frontiers | Frontiers in Artificial Intelligence

CONCEPTUAL ANALYSIS
published: 24 May 2022
doi: 10.3389/frai.2022.796741

Beyond the Benchmarks: Toward Human-Like Lexical Representations

Suzanne Stevenson[1]* *and Paola Merlo*[2]

[1] Department of Computer Science, University of Toronto, Toronto, ON, Canada, [2] Linguistics Department, University of Geneva, Geneva, Switzerland

To process language in a way that is compatible with human expectations in a communicative interaction, we need computational representations of lexical properties that form the basis of human knowledge of words. In this article, we concentrate on word-level semantics. We discuss key concepts and issues that underlie the scientific understanding of the human lexicon: its richly structured semantic representations, their ready and continual adaptability, and their grounding in crosslinguistically valid conceptualization. We assess the state of the art in natural language processing (NLP) in achieving these identified properties, and suggest ways in which the language sciences can inspire new approaches to their computational instantiation.

Keywords: computational linguistics, natural language processing, lexical semantics, lexicon structure, human lexical representations, cross-linguistic generalization

OPEN ACCESS

Edited by:
Alessandro Lenci,
University of Pisa, Italy

Reviewed by:
Nasredine Semmar,
CEA Saclay, France
James Pustejovsky,
Brandeis University, United States

***Correspondence:**
Suzanne Stevenson
suzanne@cs.toronto.edu

Specialty section:
This article was submitted to Frontiers in Artificial Intelligence, a section of the journal Frontiers in Artificial Intelligence

Received: *17 October 2021*
Accepted: *19 April 2022*
Published: *24 May 2022*

Citation:
Stevenson S and Merlo P (2022) Beyond the Benchmarks: Toward Human-Like Lexical Representations. Front. Artif. Intell. 5:796741. doi: 10.3389/frai.2022.796741

1. INTRODUCTION

The field of computational linguistics (CL) has exploded recently—especially the work characterized as "NLP," which has become almost synonymous with "machine learning approaches applied to large text datasets." The practical successes have been rampant (e.g., Collobert and Weston, 2008; Mikolov et al., 2013a; Bahdanau et al., 2015; Vaswani et al., 2017; Devlin et al., 2018). But the progress on task-oriented measures and benchmark performance has come at a price. CL as a field has long benefited from bringing together insights from theoretical linguistics, psycholinguistics, and other of the language sciences, to inform computational methods for automatically processing language. This inherently interdisciplinary approach has over time helped to ensure that computational systems are grounded in firm scientific understanding of the nature of human language. Periodically, however, the success of a particular computational approach has threatened this interdisciplinarity by seeming to obviate the need for drawing on other disciplines; this phenomenon has perhaps been most famously captured by the saying from Fred Jelinek, "Whenever I fire a linguist our system performance improves".[1]

[1] In his LREC 2004 talk, "Some of my best friends are linguists," Jelinek indicates the quote is from a talk entitled "Applying Information Theoretic Methods: Evaluation of Grammar Quality," given at the Workshop on Evaluation of NLP Systems, Wayne PA, December 1988. The focus on statistical approaches as "the answer" in 1990s NLP is further illustrated by the loss of the context of this quip: while stressing that data-driven prediction models were key, Jelinek also emphasized the importance of input from linguistics on data annotation and model structure.

We are today in NLP seeing a similar emphasis on performance and an associated focus on a particular class of algorithms.[2] This exclusionary focus is unfortunate, as it has meant that CL has weakened its crucial connections to the other language sciences, and thereby lost some of the underpinnings and guidance that comes from a comprehensive scientific understanding of language as an essential and uniquely human ability.[3] Although recent NLP work can achieve performance on benchmarks that was unheard of only a short time ago, one might wonder how much such research is actually furthering progress on the overarching goal of matching broad human abilities in linguistic communication.

In this article, we consider this issue in the context of research on lexical semantics. We adopt this focus for two reasons. First, lexical items—or words[4]—are the locus of fundamental semantics, as well as of combinatorial properties that underlie their composition into the larger units of meaning used in communication. As such, words are a basic building block of language, and adequately capturing lexical semantics is critical to computational systems for processing language. Second, words generally—and lexical semantics in particular—have received much attention in recent NLP, and are the focus of many of the reported successes noted above.

Specifically, our aim here is to examine essential desiderata for computational approaches to lexical semantics. In order to process language in a way that is compatible with human expectations in a communicative interaction, we need computational representations of lexical properties that adequately capture human knowledge of words. In this context, we discuss the concepts and issues that underlie the scientific understanding of the human lexicon and key defining properties (Section 2); assess the state of the art in NLP in achieving the identified properties (Section 3); and suggest ways in which the language sciences can inspire new approaches to their computational instantiation (Section 4).

2. THE HUMAN LEXICON

The human capacity for language is founded in very powerful cognitive mechanisms that underlie general intelligence: the ability to (multiply) categorize stimuli into richly structured representations, and to continually learn and readily adapt to novel stimuli (e.g., Langacker, 1987; Croft and Cruse, 2004; Goldberg, 2006). Moreover, language is grounded in universal human experience, such that these categorization and generalization mechanisms operate over a level of universal (crosslinguistically valid) conceptual grounding (e.g., Berlin and Kay, 1969; Bowerman and Choi, 2001; Levinson et al., 2003; Regier et al., 2007; Majid et al., 2008; Gentner and Bowerman, 2009). We briefly discuss the implications of each of these three properties for the human lexicon.

First, human lexical representations and the lexicon itself exhibit a rich semantic structure, encoding a multitude of semantic relations among words. In addition to semantic similarity and semantic relatedness,[5] people are sensitive to taxonomic relations, part-whole relations, entailment, subsumption, hyponymy, and many others, which organize the meaning of words and their relation to each other in a multiply connected structure (e.g., Collins and Loftus, 1975; Pustejovsky, 1995; Miller, 1998; Hale and Keyser, 2002; Jones et al., 2015). In addition to the structured relations among them, words also have rich internal semantic structure (e.g., Cruse, 1986; Pustejovsky, 1995; Croft and Cruse, 2004). Moreover, the commonalities along various semantic dimensions can form the basis for classes of words that have shared linguistic behavior, thus serving as a critically important means for organizing further grammatical knowledge (e.g., Levin, 1993; Croft, 1994; Baker, 2003). Lexical representation is further complicated by lexical ambiguity: Most words have multiple meanings (Bréal, 1897), with a high degree of variability in the extent and manner in which those meanings are related (e.g., Nunberg, 1979; Bartsch, 1984; Williams, 1992; Geeraerts, 1993; Tuggy, 1993), which people are sensitive to (e.g., Rodd et al., 2002; Klepousniotou et al., 2008; Armstrong and Plaut, 2016). Words are thus linked to each other by elaborate networks of semantic relations that are crucial to their felicitous use and combination.

Second, human lexical representations are malleable: in addition to being multiply ambiguous, they are readily amenable to meaning shifts in context, and frequently undergo semantic change, taking on new senses. This online adaptability is the key to successful interaction. People not only easily access different aspects of meanings in different contexts, they construct nuanced interpretations in conjunction with conversational partners (e.g., Clark and Clark, 1979; Langacker, 1987; Brennan and Clark, 1996; Cruse, 2000; Kintsch, 2001; Croft and Cruse, 2004; Zawada, 2006). Such representations are not always fleeting: linguistic creativity entails that people frequently generate new usages of words and shifts in meaning, and interlocutors adjust their lexical knowledge in response to such novel usages (e.g., Langacker, 1987; Croft and Cruse, 2004; Goldberg, 2006). While it has long been recognized that children have inductive biases to help them

[2]Perhaps the quote that best sums up the current situation is "deep learning is going to be able to do everything" (Geoff Hinton, November 2020, when asked about the ability "to replicate all of human intelligence," including language processing; https://www.technologyreview.com/2020/11/03/1011616/ai-godfather-geoffrey-hinton-deep-learning-will-do-everything/). For a recent informative debate on the tension—and potential for complementarity—between neural approaches and linguistic theory, see, for example, the perspective article of Pater (2019) and its associated commentaries.

[3]We are not alone in decrying the negative effects of the singular focus on (and sometimes exaggerated claims about) large language models, and associated benchmark tasks. See, for example, Church (2017), Bender and Koller (2020), and Church and Liberman (2021), and the carefully articulated recent discourse at https://twitter.com/emilymbender/status/1430944351358648324.

[4]While recognizing that the definition of "word" can be problematic and may have differing instantiations across languages, for simplicity here we will refer to lexical items as "words," assuming this includes relevant morphemes, lexicalized multi-word expressions, etc.

[5]We fully recognize that semantic similarity and semantic relatedness are different constructs (e.g., Budanitsky and Hirst, 2006, among many others), and perhaps different as well from distributional similarity and relatedness (e.g., Sahlgren, 2008). Since the distinction between similarity and relatedness is not of high relevance to most points being made here, we will use "semantic similarity" to encompass the idea of closeness in semantic space, except where explicitly noted.

learn from small amounts of data (e.g., Clark, 1987; Markman, 1987; Samuelson and Smith, 1999), the dynamic nature of the human lexicon means that adults also are continually adapting their lexical representations.

Third, and finally, lexical semantic knowledge is built on universal principles that are grounded in fundamental human perceptual and conceptual experiences that hold across languages (e.g., Goddard and Wierzbicka, 1994; Haspelmath, 1997; Regier et al., 2007; Majid et al., 2008; Majid and Van Staden, 2015; Kemp et al., 2018). The result is that languages show constrained variation in their lexical semantic systems. For example, languages vary widely in the precise lexical divisions they adopt in a domain (such as how to carve up the continuous color spectrum into basic color terms), differentially making a trade-off between expressivity of the terms and efficiency in their lexicons (e.g., Kemp et al., 2018; Zaslavsky et al., 2018). However, considerations of "cognitive naturalness" of lexical categories greatly constrain the observed variation across languages, such that human lexicons follow common organizational principles (e.g., Berlin and Kay, 1969; Levinson et al., 2003; Gentner and Bowerman, 2009; Xu et al., 2020). Moreover, people benefit (or suffer!) from "transfer effects" in learning a new language, or in lexical access in the context of a multilingual lexicon (e.g., Van Hell and de Groot, 1998; Degani et al., 2011). Thus, the universality of the cognitive/conceptual basis of language leads to predictions about expected crosslinguistic commonalities and areas of difference.

