Aristotle in Space: Syllogistic Entailment for Distributional Models

Caleb Kisby

1 Introduction

In traditional formal semantics, the meaning of a word or sentence is its denotation given by a formal model. Traditional semantics is particularly good at modelling the *compositionality* of language (the meaning of a sentence is composed of the meaning of its parts), as well as *entailment* between sentences.

Despite its explanatory success, this approach requires a slow by-hand analysis of each word in the lexicon. And, as Baroni et al. argue in (Baroni et al., 2014), many problems that arise in these analyses are due to *content words* in the lexicon rather than grammatical terms. Because of this, semanticists have more recently turned to semantics in which the meaning of a word can be derived from its linguistic context. This *distributional semantics* offers a data-driven way to learn words (especially content words) from a corpus of natural language sentences.

In its purest form, distributional semantics fails to account for compositionality; the linguistic context of a sentence is not a function of the contexts of its parts. In addition, it is unclear how sentence entailment can be determined at all by the sentences' contexts. In response to the former issue, there has been a plethora of proposals for *compositional distributional semantics* (Coecke et al., 2010), (Grefenstette and Sadrzadeh, 2011), (Socher et al., 2012), (Balkır, 2014), (Paperno et al., 2014), (Polajnar et al., 2015), (Clark et al., 2016). However, none of these proposals suggest a clear account of entailment (Emerson, 2020). This so-called Entailment Problem — how we can determine whether a sentence s follows from a set of sentences Γ using primarily distributional means — remains a difficult open problem.

Our contribution. In this paper, we provide a distributional framework for entailment in the restricted case of *syllogistic* entailment. That is, we only handle the basic relational quantifiers $\mathsf{All}, \mathsf{Some}, \mathsf{No}$ used in Aristotle's account of syllogistic reasoning. By focusing our attention on such a simple system, we of course reduce the magnitude of the problem. But we have chosen syllogistic language in particular because it is a perfect playground for set-theoretic notions of membership, containment, and intersection which form the basis for Montague-style formal semantics. Our semantics will heavily rely on distributional reimaginings of these set-theoretic ideas, with the upshot that our choice of entailment has a near-classical presentation.

Related work on the Entailment Problem is recent and sparse. An early paper (Baroni et al., 2012) demonstrates, within a distributional semantics, entailment between quantifier phrases (e.g. "many dogs $\models$ some dogs"), but does not handle *sentences* with quantifiers. (Sadrzadeh et al., 2018) and (Balkir et al., 2015) show that Coecke et al.'s well-known tensorial compositional semantics facilitates a notion of entailment between two sentences, but the sentences they consider are logically simple (they leave open connectives and quantifiers). Comparatively, our entailment framework handles both quantifiers and full sentences, albeit at the cost of a toy syntax.

Our paper's relation to a more recent proposal (Bernardy et al., 2018), (Bernardy et al., 2019) is more intimate. The scope of their framework is ambitious; it aims to cover classically valid and probabalistic inference, including inference involving connectives, generalized quantifiers, generics, etc. We directly make use of some of their key insights, especially with regards to re-imaginings of set-theoretic principles (see Section 3). But, in part because of its ambition, their framework doesn't ruminate on these choices for very long. In this

sense, our paper can be seen as an elaboration of this subset of their insights; by isolating these ideas in a toy context, we aim to dwell on these choices and the assumptions needed to support them. In addition, we do not match the probabilistic flavor of their semantics, opting instead to closely mirror classical syllogistic semantics (as we will see).

Organization. This paper is organized as follows. In Section 2, we introduce the basics of syllogistic and distributional semantics. We will then propose our entailment framework in Section 3. In Section 4, we present our implementation of this system along with a link to its code.

Note to Larry: Now, although I've thoroughly debugged this code, I have only tested it on small examples typed by hand. This might be sufficient for this class, although it obviously isn't enough for a conference. I have given some thought to what tests need to be done, but I've decided to code and run them (along with fine-tuning anything written here which does not quite work) *after* this class deadline. Section 4 also includes a description of these planned tests — although they are in present tense, understand that I am not claiming to have already done them yet. For now, I will also include some discussion of the small test I *have* done in Section 5.

Finally, we summarize what we have done and offer suggestions for future work in Section 6.

