A Geometric Method for Context Sensitive Distributional Semantics

by

Stephen McGregor

A thesis submitted to Queen Mary University of London for the degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary, University of London
United Kingdom

September 2017

I add this:

I hereby grant permission to anyone to do anything they so please with the text of this thesis and any information they derive from it or meaning they find in it, with or without acknowledgement of the source.

Abstract

This thesis describes a novel methodology, grounded in the distributional semantic paradigm, for building context sensitive models of word meaning, affording an empirical exploration of the relationship between words and concepts. Anchored in theoretical linguistic insight regarding the contextually specified nature of lexical semantics, the work presented here explores a range of techniques for the selection of subspaces of word co-occurrence dimensions based on a statistical analysis of input terms as observed within large-scale textual corpora. The relationships between word-vectors that emerge in the projected subspaces can be analysed in terms of a mapping between their geometric features and their semantic properties. The power of this modelling technique is its ability to generate ad hoc semantic relationships in response to an extemporaneous linguistic or conceptual situation.

The product of this approach is a generalisable computational linguistic methodology, capable of taking input in various forms, including word groupings and sentential context, and dynamically generating output from a broad base model of word co-occurrence data. To demonstrate the versatility of the method, this thesis will present competitive empirical results on a range of established natural language tasks including word similarity and relatedness, metaphor and metonymy detection, and analogy completion. A range of techniques will be applied in order to explore the ways in which different aspects of projected geometries can be mapped to different semantic relationships, allowing for the discovery of a range of lexical and conceptual properties for any given input and providing a basis for an empirical exploration of distinctions between the semantic phenomena under analysis. The case made here is that the flexibility of these models and their ability to extend output to evaluations of unattested linguistic relationships constitutes the groundwork for a method for the extrapolation of dynamic conceptual relationships from large-scale textual corpora.

This method is presented as a complement and a counterpoint to established distributional methods for generating lexically productive word-vectors. Where contemporary vector space models of distributional semantics have almost universally involved either the factorisation of co-occurrence matrices or the incremental learning of abstract representations using neural networks, the approach described in this thesis preserves the connection between the individual dimensions of word-vectors and statistics pertaining to observations in a textual corpus. The hypothesis tested here is that the maintenance

of actual, interpretable information about underlying linguistic data allows for the contextual selection of non-normalised subspaces with more nuanced geometric features. In addition to presenting competitive results for various computational linguistic targets, the thesis will suggest that the transparency of its representations indicates scope for the application of this model to various real-world problems where an interpretable relationship betweendata and output is highly desirable. This, finally, demonstrates a way towards the productive application of the theory and philosophy of language to computational linguistic practice.

Glossary

base space A high dimensional, sparse vector space of word-vectors, delineated in terms of dimensions of co-occurrence statistics.

context The situation – environmental, cognitive, perceptual, linguistic, and otherwise – in which an agent finds itself and applies language to meaning.

contextual input A set of words characteristic of a conceptual category or semantic relationship used to generate a subspace for the modelling of semantic phenomena.

dimension selection The process of contextually choosing a subset of dimensions in order to project a subspace from a base space.

co-occurrence The observation of one word in proximity to another in a corpus.

co-occurrence statistic A measure of the tendency for one word to be observed in proximity to another across a corpus.

co-occurrence window The boundary defining the proximity within which two words are considered to be co-occurring, typically a distance in terms of words within a sentence.

methodology The process of building base spaces from observations of co-occurrences within a corpus and contextually projecting subspaces through dimension selection.

model An application of methodology to a particular linguistic task or experiment, sometimes including task specific statistical analysis techniques.

subspace A context specific lower-dimensional projection from a base space, effectively mapping semantic relationships to a context by way of the geometric relationships between word-vectors.

word-vector A high-dimensional geometrically situated semantic representation of a word, constructed as an array of co-occurrence statistics.

Table of Contents

List of Figures

List of Tables

Chapter 3

Background

In this chapter, I will undertake the daunting task of outlining the scholastic background to my own research. I say this task is daunting because of the ambitious scope of my project: I intend to present a system which is both technically innovative and theoretically robust, and so I am faced with the double responsibility of providing an overview of the theory of concepts, representations, and semantics as well as a survey of ongoing work in the highly productive computational linguistic domain of distributional semantics. By achieving the right balance between theory and practice, I hope to lay the groundwork for a project that is suited for and enhanced by application to tasks developed within the field of natural language processing, but at the same time provides an empirical basis for making further theoretical commitments about the plausible operations of linguistic agents.

The theoretical background for my project in particular will lead to the development of an inventory of what Gallie (1956) has called *essentially contested concepts*, words and corresponding ideas that are more likely to invite debate and academic dissent than to offer resolution. As Deacon (2011) has put it in his biologically grounded account of the emergence of goal-directed behaviour, "Such concepts as information, function, purpose, meaning, intention, significance, consciousness, and value are intrinsically defined by their fundamental incompleteness," (ibid, p. 23). But, as Gallie points out in the context of the social sciences, these words are nonetheless important and can be useful components of a productive discourse, just so long as we are not overambitious in our claims to have arrived at some sort of conclusion about their objective definitions. Instead, I propose that the ideas of *information*, *meaning*, *creativity*, *representations*, and *concepts* should be viewed as boundary conditions for the empirical work that will be the primary focus of this thesis, delineating the conceptually territory from which my approach arises and to which it ultimately seeks to contribute. Rather than claiming to offer any particularly visionary insight into the complex and, in general, ancient questions that foment at the perimeter of my technical work, I hope to illustrate that my research is in communication with a robust philosophical tradition and could in principle provide an empirical basis for future contributions to this discourse. Sections 3.1, 3.2, and 3.3 will deal with this.

Then, with the theoretical apparatus of my research in place, Section 3.4 will outline the technical background for the computational implementation of lexical semantic modelling that I have developed. One of my goals in this Chapter is to map out a robust correspondence between the theory of language and mind to the practice of statistic semantic model building. As will be seen both in this chapter and later in my thesis when I offer more detailed background on the experiments I use to test my methodology, there has already been appreciable thought given on the part of computational linguists to the theoretical background supporting existing models and systems described in the literature, with cognitive linguists in particular providing a useful basis for conceptual modelling, and it is not my intention to suggest that my own contribution is in some sense conceptually superior. I do, however, believe that there are some novel and valuable connections made in this manuscript, in particular with the philosophical discourse surrounding matters of representation and intentionality as well as the pragmatic approach to conceptualisation. What we will finally reach by the end of the chapter is a starting point, situated in the familiar computational linguistic domain of distributional semantics, for considering how to apply theoretical insight into the contextuality of semantics to computationally tractable lexical representations.

3.1 Meaning Making

At its heart, this thesis is about the emergence of meaning from data, and in this regard it sits atop a tradition of analytic enquiry into the nature of being itself. The very question of how meaningfulness can come about in a material universe has been arguably the unifying theme of modern Western Philosophy, spanning from the *cogito* of Descartes (1911) to the phenomenology of Husserl (2001) and Heidegger (1962), by way of empiricism (Hume, 2000; Locke, 1997), transcendental idealism (Kant, 1996), pure idealism (Hegel, 1989), and intentionality (Brentano, 1995), to delineate just one of the countless pathways through the rich tradition of ideas about minds. Broadly speaking, I intend to present a philosophically motivated, empirically oriented project that, without making controversial commitments or overambitious overtures, sits comfortably with Wittgenstein's (1967) idea that "only the act of meaning can anticipate reality," (ibid, ¶188), which I will interpret to suggest that meaning is somehow properly in the world, not only in some immaterial, nominally mental space—but also that there really is such a thing as meaning, that it is not merely a convenient fiction of an otherwise behaviouralist ontology.

