Journal of Neurolinguistics xxx (2010) 1–29

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Formal Neurosemantics. Logic, meaning and composition in the Brain

Daniele Panizza [a,b,*]

[a] Department of Cognitive Science, University of Trento, Corso Bettini n. 31, 38068 Rovereto, Italy
[b] Fondazione Marica De Vincenzi, ONLUS, Italy

ARTICLE INFO

Article history:
Received 13 December 2009
Received in revised form 15 November 2010
Accepted 30 November 2010

Keywords:
Semantics
Compositionality
Neurosemantics
Semantic processing
Meaning

ABSTRACT

In the last century philosophers, mathematicians and linguists put much effort in building formal models to describe and explain the complexity of language and the meaning of words. Concepts such as truth value, compositionality, recursion, predication and logical entailment have become well known in the linguistic field of formal semantics. In the last decades, on the other hand, neuropsychologists, physicians and cognitive scientists started developing methodologies to investigate how different kinds of information are processed in real time by the brain. Electroencephalography (EEG), functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) allow us to inspect how and where many kinds of stimuli, including words and sentences, commit neuronal populations to work. In the current paper we review some recent experimental studies investigating the linguistic mechanisms postulated by formal theories of meaning in the brain of language speakers.

In particular, we will focus on the processing of Negative Polarity Items, which are terms such as *ever* or *any* that require specific semantic demands in order to be correctly used - or understood - in a sentence. Then we will explore the compositional aspects of meaning contrasted to the world knowledge based ones, and we will compare some theories that challenge or find evidence for distinct neuronal substrates handling these mechanisms. Finally, we will briefly review some formal semantics construals that have been studied through neurolinguistics methods, such as modal subordination.

* Dipartimento di Scienze della Cognizione e della Formazione, Corso Bettini n. 31, 38068 Rovereto (TN), Italy. Tel.: +39 0464 808668; fax: +39 0464 808655.
E-mail address: daniele.panizza@gmail.it.

0911-6044/$ – see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.jneuroling.2010.11.001

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

After this survey, we will conclude that there is neuroscientific evidence that the human brain implements semantic representations and operations that have the following properties: they are abstract, symbolic and grammar-driven. The challenge for the future research, then, will be to figure out how brain functions cooperate and interact with other cognitive systems, in dealing with this kind of information structures.

© 2010 Published by Elsevier Ltd.

1. Introduction

The main goal of cognitive neuroscience is to explain how the brain governs and produces the complex behavior of humans. Since the "cognitive revolution" (cf. Pinker, 2002), it is a widely shared belief that many behavioral manifestations rely on internal representations and processes, which are not overtly visible through observation. It is typically the case that complex behavior is the end product of a set of computations that are "silently" performed by the brain in a fast, efficient and successful way.

At the very beginning of the cognitive inquiry the only means of inferring that the behavioral response to a given stimulus requires more cognitive effort than the same response to another stimulus was to measure reaction times. When a stimulus is harder to process, it requires more - or more costly - computations to be performed by the brain. But what could be said about different stimuli requiring the same amount of time to be processed? What if such stimuli presented in different tasks, modalities and experimental conditions, would yield the same response within the same amount of time? Almost nothing. With the advent of neuroscience methodologies, such as event related potentials (ERPs), functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG), things changed significantly. These techniques allow the study of cognitive functions in real time, and in healthy subjects. Furthermore, they permit investigation of the timing and the localization of neural networks that are active in different tasks, even when such tasks do not elicit different reaction times. This has made modern neuroimaging a fundamental tool for the evaluation of theoretical models of diverse cognitive functions. In the present paper some works using neuroscientific and psycholinguistic methods to investigate natural language semantics will be reviewed. More precisely, this review focuses on experiments exploring how our brain deals with representations and operations that have been claimed, in the framework of linguistics, to rely on symbolic and abstract types of information, and bearing a special relationship with grammar. Before discussing experimental data, however, we will briefly characterize the main properties of such representations. Afterwards, in this introduction, we will discuss some influential proposals on the architecture of language in the brain, which will lead us to an ongoing debate regarding the nature on how the linguistic meaning is implemented in brain networks. The aim of this cursory overview is to provide a theoretical motivation for the experimental reviews we will analyze in the rest of the paper.

Language can be thought of as a code exploited by humans in order to communicate in a very fast and efficient way. To master the rules of such a code is to know how to speak. A speaker of a language must possess the abilities to assemble linguistic tokens (i.e. words) into more complex structures (i.e. sentences) following syntactic rules. She must also know what these tokens signify, in order to reconstruct a meaningful representation of the linguistic message, which is the ultimate goal of the communicative process. Remarkably, we can speak about any kind of concrete object, abstract idea, past or future event, emotion, state of affairs, etc. This is possible because the agents of communication share the knowledge of the rules (syntax) of the code they are using (their language), the meaning of each word (lexicon) but also the extra-linguistic concepts about the world (conceptual and world knowledge). Understanding linguistic communication is therefore possible because a given word conventionally evoke the same concept in any speaker of the same language, due to the link that ties lexical items (words) to conceptual representations.

According to ideas coming from the framework of formal semantics and philosophy of language, illustrated in a short survey in the next section of the present paper, something important appears to be

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

missing in the picture of language and communication just laid out. Namely, human speakers, in order to be able to process and interpret linguistic stimuli, must be equipped with a repertoire of objects and operations laying in an intermediate level between syntactic and extra-linguistic conceptual structures. These objects can be characterized by the following distinctive properties. First, they are *abstract*, in a sense that they are not tied to a specific existing entity or sensorial modality. Second, they are *symbolic* in nature, which means that they can be manipulated through mathematical, logical and formal operations. Third, they are *grammar-driven*. That is, their instantiation is directed by grammatical rules and they structure is derived from the syntactic structure of the sentence. Under this view, the ontology of language semantics is not just a collection of associative links between words and concepts, but it must include a highly structured set of objects undergoing formal - i.e. mathematical, logical - compositional operations. Thus, according to this view, the existence of lexical associations between words and non-linguistic concepts is not in doubt but the lexicon must be supported by a cognitive apparatus that is able to deal with abstract, symbolic and grammar-driven representations.

The main thesis we will try to defend in this paper is that objects and operations having the properties that we indicated, as described in detail by formal semantics, are actually implemented in computations and representations in our brain. Although the goal of linguistic theories is by no mean to provide explicit models of how language is handled by brain networks, we believe there can be a fruitful exchange of information between formal linguistics and neuroscience. As theoretical models of language are not just a mere description of linguistic facts, they should be able to offer a rather precise definition of the kind of information structures and computations that our brain must entertain, in order to produce the observed linguistic behavior. As anticipated, in order to better understand how symbolic semantic representations may be implemented in brain networks and interplay with other linguistic and extra-linguistic functions, we briefly illustrate some neuroscientific models on the architecture of language in the brain.

The way our brain processes language has been investigated for more than one century. In the nineteenth century two brain loci in the left hemisphere were identified as language areas. The first neurophysiological model of language was developed on the lesional studies conducted by two pioneer neuropathologists: Carl Wernicke and Paul Broca. This model maintains that the Left Inferior Frontal Gyrus (LIFG, Broca's area) contains the articulatory motor programs involved in language production, whereas the Superior Temporal Gyrus (STG, Wernicke's area) contains the phonological representation of spoken words, necessary for language comprehension. The modern version of this model (cf. Geschwind, 1970) suggests that Wernicke's area and surrounding regions are not only deputed to the phonological decoding of words, but also to their semantic analysis. Functional specialisation of Broca's area has been changed in the recent years too. Grounded in experimental studies conducted on patients having this area impaired, Grodzinsky and Amunts (2006) and Grodzinsky and Santi (2008) proposed that core syntactic operations such as syntactic covert movement are specifically housed in the Broca's area. A very influential cognitive model that supports the functional localization of language in the brain is Fodor's modularity theory (see Fodor, 1983). According to this view, some high-level cognitive functions such as those involved in language understanding and production are domain-specific, informationally encapsulated and implemented in fixed modular architecture. The modular view of language maintains that different linguistic functions such as phonological decoding, syntactic operations and semantic interpretation rely on separate cognitive systems operating in a serial, bottom-up and mandatory fashion. For example, the semantic interpretation of a sentence may only occur after the "syntax module" has properly constructed the sentential syntactic structure. Thus modularity provided support for the following hypotheses concerning language architecture. First, symbolic semantic objects are grammar-driven in that they are directly read off of syntactic structures. Second, distinct linguistic functions are implemented in separated areas of the brain and they deal with different ontologies. For what concerns formal semantics, symbolic structures can be identified from their informational properties: they are distinguishable from syntactic structures, but also from extra-linguistic conceptual representations, which are not necessarily symbolic and grammar-driven.

In the last decade a new way of conceiving of the brain–language relationship, in direct opposition to modularity theory has become very popular in the field of neuroscience. According to this view, language functions, instead of being served by autonomous and independent modules, are argued to emerge from the interaction of distributed brain networks (cf. Pulvermüller, 2005, among many). Such

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

networks include sensorial, motor and pre-motor areas, which were previously argued not to contribute directly to linguistic processing. Different arguments have been proposed to support this view. First, it has been shown that Broca's and Wernicke's areas are directly linked with sensorial (auditory), associative (prefrontal) and motor (dorsal and ventral pre-motor cortex and primary motor cortex) areas through adjacent as well as long-distance cortical connections (cf. Makris et al., 1999). However, this broadly interconnected network does not seem to be specific to human language. Homologues of such a system are found in monkeys, displaying the same anatomical properties and providing a tight cross-species parallelism (cf. Bremmer et al., 2001; Romanski et al., 1999). In primates this network is geared to action control and execution. This suggests that even in humans, language understanding might rely heavily on the motor system. Indeed, it has been shown that perceiving an action word (e.g. *lick* and *kick*, see Buccino et al., 2001) activates very rapidly (100–200 ms) the part of the motor cortex responsible for the motor control and execution of that action, in a somatotopic fashion. That is, hearing the word 'lick' activates the part of the motor system controlling the movements of the tongue, whereas hearing the word 'kick' activates neuronal assemblies deputed to control execution of leg movements. In addition, listening to sentences describing actions modulates the neural motor response inducted by TMS (Fadiga, Craighero, Buccino, & Rizzolatti, 2002). This suggests that language comprehension emerges from the interaction of distributed cortical networks recruiting both language areas and those areas committed to processing of the real actions described linguistically. Insofar as the importance of STG and LIFG for language processing is not denied, along with the embodied view of language the concept of "language areas" should be abandoned. Parts of the brain that were thought to be highly specialized for linguistic functions appear, as a matter of facts, to be multifunctional. Broca's area, for instance, besides being associated with syntactic processing (Grodzinsky & Santi, 2008) as well as with verbal working memory (Caplan & Waters, 1999) seems to host extra-linguistic functions. Namely, it has been claimed to house mechanisms associating action observation and execution in humans (Fadiga & Craighero, 2006; Rizzolatti & Arbib, 1998) and its homologue in monkeys (area F5) is argued to contain mirror neurons, which fire during the execution of a gesture but also during the observation of the same gesture performed by conspecifics (cf. Rizzolatti & Craighero, 2004).

Proponents of the embodied view of language (Barsalou, 1999; Feldman & Narayanan, 2004; Gallese & Lakoff, 2005; Glenberg & Roberston, 1999; Johnson, 1987; Kaschak & Glenberg, 2000; Lakoff & Johnson, 1999; MacWhinney, 1998) do not just maintain that motor system grounds both language understanding and action perception and control. They claim that linguistic meaning is processed by the brain in the very same motor representations involved in perceiving, controlling and executing a given motor program. The cognitive mechanism underlying the understanding of action words is the internal *simulation* of that action performed by the motor system. Through such mechanism we are able to build novel conceptual representations from those we have already experienced. But we are also able to form an abstract concept from the exploitation of a concrete, grounded one (cf. Barsalou, 1999).

The idea that meaning is embodied, grounded in the motor system, does not a priori rule out the existence of ungrounded symbolic representations and processes. Notice that examples of ungrounded representations are semantic ones as we previously defined them (i.e. abstract, symbolic and grammar-driven). Even if we maintain that 'real understanding', the activation of conceptual structures from linguistic input, is directly caused by the neural activation of sensorial or motor multimodal areas (cf. Mahon & Caramazza, 2005, 2008, for arguments against this claim), we might still endorse the idea that symbolic and abstract - hence amodal and ungrounded - operations are performed in order to guide the activation of the appropriate extra-linguistic representations. That is, they may well be needed to figure out which grounded - motor, sensorial, multimodal etc. - representation should be recruited, in order to arrive at the ultimate comprehension of the linguistic message (see the second section of this paper).