3. ASSESSING THE LEXICAL REPRESENTATIONS IN NLP

The identified properties of the human lexicon—richly structured representations, ready and continual adaptability, and universality—have been differentially highlighted at different stages of development in NLP, but have rarely been addressed comprehensively. In the first subsection below, we briefly outline some of the relevant history of computational lexical semantics, presenting the progression of ideas with reference to these key properties. In the second subsection, we discuss ways in which the current state-of-the-art in lexical semantic representation continues to fall short of the identified properties of the lexicon that support successful human communication.

3.1. From Early Structured to Distributional to Neural Approaches

Achieving broad coverage lexical knowledge has long been recognized as a critical step to achieving language processing at scale (i.e., beyond narrow domains or circumscribed tasks). Early approaches to large-scale lexical resources focused on highly structured lexical representations, as in, for example, WordNet (Beckwith et al., 1991; Fellbaum, 1998), FrameNet (Baker et al., 1998; Fillmore and Atkins, 1998), VerbNet (Levin, 1993; Kipper, 2005), and PropBank (Palmer et al., 2005). The structure of such lexicons is not only a practical organizational technique: crucially, lexical items derive their nuanced semantics in part through the elaboration of multiple semantic and/or syntactic relations among them. For example, Wordnet organizes words into *synsets* that group roughly synonymous words, and then links these synsets with hypernym/hyponym links (among other semantic relations) to indicate a basic taxonomic structure over meanings. Due to ambiguity, words can appear in multiple synsets, leading to a complex network structure. For example, one of the synsets of *newspaper* is {newspaper, paper} whose hypernym is {press, public press}, while another synset is {newspaper, newsprint}, whose hypernym is {paper}, where the word *paper* in different senses is both a synonym and a hypernym of *newspaper*. These resources thus capture rich semantic structure that has supported a range of applications, such as word sense disambiguation (e.g., Patwardhan et al., 2003), semantic parsing (e.g., Das et al., 2014), and question-answering (e.g., Clark et al., 2018). However, while these resources have been very successfully deployed for key tasks in NLP, they are difficult to adapt dynamically, and require considerable manual effort to transfer to other languages, because of the necessity for elaborating the multiple senses and/or semantic relations for each word (e.g., Vossen, 1998; Burchardt et al., 2009). In short, the very richness of their structure makes it resource-intensive to extend them within or across languages.

In response to these shortcomings, automatic lexical acquisition was identified as key to further progress in CL (e.g., Ellison, 1997; Baldwin et al., 2005; Armstrong et al., 2010). Computational work in lexical semantics in the 1990s and 2000s had two prominent strands: learning of the structured relations among words (a key source of the power of the above resources), as well as learning the meaning of individual words. In both cases, the focus on learning from data was intended to address both the need for adaptability and the desire for crosslinguistic breadth and validity.

The first strand of work in data-driven lexical acquisition concentrated on structured lexical representations. For example, much work aimed to learn various semantic relations among words, such as hyponymy, synonymy, part-whole, etc. (e.g., Hearst, 1992; Riloff, 1996; Girju et al., 2006). For example, the simple but highly effective technique of "Hearst patterns" used common phrases to automatically infer taxonomic relations among words; e.g., "HYPERNYM, *such as* HYPONYM" (*fruit, such as apples and bananas*) or "HYPONYM *and other* HYPERNYM" (*apples and other fruit*). Another important focus was on automatically acquiring the rich information about predicates (such as argument structure and verb or adjective classes). These methods used statistics over the syntactic patterns of predicates to automatically classify them into known semantic classes, or even to discover such classes, in order to generalize known combinatory properties of lexical items to novel or previously unseen words (Merlo and Stevenson, 2001, 2005; Stevenson and Joanis, 2003; Boleda et al., 2004; Korhonen and Briscoe, 2004; Schulte im Walde, 2006; Li and Brew, 2008; Sun and Korhonen, 2009). In addition to achieving adaptability within a language, some research was driven by the goal of crosslinguistic adaptability as well. For example, some approaches exploited crosslinguistic similarities to extend methods developed for English to new languages (Merlo et al., 2002; Padó and Lapata, 2005; Snyder and Barzilay, 2008;

Burchardt et al., 2009; Samardžić and Merlo, 2010). Other work leveraged multi-lingual resources further, by using knowledge of crosslinguistic variation as a way to improve results *within* a language. For example, Tsang et al. (2002) exploited bilingual corpus data to learn a semantic distinction in English that is not morphologically marked in English, but is so marked in Chinese. Despite these various advances in automatic lexical acquisition, a challenge remained for structured lexical approaches: These methods relied on identifying surface correlates of the deeper semantic properties to be learned, which often had to be done manually.

The approaches above were using distributional patterns to learn a set of semantic relations or an assignment into a (typically pre-conceived) structured representation. At the same time, researchers were increasingly considering distributional cues as capable of comprising the semantic representation itself. A wealth of work on distributional semantic models (DSMs) was inspired by early views in linguistics and philosophy that meaning is determined by use in context (Wittgenstein, 1953; Harris, 1954; Firth, 1957), and by computational cognitive modeling approaches to capturing meaning based on word contextual associations (e.g., Lund and Burgess, 1996; Landauer and Dumais, 1997). In contrast to the structured lexical approaches described above, the distributional hypothesis promised a data-driven representation of semantics that would avoid both the manual work and the need for explicit assumptions about semantics that may not generalize across domains, genres, and languages. Moreover, such representations had the potential to capture the various senses of a lexical item, which could be disambiguated in composition with co-occurring words (e.g., Landauer and Dumais, 1997; Kintsch, 2001; Erk and Padó, 2008; Mitchell and Lapata, 2008; Van de Cruys et al., 2011). Many types of DSM approaches have been explored in CL, considering various context sizes (e.g., number of words, or neighboring words vs. documents), contextual relations (e.g., word co-occurrence vs. dependency relations), and statistical measures of word–context association (Schütze, 1994; Padó and Lapata, 2007; Erk and Padó, 2008; Mitchell and Lapata, 2010). DSMs have generally yielded semantic representations that perform well on semantic similarity benchmarks and in a range of downstream NLP tasks (Schütze, 1994; Landauer and Dumais, 1997; Baroni and Lenci, 2010). By the mid 2000s, DSMs had become a prominent means of lexical semantic representation in CL (e.g., Lenci, 2008).

More recently, the increased power of statistical methods and neural network approaches have enabled DSMs to exploit the promise of the distributional hypothesis to a high degree (Collobert and Weston, 2007; Mikolov et al., 2013c; Pennington et al., 2014; Pereira et al., 2016), and the techniques have been successfully applied across many languages (Bojanowski et al., 2017). Moreover, recent methods have extended the basic framework to integrate with neural language models, thereby achieving adaptability of meanings in local (sentence-level) contexts for many languages (e.g., ELMo, BERT, mBERT; Devlin et al., 2018; Peters et al., 2018; Wu and Dredze, 2019). The broad practical successes of neural approaches to learning word meaning and integrating lexical semantics with other NLP tasks has led to their current dominance in the field.

3.2. Current Limitations in Matching Human Lexical Properties

Despite their success, distributional semantic representations—"word embeddings"—are still far from capturing human-like lexical abilities, along all the dimensions of structure, adaptability, and universality. First, current word embeddings do not encode all of the rich semantic properties and relations that we know humans are sensitive to (e.g., Rubinstein et al., 2015; Boleda et al., 2017; Grand et al., 2018). For example, Rubinstein et al. (2015) found that word embeddings captured taxonomic knowledge ('is a fruit', 'is an animal') much better than they did attributive properties of word meanings ('is yellow,' 'is round'). With regard to ambiguity, while evidence suggests that distributional word representations can capture multiple meanings of a word (Burgess, 2001; Kintsch, 2001; Mu et al., 2017; Arora et al., 2018; Beekhuizen et al., 2019), much remains to be explored about whether and how they might do so (Reisinger and Mooney, 2010; Li and Jurafsky, 2015; Jamieson et al., 2018). In addition, while much earlier lexical acquisition work successfully learned verb argument structures and their surface expression, experiments on context-aware embeddings have shown inconsistent performance in predicting the valid usages of verbs (e.g., Kann et al., 2019; Warstadt et al., 2019). While there is legitimate skepticism that purely text-based distributional methods can truly learn human-like meanings (e.g., Sahlgren, 2008; Bender and Koller, 2020), there is also much room for them to extend their capabilities beyond solely similarity-based semantic space.

Second, contextualized word embeddings have shown some success at exhibiting nuance of meaning in context (e.g., Choi et al., 2017; Ethayarajh, 2019; Hofmann et al., 2020). However, at least some approaches are overly sensitive to irrelevant factors (e.g., word order variation that does not change meaning), such that very close paraphrases are not assigned close embeddings (Shi et al., 2019). Further research will need to assess how well current approaches to contextualized understanding of words matches that of people. Moreover, while research on historical semantic change has thrived using historical embeddings (e.g., Hamilton et al., 2016; Lu et al., 2019), little attention has been paid to shorter-term sense change, with some caveats for using word embeddings in this task (Del Tredici et al., 2018). In addition, while there has been much focus on one-shot or few-shot learning as a means for adapting the knowledge of large-scale models (e.g., Li et al., 2006; Ritter et al., 2017; Brown et al., 2020; Schick and Schütze, 2021), recent work has discussed that "few-shot" learning is not as data-lean as it may seem (Perez et al., 2021). Thus, although their foundation in learning from data holds the promise of adaptability, the data requirements of neural approaches can limit their ability to adapt on-the-fly in the way that people can.