With this in mind, the project I describe here is broadly conversant with Floridi's (2011) pursuit of a *theory of strongly semantic information*, by way of which he arrives at a quantitative model of meaning.[1] The idea that observable data can be computational transformed into information is underwritten by the Information Theory of Shannon and Weaver (1949), which seeks without making any philosophical claims about knowledge or beliefs to formalise the measurement of what can be known in terms of the unexpectedness associated with sets of observations (see Pierce, 1980, for a thorough treatment). An early attempt to import technical insight from signal processing into the study of meaning can be found in Carnap and Bar-Hillel's (1952), who use Shannon-type metrics as the basis for quantifying the inferential properties associated with the semantic content of sentences, followed by Dretske (1981), who describes the formation of meaningful concepts in terms of the development of internal semantic structures that evolve to indicatively correspond with quantifiable informational situations in an environment. Subsequent forays in a *situation logic* designed to model semantic information content in a way which is simultaneously measurable and context specific (Barwise and Perry, 1983) have contributed to the resolution of computationally amendable formalisms, both in the tradition of Shannon and the semantics that have followed from Montague (1974), with the environmentally grounded approach to cognition which will be discussed presently.

At the more ambitious extent of the spectrum, the likes of Koch (2004) and Tononi (2008) have put forward theories attempting to quantify consciousness itself, generally in terms of the differentiable components of complex dynamic systems. *Consciousness*, however, is one of the aforementioned essentially contested terms, so instead of taking a stance here, I will take the easier route of simply acknowledging that there is a *hard problem* to be solved, to use the jargon of Chalmers (1996), and it should be perfectly possible to do good empirical work without necessarily taking sides in the fraught debate over the computability of the subjective experience of existence—or rather, perhaps an effective empirical approach comes about precisely from recognising the intractability of the debate in the first place. So here I will propose to use the notion of *creativity* as a kind of representative for the entire idea that being a cognitive agent has something to do with the production of meaning in reaction to the rampant stimulus provided by a dynamic and unpredictable cognitive *umwelt* (von Uexküll, 1957). In the spirit of Koestler (1964), then, and his model of creativity as "a new synthesis of previously unconnected matrices

[1] Unlike Fredkin (2003) and, more popularly, Bostrom (2014), Floridi navigates a middle way towards a computational model of semantics without committing to outright digital ontology.

of thought," (ibid, p. 182), I will offer a general definition of creativity as the act of meaning making in a universe of heterogeneous environmental data, and I will further assert that modelling this type of cognitive activity is, in a general sense, the target of my research.[2]

This then pushes my research into the broad domain of *computational creativity*, a field outlined in the seminal work of Boden (1990) and subsequently formalised in terms of "behaviour exhibited by natural and artificial systems, which would be deemed creative if exhibited by humans," (Wiggins, 2006, p. 206). The thrust of this work and the theory and practice that have sprung up around it involves treating creativity in terms of state spaces of combinatory components susceptible in the most productive cases to transformational transgressions of the rules for traversing the space, resulting in artefacts (and, arguably, processes) which can be evaluated in terms of their novelty and value (see Colton, 2008; Jordanous, 2012; Ritchie, 2007, among others for interesting theoretical work on the evaluation of computational creativity). If meaning making is to be construed in terms of creativity, and creativity is in turn modelled as a process of combination and composition, then at the root of the computational application of a theory of data, information, and meaning we encounter another essentially contested concept, namely, that of *representation*.

Representations have played a roll in philosophy of mind certainly since Descartes (1911) and Hobbes (1651), and by any but the most abstracted interpretation at least since Plato (1892)—perhaps they are a necessary passage in any movement towards a robust theory of mind (if, in fact, such a theory is even desirable—*cf* Rorty, 1979). The recent trend in philosophy, however, not to mention in empirically fastidious fields such as cognitive science and psychology, has been towards a resolute materialist reductionism, to such an extent that Rowlands (2010) reports that in the current cognitive scientific milieu, "even the word 'Cartesian' is often used as a term of abuse," (ibid, p. 12). This has been bad news for representations which, when applied to a theory of mind, can degrade into a homuncular regression that Dennett (1991) has described as the *Cartesian theatre*: if something is being represented, and something is doing the representing, who or what is at the receiving end of the process? The embodied and enactivist school of thought instigated by Maturana and Varela (1987) and pursued by, for instance, Haugeland (1993) and Thompson (2007), has led to the reanimation of discourse regarding the nature of mind from a perspective that does not take the *explanatory gap* (Levine, 1983) between what is subjectively experienced and what is objectively described for granted. Subsequently

[2]Creativity is itself, as Colton et al. (2014) have pointed out specifically in the context of computational approaches, an essentially contested concept, but, in the spirit of Gallie (1956), I will presume that there is significant value in identifying creativity as a boundary condition of sorts for the range of activities that I wish to explore without reaching a conclusive definition of the concept.

Gelder (1995) has outlined the premise of a mathematically tractable model of non-representational cognitive systems described in terms of dynamically coupled differential equations, while the emergentist system theory of biosemioticians like Kauffman (1995), Hoffmeyer (1997), and Pattee (2001) have provided fertile material for the sophisticated and evolutionarily plausible cognitive model of Deacon (2011).

But these anti- or post-representationalist approaches to cognition tend to unravel a bit when it comes to saying anything about language. In this particularly well travelled domain, the type theory of Whitehead and Russell (1927) and Church (1940) still holds a certain sway, with the subsequent formalisms of intensional semantics (Fox and Lappin, 2005) treating language as an ineluctably symbolic phenomenon. As such, there is an overt representationalism that is more or less necessarily at play in the symbolic commitments made by any sustainable theory of semantics, particularly in the context of natural language. Regardless of whether the representations in question are strictly in the mind, a theory promoted by Fodor (2001), or are in some sense in the world in line with the philosophy of Putnam (1975), it becomes difficult to imagine an operational model of semantics which doesn't fall back on structures which are to some extent extracted from the reality that they denote.

McGregor et al. (2014) have presented something of a start towards addressing or, perhaps more to the point, avoiding this issue (and the issue has been subsequently explored by Coeckelbergh (2016), in both cases specifically with reference to computational creativity). The idea put forward there is that, in the context of computational creativity in particular, it should be acceptable to take seriously the evident efficacy of talking about representations when talking about cognitive processes without necessarily making a commitment to the fundamental reality of such representations. I will stick to this position in the work presented in this thesis: by starting with the assumption that representations are a useful, maybe even necessary, component when talking about semantics and meaning, I maintain that we might eventually arrive at a more satisfying resolution of why this kind of structure has held such sway over the modern Western tradition of analytic philosophy in particular, and whether this influence is fundamental or just incidental. I don't claim to come close to actually answering this hard question, but I do think that there will be apparent merit in taking my methodology seriously as an empirical tool for gaining some sort of theoretical traction in this regard. So, in summary, in the following chapters, I will be describing a methodology which traffics in a particular theoretically motivated variety of meaning bearing representation, without making any commitment as to the essentialism of that device; the desideratum of these representations is that they be susceptible to the environmental situatedness that is clearly an important component of any effective cognitive or linguistic model. My contention here is that sound theoretical grounding based on insight from cognitive science should grant my models a degree of at least temporary immunity from accusations of dualism.

