

Don't Blame Distributional Semantics if it can't do Entailment

Matthijs Westera

Gemma Boleda

Universitat Pompeu Fabra, Barcelona, Spain

{firstname.lastname}@upf.edu

Abstract

Distributional semantics has had enormous empirical success in Computational Linguistics and Cognitive Science in modeling various semantic phenomena, such as semantic similarity, and distributional models are widely used in state-of-the-art Natural Language Processing systems. However, the theoretical status of distributional semantics within a broader theory of language and cognition is still unclear: What does distributional semantics model? Can it be, on its own, a fully adequate model of the meanings of linguistic expressions? The standard answer is that distributional semantics is not fully adequate in this regard, because it falls short on some of the central aspects of *formal* semantic approaches: truth conditions, entailment, reference, and certain aspects of compositionality. We argue that this standard answer rests on a misconception: These aspects do not belong in a theory of expression meaning, they are instead aspects of *speaker* meaning, i.e., communicative intentions in a particular context. In a slogan: words do not refer, speakers do. Clearing this up enables us to argue that distributional semantics on its own is an adequate model of expression meaning. Our proposal sheds light on the role of distributional semantics in a broader theory of language and cognition, its relationship to formal semantics, and its place in computational models.

Keywords: distributional semantics, expression meaning, formal semantics, speaker meaning, truth conditions, entailment, reference, compositionality, context

1 Introduction

Distributional semantics has emerged as a promising model of certain ‘conceptual’ aspects of linguistic meaning (e.g., Landauer and Dumais 1997; Turney and Pantel 2010; Baroni and Lenci 2010; Lenci 2018) and as an indispensable component of applications in Natural Language Processing (e.g., reference resolution, machine translation, image captioning; especially since Mikolov et al. 2013). Yet its theoretical status within a general theory of meaning and of language and cognition more generally is not clear (e.g., Lenci 2008; Erk 2010; Boleda and Herbelot 2016; Lenci 2018). In particular, it is not clear whether distributional semantics can be understood as an actual model of expression meaning – what Lenci (2008) calls the ‘strong’ view of distributional semantics – or merely as a model of something that *correlates* with expression meaning in certain partial ways – the ‘weak’ view. In this paper we aim to resolve, in favor of the ‘strong’ view, the question of what exactly distributional semantics models, what its role should be in an overall theory of language and cognition, and how its contribution to state of the art applications can be understood. We do so in part by clarifying its frequently discussed but still obscure relation to formal semantics.

Our proposal relies crucially on the distinction between what linguistic *expressions* mean outside of any particular context, and what *speakers* mean by them in a particular context of utterance. Here, we term the former **expression meaning** and the latter **speaker meaning**.¹ At least since Grice 1968 this distinction is generally acknowledged to be crucial to account for how humans communicate via

¹English inconveniently conflates what speakers do and what expressions do in a single verb “to mean”. In other languages the two types of ‘meaning’ go by different names, e.g., in Dutch, sentences ‘betekenen’ (mean, lit. ‘be-sign’ or ‘signify’) while speakers ‘bedoelen’ (mean, lit. ‘be-goal’).

language. Nevertheless, the two notions are sometimes confused, and we will point out a particularly widespread confusion in this paper. Consider an example, one which will recur throughout this paper:

- (1) The red cat is chasing a mouse.

The expression “the red cat” in this sentence can be used to refer to a cat with red hair (which is actually orangish in color) or to a cat painted red; “a mouse” to the animal or to the computer device; and in the right sort of context the whole sentence can be used to describe, for instance, a red car driving behind a motorbike. It is uncontroversial that the same expression can be used to communicate very different speaker meanings in different contexts. At the same time, it is likewise uncontroversial that not *anything goes*: what a speaker can reasonably mean by an expression in a given context – with the aim of being understood by an addressee – is constrained by its (relatively) context-invariant expression meaning. An important, long-standing question in linguistics and philosophy is what type of object could play the role of expression meaning, i.e., as a context-invariant common denominator of widely varying usages.

There exist two predominant candidates for a model of expression meaning: distributional semantics and formal semantics. **Distributional semantics** assigns to each expression, or at least each word, a high-dimensional, numerical vector, one which represents an abstraction over occurrences of the expression in some suitable dataset, i.e., its *distribution* in the dataset. **Formal semantics** assigns to each expression, typically via an intermediate, logical language, an interpretation in terms of reference to entities in the world, their properties and relations, and ultimately truth values of whole sentences.² To illustrate the two approaches, simplistically (and without intending to commit to any particular formal semantic analysis or (compositional) distributional semantics – see Section 5):

- (2) The red cat is chasing a mouse.

Formal semantics: $\iota x(\text{RED}(x) \wedge \text{CAT}(x) \wedge \exists y(\text{MOUSE}(y) \wedge \text{CHASE}(x, y)))$

Distributional semantics: ↗ ↘ ↙ → ↓ ↖ ← (i.e., a vector for each word)

Distributional and formal semantics are often regarded as two models of expression meaning that have *complementary strengths and weaknesses* and that, accordingly, must somehow be combined for a more complete model of expression meaning (e.g., Beltagy et al. 2013; Erk 2013; Baroni et al. 2014; Asher et al. 2016; Boleda and Herbelot 2016). For instance, in these works the vectors of distributional semantics are regarded as capturing lexical or conceptual aspects of meaning but not, or insufficiently so, truth conditions, reference, entailment and compositionality – and vice versa for formal semantics.³

Contrary to this common perspective, we argue that distributional semantics on its own can in fact be a fully satisfactory model of expression meaning, i.e., the ‘strong’ view of distributional semantics in Lenci 2008. Crucially, we will do so *not* by trying to show that distributional semantics can do all the things formal semantics does – we think it clearly cannot, at least not on its own – but by explaining that a semantics *should not* do all those things. In fact, formal semantics is mistaken about its job description, a mistake that we trace back, following a long strand in both philosophical and psycholinguistic literature, to a failure to properly distinguish speaker meaning and expression meaning. By clearing this up we aim to contribute to a firmer theoretical understanding of distributional semantics, of its role in an overall theory of communication, and of its employment in current models in NLP.