Among the proponents of the embodied view of language some strong claims have been made which directly challenge this idea. According to Glenberg and Robertson (1999), the hypothesis that meaning arises from the syntactic combination of abstract, amodal symbols is psychologically not adequate. Their aversion to formal objects is exemplified in the following statement, which they quote from Edelman (1992): the abstract symbol view of meaning 'is one of the most remarkable misunderstandings in the history of science'. Instead of ungrounded symbols, Glenberg and Roberson (2000);

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

(cf. also Kaschak & Glenberg, 2000) propose a cognitive theory of linguistic meaning based on grounded symbols such as perceptual symbols as described by Lakoff and Johnson (1999) and Barsalou (1999). In line with these claims, Gallese and Lakoff (2005) propose that all grammatical concepts should have the properties of *cogs* (Narayanan, 1997). Namely, they are simulated in secondary brain areas and permit to draw inferences via simulation mechanisms. However, for neural structures in secondary areas are inseparable in behavior from the primary structures that they are connected to, all grammar structures must be embodied, implemented in sensory and motor circuits and descending directly from primates neural systems (Gallese & Lakoff, 2005). Thus they claim: 'Neither semantics nor grammar is symbolic, in the sense of theory of formal systems, which consists of rules for manipulating disembodied meaningless symbols' (Gallese & Lakoff, 2005).

In the present paper we attempt to demonstrate that these claims are seriously problematic from both theoretical and empirical perspectives. The first goal is to show that our brain must be able to deal with the kind of semantic ontology we described above. We will not bring arguments in favor or against a specific theoretical proposal, but rather, we will try to demonstrate that, first, our brain implements representations and operations described by formal semantics and, secondly, that their processing activates different neural circuits with respect to both syntax and extra-linguistic conceptual knowledge. One key point we advance is that the existence of ungrounded semantic representations is not necessarily dependent on the validity of a specific neuroscientific model of brain and language. For example, does our cognitive system work in a modular or parallel fashion, be language functionally specialized in fixed neural architecture or broadly distributed in several areas, we claim that the existence of symbolic and abstract representations is independently motivated. Instead, we believe it is a challenging goal for neuroscientific models of language to determine how symbolic functions are carried out by the brain and by which neural networks they are performed.

The second goal of this work is to describe the functioning of symbolic processes by reviewing experimental works that investigate their timing and localization through neuroscientific methodologies, as well as the interaction between symbolic functions and extra-linguistic ones.

We will try to achieve these goals through the following roadmap. In the next section we present a brief introduction to the core ideas of formal semantics, which will become useful to unfamiliar readers for understanding the rest of the paper. A reader who is already expert in such notions may skip this section. The following section contains a critique of the idea that the perceptual symbols can entirely substitute the abstract ones. By analyzing the interaction between negation and logical connectives we show that grounded symbols cannot, alone, predict the correct interpretation of propositions containing these functors. The following sections review experimental studies investigating different semantic phenomena and their processing during online sentence comprehension. It will be shown that semantic compositional representations and processes generate neurophysiological patterns of activation that are distinguishable from extra-linguistic conceptual functions as well as syntactic processes. For instance, processing ungrammatical sentences containing illicit occurrences of Negative Polarity Items (e.g. the word *ever*) generates a complex pattern of electrophysiological results that can be successfully explained relying on theoretical models of grammar and its interaction with formal semantics. Furthermore, we will argue that abstract grammar-driven symbols are a powerful theoretical tool to account for how our brain processes sentential operators (e.g. negation) and modal markers. With respect to modality, we will see that structural constraints that tie features of meaning to discourse structures offer a promising account of observed experimental data. Finally, we will take into account some modern proposals on how different sources of information are considered during language comprehension (cf. Hagoort & Van Berkum, 2007; Kuperberg, 2007) and we will face some challenges and implications that such proposals lead to. The materials laid out in this review should be sufficient to assess the claim that it is possible, indeed valuable, to integrate models on sentence comprehension and new ideas on language architecture in the brain with the existence of symbolic and abstract semantic representations.

2. Formal semantics: truth, compositionality and information structures

Semantics is the study of meaning in natural language. An influential idea on meaning that goes back to philosophers like Wittgenstein and has penetrated modern semantics is that

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

(1) To know the meaning of a sentence is to know its truth conditions

This statement synthesizes two important points. First, the main units of analysis in modern semantics are sentences and the relations they contract with one another, rather than single words. Second, the semantic contribution of words is to be understood in terms of their contribution to truth conditions. A standard way to execute the idea embodied in (1) is due to Tarski (1935). According to Tarski, a semantic for a language L should derive for each sentence S of L the conditions under which it is true. In the case of simple sentences, truth conditions take the following trivial-looking form:

(2) "the snow is white" is true if and only if the snow is white

On the left hand side of the biconditional we refer to a particular sentence; on the right hand side we use it, thereby relying on the semantic competence of the speaker. The apparent triviality of statements of this form was discussed, by Davidson (1967), among others, who pointed out that such formal theories of meaning reveal very little about the conditions under which *individual* sentences are true. The proper goal of formal semantics, according to Davidson, may be identified "in relating the known truth conditions of each sentence to those aspects ("words") of the sentence that recur in other sentences, and can be assigned identical roles in other sentences" (Davidson, 1967). The point that Davidson makes through such considerations is that language presents *regularities*. These regularities concern classes of words and sentences and are part of the implicit knowledge that speakers have of their native language. For instance, anyone would share the intuition that "dog", "house" and "affection" have something in common, which is different from what "build", "bark" and "wonder" have in common. Trivially, one can say "the dog" or "the house" but not "the build" or "the bark", as dog is a noun and build is a verb. Moreover, we can individuate important differences among words of the same class (e.g. "dog" vs. "affection" and "bark" vs. "build"). Notice, for example, that one can have "affection for children" but not a "dog for children". Further, one can "build a house" but one cannot "bark a house". Many theories of natural language semantics - and not just those that belong to formal semantics - postulate that words may refer to different sorts of entities such as concrete individuals (dogs and houses), attitudes (affection), concrete actions (to build, bark) and mental ones (to wonder). Such approaches typically maintain that each lexical entry has its own argument grid that specifies how many, and what kind of arguments it takes. The noun "affection" requires an argument specifying the complement of one's affection, whereas the noun "dog" clearly does not.

What is somehow special about formal semantics, relative to other theories on linguistic meaning, is that it is geared towards handling the logical properties of words and sentences. This idea can be illustrated with certain classes of words, namely *quantifiers* and *connectives*, which are present in every natural language. Take a sentence like the following:

(3) There is some kid who is eating a sandwich.

One way to spell out the truth conditions of (3) is by mapping it into a formal language with a simple syntactic structure explicitly designed to bring out the fundamental logical relations. Such a language contains variables (x, y), which range over individuals (me, you, kids, sandwiches), and predicates that take variables as arguments. It furthermore contains operators capable of binding variables. For example, both *some* and the indefinite article *a* are translated with the existential quantifier ($\exists$), which informally is to be read as "there exist at least one individual". The expressions of such a formal language are then assigned truth conditions relative to a model (i.e. abstract set-theoretic structures that code 'the way the world is', i.e. data structures of sorts)[1]. Once we have done all this work properly, we end up with a formula, such as (4), that can be taken as an explicit representation of the conditions under which a sentence like (3) is true.

[1] A model can be conceived as a simplification - a coding into a symbolic format - of a real world context or situation.

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

(4) $\exists x \exists y$ [kid(x) $\wedge$ sandwich(y) $\wedge$ eat (x,y)]

Formula (4) states that (3) is true if and only if there is an individual *x* who is a kid, and there is an individual *y* that is a sandwich such that *x* eats *y*. Therefore, as soon as our model provides a boy standing in an eating relation with a sandwich, (3) is true.

To get some grasp for the power of this method, let us consider slightly more complex sentences including phrasal connectives.

(5) a Some boy is eating a sandwich or an orange.
b $\exists x \exists y \exists z$ [kid(x) $\wedge$ sandwich(y) $\wedge$ orange(x) $\wedge$ (eat(x,y) $\vee$ eat(x,z))]

(6) a Some boy is eating a sandwich and an orange.
b $\exists x \exists y \exists z$ [kid(x) $\wedge$ sandwich(y) $\wedge$ orange(z) $\wedge$ (eat(x,y) $\wedge$ eat(x,z))]

First, to know under which conditions sentences such as (5) and (6) are true, we need a way to implement the meaning of the connectives *or* and *and* in our formal language. What does this job in formal semantics is the *truth functions*, which takes truth values as input, and yields a truth value as output. We may rely on first-order predicate logic and assign to *and* and *or* the truth tables of logical conjunction and disjunction, respectively, with the former being true when all the conjuncts are true and the latter being true when at least one of the disjuncts is true. Now we can compare (5) to (6). The first part of their logical form is identical. There is at least a kid *x*, a sandwich *y* and an orange *z*. The second part of (5b) says that the kid eats both the orange and the sandwich, whereas (6b) says that he can also eat just one of them. Thus, a straight intuition coming from the comparison between (5b) and (6b) is that the former is more restrictive than the latter. More precisely, (5b) is true in a subset of situations in which (6b) is true, hence the former is logically stronger. Roughly speaking, if there is a boy eating both a sandwich and an orange, it plainly follows that there is also a boy eating one of them. This meaning relation that holds between (6a) and (5a) is called *entailment*. A proposition *A* entails another proposition *B* if the truth of *A* implicates - i.e. ensures, entails - the truth of *B*. Most importantly, this relation between (5a) and (6a) - like any other types of formal relationships among sentences - holds regardless of the context in which they are evaluated. Namely, the entailment relation between (5a) and (6a) is due to the (logical) meaning of the connectives *or* and *and*, rather than on the kind of scenario in which this sentence is uttered and verified.

Taking stock, the core idea of formal semantics is to have, first, a syntactic analysis of the sentence consisting in a set of conversion rules that translate the linguistic input into a symbolic structure (i.e. the phrase or syntactic structure, or Logical Form) where the formal relations (e.g. dependency, scope relations, argument structure, agreement etc.) between the expressions are made explicit. The interpretation process, then, takes as input this structured set of elements, and computes the truth conditions of the proposition by composing the elements contained in this structure - such as individuals, variables, predicates, etc. - following a finite set of conversion rules. All this work proceeds in a compositional fashion, namely, the meanings of sentences are derived from the meanings of their composite parts.

A formal language, like our toy language used in (4–6b), is characterized by its vocabulary and syntax. The vocabulary is what determines the basic expressions a language contains. The syntax is a number of explicit rules that say how expressions may be combined with each other, thus creating other expressions. We may think of this as a program written in a programming language, where the vocabulary contains the objects of our program (e.g. strings, integers, matrices, values etc.) and the syntax is the set of operations and functions that manipulate these objects. These ideas have been formalized precisely by Montague (1970), who provided a set of syntactic and semantic rules for a formal language that could be considered a subset (or fragment) of ordinary English. Montague's contention was that, although natural languages contain vagueness and idiosyncrasies, there is no fundamental theoretical difference between the syntax-semantics relation in a language like English, and that of a formal language such as first-order predicate logic.

Here come problems. Though the meaning of quantifiers, connectives and logical operators (such as negation) in natural languages suggests that there exist some syntactic and semantics rules that lead us to translate (3), (5a) and (6a) into (4), (5b) and (6b), respectively, and give them an interpretation, it

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

seems that there is a discrepancy between natural languages and formal ones. With respect to this issue, in the modern formal semantics framework the grammars of formal (artificial) languages are *models* of the grammar of natural languages, which are realized in cognitive systems that are distinct from the directly observable human linguistic behavior they help to explain (Chierchia & McConnell-Ginet, 2000).

Furthermore, it is often the case that the first-order predicate logic, like the one described in this section, is too weak to account for the variety of meaning that can be expressed by using, for instance, modal verbs (can, must), attitude verbs (believe, fear, surprise), connectives (but, although), counterfactuals and so forth. Therefore current formal models often rely on more complex logic frameworks, such as modal or intensional logic, which include variables ranging over possible worlds besides variables ranging over individuals and truth values. We will come back to this topic in the paragraph that deals with modality in natural language and its processing in the brain, as the aim of this cursory introduction was to provide the reader with some notions to better understand the following studies we will review.

3. Formal vs. perceptual symbols in the interpretation of logical connectives

In the preceding section we have seen how the formal treatment of natural language semantics is grounded in some core ideas, such as syntactic rules, semantic rules and the way the former are translated into the latter ones in a compositional fashion. In a formal language, every logical operator (such as *and*, *or* and *not*) is interpreted as a truth function. These functions can be conceived of as inferential rules providing our system with an inferential capacity. The key assumption linking the study of formal languages to that of natural ones is that human speakers must have such an abstract capacity in order to have intuitions about the truth and meaning relations of sentences such as (4), (5) and (6). Barsalou (1999); (cf. also Barsalou, Simmons, Barbey, & Wilson, 2003) suggests that abstract and symbolic capacities of humans may emerge from the cognitive manipulation of analog and perceptual symbols. For instance, he proposes an intuitive concept of truth that relies on the comparison between internally simulated situations and perceived ones. To figure out the truth of a sentence such as (3) ("There is some kid who is eating a sandwich") people might represent an internal frame of a boy eating a sandwich, focus on the critical portion of the frame and thus compare it to the actual situation they are attending. If there is a match between the simulated and perceived situations the sentence results true. If, instead, the actual scenario differs from the internal representation, the sentence is false. Then, Barsalou proposes a way to implement the meaning of logical connectives such as negation and disjunction via these mechanisms. Several authors (e.g. Glenberg, Gallese and Lakoff; see the Introduction) consider Barsalou's proposal as the demonstration that amodal representations are not needed to understand the meaning of language. We will try to show, in this section, that any mechanism of sentential verification with respect to an internal simulated representation must display the formal properties of logical connectives, in order to make the right predictions as to the truth or the falsity of a sentence. This argument builds on the interaction among logical functors.

