

Formal Distributional Semantics: Introduction to the Special Issue

Gemma Boleda*
University of Trento

Aurélie Herbelot**
University of Trento

Formal Semantics and Distributional Semantics are two very influential semantic frameworks in Computational Linguistics. Formal Semantics is based on a symbolic tradition and centered around the inferential properties of language. Distributional Semantics is statistical and data-driven, and focuses on aspects of meaning related to descriptive content. The two frameworks are complementary in their strengths, and this has motivated interest in combining them into an overarching semantic framework: a “Formal Distributional Semantics.” Given the fundamentally different natures of the two paradigms, however, building an integrative framework poses significant theoretical and engineering challenges. The present issue of Computational Linguistics advances the state of the art in Formal Distributional Semantics; this introductory article explains the motivation behind it and summarizes the contributions of previous work on the topic, providing the necessary background for the articles that follow.

1. Introduction

The 1960s and 1970s saw pioneering work in formal and computational semantics: On the formal side, Montague was writing his seminal work on the treatment of quantifiers (Montague 1974) and on the computational side, Spärck-Jones and colleagues were developing vector-based representations of the lexicon (Spärck-Jones 1967). At the same time, Cognitive Science and Artificial Intelligence were theorizing in which ways natural language understanding—seen as a broad, all-encompassing task—might be modeled and tested (Schank 1972; Winograd 1972). The experiments performed back then, however, suffered from a lack of resources (both in terms of data and computing capabilities) and thus only gave limited support to the developed theories.

Fifty years later, Formal Semantics (FS) and vectorial models of meaning—commonly referred to as “Distributional Semantics” (DS)—have made substantial progress. Large machine-readable corpora are available and computing power has grown exponentially. These developments, together with the advent of improved

* CIMEC (Università di Trento), Palazzo Fedrigotti, C.so Bettini 31, 38068 Rovereto, Italy.
E-mail: gemma.boleda@unitn.it.

** CIMEC (Università di Trento), Palazzo Fedrigotti, C.so Bettini 31, 38068 Rovereto, Italy.
E-mail: aurelie.herbelot@unitn.it.

Submission received: 20 June 2016; revised version received: 10 September 2016; accepted for publication: 15 September 2016.

doi:10.1162/COLI.a_00261

machine learning techniques (LeCun, Bengio, and Hinton 2015), have brought back the idea that Computational Linguistics should work on general language understanding, that is, on theories and models that account both for language use as a whole, and for the associated conceptual apparatus (Collobert and Weston 2008; Mikolov, Joulin, and Baroni 2015; Goodman, Tenenbaum, and Gerstenberg 2015; Erk 2016).

This special issue looks at this goal from the point of view of developing a semantic framework that, ideally, would encompass a wide range of the phenomena we might subsume under the term “understanding.” This framework, **Formal Distributional Semantics (FDS)**, takes up the challenge from a particular angle, which involves integrating Formal Semantics and Distributional Semantics in a theoretically and computationally sound fashion. To show why the integration is desirable, and, more generally speaking, what we mean by general understanding, let us consider the following discourse:

- (1) The new postdoc doesn’t work for Kim: she writes papers on semantics. And... uh... on those neural nets that have loops.

Processing the meaning of those sentences requires a number of complex linguistic abilities. Hearers must be able to retrieve the descriptive content of words such as *postdoc* or *write* (i.e., have knowledge of the properties of the involved individuals and events). They must also correctly process function words and structural markers, such as *the*, *and*, and bare singular/plural constructions to understand which entities are involved in the discourse: (a) a unique postdoc, (b) a person named Kim, (c) a (probably relatively small) number of entities that are papers, and (d) abstract concepts referred to generically (*semantics*, *neural nets*). They must carry out some form of composition over the constituents of each sentence, for example, shifting the meanings of *write* and *paper* towards ‘authoring’ and ‘scientific article.’ Finally, they must be prepared to perform a range of inferences, for instance: the new postdoc is knowledgeable about semantics and what is probably Recurrent Neural Networks; she cannot be working for Kim, who is a syntactician, but she perhaps works for Sandy, who does Machine Learning.

To this day, no single semantic framework has been proposed that would naturally cater to all of these phenomena. Instead, Formal and Distributional Semantics have focused on—and been extremely successful in—modeling particular aspects of meaning. Formal Semantics provides an account of the inferential properties of language and of compositionality based on the formalization of the relations between the distinct entities and events referred to in a linguistic expression (Montague 1974; Partee 2008). The various strands of FS also offer philosophically grounded theories of meaning. However, the framework struggles with descriptive content, despite the large amount of work done on lexical semantics and formal ontology (Dowty 1991; Pustejovsky 1995; Pinkal 1995; Guarino, Pribbenow, and Vieu 1996; Kennedy and McNally 2005; Roßdeutscher and Kamp 2010; Asher 2011, among others). This comes from the fact that, being focused on a particular type of logical entailment, FS naturally limits the type of phenomena that it covers, especially at the lexical level. This, in turn, affects its psychological plausibility. Distributional Semantics, on the other hand, has made good progress in modeling the descriptive content of linguistic expressions in a cognitively plausible way (Lund, Burgess, and Atchley 1995; Landauer and Dumais 1997), but faces serious difficulties with many of the phenomena that Formal Semantics excels at, such as quantification and logical inference.

Because of the complementary strengths of the two approaches, it has been suggested that much could be gained by developing an overarching framework (Coecke, Sadrzadeh, and Clark 2011; Beltagy et al. 2013; Erk 2013; Garrette, Erk, and Mooney 2014; Grefenstette 2013; Lewis and Steedman 2013; Baroni, Bernardi, and Zamparelli 2015). A Formal Distributional Semantics thus holds the promise of developing a more comprehensive model of meaning. However, given the fundamentally different natures of FS and DS, building an integrative framework poses theoretical and engineering challenges. This introductory article provides the necessary background to understand those challenges and to situate the articles that follow in the broader research context.

2. Formal Semantics

Formal Semantics is a broad term that covers a range of approaches to the study of meaning, from model-theoretic (Montague 1974; Partee 2008) to proof-theoretic semantics (Gentzen 1935). Although a comprehensive overview of those different strands is beyond the scope of this introduction, we will present here the various formal concepts that have been discussed as being desirable in the FDS context.