Finally, it is not yet clear how "universal" are the current distributional semantic spaces. Word embedding spaces show a crosslinguistically similar structure (Mikolov et al., 2013c),

but even the most successful cross-language word embedding techniques learn the monolingual spaces separately, and only in a second step map the two languages onto each other (Artetxe et al., 2017, 2018; Lample et al., 2018). Other effective approaches have depended on sentence-aligned parallel corpora to support cross-lingual embeddings (e.g., Gouws et al., 2015; Levy et al., 2017; Zennaki et al., 2019, among others). These kinds of techniques may lead to multilingual spaces that show the influence of the particular resources and languages used. In short, there has been much work on multilingual approaches, but multilingual does not necessarily equal universal, which implies a common conceptual representation across languages. For example, one approach has required manually-specified conceptual categories to show improvements on both similar and distant languages (Wang et al., 2019). Further insights from linguistic and cognitive constraints on what is a valid lexical representation or a structured lexicon may further enable true crosslinguistic generalization.

In summary, computational approaches based on lexical and grammatical theories have developed richly structured lexicons, but achieving adaptability and crosslinguistic validity in such frameworks requires much manual effort. By inducing representations from data, current distributional semantic approaches have the potential to be fully adaptable, and generalizable across languages without the manual effort of earlier NLP systems. However, distributional research has largely focused on semantic similarity as the sole organizing principle of the learned knowledge, with less attention to the many other semantic relations encoded in the human lexicon. Moreover, despite their fundamental basis in learning, the proposed methods cannot adapt dynamically due to cognitively unrealistic training data requirements. Finally, although the learning methods are in principle generalizable across languages, they lack the biases to capture human conceptual underpinnings. As it stands, overcoming the weakness of the conceptual biases requires extremely large training data sets, available only for a few languages.

4. INSPIRATION FROM HUMAN LEXICAL ABILITIES

Early work on lexical resources and automatic lexical acquisition had a strong basis in linguistic and psycholinguistic theory and insights. These connections have become more tenuous in recent NLP, despite earlier recognition that work on distributional representations in both CL and cognitive science can inform each other (see, e.g., Lenci, 2008), and despite continued work at the intersection of the two fields.[6] A more concerted effort is required to bring linguistic and psycholinguistic understanding together with recent data-driven approaches in order to achieve more human-like lexical representations and abilities. Here we describe some relevant cognitively-inspired work from recent years, and suggest how such work can inform future directions in NLP to address the properties of the human lexicon.

4.1. Structure in Lexical Representations and Learning

Word embeddings are largely founded on the notion of semantic similarity, and ensuring that word vector similarities match human judgments has been an important goal (e.g., Baroni et al., 2014; Pereira et al., 2016; An et al., 2018; Grand et al., 2018; Iordan et al., 2022). Less attention has been paid to whether the actual structure of a DSM's similarity space matches what is known about the human lexicon. For example, while work in CL has noted that different types or levels of similarity may be captured in DSMs—first-order similarity reflecting word associations, and second-order similarity reflecting substitutability (e.g., Schütze and Pedersen, 1993; Grefenstette, 1994; Levy et al., 2015)—less attention has been paid to whether and how these finer-grained notions of similarity within current word embeddings match human lexical processing. Some recent work has addressed this issue (e.g., Beekhuizen et al., 2019; Chronis and Erk, 2020; Samir et al., 2020). For example, Samir et al. (2020) demonstrate that using different combinations of the input and output matrices of the word2vec algorithm not only mimics the two kinds of similarity, but does so in a way that matches human behavioral data on semantic priming and lexical decision. However, other properties of human similarity judgments—such as asymmetries in word associations or violations of the triangle inequality (w_1 similar to w_2, and w_2 similar to w_3, do not imply w_1 similar to $w3$; cf. *asteroid*, *belt*, and *buckle*, Griffiths et al., 2007)—are not consistently captured in embedding spaces (Griffiths et al., 2007; Nematzadeh et al., 2017; Rodriguez and Merlo, 2020). Building on the insight from Griffiths et al. (2007) that interpretation of a word within the context of a topic can resolve some of these mismatches with human judgments by appropriately disambiguating the words, one avenue for the future may be to consider word embeddings that are topically-constrained (such as in Iordan et al., 2022).

Word embeddings also fail to reflect other linguistically-relevant types of similarity that play a role in human language processing. For example, when faced with long-distance dependencies between two feature-sharing items in a sentence (such as those found in questions, relative clauses, pronoun anaphora, and other frequent phenomena), people exhibit effects of interference if there is a third similar element in the sentence (Rizzi, 2004; Franck et al., 2015). However, this effect of similarity interference is not correlated to the similarity of words calculated statically in a vector space or even dynamically in a neural network model of processing (Merlo and Ackermann, 2018; Merlo, 2019). The general picture that emerges from all these studies is that word similarity is a rich construct of the human lexicon, and while word embedding spaces represent some fundamental properties of semantic similarity, more nuanced notions, and some grammatically-relevant aspects, may not emerge from such representations.

Moreover, work in psycholinguistics has shown that human access and interpretation of a word are influenced by its

[6] See, e.g., the proceedings of regularly-occurring workshops such as CogACLL, CogALex, and CMCL, as well as tracks on linguistics and psycholinguistics in the major CL conferences.

semantic neighborhood—the structural layout in semantic space of semantically similar words (e.g., Burgess, 1998; Buchanan et al., 2001). Such considerations are especially important for understanding how ambiguous words encode their varied semantics. Recent work has shown that some, but not all, oft-used word embeddings reflect a difference in neighborhood structure between homonyms (words with multiple unrelated meanings) and polysemes (words with multiple related senses) (Beekhuizen et al., 2018, 2021), corresponding to experimental differences in human processing of ambiguous words (Rodd et al., 2002; Hino et al., 2006). Other work has shown that contextualized (token) representations of abstract and concrete words differ in their pattern of dispersion across different layers of a language model (Chronis and Erk, 2020), again demonstrating the potential richness of lexical semantic structure. Such work illustrates that representational adequacy of distributional semantic models should consider finer-grained details than a match to human similarity judgments.

Beyond the similarity structure of word embeddings, it remains unclear how much current models capture the many other semantic relations that people are sensitive to (e.g., Köper et al., 2015; Santus et al., 2016; Ettinger, 2020). For example, the extent to which distributional semantic spaces represent more abstract semantic properties is an open question (Baroni and Lenci, 2008; Rubinstein et al., 2015; Hollis and Westbury, 2016; Hollis et al., 2017). Abstract semantic classes, such as verb or adjective classes, play a crucial role in theories of how human lexical knowledge encodes knowledge of grammar (e.g., Levin, 1993; Paradis, 2001; Morzycki, 2012), and so it is important that lexical representations support organization of such classes. For example, semantic verb classes are an important means for generalizing knowledge of argument structure: learning that a new verb *gorp* reflects a change of state will enable an English speaker to know that if you can say *Jane gorped the cookie*, you can also say *The cookie gorped*. Such classes often capture commonalities at a higher level of abstraction than the simple within-domain similarity that is typically demonstrated in word embeddings; for example, the change of state class covers verbs as dissimilar as *age*, *blacken*, *crumble*, *deflate*, and *energize* (Levin, 1993). Some recent work has demonstrated the ability of word embeddings to capture an abstract semantic class of adjectives that, like verb classes, also has ramifications for appropriate use of the words in grammatical constructions (Samir et al., 2021). However, even for the "poster child" task of solving linguistic analogies, which has showcased the semantic abilities of modern distributional representations (Mikolov et al., 2013b), higher levels of abstraction can be a challenge. It has been shown that for more abstract relations (ones that go beyond within-domain similarity) it is difficult to achieve reasonable performance in these tasks (Rogers et al., 2017), requiring more explicit knowledge of abstract classes (Drozd et al., 2016) or an additional learning component to extract the relevant dimensions of comparison (as in Lu et al., 2019).

Better understanding of learning algorithms may be required to achieve the kind of rich and abstract structure that human lexical knowledge demands. Again, insight may be drawn from cognitive principles. Analogies to human processes of memory and attention abound in neural architectures, yielding interesting and powerful mechanisms to guide the information flow through the network (e.g., Hochreiter and Schmidhuber, 1997; Vaswani et al., 2017). These mechanisms take inspiration from human cognition in an intuitive and loose sense, but generally do not distinguish the different types of mechanisms—such as working memory vs. episodic or semantic memory—found in humans. Closer modeling of the more structured findings from psychology and cognitive science might bring further fruits. For example, recent modeling of human reading processes in neural architectures has yielded finer-grained understanding of attention to words in language models (Sood et al., 2020; Hahn and Keller, 2021). Other work has noted that structured memory, as in humans, may be required for the kind of meaningful compression in learning that is necessary for successful abstraction over input stimuli: by disentangling computation and storage (which are intertwined in the weight parameters of most neural networks), richer storage mechanisms can be achieved that support both faster retrieval, and forgetting in support of abstraction (Nematzadeh et al., 2020). Other research considering cognitive factors in communication has found that human lexical organization is subject to optimization of the trade-off between complexity and accuracy (e.g., Kemp et al., 2018; Zaslavsky et al., 2018, among many others). The same principle of an information bottleneck has also been shown to help explain hierarchical structure in DNN layers (Tishby and Zaslavsky, 2015). Altogether, studies such as these suggest that drawing clearer connections between human principles of communication and current learning approaches may lead to more human-like lexical representations.

4.2. Adaptability and Learning From Small Amounts of Data

Adaptability requires generalization. But current data-driven NLP models do not generalize well to new problems or instances out of the training distribution (Ettinger et al., 2017; Belinkov and Bisk, 2018; Schölkopf, 2019). People are not as susceptible to overfitting, at least partly because they have strong prior biases, grounded in the actual causal structure of the problem. One possible approach for developing more robust methods, then, is to pay more attention to causal chains in the generative process that give rise to the data and not just to correlations in the data (Schölkopf et al., 2012; Lake et al., 2017; Bengio et al., 2020). Drawing on such causal knowledge should enable methods that support appropriate generalization, and improved adaptability.

