Distributional semantics in linguistic and cognitive research

Alessandro Lenci

On croit encore aux idées, aux concepts,
on croit que le mots désignent des idées,
mais ce n'est pas forcément vrai,
peut-être n'y a-t-il pas vraiment d'idées,
peut-être n'y a-t-il réellement que des mots,
e le poids propre aux mots.
[…]
En nous, donc, il n'y aurait eu aucune idée,
aucune logique, aucune cohérence?

J. Littell, *Les Bienveillantes* [*]

The hypothesis that word co-occurrence statistics extracted from text corpora can provide a basis for semantic representations has been gaining growing attention both in computational linguistics and in cognitive science. The terms *distributional*, *context-theoretic*, *corpus-based* or *statistical* can all be used (almost interchangeably) to qualify a rich family of approaches to semantics that share a "usage-based" perspective on meaning, and assume that the statistical distribution of words in context plays a key role in characterizing their semantic behavior. Besides this common core, many differences exist depending on the specific mathematical and computational techniques, the type of semantic properties associated with text distributions, the definition of the linguistic context used to determine the combinatorial spaces of lexical items, etc. Yet, at a closer look, we may discover that the commonalities are more than we could expect *prima facie*, and that a general model of meaning can indeed be discerned behind the differences, a model that formulates specific hypotheses on the format of semantic representations, and on the way they are built and processed by the human mind.

Methods for computational analysis of word distributional properties have been developed both in computational linguistics and in psychology. Because of the different aims of each field, these lines of research have typically proceeded totally in a parallel fashion, often ignoring each other. The drawbacks of this situation are clear: many

Rivista di Linguistica 20.1 (2008), pp. 1-31 *(ricevuto nell'ottobre 2008)*

potential synergies are lost, together with the opportunity to carry out in-depth analyses of the real impact that distributional methods can have on semantic research, and to achieve a better understanding of the epistemological foundations of such methods, as well as of their explanatory effectiveness and limits.

The purpose of this special issue is to foster a critical debate on distributional semantics, by bringing together original contributions from leading computational linguists, lexical semanticists, psychologists and cognitive scientists. The general aim is to explore the implications of corpus-based computational methods for the study of meaning. Distributional approaches raise the twofold question of the extent to which lexical properties can be reduced to their combinatorial behavior, as represented by their syntagmatic distribution in texts, and of the causal role of the contexts in which words are produced and observed in determining the structure and organization of semantic representations at the cognitive level.

1. What is distributional semantics?

A major issue for the study of word meaning is to devise precise identity criteria for the semantic content of words. Indeed, following Quine's well-known precept "No entity without identity" (Quine 1969:23), we can not hope to soundly investigate lexical meaning unless we are able to specify under which conditions two words can be said to have the same meaning or – if we regard the notion of synonymy too strong – to be *semantically similar* [1]. Either explicitly or implicitly, semantic similarity has a crucial role in any linguistic and psychological investigation on meaning. Rich empirical evidence has been accrued on the way the degree of semantic similarity between words affect how they are processed or stored in the mental lexicon [2]. Moreover, when we base our linguistic generalizations on semantic paradigmatic classes of expressions, e.g. the class of verbs of movement or the class of abstract nouns, we rely on semantic similarity to identify the expressions belonging to the same class.

The hallmark of any model of distributional semantics is the assumption that the notion of semantic similarity, together with the other generalizations that are built upon it, can be defined in terms of linguistic distributions. This has come to be known as the *Distributional Hypothesis* (DH), which can be stated in the following way:

Distributional Hypothesis
The degree of semantic similarity between two linguistic expressions *A* and *B* is a function of the similarity of the linguistic contexts in which *A* and *B* can appear.

Therefore, according to the DH, *at least certain aspects of the meaning of lexical expressions depend on the distributional properties of such expressions, i.e. on the linguistic contexts in which they are observed*. If this is true, by inspecting a significant number of linguistic contexts representative of the distributional and combinatorial behavior of a given word, we may find evidence about (some of) its semantic properties. A key issue is how this functional dependence between word distributions and semantic constitution is made explicit and explained, i.e. whether we conceive it to be merely *correlational* or instead a truly *causal* relation. As shown in §3, the possible answers to this issue determine large differences within the field of distributional semantics, and dramatically change the implications that we may draw on the origin and format of semantic representations in the human mind. Indeed, stronger or weaker versions of the DH can be provided, and this variation can partly be explained on the grounds of its historical roots.

1.1. The mixed fortunes of distributional semantics

Nowadays, distributional semantics has gained "popularity" especially in computational linguistics and, to a lesser extent, in psychology. This often leads us to forget that actually its roots are firmly grounded in the linguistic tradition. In his contribution 'The distributional hypothesis' in this issue, Sahlgren correctly connects the DH with the analysis procedure advocated by post-bloomfieldian American structuralists, such as Charles Hockett, Martin Joos, George Trager, and especially Zellig Harris. Actually, the history of the DH begins outside the realm of semantics, and precisely in the context of Harris' proposal of *distributional analysis* as the cornerstone of linguistics as a scientific enterprise:

In both the phonologic and morphologic analyses the linguist first faces the problem of setting up relevant elements. To be relevant these elements must be set up on a distributional basis: *x* and *y* are included in the same element *A* if the distribution of *x* relative to the other elements *B*, *C*, etc. is in some sense the same as the distribution of *y*. Since this assumes that the other elements *B*, *C*, etc., are recognized at the time when the definition of *A* is determined, this operation can be carried out without some arbitrary point of departure only if it is carried out for all the elements simultaneously. The elements are thus determined relatively to each other, and on the basis of the distributional relations among them (Harris 1951:7).

As Goldsmith (2005) points out, Harris' distributional procedure was first introduced for phonemic analysis, and then turned into a general methodology to be applied to every linguistic level. The distributional procedure is regarded by Harris as a way for linguists to ground their analyses on firm methodological bases, and primarily to avoid any argument based on meaning as an identity criterion for linguistic elements. For instance, in a footnote to the quotation above, Harris argues that "Objection might be raised here to the effect that meaning considerations too, are involved in the determinations of elements"(Harris 1951:7, fn 4). He replies that it is instead meaning identity that can be explained on distributional grounds:

It may be presumed that any two morphemes *A* and *B* having different meanings, also differ somewhere in distribution: there are some environments in which one occurs and the other does not (*Ibidem*).

According to Harris, the semantic similarity between two words is, in fact, a function of the degree of the similarity of their "linguistic environments", i.e. of the degree to which they can occur in similar contexts. Therefore, the near-synonymy between *oculist* and *eye-doctor* depends on the possibility to use these words interchangeably in most linguistic contexts (Harris 1954:157). Harris inherits from Bloomfield the refusal of meaning as an *explanans* in linguistics. However, at the same time he reverses the direction of the methodological arrow, and claims that similarity in distributions should be instead taken as an *explanans* for meaning itself, and therefore used to build paradigmatic classes out of distributionally semantic similar linguistic expressions. While for Bloomfield meaning is doomed to remain well beyond the boundaries of linguistic research [3], Harris seems to accept the possibility that semantic analysis might also receive a solid empirical foundation by the distributional approach. Meaning can become part of the linguistic science, at least for those aspects that can be defined through the very same method with which any linguistic entity should be defined, namely distributional analysis procedures.

The progressive abandonment of the structuralist distributional methodology under the attack brought by generative linguistics determined a downplay of the interest in contextual distributions as a primary key for investigating meaning. Distributional methods in semantics have remained squeezed between the hammer of the "cognitive revolution" in linguistics, and the anvil represented by formal, model-theoretic semantics. The increasing emphasis in generative grammar towards I-language as the target for linguistic inquiry and the attention towards the internalized competence of an ideal speaker entailed that the *explanans* for linguistic structures – included semantic ones – had to be found in some cognitive principles, specifically in those governing the Universal Grammar in the language faculty, rather than in the distributional constraints of linguistic constructions. Moreover, corpus-based statistical distributions were totally dismissed as a reliable source of linguistic evidence, together with the rejection of probability as a suitable formal model for grammar description.