2 Background

2.1 Modern syllogistic semantics

We now introduce what we refer to as the *classical* semantics and entailment for syllogisms. Note that we are not using Aristotle's original account of meaning and inference. Instead, we make use of a modern resurrection of syllogistic logic whose semantics are calibrated towards model-theoretic sensibilities. In particular, we use the system first given by (Corcoran, 1972), although our presentation is more directly inspired by (Moss, 2008).

In this setup, nouns are denoted as terms $w, x, y, z, \ldots$. Sentences are of the form $\mathsf{All}\ x\ y$, $\mathsf{Some}\ x\ y$, and $\mathsf{No}\ x\ y$ (read "all x are y," "some x are y," and "no x are y"). (We will often write $\mathsf{Q}\ x\ y$ to abstractly denote one of these three relational quantifiers.) Note that we exclude the sentence form "some x are *not* y," used by Aristotle (Corcoran, 1972), for notational convenience. In fact, our syntax excludes much of what a semanticist might want in syllogisms, but these components (e.g. verbs and other relational quantifiers) can be integrated naturally alongside this bare syntax (Moss, 2010), (Moss and Raty, 2018).

Now for the semantics. Like Montague-style semantics, all expressions are interpreted in a *model*. Models $\mathcal{M}$ in this case consist of a universe M, along with an interpretation function $[\![\ ]\!]$. The interpretation $[\![x]\!]$ of a noun x in a model M is a set of entities from the universe. For example, $[\![dog]\!]$ is the set of all dogs in M. Our sentences are interpreted via a satisfaction relation $\models$ as follows:

$$\begin{array}{lll} \mathcal{M} \models \mathsf{All}\ x\ y & \text{iff} & [\![x]\!] \subseteq [\![y]\!] \\ \mathcal{M} \models \mathsf{Some}\ x\ y & \text{iff} & [\![x]\!] \cap [\![y]\!] \neq \emptyset \\ \mathcal{M} \models \mathsf{No}\ x\ y & \text{iff} & [\![x]\!] \cap [\![y]\!] = \emptyset \end{array}$$

Let Γ be a set of sentences, in this same basic syntax. Entailment between sentences is exactly the sort of entailment semanticists are used to: $\Gamma \models \mathsf{Q}\ x\ y$ if and only if for every model $\mathcal{M}$, if $\mathcal{M} \models s$ for every sentence $s \in \Gamma$, then $\mathcal{M} \models \mathsf{Q}\ x\ y$.

2.2 Distributional semantics

Distributional semantics refers to a class of proposals for a semantics which takes as its primary principle the *distributional hypothesis:* that words in similar contexts have similar meaning. The literature on distributional semantics is *vast*, and it's easy to get lost – an excellent roadmap is (Baroni et al., 2014) by Baroni et al., which we closely follow.

In the purest form of distributional semantics, the meaning of a word *is* its occurrence in linguistic contexts. Typically, a word w is represented by a vector whose components are each a function of w's co-occurrences with another word, within a given window. These vectors are easily learned from a corpus, typically by counting frequencies of co-occurring words. It is common practice to then reduce the dimension of these vectors in order to filter out less robust aspects of meaning.

As for the meaning of phrases above the word, we have two options. The first, most common representation of a phrase is as a *vector mixture*, i.e. an operation on two vectors which produces a vector. For example, we could take the meaning of "sold house" to be the sum of the vectors for "sum" and "house," resulting in a vector which has embedded in it aspects of both word meanings. As Baroni et al. argue in (Baroni et al., 2014), this approach is entirely inappropriate for phrases involving quantifiers (and any other grammatical word). They point

out that the meaning of, e.g., "all dogs" should not simply be the mixture of "dog"-like things and "all"-like things.

Since quantifiers are at the heart of our language fragment, we take the alternative option, suggested by (Baroni et al., 2014): We represent content words (in our case, common nouns) as vector embeddings, and treat quantifiers All, Some, and No as *operations* on these vectors. On the surface, it may seem that we are cheating logical structure into an otherwise distributional system. But we are not too concerned about this, since 1) content words are the main phenomena we want our distributional semantics to handle, and 2) it is not clear, given the above account, that we can avoid assigning logical structure to quantifiers.

3 Our framework

3.1 Overview

The key feature of classical semantics is its emphasis on formal (set-theoretic) models. As we've seen, once phrases are given interpretations within a model, a natural notion of entailment falls out. We would like to preserve as much of this classical intuition about models as we can, so that we might inherit the classical entailment in our distributional setting.

