that healthy right handed subjects exhibiting a leftward line bisection bias on a lateralized lexical decision task had a recognition advantage for negative over positive emotional words. They suggest that functional hemispheric differences state variables may be less decisive than the trait variable of lateral hemispatial attention, and propose a reconsideration of 'hemisphericity'. Their findings also have complex implications for the interaction between cortical (anterior and posterior) and subcortical structures in the mediation of both the production of emotions and perception.

In this way, a number of indications from current-concerns related data (see Klinger 1996) have suggested that a critical property of current concerns is to dispose subjects to respond emotionally to cues associated with corresponding goal pursuits. The emotional response induces a number of levels of cognitive processing. That is to say, the emotional responsivity, which is itself based on goal commitments, mediates the effects on cognition.

4.3. Future work

Future research on cognition has to develop realistic models where non-conscious emotion and lexical-semantic analysis interact with cognitive processing. That the classical symbolic paradigm can readily meet this demand is questionable. Semantic knowledge is thought to be mainly conscious and related to "ideas". It is doubtful that this conceptualization would be realistic concerning non-conscious proto-emotional responses.

The neural network –paradigm assumes a non-conscious level of reaction that may activate any response-algorithm. These algorithms are not limited in number or in quality. Hypothetically any feature of a semantic component can be identified by an algorithm at a pre-conscious level.

The problem, As it seems, the problem would be how to identify arousing contentin lexemes. Lexical decision tasks are able to recognize that something is motivating attention by measuring time response, however, it is unable to analyze semantic knowledge in detail. The experiments cannot tell what component in negative words lead to affective arousal – nor explain why.

One methodological innovation is, for example, the work of Pitterman et al. (2010). They combined speech-based emotion recognition with adaptive human-computer modeling. Having the robust recognition of emotions from speech signals as their goal, the authors analyze the effectiveness of using a plain emotion recognizer, a speech-emotion recognizer combining speech and emotion recognition, and multiple speech-emotion recognizers at the same time. The semi-stochastic dialogue model employed relates user emotion management to the corresponding dialogue interaction history and allows the device to adapt itself to the context, including altering the stylistic realization of its speech. Yet, the semantic content of lexemes remains unanalyzed.

In this case a mixed-methods approach to semantics seems to be worth to pursue. Vanhatalo (2002a, 2002b, 2004, 2005) succeeded in demonstrating that a cluster of speech act –synonyms (engl. nag) are distinguishable in regards to the concept of 'fairness'. Anaphora is asemantic component not mentioned in the literature or in dictionaries, but discovered through open-ended questionnaires, and then quantified trough population tests. Fairness is atrait non-consciously analyzed in social settings. Proto-emotional responses to perceived unfairness regulate serotonin reactions (Crockett & al. 2008). Thus the mixed-method approach seems to be able to detect and discover relevant semantic components for further studies on the connections between emotion, lexical-semantic knowledge and cognition.

Pragmatic features may also be quantifiable through a mixed-methods approach. Iza & Konstenius (2010) were able to show though a cross-linguistic (Spain-Finland) experiment, that all negative person-related adjectives would be interpreted 2-degrees (on a scale of 10) more negative if uttered by an older person (vs. a peer), and all positive person-related adjectives would be interpreted 2,5 degrees more positive if uttered by a peer. This revels that pragmatic features should be identified as interfering variables when discussing proto-emotional responses interferences with lexical-semantic content.

5. DISCUSSION

Some authors argue for the existence of multiple modality-specific semantic systems, each responsible for processing stimuli in a specific modality (visual, verbal, auditory-non-verbal, tactile, etc.). The opposing opinion is that there is a single multimodal semantic system, responsible for all semantic processing. Let see Shallice's (1988) three arguments in favor of multiple modality specific semantic systems:

(i) The occurrence of cases of multiple modality specific aphasia points to multiple semantic systems. In modality-specific aphasia, stimuli in all modalities but one (tactile), can be named by the patient. In such a case, the patient can name a picture, or an auditory-described object, but cannot name an object he is exploring tactilely. A possible explanation is that there exist multiple modality-specific semantic representation systems, and that there are impairments in the tactile system, in this case, or in the transmission from the tactile system to the verbal system.