3.2 Concepts

As Searle (1983) points out, representations have intentional content: they have to be about things, whether or not they take the form of materially or abstractly transportable entities like words or icons. The intentionality of representations invites the addition of another term to our growing catalogue of essentially contestable concepts, this time the word *concept* itself, which I will take to refer to the cognitive aspects of the things indicated by representations. The idea that concepts are interactive structures of the mind (Fodor, 2008; Margolis and Laurence, 2007) has been productive in aligning cognitive science with computational modelling (Boden, 2006). If concepts can be modelled as rule bound composite symbolic entities, then a symbol manipulating, constraint satisfying device should provide the right kind of architecture for simulating productive interactions between conceptual representations. This type of modelling has proven practically effective in, for instance, the structured ontology of Lenat (1995) and the graph theoretical work of Sowa (2006).

There is discord afoot, however, amongst researchers interested in modelling concepts, parallel to a certain extent to the debate over mental representations outline in the previous section. The net result of this tension has been the generation of a kind of negative space: where philosophers like Fodor and Pylyshyn (1988) have made a convincing case against treating concepts as associationist networks, more recent cognitive scientific research from the likes of Hutto (2001) and Chemero (2009) offers a likewise compelling rebuke to any theory of mind that falls back on a framework of symbolic conceptual representations. What remains is a clearly developed picture of what cannot constitute a concept in a cognitive model, but a much more murky impression of what positively does count as a thought or a perception and so forth. A remedy of sorts is offered by Gibson (1979), with his view of cognition in terms of the direct perception of environmental *affordances* of opportunities for action in a situation. Clark (1997) has expanded upon this to arrive at a notion of *action-oriented representations* which outsource much of the computational load of conceptualisation to the physical and spatial domain of a cognitive agent's environment.

Here Kant (1996) has proved to be, perhaps not surprisingly, especially profound: the Kantian notion of a domain of *conception* that is supervenient upon an underlying field of emphintuition which is in turn grounded in the essentially geometric nature of reality provides a philosophically robust starting point for a spatial model of conceptualisation. By positioning conceptual models geometrically, the components of concepts which give them the composability that symbolic models afford while at the same time maintaining some degree of contact with the potentially physical context of space. The work of Gärdenfors (2000) is particularly germane here, and will serve as a primary point of reference for the methodology that I present in this thesis. By modelling concepts in terms of convex regions within conceptual spaces defined by interpretable dimensions representing attributes of the concepts themselves, Gärdenfors provides a plausible intermediary between the low-level stimulus to which a cognitive agent is exposed in an environment and the high-level symbols that become the representational currency of thought and communication: stimuli provide the data which becomes the values defining the points in a symbolically realisable conceptual space. More recent work has explored the way that a conceptual space model can be applied to lexical semantics in order to provide a geometric grounding for the categorical nature of language composition (Gärdenfors, 2014).

The environmental grounding of a conceptual model further provides a mechanism for understanding the important role of *context* in cognition. Here Barsalou's (1992) work modelling concepts in terms of *frames* offers a valuable perspective on the way that particular conceptual schemes are activated in response to situations in the world. Barsalou's approach facilitates notions of prototypicality and periphery that emerge in the course of online, context sensitive conceptualisation, once again at least hinting at a spatial component of this cognitive framework. Also of note is the *conceptual blending* approach of Fauconnier and Turner (1998), which makes use of a spatial theory of mind to develop a framework of conceptualisation as integration between frames of representation. This approach has been applied in the domain of computational creativity in particular, to the generation of language in the case of Veale (2012) and to automatic software generation by Znidarsic et al. (2016). And it is also worth mentioning the *global workspace* framework proposed by Baars (1988), which models cognition as a multi-agent system in which functional components compete and collaborate to forge a situated cognitive gestalt: this approach has been adopted by Shanahan (2010) in his work on cognitive robotics and by Wiggins (2012) again in the domain of computational creativity. A common and significant theme here is the dynamism and distribution inherent in all these approaches, contravening conceptual models that resort to static and hierarchical representational regimes.

Ultimately, I think we have to take seriously Davidson's (1974) case against the idea of conceptual models in the first place. Davidson's point is not so much that there is no such thing as a concept – that would be a fatuous claim – as that concepts are an artefact of the way that cognitive, and in particular linguistic, agents use meaning bearing representations to structure thought and communicate about experience. At first glance this view of concepts might appear as facile as the denial of the existence of concepts is fatuous: obviously concepts have something to do with having thoughts, and it is probably impossible and certainly pointless to imagine a universe in which there are concepts but there are not cognitive agents. But the subtlety of Davidson's point is that there is a dynamic between conceptual models and representational structures which belies any kind of relationship of supervenience and complicates attempts to explain cognition in terms of levels of materialistic abstraction—as, in their own distinguished and insightful ways, Floridi (2011) and Deacon (2011) have each done. This dynamic turn invites a consideration of language as a concept supporting structure, and so sets us up for the next section of this survey of the established theory and practice surrounding my own work.

What we are then left with is the impetus for a computational approach which should be situationally dynamic and contextually sensitive. With this in mind, the methodology that is the focus of this thesis will be characterised by semantic representations that are designed to be understood as conceptually productive, contextually generated perspectives on spaces defined in terms of statistical data about language use. By using quantitative data to project representations into spaces that can be manipulated in an open ended way in response to a context which in principle can be arbitrarily defined, I will seek to mirror a theory of situated cognition permitting for the emergence of concepts in the course of the dynamics between agent and environment. As with my treatment of semantic representations themselves, I don't claim to be describing a methodology for conceptual modelling which is necessarily plausible on the level of physical or biological processes; instead I take certain assumptions about conceptual spaces for granted, and so there is an element of abstraction necessarily at play here. Once again, though, my stance is that allowing for some *a priori* assumptions about what is conceptually permissible provides a sound basis for getting on with the practical work of designing data driven experiments based on conceptual models and then turning around to apply the experimental outcomes to a productive reconsideration of theoretical assumptions.

3.3 Words

What has come to be known as the Cognitive Revolution finds its origin in, among other things, Chomsky's (1959) pointed denouncement of Skinner's (1957) attempt to apply psychological behaviouralism to the study of language. Chomsky's point is that language can only properly be understood as a specialised faculty that is in some way, more than just a mode of stimulus and response, internal to the cognition of a linguistic agent: in order to effectively model language, we have to build some sort of notion of minds populated by cognitive content and attendant intentionality into the equation. For Chomsky and some of his acolytes, the logical extension of this view has been the development of a programme founded on the idea that language is itself an inborn characteristic peculiar to human cognition, certainly neurologically specific and quite possibly genetically encoded (Chomsky, 1986; Fodor, 2001; Pinker, 1994). A significant component of this project has been the development of various formulations designed to systematically encapsulate the conditions generally determining the parameters of natural languages, but for every attempt to categorically describe the particulars of human communication, linguistic anthropologists such as Levinson (2001) and Everett (2005) turn around and discover a group of language users who provide the exception which in the case of a scientific approach to language really does disprove the rule.

The movement against Chomskyan nativism has tended to swing towards what is arguably an even more fundamentally cognitive theory of language, often characterised by interpretations of Sapir (1970) and Whorf (2012) as a jointly declaring that language is, to a greater or lesser extent, actually the foundation upon which thought and attendant cultural spheres are built. More generally, the field of cognitive linguistics has emerged in response to the mainstream linguistic stance supporting theories of universal grammars, and a battery of interrelated linguistic models have emerged from the idea that language is, along with various other aspects of human behaviour, broadly wrapped up in and symptomatic of the general condition of having a mind rather than a compartmentalised cognitive faculty (Croft and Cruse, 2004). Of particular relevance here is the *cognitive grammar* of Langacker (1987), which proposes to overcome the divide between syntax and semantics by treating phonological and morphemic components of language as inextricably intertwined with semantics in ways that supersede evident distinctions across what Langacker calls *grammatical classes* (conventionally, parts of speech, basically). Also of note are the *image schema* of Lakoff (1987) and Johnson (1990), who, by focussing their analysis on the way that preposition usage in particular suggests distinct culturally specific embodied models of the world, developed environmentally and biologically grounded frameworks for productive semantic composition.