2 What we mean by distributional semantics

By distributional semantics we mean, in this paper, a broad family of models that assign (context-invariant) numerical vector representations to words, which are computed as abstractions over occur-

²Our formulation covers only the predominant, model-theoretic (or truth-conditional, referential) type of formal semantics, not, e.g., proof-theoretic semantics. We concentrate on this for reasons of space, but our proposal applies more generally.

³To clarify: when it is said that distributional semantics falls short, this pertains to distributional semantics on its own, i.e., a set of word vectors, combined perhaps with some basic algebraic operations or, at most, a simple classifier. By contrast, when distributional semantics is incorporated in a larger model (see section 2) the resulting system as a whole can be very successful.

rences of words in contexts. Implementations of distributional semantics vary, primarily, in the notion of context and in the abstraction mechanism used. A **context** for a word is typically a text in which it occurs, such as a document, sentence or a set of neighboring words, but it can also contain images (e.g., Feng and Lapata 2010; Silberer et al. 2017) or audio (e.g., Lopopolo and Miltenburg 2015) – in principle any place where one may encounter a word could be used. Because of how distributional models work, words that appear in similar contexts end up being assigned similar representations. At present, all models need large amounts of data to compute high-quality representations. The closer these data resemble our experience as language learners, the more distributional semantics is expected to be able in principle to generate accurate representations of – as we will argue – expression meaning.

As for the **abstraction mechanism** used, Baroni et al. (2014) distinguish between classic “count-based” methods, which work with co-occurrence statistics between words and contexts, and “prediction-based” methods, which instead apply machine learning techniques (artificial neural networks) to induce representations based on a prediction task, typically predicting the context given a word. For instance, the Skip-Gram model of Mikolov et al. (2013) would, applied to example (1), try to predict the words “the”, “red”, “is”, “chasing”, etc. from the presence of the word “cat” (more precisely, it would try to make these context words more likely than randomly sampled words, like “democracy” or “smear”). By training a neural network on such a task, over a large number of words in context, the first layer of the network comes to represent words as vectors, usually called *word embeddings* in the neural network literature. These word embeddings contain information about the words that the network has found useful for the prediction task.

In both count-based and prediction-based methods, the resulting vector representations encode abstractions over the distributions of words in the dataset, with the crucial property that words that appear in similar contexts are assigned similar vector representations.⁴ Our arguments in this paper apply to both kinds of methods for distributional semantics.

Word embeddings emerge not just from models that are expressly designed to yield word representations (such as Mikolov et al. 2013). Rather, any neural network model that takes words as input, trained on whatever task, must ‘embed’ these words in order to process them – hence any such model will result in word embeddings (e.g., Collobert and Weston 2008). Neural network models for language are trained for instance on language modeling (e.g., word prediction; Mikolov et al. 2010; Peters et al. 2018) or Machine Translation (Bahdanau et al., 2015). As long as the data on which these models are trained consist of word-context pairs, the resulting word embeddings qualify, for present purposes, as implementations of distributional semantics, and our proposal in the current paper applies also to them. Of course some implementations within this broad family may be better than others, and the type of task used is one parameter to be explored: It is expected that the more the task requires a human-like understanding of language, the better the resulting word embeddings will represent – as we will argue – the meanings of words. But our arguments concern the theoretical underpinnings of the distributional semantics framework more broadly rather than specific instantiations of it.

Lastly, some implementations of distributional semantics impose biases, during training, for obtaining word vectors that are more useful for a given task. For instance, to obtain word vectors useful for predicting lexical entailment (e.g., that being a cat entails being an animal), Vulić and Mrkšić (2017) impose a bias for keeping the vectors of supposed hypernyms, like “cat” and “animal”, close together (more precisely: in the same direction from the origin but with different magnitudes). This kind of approach presupposes, incorrectly as we will argue, that distributional semantics *should* account for entailment. It results in word vectors that are more useful for a particular task, but the model will be worse as a model of expression meaning. We will return to this type of approach in section 3.2.

⁴Both methods also share the characteristic that the dimensions of the high-dimensional space are automatically induced, and hence not directly interpretable (this is the main way in which they are different from traditional semantic features; see Boleda and Erk 2015). As a consequence, much work exploring distributional semantic models has relied not on the dimensions themselves but on geometric relations between words, in particular the notion of similarity (e.g., measured by cosine; as an anonymous reviewer notes, such technical notions of similarity need not completely align with semantic similarity in a more intuitive sense).

3 Distributional semantics as a model of expression meaning

We present two theoretical reasons why distributional semantics is attractive as a model of expression meaning, before arguing in section 4 that it can also be sufficient.

3.1 Reason 1: Meaning from use; abstraction and parsimony

We take it to be uncontroversial that what expressions mean is to be explained at least in part in terms of how they are used by speakers of the relevant linguistic community (e.g., Wittgenstein 1953; Grice 1968).⁵ A similar view has motivated work on distributional semantics (e.g., Lenci 2008; also at its conception, e.g., Harris 1954). For instance, what the word “cat” means is to be explained at least in part in terms of the fact that speakers have used it to refer to cats, to describe things that resemble cats, to insult people in certain ways, and so on. Note that the usages of words generally resist systematic categorization into definable senses, and attempts to characterize word meaning by sense enumeration generally fail (e.g., Kilgarriff 1997; Hanks 2000; Erk 2010; cf. Pustejovsky 1995).

A minimal, parsimonious way of explaining the meaning of an expression in terms of its uses is to say simply that **the meaning of an expression is an abstraction over its uses**. Such abstractions are, of course, exactly what distributional semantics delivers, and the view that it corresponds to expression meaning is what Lenci (2008) calls the ‘strong’ view of distributional semantics. Distributional semantics is especially parsimonious because it relies on (mostly) domain-independent mechanisms for abstraction (e.g., principal components analysis; neural networks). Of course not all implementations are equally adequate, or equally parsimonious; there are considerable differences both in the abstraction mechanism relied upon and in the dataset used (see section 2). But the family as a whole, defined by the core tenet of associating with each word an abstraction over its use, is highly suitable in principle for modeling expression meaning. This makes the ‘strong’ view of distributional semantics attractive.

