Tracing Theories of Meaning Representation

Angélique Charles-Davis

1. Introduction

The issue of defining linguistic meaning and creating representations using these definitions has its origins in philosophy, but has come to be tremendously influential in the tremendous success of today's Large Language Models (Boleda, 2020; Frege, 1892; Partee, 1996). This paper examines two approaches to meaning representation: formal semantics, a logic-based model, theoretic perspective on meaning, and distributional semantics, a structural, statistical approach. We explore their linguistic origins, their treatment of semantic phenomena, and whether or how they should be resolved in the context of the broader linguistic landscape.

2. The Rise of Formal Semantics

The inception of formal semantics as it is understood today is marked by Richard Montague's *Universal Grammar*. Montague famously began his work with the statement:

> "There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians" (Montague, 1970, p. 373)

Montague's claim, and the rigorous formal work that accompanied it, stood alone in a time during which logicians and philosophers seemed to agree that the structure of formal logic could not accommodate the ambiguities and the idiosyncracies of natural language (Partee, 1996). However, it was preceded by a philosophical tradition, launched by Frege, that had already begun to treat language as a system that exists separately from the minds that make use of it, and whose semantic components can therefore be independently described and formalized in various theories of meaning.

In *On Sense and Reference* Frege articulates his meaning theory as:

> "A proper name (word, sign, sign-compound, expression) expresses its sense, and designates or signifies its nominatum." (Frege, 1892, p. 189)

For any given word or compound, its reference, or nominatum, is its physical manifestation in the world, while its sense, or meaning, is the manner in which this referent is

designated (Frege, 1892). In this way, 'Clark Kent' and 'Superman' can be taken to have different meanings and senses, even though they refer to the same person, simply because they *need not* refer to the same person. The work of Frege and others including Carnap and Tarski helped to relax the canonical linguistic notion that language was an issue of psychology, and instead began to suggest that it could be an issue of math or modeling (Partee, 1996).

The model-theoreticity that Montague would enlist in his grammar also emerged from work in philosophical logic. Logician-philosophers such as Kripke and Kanger worked to define modality as a function over possible worlds, thus enriching understanding of how possible states of affairs influence truth conditions. This was coupled with the notion of possible models of a language, which are used directly to determine semantic entailment. The result of these paired intuitions about how meaning is determined relative to a model was that given any logical statement, both its truth-conditional and intensional meaning could be derived (Gamut, 1991). In the statements "All John's pets are mammals" and "All John's dogs are mammals," we derive the truth value of the first using the model of the language, and in a model in which both sentences are true, we distinguish their meanings using possible worlds, the former being contingently true, and the latter necessarily true. This also allows for models of tensed languages, in which we draw distinctions between the model at different moments in time, while still integrating sets of possible worlds at each moment in time.

While applications of model-theoretic techniques in modal logic were extended to explain the logical underpinnings of tense, indexicality, demonstratives and other intensional features of language, the division remained between the formalisms logicians were studying and the natural language from which the issues of interest derived (Partee, 1996).

Montague's *English as a Formal Language* explicitly laid out his project to formalize the syntax and semantics of English by building on existing practice in logical theory. In his subsequent *Universal Grammar* and *PTQ*, he details the procedures for doing so. According to Montague, in order for a grammar to be a sufficient account of a language, it must consist of a syntax algebra, a semantics algebra, and a homomorphism between the two. The existence of this homomorphism entails compositionality, and it is this feature of the grammar that lends it such theoretical power (Montague, 1970; Partee, 1996). Syntax must provide a structural basis for semantic interpretation, and must do so up to any meaningful fragment size.