Outside generative grammar, cognitive linguistics has stressed a conceptualist view of semantics, according to which the meaning of a lexical expression is a particular conceptualization of an entity or situation it is able to evoke. The definition of conceptualization given by Langacker is illuminating to understand the particular view of semantics favored in cognitive linguistics:

> The term conceptualization is interpreted broadly as embracing any kind of mental experience. It subsumes (a) both established and novel conceptions; (b) not only abstract or intellectual "concepts" but also sensory, motor and emotive experience; and (c) full apprehension of the physical, social, cultural, and linguistic context (Langacker 1998:3).

Although the linguistic context appears as one of the ingredients of human conceptualization, the emphasis of cognitive semantics is on an intrinsically embodied conceptual representation of aspects of the world, grounded in action and perception systems. The distributional constraints to which linguistic constructions obey are intended to receive a functional explanation in terms of the principles governing our processes of conceptualizing the world. Therefore, it is embodied conceptualization to be the source of meaning and to explain linguistic distributions, rather than the other way round. We will come back to this issue in §3.

Logico-philosophical and formal models of language have always emphasized a denotational approach to semantics, within the tradition of model-theoretic and referential semantics of Gottlob Frege, Alfred Tarski, Rudolf Carnap, Donald Davidson, Richard Montague, among many others. The basic tenet of this view is best described by David Lewis's statement that "Semantics with no treatment of truth-conditions is not semantics" (Lewis 1972:169). Here the polemical target is any semantic analysis that tries to analyze meaning in terms of some combination of symbols that are claimed to stand for conceptual primitives. However, Lewis' statement also indirectly concerns distributional approaches to meaning, like the one advocated by Harris. Even if we admitted that distributional analysis tells us something interesting at all about language, this could not be said to be something about meaning, because grammars are "abstract semantic systems, whereby symbols are associated with aspects of the world" (*Ibidem*:170). *Mutatis mutandis*, this is coherent with the definition of semantics given by Charles Morris, as the study of the "relations of signs to the objects to which the signs are applicable" (Morris 1938:6). Therefore, their huge differences notwithstanding, both cognitive approaches and model-theoretic ones agree on refusing distributional semantics because meaning can not be explained in terms of language-internal word distributions, but needs to be anchored onto extra-linguistic entities, being them either conceptual representations in the speakers' mind or objects in the world.

Differently from its destiny in theoretical linguistics, the idea that the distributional analysis of linguistic contexts is the key to understand word meaning has been kept alive and has even flourished within the *corpus linguistics* tradition. This fact is best summarized by the well-known slogan by J.R. Firth "You shall know a word by the company it keeps" (Firth 1957:11). Indeed, in corpus linguistics there is hardly any need to motivate the DH as a methodological principle for semantic analysis. Rather to the contrary, this is often claimed to be the unique possible source of evidence for the exploration of meaning. This position is best represented by Kilgarriff (1997:112), who claims that "Where 'word senses' have a role to play in a scientific vocabulary, they are to be construed as abstractions over clusters of word usages". The distributional method is indeed common in lexicography, which keeps an unbreakable tie with corpus linguistics. Corpora and statistical methods to analyze the word behavior in contexts (e.g. concordances, association measures, etc.) are parts and parcels of the lexicographer's toolbox. Moreover, it is not a case that the fortune of distributional semantics in computational linguistics has almost coincided with the *neo-empiricist turn* that occurred in this discipline starting from the late '80s, and that has been characterized by the renaissance and, later on, the progressive predominance of corpus-based statistical methods for language processing. Within this new methodological paradigm, it has been natural to apply statistical methods not only to part-of-speech tagging or to syntactic parsing, but also to lexical semantic analysis.

In psychology, the DH finds one of its strongest and explicit assertions (under the name of *Contextual Hypothesis*) in the work by George Miller and Walter Charles, who argue for a "usage-based" characterization of semantic representations:

> What people know when they know a word is not how to recite its dictionary definition – they know how to use it (when to produce it and how to understand it) in everyday discourse [...]. Knowing how to use words is a basic component of knowing a language, and how that component is acquired is a central question for linguists and cognitive psychologists alike. The search for an answer can begin with the cogent assumption that people learn how to use words by observing how words are used. And because words are used together in phrases and sentences, this starting assumption directs attention immediately to the importance of context (Miller & Charles 1991:4).

It is impossible not to relate this view of meaning with what Wittgenstein claims in his *Philosophical Investigations*, i.e. that "the meaning of a word is its use in the language" (Wittgenstein 1953:20). Miller and Charles try to turn this general claim into a more operative "context-based" characterization of word meaning. In fact, they argue that repeated encounters of a word in the various linguistic contexts eventually determine the formation of a *contextual representation*, defined as follows:

> the cognitive representation of a word is some abstraction or generalization derived from the contexts that have been encountered. That is to say, a word's contextual representation is not itself a linguistic context, but is *an abstract cognitive structure that accumulates from encounters with the word in various (linguistic) contexts*. The information that it contains characterizes a class of contexts (*Ibidem*:5; the emphasis is mine).

The contextual representation proposed by Miller and Charles includes all "syntactic, semantic, pragmatic and stylistic conditions" (*Ibidem*) governing the use of a word. Although a direct parallel exists with Harris' distributional analysis, yet there is a crucial element of difference between the two positions, which is highly relevant for the present discussion. Harris is, in fact, a strong anti-psychologist, and his distributional analysis is meant as a scientific method for linguistic research, without any necessary entailment for cognitive analysis (which remains beyond Harris' goals as a linguist). Conversely, the DH for Miller and Charles is a specific assumption about the cognitive format and origin of semantic representations. It is a claim about the type of information they contain, and the way they are built by a learner.

The primary and most specific goal of Miller and Charles is to use contextual representations to provide an empirical characterization of semantic similarity, as a central notion in psycholinguistic research. In their view, semantic similarity (as a cognitive notion) is to be treated as a dependent variable of the contexts in which words are used, i.e. as a function of their contextual representations. This idea had already been explored by Rubinstein & Goodenough (1965), who carried out empirical investigations to test the hypothesis of a correlation between word synonymy and similarity of contexts, the latter being formalized as the proportion of words appearing in the contexts of the two synonyms.

As Miller and Charles also point out, the mechanisms leading to the formation of contextual semantic representations can be viewed as a specific instance of general cognitive associative mechanisms recording the statistical co-occurrence between stimuli. For instance, Jenkins (1954) says that "intraverbal connections arise in the same manner in which any skill sequences arises, through repetition, contiguity, differential reinforcement" (Jenkins 1954:112). This is also consistent with the importance assigned in neo-behaviorist psychology to word associations – i.e. the first word produced by a subject in response to a word stimulus – as evidence to analyze the organization of language. In turn, word associations are interpreted by Jenkins and others as the result of the relative distribution between stimulus and response in linguistic contexts, and the stimulus-response association strength is related to the distributional similarity between the two words (cf. Schulte im Walde & Melinger this issue, 'An in-depth look into the co-occurrence distribution of semantic associates'). The fortune of the DH in cognitive science was surely affected, both positively and negatively, by this strong correlation with associative learning. Actually, word associations have often been regarded as a mere epiphenomenal feature of lexical organization. This fact, together with the idea that simple associative mechanisms do not suffice to explain cognitive dynamics, has often produced a suspicious and skeptical, attitude towards the DH itself, on the grounds of the assumption that "real semantics" actually resides in the conceptual system, rather than in the associations that can be established between words observed in the same linguistic contexts. Conversely, the interest in the DH has stayed alive in those cognitive paradigms, like *Parallel Distributed Processing* (PDP) models, that have emphasized the importance of associative learning for language acquisition, together with the role of the statistical regularities extracted out of the linguistic input (Borovsky & Elman 2006).