Formal Semantics is so-called because it has adopted some of the tools standardly used to describe formal languages: Notably, Montague proposed a way to express semantic composition with respect to a model using intensional logic (Montague 1974). Relying on a well-honed logical apparatus, FS has developed prominent models of many linguistic phenomena, from quantification to modality (see Dowty, Wall, and Peters 1981 and Cann 1993 for overviews). It caters to ontological matters (what there is in the world), reference to entities (how we talk *about* things), meaning at the higher constituent level (composition), interpretation at the sentential level (e.g., by giving propositions a truth value), and—crucially—sophisticated accounts of the logical inferences that can be drawn from a particular sentence.

One intuitive way to describe the world in a logic is via a **model**, usually provided in terms of sets (Dowty, Wall, and Peters 1981). For instance, a world with three postdocs will contain a set of three entities sharing the property of being a postdoc. This set may be a subset of the larger set of humans. Generally, models are provided for mini-worlds corresponding to a particular situation or “state-of-affairs,” as it would clearly be impractical to describe the world in its entirety, even with a limited vocabulary.

With a model at our disposal, we must explain how words come to relate to its elements. The two notions of **extension** and **intension** are key to this explanation. The **extension**, or denotation, of a linguistic expression are the entities it refers to in the model. For instance, the extension of *postdoc* is the set of all postdocs in the model under consideration. This simple correspondence is complicated by the fact that there is not a straightforward one-to-one relation between words of a language and sets of a model. Some expressions may refer to the same entity but have different semantic content. For instance, although *the new postdoc* and *the author of the neural net paper* may refer to the same entity in a particular universe of discourse, they encapsulate different properties (Frege 1892). Further, knowledge of the properties of a named entity does not imply knowledge of its extension, and vice versa (Jones 1911): It is possible to know that someone wrote a particular paper about neural networks and not be able to recognize them at a conference. The logical notion of **intension** contributes to solving this issue by providing a function mapping possible worlds to extensions. The availability of such a function allows us to posit worlds where two linguistic expressions have the same extension, and others where they do not, clearly separating extensions from linguistic expressions.

Beyond providing reference, natural languages are **compositional**. The meaning of a complex expression is derivable from the meaning of its parts in a systematic and productive way. A compositional formal semantics framework provides semantic representations of linguistic expressions in a logic, and rules for combining them. So for any expression, it is possible to identify a number of potentially distinct individuals (or sets thereof), and the relationships between them. For instance, the sentence *The new postdoc has written several articles* can be transformed (simplifying somewhat) into the following logical form:

$$(2) \quad \exists x, y [new(postdoc(x)) \wedge article^*(y) \wedge write(x, y)]$$

showing that for this state of affairs, the extension of the relevant subsets of postdocs and articles (x and y ; the $*$ sign is used to indicate a plurality) are linked through a writing relation.

A complete account of compositionality relies heavily on being able to interpret function words in the sentence. FS gives a sophisticated formalization of quantifiers (\exists in our example) that lets us select subsets of entities and assign them properties: For example, *some* entities in the set of articles, denoted by the variable y , are written by x . FS, in virtue of having access to a detailed model of a world, has generally been very successful in formalizing the meaning of logical operators, particularly quantifiers (including matters of plurality and genericity), negation, modals, and their application to entities and events.

Beyond its descriptive power, FS also has tools to *interpret* the meaning of words and sentences with respect to a particular world. In truth-theoretic semantics, the meaning of a sentence is a function from possible worlds to **truth values**. Obtaining the truth value of a proposition with respect to a given world relies on the notion of *satisfaction* (a predicate can be truthfully applied to a term if the corresponding property in a world applies to the referent of the term): In Tarski's words, the proposition *snow is white* is true if snow is white (Tarski 1944). In contrast to truth-theoretic approaches, probabilistic logic approaches assume that a speaker assigns a probability distribution to a set of possible worlds (Nilsson 1994; Pinkal 1995; van Benthem, Gerbrandy, and Kooi 2009; van Eijck and Lappin 2012).

Being able to give an interpretation to a sentence with respect to either truth or speaker belief is an essential part of explaining the *implicit* aspects of meaning, in particular **inference**. Consider the following example. If the set of postdocs is in the set of human beings, then the sentence *a postdoc is writing*, if true, will entail the truth of the sentence *a human is writing*. Speakers of a language routinely infer many facts from the explicit information that is given to them: This ability is in fact crucial to achieving communication efficiency. By logically relating parts of language to parts of a model, model theory ensures that inference is properly subsumed by the theory: If the set of postdocs is fully included in the set of humans, it follows from the definition of extension that we can refer to a set of postdocs as "humans." In practice, though, inference in model-theoretic semantics is intractable, as it often relies on a full search over the model. This is where proof theory helps, by providing classes of inference that are algorithmically decidable. For instance, whereas it is necessary to search through an entire model to find out what it knows about the concept *postdoc*, proof theory has this information readily stored in, for instance, a type (e.g., a postdoc may be cast as an individual entity which is a human, holds a doctorate, etc.). This decomposition of content words into formal

type representations allows for a range of functions to be directly applied to those types.

However, formal approaches fail to represent content words in all their richness (and by extension, the kind of inferences that can be made over lexical information). A formal ontology may tell us that the set of postdocs is fully included in the set of humans (model theory), or that the type *postdoc* encapsulates a logical relation to the type *human* in its definition (proof theory), but this falls short of giving us a full representation of the concept. For instance, imagine formalizing, in such a system, the distinction between near-synonyms such as *man/gentleman/chap/lad/guy/dude/bloke* (Edmonds and Hirst 2002; Boleda and Erk 2015). Although all these words refer to male humans, they are clearly not equivalent: For instance, *man* is a general, “neutral” word whereas *chap*, *lad*, and others have an informal connotation as well as links to particular varieties of English (e.g., British vs. American). This is not the type of information that can naturally be represented in either a model- or proof-theoretic structure.