An alternative to the ‘strong’ view is what Lenci (2008) calls the ‘weak’ view: that an abstraction over use may be *part* of what determines expression meaning, but that more is needed. This view underlies for instance the common assumption that a more complete model of expression meaning would require integrating distributional and formal semantics (e.g., Beltagy et al. 2013; Erk 2013; Baroni et al. 2014; Asher et al. 2016; Boleda and Herbelot 2016). But in section 4 we argue that the notions of formal semantic, like reference, truth conditions and entailment, do not belong at the level of expression meaning in the first place, and, accordingly, that distributional semantics can be sufficient as a model of expression meaning. Theoretical parsimony dictates that we opt for the least presumptive approach compatible with the empirical facts, i.e., with what a theory of expression meaning should account for.

Some authors equate the meaning of an expression not with an abstraction over all uses, but only *stereotypical* uses: what an expression means would be what a stereotypical speaker in a stereotypical context means by it (e.g., Schiffer 1972; Bennett 1976; Soames et al. 2002). This approach is appealing because it does justice to native speaker’s intuitions about expression meaning, which are known to reflect stereotypical speaker meaning (see Section 4). However, several authors have pointed out that stereotypical speaker meaning is ultimately not an adequate notion of expression meaning (e.g., Bach 2002; Recanati 2004). To see just one reason why, consider the following arbitrary example:

(3) Jack and Jill got married.

A stereotypical use of this expression would convey the speaker meaning that Jack and Jill got married *to each other*. But this cannot be the (context-invariant) meaning of the expression “Jack and Jill got married”, or else the following additions would be redundant and contradictory, respectively:⁶

⁵For compatibility with a more cognitive, single-agent perspective of language, such as *I-language* in the work of Chomsky (e.g., 1986), this could be restricted to the uses of a word as experienced by a single agent when learning the language.

⁶An anonymous reviewer rightly points out that this presupposes that notions like redundancy and contradiction apply to expression meanings. We think they don’t (see Section 4), at least not in their strictly logical senses, but they would if expression meaning were to be construed as stereotypical speaker meaning, which is the position we are criticizing here.

- (4) Jack and Jill got married *to each other*.
- (5) Jack and Jill got married *to their respective childhood friends*.

Hence the stereotypical speaker meaning of (3) cannot be its expression meaning. For many more examples and discussion see Bach 2002. Another challenge for defining expression meaning as stereotypical speaker meaning is that of having to define “stereotypical”. It cannot be defined simply as the most frequent type, because that presupposes that uses can be categorized into clearly delineated, countable types. Moreover, an ‘empty’ context is a context too, and not the most stereotypical one.

Summing up: what an expression means depends on how speakers use it, but the uses of an expression more generally resist systematic categorization into enumerable senses, and selecting a stereotypical use isn’t adequate either. Equating expression meaning with an abstraction over all uses, as the ‘strong’ view of distributional semantics has it, is more adequate, and particularly attractive for reasons of parsimony.

3.2 Reason 2: Distributional semantics as a model of concepts

Another reason why distributional semantics is attractive as a model of expression meaning is the following. As mentioned in section 1, distributional semantics is often regarded as a model of ‘conceptual’ aspects of meaning (e.g., Landauer and Dumais 1997; Baroni and Lenci 2010; Boleda and Herbelot 2016). This view seems to be motivated in part empirically: distributional semantics is successful at what are intuitively conceptual tasks, like modeling word similarity, priming and analogy. Moreover, it aligns with the widespread view in philosophy and developmental psychology that abstraction over instances is a main mechanism of concept formation (e.g., the influential work of Jean Piaget). Let us explain why concepts, and in particular those modeled by distributional semantics (because there is some confusion about their nature), would be suitable representatives of expression meaning.

It is sometimes assumed that the word vector for “cat” should model the concept CAT (we discuss some work that makes this assumption below). This may be a ‘true enough’ approximation for practical applications, but theoretically it is, strictly speaking, on the wrong track. This is because the word vector for “cat” does not model the concept CAT – that would be an abstraction over occurrences of *actual cats*, after all. Instead, the word vector for “cat” is an abstraction over occurrences of *the word*, not the animal, hence it would model the concept of *the word* “cat”, say, THEWORDCAT. The extralinguistic concept CAT and the linguistic concept THEWORDCAT are very different. The concept CAT encodes knowledge about cats having fur, four legs, the tendency to meow, etc.; the concept THEWORDCAT instead encodes knowledge that the word “cat” is a common noun, that it rhymes with “bat” and “hat”, how speakers have used it or tend to use it, that the word doesn’t belong to a particular register, and so on.⁷

Our distinction between THEWORDCAT and CAT, or between linguistic and extralinguistic concepts, is not new, and word vectors are known to capture the more linguistic kind of information, and to be (at best) only a proxy for the extralinguistic concepts they are typically used to denote by a speaker (e.g., Miller and Charles 1991). But it appears to be sometimes overlooked. For instance, the assumption that the word vector for “cat” would (or should) model the extralinguistic concept CAT is made in work using distributional semantics to model entailment, e.g., that being a cat entails being an animal (e.g., Geffet and Dagan 2005; Roller et al. 2014; Vulić and Mrkšić 2017). But clearly the entailment relation holds between the extralinguistic concepts CAT and ANIMAL – being a cat entails being an animal – *not* between the linguistic concepts THEWORDCAT and THEWORDANIMAL actually modeled by distributional semantics: being the word “cat” does not entail (in fact, it excludes) being the word “animal”. Hence these approaches are, strictly speaking, theoretically misguided – although their conflation of linguistic and extralinguistic concepts may be a defensible simplification for practical purposes.