In summary, the ups and downs of the DH as a methodological hypothesis to investigate meaning have strictly followed the swinging fortunes of empiricists approaches in linguistics and in cognitive science. The idea that context-based distributions can be used as a basis for semantic representations makes sense only within the perspective of admitting that the probabilistic analysis of linguistic distributions has a significant role for language organization and for its dynamics. However, even within an empiricist, "usage-based" perspective, the DH also has to defend itself from another major charge, i.e. that linguistic distributions can not be taken as an *explanans* for meaning, since this is to be searched in the processes governing our conceptualization of the world. We will come back to this issue in §3.

1.2. The essence of distributional semantics

At the outset of this paper, I have mentioned some terms that are commonly used to qualify the semantic approaches based on the DH: *distributional*, *context-theoretic*, *corpus-based* or *statistical*. We can also add the terms *vector semantics*, *word* or *semantic spaces*, and *geometrical models of meanings* (Widdows 2004), which are particularly popular in computational linguistics. After the brief history of the DH in the previous section, it is fair to say that the names that best represent this approach to meaning are *distributional* and *context-theoretic semantics*. In fact, the essence of such models resides in the idea that word meaning depends on the contexts in which words are used, and that at least parts of a word content can be characterized by its contextual representation, to be defined as an abstraction over the linguistic contexts in which a word is encountered. All the other terms emphasize important aspects that are surely relevant to characterize distributional semantic models, but still play a more contingent role.

Corpora are crucially connected to distributional semantics because, as repositories of linguistic usages, they represent the primary source of information to identify the word distributional properties. The role of corpora has also been enhanced by the current availability of huge collections of texts (up to billions of words), as well as of increasingly sophisticated computational linguistic techniques to automatically process them, and extract the relevant context feature to build distributional semantic representations. On the other hand, it is worth emphasizing that the relationship between textual data and context-based representations, although strong, should not be considered to be exclusive. Indeed, beyond what they call the "narrow" definition of context intended as linguistic "co-text", Miller and Charles also admit a broader definition of context, including linguistic and extralinguistic information alike. Therefore, the contextual hypothesis is not restricted by definition to features extracted from linguistic contexts alone, and contextual representations of words may well be imagined to contain a much wider array of information about the contexts and situations in which a word is used, e.g. information about the participants in the communicative situations, visual features of the entities populating the context, etc. We can therefore claim that the use of linguistic contexts in current distributional approaches should be taken just as an approximation of a wider notion of language usage. The main reason that justifies this approximation is that our computational techniques to extract and model features from linguistic contexts are much more advanced than our methods to extract features from the extralinguistic contexts (e.g. from images or from video-recordings). However, the fact that most methods for distributional semantics are currently applied only to data extracted from corpora does not preclude the integration of contextual information in the broader sense of the term, mixing linguistic and extralinguistic information in contextual representations. For instance, Hare et al. (2008) and Mathe et al. (2008) present computational models in which mathematical techniques similar, if not identical, to those currently adopted in corpus-based distributional semantics are applied to extract semantic information from word co-occurrences with other words as well as with particular image features. It is therefore possible to foresee a not far future in which context-theoretic semantics will be set free from its exclusive bonds with text corpora, and will be turned into a method to explore the interrelation between linguistic and extralinguistic information in feeding the word semantic content. This also shows that, their strong links notwithstanding, distributional semantics should not be automatically reduced to a kind of corpus linguistics. Indeed, the scope of contextual semantic representations is potentially broader than simply recording linguistic distributions. However, it is also worth emphasizing that, even if currently distributional semantics is indeed corpus-based, this does not prevent it to tackle crucial aspects of the format and origin of word meaning. In fact, we must not forget that language includes terms like *walk* or *dog* – whose meaning is unlikely to be learnt without having some specific sensory-motor experience with the objects and events to which they refer – side by side with terms like *understand*, *society*, *idea*, etc. whose semantic content we most probably acquire via linguistic experiences, i.e. by observing how they are used in language. As we will see in §3, applying purely corpus-based distributional methods allows us to tackle the crucial issue of how and to what extent feature extracted from the linguistic input shape meaning. Therefore, even the contingent restriction of distributional methods to language data may be turned into a source for interesting research questions about meaning.

Statistics is a key component of distributional semantics, because it is used to abstract the relevant features of the contexts that enter into forming the contextual representations of words. The contextual representation of a word like *book* is not simply the sum of all the contexts in which I have encountered *book*, not differently from the fact that my concept of book is not simply the sum of the book instances I have encountered in my life. Statistical analysis is, therefore, one of the main mathematical tools that are currently used to single out the most salient contextual features to characterize a word distributional behavior. Actually, other mathematical techniques are used as well. For instance, the mathematical backbone of *Latent Semantic Analsysis* (LSA, Landauer & Dumais 1997), one of the most popular models of distributional semantics, is *Singular Value Decomposition*, which is a well-known linear algebra technique, akin to *Principal Component Analysis*, to extract the most informative dimensions in a matrix of data. The very names *vector semantics*, *word* or *semantic spaces*, and *geometrical models of meanings* merely highlight specific mathematical techniques used to formalize the notion of contextual representation, rather than introducing any substantial element of novelty into distributional semantics. Vectors are indeed the most useful numerical data structure that can be used to formalize contextual representations: the sequence of numbers forming a vector encode the statistical association strength between a word and a certain context or distributional feature. Since n-dimensional vectors represent the coordinates of points in a n-dimensional space, if we associate a word with a contextual representation and we formalize the latter as a vector, we can also conceive words as points in a "distributional space", i.e. a space whose dimensions are provided by the relevant linguistic contexts, and in which the position of a word-vector is determined by its statistical distribution in each context. In turn, if we adopt the DH and assume that semantic distance between words is a function of their distributional similarity, we can interpret this distributional vector space as a *semantic space*, in which distances between points correspond to semantic distances between the corresponding words. By measuring the distances between the word vectors in this geometric space, we can then use such distances as proxies for the similarity relations between the words.

Mathematical and computational techniques are important ingredients of distributional semantics exactly because they allow us to turn the informal notion of contextual representation into empirically testable semantic models. However, at the same time, they should not hide the real core features of the semantic representations they contribute to design:

1. lexical semantic representations are inherently "*context-based*" and therefore "context-sensitive": the (linguistic) contexts in which words occur or are observed enter into their semantic constitution;
2. lexical semantic representations are inherently *distributed*, in the sense that meaning derives from the way a word interacts with the different contexts in which it appears, each encoded as a particular vector dimension. The semantic content of a word lies in its global distributional history, rather than in some specific set of semantic features or meaning components;
3. lexical representations are inherently *quantitative* and *gradual*. The meaning of a word is not represented through "conceptual symbols", but in terms of its statistical distribution in various linguistic contexts. Words will differ not only for the contexts in which they appear, but also for the salience of these contexts in characterizing their combinatorial behavior.

Conjunctively, these features of meaning make distributional semantic representations very different from those that are commonly used in lexical and formal semantics. The closest relatives of such type of representations are those used in connectionist models. For instance, Rogers et al. (2004) within the PDP paradigm propose a computational model of semantic memory in which conceptual representations are instantiated as points in the high-dimensional space defined by a neural network. As the authors themselves acknowledge, this approach is very similar to those pursued in distributional semantics, since they both "extract high-order co-occurrence statistics across stimulus events in the environment" (Rogers et al. 2004:232).