But a more fundamental problem is that, by being a logical system geared towards explaining a particular type of inference, FS naturally limits what it encodes of human linguistic experience. For instance, analogical reasoning is outside of its remit, despite it being a core feature of the human predictive apparatus. Knowing a complex attribute of *postdoc* (e.g., that a postdoc is more likely to answer an e-mail at midnight on a Sunday than 8 am on a Monday) can warrant the belief that third-year Ph.D. students work on Sunday evenings. This belief does not derive from a strict operation over truth values, but is still a reasonable abductive inference to make.

In relation to this issue, it should also be clear that, at least in its simplest incarnation, model theory lacks cognitive plausibility: It is doubtful that speakers hold a detailed, permanent model of the world “in their head.” It is similarly doubtful that they all hold the *same* model of the world (see Labov [1978] on how two individuals might disagree on the extension of *cup* vs. *mug*; or Herbelot and Vecchi [2016] on quantificational disagreements). Still, people are able to talk to each other about a wide variety of topics, including some which they are not fully familiar with. In order to explain this, we must account for the way humans deal with partial or vague knowledge, inconsistencies, and uncertainties, and for the way that, in the first place, they acquire their semantic knowledge. Although some progress has been done on the formalization side—for instance, with tools such as supervaluation (Fine 1975), update semantics (Veltman 1996), and probabilistic models (Nilsson 1994)—much work remains to be done.

3. Distributional Semantics

Distributional Semantics (Turney and Pantel 2010; Clark 2012; Erk 2012) has a radically different view of language, based on the hypothesis that the meaning of a linguistic expression can be induced from the **contexts** in which it is used (Harris 1954; Firth 1957), because related expressions, such as *postdoc* and *student*, are used in similar contexts (*a poor -, the - struggled through the deadline*). In contrast with Formal Semantics, this provides an operational learning procedure for semantic representations that has been profitably used in computational semantics, and more broadly in Artificial Intelligence (Mikolov, Yih, and Zweig 2013, for instance) and Cognitive Science (Lund, Burgess, and Atchley 1995; Landauer and Dumais 1997, and subsequent work). The two key points that underlie the success of Distributional Semantics are (1) the fact that DS is able to acquire semantic representations directly from natural language data, and (2) the fact that those representations suit the properties of lexical or conceptual aspects of meaning,

thus accounting well for descriptive content both at the word level and in composition (as we will see next). Both aspects strengthen its cognitive plausibility.

In Distributional Semantics, the meaning representation for a given linguistic expression is a function of the contexts in which it occurs. Context can be defined in various ways; the most usual one is the linguistic environment in which a word appears (typically, simply the words surrounding the target word, but some approaches use more sophisticated linguistic representations encoding, e.g., syntactic relations; Padó and Lapata [2007]). Figure 1(a) shows an example. Recently, researchers have started exploring other modalities, using, for instance, visual and auditory information extracted from images and sound files (Feng and Lapata 2010; Bruni et al. 2012; Roller and Schulte Im Walde 2013; Kiela and Clark 2015; Lopopolo and van Miltenburg 2015).

Distributional representations are vectors (Figure 1(c)) or more complex algebraic objects such as matrices and tensors, where numerical values are abstractions on the contexts of use obtained from large amounts of natural language data (large corpora, image data sets, and so forth). The figure only shows values for two dimensions, but standard distributional representations range from a few dozen to hundreds of thousands of dimensions (cf. the dots in the figure). The semantic information is **distributed** across all the dimensions of the vector, and it is encoded in the form of **continuous** values, which allows for very rich and nuanced information to be expressed (Landauer and Dumais 1997; Baroni and Lenci 2010).

One of the key strengths of Distributional Semantics is its use of well-defined algebraic techniques to manipulate semantic representations, which yield useful information about the semantics of the involved expressions. For instance, the collection of words in a lexicon forms a vector space or *semantic space*, in which semantic relations can be modeled as geometric relations: In a typical semantic space, *postdoc* is near *student*, and far from less related words such as *wealth*, as visualized in Figure 2. The words (represented with two dimensions *dim1* and *dim2*; see Figure 2, left) can be plotted as vectors from the origin to their coordinates in the dimensions (Figure 2, right). The visually clear vector relationships can be quantified with standard measures such as cosine similarity, which ranges (for positive-valued vectors) between 0 and 1: The cosine similarity between *postdoc* and *student* in our example is 0.99, and that of *postdoc* and *wealth* is 0.37. The same techniques used for two dimensions work for any number of dimensions, and thus we can interpret a cosine of 0.99 for two 300-dimensional vectors as signalling that the vectors have very similar values along almost all the dimensions. Also note that distributional representations are naturally **graded**: Two vectors can be more or less similar, or similar in certain dimensions but not others. This

The **postdoc** shook her head.

Any grad student or **postdoc** he'd have would be a clonal copy of himself.

That tattoo is actually a custom-made cartoon generated by the **postdoc** who led this work.

During that **postdoc**, I didn't publish anything.

(a)

 $\implies 0.43679 \ 1.93841 \dots$

(b)

(c)

Figure 1

Distributional Semantics: The linguistic contexts in which an expression appears, for example, the words in the *postdoc* sentences in (a), are mapped to an algebraic representation (see the vector in (c)) through a function, represented by the arrow in (b). The resulting representations (in the example, the vector for *postdoc*) are thus abstractions over contexts of use. Examples adapted from the COCA corpus.

	dim1	dim2
postdoc	0.43679	1.93841
student	0.71038	1.76058
wealth	1.77337	0.00012

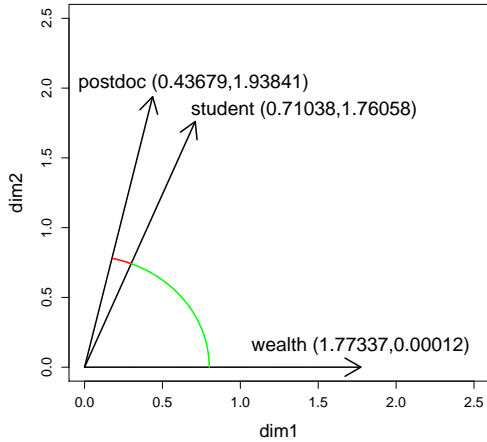


Figure 2 Semantic distance as geometric distance: A toy lexicon (left) and its representation in semantic space (right). The distance (see arcs) between the vectors for *postdoc* and *student* is smaller than the distance between the vectors for *postdoc* and *wealth*. The converse is true for similarity as opposed to distance.

is in accordance with what is known about conceptual knowledge and its interaction with language (Murphy 2004).