Analagous to our definition of a classical model as a universe along with an interpretation $[\![\ ]\!]$ from nouns to sets, we take a *distributional model* to be a *vector space*[1] along with an interpretation $[\![\ ]\!]$ from nouns to *vectors*. (We will slightly modify this choice in a moment). Of course, we acknowledge that the choice of vector $[\![x]\!]$ for a word x has been extracted from some corpus D of natural language sentences. In fact, we *identify* corpora as choices of vector embeddings, and we will simply write D to denote both the corpus and its corresponding word embeddings. Furthermore, not just any corpus (or distributional model) D will do for our semantics; we will impose further constraints on D once it is clear what assumptions we need to make.

In the next section we will present a re-imagining of the interpretation of terms from the distributional model point of view. As we mentioned before, much of this section is based on key insights from (Bernardy et al., 2019). We will then analyze some of the assumptions that justify this re-imagined semantics, which will constrain the kinds of domains that work with our approach. Finally, we will show that our system inherits the classical semantics and entailment from syllogistic logic (albeit with distributional underpinnings).

[1]Following suit with much of the distributional semantics literature, we use the term "vector space" loosely in this paper; we simply mean a set of vectors along with addition, dot product, and scalar multiplication. We do not require that the set is closed under these operations.

3.2 Entities and membership

At the heart of syllogistic semantics is the fundamental distinction between *entities* and *sets* of entities. Common nouns, at the forefront, are given interpretation as sets, and this choice facilitates the interpretation of All, Some, and No as operations on sets. Entities, instead, are not interpreted at all, but play an important role in evaluating these set operations.

Distributional semantics makes no such distinction between entities and other noun phrases. As we have presented it, all content words are assigned the same 'vector' type for their interpretation. A basic starting point in our reworking of classical semantics is to introduce this distinction. One very natural way of doing this, as suggested by (Bernardy et al., 2019), is to *lift* the type of common nouns from vectors to vector spaces. Entities remain interpreted as individual vectors, so that common nouns contain entities. Note that our interpretation function must now either map nouns to vectors or to subspaces, depending on whether the noun is an entity.

Given a common noun x, we now need to determine that space $[\![x]\!]$ so that if our corpus D contains information that implies an entity e is a member of x, then $e \in [\![x]\!]$. Bernardy et al. suggest using a measure of similarity between e and x: For each such noun x, assign a threshold θ_x, and then let

$$[\![x]\!] = \{e \mid e \cdot x > \theta_x\}$$

Although this gives us the correct type for $[\![x]\!]$, it is not immediately clear why similarity should give rise to membership in this way. In the next section, we consider an assumption on our dataset D that allows us to justify this move.

3.3 Restricting our models

Our semantics for terms makes use of the similarity between an entity e and a common noun x in order to determine e's membership in x. But high similarity between e and x only implies a high frequency of their co-occurrence in D. Of course, many co-occurrences are precisely due to membership, especially when one of the words is an entity.

For example, we might expect "Joe Biden" and "president" to frequently occur, since text about the entity Joe Biden often discusses his membership in the "president" category.

Unfortunately, there are many other relationships between e and x that might also be indicated by a co-occurrence: e knows of x, e sees x, e read about x, etc. It would be a mistake for our system to conclude, from the high frequency of Nixon's famous statement "I am not a crook" in a 1974 news corpus, that former president Nixon was a crook. We cannot even conclude that the *corpus* implies this since, despite the actual facts about the Watergate scandal, such a corpus implies the negation.

Are there corpora for which our method of membership works? Clearly, such a corpus D needs to satisfy the following constraint: For each word x, there is a threshold θ_x such that if the siilarity between e and x exceeds that threshold, then it holds in D that e is a member of x. In other words, we need many (perhaps most) co-occurrences of e and x to be due to membership, so that we may filter out other potential relationships with our threshold. An easy choice is any text whose express intent is to provide membership relations – good choices might include encyclopedic and (perhaps) news sources.

Going forward, when we write the universal *for all* D, we mean only to consider all D satisfying the above condition. Without a constraint like this, we see little hope in extracting logical knowledge from the distribution of words in text.