A general methodological commitment of cognitive linguistics is the qualitative analysis of instances of language use applied to the development of critically rich models of how conceptual and linguistic representations interface in the course of situated cognition. It should not be presumed, however, that cognitive linguists take semantic and conceptual representations to be identical or even isomorphic, and in fact Evans (2009) argues specifically that it is the nebulousness of the relationship between these domains that gives language its particular qualities of looseness and ambiguity by which lexical representation can be deployed in context specific ways to achieve an open-ended expressivity. This aspect of semantics is particularly evident in the phenomenon of figurative language, and the study of metaphor has been an especially successful pursuit here, with a valuable compendium of the productive era from the late 1970s through the 1980s assembled by Ortony (1993). Exemplary theoretical work grounding the seemingly unlimited generative capacity of figurative language in a robustly cognitive approach to linguistics includes the *interaction* view of Black (1955,7) and the *reconstructivist* stance of Ortony (1975). It is the *cognitive metaphor* approach of Lakoff and Johnson (1980), however, which stands out most of all here, not least because it has provided the most consequential material for latter day computational research into metaphor classification and interpretation (Shutova, 2015). The description of metaphor in terms of isomorphic mappings between conceptual domains lends itself to precisely the type of symbolic manipulation of information structures that have characterised traditional AI, and, as it turns out, can also provide a theoretical grounding for sophisticated statistical modelling of lexical semantics (Shutova et al., 2012).

Statistical approaches to lexical semantic modelling will be surveyed in more detail in the following section, but a brief overview of information processing applications of the theory surrounding metaphor seems appropriate here. Some early computational approaches to metaphor maintained an essentially formal character: van Genabith (2001) proposed a type theoretical model for describing metaphor. Information processing approaches have, though, been by and large data-driven, understandably utilising the processing power of symbol manipulating machines—and these data-driven approaches have generally had some sort of connection with the cognitive linguistic stances on metaphor. So, for instance, Thomas and Mareschal (1999) describe an information processing network which selectively projects features, inspired by the previously mentioned interaction view of metaphor developed by Black (1977). In terms of theoretical grounding, Shutova (2010) identifies the *selectional preference violation* approach of Wilks (1978) as especially influential, perhaps because it was formulated specifically as an information processing mechanism. A notable early effort from Fass (1991) is derived from this theoretical background, with correspondences in the selectional preference of the arguments of verbs used to detect metonymic versus metaphoric uses of language.

The mainstream of metaphor modelling has subsequently been characterised by symbol manipulating approach and, in the spirit of the conceptual metaphor model, has involved mapping between conceptual schemes (Indurkhya, 1997), often domain specific, with the underlying assumption that mappings between domains correlates with the conceptual metaphor model (Narayanan, 1999). Typical symbolic approaches to metaphor modelling involve the construction of an ontology defined by features which can be mapped between elements. The ATT-Meta system (Lee and Barnden, 2001), with its faculty for backchaining inferences across conceptual domains, is exemplary, and has furthermore been expanded into a metaphor generating system employing a combination of distributional semantic and incremental grammar techniques (Gargett and Barnden, 2013). Other symbolic approaches are notable for their recourse to pre-formulated knowledge bases such as WordNet (Veale et al., 2015), or the web at large in conjunction with other resources (Veale and Hao, 2007).

Symbolic approaches have tended to focus on the interpretation of metaphor by way of models of trans-conceptual mappings, but in another aspect of computational work, that of metaphor identification, statistical approaches have proved particularly effective.[3] An early example is the TroFi model of Birke and Sarkar (2006), which uses a clustering algorithm trained on a set of tagged sample sentences to disambiguate between literal and non-literal verb use, followed by Utsumi (2011), who explores clustering in the context of distributional semantics. Indeed, many of these statistical approaches (see Turney et al. 2011, Dunn 2013 for a comparison of distributional semantic and symbolic models, Shutova et al. 2012 for an overview of statistical models in particular) have employed the techniques of distributional semantics, which will be discussed in the next section: here Kintsch's (2000) model of metaphoric interpretation as a contextually selective traversal of the space between word-vectors is seminal. A notable recent instance of a statistical model for metaphor identification involving an application of compositional distributional semantics is described by Gutiérrez et al. (2016), of particular note here as the dataset presented by those authors will be used to evaluate the model at the heart of this thesis (see Chapter **??** for a more detailed description). Returning to the cognitive linguistic foundations of computational approaches to metaphor, Tsvetkov et al. (2014) go so far as to propose that their results derived from the statistical construction of what they construe as conceptual features associated with lexical representations "support the hypothesis that metaphors are conceptual, rather than lexical, in nature," (ibid, p. 248).

There is another theoretical twist which must be mentioned here, however, and it comes once again from Davidson (1978), this time by way of his controversial claim that the meaning of metaphoric propositions should always be taken at face value. Part of Davidson's point is that there is a pragmatic distinction to be drawn between what the metaphor means, which is to some extent in the language, and what the metaphor communicates, which is on the other hand in the world.[4] The presumption in both

[3]Shutova (2013) suggests that computational identification and interpretation of metaphor, in line with psychological analysis, should be considered a joint task.

[4]Davidson's account, which is famous or perhaps notorious amongst theoretical linguists, is notable in its absence from the computational literature, though it has recently been acknowledged at least in

conventional semantic views of metaphor such as Searle's (1979) as well as the more strongly cognitivist stances discussed above is that metaphor necessarily involves the projection of some aspect of meaning from one conceptual domain to another, but the point that Davidson raises is that there is a limit to the cognitive content that can be propositionally conveyed by language, and metaphor often reveals that limit. To borrow a popular example from the discourse surrounding relevance theory (Carston, 2012; Jr and Tendahl, 2006, for example), there is a lurking breakdown in interpretation when we try to apply any sort of transference view of metaphor to a statement such as "my boss is a bulldozer": presuming a small degree of contextual knowledge, we might easily understand that the speaker means the boss in question is inappropriately insensitive or aggressive in dealing with employees—but it is hardly clear what actual properties of BULLDOZER are transferred to BOSS, particularly in a situation which might very well not even be physical.

To address this issue, Carston (2010) proposes that metaphor necessarily involves the generation of *ad hoc concepts* that come about in the process of making a lexical mapping from one domain of encyclopaedic knowledge to another. Drawing on Barsalou's (1993) notion that language produces concepts in a way that is inherently *flexible* and *haphazard*, *ad hoc* concepts offer a relevance theoretical account of the way in which language always pragmatically, situationally specifies the semantic content of an utterance (Sperber and Wilson, 1995). This accommodates the *deflationary* view of metaphor put forward by Sperber and Wilson (2012), which holds that metaphor merely occupies an especially inferential extent of a spectrum of meaning making and interpreting activities. At stake here is the idea that language is not so much a system for codifying propositions about the world as a mechanism for achieving optimal communication of cognitive content, with the important proviso that cognition itself is primed for a perpetually unfolding contextualisation of the environmental stimuli available to an agent. This ultimately means that metaphor is able to be more than just a highly efficient way of encoding propositions about concepts; it can, even in relatively mundane instances, extend itself into domains bordering on the phenomenological, a stance eloquently summed up by Reimer (2001) in her apologetic exegesis of Davidson: "For the goal of the metaphor-maker is not to get the hearer to see that something is the case, to grasp some deep and subtle truth, but to see something in a certain way, and seeing something in a certain way is simply not the sort of thing that can be given literal expression," (ibid, p. 150).