Barsalou (1999) explains how the meaning of *or* can be implemented in a productive system based on perceptual simulation in three stages, as summarized below.

(7) OR:
- a) construct an internal representation of the proposition via perceptual simulation
- b) simulate two separate sub-events alternating the entity under disjunction while holding the other elements of the frame constant
- c) verify whether the actual scenario matches the simulated frame

Again, the last stage of this algorithm is the comparison between the internal simulated frame and the observed situation. The operational import of disjunction is captured by (7b), where the two disjuncts are alternated in two separate sub-events. Barsalou does not provide the meaning of *and*, but we can think of how it could be implemented building on (7). Whereas the initial (7a) and final (7c) stages are presumably identical, the intermediate step must be changed in order to account for the conjunctive interpretation.

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

(8) AND:
a) = (7a); c) = (7c)
b) simulate one frame containing both entities.

Let us go back to the affirmative sentences in (5a) and (6a) to test whether the algorithms in (7) and (8) work for these cases. Sentence (5a) is usually uttered in the circumstances in which the speaker does not know for certain whether there is a kid eating a sandwich or an orange. Either one, if it is being eaten by the kid, describes the meaning of this sentence. It seems, thus, that the algorithm in (7) would work perfectly for (5a). In both cases the actual scenario would match the simulated frame, which alternates the two sub-events where a boy is eating either a sandwich or an orange. As for the conjunctive proposition in (6a), here the speaker knows that a boy is eating both an orange and a sandwich. Therefore this is the only situation satisfying the meaning of the conjunction. The algorithm we sketched out in (8) would work for this case, as conjoining the two kinds of food being eaten by some boy would exactly match the scenario described by (6a). We have demonstrated, so far, that algorithms such those in (7) and (8) may account for the meaning of logical functors, grounding on simple operations applied to perceptual symbols such as "alternating" or "conjoining" sub-events.

The next step of this argument is to show that the algorithms in (7) and (8) run into serious problems in cases where connectives such as *and* and *or* interact with other logical functors, such as negation. Let us consider sentences (9a) and (10a), which are identical to (5a) and (6a) with the exception of the presence of a negative quantifier (*no*, which means "there is no individual") applying to *boys*.

(9) a No boys are eating a sandwich or an orange.
b $\neg\exists x \exists y \exists z$ [kid(x) $\wedge$ sandwich(y) $\wedge$ orange(x) $\wedge$ (eat (x,y) $\vee$ eat (x,z))]

(10) a No boys are eating a sandwich and an orange.
b $\neg\exists x \exists y \exists z$ [kid(x) $\wedge$ sandwich(y) $\wedge$ orange(z) $\wedge$ (eat (x,y) $\wedge$ eat (x,z))]

Sentence (9a), under its more natural interpretation[2], means that there is no boy eating a sandwich and there is no boy eating an orange either. The last part of formula (9b), once applied to the negated existential predication ($\neg\exists x$), accounts for such meaning. That is, the disjunction under negation is true just in case both disjuncts are false.

If we apply the Barsalou's algorithm in (7) to interpret the disjunction in (9a) we will obtain the wrong meaning. Recall that, according to Barsalou, negation specifies that there is a mismatch between the simulated and the perceived situation. Thus, we first apply (7) to sentence (9a) and we get two alternative sub-events in which some boy is eating a sandwich (the first sub-event) and some boy is eating an orange (the second sub-event). Then we are to check whether the actual scenario displays a different situation. If no boy is eating sandwiches or oranges the meaning of (9a) turns out to be true and (7) winds up working fine. But what if we are attending to a situation where some boy is eating both an orange and a sandwich? Negation applied to the result of (7) would be satisfied in this case, as the actual scenario does not mach either any one of the alternative disjunctive frames, where some boy is eating only one kind of food. However, this is not what (9a) means, in that such a situation is clearly ruled out.

Barsalou's algorithm for disjunctive meaning fails thereby to predict the right meaning of *or* when embedded under negation. The reason underlying its failure is that (7) implements just the meaning of the exclusive interpretation of *or*. When disjunction is embedded under negation, people tend to assign *or* an inclusive interpretation (cf. Chierchia, 2004). The author nevertheless admitted that his proposal was aimed to offer a viable way to implement one of the possible meanings of *or* and negation. For, both

[2] Horn (1989), Levinson (2000) and Chierchia (2004) maintain that the exclusive interpretation of *or* is due to a pragmatic inference (Scalar Implicature) that is generally not computed (or suspended) under negation and conditionals, as empirically found by Guasti et al. (2005).

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

negation and disjunction might be polysemous. Instead of appealing to polysemy to resort to the meaning of *or* under negation, we put forward an easy way to fix the algorithm in (7) rendering it compatible with the inclusive interpretation, by modifying (7b) as follows:

(7b') simulate two separate sub-events including only one of the entities under disjunction while holding the other elements of the frame and consider a third sub-event including both entities. Alternate the three sub-events.

The new algorithm in (7b') includes the possibility that both disjuncts are true. Remarkably, the meaning of *or* now is identical to that described by the first-order logic truth tables. That is, a disjunction is false when both disjuncts are false, otherwise it's true. Once we apply the modified version of (7) (containing 7b') to (9a) and negation, we end up getting the right meaning. As now (9a) is false in a situation where some boy is eating both a sandwich and an orange.

Let us now turn to conjunction, and see whether the algorithm in (8), exploiting perceptual symbols, works in interaction with negation. Sentence (10a) intuitively means that no boys are both eating a sandwich and an orange, but it leaves open a possibility that some boy is eating either just a sandwich or just an orange. After the application of (8) to (10a) we end up having a simulated frame with a boy eating both a sandwich and an orange. Such frame has to mismatch the perceived scenario after negation is applied to (8). Thus, if some boy eats just one type of food, or nothing at all, (10a) is true, for such scenarios would be different from the simulated situation. In fact (8) seems to work for rendering the meaning of *and* in both affirmative sentences and under negation. Even under negation (8) shares the same truth conditions of the formula in (10b), which, because of the truth tables of negation applied to disjunction, is false only when both conjuncts are true.

Let us imagine, now, that the actual scenario presents some boy eating an orange, a sandwich and also a banana. Such a scenario would differ from the simulated frame as much as the one in which some boy is eating just an orange. Hence it mismatches the internal simulated representation and makes (10a) true because of negation. Would it be true, thus, that no boys are eating an orange and a sandwich if, indeed, some boy is eating a sandwich, an orange and a banana? Trivially not. This fact is easily captured by the properties of propositional logic and it can be easily explained through the set theory. As shown in Fig. 1, if negation applies to a set of individuals, it will apply to any of its subsets as well.

Conjunction, in set theory, is represented as set intersection. Hence, the set of boys eating an orange, a sandwich and a banana is a subset of the set of boys eating an orange and a sandwich (which is the intersection between the set of boys eating an orange and the set of boys eating a sandwich), as shown in Fig. 1a. Once we apply negation to the former set, the latter one is excluded automatically (see Fig. 1b), for negation is downward monotonic. If we appeal to the logic-blind cognitive operations such as "mismatch verification" there is no clear way to capture these properties, which are pervasive in natural languages. If we modify the operation "check whether there is a mismatch between representations *A* and *B*" into "check whether *A* has the same number as or more perceptual symbols than *B*", it might work for a specific case (e.g. conjunction) but will predict the wrong meaning for another one (e.g. disjunction). On the other hand, propositional logic formulas such as (9b) and (10b) do explicitly obey to the principles of first-order logic. Therefore, the only improvement that would make cognitive algorithms like (8) work for any sentential structure is adding to them such formal properties. There are several examples of this sort in natural languages (e.g. the interaction of other quantifiers such as *every*, *many*, *few* with negation) suggesting that language speakers should possess the capacity of computing these logical relations during the interpretation of a linguistic message, in a relatively fast and effortless way. One clear way to do this is to construct sentential representations such as (9b) and (10b), which explicitly state the formal relations between sentences and their elements. Such representations might well be the guide for further simulation of internal frames, coded in a perceptual format and used to verify the meaning of a sentence against the actual scenario.

In this section we have shown that perceptual symbols alone, as proposed by Barsalou (1999), cannot predict the right meaning of language connectives and functors such as *not*, *and* and *or*. This means that they cannot be used as an argument against the existence of, i.e. they cannot substitute, abstract and symbolic and grammar-driven representations and operations. Furthermore, the only way to improve logic-blind algorithms in the sense of Barsalou is to make them as close to logical connectives as possible. Even in the case where they share the same truth-conditional properties, they

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

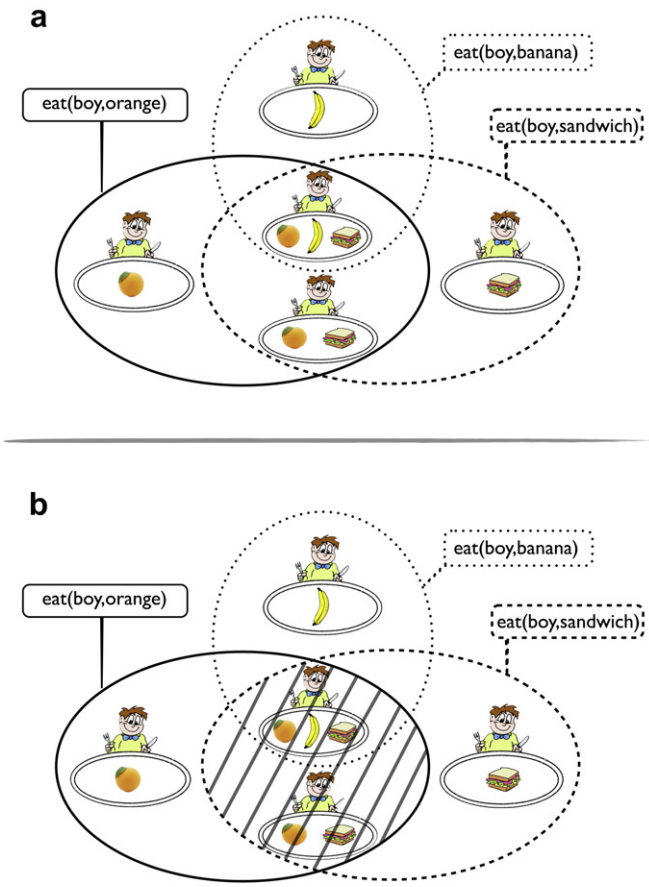

Fig. 1. a) sets of situations where sentence (10a) is evaluated; b) sets of situations where sentence (10a) is true, after negation is applied.

will fail to account for the linguistic meaning exactly at the point in which they cannot deal with logical properties of connectives and quantifiers. In the next sections we will present some studies investigating these kind of representations and operations, which will provide some evidence for the claim that our brain is indeed exploiting such informational structures.

4. Compositional semantics and world knowledge in the brain

According to a different view, embraced by several cognitive scientists (cf. Jackendoff, 2002; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Hagoort & Van Berkum, 2007), our brain is equipped with the capacity to deal with the logical symbols and compositional operations of the formal semantics framework, but there is no compelling evidence for a specialized semantic system handling those abstract computations. Symbolic relations among the objects of meaning are handled by the same systems that carry out broader non-linguistic cognitive functions such as representing events, actions, visual objects, intentions, encyclopaedic knowledge, episodic memory and emotions.

Hagoort et al. (2004) conducted several experiments to investigate whether the semantic interpretation of a sentence is separate from - and precedes in time - the integration of the meaning with

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

the world knowledge information. They presented subjects with sentences such as (11), where the critical word (typed in italics) varied across three experimental conditions.

(11) The Dutch trains are *yellow/white/sour* and very crowded.

The control condition is a natural Dutch sentence stating something true about the world, that is, the Dutch trains are yellow and crowded. The second condition was dubbed *world knowledge violation*, for it referred to a situation in which Dutch trains are white, which does not match the real state of affairs. The third condition constituted the *semantic violation*, where the adjective *sour*, usually related to taste and food, was used as a predicate of *trains*. It should not compose, though, with the noun *trains*, as their semantic features do not match, hence it was predicted to elicit a lexical/compositional semantic mismatch. In an Event Related Potentials (ERP) experiment the electrophysiological waves of the subjects were recorded while they were attending to the sentence stimuli on a computer screen. The ERP methodology in sentence processing is valuable for two reasons. First, it permits to compare the brain response elicited by different linguistic stimuli with a very high temporal resolution. If they elicit two different electrophysiological effects we may conclude that they are processed by different brain networks. Second, the ERP effects elicited through the experimental manipulation may be reconducted to brain components that have already been reported and frequently investigated in the literature. As a consequence, an experimental manipulation affecting the electrophysiological component that has previously been associated to a specific cognitive mechanism, may be attributed to influence of the same mechanism.