3.4 Semantics and Entailment

With our models and interpretations in hand, we can at last present our semantics and entailment. Our semantics for syllogisms is exactly the classical semantics, but with $\subseteq, =, \cap, \emptyset$ operating now on subspaces containing vectors rather than mere sets containing elements.

$$\begin{array}{lll} D \models \mathsf{All}\ x\ y & \text{iff} & [\![x]\!] \subseteq [\![y]\!] \\ D \models \mathsf{Some}\ x\ y & \text{iff} & [\![x]\!] \cap [\![y]\!] \neq \emptyset \\ D \models \mathsf{No}\ x\ y & \text{iff} & [\![x]\!] \cap [\![y]\!] = \emptyset \end{array}$$

Similarly for entailment: $\Gamma \models \mathsf{Q}\ x\ y$ if and only if for every D, if $D \models s$ for each $s \in \Gamma$, then $D \models \mathsf{Q}\ x\ y$. Notice that there is a subtle difference between this entailment and the classical one, namely that we are now quantifying over those distributional models that satisfy the constraint in the previous section.

In practice, the universal quantification in the definition of entailment causes some trouble. The issue is that it is often difficult to argue that *all* corpora D satisfy the sentences in Γ. To work around this for practitioners, who are typically dealing with a single dataset, we offer the following sleight of hand: Pick a *particular* corpus D, but impose the assumption that D is an *arbitrary* representative corpus. We then check that $D \models \Gamma$ implies $D \models \mathsf{Q}\ x\ y$. For this to be theoretically equivalent to our definition, D must have no characteristics at all which distinguish it from any other relevant corpus. But for practical use, we suggest taking D to be very large and typical for the chosen domain.

4 Implementation

4.1 Overview of system design

In order to test our choices above, we have implemented these semantics and test its ability to perform inference. We have chosen to test a news domain, as well as an encyclopedia domain, in line with our constraints on domains. Our code is available on Github[2].

Here is a high-level overview. First, we use a small named entity recognition dataset in order to distinguish between entities and category nouns. We use the Wikigold dataset for our encyclopedia domain (Balasuriya et al., 2009), and we use the entity-tagging portion of the GMB dataset for our news domain (Bos et al., 2017). Both of these datasets are small, as is typical for named entity corpora. But in order for our word embeddings to be meaningful, they need to be learned via a large corpus. Recent developments in entity tagging have resulted in methods for generating large datasets (Menezes et al., 2019), but we do not make use of this during testing. Instead, we merely borrow the word embeddings from a larger corpus in the same general domain. We use two pretrained vector models available via Python's `gensim` package (Řehůřek and Sojka, 2010): GloVe for our encyclopedia domain (Pennington et al., 2014), and Google News for our news domain. Observe that our dataset D's vocabulary is the *intersection* of the tagging dataset and the large dataset's respective vocabularies.

Next, we calculate thresholds θ_x for each word x. This is a bit of an art; we need θ_x to be high enough that few non-x entities are members of x, yet low enough that we do not exclude *actual* members of

[2]https://github.iu.edu/cckisby/distributional-syllogistic

x. Unfortunately, noise will occur regardless of the choice of θ_x. And so we stuck with an arbitrary strict cutoff of $\theta_x = \max\{e \cdot x \mid e \text{ is an entity}\} - 2$.

We can evaluate the semantics of All, Some, and No straightforwardly from here. We consider two ways we can test this system. First, we can directly test its ability to perform various entailments $\Gamma \models \mathsf{Q}\ x\ y$. Alternatively, we can test more indirectly whether our system's evaluations $D \models \mathsf{Q}\ x\ y$ result in judgements that intuitively hold. We do both, and describe each test in more detail in the following sections.

Disclaimer: The following sections mention big, systematic tests. I have not implemented these yet, since I came up with the specific tests within the past week. But I decided to write this design section as if I do currently run these tests. So even though these are written in present-tense, it helps to read them as future TODO items.

4.2 Testing entailment

To test entailment, typically semanticists would use a standard benchmark dataset (such as SICK or FraCaS). In addition, (Bernardy et al., 2019) provides another test suite for inferences, some of which are syllogistic in nature. Unfortunately, our syntax is not expressive enough even to make partial use of these datasets, since almost no test cases consist strictly of the sentence forms for All, Some, and No.