There have been many proposals to integrate formal and distributional semantics (e.g., Beltagy et al. 2013; Erk 2013; Baroni et al. 2014; Asher et al. 2016), and a similar confusion exists in at least some of them (Asher et al., 2016; McNally and Boleda, 2017). We are unable within the scope of the cur-

⁷To clarify: the difference persists even if the notion of context in distributional semantics is enriched to include, say, pictures of cats, or even actual cats. The distributions it models would still be distributions *of words*, not of things like cats.

rent paper to do justice to the technical sophistication of these approaches, but for present purposes, impressionistically, the type of integration they pursue can be pictured as follows:

(6) The red cat is chasing a mouse.

Formal semantics: $\iota x(\text{RED}(x) \wedge \text{CAT}(x) \wedge \exists y(\text{MOUSE}(y) \wedge \text{CHASE}(x, y)))$

Distributional semantics: $\nearrow \searrow \swarrow \rightarrow \downarrow \nearrow \leftarrow$ (i.e., a vector for each word)

Possible integration: $\iota x(\searrow(x) \wedge \swarrow(x) \wedge \exists y(\leftarrow(y) \wedge \downarrow(x, y)))$ (very simplistically)

Again, this may be a ‘true enough’ approximation, but it is theoretically on the wrong track. The atomic constants in formal semantics are normally understood (e.g., Frege 1892 and basically anywhere since) to denote the extralinguistic kind of concept, i.e., CAT and not THEWORDCAT. Put differently, entity x in example (6) should be entailed to be a cat, not to be the word “cat”. This means that the distributional semantic word vectors are, strictly speaking, out of place in a formal semantic skeleton like in (6).⁸

In short, distributional semantics models linguistic concepts like THEWORDCAT, not extralinguistic concepts like CAT. But this is not a shortcoming; it makes distributional semantics more adequate, rather than less adequate, as a model of expression meaning, for the following reason. A prominent strand in the literature on concepts conceives of concepts as *abilities* (e.g., Dummett 1993; Bennett and Hacker 2008; for discussion see Margolis and Laurence 2014). For instance, possessing the concept CAT amounts to having the ability to recognize cats, discriminate them from non-cats, and draw certain inferences about cats. The concept CAT is, then, the starting point for *interpreting* an object as a cat and draw inferences from it. It follows that the concept THEWORDCAT is the starting point for interpreting a word as the word “cat” and drawing inferences from it, notably, inferences about what a speaker in a particular context may use it for: for instance, to refer to a particular cat.⁹ Thus, **the view of distributional semantics as a model of concepts, but crucially concepts of words, establishes word vectors as a necessary starting point for interpreting a word.** This is exactly the explanatory job assigned to expression meaning: a context-invariant starting point for interpretation. Not coincidentally, for neural networks that take words as input, distributional semantics resides in the first layer of weights (see Section 2).

Summing up, this section presented two reasons why distributional semantics is attractive as a model of expression meaning. The next section considers whether it could also be *sufficient*.

4 Limits of distributional semantics: words don’t refer, speakers do.

In many ways the standard for what a theory of expression meaning ought to do has been set by formal semantics. Consider again our simplistic comparison of distributional semantics and formal semantics:

(7) The red cat is chasing a mouse.

Formal semantics: $\iota x(\text{RED}(x) \wedge \text{CAT}(x) \wedge \exists y(\text{MOUSE}(y) \wedge \text{CHASE}(x, y)))$

Distributional semantics: $\nearrow \searrow \swarrow \rightarrow \downarrow \nearrow \leftarrow$ (i.e., a vector for each word)

The logical formulae into which formal semantics translates this example are assigned precise interpretations in (a model of) the outside world. For instance, RED would denote the set of all red things, CAT

⁸The mathematical techniques of the aforementioned approaches do not depend for their validity on the exact nature of the vectors. We hope that these techniques can be used to represent not expression meaning but speaker meaning (see section 4), provided we use vector representations of the distribution of actual cats, instead of the word “cat”.

⁹This is because how a speaker may use a word is constrained by how speakers have used it in the past – a trait of linguistic convention. Since the concept THEWORDCAT reflects uses of “cat” in the past, among which are referential uses, it constrains (hence warrants inferences about) what it may be used by a given speaker to refer to. (To clarify: this does not imply that the actual or potential referents of a word are actually part of its meaning – see Section 4.) The same holds for the distributional semantic word vector for “cat”, although instantiations of distributional semantics may differ in how much referentially relevant information they encode. Presumably, more information of this sort is encoded when reference is prominent in the original data, for instance when a distributional semantic model is trained on referential expressions grounded in images (Kazemzadeh et al., 2014); otherwise such information needs to be induced from patterns in the text alone (like any other semantic information in text-only distributional semantics).

the set of all cat-like things, CHASE a set of pairs where one chases the other, the variable x would be bound to a particular entity in the world, etc., and the logical connectives can have their usual truth-conditional interpretation.¹⁰ In this way formal semantics accounts for reference to things in the world and it accounts for truth values (which is what sentences refer to; Frege 1892). Moreover, referents and truth values across possible worlds/situations in turn determine truth conditions, and thereby entailments – because one sentence entails another if whenever the former is true the latter is true as well.¹¹ By contrast, distributional semantics on its own (cf. footnote 3) struggles with these aspects (Boleda and Herbelot 2016; see also the work discussed in section 3.2 on entailment), which has motivated aforementioned attempts to integrate formal and distributional semantics (e.g., Beltagy et al. 2013; Erk 2013; Baroni et al. 2014; Asher et al. 2016; Boleda and Herbelot 2016). Put simply, distributional semantics struggles because there are no entities or truth values in distributional space to refer to. Nevertheless, we think that this isn't a shortcoming of distributional semantics; **we argue that a theory of expression meaning shouldn't model these aspects.**¹²

We think that these referential notions on which formal semantics has focused are best understood to reside at the level of speaker meaning, not expression meaning. In a nutshell, **our position is that words don't refer, speakers do** (e.g., Strawson 1950) – and analogously for truth conditions and entailment. The fact that speakers often refer *by means of* linguistic expressions doesn't entail that these expressions must in themselves, out of context, have a determinate reference, or even be capable of referring (or capable of entailing, of providing information, of being true or false). Parsimony (again) suggests that we do not assume the latter: To explain why a speaker can use, e.g., the expression “cat” to refer to a cat, it is sufficient that, in the relevant community, that is how the expression is often used. It is theoretically superfluous to assume in addition that the expression “cat” itself refers to cats.