Hagoort et al. reported a well-known negative component, the N400, in association with both the semantic ("Dutch trains are sour") and world knowledge ("Dutch trains are white") violation. This electrophysiological component (a negative deflection arising around 250 ms and rising its peak at around 400 ms, on centro-posterior sites) is often found in association with semantic violations, difficult contextual integration and unpredictable continuations (cf. Kutas & Hillyard, 1980; cf. Kutas & Federmeier, 2000 for a review). Amplitude and latency of the N400 were the same across the two violations, therefore the authors concluded that the brain processed semantic and world knowledge incongruencies the same way, and during the same time window. Furthermore they replicated this experiment, employing the same sentences, with functional Magnetic Resonance Imaging (fMRI) technique. This methodology presents a lower temporal resolution than the ERPs, but it has a very high spatial localization capacity. What they found was that both semantic and world knowledge violations induced an increased activity in the Left Inferior Prefrontal Cortex (LIPC, Brain areas 45 and 47). From the evidence coming from these two experiments the authors concluded that both word meaning and world knowledge are integrated very rapidly in the brain. Moreover, their results suggest that it is not the case that the meaning of a sentence is first constructed independently and then verified in relation to the knowledge of the world, but these processes seem to be carried out simultaneously, and by the same neuronal networks.

One objection that we can offer against this conclusion is that Hagoort et al. do not distinguish between lexical semantics and compositional semantics features. They assume that a sentence like "trains are sour" winds up being impossible to compose, therefore its processing should result in an overload of the brain network, if it exists, which deals with abstract formal semantics features. Their conclusion is that such a specialized network does not in fact exist. This idea has been challenged by Pylkkänen, Oliveri and Smart (2009), who also developed an experimental study that aimed to contrast semantics and world knowledge processing in the brain. They employed the magnetoencephalography (MEG) technique, which offers both a high temporal and spatial resolution and it is thereby optimal means for tracking the timing of processes taking place within small portions of the brain. Pylkkänen et al. point out that the proposition "trains are sour" is not impossible to compose, as the words *train* and *sour* denote property of individuals (i.e. being *train* and being *sour*) and they may be composed, indeed, resulting in an awkward meaning. Such meaning may be classified as world knowledge violations exactly like "Dutch trains are white". Pylkkänen et al., on the other hand, employed English verbal un-prefixation, which is an operation by which a verb predicate receives a reversative meaning. The constituent "unbutton a shirt", for instance, means something like "undo the result of buttoning the shirt". This operation is semantically constrained, in that it requires an accomplishment verb, that

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

is, a verb that has a particular event structure consisting of a process that leads up to a change of state (see Dowty, 1979). Examples of their stimuli are below, where "uncorked" is a well-formed verb and may be a predicate of "wine" but not of "thirst", whereas "unchilled" generates an ill-formed proposition when used in a continuous verb tense as in (12), for the reason outlined above.

(12)the wine/thirst was being uncorked/unchilled

Notice that it is far from impossible to conceive of the meaning of "to unchill" (e.g. something like "to undo the effect of chilling something" or "warm up"), but this operation is argued to be banned from our grammar when used in a proposition such as (12), yet for reasons concerning the aspectual form of the verb. Pylkkänen et al. focused their attentions on two brain areas, namely the LIPC and the ventromedial Prefrontal Cortex (vmPFC). The former is an area involved in semantic processing (cf. Petersen, Fox, Posner, Mintun, & Raichle, 1988) and investigated deeply by Hagoort et al. (2004), whereas the latter has recently been found as the generator of the Anterior Midline Field (AMF). The AMF is an MEG component that is a candidate neural correlate of semantic composition, since it showed increased amplitudes in semantic well-formed but hard to compose sentences, such as "begin the book", where there is a resolvable type-mismatch between an eventive verb and an entity-denoting object (Pylkkänen & McElree, 2007; Pylkkänen, Oliveri et al., 2009; Pylkkänen, Martin, McElree, & Smart, 2009), and "the clown jumped for 10 min", where a verb describing a punctual event generates a repetitive reading because of the presence of a durative adverb (Brennan & Pylkkänen, 2008). The authors reported a higher activation in the vmPFC for semantic violations ("the wine was being unchilled"), compared to control sentences ("the wine was being uncorked"), but not for world knowledge violations ("the thirst was being uncorked"). On the other hand, the LIPC was more active for both semantic and world knowledge violations, as already found by Hagoort et al. (2004). Furthermore, the effect in the vmPFC was observed at 225–300 ms and at 325–350 ms whereas the increase of amplitude in the LIPC occurred only at 300–350 ms for both the violations.

These studies report, for the first time, the specific activation of a brain area that is elicited by a violation or the higher complexity of the compositional semantic structure of sentences. Furthermore, Pylkkänen, Oliveri et al., 2009; Pylkkänen, Martin et al., 2009 found that the greater activation in the vmPFC preceded that in the LIPC by more than 50 ms. This might suggest that compositional operations start being performed by the brain slightly earlier than those devoted to the integration between meaning and knowledge.

As we noticed in the introduction of the paper, the LIPC is involved in several linguistic processing domains, other than semantic integration, such as syntax, lexical-semantic processing and speech production (cf. Grodzinsky & Amunts, 2006) and it partially overlaps with the Broca's area (broadman areas 44 and 45), whose function is still under debate (cf. Grodzinsky & Santi, 2008).

Taking stock, there is some neuroscientific evidence suggesting that processes aimed to combine the meaning of constituents generate different activation patterns, compared to processes dealing with integrating conceptual representations and world knowledge. This suggests that these processes, contrary to what has been claimed by Hagoort et al. (2004) are carried out by different neuronal networks.

5. Formal meaning and entailment relations in the brain, the case of Negative Polarity Items

The licensing of Negative Polarity Items (NPIs) is a linguistic phenomenon that has recently drawn the attention of neuroscience. The reason why we will discuss some experimental studies investigating this issue is that there is an intimate connection between NPI licensing and features of meaning, such as entailment patterns, which have been studied in the field of formal semantics.

NPIs are a class of words, present in almost every natural language, which generally occur under negation or negative contexts. In English, for instance, words such as *ever* and *any* - as well as *any*-compounds like *anymore* or *anything* - and expressions such as *at all* and *lift a finger* are NPIs. Because of their distributional association with negation, some linguistic theories (Klima, 1964; Progovac, 1992) maintained that NPI licensing is a syntactically controlled phenomenon. Namely, a word such as *ever* should check for a syntactic feature (+neg) carried by negative markers (e.g. *not*) and negative

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

operators (e.g. *to doubt*). Thus, when NPIs occur under the scope of negation, as in (13a), the result is grammatical sentences, whereas when they occur in positive context like (13b), the outcome is ungrammatical sentences.

(13) a John did not *ever* drink a beer.
b * John did *ever* drink a beer.

Under different accounts (Chierchia, 2004, 2006; Krifka, 1995; Ladusaw, 1979) NPI licensing is argued to be a mechanism that is driven by semantic factors. The critical observation leading these authors to draw this conclusion is that NPIs may well occur in certain constructions (e.g. the antecedent of conditionals as in (14a), the first argument of universal quantifiers as in (14b), interrogative clauses, before sentences etc.) that are not negative, though sharing some properties with proper negative contexts.

(14) a If John *ever* drank a beer, he would become alcoholic.
b Everyone who *ever* drank a beer should try this pale ale.

The property shared by (13a) and (14a and b) is called *downward entailment*, and it is about the inference pattern associated with kind of propositions. A downward entailing context licenses entailing inferences from a set to a subset, whereas an upward entailing context, such as an affirmative proposition, licenses the opposite pattern of inferences (i.e. from a set to a superset). For example, if we take (14a) as true, it follows that even if John drank a specific kind of beer, say dark beer, he would become alcoholic. Note that the set predicated by *dark beer* is a subset of that predicated by *beer*[3]. The generalization proposed by Ladusaw (1979), Krifka (1995) and Chierchia (2004, 2006) states that NPIs are licensed in downward entailing context. This generalization has consequences for processing. During the comprehension of a sentence containing an NPI the semantic property of the linguistic context in which the NPI occurs should be checked online. If the proposition does not display the semantic characteristic needed for a correct interpretation of the NPI, its meaning will result in a logical contradiction (cf. Chierchia, 2004; Krifka, 1995). For this reason NPI violations are argued to constitute a classic case of semantic violations.

In the last five years several experimental studies (Drenhaus, Graben, Saddy, & Frisch, 2006; Saddy, Drenhaus, & Frisch, 2004; Xiang, Dillon, & Phillips, 2008) were conducted to investigate the reaction of the human brain to NPI violations, which are sentences such as (13b) in which an NPI occurs in a non-licensing context. Additionally, other studies (Drenhaus, Blaszczak & Schütte, 2007; Vespignani, Panizza, Zandomeneghi & Job, 2009; cf. also Panizza, 2009) also investigated the processing of NPIs or NPI-like words in different grammatical contexts. There are several well-known ERP components elicited by linguistic stimuli that require enhanced processing or some sort of anomaly detection and repair. We introduced in the previous section the N400 as an index of semantic processing overload and difficult contextual integration. Another well-known electrophysiological component is the P600. The P600 is a positive deflection arising from about 600 ms up to 900 ms on the posterior sites after the presentation of the critical word. It is elicited by syntactic anomalies, higher syntactic processing and syntactically complex expressions such as ambiguous sentences (Hagoort, Brown, & Groothusen, 1993; Hahne & Friederici, 1999; Kaan, Harris, Gibson, & Holcomb, 2000; Osterhout & Holcomb, 1992).

If NPI licensing is a syntactically controlled phenomenon, one clear prediction is that NPI violations should elicit a P600 effect like other kinds of syntactic violations. If, on the other hand, NPI licensing relies on semantic properties of the propositional environment (i.e. entailment), ungrammatical occurrences of NPIs might well elicit other kinds of neuropsychological components (e.g. N400-like effects). To test these empirical questions some experimental studies have been conducted on NPI violations in different languages using ERP methodology. Table 1 below presents a summary of the results coming from these studies.

[3] An upward entailing context, such as an affirmative proposition like "John drank a beer" licenses the opposite pattern of inference. Namely, it entails a sentence that predicates something about a superset, such as "John drank something" and is entailed by a sentence predicating something about a proper subset, such as "John drank a dark beer".

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

Table 1
Results from studies on NPI processing conducted with ERPs.

Experimental sentence	Index reported
(15) *Ein Mann, der einen Bart hatte, was jemals froh. *A man who a beard had was ever happy (Drenhaus et al., 2006; Saddy et al., 2004)	N400, (P600)
(16) *Most restaurants that the local newspapers have recommended in their dining reviews have ever gone out of business. (Xiang et al., 2008)	P600
(17) *Sul giornale si legge che il presidente ha mai avuto un'amante. *The newspaper reports that the president has ever had a lover.	N400, FP600 P600
(18) Sul giornale si legge che il presidente *mai* ha avuto un'amante. The newspaper reports that never the president has had a lover. (Vespignani et al., 2009; Panizza, 2009)	N400, FP600
(19) *Der Lehrer hat den Schüler jemals geschlagen. *The teacher has ever hit the student.	N400, P600
(20) *Ein Lehrer hat den Schüler jemals geschlagen. *A teacher has ever hit the student.	N400, P600
(21) Welcher Jäger hat den Angler jemals gestört? Which hunter has ever disturbed the fisherman? (Drenhaus et al., 2007)	N400