How are distributional representations obtained? There are in fact many different versions of the function that maps contexts into distributional representations (represented as an arrow in Figure 1(b)). Traditional distributional models are *count*-based (Baroni, Dinu, and Kruszewski 2014): They are statistics over the observed contexts of use, corresponding for instance to how many times words occur with other words (like *head*, *student*, *researcher*, etc.) in a given sentence.¹ More recent neural network-based models involve *predicting* the contexts instead (Collobert and Weston 2008; Socher et al. 2012; Mikolov, Yih, and Zweig 2013). In this type of approach, semantic representations are a by-product of solving a linguistic prediction task. Tasks that are general enough lead to general-purpose semantic representations; for instance, Mikolov, Yih, and Zweig (2013) used language modeling, the task of predicting words in a sentence (e.g., *After graduating, Barbara kept working in the same institute, this time as a .*). In a *predictive* setup, word vectors (called **embeddings** in the neural network literature) are typically initialized randomly, and iteratively refined as the model goes through the data and improves its predictions. Because similar words appear in similar contexts, they end up with similar embeddings; those are essentially part of the internal representation of the model for the prediction of a particular word. Although predictive models can outperform count models by a large margin (Baroni, Dinu, and Kruszewski 2014), those improvements can be replicated with finely tuned count models (Levy, Goldberg, and Dagan 2015).

Distributional models have been shown to reliably correlate with human judgments on semantic similarity, as well as a broad range of other linguistic and psycholinguistic phenomena (Lund, Burgess, and Atchley 1995; Landauer and Dumais 1997; Baroni and Lenci 2010; Erk, Padó, and Padó 2010). For instance, Baroni and Lenci (2010) explore synonymy (*puma-cougar*), noun categorization (*car IS-A vehicle; banana IS-A fruit*),

¹ Transformations are typically applied to counts; see Turney and Pantel (2010).

selectional preference (*eat topinambur* vs. **eat sympathy*), analogy (*mason* is to *stone* like *carpenter* is to *wood*), and relation classification (*exam-anxiety*: CAUSE-EFFECT), among others. Also, it has been shown that, at least at the coarse-grained level, distributional representations can model brain activation (Mitchell et al. 2008a). This success is partially explained by the fact that vector values correspond to how people use language, in the sense that they are abstractions over context of use.

Table 1 illustrates the kinds of meaning nuances that are captured in distributional models, showing how they reflect the semantics of some of the near-synonyms discussed in Section 2. The table shows the nearest neighbors of (or, words that are closest to) *man/chap/lad/dude/guy* in the distributional model of Baroni, Dinu, and Kruszewski (2014). The representations clearly capture the fact that *man* is a more general, neutral word whereas the others are more informal, as well as other aspects of their semantics, such as the fact that *lad* is usually used for younger people.

In recent years, distributional models have been extended to handle the semantic composition of words into phrases and longer constituents (Baroni and Zamparelli 2010; Mitchell and Lapata 2010; Socher et al. 2012), building on work that the Cognitive Science community had started earlier on (Foltz 1996; Kintsch 2001, among others). Although these models still do not account for the full range of composition phenomena that have been examined in Formal Semantics, they do encode relevant semantic information, as shown by their success in demanding semantic tasks such as predicting sentence similarity (Marelli et al. 2014). Compositional Distributional Semantics allows us to model semantic phenomena that are very challenging for Formal Semantics and more generally symbolic approaches, especially concerning content words. Consider polysemy: In the first three sentences in Figure 1(a), *postdoc* refers to human beings, whereas in the fourth it refers to an event. Composing *postdoc* with an adjective such as *tall* will highlight the human-related information in the noun vector, bringing it closer to *person*, whereas composing it with *long* will highlight its eventive dimensions, bringing it closer to *time* (Baroni and Zamparelli 2010, Boleda et al. 2013 as well as Asher et al. and Weir et al., this issue); crucially, in both cases, the information relating to research activities will be preserved. Note that in this way DS can account for polysemy without sense enumeration (Kilgarriff 1992; Pustejovsky 1995).

For all its successes at handling lexical semantics and composition of content words, however, DS has a hard time accounting for the semantic contribution of function words (despite efforts such as those in Grefenstette [2013], Hermann, Grefenstette, and Blunsom [2013], and Herbelot and Vecchi [2015]). The problem is that the kind of fuzzy, similarity-based representation that DS provides is very good at capturing conceptual or generic knowledge (postdocs do research, are people, are not likely to

Table 1

Near-synonyms in semantic space: The words closest to *man*, *chap*, *lad*, *dude*, and *guy* in the distributional model of Baroni, Dinu, and Kruszewski (2014), based on the UKWaC. Taken from Baroni (2016).

man	chap	lad	dude	guy
woman	bloke	boy	freakin'	bloke
gentleman	guy	bloke	woah	chap
gray-haired	lad	scouser	dorky	doofus
boy	fella	lass	dumbass	dude
person	man	youngster	stoopid	fella

be wealthy, and so on), but not at capturing episodic or referential information, such as whether there are one or two postdocs in a given situation. This information is of course crucial for humans—and for computational linguistic tasks. Applying distributional models to entailment tasks such as Recognizing Textual Entailment, for instance, can increase coverage, but it can lower precision (Beltagy et al. 2013). Entailment decisions require predicting whether an expression applies to the same referent or not, and these systems make trivial mistakes, such as deciding that *tall postdoc* and *short postdoc*, being distributionally similar, should co-refer.

Given the respective strengths and weaknesses of FS and DS, a natural next step is to work towards their integration. The next section describes the state of the research in this area.

4. Formal Distributional Semantics

The task of integrating formal and distributional approaches into a new framework, Formal Distributional Semantics (FDS), can be approached in various ways. Two major strands have emerged in the literature, which differ in what they regard as their basic meaning representation. One can be seen as Formal Semantics supplemented by rich lexical information (in the form of distributions). We will call this approach “F-first” FDS. The other puts distributional semantics center stage and attempts to reformulate phenomena traditionally dealt with by Formal Semantics in vector spaces (e.g., quantification, negation). We will refer to it as “D-first” FDS.