Now, most work in formal semantics would acknowledge that “cat” out of context doesn't refer to cats, and that its use in a particular context to refer to cats must be explained on the basis of a less determinate, more underspecified notion of expression meaning. More generally, expressions are well-known to underdetermine speaker meaning (e.g., Bach 1994; Recanati 2004), as basically any example can illustrate (e.g., (1) “red cat” and (3) “got married”). However, this alone does not imply that the notions of formal semantics are inadequate for characterizing expression meaning; in principle one could try to define, in formal semantics, the referential potential of “cat” in a way that is compatible with its use to refer to cats, to cat-like things, etcetera. And one could define the expression meaning of “Jack and Jill got married” in a way that is compatible with them marrying each other and with each marrying someone else.¹³ What is problematic for a formal semantic approach is that the *ways* in which expressions underdetermine speaker meaning are not clearly delineated and enumerable, and that there is no symbolically definable common core among all uses.¹⁴ This argument was made for instance by Wittgenstein (1953), who notes that the uses of an expression (his example was “game”) are tied together

¹⁰In fact, the common reliance on an intermediate formal, logical language is not what defines formal semantics; what matters is that it treats natural language itself as a formal language (Montague, 1970), by compositionally assigning precise interpretations to it – and this can be done directly, or indirectly via translation to a logical language as in our example.

¹¹There are serious shortcomings to the formal semantics approach, some of which we discuss below, but others which aren't relevant for present purposes. An important criticism that we won't discuss is that the way in which formal semantics assigns interpretations to natural language relies crucially on the manual labor of hard-working semanticists, which does not scale up.

¹²Truth conditions, entailments and reference are just three sides of the same central, referential tenet of formal semantics, and what we will say about reference in what follows will apply to truth conditions and entailment, and vice versa. An anonymous reviewer draws our attention also to the logical notions of satisfiability and validity, i.e., possible vs. necessary truth. Our proposal applies to these notions too, regardless of whether they are understood in terms of quantification over possible ways the world may be, or in terms of quantification over possible interpretations.

¹³For instance, an anonymous reviewer notes that richer logical formalisms such as dependent type theory are well-suited for integrating contextual information into symbolic representations.

¹⁴Similarly, Bach (2005, among others) has criticized the common approach in formal semantics of incorporating, in definitions of expression meaning, ‘slots’ where supposed context-sensitive material is to be plugged in. The meaning of a scalar adjective like “big”, for instance, would contain a slot for ‘standard of comparison’ to be filled by context in order to explain why the same thing may be described as “big” in one context but not in another (e.g., Kennedy 2007). Bach (2005) notes that this type of approach does not generalize to all the ways in which expression meaning underdetermines speaker meaning; the meaning of each expression would essentially end up being a big empty slot, to be magically filled by context.

not by definition but by family resemblance. More recent iterations of this argument can be found in criticisms of the “classical”, definitional view of concepts (e.g., Rosch and Mervis 1975; Fodor et al. 1980; Margolis and Laurence 2014), and in criticisms of sense enumeration approaches to word meaning (e.g., Kilgarriff 1997; Hanks 2000; Erk 2010; cf. Pustejovsky 1995), which we already mentioned briefly before: it is unclear what constitutes a word sense, and no enumeration of senses covers all uses.

The only truly common core among all uses of any given expression is that they are all, indeed, uses of the same expression. Hence, if expression meaning is to serve its purpose as a common core among all uses, i.e., as a context-invariant starting point of semantic/pragmatic explanations, then it must reflect all uses. As we argued in section 3, distributional semantics, conceived of as a model of expression meaning (i.e., the ‘strong’ view of Lenci 2008), embraces exactly this fact. This makes the representations of distributional semantics, but not those of formal semantics, suitable for characterizing expression meaning. By contrast, (largely) discrete notions like reference, truth and entailment are useful, at best, at the level of *speaker* meaning – recall that our position is that words don’t refer, speakers do (Strawson, 1950).¹⁵ That is, one can fruitfully conceive of a particular speaker, in some individuated context, as intending to refer to discrete things, communicating a certain determinate piece of information that can be true or false, entailing certain things and not others. This still involves considerable abstraction, as any symbolic model of a cognitive system would (Marr, 1982); e.g., speaker intentions may not always be as determinate as a symbolic model presupposes. But the amount of abstraction required, in particular the kind of determinacy of content that a symbolic model presupposes, is not as problematic in the case of speaker meaning as for expression meaning. The reason is that a model of speaker meaning needs to cover only a single usage, by a particular speaker situated in a particular context; a model of expression meaning, by contrast, needs to cover countless interactions, across many different contexts, of a whole community of speakers. The symbolic representations of formal semantics are ill-suited for the latter.