(ii) A patient, AR, presented with semantic access dyslexia. In a word-reading task, AR was aided more by a verbal cue than by a picture. Object naming was also difficult for AR. Shallice offers the explanation that AR could not transmit information between his visual semantic system and his verbal system. Thus, his reading was not aided by a picture, and his picture-naming was not aided by a verbal cue.

(iii) The study of aphasic patients who have lost information from their semantic systems. Such patients often show differences in performance between modalities. For example, a patient may be able to describe a visually-presented object, but not the auditory-presented word. In such a case, the patient's visual semantic system may be intact, while information has been lost from his auditory-verbal semantic system.

On the other hand, there exists evidence in support of a single multi-modality semantic system. Such a system would be responsible for all semantic processing, regardless of input or output modality. Hillis et al (1995) mention a patient, KE, who made semantically related errors in various tasks, (oral reading, written naming, word-picture matching and naming from tactile input). Analysis of KE's performance revealed similar frequency and type of error across all modalities of input and output. This qualitatively and quantitatively similar level of performance across modalities suggests the possibility of damage to a single semantic processing system, responsible for all lexical-semantic processing. Such damage would result in impairment of processing in all tasks involving the semantic system.

Does processing word referring to *events* involve the activation of modality-related information even when we justlisten? In the case of PET listening attentively to blocks of words (motion *skate*, sensory *taste*):

(i) Premotor/motor areas(BA 4/6) activations for motor words;

(ii) Multimodal temporal basal areas (BA 20/36) for sensory words.

Analyzing the regions of interest, listening (the most automatic task) to motor words activates primary motor cortex. This suggests that we retrieve non-linguistic information specific to modality and there is not effect for sensory words in basal temporal areas.

Finally, knowledge about word is organized according to grammatical class (nouns & verbs): (i) there have been described a phasic patients who are selectively impaired neither for nouns, nor for verbs and vice versa; (ii) Areas of specific activation for verbs have been reported.

However, some studies confounded the semantic distinction between objects and events and the grammatical distinction between nouns and verbs. In an automatic task, listening to words, a common neural system underlies the processing of nouns and verbs, once semantics is controlled.

So therefore, the issue of a single multi-modality semantic system *vs*. multiple modality-specific systems may only be resolved through further research. If conclusive evidence is obtained to support one hypothesis or well the other, language processing models may have to be adapted to represent correctly the structure and function of the semantic component.

There have been Very few computationally detailed semantic-based models of emotional inference have been proposed yet. In particular, very little work has been done in this regard within the framework of cognitive architectures (Newell, 1990; Sun, 2002). From this work, we can envisage a semantic computational model with some attributes: (i) mechanisms and process-based aspects of different psychological processes with words; (ii) a cognitive architecture for integrating the semantic representation into a broader context and with other psychological processes; (iii) it should be grounded on essential motivations and emotions of human behavior and action (Sun, 2009).

An example of such an architecture is CLARION. It has been successful in simulating artificial grammar learning tasks, dynamic control tasks, categorical inference tasks and certain mental disorders (Sun et al, 2011). It also works on reasoning tasks, social simulation tasks, as well as meta-cognitive, and motivational tasks (Helie& Sun, 2010; Sun & Zhang, 2006).

Within this architecture, there is the constant interaction among four subsystems within a cognitive agent: (i) motivational (there is a set of basic motives or drives which are universal across individuals); (ii) meta-cognitive; (iii) action-centered;(iv) non-action centered. Individual differences may be explained by the differences inthedrive activations in different situations by different subjects (it is explained by differential drive strengths). The drives lead to the setting of goals and a number of major cognitive parameters by the meta-cognitive subsystem. Individual differences on drive strengths are reflected in the resulting goals and major cognitive parameters in consequence. On the basis of this selection, a cognitive agent makes its own action decisions, within the action-centered subsystem. Thus, these actions reflect individual differences as well as contextual factors.

In this concern, we could hypothesize that mental processes may be the result of complex interactions among a large set of psychological entities, elements and components, such as emotional inference, semantic representation and so forth. Therefore, a detailed computational simulation based on this cognitive architecture can provide a coherent account of the underlying complex dynamics and an integration of existing empirical or clinical approaches.

6. ACKNOWLEDGEMENTS

This work has been supported by the project (FFI2009-08999) of Spanish MICINN.

7. REFERENCES

[1]. Alishahi, A. & Stevenson, S. (2007). A computational usage-based model for learning general properties of semantic roles. In Proceedings of the 2nd European Cognitive Science Conference, Delphi, Greece.