Detailed linguistic analyses and psycholinguistic studies can provide information on the causal structure that is likely to underlie the observed distributions. For example, recent work in linguistics has investigated the causes of variation in the expression of causative constructions in several languages—corresponding to the alternation in English between *Kiva broke the vase* (with the causal agent specified) and *the vase broke* (Alexiadou, 2010; Haspelmath et al., 2014; Heidinger, 2015). A superficial correlation has been found between the distribution of verb form (length of the causative alternative of the verb) and its frequency (Haspelmath et al., 2014). However,

further investigation has identified the perceived probability of external causation—is the event spontaneous or not—as a better explanation of the patterns of crosslinguistic data (Samardžić and Merlo, 2018). Taking this latent factor into account is shown to inform generalization, achieving improved prediction of which verbs can occur in which causative constructions (Samardžić and Merlo, 2018). In another example, Yu et al. (2020) proposed a probabilistic model building on linguistic analyses of denominalization—i.e., use of a noun as a verb (Clark and Clark, 1979). Human-like interpretation of novel uses—inferring that *porch the newspaper* likely means "throw the newspaper onto the porch"— depends on a latent frame (topic or scenario) variable in the model. Yu et al. (2020) demonstrate that this latent variable enables the model to outperform BERT in predicting the appropriate paraphrase for novel denominal verbs. This work suggests that appropriately modeling the causal structure of a phenomenon can outweigh even the massive knowledge encoded in a recent language model.

Such predictive generalization is key to achieving the on-the-fly adaptability that people exhibit. Historical corpora and other resources (including associated historical embeddings) have facilitated work on lexical change over some period of time (Hamilton et al., 2016; Lu et al., 2019), but less attention has been given to the rapid adaptation to novel nuances of meaning and novel constructions (e.g., Cook et al., 2014; Del Tredici et al., 2018; Ryskina et al., 2020; Watson et al., 2021). People continually produce words in new meanings and in creative usages of constructions, and interlocutors quickly extend their lexical knowledge to grasp the novel interpretations. This "one-shot" learning in people has not been achieved in recent NLP systems, which, as noted above, do not actually use just small amounts of training data in so-called "few-shot" learning (Perez et al., 2021). Psycholinguists have proposed a number of biases that enable children to learn words with few exposures; what principles govern the ability of adults to similarly adapt quickly and generalize over small amounts of data?

Much psycholinguistic work aims to elucidate the cognitive mechanisms that enable people to generalize their lexical knowledge in producing and interpreting novel usages of words. By understanding the cognitive processes at play when people form generalizations, work in NLP can better identify the factors and mechanisms required to achieve human-like abilities. For example, generalization of constructions to new words—such as saying "don't try to batman your way into it"—is viewed as a process of category extension (i.e., seeing a construction, such as "VERB one's way into NOUN," as a category of usages). This process is influenced by factors such as similarity of the novel item to observed instances of the construction, and the frequency and variability of the latter—factors which support easier extension to new usages (e.g., Bybee and Eddington, 2006; Suttle and Goldberg, 2011; Perek, 2016). Recently, Watson et al. (2021) have demonstrated that these principles hold in creative usages in large-scale social media data – specifically, in novel usages of denominal verb constructions in an online discussion platform. For example, novel usages such as "I am a man (...usually all *flannelled up*)" tend to have high similarity to existing usages (*gear up*, *glove up*, *mask up*, *sweater up*, ...) that form a broad and frequent class. Moreover, Watson et al. (2021) find that novel usages cluster around other novel usages (*flannel* is similar to *sweater*), confirming that the exemplar-driven innovation found in historical analyses (e.g., Habibi et al., 2020; Yu et al., 2020) plays a role in dynamic adaptation of language. This is an important point, because one-off usages, rather than being statistical noise, serve as informative signals to people of legitimate creativity. Research is needed to see how such biases suggested by cognitive principles might be built into neural models of meaning acquisition and extension (e.g., compare McCoy et al., 2020), to ensure the level of lexical adaptability observed in human communication.

4.3. Truly Crosslinguistic Generalization

Semantic typology has contributed significantly to our understanding of the crosslinguistic foundations for human lexical semantics. Clearly, languages vary widely in how they "carve up" a semantic space with words—e.g., some having a single word for two concepts for which others have distinct words (English *on* [SUPPORT] vs. Dutch *aan* [TENUOUS SUPPORT] and *op* [STABLE SUPPORT]). Despite this lack of alignment in the world's lexicons—with various one-to-many or even many-to-many mappings attested between languages—detailed linguistic analyses of various semantic domains have revealed consistent commonalities in how languages label concepts with words (e.g., Berlin and Kay, 1969; Haspelmath, 1993; Levinson et al., 2003; Majid et al., 2008; Gentner and Bowerman, 2009). More recently, large-scale work has confirmed that languages exhibit universal tendencies in lexical structure across a wide variety of semantic domains (Youn et al., 2016; Thompson et al., 2018). However, while NLP has effectively drawn on linguistic typology in other areas (such as morphology and syntax), little research has considered how to incorporate the insights from lexical semantic typology to inform and constrain computational approaches to meaning (e.g., Bender, 2016; Dubossarsky et al., 2019).

The creation of multilingual semantic spaces is one area that may benefit from typological considerations, especially an examination of how well such spaces capture the crosslinguistic principles that underlie human lexicons. Much of the richness of lexical structure, and differences across languages, arise from variation in polysemy—how languages differently package up related meanings into ambiguous words. For example, while the word for "tongue" in English, Hebrew, and Russian refers to both the physical organ and a language, only in English and Hebrew does it also refer to a piece of land that protrudes into the sea (Navigli and Ponzetto, 2010). Such misalignments pose serious challenges for NLP, since ambiguous words in one language can map very differently to words in another language. Rabinovich et al. (2020) showed that a multilingual semantic space could capture the similarity structure among concepts that match human patterns of such polysemies. Ensuring that multilingual spaces have such properties may enable them to better support automatic alignment across languages in future NLP systems.

One issue raised by such work is the extent to which multilingual spaces show bias from certain languages, since they generally rely on monolingual spaces or bilingual resources, in

which some languages are likely over-represented (e.g., Artetxe et al., 2017, 2018; Lample et al., 2018). Interestingly, Merlo and Rodriguez (2019) show that multilingual spaces exhibit transfer effects—where the source language influences the semantic space of the target language—that are consistent with the cross-language influences seen in human bilinguals. This is intriguing, as it tells us that we can learn much from the broad literature on lexical semantic transfer effects in humans (e.g., Van Hell and de Groot, 1998; Degani et al., 2011). For example, even highly skilled human translators exhibit statistically detectable evidence of the source language in their target word choice (Rabinovich et al., 2017). By understanding more clearly how a source language can influence a target language, better means might be determined for anticipating bias in transfer learning and alleviating it. Such considerations are becoming increasingly important: work in NLP is heavily invested in so-called "foundation models," which are largely focused on English due to their data and/or computation requirements (Bender et al., 2021; Bommasani et al., 2021). Methods for transferring such large-scale knowledge to a broad range of diverse languages will be necessary, and informed debiasing—drawing on knowledge of crosslinguistic tendencies and divergences—could be key to such efforts.

In addition to considering transfer between languages, NLP could also benefit from additional insights into the universal basis of lexical semantics. A key finding in semantic typology is the observation that, the more frequently (across languages) that two meanings are referred to by a single word, the more likely those meanings are to be ("universally") semantically similar (e.g., Levinson et al., 2003; Gentner and Bowerman, 2009). Building on this insight, recent computational modeling work has shown that patterns in crosslinguistic data can reveal conceptual similarities that can form the basis of a "universal" semantic space for various lexical domains (Beekhuizen et al., 2014, 2017; Beekhuizen and Stevenson, 2018). In contrast to the typical multilingual approach in NLP of aligning a collection of monolingually-derived spaces, here a common semantic space across languages is founded in the dimensions of meaning that emerge from the crosslinguistic lexical patterns of aligned word usages. Such representations can reveal important properties of a distributional semantic space that conforms to typological principles. For example, building on insights from such a semantic space, Rabinovich et al. (2019) derived typologically-predicted patterns of human use of semantically-nuanced words, and demonstrated that some neural language models mimic these patterns. Practical limitations have prohibited the token-level, word-aligned techniques of Beekhuizen and colleagues from deployment for large-scale broad-coverage lexicons. However, there has been success in supersense tagging using coarser-grained type-level representations based on sentence alignments (Zennaki et al., 2019). Such results suggest that discovering similar methods for finer-grained representations that can scale is a promising avenue to pursue. Moreover, because their success depends on having a representative sample of languages, these kinds of approaches can inform how to sample languages efficiently to capture the broad crosslinguistic regularities (e.g., Stoll and Bickel, 2013; Beekhuizen and Stevenson, 2015; Beekhuizen et al., 2017). Thus, drawing on typological principles could extend the repertoire of NLP approaches to creation of multilingual spaces that truly generalize across languages.

5. CONCLUDING THOUGHTS

The research from adjoining fields of linguistics, psycholinguistics, and cognitive science provides many challenging targets—as well as many sources of inspiration—for learning more structured, adaptable, and generalizable models in NLP. We have highlighted a broad range of interdisciplinary work that indicates how these high-level goals translate into more specific questions and hypotheses about computational approaches to word meaning. Such studies have informed the current understanding of human-like lexical representations and the algorithms that can achieve them, and have highlighted the possibilities for future research aimed at bringing these insights into NLP.