Despite the foregoing considerations being prominent in the literature, formal semantics has continued to assume that referents, truth conditions, etc., are core aspects of expression meaning. The main reason for this is the traditional centrality of supposedly ‘semantic’ intuitions in formal semantics (Bach, 2002), either as the main source of data or as the object of investigation (‘semantic competence’, for criticism see Stokhof 2011). In particular, formal semantics has attached great importance to intuitions about truth conditions (e.g., “semantics with no treatment of truth conditions is not semantics”, Lewis 1972:169), a tenet going back to its roots in formal logic (e.g., Montague 1970 and the earlier work of Frege, Tarski, among others). Clearly, if expressions on their own do not even *have* truth conditions, as we have argued, these supposedly semantic intuitions cannot genuinely be about expression meaning. And that is indeed what many authors have pointed out. Strawson (1950); Grice (1975); Bach (2002), among others, have argued that **what seem to be intuitions about the meaning of an expression are really about what a stereotypical speaker would mean by it** – or at least they are heavily influenced by it. Again example (3) serves as an illustration here: intuitively “marry” means “marry each other”, but to assume that this is therefore its expression meaning would be inadequate (as we discussed in section 3.1). But we want to stress that this is not just an occasional trap set by particular kinds of examples; just being a bit more careful doesn’t cut it. It is the foundational intuition that expressions can even *have* truth conditions that is already inaccurate. Our intuitions are *fundamentally* not attuned to expression meaning, because expression meaning is not normally what matters to us; it is only an instrument for conveying speaker meaning, and, much like the way we string phonemes together to form words, it plays this role largely or entirely without our conscious awareness. The same point has been made in the more psycholinguistic literature (Schwarz, 1996), occasionally in the formal semantics/pragmatics literature (Kadmon and Roberts, 1986), and there is increasing acknowledgment of this also in experimental pragmatics, in particular of the fact that participants in experiments imagine stereotypical contexts (e.g., Westera and Brasoveanu 2014; Degen and Tanenhaus 2015; Poortman 2017).

Summing up, the standard that formal semantics has set for what a theory of expression meaning

¹⁵We are not discussing another long-standing criticism of formal semantics, namely that referring (and asserting something that can be true or false) is not all that speakers do with language (e.g., Austin 1975; Searle 1969). We do not claim that formal semantics would be *sufficient* as a model of speaker meaning; only that its notions are more adequate there than at the level of expression meaning.

ought to account for, and which makes distributional semantics appear to fall short, turns out to be misguided. Reference, truth conditions and entailment belong at the level of speaker meaning, not expression meaning. It entails that distributional semantics on its own need not account for these aspects, either theoretically or computationally; it should only provide an adequate starting point. Interestingly, this corresponds exactly to its role in current neural network models, on tasks that involve identifying aspects of speaker meaning. Consider the task of visual reference resolution (e.g., Plummer et al. 2015), where the inputs are a linguistic description plus an image and the task is to identify the intended referent in the image. A typical neural network model would achieve this by first activating word embeddings (a form of distributional semantics; Section 2) and then combining and transforming these together with a representation of the image into a representation of the intended referent – speaker meaning.

5 Compositionality

Language is compositional in the sense that what a larger, composite expression means is determined (in large part) by what its components mean and the way they are put together. Compositionality is sometimes mentioned as a strength of formal semantics and as an area where distributional semantics falls short (a.o. Beltagy et al., 2013). But in fact both approaches have shown strengths and weaknesses regarding compositionality (see Boleda and Herbelot 2016 for an overview). To illustrate, consider again:

(8) The red cat is chasing a mouse.

In this context the adjective “red” is used by the speaker to mean something closer to ORANGE (because the “red hair” of cats is typically orange), unlike its occurrence in, say, “red paint”. Distributional semantics works quite well for this type of effect in the composition of content words (e.g., Baroni et al. 2014; McNally and Boleda 2017), an area where formal semantics, which tends to leave the basic concepts unanalyzed, has struggled (despite efforts such as Pustejovsky 1995). Classic compositional distributional semantics, in which distributional representations are combined with some externally specified algorithm (which can be as simple as addition), also works reasonably well for short sentences, as measured for instance on sentence similarity (e.g., Mitchell and Lapata 2010; Grefenstette et al. 2013; Marelli et al. 2014). But for longer expressions distributional semantics on its own falls short (cf. our clarification of “on its own” in footnote 3), and this is part of what has inspired aforementioned works on integrating formal and distributional semantics (e.g., Coecke et al. 2011; Grefenstette and Sadrzadeh 2011; Beltagy et al. 2013; Erk 2013; Baroni et al. 2014; Asher et al. 2016).

However, that distributional semantics falls short of accounting for full-fledged compositionality does not mean that it cannot be a sufficient model of expression meaning. For that, it should be established first that compositionality wholly resides at the level of expression meaning – and it is not clear that it does. Let us take a closer look at the main theoretical argument for compositionality, the argument from *productivity*.¹⁶ According to this argument, compositionality is necessary to explain how a competent speaker can understand the meaning of a composite expression that they have never before encountered. However, in appealing to a person’s supposed understanding of the meaning of an expression, this argument is subject to the revision proposed in Section 4: it reflects speaker meaning, not expression meaning. More correctly phrased, then, the type of data motivating the productivity argument is that a person who has never encountered a speaker uttering a certain composite expression, is nevertheless able to understand *what some (actual or hypothetical) speaker would mean by it*. And this leaves undetermined where compositionality should reside: at the level of expression meaning, speaker meaning, or both.

To illustrate, consider again example (8), “The red cat is chasing a mouse”. A speaker of English who has never encountered this sentence will nevertheless understand what a stereotypical speaker would mean by it (or will come up with a set of interpretations) – this is an instance of productivity. One explanation for this would be that the person can compositionally compute an expression meaning for the

¹⁶To clarify: the issue here is not whether *distributed* representations can be composed, but whether *distributional* representations – i.e., abstractions over distributions of use – can and should be composed. Sophisticated approaches exist for composing *distributed* representations (notably the tensor product approach of Smolensky 1990).

whole sentence, and from there infer what a speaker would mean by it. This places the burden of compositionality entirely on the notion of expression meaning. An alternative would be to say that the person first infers speaker meanings for each word (say, the concept CAT for “cat”),¹⁷ and then composes these to obtain a speaker meaning of the full sentence. This would place the burden of compositionality entirely on the notion of speaker meaning (cf. the notion of *resultant procedure* in Grice 1968; see Borge 2009 for a philosophical argument for compositionality residing at the speaker meaning level). The two alternatives are opposite extremes of a spectrum; and note that the first is what formal semantics proclaims, yet the second is what formal semantics does, given that the notions it composes in fact reside at the level of speaker meaning (e.g., concepts like CAT as opposed to THEWORDCAT; and the end product of composition in formal semantics is typically a truth value). There is also a middle way: The person could in principle compositionally compute expression meanings for certain intermediate constituents (say, “the red cat”, “a mouse” and “chases”), then infer speaker meanings for these constituents (say, a particular cat, an unknown mouse, and a chasing event), and only *then* continue to compose these to obtain a speaker meaning for the whole sentence. This kind of middle way requires that a model of expression meaning (distributional semantics) accounts for some degree of compositionality (say, the direct combination of content words), with a model of speaker meaning (say, formal semantics) carrying the rest of the burden. The proposal in McNally and Boleda (2017) is a version of this position.