[2]. Altarriba, J. & Canary, T.M. (2004). The influence of emotional arousal on affective priming in monolingual and bilingual speakers. Journal of Multilingual and Multicultural Development 25 (2): 248-265.

[3]. Bachorowski, J.A. & Owren, M.J. (1995). Vocal expression of emotion: acoustic properties of speech are associated with emotional intensity and context. Psychological Science 6 (4): 219-233.

[4]. Ballard, D. (1986) Parallel logical inference and energy minimization, Technical Report 142, Department of Computer Science, University of Rochester, Rochester, NY.

[5]. Barsalou, L.W. (2005). Continuity of the conceptual system across species. Trends in Cognitive Sciences, 9, 309-311.

[6]. Brachman, R.J. & Schmolze, J.G. (1985) An overview of the KL-ONE knowledge representation system, Cognitive Science 9, 171-216.
[7]. Brown, PF, Della Pietra, SA, Della Pietra, VJ & Mercer, RL (1993). The mathematics of statistical machine translation: parameters estimation. Computational Linguistics 19 (2), 263-311.
[8]. Burgess, C., & Lund, K. (1997). Representing abstract words and emotional connotation in high-dimensional memory space. Proceedings of the Cognitive Science Society (p. 61 - 66). Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
[9]. Canli, T., Sivers, H., Thomason, M.E., Whitfield-Gabrieli, S., John, D.E. & Gotlib, I.H. (2004). Brain activation to emotional words in depressed vs healthy subjects. Neuro Report 15 (17): 2585-2588.
[10]. Charniak, E. (1981) The case-slot identity theory, Cognitive Science 5, 285-292.
[11]. Chater, N. & Oaksford, M. (1989) Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn, Research Paper EUCCS/RP-27.
[12]. Crockett M.J., Clark, L. Tabibnia, G., Lieberman, M.D. & Robbins, T.W. (2008) Serotonin Modulates Behavioral Reactions to Unfairness. Science 320 (5884): 1739.
[13]. Damasio, H., Grabowski, T.J., Tranel, D., Ponto, L.L.B., Hichwa, R.D., & Damasio, A.R. (2001). Neural correlates of naming actions and naming spatial relations. NeuroImage, 13, 1053-1064.
[14]. Devlin, J., Gonnerman, L., Andersen, E. & Seidenberg, M.S. (1998). Category specific semantic deficits in focal and widespread brain damage: a computational account. Journal of Cognitive Neuroscience 10: 77-94.
[15]. Dyer, M.G. (1988) Symbolic NeuroEngineering for natural language processing: A multilevel research approach, Technical Report UCLA-AI-88-14, Artificial Intelligence Laboratory, Computer Science Department, UCLA.
[16]. Estes, Z. & Adelman, J.S. (2008). Automatic vigilance for negative words in lexical decision and naming. Emotion 8 (4): 441-444.
[17]. Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality- specificity and emergent category-specificity. Journal of Experimental Psychology: General, 120, 339-357.
[18]. Fazly, A., Alishahi, A. & Stevenson, S. (2008). A probabilistic incremental model of word learning in the presence of referential uncertainty. Proceedings of the 30th Annual Conference of the Cognitive Science Society, Washington, D.C.
[19]. Feldman, J.A. (1982) Dynamic connections in neural networks, Bio-Cybernetics 46, 27-39.
[20]. Feldman, J.A. (Ed.) (1985) Special issue in connectionist models and their applications, Cognitive Science 9(1), 1-169.
[21]. Feldman, J.A. & Ballard, D.H. (1982) Connectionist models and their properties, Cognitive Science 6, No. 3, 205-254.