A surprising outcome of the majority of these studies is that NPI violations elicited an N400-like effect. As mentioned previously, the N400 is a well studied electrophysiological component and it is yielded by a word inducing a semantic contrast. This contrast is usually due to a difficult contextual integration, unexpected continuation, low lexical association between the trigger word and the preceding material and implausible meaning (Kutas & Federmeier, 2000; Kutas, Van Petten, & Kluender, 2006). Though, as also pointed out by Saddy et al. (2004), the kind of deviance produced by NPI violations is rather different from that induced by semantic/pragmatic implausibility, and this idea is clearly supported by the fact that sentences like (13b) sound completely ungrammatical rather than semantically odd. One straightforward way to account for these findings is to explain the N400 effects under the hypothesis that NPI violations implicate a semantic incongruency (Chierchia, 2006; Krifka, 1995). From this idea it does not necessarily follow that NPI licensing is controlled exclusively by semantic mechanisms. In fact almost all of these studies have also reported a P600 effect for the unlicensed NPIs, which probably indexes a structural repair of the ill-formed sentence, as it is commonly interpreted as a marker of a syntactic mismatch. Thus, the way our brain processes NPI violations indeed shows that both semantic and syntactic functions are involved while we cope with such anomalies. Nevertheless there results are not confined to the processing of illicit sentences only. Curiously, Drenhaus et al. (2006) and Vespignani et al. (2009) found an enhanced N400 for grammatical occurrences of NPIs and NPI-like words, when occurring in certain linguistic constructions. Drenhaus et al. reported a more pronounced N400 for sentences like (21) where the NPI *jemals* was licensed by a Wh-pronoun (*welcher*, a German word for *which*) compared to its occurrence under negation. The authors interpreted this result as indicating that *whelcher* is a weak licensor, therefore it requires more semantic processing to integrate the meaning of *jemals* in the interrogative clause than in the negative construction. Vespignani et al. found an N400 followed by a frontal and early positive deflection (FP600) in response to the presentation of an NPI-like word (i.e. *mai*, commonly defined as N-word, which in Italian can mean both *ever* and *never*) occurring in a grammatical sentence. Critically, *mai* in Italian, like other N-words in Romance languages, has a negative meaning only when it occurs in pre-verbal position and in an upward entailing context, otherwise it is interpreted exactly like an NPI

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

with a positive meaning (such as *ever* in English). When *mai* occurred in pre-verbal position, thus, it elicited the same components (i.e. N400 and FP600) that were also found in the condition where *mai* occurred in postverbal position without a licensing context, except for a posterior P600 that was only found in the latter ungrammatical condition. This led the authors to conclude that in languages such Italian the parser exploits certain semantic features of the sentence (i.e. the entailing properties) to figure out whether *mai* has to be assigned a negative meaning. If the context is positive, upward entailing, and *mai* occurs in pre-verbal position, the N-word is interpreted with a negative meaning. Thus, this process is argued to require more computational resources, resulting in an N400 plus an FP600. This suggests that N400 effects caused by NPI processing are not elicited by anomalous propositions only, but they may well be a neuropsychological correlate of successful, though effort-demanding, comprehension processes in grammatical sentences.

Taking stock, converging evidence from linguistic models and experimental studies employing the ERPs technique suggests that entailing relations are a good candidate for explaining the working of NPI licensing mechanisms. Higher semantic processing, witnessed by the presence of enhanced N400-like components, is often detected when propositional semantic requirements are not met (regardless of whether the sentence can or cannot be assigned a regular interpretation).

One proviso must be stated here. Entailing relations among propositions may be derived from symbolic - i.e. ungrounded - sentential representations, as well as from grounded ones (see Barsalou, 1999). Notwithstanding this consideration, it is hard to think of a way to compute the entailment properties of a given sentence before having accessed the whole propositional content without relying on its structural (grammatical) features. Take, for instance, (14a) and consider the point in which the word *ever* occurs. The knowledge that the conditional antecedent introduces a downward entailing context, at this point, cannot be derived from the entire propositional content because, trivially, the reader has only encountered the fragment "if John…". Nonetheless readers become aware of the anomaly as soon as they hit the critical word *ever* and it's reading immediately generates processing differences as compared to grammatical occurrences of the same word. This reasoning applies to other examples of NPI violations such as in (13a) and (14b).

Formal theories of meaning offer a straightforward account to overcome this problem. Namely, the upcoming phrasal material can be previously predicted and incorporated in the phrasal structure with abstract objects such as variables representing, for instance, individual entities and predicates. In line with this reasoning, the interpretation of any verbal predicate occurring after the words "If John…" will be constrained by the semantics of *if*, regardless of its lexical properties. If our brain is equipped with this capacity, it should be able to perform semantic computations based on the entailment relations of the sentence *before* having accessed all of the linguistic material, which is exactly what both our intuitions and processing experiments on NPIs tell us. If, instead, our brain can only elaborate on the meaning of a sentence *after* producing a simulation of the situation described, for instance, using grounded representations, we are to claim that NPIs violations are syntactic anomalies and the results we just reviewed remain unexplained.

Remarkably, run-off-the-mill syntactic incongruencies are not distinguishable from NPI violations on the basis of the observed behavior. Namely, speakers judge the latter ones as ungrammatical along with classic syntactically ill-formed sentences (e.g. argument structure violation, agreement mismatch etc.). Thus, the case of NPI violations offers compelling evidence for the idea that linguistic functions associated with syntactic and semantic processing do generate different neurophysiological activation patterns, even when they produce the same behavioral effect.

What is yet to be established, however, is whether these semantic features, associated with the meaning of NPIs, are handled by the same cognitive system that deals with non-linguistic concepts, world knowledge, semantic memory etc. One answer, on purely speculative grounds, is "no". Entailing relation and logical properties may be immediately read from an abstract and symbolic propositional representation, and they are, in principle, not sensitive to the world knowledge and contextual (in a sense of emotions, intentions, memory and visual representations) environment. Indeed, when violated, they cause ungrammaticality but not pragmatic implausibility. Another possible answer, grounded in the experimental investigation of NPIs, is positive. It relies on the N400 component that was found in association with contextual/world knowledge incongruencies as well as with NPIs violation. However, it is well known that this electrophysiological component has different sources in

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

the brain (Lau, Poeppel, & Phillips, 2008), and to provide a more reliable answer to this dilemma more research is needed, with more subtle and advanced experimental techniques.

6. How the brain handles the truth, the falsity and the negation: a challenge for formal semantics?

As we discussed in the first chapter the notion of truth is central to the framework of formal semantics: the truth conditions of a sentence *are* the meaning of the sentence. In this framework the notion of truth is to be understood in purely abstract terms. Namely, it is detached from the world as well as from its sensory-based representation in our brain. The truth conditions of a sentence can be thought of as depending solely on the inner relations between the words and the constituents of which it is composed, no matter how we wind up evaluating them in the real world or in our cognitive domain. This is the very reason why formal theories of meaning are often viewed in competition with theories grounded in simulation mechanisms (cf. Gallese & Lakoff, 2005; Glenberg & Robertson, 1999 the introduction of this work). As we anticipated, in this paper we advance a way to overcome this theoretical contraposition. One of the main claims we make is that constructing the meaning of a proposition by assembling its constituents in a compositional fashion is a necessary characteristic that our cognitive system must possess in order to arrive at the evaluation of the sentence meaning against the conceptual knowledge.

The critical point we tried to address is whether grammar-driven abstract symbols and operations are really exploited by our brain in solving this task. We presented, so far, both theoretical and experimental evidence to address this point. If we are right there are some interesting issues that emerge, which we focus on in this section. One is the question about when exactly during the online interpretation process does our brain show sensitivity to symbolic and compositional operations geared to assemble sentential constituents. Another question is the following: when does the brain become sensitive to the truth (or the falsity) of a proposition? In other terms, assuming that as soon as we process the meaning of a sentence we automatically evaluate it with respect to a contextual scenario - be it real or fictitious - when should this happen? Going back to the case of "the yellow/white Dutch trains", the work by Hagoort et al. (2004) suggests that we evaluate the meaning of each word very rapidly, as soon as we read it. However, the effects they reported are not necessarily associated with the proper evaluation of the truth of the linguistic message. Indeed, they might be due to the intrinsic complexity of the association between either the lexical items or the concepts evoked by the sentences employed in their experiments. For instance, for Dutch speakers the lexical items *train* and *yellow*, as well as the concepts they evoke, are more closely associated than *train* and *white*. For, in the Netherlands, yellow trains are far more common than white ones. Consider, furthermore, that the brain is extremely fast in integrating conceptual and world knowledge information while processing linguistic stimuli. A rule of thumb regarding such integration processes is that the higher the conceptual discrepancy between the critical word and the world knowledge is, the bigger the brain response effect that corresponds to the processing of that word.

In order to investigate the import of the truth and verification mechanisms during sentence processing some studies exploited the operator that in natural languages inverts the truth value of a sentence, namely negation. In the following paragraphs we will try to address the questions previously raised by reviewing experimental works that explore the processes of evaluating the meaning of true and false propositions in affirmative as well as in negated forms. One of our goals is to test whether the way such processes are performed by the brain poses any challenge for the thesis we defend in the current work.

An influential study on verification mechanisms was conducted more than twenty years ago by Ratcliff and McKoon (1982), who measured the time and accuracy of the answers given by the subjects at a certain delay after the presentation of the sentence (response signal procedure). The task was to verify the truth of a synonymous ("a carpet is a rug"), opposite ("a lion is a tiger"), anomalous ("a captain is a sandwich"), category-member ("a bird is a robin") and member-category ("a robin is a bird") sentences. What they found was that in category-member statements there was an increasing tendency to respond yes in early stages, replaced by an increasing tendency to respond no later on. This was one of the first experimental pieces of evidence showing that early processing of linguistic material

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

is mostly affected by lexical properties of words (e.g. bird is highly associated with robin), whereas the propositional meaning takes some time to be comprehended and, thus, affects late decision times. In contrast to this, opposite statements did not show a greater tendency to respond yes than anomalous statements did, with the former containing highly related items (lion-tiger) and the latter not (captain-sandwich).

Fischler, Bloom, Childers and Perry (1983) employed the ERP methodology to investigate verification mechanisms performed by the brain in real time. They presented subjects with sentences like "a robin is (not) a bird" vs. "a robin is (not) a tree". Note that in the former sentence the presence of negation makes its meaning false, whereas the latter sentence is made true by negation (i.e. it is true that a robin is not a tree). In previous studies where reaction times were monitored (Clark & Chase, 1972; Collins & Quillian, 1969) an interaction was found between the form of a sentence (affirmative/negative) and the truth value (true/false). That is, affirmative false sentences required more time to be evaluated than affirmative true ones did, whereas with negative sentences the pattern was reversed. Interestingly, the results from Fischler, Bloom, Childers, Roucos, and Perry (1983) mirrored the reaction times studies, in that they found a negative enhanced deflection at 250–450 ms for affirmative false vs. true sentences, and for negative true vs. false sentences. They interpret these results by stating that this negativity is likely to reflect the semantic mismatch between the two words (robin vs. bird/tree) rather than the truth - or the falsity - of the sentence taken as a whole. Therefore these results support the idea that the meaning of a negated sentence is fully understood in a subsequent stage, after the representation of the positive version of the negative sentence is built and evaluated.

Kounios and Holcomb (1992) found similar ERP results, where an N400 was elicited by category and relatedness properties of words presented both in isolation and embedded in sentences (i.e. "some dogs are animals/clothes" vs. "some animals/clothes are dogs"). They concluded that N400 effects are apparently non-sensitive to sentence truth value and would reflect non-decisional and non-propositional aspects of semantic processing - which might nonetheless be performed in sentence processing at very early stages. On the other hand, by manipulating the quantifier ("some/all dogs are animals") they found sentence truth effect affecting reaction times (i.e. higher reaction times for false propositions). Thus, the authors concluded that such processes are at play in concurrence with decision processes, after the sentence is read.

So far, the picture of processing of negation and evaluation of the truth of a proposition is quite coherent. First, in the very early stages of language comprehension (up to 400–600 ms) our brain is sensitive only to the lexical or conceptual properties of words. It is insensitive, however, to the truth-conditional meaning as well as to the switch of truth value induced by a negative operator. During the second stage, then, these factors become prominent in sentence comprehension, and surface in higher reading times associated with negation and falsity. This view has been dubbed the "two-stages hypothesis" of comprehension and it brings two important implications about how the brain constructs the linguistic meaning. First, the propositional content takes some time to be accessed. With respect to the formal theories of meaning this picture is by no means incompatible, and leaves room for some speculations. One is that the compositional processes require a certain amount of time to yield an interpretable output, which will be subjected to a process of evaluation against the conceptual knowledge. Another speculation is that compositional processes are somehow silently performed in the brain, in that they do not yield prominent effects that can be detected with neuropsychological techniques.

Secondly, since negative deflections (N400) are often found in association with category, relatedness and plausibility mismatches - regardless of the propositional content of a sentence (i.e. negation or falsity) - it seems that these two information streams are independent from one another.

However, there are some issues, raised by studies discussed in the preceding sections, which are clearly in contrast with this picture. Recall, for instance, that NPIs violations elicit N400 effects starting at 250 ms after the presentation of the critical word, which cannot be explained in terms of lexical/conceptual association among linguistic items. In order to reconcile the two-stage hypothesis of comprehension with these findings we might say that compositional processes actually start as early as lexical/associative ones, and their disruption does yield neuropsychological effects. In relation with this, note that violations of compositional mechanisms (e.g. un-prefixation in English, cf. Pylkkänen, Oliveri et al., 2009; Pylkkänen, Martin et al., 2009) produce specific MEG activations after 320 ms,

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

which may also be reflected in N400 effects. All this can be accounted for by maintaining that the N400 component is actually the product of several sub-components generated by different brain sources, which might be associated, in turn, with processes sensitive to diverse representations and information streams (e.g. lexical relatedness and plausibility vs. compositional mechanisms and entailment relations).