In the following, we provide an overview of both strands of FDS, and briefly discuss how they contribute to building a full semantics.

4.1 F-first FDS

In a F-first FDS, a classical formal semantics is chosen to represent the meaning of propositions. A popular choice is to posit a semantics expressible in first-order logic with a link to some notion of model (Garrette, Erk, and Mooney 2011; Beltagy et al. 2013; Lewis and Steedman 2013; Erk 2016; Beltagy et al., this issue). The availability of a model affords logical operations over individual entities, ranging from quantification to inference. It also gives a handle on matters of reference, by providing a notion of denotation—although *what* exactly words denote in a FDS is not entirely clear (see Section 4.3). Finally, the use of an existing logic allows the semantics to rely on the standard formal apparatus for composition, supplemented by techniques for integrating distributional information (Garrette, Erk, and Mooney 2011; Beltagy et al. 2013; see also this issue: Asher et al., Beltagy et al., Rimell et al., Weir et al.).

In any FDS, the core of the distributional component is a set of content word representations in a semantic space. The assumption is that those representations carry the *lexical* meaning of the sentence or constituent under consideration—and by extension, some useful world knowledge. In a F-first FDS specifically, the distributional component acts as a layer over logical forms, expressed as, for example, lexical rules or meaning postulates (Beltagy et al. 2013). This layer provides a way to compute similarity between lexical terms and thereby propose inferences based on likely paraphrases. Garrette, Erk, and Mooney (2011), for instance, use distributional similarity in inference rules to model a range of implications such as the following (from the original paper):

- (3) Ed has a convertible \models Ed owns a car.

One natural question F-first FDSs must solve is how exactly distributional vectors fit into the traditional formal semantic apparatus (as described in Section 2). Erk (2013) suggests that they might be seen as the intensional part of the semantics, and related to the notion of “concept.” This, however, calls for an explanation of the interaction between extension (modeled by the logic) and intension (modeled by the semantic space). A more recent proposal (Erk 2016) is to use distributional data as observable evidence for the likelihood of a given model, under the assumption that distributional similarity overlap correlates with denotational overlap. One interesting aspect of this work is that it addresses questions about the kind of semantic theory assumed by a particular framework (e.g., Erk [2016] that can be seen as a non-truth-theoretic, probabilistic semantics). We discuss this issue in more detail in Section 4.3.

4.2 D-first FDS

In a D-first FDS, logical operators are directly expressed as functions over distributions, bypassing existing logics. The solutions proposed for this task vary widely. Quantifiers, for instance, may take the form of (a) operations over tensors in a truth-functional logic (Grefenstette 2013); (b) a mapping function from a distributional semantics space to some heavily underspecified form of a “model-theoretic” space (Herbelot and Vecchi 2015); (c) a function from sentences in a discourse to other sentences in the same discourse (Capetola 2013) (i.e., for universal quantification, we map *Every color is good* to *Blue is good, Red is good...*). Other phenomena such as negation and relative pronouns have started receiving treatment: In Hermann, Grefenstette, and Blunsom (2013), negation is defined as the distributional complement of a term in a particular domain; similarly, Kruszewski et al. (this issue) show that distributional representations can model alternatives in “conversational” negation; a tensor-based interpretation of relative pronouns has been proposed in Clark, Coecke, and Sadrzadeh (2013) and is further explored in this issue (Rimell et al.). Some aspects of logical operators, in fact, can even be directly derived from their vectors: Baroni et al. (2012) show that, to a reasonable extent, entailment relations between a range of quantifiers can be learned from their distributions. Defining distributional logical functions in a way that accounts for all phenomena traditionally catered for by Formal Semantics is, however, a challenging research program that is still in its first stages.

D-first FDS also has to define new composition operators that act over word representations. Compositionality has been an area of active research in distributional semantics, and many composition functions have been suggested, ranging from simple vector addition (Kintsch 2001; Mitchell et al. 2008b) to matrix multiplication (where functional content words such as adjectives must be learned via machine learning techniques; Baroni and Zamparelli 2010, Paperno and Baroni 2016) to more complex operations (Socher et al. 2012). Some approaches to composition can actually be seen as sitting between the F-first and D-first approaches: In Coecke, Sadrzadeh, and Clark (2011) and Grefenstette and Sadrzadeh (2011), a CCG grammar is converted into a tensor-based logic relying on the direct composition of distributional representations.

In contrast with their F-first counterparts, D-first FDSs regard distributions as the primary building blocks of the sentence, which must undergo composition rules to get at the meaning of longer constituents (see above). This set-up does not clearly distinguish the lexical from the logical level. This has advantages in cases where the lexicon has a direct influence over the structure of a constituent: Boleda et al. (2013), for instance, show that there is no fundamental difference in the difficulty of modeling intersective vs. non-intersective adjectives in a D-first system.

At first glance, building sentences out of distributions seems to lead to a “conceptual” semantics (Erk 2013; McNally 2013) without a notion of entity (and therefore of set or model). Seen from another angle, however, distributional information can play a role in building entity representations that are richer than the set constituents of formal semantics. In the way that the distribution of *postdoc* is arguably a better representation of its meaning than the extension of *postdoc* (the set of postdocs in a world), there are reasons for wanting to develop rich distributional models of instances, and relate them back to a traditional idea of reference. Recent work has started focusing on this question: Gupta et al. (2015) show that vectors of single entities (cities and countries) can be linked to referential attributes, as expressed by a formal ontology; Herbelot (2015) gives a model for constructing entity vectors from concept-level distributions; Kruszewski, Paperno, and Baroni (2015) map distributions to “boolean” vectors in which each dimension roughly corresponds to an individual (noting that in practice, however, the induced dimensions cannot be so straightforwardly related to instances). Notably, the approach of Lewis and Steedman (2013), although F-first in nature, also relies on the distributional similarity of named entity vectors to build a type system.

4.3 The S in FDS

Although it is straightforward to highlight the complementary strengths of Formal and Distributional Semantics, it is not so trivial to define the result of their combination, either philosophically or linguistically. In fact, if semantics is synonymous with ‘theory of meaning,’ it is safe to say that Formal Distributional Semantics has not yet reached the status of a fully fledged semantics. Nor has, one should add, Distributional Semantics. This has prompted publications such as Erk’s “Towards a semantics for distributional representations” (Erk 2013), which query what aspect of meaning might be modeled by points in a vector space. The solution proposed in that paper involves linking both distributions and traditional intension to mental concepts, using an F-first FDS.