Whether this particular type of representation is actually adequate to explain meaning and to account for its dynamics is obviously an empirical issue. Here, I would just like to add a few words to dismiss a very common stereotype that establishes a simplistic equation between distributional semantics on the one hand, and anti-nativist stances in language and cognition on the other. Distributional models are indeed empiricists, since they claim that (at least parts of) word semantic properties depend on the way words are used, and therefore can be inductively derived from the statistical analysis of language data. However, anti-nativist positions are not a logical consequence of adopting a distributional view of meaning. Admittedly, many distributional models (e.g. Li et al. 2000 and Borovsky & Elman 2006) extract meaning features out of "raw" linguistic contexts, conceived as an unstructured window of tokens surrounding a given word, and use these computational models as an argument against the need to postulate innate principles governing language acquisition. On the other hand, computational distributional methods such as Padó & Lapata (2007) and Baroni & Lenci (2009) adopt a more "syntactically savvy" notion of linguistic contexts, in which semantic properties are reconstructed by analyzing the statistical distribution of words into specific syntactic configurations. These latter models are in principles not incompatible with the idea that there are syntactic *a priori* acting as the scaffoldings that guide distributional analysis. Therefore, the DH and nativist assumptions in cognition are not inconsistent at all. For instance, the "syntactic bootstrapping" approach to word acquisition (Lidz et al. 2003) argues that syntax guides young children's interpretations during verb learning, and that one important source of information lies in the systematic relationships between verb meaning and syntactic structure. This hypothesis, which is strongly reminiscent of the DH, is however typically explored within a general universalist and nativist position on the principles governing the mapping between syntax and lexical semantics [4].

2. The two souls of the Distributional Hypothesis

Because of its history and different roots, distributional semantics is a manifold program for semantic analysis, which is pursued in disciplines as different as computational linguistics and psychology. The goals of computational methods adopting the DH are equally various: thesaurus and lexicon building, word-sense disambiguation, terminology extraction, models for vocabulary learning, models of semantic priming and semantic deficits, etc. Given its broad scope of applications, it is actually important to distinguish between a *weak* and a *strong* version of the DH. These versions differ not so much for their basic methodology, but rather for the status that is assigned to contextual representations in the analysis of meaning. Both these sides of the DH are represented in the contributions in this issue.

2.1. The "weak" Distributional Hypothesis: exploring word meaning with distributional analysis

In its weak version, DH is a quantitative method for semantic analysis, akin to Harris' distributional procedures. Word distributions in contexts are investigated to identify the semantic paradigmatic properties of these expressions. The idea is that word meaning (whatever this might be) determines the combinatorial behavior of words in contexts, and the semantic features of lexical expressions act as constraints governing their syntagmatic behavior. Consequently, by inspecting a relevant number of distributional contexts, we may hope to be able to identify those aspects of meanings that are shared by words with similar contextual distributions and that can be used to explain these very distributions. It is worth emphasizing that under this version, the DH only assumes the existence of a *correlation* between semantic content and linguistic distributions, and exploits such correlation to get at a better understanding of the semantic behavior of lexical items. Assuming this weak version of the DH does not entail assuming that word distributions are themselves constitutive of the semantic properties of lexical items at a cognitive level. It rather corresponds to taking semantics as a kind of "latent variable" which is responsible for the linguistic distributions that we observe, and that we try to uncover by inspecting a significant number of such distributions.

The weak DH is, thus, compatible with many research programs in theoretical linguistics. For instance, Levin (1993) claims that the semantic properties of verb argument structures determine the way in which such structures are realized syntagmatically, and the array of possible syntactic alternations that a verb participate in. The so-called Levin's Classes of English verbs are indeed identified on a distributional basis, and the contexts are the possible alternations in which verbs are found. Verbs are grouped into equivalence classes depending on the types of alternations they share. Then, these classes are inspected to identify the common semantic elements that are shared by their elements, and that may be regarded as an explanation for that particular class (Levin & Rappaport Hovav 2005). Under this view, similarity in distribution is taken to be an overt consequence of some deep semantic property that explains it. Various corpus-based, computational versions of Levin's methodology have been proposed, such as for instance Merlo & Stevenson (2001) for English, Schulte im Walde (2006) for German and Lenci (to appear) for Italian. In all these cases, computational distributional analysis is used to find paradigmatic classes of verbs on the grounds of their corpus-based behavior.

The contributions by Fazly & Stevenson, by Pustejovsky & Jezek, and by Rumshisky in this issue all adhere to the weak version of the DH, and show how the distributional method can be used to analyze different lexical semantic phenomena. Fazly & Stevenson ('A distributional account of the semantics of multiword expressions') argue that various properties of multiword expressions, such as non-compositionality, rigidity, etc., can receive a distributional interpretation. In fact, corpus-based statistical methods have long been used to automatically identify multiword expressions (together with collocations, as their closest relatives). Now, Fazly & Stevenson prove that distributional indicators can also be used to discriminate between different semantic classes of multiword expressions. The results of their experiments show that the DH can successfully be applied to achieve a better understanding of the internal structure of that complex phenomenon represented by multiword expressions.

Both Pustejovsky & Jezek ('Semantic coercion in language: beyond distributional analysis') and Rumshisky ('Resolving polysemy in verbs: contextualized distributional approach to argument semantics') apply the DH within the theory of the *Generative Lexicon* (GL; Pustejovsky 1995, 2001), and provide interesting examples of the advantages of using the DH in close cooperation with a solid theory of lexical semantics. Pustejovsky & Jezek use distributional data to investigate various aspects of lexical coercion phenomena in English and Italian [5], while Rumshisky presents a method to discriminate the senses of polysemous verbs based on the notion of distributional similarity. These contributions succeed in showing that the combinatorial properties of lexemes can be used to give a more robust empirical foundation to various GL theoretical constructs, such as the *Qualia Structure*, the compositional operations proposed in Pustejovsky (1995) and Pustejovsky (2006), and the "dot-types" that represent regular polysemy alternations [6]. Even more importantly, this way we can falsify the quite common view that distributional semantics is a sort of "theory-free" way of looking at meaning. Although this might indeed be an option, yet it is not a necessary one. Actually, GL is able to provide interesting interpreting keys for distributional information, and further progress in semantic research will surely be achieved by enhancing the synergies between distributional methods and formal theoretical analysis.

2.2. The "strong" distributional hypothesis: the constitutive role of word distributions for semantic representations

In its strong version, the DH is a *cognitive hypothesis* about the form and origin of semantic representations, as countenanced by Miller and Charles. Repeated encounters with words in different linguistic contexts eventually lead to the formation of a contextual representation as an abstract characterization of the most significant contexts with which the word is used. Crucially, assuming the strong DH entails assuming that word distributions in context have a specific *causal role* in the formation of the semantic representation for that word. Under this version, the distributional behavior of a word in contexts is not only taken as a way to get at its semantic behavior, but indeed as a way to explain its semantic content at the cognitive level.

Distributional semantic representations have been used to model a variety of psychological phenomena such as similarity judgments, semantic and associative priming, semantic deficits, semantic memory deterioration, etc. (Rubinstein & Goodenough 1965, Miller & Charles 1991, Vigliocco et al. 2004, Rogers et al. 2004, Jones et al. 2006). Distributional techniques have also been applied to model child lexical development as a bootstrapping process in which lexical and grammatical categories are extracted from the statistical distributions in the adults' input (Li et al. 2004, Baroni et al. 2007). Some of the most influential models for distributional semantics, such as *Latent Semantic Analysis* (LSA; Landauer & Dumais 1997) and *Hyperspace Analogue to Language* (HAL; Burgess & Lund 1997) have in fact been developed for cognitive and psychological research, and have been claimed to represent cognitively plausible models for the way semantic representations are learnt extracting regular co-occurrence patterns from the linguistic input (cf. also the recent Landauer et al. 2007, *Handbook for Latent Semantic Analysis*, for various applications of distributional methods to cognitive research). Some contributions have also come from the connectionist perspective. For instance, Farkas & Li (2000) and Li et al. (2004) propose an incremental version of

HAL with a recurrent neural network trained with Hebbian learning. Borovsky & Elman (2006) also model word learning in a fairly incremental fashion, using as word representations the hidden layer activation vectors produced by a *Simple Recurrent Network*. The network is probed at different training epochs and its internal representations are evaluated under a gold standard ontology of semantic categories to monitor the progress in word learning. Although Borovsky & Elman (2006) claim to be able to simulate relevant aspects of child word learning, a major shortcoming of their work is represented by the fact that the training corpus is formed by simple artificial sentences, with a potential negative impact on the cognitive realism of the simulation. In fact, using naturalistic corpus data for training appears as a necessary condition to test distribution-based word learning under realistic conditions, avoiding any unwarranted abstraction from the noise and fragmentary character of adult-child interactions.