First, interdisciplinary research points to the need to learn more richly-structured notions of semantic similarity and other types of relations. Representations must achieve a higher level of abstraction that identifies classes of words that capture generalizable knowledge. The overarching challenge is for neural architectures to learn such structured semantic spaces. Research suggests that drawing closer connections to human cognitive mechanisms, such as memory, attention, and communicative efficiency, can lead to insight into what to store in memory and how to effectively abstract and simplify representations. Second, to achieve human levels of lexical adaptability, we must move beyond correlation to causation: systems must be sensitive to the latent causal factors of the observed effects, in order to support generalizations that mirror the structure of the problem, and thus are both more predictive and more explainable. To do so will require consideration of cognitive mechanisms such as categorization, and incorporation of human-like biases, such that learning systems can adapt dynamically given small amounts of data (even within a conversation). Finally, the goal of true crosslinguistic validity will require lexical representations that conform to a universal conceptual foundation, and multilingual semantic spaces that reflect the understood mappings between language-specific lexicons. Practical learning algorithms will need to anticipate transfer effects when using more-resourced languages to leverage knowledge for less-resourced languages. Multilingual systems will also need to draw on the known dimensions of typological, historical, and structural variation to inform small but representative language samples to ensure crosslinguistic generalization.

We have undergone a paradigm shift in natural language processing due to the ability of recent machine learning methods to effectively process huge amounts of data. But the integration of machine learning methods into computational linguistics is not new. The statistical revolution of the 1990s led to tremendous advances in a vast array of applications, from machine translation to automatic lexical acquisition to summarization and more. It also eventually led to a realization that knowledge of the language sciences—that is, deep understanding of the findings from fields like linguistics and

psycholinguistics, on both the qualitative and the quantitative properties of language—were critical to obtaining success in NLP. Today, the same marvel of very large scale language models that is having such a positive effect on our ability to generate useful applications with relatively simple fine-tuning, has the negative effect of making us forget that grounded natural language processing is far from solved. Moreover, scientific progress is held back when resources and efforts are concentrated into the single mould of NLP as generic optimization, and away from questions and techniques that are more deeply integrated with the properties of the object of study. The language sciences have long (in some cases, thousands of years) revealed subtleties of the linguistic system that may be fruitfully incorporated into current approaches in NLP as knowledge representations, inductive biases, and principles of constrained variation.

AUTHOR CONTRIBUTIONS

PM and SS contributed equally to the ideas in and writing of this article and approved the submitted version.

FUNDING

SS was supported by grant RGPIN-2017-06506 from NSERC (Canada). PM gratefully acknowledges the partial support of the NCCR Evolving Language, Swiss NSF Agreement 51NF40_180888.

REFERENCES

Alexiadou, A. (2010). "On the morpho-syntax of (anti-) causative verbs," in *Lexical Semantics, Syntax, and Event Structure*, eds M. R. Hovav, E. Doron, and I. Sichel (Oxford: Oxford University Press), 177–203. doi: 10.1093/acprof:oso/9780199544325.003.0009

An, J., Kwak, H., and Ahn, Y.-Y. (2018). "SemAxis: a lightweight framework to characterize domain-specific word semantics beyond sentiment," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Melbourne, VIC: Association for Computational Linguistics). doi: 10.18653/v1/P18-1228

Armstrong, B. C., and Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Lang. Cogn. Neurosci.* 31, 940–966. doi: 10.1080/23273798.2016.1171366

Armstrong, S., Church, K., Isabelle, P., Manzi, S., Tzoukermann, E., and Yarowsky, D. (2010). *Natural Language Processing Using Very Large Corpora. Text, Speech and Language Technology*. Dordrecht: Springer Netherlands.

Arora, S., Li, Y., Liang, Y., Ma, T., and Risteski, A. (2018). Linear algebraic structure of word senses, with applications to polysemy. *Trans. Assoc. Comput. Linguist.* 6, 483–495. doi: 10.1162/tacl_a_00034

Artetxe, M., Labaka, G., and Agirre, E. (2017). "Learning bilingual word embeddings with (almost) no bilingual data," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Vancouver, BC), 451–462. doi: 10.18653/v1/P17-1042

Artetxe, M., Labaka, G., Lopez-Gazpio, I., and Agirre, E. (2018). "Uncovering divergent linguistic information in word embeddings with lessons for intrinsic and extrinsic evaluation," in *Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018* (Brussels), 282–291. doi: 10.18653/v1/K18-1028

Bahdanau, D., Cho, K., and Bengio, Y. (2015). "Neural machine translation by jointly learning to align and translate," in *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015* (San Diego, CA).

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). "The Berkeley FrameNet project," in *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics (ACL-COLING'98)* (Montreal, QC), 86–90. doi: 10.3115/980845.980860

Baker, M. (2003). *Lexical Categories: Verbs, Nouns and Adjectives*. Cambridge University Press. doi: 10.1017/CBO9780511615047

Baldwin, T., Korhonen, A., and Villavicencio, A. (eds.). (2005). *Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition*.

Baroni, M., Dinu, G., and Kruszewski, G. (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)* (Baltimore, MD), 238–247. doi: 10.3115/v1/P14-1023

Baroni, M., and Lenci, A. (2008). Concepts and properties in word spaces. *Italian J. Linguist.* 1, 55–88.

Baroni, M., and Lenci, A. (2010). Distributional memory: a general framework for corpus-based semantics. *Comput. Linguist.* 36, 673–721. doi: 10.1162/coli_a_00016

Bartsch, R. (1984). Norms, tolerance, lexical change, and context-dependence of meaning. *J. Pragmat.* 8, 367–393. doi: 10.1016/0378-2166(84)90029-8

Beckwith, R., Fellbaum, C., Gross, D., and Miller, G. (1991). "Wordnet: a lexical database organized on psycholinguistic principles," in *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon*, ed U. Zernik (Laurence Erlbaum), 211–232. doi: 10.4324/9781315785387-12

Beekhuizen, B., Armstrong, B. C., and Stevenson, S. (2021). Probing lexical ambiguity: word vectors encode number and relatedness of senses. *Cogn. Sci.* 45, 1–37. doi: 10.1111/cogs.12943

Beekhuizen, B., Cui, C. X., and Stevenson, S. (2019). "Representing lexical ambiguity in prototype models of lexical semantics," in *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (Montreal, QC).

Beekhuizen, B., Fazly, A., and Stevenson, S. (2014). "Learning meaning without primitives: typology predicts developmental patterns," in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

Beekhuizen, B., Milić, S., Armstrong, B., and Stevenson, S. (2018). "What company do semantically ambiguous words keep? Insights from distributional word vectors," in *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (Madison, WI).

Beekhuizen, B., and Stevenson, S. (2015). "Crowdsourcing elicitation data for semantic typologies," in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (Pasadena, CA).

Beekhuizen, B., and Stevenson, S. (2018). More than the eye can see: a computational model of color term acquisition and color discrimination. *Cogn. Sci.* 42, 2699–2734. doi: 10.1111/cogs.12665

Beekhuizen, B., Watson, J., and Stevenson, S. (2017). "Semantic typology and parallel corpora: Something about indefinite pronouns," in *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (London).

Belinkov, Y., and Bisk, Y. (2018). "Synthetic and natural noise both break neural machine translation," in *International Conference on Learning Representations (ICLR)*.

Bender, E. M. (2016). Linguistic typology in natural language processing. *Linguist. Typol.* 20, 645–660. doi: 10.1515/lingty-2016-0035

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). "On the dangers of stochastic parrots: can language models be too big?," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21* (New York, NY: Association for Computing Machinery), 610–623. doi: 10.1145/3442188.3445922

Bender, E. M., and Koller, A. (2020). "Climbing towards NLU: on meaning, form, and understanding in the age of data," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics), 5185–5198. doi: 10.18653/v1/2020.acl-main.463

Bengio, Y., Deleu, T., Rahaman, N., Ke, R., Lachapelle, S., Bilaniuk, O., et al. (2020). "A meta-transfer objective for learning to disentangle causal mechanisms," in

Proceedings of the Eighth International Conference on Learning Representations (ICLR 2020).

Berlin, B., and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA: UC Press.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist*. 5, 135–146. doi: 10.1162/tacl_a_00051

Boleda, G., Badia, T., and Batlle, E. (2004). "Acquisition of semantic classes for adjectives from distributional evidence," in *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics* (Geneva: COLING), 1119–1125. doi: 10.3115/1220355.1220516

Boleda, G., Gupta, A., and Padó, S. (2017). "Instances and concepts in distributional space," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (Valencia; Stroudsburg, PA: Association for Computational Linguistics), 79–85. doi: 10.18653/v1/E17-2013

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. doi: 10.48550/arXiv.2108.07258

Bowerman, M., and Choi, S. (2001). *Shaping Meanings for Language: Universal and Language-Specific in the Acquisition of Semantic Categories*. Cambridge: Cambridge University Press. 475–511. doi: 10.1017/CBO9780511620669.018

Bréal, M. (1897). *Essai de sémantique: Science des Significations*. Hachette.

Brennan, S. E., and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol*. 22:1482. doi: 10.1037/0278-7393.22.6.1482

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). "Language models are few-shot learners," in *Advances in Neural Information Processing Systems, Vol. 33*, eds H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc.), 1877–1901.