This whole picture finds further support in recent findings coming from experimental studies on verification and negation (Ludtke, Friedrich, De Filippis, & Kaup, 2008). Ludtke et al. employed a sentence-picture verification paradigm, in which they presented subjects with German sentences describing a scene (e.g. "In front of the tower there is a ghost") followed by a picture that could be either a true description of the sentence (e.g. a tower with a ghost in front of it) or a false one (e.g. a tower with a lion in front of it). In half of the trials the experimental sentence was affirmative, whereas in the other half it was in the negative form (e.g. "In front of the tower there is no ghost"). Finally, the delay of picture presentation was varied (250 ms vs. 1500 ms). First, the authors reported an enhanced negative shift on the noun (*ghost* or *lion*) when it was preceded by a negative quantifier (*kein*, which means *no* in English vs. *ein*, which means *a*). This shows that the presence of the negative quantifier had an early impact on the processing of the noun (starting at 250 ms and extending over to 2000 ms). Then the authors reported a typical interaction *negation* by *truth* on the N400 effect, at the presentation of the picture. In this study, however, what drove this interaction was not the lexical association between items - which was kept constant - but whether the object depicted in the picture was mentioned or not. Namely, in the false affirmative and true negative trials the sentence contained, e.g., "a ghost", whereas the picture displayed a lion. Although this N400 pattern was present both when the picture was presented at 250 ms and when it was delayed to 1500 ms, it is only in the latter condition that Ludtke et al. also found a main effect of negation and truth. To interpret these results the authors maintained that the negated state of affairs is simulated first (e.g. "a ghost in front of a tower" for the sentence "in front of a tower there is no ghost"), whereas the actual state of affairs is simulated only at a later point (i.e. "no ghost in front of a tower"). Thus the full comprehension of the meaning of a negated proposition is available only after a certain amount of time, in which decisional processes are at play, whereas non-propositional semantic effects, such as the N400 elicited by not mentioned objects, show up earlier. Critically, however, this study provided evidence that even the mere processing of negation may be reflected in neurophysiological effects, and it is at odds with the idea that negation does not elicit processing differences. This is in line with what reported by other studies (Carpenter & Just, 1975; Kaup & Zwaan, 2003) that found higher reaction times and error rates with negative sentences. Such findings are also consistent with what was suggested by the experimental studies on NPI violations, where the meaning of the negation had to be processed rapidly to account for the effects emerging at the time subjects hit the critical word (Drenhaus et al., 2006; Vespignani et al., 2009; Saddy et al., 2004).

A recent work by Nieuwland and Kuperberg (2008) provides some evidence against these claims, which support the hypothesis that negation is not necessarily processed and comprehended in a later stage. In this study subjects were presented with positive and negative sentences, which could be either true or false, while their ERPs were recorded. Crucially, the main difference between this experiment and the others is that in half of the sentences negation was pragmatically felicitous (e.g. "travelling in Baghdad isn't very safe because of the war"). In these conditions the critical word (i.e. the adjective *safe*) was contrasted to a word with an opposite meaning (i.e. *dangerous*) that rendered the sentence false, and the presence of the negation was manipulated as well. The same manipulation was applied to the other half of sentences, where the negation was *not* pragmatically felicitous (e.g. "vitamins and proteins aren't very bad for your health"). The idea underlying this study, thus, is that the lack of truth-value effect in the previous experiments might be reconducted to the pragmatic infelicity of the sentences that were employed (e.g. "a lion is not a tiger").

As a matter of fact the authors reported a significant N400 effect for false propositions (truth-value effect) - either in the affirmative or in the negative form - vs. true ones, when the negation was pragmatically licensed. In contrast, when it was not pragmatically licensed, there was no main effect of truth value but an interaction between truth value and negation, due to the false words eliciting a larger N400 in affirmative sentences but not in the negative ones. The authors conclude that in sentences with a pragmatically informative meaning the truth value is evaluated incrementally and

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

negation is rapidly interpreted and incorporated into the sentence meaning. Nieuwland and Kuperberg further claim that negation, per se, does not impose any principled obstacles onto incremental and high-level language comprehension, but they acknowledge that methodological differences between this study and the previous ones discourage a more direct comparison. In this respect it is worth noting a difference between this and other studies investigating negation. Ludtke et al. (2008), for instance, found an early effect of negation right after the negative indefinite (*kein*, which means *no* in English) at the onset of the noun. Instead, in Nieuwland and Kuperberg's study there was another word (e.g. *very*) between the negative particle (*not*) and the critical word (*safe*), which presumably gave more time to the subjects to process the meaning of negation.

The main results coming from this last experiment appear to be in contrast with one of the speculations we advanced earlier in this section. That is, they suggest that the processing of propositional content and its verification is not independent from pragmatic aspects of language comprehension, in that when negation is pragmatically felicitous the truth value of the sentence affects the N400 component (i.e. larger N400 for false statements, regardless of the presence of negation). Extra-linguistic sources of information such as pragmatic plausibility may thereby boost early evaluation of the sentence while it is still being processed. Propositional verification processes, thus, might display a certain degree of flexibility. If the words already accessed lead to a coherent and very plausible interpretation, the meaning of that fragment may be evaluated prior to the reading of the whole sentence.

With respect to the two-stage hypothesis of comprehension this new picture brings some implications. First, the duration of the initial stage, in which the brain appears to be sensitive only to lexical and conceptual properties of words but insensitive to the propositional meaning, is variable. Highly salient verbal material can render N400 effects affected by truth-value manipulations (cf. Nieuwland & Kuperberg, 2008). Second, processes aimed at verifying the meaning of a sentence against conceptual knowledge may occur before every word of such sentence is accessed. Such processes in fact elicit neuroscientific effects earlier than what some experimental studies initially suggested (i.e. in tandem with late decision processes, cf. Fischler et al., 1983; Kounios & Holcomb, 1992).

At this point one key question we anticipated in the introduction needs to be addressed. Specifically, is the way our brain evaluates the truth-conditional content of a sentence in contrast with its commitment to use of symbolic, abstract and grammar-driven semantic representations in order to construct the propositional meaning? Recall that formal theories of meaning imply that some kind of semantic analysis *must* be performed prior to any possible evaluation of the meaning against conceptual knowledge or discourse contexts. In this sense, formal theories are two-stages theories of comprehension, with the first step being the compositional construction of the propositional content and the second step being its evaluation with respect to a model.

The answer we provide to this question is negative. In contrast, the semantic analysis performed in compositional fashion may account for immediate effects of negation (Ludtke et al., 2008), compositional violations (Pylkkänen, Oliveri et al., 2009; Pylkkänen, Martin et al., 2009) and NPI processing (Drenhaus et al., 2006; Saddy et al., 2004). It offers, furthermore, powerful tools for interpreting incomplete sentences. For, as we previously pointed out, a compositional structure allows one to fill the holes that will be occupied by words that have yet to be accessed, with abstract variables (e.g. placeholders). One advantage of this process is that meaning relations between partial sentences can still be computed (e.g. entailment, contradiction). Another advantage is that predictions drawn on upcoming material can be structurally constrained. Thank to this, the search space concerning hypotheses on words that are likely to occur is more limited and, hence, predictions are easier to make.

One issue that is yet to be explained is how extra-linguistic informational sources can facilitate comprehension by interacting with linguistic (grammar-driven) representations. Consider, for instance, the modularity theory as a psychological model of semantic composition. According to this model, sentential structures are computed from the bottom up. As a consequence, semantic analysis is predicted to be influenced only by lower (i.e. syntactic) computations but, critically, not by the higher level ones. Pragmatics, plausibility and conceptual structures are by all means found in such higher levels. Thus, if symbolic representations are dependent not only on syntactic derivations, we need a way to explain how they can be modulated by top-down processes. In the last section of this work focus will be put on this issue and hypotheses to overcome this problem will be advanced.

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

7. Beyond the actual truth: how the brain handles the reference to modal contexts

The expressive power of natural languages is not only limited to real entities and events. We are indeed able to speak about hypothetical worlds, fictional entities and, further, things whose existence we are not sure about, to which, nevertheless, we can be committed in some possible world existing in our mind. Philosophers of language have been interested in this topic for decades, ever since concepts such as *intensionality* and *modality* have been developed (cf. Carnap, 1947) to model how our thoughts about possible (as well as impossible) entities are conveyed in natural language. In any language spoken in the world, in fact, there are some words that are called *mood markers* (e.g. modal auxiliaries such as *can, might, must*, non-factual adverbs such as *possibly, likely*, propositional attitude verbs such as *consider, think* etc.), whose function is to frame the entities, characters and events we are talking about within a modal context. Modal contexts are possible worlds, scenarios that might - or might not - become real in the past or in the future, depending on their accessibility relation to the actual world. Probably the most famous examples are due to Montague (1970), who elaborated on the contrast between sentences such as (22a) and (22b).

(22) a John likes a unicorn
b John saw a unicorn

Whereas sentence (22a) sounds perfectly natural, and it means that John likes a fabled creature with a single straight spiralled horn projecting from its forehead (or whatever he thinks a unicorn is), (22b) sounds weird. Its weirdness is due to the fact that the verb *to see* is factive, that is, it entails the truth of its complement in the actual world. Thus, as there exist no unicorns in the actual world, (22b) describes an impossible scenario, whereas nothing prevents John from liking something that only exists in fantasy books.

A critical issue about modality is the way it interacts with grammar and discourse. In formal semantics mood markers introduce modal operators, which just like any other formal operators have their scope (i.e. a portion of the whole proposition that is subject to the effect of the operator). In sentence (22a), for example, the existence of a unicorn predicated by the existential quantifier is embedded under the scope of a modal operator (introduced by the verb *likes*). Hence it is confined to the domain of the things that John likes, which do not have to exist in the real world. Now let us consider the following example (original sentences are due to Karttunen, 1976).

(23) a John is considering writing a novel. It could end quite abruptly
b # John is considering writing a novel. It ends quite abruptly.

The sentence in (23a) sounds natural. Its meaning may spelled out as: there is a possible world accessible to the actual one in which there exists a novel that John writes and this novel ends quite abruptly. In contrast with this, (23b) is infelicitous, as in the second clause the verb *ends* predicates something about the hypothetical novel in the actual world, although its existence is only instantiated in a possible world. To account for this puzzle Roberts (1989) stated that propositions in discourse are disposed in a logical hierarchy, where non-factual (i.e. modal) clauses are subordinated to factual ones. This theory is known as *modal subordination* and Roberts developed it under the framework of Discourse Representation Theory (Kamp & Reyle, 1993). Under this framework the entities of the discourse are depicted in boxes that display hierarchical relations (i.e. a subordinate box is contained in the main one). Thus the problem in (23b) is that the descriptive content of a non-factual clause (i.e. "write a novel") is subordinated to the factual one, and the anaphoric content in the continuation of (23b) cannot trace back to its reference because it lays in a lower layer. The anaphor, however, must be either at the same level or in a subordinate one with respect to its antecedent. Thus the continuation of (23b) is deviant for structural reasons, namely, the way in which discourse entities are represented in different hierarchies in the discourse structure.

Accounts such as modal subordination, developed in the field of formal semantics, provide an explicit explanation for the contrast displayed by sentences such as (23a) vs. (23b). Moreover, they may lead to predictions with respect to how our brain reacts to these anomalies. For instance, if the reading

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

of "ends" in (23b) generates an incongruence at the conceptual level alone, an index of semantic mismatch should be observed by contrasting "end" in (23b) to its occurrence in (23a). As previously noticed, in ERP studies an N400 is often found when the meaning of a word is difficult to integrate with the propositional or conversational context (cf. Kutas & Federmeier, 2000). Therefore it is likely that an N400 is found in case our brain processes (23b) like a "pure" semantic violation (e.g. "the Dutch trains are white/sour"). On the other hand, if the deviancy of (23b) is due to incongruence in the discourse structure, other psychophysiological effects might be elicited. In the ERP literature mismatch, complexity and repair mechanisms operating at the level of propositional or discourse structure often generate positivities starting at 300 ms, usually reaching their peak around 600 ms and then prolonging for more than 1 s after the critical word is hit. Such positivities may be generated in posterior sites (classic P600) or frontal ones (FP600). As we have already seen, syntactic anomalies often elicit the former effect, whereas anomalies and higher complexity at the discourse level have been found to give rise to the latter one.

Building on these considerations, Dwivedi, Phillips, Lague-Beauvais, and Baum (2006) investigated the processing of deviant sentences such as (23b), in contrast with (23a), where the mood of the second clause makes the sentence sound natural, with ERPs. In addition their subjects were presented with sentences of the form in (24) where the mood of the first clause was also manipulated (see (23) compared to (24)).

(24) a John is writing a novel. It could end quite abruptly.
b John is writing a novel. It ends quite abruptly.

Critically, the experimental condition was that in which the first clause was non-factual and the second one was factual (i.e. such as (23b)), with the other conditions serving as controls.