One problem faced by FDS is that not every strand of Formal Semantics is naturally suited to a combination with Distributional Semantics. Truth-theoretic approaches, in particular, come from a philosophical tradition that may be seen as incompatible with DS. For instance, the Tarskian theory of truth (Tarski 1944), which is foundational to Montagovian semantics, is ontologically neutral in that it does not specify what the *actual* world is like, but still commits itself metaphysically to the fact that sentences can be said to be true or false *about* a world. On the other hand, Distributional Semantics can be seen as having emerged from a particular take on Wittgenstein’s *Philosophical Investigations* (1953). The late Wittgenstein’s view of meaning is *not* anchored in metaphysics, that is, it makes no commitment with regard to what there is in the world (or any world, for that matter). This makes it non-trivial to define exactly what denotation and reference stand for in FDS, and by extension, to give a complete account of the notion of meaning.

Some recent approaches try to solve this issue by appealing to probabilistic semantics and the notion of a speaker’s “information state” (Erk 2016). In such attempts, denotation takes place over possible worlds: A speaker may not know the actual extension of a word such as *alligator*, but their information state allows them to assign probabilities to worlds in a way that, for instance, a world where alligators are animals is more likely than a world where this is not the case. The probabilities are derived from distributional information. By moving from “global” truth to the information states of specific speakers, the formalisms of model theory can be retained without committing

oneself to truth. Overall, however, the relationships between world, perception, beliefs, and language are still far from understood and call for further research in FDS.

5. Introduction to the Articles in this Special Issue

We started this introduction with an example that illustrated the many complex layers of semantics that are necessary to understand a text fragment. This special issue contains five articles that all tackle relevant phenomena, and demonstrate the complexity of the task at hand, advancing the state of the art towards a fully-fledged Formal Distributional Semantics.

The first article, by Kruszewski et al., presents a model which explains the semantic and inferential properties of *not* in “conversational” negations, for example, *this is not a dog* (compare *the new postdoc doesn't work for Kim*). It shows that such negations can be re-written as alternatives, supporting inferences such as *it is a wolf/fox/coyote* (cf. *the new postdoc works for Sandy*). Such alternatives seem to be well modeled by the nearest neighbors of distributional vectors, indicating a strong relationship between semantic relatedness and the notion of alternative. This work opens up a range of research topics on modeling broad discourse context, implicit knowledge, and pragmatic knowledge with Distributional Semantics.

Similarly focusing on a complex semantic phenomenon, Rimell et al. investigate relative clauses, associating definitional NPs such as *a person that a hotel accommodates* with concepts such as *traveller* (cf. *a neural net that has loops* vs. *Recurrent Neural Network*). Experimenting on a challenging data set, the authors show promising results using both simple vector addition and tensor calculus as composition methods, and provide qualitative analyses of the results. Although addition is a strong performer in their experiments given the data at their disposal, the authors discuss the need for more sophisticated composition methods, including tensor-based approaches, in order to go beyond current performance limits.

Weir et al. propose a general semantic composition mechanism that retains grammatical structure in distributional representations, testing their proposal empirically on phrase similarity. The authors see composition as a contextualization process, and pay particular consideration to the meaning shift of constituents in short phrases (e.g., *paper* in the context of *postdoc* and *write*). Having access to grammatical information in word distributions allows them to define a composition framework in terms of formally defined packed dependency trees that encapsulate lexical information. Their proposal promises to open a way to distributional representations of entire sentences.

Asher et al. present a proposal for combining a formal semantic framework, Type Composition Logic (Asher 2011), with Distributional Semantics. Specifically, types are recast in algebraic terms, including the addition of coercive or co-compositional functors that in the theory account for meaning shifts. Their system shows good performance on adjective–noun composition, in particular, in modeling meaning shift (such as *new* in the context of *postdoc* as opposed to other head nouns), subsectivity (vs. intersectivity), entailment, and semantic coherence.

To finish the issue, Beltagy et al. offer a description and evaluation of a full F-first FDS system, showing excellent results on a challenging entailment task. Their system learns weighted lexical rules (e.g., if *x* is an ogre, *x* has a certain likelihood of being grumpy), which can be used in complex entailment tasks (cf. if *x* writes papers about semantics, *x* is probably knowledgeable about semantics). Their entire framework, which involves not only a semantic parser and a distributional system, but also an

additional knowledge base containing WordNet entries, is a convincing example of the general need for semantic tool integration.

6. Directions for Future Research

FDS promises to give us a much better coverage of natural language than either Formal or Distributional Semantics. However, much remains to be done; here we address some prominent limitations of current approaches and propose directions for future research. The first is probably that little has been achieved so far in terms of accounting for discourse and dialogue phenomena. So, although the sentence level is actively under research, more needs to be done to integrate effects from larger text constituents as well as from the use of language in conversation (see Bernardi et al. [2015] for more discussion).

Secondly, FDS has so far primarily concentrated on modeling general conceptual knowledge and (to some extent) entities, with less attention to instances of events and situations. Future research should delve deeper into how distributional approaches can contribute to this aspect of the semantics.

Finally, despite the fact that Distributional Semantics is by nature anchored in context, the extent to which FDS treats phenomena at the semantics/pragmatics interface is still limited (but see Kruszewski et al., this issue). The notion of “context” explored so far includes the linguistic environment and specific aspects of perception (vision, sound), but fails to integrate, for instance, meaning variations linked to speaker intent or common ground. Generally speaking, if FDS wishes to retain a strong connection to the philosophical theories of “meaning as use,” it will have to expand into what other semantics might have left to the domain of pragmatics.

Acknowledgments

We thank Katrin Erk, Ann Copestake, Julian Michael, Guy Emerson, the FLOSS group, and two anonymous reviewers for fruitful feedback on this article. And we most heartily thank the authors and reviewers of the Special Issue for their hard work, as well as the journal editor, Paola Merlo, for her guidance and support. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no 655577 (LOVe) as well as the ERC 2011 Starting Independent Research Grant no. 283554 (COMPOSES).