Buchanan, L., Westbury, C., and Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychon. Bull. Rev*. 8, 531–544. doi: 10.3758/BF03196189

Budanitsky, A., and Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. *Comput. Linguist*. 32, 13–47. doi: 10.1162/coli.2006.32.1.13

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2009). "8. Using FrameNet for the semantic analysis of German: Annotation, representation, and automation," in *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, ed H. Boas (Berlin; New York, NY: De Gruyter Mouton), 209–244. doi: 10.1515/9783110212976.3.209

Burgess, C. (1998). From simple associations to the building blocks of language: modeling meaning in memory with the HAL model. *Behav. Res. Methods Instrum. Comput*. 30, 188–198. doi: 10.3758/BF03200643

Burgess, C. (2001). "Representing and resolving semantic ambiguity: a contribution from high-dimensional memory modeling," in *On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity*, ed D. S. Gorfein (Washington, DC: American Psychological Association), 233–261. doi: 10.1037/10459-013

Bybee, J., and Eddington, D. (2006). A usage-based approach to Spanish verbs of 'becoming'. *Language* 82, 323–355. doi: 10.1353/lan.2006.0081

Choi, H., Cho, K., and Bengio, Y. (2017). Context-dependent word representation for neural machine translation. *Comput. Speech Lang*. 45, 149–160. doi: 10.1016/j.csl.2017.01.007

Chronis, G., and Erk, K. (2020). "When is a bishop not like a rook? When it's like a rabbi! Multi-prototype BERT embeddings for estimating semantic relationships," in *Proceedings of the 24th Conference on Computational Natural Language Learning (CoNLL-2020)*, 227–244. doi: 10.18653/v1/2020.conll-1.17

Church, K., and Liberman, M. (2021). The future of computational linguistics: on beyond alchemy. *Front. Artif. Intell*. 4:10. doi: 10.3389/frai.2021.625341

Church, K. W. (2017). Emerging trends: i did it, i did it, i did it, but... *Nat. Lang. Eng*. 23, 473–480. doi: 10.1017/S1351324917000067

Clark, E. V. (1987). "The principle of contrast: a constraint on language acquisition," in *Mechanisms of Language Acquisition*, ed B. MacWhinney (Hillsdale, NJ: Erlbaum), 1–33.

Clark, E. V., and Clark, H. H. (1979). When nouns surface as verbs. *Language* 767–811. doi: 10.2307/412745

Clark, P., Dalvi, B., and Tandon, N. (2018). What happened? Leveraging VerbNet to predict the effects of actions in procedural text. *arXiv preprint arXiv:1804.05435*. doi: 10.48550/arXiv.1804.05435

Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev*. 82:407. doi: 10.1037/0033-295X.82.6.407

Collobert, R., and Weston, J. (2007). "Fast semantic extraction using a novel neural network architecture," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (Prague), 560–567.

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, ICML '08* (New York, NY: Association for Computing Machinery), 160–167. doi: 10.1145/1390156.1390177

Cook, P., Lau, J. H., McCarthy, D., and Baldwin, T. (2014). "Novel word-sense identification," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (Dublin), 1624–1635.

Croft, W. (1994). Semantic universals in classifier systems. *Word* 45, 145–171. doi: 10.1080/00437956.1994.11435922

Croft, W., and Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge University Press. doi: 10.1017/CBO9780511803864

Cruse, A. (2000). "Aspects of the micro-structure of word meanings," in *Polysemy: Theoretical and Computational Approaches*, eds Y. Ravin and C. Leacock (Oxford University Press), 30–51.

Cruse, D. A. (1986). Lexical semantics. *Cambridge University Press*.

Das, D., Chen, D., Martins, A. F. T., Schneider, N., and Smith, N. A. (2014). Frame-semantic parsing. *Comput. Linguist*. 40, 9–56. doi: 10.1162/COLI_a_00163

Degani, T., Prior, A., and Tokowicz, N. (2011). Bidirectional transfer: the effect of sharing a translation. *J. Cogn. Psychol*. 23, 18–28. doi: 10.1080/20445911.2011.445986

Del Tredici, M., Fernández, R., and Boleda, G. (2018). Short-term meaning shift: a distributional exploration. *arXiv preprint arXiv:1809.03169*. doi: 10.18653/v1/N19-1210

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805 [cs.CL]*. doi: 10.48550/arXiv.1810.04805

Drozd, A., Gladkova, A., and Matsuoka, S. (2016). "Word embeddings, analogies, and machine learning: Beyond king - man + woman = queen," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (Osaka: The COLING 2016 Organizing Committee), 3519–3530.

Dubossarsky, H., McCarthy, A. D., Ponti, E. M., Vulić, I., Vylomova, E., Berzak, Y., et al. (eds.). (2019). *Proceedings of TyP-NLP: The First Workshop on Typology for Polyglot NLP*. Florence: Association for Computational Linguistics.

Ellison, T. M., (ed.). (1997). CoNLL97: computational natural language learning," in *Proceedings of the 1997 Meeting of the ACL Special Interest Group in Natural Language Learning*.

Erk, K., and Padó, S. (2008). "A structured vector space model for word meaning in context," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (Honolulu: Association for Computational Linguistics), 897–906. doi: 10.3115/1613715.1613831

Ethayarajh, K. (2019). How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. *arXiv preprint arXiv:1909.00512*. doi: 10.18653/v1/D19-1006

Ettinger, A. (2020). What BERT is not: lessons from a new suite of psycholinguistic diagnostics for language models. *Trans. Assoc. Comput. Linguist*. 8, 34–48. doi: 10.1162/tacl_a_00298

Ettinger, A., Rao, S., Daumé III, H., and Bender, E. M. (2017). Towards linguistically generalizable NLP systems: a workshop and shared task. *arXiv preprint arXiv:1711.01505*. doi: 10.18653/v1/W17-5401

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/7287.001.0001

Fillmore, C. J., and Atkins, B. T. S. (1998). "Framenet and lexicographic relevance," in *Proceedings of the First International Conference on Language Resources and Evaluation* (Granada).

Firth, J. R. (1957). "A synopsis of linguistic theory 1930-1955," in *Studies in Linguistic Analysis* (London: Philological Society), 1–32.

Franck, J., Colonna, S., and Rizzi, L. (2015). Task-dependency and structure dependency in number interference effects in sentence comprehension. *Front. Psychol.* 6:807. doi: 10.3389/fpsyg.2015.00807

Geeraerts, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cogn. Linguist.* 4, 223–272. doi: 10.1515/cogl.1993.4.3.223

Gentner, D., and Bowerman, M. (2009). "Chapter 34: Why some spatial semantic categories are harder to learn than others. The Typological Prevalence Hypothesis," in *Crosslinguistic Approaches to the Psychology of Language. Research in the Tradition of Dan Isaac Slobin*, eds J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, and S. Özcaliskan (New York, NY: Psychology Press), 465–480.

Girju, R., Badulescu, A., and Moldovan, D. (2006). Automatic discovery of part-whole relations. *Comput. Linguist.* 32, 83–135. doi: 10.1162/089120106776173075

Goddard, C., and Wierzbicka, A. (eds.). (1994). *Semantic and Lexical Universals: Theory and Empirical Findings*. John Benjamins Publishing. doi: 10.1075/slcs.25

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.

Gouws, S., Bengio, Y., and Corrado, G. (2015). "BilBOWA: fast bilingual distributed representations without word alignments," in *32nd International Conference on Machine Learning* (Lille), 748–756.

Grand, G., Blank, I. A., Pereira, F., and Fedorenko, E. (2018). Semantic projection: recovering human knowledge of multiple, distinct object features from word embeddings. *arXiv preprint arXiv:1802.01241*. doi: 10.48550/arXiv.1802.01241

Grefenstette, G. (1994). "Corpus-derived first, second and third-order word affinities," in *Proceedings of the Sixth Euralex International Congress*.

Griffiths, T. L., Steyvers, M., and Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychol. Rev.* 114, 211–244. doi: 10.1037/0033-295X.114.2.211

Habibi, A. A., Kemp, C., and Xu, Y. (2020). Chaining and the growth of linguistic categories. *Cognition* 202:104323. doi: 10.1016/j.cognition.2020.104323

Hahn, M., and Keller, F. (2021). Modeling task effects in human reading with neural attention. *arxiv.org/abs/1808.00054*. doi: 10.48550/arXiv.1808.00054

Hale, K., and Keyser, S. J. (2002). *Prolegomena to a Theory of Argument Structure, Vol. 39 of Linguistic Inquiry Monograph*. Cambridge, MA: MIT Press.

Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2016). "Diachronic word embeddings reveal statistical laws of semantic change," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1489–1501. doi: 10.18653/v1/P16-1141

Harris, Z. (1954). Distributional structure. *Word* 10, 146–162. doi: 10.1080/00437956.1954.11659520

Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. *Causat. Transit.* 23:87. doi: 10.1075/slcs.23.05has

Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford University Press.

Haspelmath, M., Calude, A., Spagnol, M., Narrog, H., and Bamyac., E. (2014). Coding causal-noncausal verb alternations: a form-frequency correspondence explanation. *J. Linguist.* 50, 587–625. doi: 10.1017/S0022226714000255

Hearst, M. A. (1992). "Automatic acquisition of hyponyms from large text corpora," in *Coling 1992 Volume 2: The 15th International Conference on Computational Linguistics*. doi: 10.3115/992133.992154

Heidinger, S. (2015). Causalness and the encoding of the causative-anticausative alternation in French and Spanish. *J. Linguist.* 51, 562–594. doi: 10.1017/S0022226714000607

Hino, Y., Pexman, P. M., and Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: are they due to semantic coding? *J. Mem. Lang.* 55, 247–273. doi: 10.1016/j.jml.2006.04.001

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hofmann, V., Pierrehumbert, J. B., and Schütze, H. (2020). Dynamic contextualized word embeddings. *arXiv preprint arXiv:2010.12684*. doi: 10.18653/v1/2021.acl-long.542

Hollis, G., and Westbury, C. (2016). The principals of meaning: extracting semantic dimensions from co-occurrence models of semantics. *Psychon. Bull. Rev.* 23, 1744–1756. doi: 10.3758/s13423-016-1053-2

Hollis, G., Westbury, C., and Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *Q. J. Exp. Psychol.* 8, 1603–1619. doi: 10.1080/17470218.2016.1195417

Iordan, M. C., Giallanza, T., Ellis, C. T., Beckage, N. M., and Cohen, J. D. (2022). Context matters: Recovering human semantic structure from machine learning analysis of large-scale text corpora. *Cogn. Sci.* 46:e13085. doi: 10.1111/cogs.13085

Jamieson, R. K., Avery, J. E., Johns, B. T., and Jones, M. N. (2018). An instance theory of semantic memory. *Comput. Brain Behav.* 1, 119–136. doi: 10.1007/s42113-018-0008-2

Jones, M. N., Willits, J., and Dennis, S. (2015). "Models of semantic memory," in *Oxford Handbook of Mathematical and Computational Psychology*, eds J. R. Busemeyer, Z. Wang, J. T. Townsend, and A. Eidels (Oxford University Press), 232–254. doi: 10.1093/oxfordhb/9780199957996.013.11

Kann, K., Warstadt, A., Williams, A., and Bowman, S. R. (2019). "Verb argument structure alternations in word and sentence embeddings," in *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, 287–297.