The foregoing shows that the productivity argument for compositionality falls short as an argument for compositionality *of expression meanings*; that is, **compositionality may well reside in part, or even entirely, at the level of speaker meaning**. We will not at present try to settle the issue of where compositionality resides – though we favor a view according to which compositionality is multi-faceted and doesn’t necessarily reside exclusively at one level.¹⁸ What matters for the purposes of this paper is that the requirement imposed by formal semantics, that a theory of expression meaning should account for full-fledged compositionality, turns out to be unjustified.

6 Outlook

We presented two strong reasons why distributional semantics is attractive as a model of expression meaning, i.e., in favor of the ‘strong’ view of Lenci 2008: The parsimony of regarding expression meaning as an abstraction over use; and the understanding of these abstractions as concepts and, thereby, as a necessary starting point for interpretation. Moreover, although distributional semantics struggles with matters like reference, truth conditions and entailment, we argued that a theory of expression meaning *should not* account for these aspects: words don’t refer, speakers do (and likewise for truth conditions and entailments). The referential approach to expression meaning of formal semantics is based on misinterpreting intuitions about stereotypical speaker meaning as being about expression meaning. The same misinterpretation has led to the common view that a theory of expression meaning should be compositional, whereas in fact compositionality may reside wholly or in part (and does reside, in formal semantics) at the level of speaker meaning. Clearing this up reveals that distributional semantics is the more adequate approach to expression meaning. In between our mostly theoretical arguments for this position, we have shown how a consistent interpretation of distributional semantics as a model of expression meaning sheds new light on certain applications: e.g., distributional semantic approaches to entailment and attempts at integrating distributional and formal semantics.

¹⁷We discuss this here as a hypothetical possibility; to assume that individual words of an utterance can be assigned speaker meanings may not be a feasible approach in general.

¹⁸The empirical picture is undecided in this regard: just because distributional semantics appears to be able to handle certain aspects of compositionality, that doesn’t mean it should. After all, word vectors like “cat” have been quite successfully used as a proxy for extra-linguistic concepts like CAT, even though as we explained this is strictly speaking a misuse (conflating CAT and THEWORDCAT; see section 3.2). Perhaps the moderate success of distributional semantics on for instance adjective-noun composition like “red cat” reflects the fact that the extra-linguistic concepts RED and CAT compose (speaker meaning), even if the linguistic concepts THEWORDRED and THEWORDCAT don’t (expression meaning).

Acknowledgments

We are grateful to the anonymous reviewers for their valuable comments. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 715154), and from the Spanish Ramón y Cajal programme (grant RYC-2015-18907). This paper reflects the authors’ view only, and the EU is not responsible for any use that may be made of the information it contains.



References

- Asher, N., T. Van de Cruys, A. Bride, and M. Abrusán (2016). Integrating type theory and distributional semantics: a case study on adjective–noun compositions. *Computational Linguistics* 42(4), 703–725.
- Austin, J. L. (1975). *How to do things with words*, Volume 88. Oxford university press.
- Bach, K. (1994). Conversational implicature. *Mind and Language* 9, 124–62.
- Bach, K. (2002). Seemingly semantic intuitions. In J. Campbell, M. O’Rourke, and D. Shier (Eds.), *Meaning and Truth*. New York: Seven Bridges Press.
- Bach, K. (2005). Context ex machina. *Semantics versus pragmatics* 1544.
- Bahdanau, D., K. Cho, and Y. Bengio (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of ICLR Conference Track*, San Diego, CA.
- Baroni, M., R. Bernardi, and R. Zamparelli (2014). Frege in space: A program of compositional distributional semantics. *LiLT (Linguistic Issues in Language Technology)* 9.
- Baroni, M., G. Dinu, and G. 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 (Volume 1: Long Papers)*, Volume 1, pp. 238–247.
- Baroni, M. and A. Lenci (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics* 36(4), 673–721.
- Beltagy, I., C. Chau, G. Boleda, D. Garrette, K. Erk, and R. Mooney (2013). Montague meets markov: Deep semantics with probabilistic logical form. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity*, Volume 1, pp. 11–21.
- Bennett, J. (1976). *Linguistic Behavior*. Cambridge University Press.
- Bennett, M. R. and P. M. S. Hacker (2008). *History of cognitive neuroscience*. John Wiley & Sons.
- Boleda, G. and K. Erk (2015). Distributional semantic features as semantic primitives—or not. In *AAAI Spring Symposium on Knowledge Representation and Reasoning, Stanford University, USA*.
- Boleda, G. and A. Herbelot (2016). Formal distributional semantics: Introduction to the special issue. *Computational Linguistics* 42(4), 619–635.
- Borge, S. (2009). Intentions and compositionality. *SATS* 10(1), 100–106.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger.