The compositionality inherent to Montague's theory was attractive to linguists for multiple reasons. Firstly, it helped to account for semantic competence and productivity in humans. Secondly, Montague's work coincided with the 'linguistic wars' between generative and interpretive linguistics and offered a fruitful alternative outlet. While both believed that surface syntactic structure emerged from a deep internal mental structure via a series of transformations, they differed in how deep they believed the semantic interpretation of these syntactic structures had to lie (Partee, 1996). What Montague's grammar

offered was a conservative and syntax-driven underlying structure from which semantic interpretation could be readily performed. The underlying syntactic form could be demystified as something inherently logical and which operated alongside semantic work to generate meaning. This semantic work became something to be studied. In a departure from a Chomskyan tradition in which syntax was scrutinized with little to no regard for semantics, linguists began to involve themselves in the philosophical tradition and generate compelling accounts for semantic phenomena such as tense and aspect, mass nouns, questions, and modality (Janssen & Zimmermann, 2021; Partee, 1996; Speaks, 2024).

By the end of his contributions, Montague's grammar consisted of a categorial grammar representing the syntactic algebra and a lambda calculus implementation of the semantics, which was particularly well suited to representing the basic process of function application that underlies his semantic system. Adjectives are functions that are applied to nouns, noun phrases are generalized quantifier functions, and so on (Janssen & Zimmermann, 2021; Partee, 1996).

Montague Grammar remains a touchstone in contemporary formal semantics, which largely consists of differing approaches to solving empirical linguistic questions, leaving behind many of the foundational philosophical questions that drove the field's inception (Partee, 1996).

3. The Rise of Distributional Semantics

The general shape of distributional semantics as it exists now arose from a blending of early work in information theory, late American structuralism, and anthropology. Namely, the statistical modeling of Claude Shannon and the corpus-based, empiricist work of Zellig Harris and J.R. Firth each contributed to the development of a meaning representation paradigm based in statistical information about context (Brunila & LaViolette, 2022).

Shannon's 1948 paper, *A Mathematical Theory of Computation*, laid the groundwork for understanding language as a statistical system. Shannon introduced the idea of n-grams, strings of n words or letters, and described language generation as the process of selecting the most likely n-gram given an (n-1)-gram and a corpus. He also highlighted redundancy in language, which significantly constrains the number of candidates in any instance of language generation. The information content of a linguistic unit he defined as a negative log probability of its occurrence, such that low probability events carry more information that high probability events. In other words, systems with lower redundancy will generally have more potential information to convey (Shannon, 1948).

Across modern NLP research, both Harris and Firth are regularly cited as originators of the theory behind the embeddings that power models. Firth's famous quote: "You shall know a word by the company it keeps," is invoked freely and frequently to essentialize the broader argument that collocation can not only serve as a proxy for meaning, but is itself what constitutes it (Brunila & LaViolette, 2022).

While both Harris and Firth's views are grounded in usage-based representations that highlight collocation, this is only a small part of the story. Harris rose to prominence during World War II efforts to mechanize linguistics. He operated from a structuralist foundation, in which language was a structural system of formal objects characterized by relations between them. Where Harris distinguished himself from earlier structuralists was in his insistence that natural language corpora were the only means by which a linguist could access these relationships (Brunila & LaViolette, 2022). At all levels of linguistic analysis, observing 'environments' of use allowed the linguist to group elements into structurally equivalent classes that share the same distributional rules. Meaning therefore, is not external to language as some kind of correlate to linguistic form, but is rather constituted by linguistic form (Harris, 1954). If contextual information is statistical, we can understand Harris' conception of meaning in light of Shannon's information theory. Redundancy is collocation, and the resulting information is semantic.

Coming from a background in anthropology, Firth rejected Harris' division between meaning and society. He wanted to understand meaning as a multimodal system, collocation being only one means by which the broader syntagmatic relations, or organization of form units that yield meanings, could begin to be accessed. In addition to collocation, crucial to meaning-finding was the broader social context when applied to concrete use of language, including the features of the participants involved in discourse, their actions, the place, and the relevant objects (Firth, 1968). Firth's overall system left much to be desired in terms of rigor, but his students would go on to add statistical extensions to enrich his core notion that, as in accordance with Harris, linguistic elements could be organized hierarchically and assigned structural classes based on distribution, with one level of distributional analysis being social context (Brunila & LaViolette, 2022).

Although the distributional semantics of Harris and Firth predate the formalisms of Montague, it was not until after the late 1980's that a resurgence in empiricist methods pushed them to the forefront of computational linguistics, culminating in the introduction of word2vec and embedding-based language models (Boleda & Herbelot, 2016). The performance of these models and their far-reaching applications have cemented distributional semantics as a powerful meaning theory.