References

Asher, Nicholas. 2011. *Lexical Meaning in Context: A Web of Words*. Cambridge University Press.

Baroni, Marco. 2016. Composes: An executive summary. Talk at the Composes Workshop at ESSLLI 2016.

Baroni, Marco, Raffaella Bernardi, Ngoc-Quynh Do, and Chung-chieh Shan. 2012. Entailment above the word level in distributional semantics. In *Proceedings of*

the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 23–32, Avignon.

- Baroni, Marco, Raffaella Bernardi, and Roberto Zamparelli. 2015. Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technology*, 9:5–110.
- Baroni, Marco, Georgiana Dinu, and Germán Kruszewski. 2014. Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014)*, pages 238–247, Baltimore, MD.
- Baroni, Marco and Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.
- Baroni, Marco and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP2010)*, pages 1183–1193, MIT, MA.

- Beltagy, Islam, Cuong Chau, Gemma Boleda, Dan Garrette, Katrin Erk, and Raymond Mooney. 2013. Montague meets Markov: Deep semantics with probabilistic logical form. In *Second Joint Conference on Lexical and Computational Semantics (*SEM2013)*, pages 11–21, Atlanta, GA.
- Bernardi, Raffaella, Gemma Boleda, Raquel Fernández, and Denis Paperno. 2015. Distributional semantics in use. In *Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics (LSDSem)*, page 95, Lisbon.
- Boleda, Gemma, Marco Baroni, The Nghia Pham, and Louise McNally. 2013. Intensionality was only alleged: On adjective-noun composition in distributinal semantics. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*, pages 35–46, Potsdam.
- Boleda, Gemma and Katrin Erk. 2015. Distributional semantic features as semantic primitives—or not. In *AAAI Spring Symposium on Knowledge Representation and Reasoning*, Stanford, CA.
- Bruni, Elia, Gemma Boleda, Marco Baroni, and Nam-Khanh Tran. 2012. Distributional semantics in technicolor. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL2012)*, pages 136–145, Jeju Island.
- Cann, Ronnie. 1993. *Formal Semantics*. Cambridge University Press.
- Capetola, Matthew. 2013. Towards universal quantification in distributional semantic space. In *Joint Symposium on Semantic Processing (JSSP2013)*, pages 75–79, Trento.
- Clark, Stephen. 2012. Vector Space Models of Lexical Meaning. In Shalom Lappin and Chris Fox, editors, *Handbook of Contemporary Semantics – second edition*, pages 493–522, Wiley-Blackwell.
- Clark, Stephen, Bob Coecke, and Mehrnoosh Sadrzadeh. 2013. The Frobenius Anatomy of Relative Pronouns. In *The 13th Meeting on the Mathematics of Language (MoL2013)*, pages 41–51, Sofia.
- Coecke, Bob, Mehrnoosh Sadrzadeh, and Stephen Clark. 2011. Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis: A Festschrift for Joachim Lambek*, 36(1–4):345–384.
- Collobert, Ronan and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine Learning*, pages 160–167, Helsinki.
- Dowty, David. 1991. Thematic Proto-Roles and Argument Selection. *Language*, 67(3):547–619.
- Dowty, David R., Robert Wall, and Stanley Peters. 1981. *Introduction to Montague Semantics*. Springer.
- Edmonds, Philip and Graeme Hirst. 2002. Near-Synonymy and Lexical Choice. *Computational Linguistics*, 28(2):105–144.
- Erk, Katrin. 2012. Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, 6(10):635–653.
- Erk, Katrin. 2013. Towards a semantics for distributional representations. In *Proceedings of the Tenth International Conference on Computational Semantics (IWCS2013)*, pages 1–11, Potsdam.
- Erk, Katrin. 2016. What do you know about an alligator when you know the company it keeps? *Semantics and Pragmatics*, 9:17–1.
- Erk, Katrin, Sebastian Padó, and Ulrike Padó. 2010. A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics*, 36(4):723–763.
- Feng, Yansong and Mirella Lapata. 2010. Visual information in semantic representation. In *Proceedings of the 2010 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT2010)*, pages 91–99, Los Angeles, CA.
- Fine, Kit. 1975. Vagueness, truth and logic. *Synthese*, 30(3):265–300.
- Firth, John Rupert. 1957. *A Synopsis of Linguistic Theory, 1930–1955*. Philological Society, Oxford.
- Foltz, Peter W. 1996. Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, & Computers*, 28(2):197–202.
- Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50.
- Garrette, Dan, Katrin Erk, and Raymond Mooney. 2011. Integrating logical representations with probabilistic information using Markov logic. In *Proceedings of the Ninth International Conference on Computational Semantics*, pages 105–114, Oxford.
- Garrette, Dan, Katrin Erk, and Raymond Mooney. 2014. A formal