Kemp, C., Xu, Y., and Regier, T. (2018). Semantic typology and efficient communication. *Annu. Rev. Linguist.* 4, 109–128. doi: 10.1146/annurev-linguistics-011817-045406

Kintsch, W. (2001). Predication. *Cogn. Sci.* 25, 173–202. doi: 10.1207/s15516709cog2502_1

Kipper, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon* (Ph.D. thesis). University of Pennsylvania, Philadelphia, PA, United States.

Klepousniotou, E., Titone, D., and Romero, C. (2008). Making sense of word senses: the comprehension of polysemy depends on sense overlap. *J. Exp. Psychol.* 34, 1534–1543. doi: 10.1037/a0013012

Köper, M., Scheible, C., and Schulte im Walde, S. (2015). "Multilingual reliability and "semantic" structure of continuous word spaces," in *Proceedings of the 11th International Conference on Computational Semantics*, 40–45.

Korhonen, A., and Briscoe, T. (2004). "Extended lexical-semantic classification of English verbs," in *Proceedings of the Computational Lexical Semantics Workshop at HLT-NAACL 2004* (Boston, MA), 38–45. doi: 10.3115/1596431.1596437

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behav. Brain Sci.* 40:e253. doi: 10.1017/S0140525X16001837

Lample, G., Conneau, A., Ranzato, M., Denoyer, L., and Jégou, H. (2018). "Word translation without parallel data," in *6th International Conference on Learning Representations* (Vancouver, BC), 1–14.

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. *Psychol. Rev.* 104, 211–240. doi: 10.1037/0033-295X.104.2.211

Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

Lenci, A. (2008). Distributional approaches in linguistic and cognitive research. *Italian J. Linguist.* 20, 1–31.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL: University of Chicago Press.

Levinson, S., Meira, S., Language, T., and Group, C. (2003). "Natural concepts" in the spatial topological domain—adpositional meanings in crosslinguistic perspective: an exercise in semantic typology. *Language* 79, 485–516. doi: 10.1353/lan.2003.0174

Levy, O., Goldberg, Y., and Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Trans. Assoc. Comput. Linguist.* 3, 211–225. doi: 10.1162/tacl_a_00134

Levy, O., Søgaard, A., and Goldberg, Y. (2017). "A strong baseline for learning cross-lingual word embeddings from sentence alignments," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (Valencia: Association for Computational Linguistics), 765–774. doi: 10.18653/v1/E17-1072

Li, F.-F., Fergus, R., and Perona, P. (2006). One-shot learning of object categories. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 594–611. doi: 10.1109/TPAMI.2006.79

Li, J., and Brew, C. (2008). "Which are the best features for automatic verb classification," in *Proceedings of ACL-08: HLT* (Columbus, OH), 434–442.

Li, J., and Jurafsky, D. (2015). "Do multi-sense embeddings improve natural language understanding?," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, eds L. Márquez, C. Callison-Burch, J. Su, D. Pighin, and Y. Marton (The Association for Computational Linguistics), 1722–1732. doi: 10.18653/v1/D15-1200

Lu, H., Wu, Y. N., and Holyoak, K. J. (2019). Emergence of analogy from relation learning. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4176-4181. doi: 10.1073/pnas.1814779116

Lund, K. and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput*. 28, 203–208. doi: 10.3758/BF03204766

Majid, A., Boster, J. S., and Bowerman, M. (2008). The cross-linguistic categorization of everyday events: a study of cutting and breaking. *Cognition* 109, 235–250. doi: 10.1016/j.cognition.2008.08.009

Majid, A., and Van Staden, M. (2015). Can nomenclature for the body be explained by embodiment theories? *Top. Cogn. Sci.* 7, 570–594. doi: 10.1111/tops.12159

Markman, E. M. (1987). "How children constrain the possible meanings of words," in *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization, Vol. 1*, ed U. Neisser (New York, NY: Cambridge University Press), 255–287.

McCoy, R. T., Grant, E., Smolensky, P., Griffiths, T. L., and Linzen, T. (2020). Universal linguistic inductive biases *via* meta-learning. *arXiv preprint arXiv:2006.16324*. doi: 10.48550/arXiv.2006.16324

Merlo, P. (2019). "Probing word and sentence embeddings for long-distance dependencies effects in French and English," in *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP* (Florence: Association for Computational Linguistics), 158–172. doi: 10.18653/v1/W19-4817

Merlo, P., and Ackermann, F. (2018). "Vectorial semantic spaces do not encode human judgments of intervention similarity," in *Proceedings of the 22nd Conference on Computational Natural Language Learning* (Brussels: Association for Computational Linguistics), 392–401. doi: 10.18653/v1/K18-1038

Merlo, P., and Rodriguez, M. A. (2019). "Cross-lingual word embeddings and the structure of the human bilingual lexicon," in *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)* (Hong Kong), 110–120. doi: 10.18653/v1/K19-1011

Merlo, P., and Stevenson, S. (2001). Automatic verb classification based on statistical distributions of argument structure. *Comput. Linguist*. 27, 373–408. doi: 10.1162/089120101317066122

Merlo, P., and Stevenson, S. (2005). "Structure and frequency in verb classification," in *Volume Collecting Papers From the 30th Incontro di Grammatica Generativa*, eds L. Bruge, G. Giusti, N. Munaro, W. Schweikert, and G. Turano (Venice).

Merlo, P., Stevenson, S., Tsang, V., and Allaria, G. (2002). "A multi-lingual paradigm for automatic verb classification," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)* (Philadelphia, PA), 207–214. doi: 10.3115/1073083.1073119

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *CoRR, abs/1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). "Distributed representations of words and phrases and their compositionality," in *Neural and Information Processing System (NIPS)*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013c). "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems 26*, eds C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger (Lake Tahoe: Curran Associates, Inc.), 3111–3119.

Miller, G. A. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Mitchell, J., and Lapata, M. (2008). "Vector-based models of semantic composition," in *Proceedings of ACL-08: HLT* (Columbus, OH), 236–244.

Mitchell, J., and Lapata, M. (2010). Composition in distributional models of semantics. *Cogn. Sci*. 34, 1388–1429. doi: 10.1111/j.1551-6709.2010.01106.x

Morzycki, M. (2012). Adjectival extremeness: degree modification and contextually restricted scales. *Nat. Lang. Linguist. Theory* 30, 567–609. doi: 10.1007/s11049-011-9162-0

Mu, J., Bhat, S., and Viswanath, P. (2017). "Geometry of polysemy," in *Proceedings of the International Conference on Learning Representations*.

Navigli, R., and Ponzetto, S. P. (2010). "BabelNet: building a very large multilingual semantic network," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (Uppsala: Association for Computational Linguistics), 216–225.

Nematzadeh, A., Meylan, S. C., and Griffiths, T. L. (2017). "Evaluating vector-space models of word representation, or, the unreasonable effectiveness of counting words near other words," in *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (London).

Nematzadeh, A., Ruder, S., and Yogatama, D. (2020). "On memory in human and artificial language processing systems," in *Proceedings of ICLR Workshop on Bridging AI and Cognitive Science*.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: polysemy. *Linguist. Philos*. 3, 143–184. doi: 10.1007/BF00126509

Padó, S., and Lapata, M. (2005). "Cross-linguistic projection of role-semantic information," in *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (Vancouver, BC), 859–866. doi: 10.3115/1220575.1220683

Padó, S., and Lapata, M. (2007). Dependency-based construction of semantic space models. *Comput. Linguist*. 33, 161–199. doi: 10.1162/coli.2007.33.2.161

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: an annotated corpus of semantic roles. *Comput. Linguist*. 31, 71–105. doi: 10.1162/0891201053630264

Paradis, C. (2001). Adjectives and boundedness. *Cogn. Linguist*. 12, 47–65. doi: 10.1515/cogl.12.1.47

Pater, J. (2019). Generative linguistics and neural networks at 60: foundation, friction, and fusion. *Language* 95, e41?-e74. doi: 10.1353/lan.2019.0009

Patwardhan, S., Banerjee, S., and Pedersen, T. (2003). "Using measures of semantic relatedness for word sense disambiguation," in *International Conference on Intelligent Text Processing and Computational Linguistics* (Berlin: Springer), 241–257. doi: 10.1007/3-540-36456-0_24

Pennington, J., Socher, R., and Manning, C. D. (2014). "Glove - global vectors for word representation," in *2014 Conference on Empirical Methods in Natural Language Processing* (Doha), 1532–1543. doi: 10.3115/v1/D14-1162

Pereira, F., Gershman, S., Ritter, S., and Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cogn. Neuropsychol*. 33, 175–190. doi: 10.1080/02643294.2016.1176907

Perek, F. (2016). Using distributional semantics to study syntactic productivity in diachrony: a case study. *Linguistics* 54, 149–188. doi: 10.1515/ling-2015-0043

Perez, E., Kiela, D., and Cho, K. (2021). True few-shot learning with language models. *CoRR, abs/2105.11447*. doi: 10.48550/arXiv.2105.11447

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., et al. (2018). "Deep contextualized word representations," in *Proceedings of NAACL*. doi: 10.18653/v1/N18-1202

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Rabinovich, E., Ordan, N., and Wintner, S. (2017). "Found in translation: reconstructing phylogenetic language trees from translations," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 530–540. doi: 10.18653/v1/P17-1049

Rabinovich, E., Watson, J., Beekhuizen, B., and Stevenson, S. (2019). "Say anything: automatic semantic infelicity detection in L2 English indefinite pronouns," in *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)* (Hong Kong), 77–86. doi: 10.18653/v1/K19-1008

Rabinovich, E., Xu, Y., and Stevenson, S. (2020). "The typology of polysemy: a multilingual distributional framework," in *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Regier, T., Kay, P., and Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proc. Natl. Acad. Sci. U.S.A*. 104, 1436–1441. doi: 10.1073/pnas.0610341104

Reisinger, J., and Mooney, R. J. (2010). "Multi-prototype vector-space models of word meaning," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (Los Angeles, CA: Association for Computational Linguistics), 109–117.