Although under different fashions, all these models adhere to the strong DH, and assume that the encounters of a word in different linguistic environments have an effect on the semantic representations of these words, e.g. on the similarity relationships that are established among them in the mental lexicon. The contributions by Baroni & Lenci, by Schulte im Walde & Melinger and by Onnis et al. in this issue also commit to a strong version of the DH. Baroni & Lenci ('Concepts and properties in word spaces') focus on the notion of *property* of a concept, such as for instance *being an animal* for *dog* or *having wheels* for *car*, and investigate whether distributional models are able to produce reasonable property-based descriptions of concepts, akin to those elicited from humans. Their result show important similarities between distributional models and human-generated properties, but also some striking differences, which also concern alternative implementations of the DH. For instance, there emerges a patent difficulty for distributional models to extract some types of properties as the parts of an object or its prototypical color (e.g. *being yellow* for *banana*). Thus, Baroni & Lenci point out some potentially critical aspect of distributional semantics, at the same time stressing the need for a careful evaluation and analysis of the type of semantic information that can reasonably be acquired from linguistic distributions.

Schulte im Walde & Melinger ('An in-depth look into the co-occurrence distribution of semantic associates') instead focus on word free semantic associations, namely words (e.g. *water*) that are called to mind in response to a given stimulus (e.g. *swim*). Specifically, Schulte im Walde & Melinger test the so-called *co-occurrence hypoth-*

esis, holding that semantic associations are related to the textual co-occurrence of the stimulus-response pairs. After collecting association norms for German verbs, the two authors show that the co-occurrence hypothesis (actually, an instance of the strong DH) is confirmed by their data, and investigate a large array of distributional variables concerning associate words, e.g. their part of speech, the distance from the stimulus word in linguistic contexts, etc. Their contribution also suggests that many issues still need to be explored on this topic, to get at a better understanding of the possible distributional roots of word associations, e.g. the different role of strong syntagmatic co-occurrence vs. paradigmatic associations.

Onnis et al. ('Generalizable distributional regularities aid fluent language processing: The case of semantic valence tendencies') argue that contextual distributional features are able to determine the positive or negative valence of a predicate. According to their hypothesis, the fact that a verb like *cause* tends to occur in English with negative events molds the semantic constitution of this verb and activates specific semantic expectations during sentence processing. In a series of experiments, they show that indeed semantic valence can be interpreted as a distributional phenomenon, with a direct impact on mechanisms of sentence processing, thereby adding considerable psychological reality to the information derived from the analysis of linguistic contexts. The work by Onnis et al. also shows that contextual representations like those advocated by Miller and Charles may also be the basis to explain more fine-grained aspects of word meaning, such as for instance connotative dimensions.

3. How semantic is distributional semantics?

The contextual representations advocated by Miller and Charles and the strong version of the DH commit to the cognitive plausibility of distributional semantic structures. Under this view, word meaning is, at least in part, determined by its combinatorial behavior in linguistic contexts. Although empirical evidence has shown that various semantic phenomena can, in fact, be modeled under the assumption that cognitive semantic representations have a distributional basis, yet the idea that meaning might consist of abstractions from purely linguistic co-occurrence patterns has raised many critiques (cf. also §1.1). Skeptics towards the cognitive plausibility of the DH and the possibility for it to provide an empirical foundation for semantics generally claim that whatever distributional information can tell us about a word, this can not be its meaning. The main reason for this negative attitude resides in the difficulty (or even sometimes impossibility) for distributional semantics to satisfactorily address core issues concerning semantic representations, above all *compositionality*, *inference* and *reference (grounding)*. The key question, obviously, is whether these weak points depend on contingent features of current models implementing the DH, or instead are inherently related to the structure of distributional representations, thereby impairing their explanatory adequacy for meaning and its dynamics. Indeed, the consequence drawn from the fact that distributional models are unable to tackle one or more of these aspects is usually that *distributional semantics is not semantics at all*. I will briefly go through each of these issues, before arguing that this conclusion is too strong and not fully justified.

3.1. Compositionality

It is commonly stated that the DH mainly concerns lexical meaning. However, it is also true that part of the meaning of a lexical expression consists in its ability to semantically compose with other linguistic expressions to form the meaning of a complex linguistic structure. Compositionality lies at the heart of the scientific debate in formal semantics (Partee 1984), but it is actually an issue that any theory of lexical meaning must address, as also argued by Pustejovsky (1995). A model for semantic representations, like the one proposed by the DH, should therefore be able to explain how the meaning of a complex expression can be built from the meanings of its components, and at the same time should be able to model the semantic constraints governing the range of expressions with which a given lexeme can compose.

Its centrality notwithstanding, compositionality and the problems that it raises often remain out of the focus of mainstream distributional semantics. Moreover, it is still an open issue whether vector-based distributional representations are really able to provide a satisfactory account of the way meaning is built compositionally. Landauer & Dumais (1997) propose a simple model of semantic composition for distributional representations based on vector summation. Under this view, the meaning of the sentence *The dog bit the man* is also represented as a vector, obtained simply by summing the vectors associated to the lexical words *dog*, *bit* and *ball* (leaving aside issues concerning tense, etc.). However, this model is patently inadequate, because the sum is a commutative operation and therefore the sentence *The dog bit the man* turns out to be semantically equivalent to *The man* *bit the dog*. Recently, Erk & Padó (2008) and Mitchell & Lapata (2008) have shown that by adopting a more sophisticated model of vector composition and a more abstract representational structure for word distributions that incorporates syntactic dependences, it is possible to solve some of these problems, and clearly differentiate the two sentences using contextual representations. Kintsch (2001) also proposes a computational model based on LSA that is able to distinguish literal predications like *This fish is a shark*, from metaphorical ones, like *This lawyer is a shark*. These researches therefore show that, although many problems still remain to be solved, compositionality is indeed an aspect of meaning that can in principle be tackled with contextual representations, and actually this is an important issue in the agenda of distributional semantics nowadays (cf. also Jones & Mewhort 2007).

3.2. Lexical inference

If we know the meanings of the English words *buy*, *acquire*, *car* and *vehicle*, we are also able to recognize the validity of the following inferential relations:

(1) a. Google *bought* a new company → Google *acquired* a new company
b. John drives a *car* → John drives a *vehicle*

Moreover, knowing the meaning of these words also entails recognizing that these inferences are licensed on different grounds. In fact, (1a) is justified by the existence of a *paraphrase* or *synonymy* relation between *buy* and *acquire*, while (1b) depends on the fact that *vehicle* is a *hyperonym* of *car*. This difference is also responsible for the following semantic facts:

(2) a. Google *acquired* a new company → Google *bought* a new company
b. John drives a *vehicle* *→ John drives a *car*
c. John drives a *car* *→ John drives a *van*

In this case, (2a) is a sound inference because synonymy is a symmetric relation (i.e. if A is a synonym of B, then B is a synonym of A), while (2b) does not hold because hyperonymy is an asymmetric relation. Moreover, (2c) shows that synonymy and hyperonymy should also be distinguished from a third relation, *co-hyponymy*, linking *car* and *van* as both hyponyms of *vehicle*. Thus, although *buy* is semantically very similar to *acquire* on the one hand, and *car*, *vehicle* and *van* are all somehow semantically related, the notion of “semantic similarity” actually covers different types of semantic relations, each with different logical properties and each licensing different inferences.