With all this in mind, we arrive at a further specification for the boundary conditions of our computational semantic model: in addition to being a representational system with a capacity for summoning context specific relationships between lexical semantic

passing by Veale (2016).

entities, it should also be able to generate new conceptual representations in an *ad hoc* manner. This implicates the modelling of conceptual spaces that are not merely invoked by the process of specification inherent in communication, but actually generated in the course of lexical dynamics. And the situated, even arbitrary production of conceptual relationships in turn suggests, beyond just the activation of existing or implicit networks of association between semantically tractable entities, the online creation of entirely new connections and correspondingly of new ideas: put simply, the open-ended generation of conceptual spaces is the machinery of meaning-making. It seems more or less impossible to imagine a regime of strictly symbolic representations which could fulfil these requirements, because symbols necessarily come with the logic and extent of their combinatory potentials, setting the constraints for the state space of their potential for interactive conceptualisation, more or less built in. Instead, I propose that a statistical approach, in which lexical semantic representations are defined in terms of observations of symbols in use rather than rules applied directly to symbols, will offer the right kind of flexibility and dynamism for modelling the situated nature of concepts and the rampant looseness inherent in the relationship between words and objects of the mind.

3.4 Data

Finally, arriving at the technical background for the instantiation of the system of context sensitive, semantically productive representations outlined above, the research described in this thesis is grounded in recent and ongoing success in the paradigm of *distributional semantics*. The tradition of word-counting in order to predict sequences in language traces its roots back to the fastidious work of Andrei Markov, who tabulated co-occurrences of characters in Pushkin's *Eugene Onegin* by hand (Basharin et al., 2004), and Shannon and Weaver (1949) propose a comparable application in their seminal work on information theory. The idea of applying co-occurrence statistics to semantic applications is central to Harris's (1954) work examining "meaning as a function of distribution," (p. 155); the various consequent formulations of the *distributional hypothesis* have been outlined by Sahlgren (2008), with Pantel's (2005) asseveration that "words that occur in the same contexts tend to have similar meaning," (ibid, p. 126) being representative.[5] Theoretically speaking, computational linguists have ambitiously sought to ground dis-

[5]Scholars frequently cite Firth's (1959) quip "you shall know a word by the company it keeps," (ibid, p. 179) as being foundational in the field. I contend that Firth was referring in this passage specifically to the study of idiomaticity, particularly the way that idioms ossify culturally through repeated use, and this in the context of a larger proposal for a heterogeneous approach to the study of linguistics more in line with the comprehensive emergent view of MacWhinney (1998) rather than anything that could be construed in terms of a computational, word-counting practice. All the same, the quote has a nice ring to it and, taken out of context, serves its purpose.

tributional semantics in the formal semantics of Frege (Baroni et al., 2014a) or indeed in the pragmatics of Wittgenstein (Grefenstette and Sadrzadeh, 2011).

Rather than indulge in speculation of what Wittgenstein might have done with a computer, I will propose a perhaps even less likely candidate as the philosophical forbearer of word-counting as a productive applied linguistic practice: the semiotics of Peirce (1932), which maintain that the very physiognomy of meaning bearing structures, or *signs* in Peirce's parlance, are semantically productive by way of their very physiognomy, and that they gain this productive structure through their ongoing contact with their environment. From his own analysis of Peirce, Eco (1976) extrapolates a notion of *unlimited semiosis* by which signs participate in an infinite regression of semantic productivity, with one sign becoming the substrate for the constitution of a subsequent sign. This begins to look, in an abstract way, a bit like the distributional semantic regime, where the sentential context in which words are found becomes the substance of interactive lexical semantic representations. Another historical touchpoint is, as Miller and Charles (1991) have pointed out, the *salva veritate* of Leibniz, by which, in terms of logical formalisms, terms are considered to be synonymous if they can be universally interchanged in logical expressions without changing the truth values of the expressions. Exporting this notion to the domain of computational linguistics, we arrive at the central dogma of distributional semantics, namely, that words can be modelled in terms of observations of their co-occurrence tendencies across large scale corpora, and furthermore that words with similar profiles can be interpreted as being likewise semantically associated.

Practically speaking, early work from, for instance, Salton et al. (1975) suggested that the information content of documents could be effectively indexed by representing them as points in a vector space whose dimensions correspond to weighted measures of word frequency within a given document. Schütze (1992) extends this insight to represent words as vectors defined by the frequencies with which they are observed to co-occur with other words in a corpus, and uses angular measures from the consequent vector space as grounds for disambiguating the senses of polysemous words. An important result of modelling words in terms of their co-occurrence profiles is that two words which have never been observed in proximity to one another might nonetheless turn out to be very close in the model and therefore very similar to each other: so, for instance, we can imagine a language in which the words for CAT and DOG are prohibited from ever being used in the same sentence, but we might still discover a semantic correspondence between the concepts because their signifiers tend to have similar patterns of usage. The conversion of raw word counts into weighted statistics, perhaps most basically through the application of term-frequency, inverse-document-frequency type metrics (Salton and Buckley, 1988) but more typically in more recent applications with information theoretical functions (Turney, 2001), has produced particularly productive co-occurence based lexical semantic representations. The geometric efficacy of passing co-occurrence statistics through logarithmic functions will be discussed in Chapters **??** and **??**. The end product of this type of approach is fundamentally that words are mapped into spatial relationships with one another, where the geometry of the space itself is to a greater or lesser extent semantically productive, and authors such as Landauer and Dumais (1997) have explored some of the psychological and philosophical ramifications of this.

The vector space approach to distributional semantics has subsequently evolved into a productive computational programme. The distributional semantic methodology usually involves the selection of a corpus, the traversal of this corpus in order to tabulate the counts of co-occurrence terms within a certain proximity of target words (typically defined in terms of a window of k words around each observation of a target word), the application of a weighting function to the resulting co-occurrence matrix, and the projection of the weighted vectors into a space (see Turney and Patel 2010 and, more recently, Clark 2015 for comprehensive overviews). Bullinaria and Levy (2012) have reported comparative results based on a variety of weighting schemes, most notably *positive pointwise mutual information* (PPMI), an information theoretical metric designed to build sparse matrices capturing the most semantically salient co-occurrence features of word-vectors. Where PPMI simply disregards co-occurrences that are observed at a frequency below the overall corpus average, Levy et al. (2015) explore a slightly more subtle techniqe of shifting their co-occurrence statistics to avoid massively negative logarithms; a similar metric will be the basis for my own methodology. The construction of distributional semantic models also often involves an additional step of dimensional reduction by way of, for instance, principal component analysis, with a particularly notable technique involving singular value decomposition described by Deerwester et al. (1990).