Riloff, E. (1996). An empirical study of automated dictionary construction for information extraction in three domains. *Artif. Intell*. 85, 101–134. doi: 10.1016/0004-3702(95)00123-9

Ritter, S., Barrett, D. G., Santoro, A., and Botvinick, M. M. (2017). "Cognitive psychology for deep neural networks: a shape bias case study," in *International Conference on Machine Learning*, 2940–2949.

Rizzi, L. (2004). "Locality and left periphery," in *The Cartography of Syntactic Structures, Number 3 in Structures and Beyond*, ed A. Belletti (New York, NY: Oxford University Press), 223–251.

Rodd, J. M., Gaskell, G., and Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: semantic competition in lexical access. *J. Mem. Lang.* 46, 245–266. doi: 10.1006/jmla.2001.2810

Rodriguez, M. A., and Merlo, P. (2020). "Word associations and the distance properties of context-aware word embeddings," in *Proceedings of the 24th Conference on Computational Natural Language Learning* (Association for Computational Linguistics), 376–385. doi: 10.18653/v1/2020.conll-1.30

Rogers, A., Drozd, A., and Li, B. (2017). "The (too many) problems of analogical reasoning with word vectors," in *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (* SEM 2017)*, 135–148. doi: 10.18653/v1/S17-1017

Rubinstein, D., Levi, E., Schwartz, R., and Rappoport, A. (2015). "How well do distributional models capture different types of semantic knowledge?," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 726–730. doi: 10.3115/v1/P15-2119

Ryskina, M., Rabinovich, E., Berg-Kirkpatrick, T., Mortensen, D. R., and Tsvetkov, Y. (2020). "Where new words are born: distributional semantic analysis of neologisms and their semantic neighborhoods," in *Proceedings of the Society for Computation in Linguistics*.

Sahlgren, M. (2008). The distributional hypothesis. *Italian J. Linguist.* 20, 33–53.

Samardžić, T., and Merlo, P. (2010). "Cross-lingual variation of light verb constructions: using parallel corpora and automatic alignment for linguistic research," in *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground* (Uppsala: Association for Computational Linguistics), 52–60.

Samardžić, T., and Merlo, P. (2018). Probability of external causation: an empirical account of cross-linguistic variation in lexical causatives. *Linguistics* 56, 895–938. doi: 10.1515/ling-2018-0001

Samir, F., Beekhuizen, B., and Stevenson, S. (2021). "A formidable ability: detecting adjectival extremeness with DSMs," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 4112–4125. doi: 10.18653/v1/2021.findings-acl.360

Samir, F., Stevenson, S., and Beekhuizen, B. (2020). "Untangling semantic similarity: Modeling lexical processing experiments with distributional semantic models," in *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Samuelson, L. K., and Smith, L. B. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition* 73, 1–33. doi: 10.1016/S0010-0277(99)00034-7

Santus, E., Gladkova, A., Evert, S., and Lenci, A. (2016). "The CogALex-V shared task on the corpus-based identification of semantic relations," in *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*, 69–79.

Schick, T., and Schütze, H. (2021). "Exploiting cloze-questions for few-shot text classification and natural language inference," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (Association for Computational Linguistics), 255–269. doi: 10.18653/v1/2021.eacl-main.20

Schölkopf, B. (2019). Causality for machine learning. *arXiv:1911.10500v2*.

Schölkopf, B., Janzing, D., Peters, J., Sgouritsa, E., Zhang, K., and Mooij, J. (2012). "On causal and anticausal learning," in *Proceedings of the 29th International Conference on Machine Learning* (Edinburgh).

Schulte im Walde, S. (2006). Experiments on the automatic induction of German semantic verb classes. *Comput. Linguist.* 32, 159–194. doi: 10.1162/coli.2006.32.2.159

Schütze, H. (1994). "Distibutional part-of-speech tagging," in *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics* (Dublin). doi: 10.3115/976973.976994

Schütze, H., and Pedersen, J. (1993). "A vector model for syntagmatic and paradigmatic relatedness," in *Making Sense of Words: Proceedings of the Conference*, 104–113.

Shi, W., Chen, M., Zhou, P., and Chang, K.-W. (2019). Retrofitting contextualized word embeddings with paraphrases. *arXiv preprint arXiv:1909.09700*. doi: 10.18653/v1/D19-1113

Snyder, B., and Barzilay, R. (2008). "Unsupervised multilingual learning for morphological segmentation," in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT)* (Columbus, OH), 737–745.

Sood, E., Tannert, S., Frassinelli, D., Bulling, A., and Vu, N. T. (2020). "Interpreting attention models with human visual attention in machine reading comprehension," in *Proceedings of ACL SIGNLL Conference on Computational Natural Language Learning (CoNLL)* (Association for Computational Linguistics), 12–25. doi: 10.18653/v1/2020.conll-1.2

Stevenson, S., and Joanis, E. (2003). "Semi-supervised verb class discovery using noisy features," in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003* (Edmonton, AB), 71–78. doi: 10.3115/1119176.1119186

Stoll, S., and Bickel, B. (2013). "Capturing diversity in language acquisition research," in *Language Typology and Historical Contingency, in Honor of Johanna Nichols, Number 104 in Typological Studies in Language*, eds B. Bickel and L. A. Grenoble (Amsterdam; Philadelphia, PA: John Benjamins), 195–216. doi: 10.1075/tsl.104.08slo

Sun, L. and Korhonen, A. (2009). "Improving verb clustering with automatically acquired selectional preferences," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (Singapore: Association for Computational Linguistics), 638–647. doi: 10.3115/1699571.1699596

Suttle, L., and Goldberg, A. E. (2011). The partial productivity of constructions as induction. *Linguistics* 49, 1237–1269. doi: 10.1515/ling.2011.035

Thompson, B., Roberts, S., and Lupyan, G. (2018). "Quantifying semantic similarity across languages," in *Proceedings of the 40th Annual Conference of the Cognitive Science Society (CogSci 2018)* (Madison, WI).

Tishby, N., and Zaslavsky, N. (2015). "Deep learning and the information bottleneck principle," in *2015 IEEE Information Theory Workshop (ITW)*, 1–5. doi: 10.1109/ITW.2015.7133169

Tsang, V., Stevenson, S., and Merlo, P. (2002). "Cross-linguistic transfer in automatic verb classification," in *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)* (Taipei), 1023–1029. doi: 10.3115/1072228.1072374

Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cogn. Linguist.* 4, 273–290. doi: 10.1515/cogl.1993.4.3.273

Van de Cruys, T., Poibeau, T., and Korhonen, A. (2011). "Latent vector weighting for word meaning in context," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (Edinburgh: Association for Computational Linguistics), 1012–1022.

Van Hell, J. G., and de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: effects of concreteness and cognate status in word association. *Bilingualism* 1, 193–211. doi: 10.1017/S1366728998000352

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in Neural Information Processing Systems, Vol. 30*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Long Beach, CA: Curran Associates, Inc.).

Vossen, P. (1998). "Introduction to EuroWordNet," in *EuroWordNet: A Multilingual Database With Lexical Semantic Networks*, ed P. Vossen (Dordrecht: Springer), 1–17. doi: 10.1007/978-94-017-1491-4_1

Wang, H., Henderson, J., and Merlo, P. (2019). "Weakly-supervised concept-based adversarial learning for cross-lingual word embeddings," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Hong Kong: Association for Computational Linguistics), 4419–4430. doi: 10.18653/v1/D19-1450

Warstadt, A., Cao, Y., Grosu, I., Peng, W., Blix, H., Nie, Y., et al. (2019). "Investigating BERT's knowledge of language: five analysis methods with NPIs," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Hong Kong: Association for Computational Linguistics), 2877–2887. doi: 10.18653/v1/D19-1286

Watson, J., Samir, F., Stevenson, S., and Beekhuizen, B. (2021). "Coin it up: generalization of creative constructions in the wild," in *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*.

Williams, J. N. (1992). Processing polysemous words in context: evidence for interrelated meanings. *J. Psycholinguist. Res.* 21, 193–218. doi: 10.1007/BF01068072

Wittgenstein, L. (1953). *Philosophical Investigations*. Blackwell Publishing.

Wu, S., and Dredze, M. (2019). "Beto, bentz, becas: the surprising cross-lingual effectiveness of BERT," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Hong Kong: Association for Computational Linguistics), 833–844. doi: 10.18653/v1/D19-1077

Xu, Y., Duong, K., Malt, B. C., Jiang, S., and Srinivasan, M. (2020). Conceptual relations predict colexification across languages. *Cognition* 201:104280. doi: 10.1016/j.cognition.2020.104280

Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., et al. (2016). On the universal structure of human lexical semantics. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1766–1771. doi: 10.1073/pnas.1520752113

Yu, L., El Sanyoura, L., and Xu, Y. (2020). "How nouns surface as verbs: Inference and generation in word class conversion," in *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Zaslavsky, N., Kemp, C., Regier, T., and Tishby, N. (2018). Efficient human-like semantic representations *via* the information bottleneck principle. *arXiv preprint arXiv:1808.03353*. doi: 10.48550/arXiv.1808.03353

Zawada, B. (2006). Linguistic creativity from a cognitive perspective. *Southern Afr. Linguist. Appl. Lang. Stud.* 24, 235–254. doi: 10.2989/16073610609486419

Zennaki, O., Semmar, N., and Besacier, L. (2019). A neural approach for inducing multilingual resources and natural language processing tools for low-resource languages. *Nat. Lang. Eng.* 25, 43–67. doi: 10.1017/S1351324918000293

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.