The key point is that current implementations of the DH are not able to account for the different lexical inferences that words can or can not license, and more specifically they are not able to account for simple semantics facts such those in (1) and (2). The words in the examples above are surely very similar from the distributional point of view: e.g. *buy* and *acquire* share a high number of syntactic subjects and direct objects. In fact, state-of-the-art distributional models are able to account for such similarity relations from the distributional point of view. However, these same models can not represent the fact that these words are indeed similar, *but under very different semantic respects*. Distributional models only place words in a common semantic space depending on their contextual representations, and measure the distances among them to account for their semantic similarity. Moreover, the distance between the words in distributional semantic spaces is a symmetric relation: if a word *A* is close to *B* in the semantic space, then *B* is also close to *A*. Thus, for instance, word distance by itself is not enough to capture the fact that *car* and *vehicle* are semantically highly related, but through an asymmetric relation, i.e. hyperonymy. Therefore, the notion of distance in vector space can be at most a proxy for a general, unspecified notion of semantic similarity, but it can not account for the fact that the space of relations linking word meanings is a highly structured one. From this point of view, distributional models have still a descriptive adequacy that is far below the one of other semantic models, such for instance WordNet (Fellbaum 1998) and semantic networks in general (Collins & Quillian 1969).

There are attempts to design algorithms for distributional semantics that are able to identify the particular type of relation linking word meanings (cf. for instance Pantel & Pennacchiotti 2006), but these models are still far from being really satisfactory, let alone to achieve a good level of cognitive plausibility in accounting for even the simplest inferences licensed by word meaning. In fact, semantic inference still lies beyond the current capability of distributional semantics (Erk 2009). We may wonder whether this is something due to contingent shortcomings of actual models, or it is rather a sign of the explanatory limits of the DH and of contextual representations.

3.3. Reference and grounding

In § 1.1, we have seen that Miller and Charles argue for the notion of distributional contextual representation on the ground that knowing the meaning of a word is knowing *how to use it*. However, knowing how to use a word involves much more than knowing how to use it *linguistically,* i.e. its distributional constraints. This fact is well described by Marconi (1997:2):

> It seemed to me that to be able to use a word is, on the one hand, to have access to a network of connections between that word and other words and linguistic expressions: it is to know that cats are animals, that in order to arrive somewhere one has to move, that an illness is something one may be cured of, and so forth. On the other hand, to be able to use a word is to know how to map lexical items onto the world, that is to be capable of both *naming* (selecting the right word in response to a given object and circumstances), and *application* (selecting the right object and circumstance in response to a given word).

Marconi calls these two sides of our knowledge of word meaning respectively *inferential* and *referential competence*. Clearly, distributional semantic representations are at most able to account for our inferential competence (at least in part, given what I have said in the previous section), but surely lacks the possibility to address the aspects of word meaning concerning *reference* to the world. Indeed, we can read lots of pages about aardvarks, and thus forming a very rich distributional representation of this word, without acquiring any skill on how to recognize such an animal in a zoo or in a movie. Distributional representations thus fall short of explaining the referential aspects of word meaning.

This point is strictly related to the strong critiques raised to distributional models by *embodied cognition approaches* in cognitive sciences. The contribution by Glenberg & Metha in this issue ('Constraint on covariation:It's not meaning') perfectly describes this position, by denying to symbol covariation, i.e. distributional information, any causal role on meaning formation. This type of critique to distributional models was first expressed by Glenberg & Robertson (2000), but has largely dominated the scientific debate on semantic representations since then. Essentially, distributional contextual representations are refused as a cognitive plausible model of meaning for two reasons: first of all, they are regarded to be intrinsically *ungrounded* symbolic representations, since they only represent statistical distributional relations between symbols in the linguistic input. This is an incontrovertible fact, because a vector representing the contextual distribution of word in a text simply records the different encounters of this word with other linguistic expressions. In that sense, other differences aside, the distributional vector representing the semantic content of the word *car* is as symbolic as any other formal structure of "conceptual symbols". Consequently, distributional representations too would fall under the "symbol grounding problem" (Harnad 1990) or "Chinese Room"-like arguments [7] (Searle 1980) that are claimed to affect any type of symbolic representation, i.e. the fact that symbol covariation alone is unable to generate meaning.

The second reason for which the plausibility of distributional representations is refuted depends on a specific hypothesis about the nature of conceptual representations, i.e. that they are intrinsically *embodied* and *grounded in the sensory-motor systems*. In fact, the current debate on the cognitive role of distributional representations runs parallel to the broader debate between models of concepts that reduce their content to embodied sensory-motor information (*Embodied Cognition Hypothesis*, ECH) and models of concepts as symbolic and abstract entities (*Abstract Cognition Hypothesis*, ACH). There are indeed many versions of the ECH, but one of the most influential models is the notion of *perceptual symbol* proposed by Barsalou (1999). According to this view, concepts (and meanings as well) are not amodal, formal symbols, but rather inherently modal entities, represented in the same perceptual systems from which we acquire experience of their instances. Accordingly, knowing the meaning of the word *aardvark* implies being able to run an "embodied simulator" that re-enacts our perceptual experiences with aardvarks. More recently, many behavioral data and neuroscientific findings have also been adduced as a proof of the fact that conceptual content consists of sensory-motor experiences (Barsalou 2003, Rizzolatti & Craighero 2004, De Vega et al. 2008).

In summary, to the extent that contextual representations remain purely linguistic structures, they are unable to address referential aspects of meanings. If concepts are instead representations that are inherently grounded in our sensory-motor experience, then distributional contextual representations seem to lose their cognitive plausibility. Within the embodied cognition perspective, only our experience in the world can actually have a causal role in determining the meaning of the words that refer to it.

4. Distributional semantics is semantics, after all

Embodied cognition has challenged the claim that distributional structures can provide an empirically and epistemologically well-motivated basis for cognitive plausible semantic representations. But is this negative conclusion really justified? Is it right to assume that distributional semantics can not play any substantial role in a cognitive explanation of meaning? I would like to conclude this paper by saying that such negative stances are not really motivated.

First of all, although there is an important array of evidence supporting the ECH, the idea that concepts should be reduced to representations grounded in sensory modalities is not totally warranted. For instance, Mahon & Caramazza (2008) argue that the same neurophysiological and behavioral evidence can be explained by a symbolic model of concepts, provided that we add to it "spreading activation" effects between the conceptual level and sensory-motor systems. Notice that ACH is compatible with the DH, because concepts are modeled as abstract symbolic entities integrating different types of information, possibly also the one coming from the distributional analysis of linguistic contexts.

Secondly, symbolic models of meaning, like those to which distributional semantics is currently associated, does not imply that all semantics is to be reduced to symbol manipulation. As Shapiro (2008) claims, Searle's "Chinese Room" argument does not necessarily force us to abandon symbolic approaches. The hypothesis that all semantic content can be simply derived from symbol transformations is a "folly" that no serious symbolist would admit. One can well imagine symbolic models of concepts that admit a causal power of linguistic distributions in determining semantic content, without endorsing the extreme case presented by Searle's argument.

Thirdly, even if embodied cognition were right in claiming the key role of non-propositional thought and of semantic structures inherently grounded in sensorial modalities, this would not entail that embodied representations were the whole story. The importance of grounded concepts does not in itself preclude the fact that contextual representations of the sort proposed by Miller and Charles and extracted from linguistic distributions *also* play a key role in the processes leading to meaning formation. Coming back to Marconi's model, if on the one hand referential competence is lacking in distributional semantics, on the other hand cognitive mechanisms specialized in the distributional analysis of linguistic input might provide the basis to explain other, equally important aspects of meaning.