Distributional semantic models have evolved out of the practical requirement for effective and efficient document retrieval based on textual queries, but the linguistic tasks subsequently tackled have included entailment (Baroni et al., 2012; Geffet and Dagan, 2005; Rimell, 2014), word sense disambiguation (Kartsaklis and Sadrzadeh, 2013; Schütze, 1998), and sentiment analysis (dos Santos and Gatti, 2014; Malandrakis et al., 2013), among other things. A particularly interesting development has been the use of linear algebraic operations on representations to facilitate language composition (Mitchell and Lapata, 2010). By treating, for instance, nouns as word-vectors and adjectives as tensors, Baroni and Zamparelli (2010) describe a model for projecting adjective-noun phrases into a vector space in which these compound linguistic entities can be compared using the same approaches applied to word-vectors. Borrowing from the mathematical arsenal of quantum mechanics, Coecke et al. (2011) conceive a correspondence between distributional semantics and formal semantics, modelling syntactic elements as vectors and tensors based on observations across a corpus that map to category theoretical components of a grammar, pushing whole sentences into vector spaces allowing for comparison between sentences and the assignment of truth values. The import of all of this is, once again, that the modelling of semantic units using high dimensional representations provides a productive and computationally tractable grounding for a variety of linguistic phenomena.

The development of high powered computers and the related advent of massive corpora of digitised textual data has facilitated another turn in the distributional semantic programme: the application of neural networks to data driven semantic modelling. Bengio et al. (2003) is an early proponent of this approach, demonstrating that the application of iteratively learned word-vectors consisting of abstract features is an effective mechanism for language modelling, followed by Collobert and Weston (2008), who use a convolutional neural network to build a vector space model suited to learning to perform a number of supervised and semi-supervised linguistic tasks including semantic modelling, language modelling, and sentence parsing. And the contribution of Mikolov et al. (2013a,1,1), dubbed `word2vec`, has been one of the most widely discussed developments in the field in recent years, offering up a highly generalisable set of models with particularly remarkable capacities for modelling the semantically significant phenomenon of analogy, which will be discussed in more detail in Chapter **??**.

The dichotomy between co-occurrence statistic based models, almost always complemented with some dimension reduction technique such as a principal component analysis, and neural network approaches has led to a productive tension in the field, summarised by Baroni et al.'s (2014b) in terms of *counting* to derive statistically defined word-vectors versus *predicting* what have sometimes been called *word embeddings* using a neural network—though it should be noted that both methodologies necessarily act on observations of word co-occurrences made in the course of the traversal of a corpus, and both types of model have been successfully configured for the kind probabilistic output involved in, for instance, language modelling. And, where Baroni et al. ultimately decide that neural network based approaches offer a more robust extrapolation of semantic representations from corpus data, Levy and Goldberg (2014) have argued that the superficial differences between the two broad methodlogies can be understood in terms of decisions regarding the tuning of the extensive range of hyperparameters inevitably associated with either type of model. Along similar lines, one of the main findings of this thesis, and a motivation for the methodology I've developed, is that, once a layer of removal from the data has been applied to statistical models through for instance singular value decomposition, they, like neural network models, become immune to context specific manipulation, because their dimensionality becomes abstract and uninterpretable.

One consequence of the collegial arms race between the two approaches has been the development of increasingly task specific systems, often coupling distributional semantic models with heuristics involving the identification of syntactic patterns or the extraction of information from pre-formulated knowledge bases. In response to this, Baroni and Lenci (2010) have described an ensemble of vector space models packed into a tensor space of potential relationships between lexical entities—a model of models of sorts, capable of selectively activating the appropriate component of its representational hyperspace based on an assessment of the task at hand. This is well motivated, and I have sought to develop a similarly generalisable methodology, but in the case of my research the generalisability arises from the ability of my models to selectively project an astronomical range of context specific semantic subspaces rather than from an extra layer of model specification. In practice, my methodology will be tested against a battery of existing tests designed by fellow researchers in the field of computational linguistics, including word relatedness (Finkelstein et al., 2002), word similarity (Hill et al., 2015), metaphor classification (Gutiérrez et al., 2016), semantic type coercion (Pustejovsky et al., 2010), and analogy completion (Mikolov et al., 2013c).

So finally we arrive at something like a way forwards towards the computational modelling of context sensitive lexical semantics. Distributional semantics provides a mechanism for the production of dynamically interactive representations based on observations of large scale textual data, offering up a malleable lexicon suited to the rampant contextualisation indicated by theoretical insight into concept production. To chart a passage through the territory mapped throughout this chapter, then, statistics reflecting the co-occurrences of words in a large scale corpus will serve as the data substantiating the informational character of dynamic lexical semantic representations which, in their interactions, will be projected into conceptually interpretable spaces that are in turn reflective of the evidently representational character of meaning making. With this apparatus in place we can now move on to the task of a theoretical description of my own methodology in Chapter **??**, followed by a technical description of the consequent computational in Chapter **??**.

References

Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

Baroni, M., Bernardi, R., Do, N., and Shan, C. (2012). Entailment above the word level in distributional semantics. In *EACL 2012, 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 23–32.

Baroni, M., Bernardi, R., and Zamparelli, R. (2014a). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technology*, 9:241–346.

Baroni, M., Dinu, G., and Kruszewski, G. (2014b). Don't count, predict! In *ACL 2014*.

Baroni, M. and Lenci, A. (2010). Distributional memory: A general framework for distributional semantics. *Computational Linguistics*, 36(4).

Baroni, M. and Zamparelli, R. (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*, pages 1183–1193.

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In Lehrer, A. and Kittay, E. F., editors, *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*, pages 21–74. Lawrence Erlbaum Associates, Hillsdale, N.J.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: manifestations of a compositional system of perceptual symbols. In Collins, A., Gathercole, S., and Conway, M., editors, *Theories of memory*, pages 29–101. Lawrence Erlbaum Associates, London.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. MIT Press, Cambridge, MA.

Basharin, G. P., Langville, A. N., and Naumov, V. A. (2004). The life and work of a.a. markov. *Linear Algebra and its Applications*, 386:3–26. Special Issue on the Conference on the Numerical Solution of Markov Chains 2003.

Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155.

Birke, J. and Sarkar, A. (2006). A clustering approach for nearly unsupervised recognition of nonliteral language. In *Proceedings of EACL-06*, pages 329–336.

Black, M. (1955). Metaphor. In *Proceedings of the Aristotelian Society*, volume 55, pages 273–294.

Black, M. (1977). More about metaphor. In Ortony, A., editor, *Metaphor and Thought*, pages 19–41. Cambridge University Press, 2nd edition.

Boden, M. A. (1990). *The Creative Mind: Myths and Mechanisms.* Weidenfeld and Nicolson, London.

Boden, M. A. (2006). *Mind as Machine: A History of Cognitive Science.* Clarendon, Oxford.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press, Oxford, UK, 1st edition.

Brentano, F. (1974/1995). *Psychology from an Empirical Standpoint.* Routledge, London. Translated by Antos C. Rancurello and D. B. Terrell and Linda L. McAlister.

Bullinaria, J. A. and Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and svd. *Behavior Research Methods*, 44(3):890–907.

Carnap, R. and Bar-Hillel, Y. (1952). An outline of a theory of semantic information. Technical Report 247, Research Laboratory of Electronics, MIT.

Carston, R. (2010). Metaphor: Ad hoc concepts, literal meaning and mental images. 110(3):297–323.

Carston, R. (2012). Metaphor and the literal/nonliteral distinction. In Allan, K. and Jaszczolt, K. M., editors, *The Cambridge Handbook of Pragmatics*, pages 469–492. Cambridge University Press.

Chalmers, D. J. (1996). *The Conscious Mind.* Oxford University Press.

Chemero, A. (2009). *Radical Embodied Cognitive Science.* The MIT Press, Cambridge, MA.

Chomsky, N. (1959). A review of b. f. skinner's verbal behavior. *Language*, 35(1):26–58.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origins, and Use.* Praeger, New York, NY.

Church, A. (1940). A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5(2):56–68.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again.* MIT Press, Cambridge, MA.