So we instead generate a custom suite of test inferences using strictly syllogistic language. We include the code for doing so along with our implementation. First, we enumerate all possible sentences $\mathsf{Q}\ x\ y$, for all x, y in our corpus. We then filter these into two categories, based on whether the sentence holds in D. We then select various sets Γ from these sentences, making sure to obtain some Γ for which $D \models \Gamma$, and others for which $D \not\models \Gamma$. For each Γ, we pick several φ for which (classically) $\Gamma \models \varphi$ and for which $\Gamma \not\models \varphi$. We can determine this using a polynomial-time algorithm for the provability relation $\Gamma \vdash \varphi$ (I couldn't find the citation for this – can you help?). Finally, we make some effort to exercise all of the inference rules of syllogistic reasoning in (Moss, 2008) by including some hand-coded inferences, based on one-step application of these rules.

4.3 Testing our model

Another approach toward testing our semantics is to check that the judgements $D \models \mathsf{Q}\ x\ y$ intuitively hold. More concretely, if we expect $\mathsf{Q}\ x\ y$ to hold in all corpora D (satisfying our earlier constraints), then $\mathsf{Q}\ x\ y$ should hold in D. In other words, we are checking whether D was picked "arbitrarily" enough to justify our reasoning at the end of Section 3. Additionally, checking that these $\mathsf{Q}\ x\ y$ hold goes a long way to convince us that the system isn't producing garbage interpretations.

Note to Larry: What we need here is a hand-coded set of syllogisms that hold in ordinary English, e.g. a dataset that looks like:

All boys are children
All girls are children
Some women are professors
No professor is unemployed
...

I haven't looked into this in detail yet; if you have any leads, I would love to hear them. My current thought is to comb the psychology literature on human performance at syllogisms. Since this area has tended (in the past) to ignore inference, I'm hoping that somebody has recorded these sorts of intuitive "off the cuff" judgements. If not, then I need to spend some time generating my own (semi-computer-assisted).

5 Results and discussion

Note to Larry: As I mentioned, I have not yet actually performed these tests – although I can do them well before the PaM deadline, if this paper still seems promising at this point. I have done a (very) small, unsystematic version of the second test, though, just to see whether this system's output makes some sense. Consider the following words taken from the Google News dataset:

president	*politician*	*boy*
girl	*child*	*fish*
bird	*animal*	*species*
man	*woman*	*adult*
country	*group*	*organization*
professor	*communist*	*working*

I typed by hand all 324 possible choices for All $x\ y$, and then used my intuition to type the correct judgement. For these words, intuition is particularly easy.

The current "as-is" system scored 86% on negative examples, and 40% on positive examples (note that I have filtered out the 18 instances of All $x\ x$,

since our system easily gets these). 40% on positive cases is disappointingly low.

You asked me before if I still believe in the assumptions we make in this paper. But I don't think I've given these ideas a fair shot yet; there are several issues with the current implementation that I would like to fix, and that I think are throwing off the score.

First, I'm not quite doing entity-tagging correctly. The tagging dataset only tags individual words, e.g. in "Hurricane Katrina," "Hurricane" and "Katarina" are each tagged as entities. My naive fix for this is to group runs of entities into a single entity, e.g. "Hurricane_Katrina." This works for simple kinds of entity phrases, but titles like "In the Mood for Love" fall through the cracks. Another issue is that some nouns, like "tornado," are currently mis-tagged in GMB as entities.

Second, the thresholds θ_x are currently being selected based on my inutition (I just check whether members of nouns are what we would expect). But in order to optimize the accuracy of the system, it would be very helpful to pick the thresholds more carefully. I'm thinking to use gradient descent to find those thresholds θ_x which result in the highest accuracy.

Finally, I have not yet tested this system on the Wikipedia domain. While writing together, Matthew suggested the idea that my domain might be the problem (e.g. nobody would mention the word "woman" near "Taylor Swift" in the news corpus due to it being a background assumption). So this idea to try the system on a corpus whose *purpose* is to define memberships is pretty recent. Before throwing in the towel, I ought to give the Wikipedia domain a chance – it's the best chance I have at extracting any membership information from word co-occurrences.

6 Conclusions and Future Work

We have presented a basic framework for determining simple (syllogistic) entailments within a distributional model. Our deliberately restrictive syntax has allowed us to elaborate on ideas from other proposals for entailment – namely, on the conditions our domain must have in order to "lift" a noun vector to a vector space. This re-imagining of set-theoretic semantics in distributional terms closely mirrors the classical account, which allows us to inherit the latter's definition of entailment.