- approach to linking logical form and vector-space lexical semantics. In Harry Bunt, Johan Bos, and Stephen Pulman, editors, *Computing Meaning*, volume 4, pages 27–48, Springer.
- Gentzen, Gerhard. 1935. Untersuchungen über das logische schließen. i. *Mathematische zeitschrift*, 39(1):176–210.
- Goodman, Noah D., Joshua B. Tenenbaum, and Tobias Gerstenberg. 2015. Concepts in a probabilistic language of thought. In *Concepts: New directions*, pages 623–653, MIT Press.
- Grefenstette, Edward. 2013. Towards a formal distributional semantics: Simulating logical calculi with tensors. In *Second Joint Conference on Lexical and Computational Semantics (*SEM)*, pages 1–10, Atlanta, GA.
- Grefenstette, Edward and Mehrnoosh Sadrzadeh. 2011. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404, Edinburgh.
- Guarino, Nicola, Simone Pribbenow, and Laure Vieu. 1996. Modeling parts and wholes. *Data & Knowledge Engineering*, 20(3):257–258.
- Gupta, Abhijeet, Gemma Boleda, Marco Baroni, and Sebastian Padó. 2015. Distributional vectors encode referential attributes. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*, pages 12–21, Lisbon.
- Harris, Zelig. 1954. Distributional structure. *Word*, 10(2-3):146–162.
- Herbelot, Aurélie. 2015. Mr. Darcy and Mr. Toad, gentlemen: distributional names and their kinds. In *Proceedings of the 11th International Conference on Computational Semantics (IWCS2015)*, pages 151–161, London.
- Herbelot, Aurélie and Eva Maria Vecchi. 2015. Building a shared world: Mapping distributional to model-theoretic semantic spaces. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*, pages 22–32, Lisbon.
- Herbelot, Aurélie and Eva Maria Vecchi. 2016. Many speakers, many worlds: Interannotator variations in the quantification of feature norms. *Linguistic Issues in Language Technology*, 13.
- Hermann, Karl Moritz, Edward Grefenstette, and Phil Blunsom. 2013. “Not not bad” is not “bad”: A distributional account of negation. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, pages 74–82, Sofia.
- Jones, Emily. 1911. *A New Law of Thought and Its Logical Bearings*. Cambridge University Press.
- Kennedy, Christopher and Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, pages 345–381.
- Kiela, Douwe and Stephen Clark. 2015. Multi-and cross-modal semantics beyond vision: Grounding in auditory perception. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*, pages 2461–2470, Lisbon.
- Kilgarriff, Adam. 1992. Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities*, 26(5-6):365–387.
- Kintsch, W. 2001. Predication. *Cognitive Science*, 25(2):173–202.
- Kruszewski, German, Denis Paperno, and Marco Baroni. 2015. Deriving Boolean structures from distributional vectors. *Transactions of the Association for Computational Linguistics*, 3:375–388.
- Labov, William. 1978. Denotational structure. In Donka Farkas, Wesley Jacobsen, and Karol Todrys, editors, *Parasession on the Lexicon*. Chicago Linguistics Society, pages 220–260.
- Landauer, Thomas K. and Susan T. Dumais. 1997. A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240.
- LeCun, Yann, Yoshua Bengio, and Geoffrey E. Hinton. 2015. Deep learning. *Nature*, 521:436–444.
- Levy, Omer, Yoav Goldberg, and Ido Dagan. 2015. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3:211–225.
- Lewis, Mike and Mark Steedman. 2013. Combined Distributional and Logical Semantics. *Transactions of the Association for Computational Linguistics*, 1:179–192.
- Lopopolo, Alessandro and Emiel van Miltenburg. 2015. Sound-based distributional models. *Proceedings of the 11th International Conference on Computational Semantics (IWCS2015)*, pages 70–75, London.

- Lund, Kevin, Curt Burgess, and Ruth Ann Atchley. 1995. Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, 17:660–665, Pittsburgh, PA.
- Marelli, Marco, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi, and Roberto Zamparelli. 2014. A sick cure for the evaluation of compositional distributional semantic models. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, pages 216–223, Reykjavik.
- McNally, Louise. 2013. Formal and distributional semantics: From romance to relationship. In Invited talk at the ‘Towards a Formal Distributional Semantics’ workshop (collocated with IWCS2013), Potsdam, Germany.
- Mikolov, Tomas, Armand Joulin, and Marco Baroni. 2015. A roadmap towards machine intelligence. arXiv preprint arXiv:1511.08130.
- Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL2013)*, pages 746–751, Atlanta, GA.
- Mitchell, Jeff and Mirella Lapata. 2010. Composition in Distributional Models of Semantics. *Cognitive Science*, 34(8):1388–1429.
- Mitchell, Tom M., Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. 2008a. Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191–1195.
- Mitchell, Tom M., Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. 2008b. Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191–1195.
- Montague, Richard. 1974. The proper treatment of quantification in ordinary English. In R. Thomason, editor, *Formal Philosophy*. Yale University Press, New Haven, CT, pages 247–270.
- Murphy, Gregory L. 2004. *The Big Book of Concepts*. The MIT Press, Cambridge, MA, 2nd edition.
- Nilsson, Nils J. 1994. Probabilistic logic revisited. *Artificial Intelligence*, 59(1-2):39–42.
- Padó, Sebastian and Mirella Lapata. 2007. Dependency-Based Construction of Semantic Space Models. *Computational Linguistics*, 33(2):161–199.
- Paperno, Denis and Marco Baroni. 2016. When the whole is less than the sum of its parts: How composition affects PMI values in distributional semantic vectors. *Computational Linguistics*, 42(2):345–350.
- Partee, Barbara H. 2008. *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee*. John Wiley & Sons.
- Pinkal, Manfred. 1995. *Logic and lexicon: The semantics of the indefinite*, volume 56. Springer Science & Business Media.
- Pustejovsky, James. 1995. *The Generative Lexicon*. MIT Press.
- Roller, Stephen and Sabine Schulte Im Walde. 2013. A multimodal LDA model integrating textual, cognitive and visual modalities. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP2013)*, pages 1146–1157, Seattle, WA.
- Roßdeutscher, Antje and Hans Kamp. 2010. Syntactic and semantic constraints in the formation and interpretation of *ung*-nouns. In A. Alexiadou and M. Rathert, editors, *The Semantics of Nominalizations Across Languages and Frameworks Interface Explorations 22*, pages 169–214, Mouton de Gruyter.
- Shank, Roger C. 1972. Conceptual dependency: A theory of natural language understanding. *Cognitive psychology*, 3(4):552–631.
- Socher, Richard, Brody Huval, Christopher Manning, and Andrew Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing (EMNLP2012)*, pages 1201–1211, Jeju Island.
- Spärck-Jones, Karen. 1967. A small semantic classification experiment using cooccurrence data. Technical Report ML 196, Cambridge Language Research Unit, Cambridge.
- Tarski, Alfred. 1944. The semantic conception of truth. *Philosophy and Phenomenological Research*, 4:341–375.
- Turney, Peter D. and Patrick Pantel. 2010. From frequency to meaning: Vector space

- models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188.
- van Benthem, Johan, Jelle Gerbrandy, and Barteld Kooi. 2009. Dynamic update with probabilities. *Studia Logica*, 93(1):67–96.
- van Eijck, Jan and Shalom Lappin. 2012. Probabilistic semantics for natural language. *Logic and Interactive Rationality (LIRA)*, 2:17–35.
- Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261.
- Winograd, Terry. 1972. Understanding natural language. *Cognitive Psychology*, 3(1):1–191.