[22]. Fodor, J.A. & Pylyshyn, Z.W. (1988) Connectionism and Cognitive architecture: A critical analysis, Cognition 28, 3-71.
[23]. Garrard, P., Lambon, MA, Hodges, JR & Patterson, K. (2001). Prototypicality, distinctiveness and intercorrelation: analyses of the semantic attributes of living and nonliving concepts. Cognitive Neuropsychology 18: 125-174.
[24]. Gollwitzer, P.M. (1996). The volitional benefits of planning. In P. M. Gollwizer and J.A. Bargh (eds), The psychology of action: linking cognition and motivation to behavior, 287-312. NY: Guilford Press.
[25]. Helie, S. & Sun, R. (2010). Incubation, insight and creative problem solving: a unified theory and a connectionist model. Psychological Review 117 (3): 994-1024.
[26]. Hillis, A.E., Rapp, B. & Caramazza, A. (1995). Constraining claims about theories of semantic memory: More on unitary versus multiple semantics Cognitive Neuropsychology 12 (2), 175 – 186.
[27]. Hinton, G.E. (1990) Mapping part-whole hierarchies into connectionist networks, Artificial Intelligence 46, 47-75.
[28]. Hinton, G.E. & Shallice, T. (1991). Lesioning an attractor network: investigations of acquired dyslexia. Psychological Review 98 (1), 74-95.
[29]. Hirsch, C. & Mathews, A. (1997). Interpretative inferences when reading about emotional events. Behavior Research and Therapy 35 (12): 1123-1132.
[30]. Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational capabilities, Proceedings of the National Academy of Sciences 79, 2554-2558.
[31]. Horst, JS, McMurray, B. & Samuelson, L.K. (2006). On-line processing is essential for learning: understanding fast mapping and word learning in a dynamic connectionist architecture. In Proceedings of CogSci'06.
[32]. Ihssen, N., Heim, S. & Keil, A. (2007). The costs of emotional attention: affective processing inhibits subsequent lexico-semantic analysis, Journal of Cognitive Neuroscience 19 (12): 1932-1949.
[33]. Iza, M & Konstenius, A. (2010). Verb representation and modularity. In: Perez, LA, Izaguirre, A, editors. Advances in cognitive science. pp. 253-280. Bilbao: UPV.
[34]. Jackendoff, R. (1990). Semantic structures. Cambridge, MA: MIT Press.
[35]. Jackendoff, R. (2010). Meaning and the Lexicon: The Parallel Architecture 1975–2010. Oxford: Oxford University Press
[36]. Kahan, T.A. & Hely, C.D. (2008). The role of valence and frequency in the emotional Stroop task. Psychonomis Bulletin & Review 15 (5): 956-960.
[37]. Kanfer, R. (1996). Self-regulatory and other non-ability determinants of skill acquisition. In P. M. Gollwizer and J.A. Bargh (eds), The psychology of action: linking cognition and motivation to behavior, 404-423. NY: Guilford Press.
[38]. Kauschke, C. & Stenneken, P. (2008). Differences in noun and verb processing in lexical decision cannot be attributed to word form and morphological complexity alone. Journal of Psycholinguistic Research 37 (6): 443-452.
[39]. Keil, F.C. (1987) Conceptual development and category structure. In U. Neisser (Ed.) Concepts and conceptual development, Cambridge: Cambridge University Press , pp. 175-200.
[40]. Keil, F.C. (1989). Concepts, kinds and cognitive development. A Bradford Book
[41]. Klinger, E. (1996). Emotional influences on cognitive processing, with implications for theories of both. In P. M. Gollwizer and J.A. Bargh (eds), The psychology of action: linking cognition and motivation to behavior, 168- 192. NY: Guilford Press.