- Coecke, B., M. Sadrzadeh, and S. Clark (2011). Mathematical Foundations for a Compositional Distributional Model of Meaning. *Linguistic Analysis: A Festschrift for Joachim Lambek* 36(1–4), 345–384.
- Collobert, R. and J. 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*, pp. 160–167. ACM.
- Degen, J. and M. K. Tanenhaus (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science* 39(4), 667–710.
- Dummett, M. (1993). *The seas of language*. Oxford University Press.
- Erk, K. (2010). What is word meaning, really?:(and how can distributional models help us describe it?). In *Proceedings of the 2010 workshop on geometrical models of natural language semantics*, pp. 17–26. Association for Computational Linguistics.
- Erk, K. (2013). Towards a semantics for distributional representations. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers*, pp. 95–106.
- Feng, Y. and M. Lapata (2010). Visual information in semantic representation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 91–99. Association for Computational Linguistics.
- Fodor, J. A., M. F. Garrett, E. C. Walker, and C. H. Parkes (1980). Against definitions. *Cognition* 8(3), 263–367.
- Frege, G. (1892). Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100(1), 25–50.
- Geffet, M. and I. Dagan (2005). The distributional inclusion hypotheses and lexical entailment. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 107–114. Association for Computational Linguistics.
- Grefenstette, E., G. Dinu, Y.-Z. Zhang, M. Sadrzadeh, and M. Baroni (2013). Multi-Step Regression Learning for Compositional Distributional Semantics. In *Proceedings of IWCS 2013 (10th International Conference on Computational Semantics)*, East Stroudsburg PA, pp. 131–142. ACL.
- Grefenstette, E. and M. Sadrzadeh (2011). Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11)*, pp. 1394–1404.
- Grice, H. P. (1968). Utterer’s meaning, sentence-meaning, and word-meaning. In *Philosophy, Language, and Artificial Intelligence*, pp. 49–66. Springer.
- Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and Semantics*, Volume 3, pp. 41–58.
- Hanks, P. (2000). Do word meanings exist? *Computers and the Humanities* 34(1-2), 205–215.
- Harris, Z. S. (1954). Distributional structure. *Word* 10(2-3), 146–162.
- Kadmon, N. and C. Roberts (1986). Prosody and scope: The role of discourse structure. In *CLS 22: Proceedings of the Parasession on Pragmatics and Grammatical Theory*, pp. 16–18. Chicago Linguistic Society.
- Kazemzadeh, S., V. Ordonez, M. Matten, and T. L. Berg (2014). Referit game: Referring to objects in photographs of natural scenes. In *EMNLP*.

- Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy* 30(1), 1–45.
- Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the Humanities* 31(2), 91–113.
- Landauer, T. K. and S. T. Dumais (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review* 104(2), 211.
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian journal of linguistics* 20(1), 1–31.
- Lenci, A. (2018). Distributional models of word meaning. *Annual review of Linguistics* 4, 151–171.
- Lewis, D. (1972). General semantics. In *Semantics of natural language*, pp. 169–218. Springer.
- Lopopolo, A. and E. Miltenburg (2015). Sound-based distributional models. In *Proceedings of the 11th International Conference on Computational Semantics*, pp. 70–75.
- Marelli, M., S. Menini, M. Baroni, L. Bentivogli, R. Bernardi, R. Zamparelli, et al. (2014). A sick cure for the evaluation of compositional distributional semantic models. In *LREC*, pp. 216–223.
- Margolis, E. and S. Laurence (2014). Concepts. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.). Metaphysics Research Lab, Stanford University.
- Marr, D. C. (1982). *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman & Co.
- McNally, L. and G. Boleda (2017). Conceptual versus referential affordance in concept composition. In *Compositionality and concepts in linguistics and psychology*, pp. 245–267. Springer.
- Mikolov, T., M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur (2010). Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*.
- Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119.
- Mikolov, T., W.-t. Yih, and G. Zweig (2013). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746–751.
- Miller, G. A. and W. G. Charles (1991). Contextual correlates of semantic similarity. *Language and cognitive processes* 6(1), 1–28.
- Mitchell, J. and M. Lapata (2010). Composition in distributional models of semantics. *Cognitive science* 34(8), 1388–1429.
- Montague, R. (1970). Universal grammar. *Theoria* 36(3), 373–398.
- Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer (2018). Deep contextualized word representations. In *Proc. of NAACL*.
- Plummer, B. A., L. Wang, C. M. Cervantes, J. C. Caicedo, J. Hockenmaier, and S. Lazebnik (2015). Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models. In *Proceedings of the IEEE international conference on computer vision*, pp. 2641–2649.
- Poortman, E. B. (2017). Concept typicality and the interpretation of plural predicate conjunction. In *Compositionality and concepts in linguistics and psychology*, pp. 139–162. Springer.

- Pustejovsky, J. (1995). *The generative lexicon*. MIT press.
- Recanati, F. (2004). *Literal meaning*. Cambridge University Press.
- Roller, S., K. Erk, and G. Boleda (2014). Inclusive yet selective: Supervised distributional hypernymy detection. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 1025–1036.
- Rosch, E. and C. B. Mervis (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology* 7(4), 573–605.
- Schiffer, S. (1972). *Meaning*. Oxford: Oxford University Press.
- Schwarz, N. (1996). *Cognition and Communication: Judgmental Biases, Research Methods and the Logic of Conversation*. Hillsdale, NJ: Erlbaum.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*, Volume 626. Cambridge university press.
- Silberer, C., V. Ferrari, and M. Lapata (2017). Visually grounded meaning representations. *IEEE transactions on pattern analysis and machine intelligence* 39(11), 2284–2297.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence* 46(1-2), 159–216.
- Soames, S. et al. (2002). *Beyond rigidity: The unfinished semantic agenda of naming and necessity*. Oxford University Press on Demand.
- Stokhof, M. (2011). Intuitions and competence in formal semantics. In B. P. and M. Glanzberg and J. Skilters (Eds.), *Formal Semantics and Pragmatics*, Number 6 in Baltic International Yearbook of Cognition, Logic and Communication, pp. 1–23. New Prairie Press.
- Strawson, P. F. (1950). On referring. *Mind* 59(235), 320–344.
- Turney, P. D. and P. Pantel (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research* 37, 141–188.
- Vulić, I. and N. Mrkšić (2017). Specialising word vectors for lexical entailment. In *NAACL 2018*.
- Westera, M. and A. Brasoveanu (2014). Ignorance in context: The interaction of modified numerals and quds. In S. D. Todd Snider and M. Weigand (Eds.), *Semantics and Linguistic Theory (SALT) 24*, pp. 414–431.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell. Edited by G.E.M. Anscombe (trans.) and R. Rhees.