In fact, there is a growing trend in cognitive sciences to find a common ground in which embodied cognition and distributional approaches to meaning could eventually meet. Barsalou et al. (2008), for instance, propose a dual model of meaning in which embodied simulation is accompanied by processed based on linguistic word co-occurrences, in the spirit of the DH, although linguistic distributions are assigned just a superficial role, with "real" deep semantic processing being carried out at the level of embodied representations. Another dual model is proposed by Louwerse & Jeuniaux (2008), who claim that language comprehension is both embodied and symbolic. According to their symbol interdependency hypothesis, "symbols link to other symbols through higher-order relationships, and they also refer to objects in the real world" (Louwerse & Jeuniaux 2008:320). Semantic representations as mixtures of sensory-motor grounded features and distributional linguistic information are also advocated by Andrews et al. (2009).

It is then possible to conclude that a promising line of research may come from assuming that between embodied cognition and the DH actually exists a sort of *division of semantic labour*, and that the empirical problem rather lies on how to divide their respective contributions in constructing meaning. This division actually concerns multiple dimensions and levels:

1. *lexical categories and domains* - lexical categories may differ for the role of distributional information shaping their semantic behavior. The most prominent example is provided by abstract terms, for which it is highly plausible to assume a more central role of distributional learning;
2. *semantic dimensions* - various aspects of meanings may differently be related to distributional information. This is also suggested by Barsalou et al. (2008) who suggest that taxonomic properties of concepts (e.g. *being an animal* for a *lion*) may be mostly derived from linguistic input. Aspects of meaning that interface with syntax (e.g. argument structure) and morphology (e.g. categories such as animacy or telicity which often influence overt markings) may also be more dependent on linguistic distributional information;
3. *semantic processes* - various types of cognitive processes may be differently sensitive to embodied or distributional symbolic processes. It remains to be explored if deep semantic processes could

also involve linguistic distributional processing, differently from what Barsalou et al. (2008) argue.

Zellig Harris conceived distributional analysis as a method to establish linguistics as a scientific enterprise. Similarly, distributional semantics, with its various implementations, is a research program that adopts new mathematical and computational methods to investigate meaning in a scientific way. Current models of distributional semantics suffer of various shortcomings, but these do not suffice to dismiss the semantic information that can be extracted with distributional analyses, such as those envisaged by Harris. Psychological research has correctly emphasized the role of experience in shaping our semantic memory. Experience is usually considered to consist in sensory-motor information extracted from our perception and action in the world. However, we should not forget that *language is also part of our experience*, it is part of our acting in the world and of our trying to know the world. It is not implausible to imagine that, like every other experience of the world, the structures of the language to which we are exposed may contribute to shape our semantic competence. The extent to which this may happen, and, more generally, the extent to which linguistic categories may have a distributional basis, is again an empirical question. Distributional semantics can be a tool for linguistics to explore these issues and to provide new contributions to cognitive science.

Address of the Author:
Alessandro Lenci, Dipartimento di Linguistica "T. Bolelli", Università di Pisa, via Santa Maria 36, 56126 Pisa.
<alessandro.lenci@ilc.cnr.it>

Notes

[*] I would like to thank the Authors that have contributed to this Special Issue of the *Italian Journal of Linguistics*. I am also grateful to Giovanna Marotta for useful comments on this paper.

[1] Nowadays in cognitive science it is more customary to talk about semantic similarity, rather than full semantic identity. In fact, differently from synonymy, semantic similarity is a gradable notion, and thus more useful and extensively applicable to human judgments and language data. Moreover, it allows us to compare a word to many other words with respect to their relative degree of similarity, e.g. we can say that *dog* is more semantically similar to *cat* than to *bird*.

[2] For instance, much evidence comes from so-called *semantic priming* effects: the time needed to recognize a "target" word is affected by its seman-

tic similarity with a "prime" word presented to the subject shortly before the target.

[3] These quotations from *Language* do not leave much uncertainty on this point: "The statement of meaning is therefore the weak point in language-study, and will remain so until human knowledge advances very far beyond its present state" Bloomfield (1933:140); "the linguist cannot define meanings" (*Ibidem*:145).

[4] Interestingly, Gleitman (2002) explicitly acknowledges the debt that "syntactic bootstrapping" owes to Harris' distributional analysis.

[5] *Coercion* is a semantic type-shifting operation that various formal lexical theories (GL included) assume to account for the fact that the semantic type of an argument noun can be changed into the one requested by a predicate. For instance, in GL a verb like *deny* expresses a predicate whose internal argument is of type PROPOSITION: e.g. *The Government denied that Mr. Smith was a spy*. However, this verb can also occur in a sentence like *The Government denied the plot*, whose direct object noun does not typically refer to a proposition. To explain why a type-mismatch like this does not result in a semantic anomaly, GL assumes that the semantic type of *plot* can be coerced into a proposition to match the type requested by the predicate.

[6] For instance, in GL *book* is assigned the complex ("dot") semantic type *phys • info*, because it refers both to a concrete object, as in *Burning books is a terrible crime*, and to its information content, as in *I found this book boring*.

[7] The "Chinese Room" is a famous thought experiment proposed by Searle to argue that understanding can not be reduced to any form of pure symbol manipulation. Searle imagines to be locked in a room and to receive batches of Chinese characters. Crucially, he does not know Chinese and for him these characters are just meaningless symbols. Searle also receives "rules" (in English) on how to combine these characters, and how to respond to Chinese characters with other Chinese characters. Suppose that, after a certain amount of training, Searle has been able to learn to combine Chinese characters and to reply to Chinese characters in such a way that his answers are *de facto* indistinguishable from those given by Chinese native speakers. The key point in Searle's argument is that even in this case it would still be true that he *does not understand* Chinese. The "Chinese Room" experiment was conceived by Searle as an argument against so-called "Strong Artificial Intelligence (AI)", i.e. the claim that "the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states." (Searle 1980:417). More in general, the "Chinese Room" is typically used to argue against the idea that intelligence, meaning and understanding can be reduced to mere syntactic manipulation of formal symbols.

Bibliographical references

ANDREWS Mark, Gabriella VIGLIOCCO & David P. VINSON 2009. Integrating experiential and distributional data to learn semantic representations. *Psychological Review* CXVI/3.

BARONI Marco, Alessandro LENCI & Luca ONNIS 2007. ISA meets Lara: A fully incremental word space model for cognitively plausible simulations of semantic learning. *Proceedings of the ACL Workshop on Cognitive Aspects of Language Acquisition*. 49-56.

BARONI Marco & Alessandro LENCI 2009. One semantic memory, many semantic tasks. *Proceedings of the EACL Workshop on Geometrical Models of Natural Language Semantics*, Athens, 31st March.

BARSALOU Lawrence W. 1999. Perceptual symbols systems. *Behavioural and Brain Sciences* XXII. 577-609.

BARSALOU Lawrence W. 2003. Semantic simulation in the human conceptual system. *Language and Cognitive Processes* XVIII. 513-562.

BARSALOU Lawrence W., Ava SANTOS, W. Kyle SIMMONS & Christine D. WILSON 2008. Language and simulation in conceptual processing. In DE VEGA Manuel et al. (2008). 245-283.

BLOOMFIELD Leonard 1933. *Language*. New York: Henry Holt and Company.

BOROVSKY Arielle & Jeffrey L. ELMAN 2006. Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language* XXXIII. 759-790.

BURGESS Curt & Kevin LUND 1997. Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes* XII. 1-34.

COLLINS Allan M. & M. Ross QUILLIAN 1969. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviour* VIII/2. 240–248.