4. Comparing the Theories of Meaning

An adequate theory of meaning should, given any natural language expression, return its meaning. In this sense, both formal and distributional approaches have found great success, but they differ in the types of meanings they return. In model-theoretic formal semantics, truth conditions are the semantic currency. Phrases 'mean' the ways that they format the world model, and the entailments that can be reached from the model, and words 'mean' the arguments or functions, which, once composed, bring us to these formattings. In distributional semantics, meaning is use. Words, phrases, and even entire

documents are said to 'mean' the contexts in which they are used.

A crucial metric for comparing these two theories is in how successfully they account for different semantic issues.

Traditional compositionality is formal semantic's most prized feature. The lambda calculus fulfills the compositional requirement that the meaning of a whole be able to be discerned from the meaning of its parts in an accessible manner. Semantic entailment is also a direct side effect of the model-theoretic paradigm (Partee, 1996). The truth conditions of one declarative sentence yield a newly constrained model under which logical further logical statements can be evaluated. 'She has a green Chevy' entails 'She has a green car' because our model encodes Chevy as a subset of car.

Where formal systems fall short is in their encoding of the lexical aspects of words. While 'man', 'guy', 'bloke' and 'chap' all fall into the same model set of male human, no one would argue with their meaning different things, whatever we take meaning to be. In a formal system, describing these shades of meaning becomes intractable, particularly without any empirical basis upon which to mount judgments (Boleda & Herbelot, 2016). It is also worth noting that while formal semantics boasts a logical entailment function, not all linguistic entailment is strictly logical, and much of it instead relies on the very descriptive meaning that formal semantics struggles to capture. If we know something about bakers, we may also know something about chefs. In the discourse context, negation tends to behave most like a graded similarity function (Kruszewski et al., 2016). As an example:

This is not a dog.

evokes a set of alternative predicates that may be being referred to. In this case, similarity to dog seems most appropriate, as other middle-sized mammals are more plausible alternatives than buses or large buildings. Given the two possible continuations:

a) This is not a dog...it is a wolf.
b) This is not a dog...it is a screwdriver

we want to assign a higher probability to (a) than to (b). However, taking on a traditional logical lens of predicate negation as complementizing the corresponding set assigns each continuation equal probability (Kruszewski et al., 2016).

The strengths of distributional representations lie in their gradedness. In a vector representation of a word, similarity can be directly computed as a geometric function, allowing for a strongly encoded notion of semantic similarity that is multidimensional and nuanced. used to strengthen accounts of both polysemy and semantic change. Where formal semantics treats the different instances of a word with multiple meanings as separate entities, a feature representation of a word allows us to model how its multiple meanings might differ, to what extent, and which aspects of its meaning remain the same across instances. Feature representations also allow us to detect and quantify semantic

change over time (Boleda, 2020). Finally, although compositionality is less obviously inherent to a distributional model, summing word vectors is one of several mechanisms that have shown success in composing phrase-level distributions. In the case of adjective composition, McNally & Boleda, 2017 found that distributional models struggle most with adjectives used to denote an atypical property of a noun, such as 'severe' for 'budget,' and propose that composition relies on both conceptual and referential information. Conceptual information about the nature of adjectives, nouns and the regularities in their pairings (i.e. phrases like 'red pen' and 'blind cook' are acceptable, while 'blind pronunciation' is not) has to be coupled with information about the referent of the phrase itself (i.e. information about the cook that could allow us to glean along which dimension they might be blind) to generate faithful representations. The difficulties a distributional model has in accounting for a 'severe budget' can therefore likely be explained by its lack of built-in referential information (severe in what way?) (McNally & Boleda, 2017).