[42]. Lachter, J. & Bever, T.G. (1988) The relation between linguistic structure and associative theories of language learning - A constructive critique of some connectionist learning models, Cognition 28, 195-247.
[43]. Lakoff, G. (1987). Women, fire, and dangerous things: What categories reveal about the mind. Chicago: University of Chicago.
[44]. Landauer, T. K. & Dumais, S. T. (1997) A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. Psychological Review, 104(2) , 211-240
[45]. Larsen, R., Mercer, K.A., Balota, D. & Strube, M.J. (2008). Nota ll negative words slow down lexical decisión and naming speed: importance of Word arousal. Emotion 8 (4): 445-452.
[46]. Li, P., Farkas, I. & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. Neural Networks 17, 1345-1362.
[47]. Lyon, K. & Chater, N. (1990) Localist and globalist approaches to concepts. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie & G. Erdos (Eds.) Lines of thinking: Reflection on the psychology of thought, Vol 1: Representation, reasoning, analogy and decision making, pp. 41-56, Chichester: Wiley.
[48]. MacKay, D.G., Shafto, M., Taylor, J.K., Marian, D.E., Abrams, L. & Dyer, J.R. (2004). Relations between emotion, memory and attention: evidence from taboo Stroop, lexical decision and immediate memory tasks. Memory & Cognition 32 (3): 474-488.
[49]. Massaro, D.W. (1988) Some criticisms of connectionist models of human performance, Journal of Memory and Language 27, 213-234.
[50]. Mathews, A. & MacLeod, C. (1994). Cognitive approaches to emotion and emotional disorders. Annual Review of Psychology, 45: 25-50.
[51]. McClelland, J.L. & Kawamoto, A.H. (1986) Mechanisms of sentence processing: Assigning roles to constituents of sentences. In McClelland & Rumelhart (Eds.) 1986, Chapter 19, pp. 272-326.
[52]. McClelland, J.L., Rumelhart, D.E. & the PDP Reseacrh Group (Eds.) (1986) Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2, Cambridge, MA: MIT Press.
[53]. Mervis, C.B. & Crisafi, M.A. (1982) Order of acquisition of subordinate-, basic-, and superordinate-level categories, Child Development 53, 258-266.
[54]. Miller, G. & Fellbaum, C. (1991). Semantic networks of English. Cognition 41, 197-229.
[55]. Moss, HE, Tyler, LK, Durrant, M & Bunn E (1998). Two eyes of a see-through: impaired and intact semantic knowledge in a case of selective deficit for living things. Neurocase 4: 291-310.
[56]. Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.
[57]. Oaksford, M., Chater, N. & Stenning, K. (1989) Connectionism, classical cognitive science and experimental psychology, Research Paper EUCCS/RP-29, Centre for Cognitive Science, University of Edinburgh.
[58]. Pinker, S. (1989). Learnability and cognition: the acquisition of verb-argument structure. Cambridge, MA: Harvard University Press.
[59]. Pinker, S. & Prince, A. (1988) On language and connectionism: Analysis of a parallel distributed processing model of language acquisition, Cognition 28, 73-193.
[60]. Pittermann, J., Pittermann, A. & Minker, W. (2010). Handling emotions in human-computer dialogues. Berlin: Springer Verlag.
[61]. Plaut, D.C. (1995). Semantic and associative priming in a distributed attractor network. Proceedings of the 17th Annual Conference of the Cognitive Science Society (pp. 37-42). Hillsdale, NJ: Lawrence Erlbaum Associates.
[62]. Plaut, DC (2002). Graded modality-specific specialization in semantics: a computational account of optic aphasia. Cognitive Neuropsychology 19: 603-639.
[63]. Plaut, D. C. & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. Cognitive Neuropsychology, 10, 377-500.
[64]. Quillian, M.R. (1969) The Teachable Language Comprehender: A simulation program and theory of language, Communications of the Association for Computing Machinery 12(8), 459-476.
[65]. Regier, T. (2005). The emergence of words: attentional learning in form and meaning. Cognitive Science 29, 819-865.
[66]. Rieman, B.C. & McNally, R.J. (1995). Cognitive processing of personally-relevant information. Cognition and Emotion 9: 325-340.
[67]. Rumelhart, D.E. & McClelland, J.L. (1986) On learning the past tenses of English verbs. In J.L. McClelland, D.E. Rumelhart & the PDP Research Group (Eds.) 1986, Chapter 18, pp. 216-271
[68]. Rumelhart, D.E., Smolensky, P., McClelland, J.L. & Hinton, G.E. (1986) Schemata and sequential thought processes in PDP models. In J.L. McClelland, D.E. Rumelhart & the PDP Research Group (Eds.) 1986.
[69]. Schneider, W. (1987) Connectionism: Is it a paradigm shift for psychology? Behavior Research Methods, Instruments, & Computers, 19 (2), 73-83.
[70]. Shallice, T. (1988). Specialisation within the semantic system. Cognitive Neuropsychology 5 (1), 133 - 142
[71]. Shapiro, S.C. (1987) Encyclopedia of Artificial Intelligence, Vol. 1 y 2, John Wiley & Sons.
[72]. Shastri, L. (1988a) Semantic networks: An evidential formalization and its connectionist realization, Morgan Kaufmann.
[73]. Shastri, L. (1988b) A connectionist approach to knowledge representation and limited inference, Cognitive Science 12, 331-392.
[74]. Shastri, L. (1989) The relevance of connectionism to AI: A Representation and Reasoning Perspective, MS-CIS-89-05, Department of Computer and Information Science, University of Pennsylvania.
[75]. [Shastri, L. (1990) Connectionism and the computational effectiveness of reasoning, Theoretical Linguistics 16 (1), 65-87.