The first informative effect reported by the authors was a frontal positivity (FP600) in the time window between 300 and 1100 ms after the onset of the verb in the anomalous condition (i.e. "ends" in (23b)) vs. the controls. Then, at the pronoun position (*it* in the second clause of (23a and b)) the sentences where the first clause introduced a hypothetical context (i.e. (23a and b)) were more negative-going in left centro-parietal sites at 500 ms, as compared to the sentences with the first clause being factual (i.e. (24a and b)). The interpretation of the frontal positivity they found given by Dwivedi et al. was that it indexes a structural reinterpretation of the sentence/discourse structure. This ERP effect has already been reported by other studies employing other manipulations of discourse and sentential complexity (see frontal P600-like components, Kaan & Swaab, 2003; Friederici, Hahne, & Saddy, 2002; Bornkessel, Schlesewsky, & Friederici, 2002). The negativity found at the pronoun after the hypothetical contexts, on the other hand, was interpreted as a correlate of an increased cognitive load, due to the extra complexity - either of semantic or of syntactic nature - of the modal context containing the anaphoric antecedent.

This study provides a clear example of how structural relations related to language interpretation can play a role in giving rise to neuropsychological effects. Recall that we put some emphasis in the introduction of the current paper on the idea that symbolic semantic representations are grammar-driven. The case of mood structures, and their violations, extends this idea from the single proposition level towards the discourse one. The hypothesis that anomalies such those in (23b) may be reconducted to the illicit structure of modal representations fits the observed data quite nicely. In theoretical linguistics phenomena that concern how issues about structure (i.e. syntactic and discourse structure) interact with issues about meaning and interpretation are dubbed *syntax-semantics interface phenomena*. Crucially, any model of language processing and comprehension that denies grammar-driven properties of semantic representations (e.g. Gallese & Lakoff, 2005; Glenberg & Robertson, 1999; cf. the introduction of this paper) run into serious problems in explaining why sentences such as (23b) are deviant and why they elicit such ERP patterns.

Furthermore, some connections can be made between the results of this study and the one we discussed earlier on N-words processing (Vespignani et al., 2009). In both of these studies the FP600 is found in association when a certain interpretation of the critical item prompts a reanalysis or a modification of the propositional structure. In the experimental sentences employed by Dwivedi et al. modal frames have to be adjusted in order to rescue the sentence towards a meaningful interpretation.

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

In those from Vespignani et al.'s study a negation was to be added to the meaning of the proposition and the FP600 was elicited even in cases where this operation was allowed by the grammar.

On speculative grounds, these results suggest that syntactic and semantic processes can mutually interact in both directions. Such a hypothesis, be it empirically confirmed or not, is useful to have a grasp of the following idea. Although syntactic and semantic representations are tightly linked - whence the grammar-driven property of symbolic objects - they are not bound to operate in a bottom-up serial fashion. As semantic representations may be derived from syntactic structures, nothing prevents them from having a backward influence on syntactic structures. In other words, even if the structure is argued to guide the interpretation, the way we process language might well go on occasion in the other direction, namely in a top-down fashion.

Finally, similar claims to those we advanced in the previous sections can be maintained with respect to intensionality and grammar. That is, having a way to formalize intensional representations and operations is extremely useful in capturing meaning properties displayed by sentences where modal markers occur. For instance, we can capture the following property: any sentence identical to (23b) except from the substitution of "end" with another factual verb - i.e. a verb entailing the truth of its complement - will be deviant as well. It is not clear, nor it has been proposed, how to account for this generalization if we try to represent the meaning of (23a and b) through simulation mechanisms. How would a conceptual or perceptual frame representing (23a) differ from the one representing (23b)? Along the lines of modal subordination, as developed in DRT framework, hierarchical properties of propositions are made explicit. Furthermore such properties may interact with the other formal features of meaning we have discussed in the previous sections.

8. Formal Neurosemantics and its implications for models of sentence processing

In the introduction of the current work we anticipated that the goal of a neuroscientific theory of meaning is to know how representations and processes committed to comprehending and interpreting linguistic stimuli manifest themselves in the brain. Our main claim is that such a theory should incorporate representations and mechanisms that display certain properties, which have been defined precisely in the framework of formal semantics. Namely, such representations should be symbolic, abstract and grammar-driven.

Current models in formal semantics (see Chierchia & McConnell-Ginet, 2000, for an overview) infer from the observed linguistic behavior the underlying structures of meaning that are handled by our cognitive systems. The capacity of speakers to deal with such structures is referred to as semantic competence. An example of this has been discussed in the section dedicated to the interpretation of phrasal connectives (*and*, *or*) in their interaction with negation. The point we made in that section is that phrasal connectives are interpreted as logical operators, hence any theoretical account geared to model the meaning of connectives in natural language must be sensitive to the principles of first-order logic, otherwise it will end up predicting the wrong interpretations.

We may call this argument the logicality of language: its property of exhibiting logic like behavior, that is, its capacity to encode logical operations. The logicality of language presupposes the symbolic and abstract nature of semantic compositional representations. Propositions, data structures and other information bearing objects typically can be manipulated through operations such as conjunction ($\wedge$), disjunction ($\vee$), negation ($\neg$) and predication (boy(x))., that are related to each other in regular ways. As abstract, they are not strictly dependent on any sensorial modality or extra-linguistic representation, although their use in communicative exchange leads to the ultimate conceptual (hence memory- and sensory-based) activation. Such logical operations can be represented (and studied) in many ways. For example, if you think of propositions as a set of situations in a logical space (see the example in Fig. 1), then operations on propositions can be represented in terms of simple theoretic functions like intersection ($\wedge = \cap$), union ($\vee = \cup$) and complementation (e.g., with universe U, $\neg$ P = U-P); properties can be modeled as functions from situations to sets, predication as set memebership, etc. One proviso must be stated at this point. Formalisms such as those of propositional logic, set theory, DRT and any kind of symbolic formalism, as they are often used in formal semantics, are just tools to capture certain relations between words and their meaning. We do not claim that somewhere in our brain there must be a neuronal population representing a particular set or a particular logical formula in

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

a particular notation. What we do claim is that our brain must be equipped with information structures and computational capabilities that are structurally similar to those of logic.

What distinguishes linguistic semantic objects from other kinds of formal entities (e.g. the ontology of an artificial language such as a programming language) in their link to natural language syntax. Their structural organization is determined by the syntactic structure of sentences: they are grammar-driven. We might use in this connection an old metaphor. Our brain should possess the properties, to some extent, of a computer compiler. Namely, it must be able somehow to take an arbitrary input (e.g. sounds, words) and produce an interpreted output (conceptual activation, actions etc.). In parallel, our semantic cognitive system can be thought of, to some extent, as a theorem prover. It should be able to extract the logical consequences of any (well formed) linguistic input and use such consequences in the pursuit of its goals.

If we are right, thus, any model of semantics that does not include computational devices parallel to those of logic is doomed to failure in accounting for how we speak and comprehend words and sentences. By the same token, any model of language architecture in the brain that cannot account for how such capacities come about has little hope of explaining how the observed linguistic behavior can result from neural computations.

Thus, unless convincing counterarguments will be brought against logicality of language, strong claims defending the inadequacy of ungrounded representations - such as those advanced by Glenberg and Robertson (1999) and Gallese and Lakoff (2005), which we presented in the introduction - turn out to be deeply mistaken.

From this point, a wide range of options open up as to the way semantic representations having the core properties outlined in this paper are implemented and processed by the brain. For instance, it could be the case that all kinds of abstract objects undergoing compositional operations, both of syntactic and semantic nature, are not clearly distinguishable with respect to the way our brain process them. An example of proposals entertaining this idea is, e.g., the one suggested by Kuperberg (2007) according to which both syntactic and semantic compositional operations fall within *combinatorial* processing streams, which are argued to affect only the P600 ERP component. Whereas the aspects of meaning based on semantic memory retrieval are claimed to affect the N400 component, and belong to a separate informational stream, which is thought to be working in parallel and in interaction with the combinatorial stream.

Alternatively, symbolic representations related to language interpretation might be implemented in broader conceptual structures, such as those capturing extra-linguistic knowledge, sensory-based concepts, memory-based representations and so forth (cf. for such kinds of proposals: Jackendoff, 2002; Hagoort et al., 2004; Hagoort & Van Berkum, 2007). Such a hypothesis is not incompatible with the existence of symbolic and abstract representations; instead it predicts that linguistic (i.e. symbolic and grammar-driven) and non-linguistic (i.e. conceptual, sensory- or memory-based) informational streams cannot be disentangled through neuroscientific investigation.

We reviewed some relevant works investigating the workings of semantic processes in the brain during sentence comprehension that point to the following conclusion. Symbolic, abstract and grammar-driven semantic representations are processed differently with respect to both syntactic structures and extra-linguistic conceptual representations. We can infer this from the fact that computational processes carried out by our brain operating on such representations elicit different neural activation patterns. Pylkkänen, Oliveri et al. (2009) and Pylkkänen, Martin et al. (2009), (cf. also Pylkkänen & McElree., 2007) found different activation for compositional semantic vs. world knowledge violations with MEG. These authors offer a very precise description of the activation they found in correspondence to semantic violations in the vmPFC, which took place at least 50 ms earlier than the activation of the LPC, suggesting that the former is likely to reflect the compositional processes apt to assemble the sentence constituents while the latter one reflects those mechanisms integrating the meaning with the conceptual/world knowledge. Several experimental studies employing ERP methodology provided converging evidence that NPI violations, which generate ungrammatical sentences, elicit N400 effects (Drenhaus et al., 2006; Vespignani et al., 2009; Saddy et al., 2004). As we discussed in detail, NPIs are words that require certain formal relations to be satisfied (i.e. a downward entailing licensing context) and when they occur in the wrong propositional context (e.g. an affirmative sentence) they are claimed to give rise to a semantic violation. Two implications from these studies are

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

critical for distinguishing syntactic vs. semantic processes. First, sentences containing illicit occurrences of NPIs do not seem to generate a sense of deviancy that is different from classic syntactic violations, as witnessed by speakers' intuitions. In spite of this, they do generate different neuropsychological patterns. Second, in Vespignani et al.'s study, which employed N-words - the counterparts of NPIs in Romance languages such as Italian (cf. Panizza, 2009), N400 and FP600 where elicited by all the experimental conditions compared to the controls. However, posterior P600 showed a selectivity to illicit occurrences of N-words, as they were only elicited in ungrammatical sentences. Posterior P600 appears therefore to be a good diagnostic tool for syntactic violations whereas anterior positivities seem to involve overload of processes at the interface between the structure (of the proposition or the discourse) and its interpretation.

The sensitivity of our brain to symbolic representations of meaning is useful to explain why NPI violations elicit N400-like components, but it also provides a straightforward explanation as to why modal structure violations elicit P600-like effects. Sentences such as "John is considering writing a novel. It ends quite abruptly" are argued to be anomalous because their discourse structure is ill-formed. Formal models have the capacity of indicating precisely why this is so. Namely, it is because the propositional frame (e.g. possible worlds, discourse representations, etc.) where the first clause is evaluated lays in the wrong structural level with respect to that in which the second clause is evaluated (cf. Dwivedi et al., 2006). This example is a clear case of close connection between syntactic structures (at the propositional and discourse level) and the semantic representation that may be derived from them.

Finally, formal models suggest that certain words, despite being high frequency items, likely impose particular demands as soon as they are encountered and processed. Negation and modal auxiliaries, for instance, introduce new objects in the phrasal structure and implicate additional operations concerning the sentential interpretation. The fact that very rapid ERP effects are elicited as rapidly as - or soon after - such words are read (see early negativites reported by Ludtke et al., 2008, with negation and by Dwivedi et al., 2006, with modal markers) provides support for this idea.

Another aspect of meaning we focused our attention on is when, during the online course of sentence interpretation, our brain is sensitive to the truth value of a proposition. Earlier works investigating this issue (cf. Kounios & Holcomb, 1992) suggested that truth-value effects are revealed after the sentence is initially read and processed. Recent works (Nieuwland & Kuperberg, 2008), in contrast, provide some evidence that this is not always the case. Pragmatic plausibility of propositional content appears to quicken the online interpretation of the sentence, giving rise to the truth-value effects regardless of the presence of negative operators, which were previously argued to constitute an obstacle against a fast computation of the sentence meaning.

The fact that extra-linguistic sources of information such as context, knowledge, sensory-based representations, emotions, expectations, intentions and so forth can immediately influence the interpretation of a given sentence has posed challenges to the modular view of language architecture. In psycholinguistics, recent works (cf. Kuperberg, 2007; Hagoort & Van Berkum, 2007) highlighted the idea that a view of language processing centered on syntax is seriously problematic. Whereas such proposals do not deny the importance of grammatical syntactic structures, they do not provide any explicit account for how symbolic representations, sensitive to logical principles, drive the sentential interpretation. Hagoort and Van Berkum (2007), for instance, argue against the traditional distinction between *context-free* rule-based combination of fixed word meaning - often dubbed 'sentence meaning', derived from compositional combination of constituents - and the pragmatic contribution of communicative context to the interpretation. They reviewed some relevant works demonstrating that speaker's knowledge, world knowledge and co-speech gestures are immediately taken into account while constructing the meaning of a sentence. In parallel, they argue against a two-step model of language comprehension where first the constituents are interpreted and, second, they are evaluated against the world knowledge.