Clark, S. (2015). Vector space models of lexical meaning. In Lappin, S. and Fox, C., editors, *The Handbook of Contemporary Semantic Theory*, pages 493–522. Wiley-Blackwell.

Coecke, B., Sadrzadeh, M., and Clark, S. (2011). Mathematical foundations for a compositional distributed model of meaning. *Linguistic Analysis*, 36(1-4):345–384.

Coeckelbergh, M. (2016). Can machines create art? *Philosophy & Technology.*

Collobert, R. and Weston, J. (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.

Colton, S. (2008). Creativity versus the perception of creativity in computational systems. In *Proceedings of the AAAI Spring Symposium on Creative Intelligent Systems*.

Colton, S., Cook, M., Hepworth, R., and Pease, A. (2014). On acid drops and teardrops: Observer issues in computational creativity. In Kibble, R., editor, *Proceedings of the 50th Anniversary Convention of the AISB*.

Croft, W. and Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge University Press.

Davidson, D. (1974). On the very idea of a conceptual scheme. In *Proceedings and Addresses of the American Philosophical Association*, volume 47, pages 5–20.

Davidson, D. (1978). What metaphors mean. In *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford, 2nd edition.

Deacon, T. W. (2011). *Incomplete Nature: How Mind Emerged from Matter*. W. W. Norton & Company, New York, NY.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal for the American Society for Information Science*, 41(6):391–407.

Dennett, D. C. (1991). *Consciousness Explained*. The Penguin Press, London.

Descartes, R. (1641/1911). *The Philosophical Works of Descartes*. Cambridge University Press. Translated by Elizabeth S. Haldane.

dos Santos, C. and Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 69–78, Dublin, Ireland.

Dretske, F. I. (1981). *Knowledge and the Flow of Information*. CSLI Publications.

Dunn, J. (2013). Evaluating the premises and results of four metaphor identification systems. In Gelbukh, A., editor, *Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part I*, pages 471–486. Springer Berlin Heidelberg.

Eco, U. (1976). *A Theory of Semiotics*. Indiana University Press, Bloomington.

Evans, V. (2009). *How Words Mean: Lexical Concepts, Cognitive Models, and Meaning Construction*. Oxford University Press.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in pirahã: Another look at the design features of human language. *Current Anthropology*, 46(4):621–646.

Fass, D. (1991). Met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.

Fauconnier, G. and Turner, M. (1998). Conceptual integration networks. *Cognitive Science*, 22(2):133–187.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., and Ruppin, E. (2002). Placing search in context: The concept revisited. *ACM Tran-*

sanction on Information Systems, 20(1):116–131.

Firth, J. R. (1959). A synopsis of linguistic theory, 1930-55. In Palmer, F. R., editor, *Selected Papers of J. R. Firth 1952-59*. Indiana University Press.

Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press.

Fodor, J. (2001). *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*. MIT Press.

Fodor, J. A. (2008). *LOT 2: The Language of Thought Revised*. Oxford University Press.

Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2):3–71.

Fox, C. and Lappin, S. (2005). *Foundations of Intensional Semantics*. Blackwell Publishing, Oxford.

Fredkin, E. (2003). The digital perspective. *International Journal of Theoretical Physics*, 42(2):145–145.

Gallie, W. B. (1956). Essentially contested concepts. *Proceedings of the Aristotelian Society*, 56:167–198.

Gärdenfors, P. (2000). *Conceptual Space: The Geometry of Thought*. The MIT Press, Cambridge, MA.

Gärdenfors, P. (2014). *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. The MIT Press.

Gargett, A. and Barnden, J. (2013). Gen-meta: Generating metaphors using a combination of ai reasoning and corpus-based modeling of formulaic expressions. In *Proceedings of TAAI 2013*.

Geffet, M. and Dagan, I. (2005). The distributional inclusion hypotheses and lexical entailment. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 107–114.

Gelder, T. V. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, 92(7):345–381.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Miffline, Boston.

Grefenstette, E. and Sadrzadeh, M. (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, Stroudsburg, PA, USA. Association for Computational Linguistics.

Gutiérrez, E. D., Shutova, E., Marghetis, T., and Bergen, B. K. (2016). Literal and metaphorical senses in compositional distributional semantic models. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.

Harris, Z. (1954). Distributional structure. *Word*, 10(23):146–162.

Haugeland, J. (1993). Mind embodied and embedded. In Houng, Y. H. and Ho, J., editors, *Mind and Cognition: 1993 International Symposium*, pages 233–267. Academica Sinica.

Hegel, G. W. F. (1816/1989). *Science of Logic*. Humanities Press, Atlantic Highlands, NJ. Translated by A. V. Miller.

Heidegger, M. (1926/1962). *Being and Time*. Basil Blackwell, Oxford. translated by John Macquarrie and Edward Robinson.

Hill, F., Reichart, R., and Korhonen, A. (2015). Simlex-999: Evaluating semantic models with genuine similarity estimation. *Computational Linguistics*, 41(4):665–695.

Hobbes, T. (1651). *Leviathan*. Andrew Cooke.

Hoffmeyer, J. (1997). *Signs of Meaning in the Universe*. Indiana University Press.

Hume, D. (1738/2000). A treatise of human nature. Oxford University Press.

Husserl, E. (1900/2001). Logical investigations, volume 1. Number v. 1 in (International library of philosophy and scientific method.). Routledge. Translated by J. N. Findlay.

Hutto, D. D. (2001). Consciousness and conceptual schema. In Pylkkänen, P. and Vadén, T., editors, *Dimensions of Conscious Experience*, pages 15–43. John Benjamins.

Indurkhya, B. (1997). Metaphor as change of representation: an artificial intelligence perspective. *Journal of Experimental & Theoretical Artificial Intelligence*, 9(1):1–36.

Johnson, M. (1990). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press.

Jordanous, A. K. (2012). *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application*. PhD thesis, University of Sussex.

Jr, R. W. G. and Tendahl, M. (2006). Cognitive effort and effects in metaphor comprehension: Relevance theory and psycholinguistics. *Mind and Language*, 21(3):379–403.

Kant, I. (1787/1996). *Critique of Pure Reason*. Hackett Publishing Company, Indianapolis, IN. Translated by Werner S. Pluhar.

Kartsaklis, D. and Sadrzadeh, M. (2013). Prior disambiguation of word tensors for constructing sentence vectors. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1590–1601.

Kauffman, S. A. (1995). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford University Press.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, 7(2):257–266.

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Roberts and Company.

Koestler, A. (1964). *The Act of Creation*. Hutchinson, London.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. University of Chicago Press.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

Landauer, T. K. and Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical Prerequisites*. Stanford University Press, Stanford, CA.

Lee, M. G. and Barnden, J. A. (2001). Reasoning about mixed metaphors within an implemented artificial intelligence system. *Metaphor and Symbol*, 16(1-2):29–42.

Lenat, D. B. (1995). Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64:354–61.

Levinson, S. C. (2001). Yélî dnye and the theory of basic color terms. *Journal of Linguistic Anthropology*, 10(1):3–55.

Levy, O. and Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 2177–2185. Curran Associates, Inc.

Levy, O., Goldberg, Y., and Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transaction of the Association for Computational Lingusitics*, 3:211–225.

Locke, J. (1689/1997). An essay concerning human understanding. Penguin, London.

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, 49:199–227.

Malandrakis, N., Potamianos, A., Elias, I., and Narayanan, S. S. (2013). Distributional semantic models for affective text analysis. *IEEE Transactions on Audio, Speech and Language Processing*, 21(11):2379–2392.