Although we only specified our semantics for a small fragment of syllogistic logic, the ideas behind our engine are applicable to any semantic system predicated on membership and set operations. An obvious next step would be to extend these semantics towards more complicated syllogistic systems, incorporating verbs and other quantifiers like "most." But syllogisms are just practice for the larger ambition: We imagine that a re-working of Montague semantics itself should invoke the notion of membership outlined here.

As we discussed earlier, our approach is only appropriate for domains, such as encyclopedias, where membership is the most common relation expressed between entities e and nouns x. But we would like to extend this system towards other domains. One (naive) way to do this might be to leverage verbs and other keywords that co-occur with e and x in order to filter out other relationships. But we leave this for future work.

Additionally, our semantics and entailment resemble those of classical semantics. But without the constraints on our domain, we can view the underlying interpretations as probabilistic. That is, there is a probability p that a co-occurrence between e and x indicates membership. It would be nice to have a detailed account of how we might re-define syllogisms and entailment with p in mind. But this is beyond the scope of this paper.

References

Dominic Balasuriya, Nicky Ringland, Joel Nothman, Tara Murphy, and James R Curran. 2009. Named entity recognition in wikipedia. In *Proceedings of the 2009 Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources (People's Web)*, pages 10–18.

Esma Balkır. 2014. Using density matrices in a compositional distributional model of meaning. *Master's thesis, University of Oxford*.

Esma Balkir, Mehrnoosh Sadrzadeh, and Bob Coecke. 2015. Distributional sentence entailment using density matrices. In *International Conference on Topics in Theoretical Computer Science*, pages 1–22. Springer.

Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do, and Chung-chieh Shan. 2012. Entailment above the word level in distributional semantics. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 23–32.

Marco Baroni, Raffaella Bernardi, Roberto Zamparelli, et al. 2014. Frege in space: A program for composi-

tional distributional semantics. *Linguistic Issues in language technology*, 9(6):5–110.

Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, and Shalom Lappin. 2018. A compositional bayesian semantics for natural language. In *Proceedings of the First International Workshop on Language Cognition and Computational Models*, pages 1–10.

Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, Shalom Lappin, and Aleksandre Maskharashvili. 2019. Bayesian inference semantics: A modelling system and a test suite. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (* SEM 2019)*, pages 263–272.

Johan Bos, Valerio Basile, Kilian Evang, Noortje J Venhuizen, and Johannes Bjerva. 2017. The groningen meaning bank. In *Handbook of linguistic annotation*, pages 463–496. Springer.

Stephen Clark, Laura Rimell, Tamara Polajnar, and Jean Maillard. 2016. The categorial framework for compositional distributional semantics. *University of Cambridge Computer Laboratory*.

Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394*.

John Corcoran. 1972. Completeness of an ancient logic. *The journal of symbolic logic*, 37(4):696–702.

Guy Emerson. 2020. What are the goals of distributional semantics? *arXiv preprint arXiv:2005.02982*.

Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011. Experimental support for a categorical compositional distributional model of meaning. *arXiv preprint arXiv:1106.4058*.

Daniel Menezes, Ruy Milidiú, and Pedro Savarese. 2019. Building a massive corpus for named entity recognition using free open data sources. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 6–11. IEEE.

Lawrence S Moss. 2008. Completeness theorems for syllogistic fragments. *Logics for linguistic structures*, 29:143–173.

Lawrence S Moss. 2010. Syllogistic logics with verbs. *Journal of Logic and Computation*, 20(4):947–967.

Lawrence S Moss and Charlotte Raty. 2018. Reasoning about the sizes of sets: Progress, problems, and prospects. In *Bridging@ IJCAI/ECAI*, pages 33–39.

Denis Paperno, Marco Baroni, et al. 2014. A practical and linguistically-motivated approach to compositional distributional semantics. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 90–99.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Tamara Polajnar, Laura Rimell, and Stephen Clark. 2015. An exploration of discourse-based sentence spaces for compositional distributional semantics. In *Proceedings of the First Workshop on Linking Computational Models of Lexical, Sentential and Discourse-level Semantics*, pages 1–11.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA.

Mehrnoosh Sadrzadeh, Dimitri Kartsaklis, and Esma Balkır. 2018. Sentence entailment in compositional distributional semantics. *Annals of Mathematics and Artificial Intelligence*, 82(4):189–218.

Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pages 1201–1211.