[76]. Shastri, L. & Ajjanagadde, V. (1989) A Connectionist System for Rule Based Reasoning with Multi-Place Predicates and Variables, MS-CIS-89-06, Computer and Information Science Department, University of Pennsylvania.
[77]. Shastri, L. & Ajjanagadde, V. (1990) From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings, Technical Report MS-CIS-90-05, University of Pennsylvania.
[78]. Shastri, L. & Ajjanagadde, V. (1992) From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony, Revised Technical Report MS-CIS-90-05, University of Pennsylvania.
[79]. Shastri, L. & Ajjanagadde, V. (1993) From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony, Behavioral and Brain Sciences 16 (4).
[80]. Shastri, L. & Feldman, J.A. (1986) Semantic nets, neural nets, and routines. In N. Sharkey (Ed.) Advances in Cognitive Science, Chichester, UK: Ellis Harwood Limited/John Wiley & Sons.
[81]. Sun, R. (2002). Duality of the mind: a bottom-up approach toward cognition. Mahwah: LEA.
[82]. Sun, R. (2009). Motivational representations within a computational cognitive architecture. Cognitive Computation 1(1), 91-103.
[83]. Sun, R, Wilson N & Mathews R (2011). Accounting for certain mental disorders within a comprehensive cognitive architecture. Cognitive Computation. Published online.
[84]. Tamagni, C., Mantei, T. & Brugger, P. (2009). Emotion and space: lateralized emotional word detection depends on line bisection bias. Neuroscience 162 (4): 1101-1105.
[85]. Thomas, LA & LaBar, KS (2005) Emotional arousal enhances word repetition priming. Cognition and Emotion 19 (7), 1027-1047.
[86]. Touretzky, D.S. (1990) BoltzCONS: dynamic symbol structures in a connectionist network, Artificial Intelligence 46, 5-46.
[87]. Vanhatalo, U. (2002a). "Naiset motkottaa aiheesta ja nalkuttaa syyttä": Kyselytestit verbien semanttisten sisältöjen arvioinnissa. [Using questionnaires to assess semantic content of verbs]. Virittäjä 106: 330–353.
[88]. Vanhatalo, U. (2002b). Population Tests in Lexicography. Proceedings of the North West Linguistics Conference 2002. Eds Geoffrey Stewart Morrison & Les Zsoldos. The Simon Fraser University Linguistics Graduate Student Association.
[89]. Vanhatalo, U. (2004). Koehenkilöiden taustan merkitys kokeellisessa leksikaalisessa semantiikassa ja sen sovelluksissa. [The importance of the backgrounds of informants in experimental lexical semantics and it's applications]. Sananjalka 46: 81–105.
[90]. Vanhatalo, U. (2005). Kyselytestit synonymian selvittämisessä. Sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan. Helsinki. [The use of questionnaires in exploring synonymy. Lexical knowledge from native speakers to electronic dictionaries] PhD thesis. Location of document (http://ethesis.helsinki.fi/julkaisut/hum/suoma/vk/vanhatalo).
[91]. Vermeulen, N., Luminet, O. & Corneille, O. (2006). Alexithymia and the automatic processing of affective information: evidence from the affective priming paradigm. Cognition and Emotion 20(1), 64-91.
[92]. Vigliocco, G., Vinson, D., Lewis, W. & Garrett, MF (2004). Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. Cognitive psychology 48, 422-488.
[93]. Vinson, D.P. & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. Behavior Research Methods 40(1), 183-1909.
[94]. Wickelgren, W.A. (1979) Chunking and consolidation: a theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system, Psychological Review 86, 44-60.
[95]. Xu, F. & Tenenbaum, JB (2007). Word learning as Bayesian inference. Psychological Review 114(2), 245-272.
[96]. Young, J. (1987). The role of selective attention in the attitude-behavior relationship. Unpublished doctoral dissertation. University of Minnesota.
[97]. Yu, C. (2005). The emergence of links between lexical acquisition and object categorization: a computational study. Connection Science 17(3-4), 381-397.
[98]. Yu, C. (2006) Learning syntax-semantics mappings to bootstrap word learning. In Proceedings of CogSci'06.