Distributional representations also lack a basis for variable or meta-level representation (Pierrehumbert, 2024; Pinker & Prince, 1988). In general, as a function of their mechanics, distribution-based language models generate more inaccurate predictions for less "bursty" words, words that appear evenly distributed over the course of a text, rather than in concentrated bursts. The burstiness of a word aligns closely with its semantic category, with most bursty words being entities and predicates, and less bursty words being higher order operators that generalize over relations (hence, not, often) (Pierrehumbert, 2024). Without a notion of a variable linguistic unit, it becomes difficult to see how the meanings of higher order operators disambiguate themselves as such, or how their distributional information reflects their ability to exert such influence uniformly on other linguistic units. Similarly, distributional models lack a mental lexicon that generalizes over word tokens, creating challenges in dealing with self-referential text and inducing morphological patterns, which, while not necessarily a semantic competence, is critical to a person's linguistic competence (Pierrehumbert, 2024; Pinker & Prince, 1988).

A second consideration for a meaning theory might be its cognitive plausibility. In distributional semantics, representations are induced from empirical language use. This process of contextual prediction mediated by surprisal is borne out in studies of human cognition (Boleda & Herbelot, 2016; Pierrehumbert, 2024; Shannon, 1948; Tomasello, 2015). Still missing from the cognitive picture, however, is referential information and by extension, communicative intent. Usage-based language acquisition theories such as those posited by Tomasello emphasize the importance of pragmatic information in acquisition. Children's minds do distributional work, but it is functionally based, with functions being induced from intention-reading while observing adult speech (Tomasello, 2015). This core notion that words exert influence on the state of the world may be better captured by a model-theoretic system. In addition to providing no mechanism for language acquisition, the overall cognitive plausibility of a fully-fledged formal semantic system is hindered by the undecidability of infinite possible world models. It seems unlikely that humans

not only hold fine-grained models of the world in their minds, but that these models all converge perfectly to ensure successful communication (Boleda & Herbelot, 2016). With these considerations, a distributional approach enhanced with some form of embodiment seems to edge out formal mental representations.

5. Towards a Cohesive Semantic Story

The two meaning theories discussed operate in largely separate spheres, although some attempts have been made to unify them in the hopes of combining the strengths of both approaches (Boleda & Herbelot, 2016; Kruszewski et al., 2016; McNally & Boleda, 2017). Distributional semantics are able to leverage the advanced computing power we have today to build large embeddings, which are fed to even larger models, which are put to work on language tasks. Pure formal semantics projects of this scale have proven largely intractable, as each irregularity and idiosyncrasy of language has to be accounted for separately under one unified logical formalism (Boleda & Herbelot, 2016).

The rise of highly complex machine learning models whose core mechanisms are distributional has generated some tensions between linguists who view language as a phenomenon to be explained and engineers who view it as a problem to be solved. On the one hand, over the course of this paper we have seen that the distributional theory from which models are built is firmly rooted in linguistic tradition. On the other, it seems that this theory is frequently invoked to lend authenticity to a field that is increasingly concerned with tweaking model structure to achieve state-of-the-art numbers, with little regard to understanding how these performance metrics can inform the foundational theories that got them there (Brunila & LaViolette, 2022). The persistent view is that by using massive amounts of distributional data over highly complex model architectures, engineers are not engaging in true linguistic work, but rather are substituting complex descriptions of form and converging on the same results as if a true model of meaning had been input (Bender & Koller, 2020).

This view only holds if you do not believe that meaning *is itself* constituted by form. If we take Harris and Firth's original view at face value, then the model is dealing directly with salient aspects of meaning. Firth would argue that extralinguistic social context, something most models do not directly compute, is also required to get to the completeness of meaning (Firth, 1968), a view shared by many researchers in the NLP community (Hovy & Yang, 2021; Nguyen et al., 2021). Authors like McNally & Boleda and Tomasello similarly raise the importance of referential information for getting at the exact mechanisms of composition and meaning-building (McNally & Boleda, 2017; Tomasello, 2015). Nonetheless, there is a strong argument to be made that the underlying structural foundations of models would meet the requirement of encoding a mode of meaning, if not all modes meaning. We must therefore distinguish a distributional hypothesis, which states that if we represent words based on their collocations then words with similar meanings,

meaning being some other, unknown thing, will have similar representations, from the original distributional meaning theory that we set out with, as these make different claims about how 'linguistic' language models truly are.