DE VEGA Manuel, Arthur M. GLENBERG & Arthur C. GRAESSER (eds.). 2008. *Symbols, Embodiment and Meaning*. Oxford, Oxford University Press.

ERK Katrin 2009. Supporting inferences in semantic space: representing words as regions. *Proceedings of IWCS 2009*.

ERK Katrin & Sebastian PADÓ 2008. A Structured Vector Space Model for Word Meaning in Context. *Proceedings of EMNLP 2008*.

FARKAS Igor & Ping LI 2001. A self-organizing neural network model of the acquisition of word meaning. *Proceedings of the 4th International Conference on Cognitive Modeling*.

FELLBAUM Christiane 1998. *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

FIRTH John R. 1957. *Papers in Linguistics*. London, Oxford University Press.

GLEITMAN Lila 2002. Verbs of a feather flock together II. In NEVIN Bruce (ed.). *The Legacy of Zellig Harris: Language and Information in the 21st Century*. Philadelphia: John Benjamins. 209-229.

GLENBERG Arthur M. & David A. ROBERTSON 2000. Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language* XLIII/3. 379-401.

GOLDSMITH John, 2005. Review to Nevin Bruce. The Legacy of Zellig Harris. *Language* LXXXI/3. 719-736.

HARE Jonathon, Sina SAMANGOOEI, Paul LEWIS & Mark NIXON 2008. Semantic spaces revisited: investigating the performance of auto-annotation and semantic retrieval using semantic spaces. *CIVR '08: The 2008 international Conference on Content-based Image and Video Retrieval*, Niagara Falls.

HARNAD Stevan 1990. The symbol grounding problem. *Physica D* XLII. 335-346.

HARRIS Zellig S. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

HARRIS Zellig S. 1954. Distributional structure. *Word*, X/2-3, 1954. 146-62 [reprinted in HARRIS Zellig S. 1970. *Papers in Structural and Transformational Linguistics*. Dordrecht: Reidel. 775-794].

JENKINS James J. 1954. Transitional organization: association techniques. In OSGOOD Charles E. & Thomas A. SEBEOK (eds.). *Psycholinguistics. A Survey of Theory and Research Problems*, Bloomington: Indiana University Press. 112-118.

JONES Michael N., Walter KINTSCH & Douglas J.K. MEWHORT 2006. High dimensional semantic space accounts of priming. *Journal of memory and Language* LV. 534-552.

JONES Michael N. & Douglas J.K. MEWHORT 2007. Representing word meaning and order information in a composite holographic lexicon. *Psychological Review* CXIV/1. 1-37.

KILGARRIFF Adam 1997. I don't believe in word senses. *Computers and the Humanities* XXXI. 91-113.

KINTSCH Walter 2001. Predication. *Cognitive Science* XXV. 173-202.

LANDAUER Thomas K. & Susan T. DUMAIS 1997. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review* CIV/2. 211-240.

LANDAUER Thomas K., Danielle S. MCNAMARA, Simon DENNIS & Walter KINTSCH (eds.). 2007. *Handbook of Latent Semantic Analysis*. Mahwah NJ, Lawrence Erlbaum.

LANGACKER Ronald 1998. Conceptualization, symbolization and grammar. In TOMASELLO Michael (ed.). *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ, Lawrence Erlbaum. 1-39.

LENCI Alessandro to appear. Argument alternations in Italian verbs: a computational study. *Atti del XLII Congresso Internazionale di Studi della Società di Linguistica Italiana*, Pisa.

LEVIN Beth 1993. *English Verb Classes and Alternations. A Preliminary Investigation*. Chicago: University of Chicago Press.

LEVIN Beth & Malka RAPPAPORT HOVAV 2005. *Argument Realization*. Cambridge: Cambridge University Press.

LEWIS David 1972. General semantics. In HARMAN Gilbert & Donald DAVIDSON (eds.). *Semantics of Natural Language*. Dordrecht: Reidel. 169-218.

LI Ping, Curt BURGESS & Kevin LUND 2000. The acquisition of word meaning through global lexical co-occurrences. *Proceedings of the 31st Child Language Research Forum*. 167-178.

LI Ping, Igor FARKAS & Brian MACWHINNEY 2004. Early lexical development in a self-organizing neural network. *Neural Networks* XVII. 1345-1362.

LIDZ Jeffrey, Henry GLEITMAN & Lila GLEITMAN 2003. Understanding how input matters: verb learning and the footprint of universal grammar. *Cognition* LXXXVII/3. 151-178.

LOUWERSE Max & Patrick JEUNIAUX 2008. Language comprehension is both embodied and symbolic. In DE VEGA Manuel et al. (2008). 309-326.

MAHON Bradford Z. & Alfonso CARAMAZZA 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology – Paris* CII. 59-70.

MARCONI Diego 1997. *Lexical Competence*. Cambridge, MA: MIT Press.

MATHE Stephan, Afsaneh FAZLY, Sven DICKINSON & Suzanne STEVENSON 2008. Learning the abstract motion semantics of verbs from captioned videos. *Proceedings of the 3rd International Workshop on Semantic Learning Applications in Multimedia (the new SLAM)*, Anchorage.

MERLO Paola & Suzanne STEVENSON 2001. Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics* XXVII/3. 373-408.

MILLER George A. & Walter G. CHARLES 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* VI. 1-28.

MITCHELL Jeff & Mirella LAPATA 2008. Vector-based models of semantic composition. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. 236-244.

MORRIS Charles W. 1938. Foundations of a theory of signs. *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press. 77-138.

PADÓ Sebastian & Mirella LAPATA 2007. Dependency-based construction of semantic space models. *Computational Linguistics* XXXIII/2. 161-199.

PANTEL Patrick & Marco PENNACCHIOTTI 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. *Proceedings of the Conference on Computational Linguistics / Association for Computational Linguistics (COLING/ACL)*. 113-120.

PARTEE Barbara H. 1984. Compositionality. In LANDMAN Fred & Frank VELTMAN (eds.). *Varieties of Formal Semantics*. Dordrecht: Foris. 281-311.

PUSTEJOVSKY James 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.

PUSTEJOVSKY James 2001. Type construction and the logic of concepts. In BOUILLON Pierette & Federica BUSA (eds.). *The Syntax of Word Meaning*. Cambridge: Cambridge University Press. 91-123.

PUSTEJOVSKY James 2006. Type theory and lexical decomposition. *Journal of Cognitive Science* VI. 39-76.

QUINE Willard V.O. 1969. *Ontological Relativity and other Essays*. New York: Columbia University Press.

RIZZOLATTI Giacomo & Laila CRAIGHERO 2004. The mirror-neuron system. *Annual Review of Neuroscience* XXVII. 169-192.

ROGERS Timothy T., Mathew A. LAMBON RALPH, Peter GARRARD, Sasha BOZEAT, James L. MCCLELLAND, John R. HODGES & Karalyn PATTERSON 2004. The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review* CXI/1. 205-235.

RUBINSTEIN Herbert & John B. GOODENOUGH 1965. Contextual correlates of synonymy. *Computational Linguistics* VIII. 627-633.

SCHULTE IM WALDE Sabine 2006. Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics* XXXII/2. 159-194.

SEARLE John 1980. Minds, brains and programs. *Behavioural and Brain Sciences* III. 417-424.

SHAPIRO Lewis 2008. Symbolism, embodied cognition, and the broader debate. In DE VEGA Manuel et al. (2008). 57-74.

VIGLIOCCO Gabriella, David P. VINSON, William LEWIS & Merrill F. GARRETT 2004. Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology* XLVIII. 422-488.

WIDDOWS Dominic 2004. *Geometry and Meaning*. Stanford, CA: CSLI.

WITTGENSTEIN Ludwig 1953. *Philosophical Investigations*. Oxford: Blackwell.