As already pointed out, we defend, to some extent, such a distinction between linguistically based symbolic and abstract semantic representations vs. extra-linguistic, memory-based, conceptual representations, which are not necessarily symbolic, abstract and grammar-driven. In this work we presented some evidence that this distinction is motivated not only on theoretical grounds but also on empirical - neuroscientific - data. If this holds, thus, one critical problem must be addressed. Traditionally, the process of sentence comprehension was thought to unfold along a bottom-up direction. Once the

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

propositional structure is built, it is given a semantic interpretation which, in turn, is evaluated against - and enriched by - pragmatic and contextual knowledge. If we admit that context may indeed influence the first stages of this process, we are obliged to provide a reasonable account for how this can ever happen.

One plausible account preserves the idea that the interpretation process takes place in two separate stages. The first stage consists of the construction of a propositional derivation from the given linguistic string, and its outcome includes all the possible derivations that are allowed by the grammatical rules. Then, if one of these derivations generates a meaning that is more plausible and informative with respect to the others, it will be favoured over the competing alternatives and it will end up being the intended meaning at the end of the day. Under this hypothesis, a cognitive computation of evaluating the plausibility of alternative derivations is required. Such an operation might be driven by lexical factors, extra-linguistic knowledge, sensory-based cues or representations stored in memory. Notice that this evaluation process does not have to be, in principle, the verification of the truth-value content of the proposition. Thus, according to this idea, understanding a sentence is tantamount to narrowing down all the possible hypothesis about its meaning to the ones displaying the highest likelihood. Notice that this might well happen before the listener heard the whole utterance, in that she might have exploited prediction processes to build a meaningful derivation prior to hearing the last part of the sentence. Such a model maintains that the first and the second step interact from the very first stages of sentence processing, without abandoning the idea of a functional and compositional dependency between formal syntactic and semantic representations.

An alternative account might be one that, in contrast, abandons completely the idea of two separate stages of the construction of sentence meaning. Syntactic and semantics structures, for instance, may be influenced by contextual factors while being built. It is well known that priming a syntactic structure affects people's comprehension (Scheepers & Crocker, 2004) as well as production (Bock, 1986; Griffin & Bock, 2000) of the following ambiguous sentences. Priming a structure increases its chance to be chosen among its competitors. Thus, priming a semantic structure may affect the likelihood of one interpretation versus other admissible ones. According to this idea symbolic representations of meaning may be enhanced or discouraged by conceptual and non-linguistic representations via a feedback mechanism channelled in the connections linking abstract symbols to conceptual domains and entities. On this view, formal representations could be actually affected by the context since the first moment of their instantiation.

Remarkably, both these hypotheses, drawn on purely speculative grounds, entertain a highly parallel and interactive model of sentence comprehension, which, nonetheless, allows the possible meanings of a given sentence to be still determined by the rules whereby syntactic structures are given an admissible interpretation by the semantics. Thus, formal theories of meaning remain crucial to predicting which derivations of a given sentence are allowed, and which ones are ruled out.

In conclusion, in the current work we aimed to present convincing evidence that our brain must implement certain representations necessary to understand language, which are symbolic, abstract and grammar-driven. In addition, we tried to describe the workings of the processes operating on this representation–in interaction with other informational streams related and unrelated to linguistic analysis–as observed by neuroscientific methodologies such as ERP, MEG and fMRI.

There are other interesting issues that are closely related to this topic. One is the question of the learnability of symbolic representations of meaning, which seem to be exploited by our cognitive system from the very early stages of language development (cf. Crain & Thornton, 1998, on this topic). Another one is the phylogenesis of symbolic capacities of humans, which hardly appears to be manifested by other species. As discussed in the introduction, some maintain the idea that semantic computations are performed by neural networks that house motor and sensorial cognitive systems and descend directly from primate homologue systems. We are not yet able to indicate exactly where and how in the brain such computations are performed. However, if we maintain that neural structures in secondary areas are inseparable in behaviour from the primary structures they are connected to (as claimed by Gallese & Lakoff, 2005), we might end up claiming that ungrounded - symbolic and abstract - computations are in fact carried out by motor and sensorial areas. Even on this rather extreme scenario, the existence of such representations and computations would still be necessary to explain

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

how we speak and interpret language, and critically, to explain the neuropsychological effects examined in this review.

Acknowledgments

This paper was inspired by the contribution to the workshop "Is a neural theory of language possible? Development of unified representations in natural and artificial systems" (Lecce, June 2007) given by Gennaro Chierchia. I want to thank him for great help, suggestions and insights he provides while I was working on this manuscript. I'm also very grateful to an anonymous reviewer, who provided extremely useful comments and critiques that helped me improve this manuscript, and Kristie Fischer, Francesco-Alessio Ursini and, above all, Svitlana Antonyuk-Yudina, who helped me to review the form and the content of this work.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.
Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences, 7*, 84–92.
Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.
Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Beyond syntax: language-related positivities reflect the revision of hierarchies. *NeuroReport, 13*, 361–364.
Bremmer, F., Schlack, A., Jon Shah, N., Zafiris, O., Kubischik, M., Hoffmann, K. P., et al. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalences between humans and monkeys. *Neuron, 29*, 287–296.
Brennan, J., & Pylkkänen, L. (2008). Processing events: behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language, 106*, 132–143.
Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience, 13*, 400–404.
Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*, 77–126.
Carnap, R. (1947). *Meaning and Necessity: A study in semantics and modal logic*. Chicago: University of Chicago Press.
Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychological Review, 82*, 45–73.
Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond*. Oxford: Oxford University Press.
Chierchia, G. (2006). Broaden your views. Implicatures of domain widening and the "Logicality" of language. *Linguistic Inquiry, 37*(4), 535–590.
Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and grammar: An introduction to semantics* (2nd ed.). Cambridge, Mass: MIT Press.
Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.
Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240–247.
Crain, S., & Thornton, R. (1998). *Investigations in universal grammar: A guide to experiments in the acquisition of syntax and semantics*. Cambridge, MA: The MIT Press.
Davidson, D. (1967). Truth and meaning. *Synthese, 17*(1), 304–323.
Dowty, D. (1979). *Word meaning and montague grammar*. Dordrecht, Netherlands: Reidel.
Drenhaus, H., Blaszczak, J., & Schütte, J. (2007). Some psycholinguistic comments on NPI licensing. In: E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung 11* (pp. 180–193). Barcelona: Universitat Pompeu Fabra.
Drenhaus, H., Graben, P., Saddy, D., & Frisch, S. (2006). Diagnosis and repair of negative polarity constructions in the light of symbolic resonance analysis. *Brain and Language, 96*, 255–268.
Dwivedi, V., Phillips, N. A., Lague-Beauvais, M., & Baum, S. (2006). An electrophysiological investigation of mood, modal context and anaphora. *Brain Research, 1117*, 135–153.
Edelman, G. M. (1992). *Bright air, brilliant fire*. New York: Penguin.
Fadiga, L., & Craighero, L. (2006). Hand actions and speech representation in Broca's area. *Cortex, 42*(4), 486–490.
Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience, 15*, 399–402.
Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language, 89*, 385–392.
Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology, 20*, 400–409.
Fodor, J. A. (1983). *The modularity of mind*. Cambridge MA: MIT Press.
Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31*, 45–63.
Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the sensory-motor system in reason and language. *Cognitive Neuropsychology, 22*, 455–479.
Geschwind, N. (1970). The organization of language and the brain. *Science, 170*, 940–944.
Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565.

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes, 28*, 1–26.
Griffin, Z. M., & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274–279.
Grodzinsky, Y., & Amunts, K. (Eds.). (2006). *Broca's region*. New York: Oxford University Press.
Grodzinsky, Y., & Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Sciences, 12.12*, 474–480.
Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes, 20*, 667–696.
Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift as an ERP-measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.
Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438–441.
Hagoort, P., & Van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B, 362*, 801–811.
Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*, 194–205.
Horn, L. R. (1989). *A natural history of negation*. Chicago, Ill: University of Chicago Press.
Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.
Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination and reason*. Chicago: University of Chicago Press.
Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98–110.
Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159–201.
Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to model-theoretic semantics of natural language, formal logic and discourse representation theory*. Boston: Kluwer Academic.
Karttunen, L. (1976). Discourse referents. In J. McCawley (Ed.), *Syntax and semantics, Vol. 7* (pp. 363–385). New York: Academic Press.
Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: the role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43*(3), 508–529.
Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 439–446.
Klima, E. (1964). Negation in English. In J. Fodor, & J. Katz (Eds.), *The structure of language*. Englewood Cliffs: Prentice Hall.
Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General, 121*, 459–479.
Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguistic Analysis, 25*, 209–257.
Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research, 1146*, 23–49.
Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203–205.
Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.). New York, NY: Elsevier.
Ladusaw, B. (1979). *Polarity sensitivity as inherent scope relations*. Austin: University of Texas.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.
Lau, E. F., Poeppel, D., & Phillips, C. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, .
Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.
Ludtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008). ERP correlates of negation in a sentence-picture-verification paradigm. *Journal of Cognitive Neuroscience, 20*, 1355–1370.
MacWhinney, B. (1998). The emergence of language from embodiment. In B. MacWhinney (Ed.), *The emergence of language from embodiment*. Mahwah, NJ: Erlbaum.
Mahon, B. Z., & Caramazza, A. (2005). The orchestration of the sensory-motor systems: clues from neuropsychology. *Cognitive Neuropsychology, 22*, 480–494.
Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis & a new proposal for grounding conceptual content. *Journal of Physiology - Paris, 102*, 59–70.
Makris, N., Meyer, J. W., Bates, J. F., Yeterian, E. H., Kennedy, D. N., & Caviness, V. S. (1999). MRI-Based topographic parcellation of human cerebral white matter and nuclei II. Rationale and applications with systematics of cerebral connectivity. *Neuroimage, 9*(1), 18–45.
Montague, R. (1970). Universal grammar. *Theoria, 36*, 373–398.
Narayanan, S. (1997). KARMA: Knowledge-based active representations for metaphor and aspect. PhD dissertation, Berkeley: Computer Science Division, University of California (http://www.icsi.berkeley.edu/~snarayan/thesis.pdf).
Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth isn't too hard to handle: an event-related potential study on the pragmatics of negation. *Psychological Science, 19*, 1213–1218.
Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–804.
Panizza, D. (2009). The processing of N-words in Italian. In V. Moscati, & E. Servidio (Eds.), *University of Siena CISCL Working Papers #3: Proceedings of XXXV Incontro di Grammatica Generativa*. Cambridge, MA: MITWPL.
Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(6157), 585–589.
Pinker, S. (2002). *The blank state. The modern denial of human nature*. New York, U.S.A.: Viking Penguin.

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001

ARTICLE IN PRESS

Progovac, L. (1992). Negative polarity: a semantico-syntactic approach. *Lingua, 86*, 271–299.
Pulvermüller, F. (2005). Brain mechanisms linking language and action. Nature reviews. The embodied cognition hypothesis. *Neuroscience, 6*, 576–582.
Pylkkänen, L., & McElree, B. (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience, 19*, 1905–1921.
Pylkkänen, L., Martin, A. E., McElree, B., & Smart, A. (2009). The anterior midline field: coercion or decision making? *Brain and Language, 108*, 184–190.
Pylkkänen, L., Oliveri, B., & Smart, A. (2009). Semantics vs. world knowledge in prefrontal cortex. *Language and Cognitive Processes, 24*, 1313–1334.
Ratcliff, R., & McKoon, G. (1982). Speed and accuracy in the processing of false statements about semantic information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 16–36.
Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience, 21*, 188–194.
Rizzolatti, G., & Craighero, L. (2004). The Mirror-Neuron System. *Annual Review of Neuroscience, 27*, 169–192.
Roberts, C. (1989). Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy, 12*, 683–721.
Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience, 2*, 1131–1136.
Saddy, D., Drenhaus, H., & Frisch, S. (2004). Processing polarity items: contrastive licensing costs. *Brain and Language, 90*, 495–502.
Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: a visual-world study. In M. Carreiras, & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 167–185). New York: Psychology Press.
Tarski, A. (1935). *The concept of truth in formalized languages*. (Ms).
Vespignani, F., Panizza, D., Zandomeneghi, P., & Job, R. (2009). Never in the wrong place: an ERPs study of unlicensed NPIs in Italian. Poster presented at the 22nd annual CUNY conference on human sentence processing, Davis, California.
Xiang, M., Dillon, B. W., & Phillips, C. (2008). Illusory licencing effects across dependency types: ERP evidence. *Brain and Language*, .

Please cite this article in press as: Panizza, D., Formal Neurosemantics. Logic, meaning and composition in the Brain, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.11.001