Overall, our two semantic stories align with human intuitions about how words work and what they mean. When asked to describe an entity, like a van, you are likely to describe it in terms of what it *is like*: "It's between a car and a truck." When asked to describe a higher order, non-bursty operator like 'not,' you are more likely to describe it in terms of what it *does*: "You put it in front of a word, and it gives you the opposite (or, more precisely, a graded opposite)." Describing either using the other descriptive mechanism feels unwieldy and unintuitive. It can be baffling how distributional representations in particular are able to achieve such success despite this friction.

This tension between 'intuitive,' rule-based theories that are aligned with our own introspection and theories that work well in practice at the expense of interpretability extends beyond meaning-modeling (Bender & Koller, 2020; Pinker & Prince, 1988). For both formal and distributional systems, we observe a performance ceiling that is primarily computational, rather than theoretical. It is difficult, therefore, to distinguish what makes a model 'good' from what makes it adequate, what makes it intuitive, or what makes it plausible. Both formal and distributional semantics are firmly grounded in a linguistic, philosophical and computational reality, neither more formal than the other, and one need not choose which to 'believe in,' or which deserves greater legitimacy.

References

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. https://www.nytimes.com/2018/11/18/technology/artific

Boleda, G. (2020). Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*. https://doi.org/10.1146/annurev-linguistics-011619

Boleda, G., & Herbelot, A. (2016). Formal Distributional Semantics: Introduction to the Special Issue. *Computational Linguistics*, 42(4), 619–635. https://doi.org/10.1162/COLI_a_00261

Brunila, M., & LaViolette, J. (2022). What company do words keep? Revisiting the distributional semantics of J.R. Firth & Zellig Harris. http://arxiv.org/abs/2205.07750

Firth, J. (1968). Linguistic analysis as a study of meaning. In F. R. Palmer (Ed.), *Selected papers of J.R. Firth (1952-59)* (pp. 12–27). Longman and Indiana University Press.

Frege, G. (1892). On sense and reference. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25–50.

Gamut, L. (1991). Logic, Language, and Meaning. In F. R. Palmer (Ed.). University of Chicago Press.

Harris, Z. S. (1954). Distributional structure. *WORD*, 10, 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hovy, D., & Yang, D. (2021, June). The importance of modeling social factors of language: Theory and practice. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty & Y. Zhou (Eds.), *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 588–602). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.49

Janssen, T. M. V., & Zimmermann, T. E. (2021). Montague Semantics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University.

Kruszewski, G., Paperno, D., Bernardi, R., & Baroni, M. (2016). There Is No Logical Negation Here, But There Are Alternatives: Modeling Conversational Negation with Distributional Semantics. *Computational Linguistics*, 42(4), 637–660. https://doi.org/10.1162/COLI_a_00262

McNally, L., & Boleda, G. (2017). Conceptual versus referential affordance in concept composition. In J. A. Hampton & Y. Winter (Eds.), *Compositionality and concepts in linguistics and psychology* (pp. 245–267). Springer International Publishing. https://doi.org/10.1007/978-3-319-45977-6_10

Montague, R. (1970). Universal grammar. *Theoria*, 36(3), 373–398. https://doi.org/10.1111/j.1755-2567.1970.tb00434.x

Nguyen, D., Rosseel, L., & Grieve, J. (2021, June). On learning and representing social meaning in NLP: A sociolinguistic perspective. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty & Y. Zhou (Eds.), *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 603–612). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.50

Partee, B. H. (1996). The development of formal semantics in linguistic theory. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 11–38). Blackwell Reference.

Pierrehumbert, J. (2024). On the distributional hypothesis - why it works and how it doesn't [Presented at ESSLLI 2024].

Pinker, S., & Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*. https://doi.org/10.1016/0010-0277(88)90032-7

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 623–656.

Speaks, J. (2024). Theories of Meaning. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2024). Metaphysics Research Lab, Stanford University.

Tomasello, M. (2015). The usage-based theory of language acquisition. In E. L. Bavin & L. R. Naigles (Eds.), *The cambridge handbook of child language* (pp. 89–106). Cambridge University Press.