Margolis, E. and Laurence, S. (2007). The ontology of concepts–abstract objects or mental representations? *Noûs*, 41(4):561–593.

Maturana, H. and Varela, F. (1987). *The Tree of Knowledge*. Shambhala, Boston, MA. Translated by Robert Paolucci.

McGregor, S., Wiggins, G., and Purver, M. (2014). Computational creativity: A philosophical approach, and an approach to philosophy. In *Proceedings of the Fifth International Conference on Computational Creativity*.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. In *Proceedings of ICLR Workshop*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pages 3111–3119.

Mikolov, T., tau Yih, W., and Zweig, G. (2013c). 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*, pages 246–251.

Miller, G. A. and Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language & Cognitive Processes*, 6(1):1–28.

Mitchell, J. and Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1439.

Montague, R. (1974). English as a formal language. In Thompson, R. H., editor, *Formal Philosophy: selected papers of Richard Montague*. Yale University Press, New Haven, CT.

Narayanan, S. (1999). Moving right along: A computational model of metaphoric reasoning about events. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence and Eleventh Conference on Innovative Applications of Artificial Intelligence*, pages 121–127.

Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory*, 25(1):45–53.

Ortony, A., editor (1993). *Metaphor and Thought*. Cambridge University Press, 2nd edition.

Pantel, P. (2005). Inducing ontological co-occurrence vectors. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 125–132, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pattee, H. H. (2001). The physics of symbols: Bridging the epistemic cut. *Biosystems*, pages 5–21.

Peirce, C. S. (1932). *Collected Papers of Charles Sanders Peirce*. Harvard University Press. edited by Charles Hartshorne and Paul Weiss.

Pierce, J. R. (1980). *An Introduction to Information Theory*. Dover, New York, 2nd edition.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. William Morrow.

Plato (1892). *The Republic*. Oxford University Press.

Pustejovsky, J., Rumshisky, A., Plotnick, A., Jezek, E., Batiukova, O., and Quochi, V. (2010). Semeval-2010 task 7: Argument selection and coercion. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 27–32.

Putnam, H. (1975). The meaning of "meaning". In Gunderson, K., editor, *Language, Mind, and Knowledge*, pages 131–193. University of Minnesota Press.

Reimer, M. (2001). Davidson on metaphor. *Midwest Studies in Philosophy*, 25:142–155.

Rimell, L. (2014). Distributional lexical entailment by topic coherence. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg.

Ritchie, G. (2007). Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17(1):67–99.

Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton University Press.

Rowlands, M. (2010). *The New Science of the Mind*. The MIT Press, Cambridge, MA.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Linguistics*, 20(1):33–54.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing Management*, 24(5):513–523.

Salton, G., Wong, A., and Yang, C. S. (1975). A vector space model for automatic indexing. In *Proceedings of the 12th ACM SIGIR Conference*, pages 137–150.

Sapir, E. (1970). *The Status of Linguistics as a Science*, pages 65–77. University of California Press.

Schütze, H. (1992). Dimensions of meaning. In *Proceedings of the 1992 ACM/IEEE conference on Supercomputing*, pages 787–796.

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.

Searle, J. R. (1979). Metaphor. In Ortony, A., editor, *Metaphor and Thought*. Cambridge University Press.

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

Shanahan, M. (2010). *Embodiment and the Inner Life: Cognition and Consciousness in the Space of Possible Minds*. Oxford University Press.

Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL.

Shutova, E. (2010). Models of metaphor in nlp. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 688–697.

Shutova, E. (2013). Metaphor identification as interpretation. In *Proceedings of *SEM 2013*.

Shutova, E. (2015). Design and evaluation of metaphor processing systems. *Computational Linguistics*, 41(4):579–623.

Shutova, E., Teufel, S., and Korhonen, A. (2012). Statistical metaphor processing. *Computational Linguistics*, 39(2):301–353.

Skinner, B. F. (1957). *Verbal Behavior*. Copley Publishing Group, Acton, MA.

Sowa, J. F. (2006). *Semantic Networks*. John Wiley & Sons, Ltd.

Sperber, D. and Wilson, D. (1995). *Relevance: Communication and Cognition*. Blackwell, 2nd edition.

Sperber, D. and Wilson, D. (2012). A deflationary account of metaphors. In Wilson, D. and Sperber, D., editors, *Meaning and Relevance*, pages 97–122. Cambridge University Press.

Thomas, M. S. C. and Mareschal, D. (1999). Metaphor as categorisation: A connectionist implementation. In *Proceedings of the AISB '99 Symposium on Metaphor, Artificial Intelligence, and Cognition*, University of Edinburgh.

Thompson, E. (2007). *Mind in Life*. Harvard University Press, Cambridge, MA.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, 215.

Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., and Dyer, C. (2014). Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 248–258. The Association for Computer Linguistics.

Turney, P. D. (2001). Mining the web for synonyms: Pmi-ir versus lsa on toefl. In *Proceedings of the 12th European Conference on Machine Learning*, pages 491–502, London, UK, UK. Springer-Verlag.

Turney, P. D., Neuman, Y., Assaf, D., and Cohen, Y. (2011). Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 680–690.

Turney, P. D. and Patel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188.

Utsumi, A. (2011). Computational exploration of metaphor comprehension processes using a semantic space model. *Cognitive Science*, 35(2):251–296.

van Genabith, J. (2001). Metaphors, logic and type theory. *Metaphor and Symbol*, 16(1-2):23–57.

Veale, T. (2012). From conceptual mash-ups to bad-ass blends: A robust computational model of conceptual blending. In *Proceedings of the Third International Conference on Computational Creativity*, pages 1–8.

Veale, T. (2016). Round up the usual suspects: Knowledge-based metaphor generation. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 34–41, San Diego, California. Association for Computational Linguistics.

Veale, T. and Hao, Y. (2007). Comprehending and generating apt metaphors: A web-driven, case-based approach to figurative language. *AAAI*, pages 1471–1476.

Veale, T., Valitutti, A., and Li, G. (2015). Twitter: The best of bot worlds for automated wit. In Streitz, N. and Markopoulos, P., editors, *Distributed, Ambient, and Pervasive Interactions: Third International Conference, DAPI*, pages 689–699. Springer International Publishing.

von Uexküll, J. (1957). A stroll through the worlds of animals and men: A picture book of invisible worlds. In Schiller, C. H., editor, *Instinctive Behavior: The Development of a Modern Concept*, pages 5–80. International Universities Press, Inc., New York City, NY.

Whitehead, A. N. and Russell, B. (1927). *Principia Mathematica*. Cambridge University Press.

Whorf, B. L. (2012). *Science and Linguistics (1940)*, pages 265–280. MIT Press.

Wiggins, G. A. (2006). Searching for computational creativity. *New Generation Com-*

puting, 24:209–222.

Wiggins, G. A. (2012). The mind's chorus: Creativity before consciousness. *Cognitive Computing*, (4):306–319.

Wilks, Y. (1978). Making preferences more active. *Artificial Intelligence*, 11(3):197–223.

Wittgenstein, L. (1953/1967). *Philosophical Investigations.* Basil Blackwell, Oxford, 3rd edition. trans. G. E. M. Anscombe.

Znidarsic, M., Cardoso, A., Gervás, P., Martins, P., Hervás, R., Alves, A. O., Oliveira, H. G., Xiao, P., Linkola, S., Toivonen, H., Kranjc, J., and Lavrac, N. (2016). Computational creativity infrastructure for online software composition: A conceptual blending use case. In *Proceedings of the Seventh International Conference on Computational Creativity, UPMC, Paris, France, June 27 - July 1, 2016.*, pages 371–379.