

A low-level treatment of generalised quantifiers in categorical compositional distributional semantics

Extended Abstract

Ondřej Rypáček^{1*} and Mehrnoosh Sadrzadeh^{2†}

¹ University of Oxford, Oxford, UK
ondrej@cs.ox.ac.uk

² Queen Mary University, London, UK
m.sadrzadeh@qmul.ac.uk

Abstract

We show how one can formalise quantifiers in the categorical compositional distributional model of meaning. Our model is based on the generalised quantifier theory of Barwise and Cooper. We develop an abstract compact closed semantics using Frobenius Algebras for these quantifiers and instantiate the abstract model in vector spaces and in relations. The former is an example for the distributional corpus-based models of language and the latter for the truth-theoretic ones. We provide explanations and toy examples and defer formal proofs and large scale empirical validation to the full version of the paper.

1 Introduction

Distributional models of natural language are based on Firth’s hypothesis that meanings of words can be deduced from the contexts in which they often occur [6]. One then fixes a context window of n words and computes frequencies of how many times a words has occurred in this window with other words. These models have been applied to various language processing tasks, for instance thesauri construction [5]. *Compositional* distributional models of meaning extend these models from words to sentences. The categorical such models [4, 1] do so by taking into account the grammatical structure of sentences and the vectors of the words in there. These models have proven successful in practical natural language tasks such as disambiguation, term/definition classification and phrase similarity, for example see [7, 8]. Nevertheless, it has been an open problem how to deal with meanings of logical words such as quantifiers and conjunctives. In this paper, we present preliminary work which aims to show how quantifiers can be deal with using the generalised quantifier theory of Barwise and Cooper [2].

Very briefly put, according to the ‘living on property’ of generalised quantifier theory, the meaning of a sentence with a natural language quantifier *Det* such as ‘*Det* Sbj Verb’ is determined by first taking the intersection of the denotation of Sbj with the denotation of

* Author supported by EPSRC Grant EP/I037512/1.

† Author supported by an EPSRC CAF grant EP/J002607/1.

subjects of the Verb, then checking if this is an element of the denotation of $Det(\text{Sbj})$. The denotation of Det is specified separately, for example, for $Det = \exists$, it is the set of non-empty subsets of the universe, for $Det = 2$ it is the set of subsets of the universe that have two elements and so on. As a result, and for example, the meaning of a sentence “some men sleep” becomes true if the set of men who sleep is non empty and the meaning of “two men sleep” is true if the set of men who sleep has cardinality 2.

In what follows, we work in the categorical compositional distributional model of [4], treat natural language quantifiers as generalised quantifiers and model them in a way similar to how we modelled relative pronouns in previous work [12, 13]. We first present a brief preliminary account of compact closed categories and Frobenius algebras over them and review how vector spaces and relations provide instances. Then, we develop a compact closed categorical semantic for quantifiers, in terms of diagrams and morphisms of compact closed categories and Frobenius Algebras over certain objects of them. We present two concrete interpretations for this abstract setting: relations and vector spaces. The former provides basis for a truth-theoretic model of meaning and the latter for a corpus-base distributional model of meaning. In the full article version of the current extended abstract, we provide details of interpretation maps, formal proofs, and results of large scale empirical evaluations.

2 Preliminaries

This subsection briefly reviews compact closed categories and Frobenius algebras. For a formal presentation, see [9, 10]. A compact closed category, \mathcal{C} , has objects A, B ; morphisms $f: A \rightarrow B$; a monoidal tensor $A \otimes B$ that has a unit I , that is we have $A \otimes I \cong I \otimes A \cong A$. Furthermore, for each object A there are two objects A^r and A^l and the following morphisms:

$$A \otimes A^r \xrightarrow{\epsilon_A^r} I \xrightarrow{\eta_A^r} A^r \otimes A \quad A^l \otimes A \xrightarrow{\epsilon_A^l} I \xrightarrow{\eta_A^l} A \otimes A^l$$

These morphisms satisfy the following equalities, sometimes referred to as the *yanking* equalities, where 1_A is the identity morphism on object A :

$$\begin{aligned} (1_A \otimes \epsilon_A^l) \circ (\eta_A^l \otimes 1_A) &= 1_A & (\epsilon_A^r \otimes 1_A) \circ (1_A \otimes \eta_A^r) &= 1_A \\ (\epsilon_A^l \otimes 1_A) \circ (1_{A^l} \otimes \eta_A^l) &= 1_{A^l} & (1_{A^r} \otimes \epsilon_A^r) \circ (\eta_A^r \otimes 1_{A^r}) &= 1_{A^r} \end{aligned}$$

These express the fact the A^l and A^r are the left and right adjoints, respectively, of A in the 1-object bicategory whose 1-cells are objects of \mathcal{C} .

A Frobenius algebra in a monoidal category $(\mathcal{C}, \otimes, I)$ is a tuple $(X, \delta, \iota, \mu, \zeta)$ where, for X an object of \mathcal{C} , the triple (X, δ, ι) is an internal comonoid; i.e. the following are coassociative and counital morphisms of \mathcal{C} :

$$\delta: X \rightarrow X \otimes X \quad \iota: X \rightarrow I$$

Moreover (X, μ, ζ) is an internal monoid; i.e. the following are associative and unital morphisms:

$$\mu: X \otimes X \rightarrow X \quad \zeta: I \rightarrow X$$

And finally the δ and μ morphisms satisfy the following *Frobenius condition*:

$$(\mu \otimes 1_X) \circ (1_X \otimes \delta) = \delta \circ \mu = (1_X \otimes \mu) \circ (\delta \otimes 1_X)$$

Informally, the comultiplication δ dispatches the information contained in one object into two objects, and the multiplication μ unifies the information of two objects into one.

Finite Dimensional Vector Spaces. These structures together with linear maps form a compact closed category, which we refer to as FdVect . Finite dimensional vector spaces V, W are objects of this category; linear maps $f: V \rightarrow W$ are its morphisms with composition being the composition of linear maps. The tensor product $V \otimes W$ is the linear algebraic tensor product, whose unit is the scalar field of vector spaces; in our case this is the field of reals \mathbb{R} . Here, there is a natural isomorphism $V \otimes W \cong W \otimes V$. As a result of the symmetry of the tensor, the two adjoints reduce to one and we obtain the isomorphism $V^l \cong V^r \cong V^*$, where V^* is the dual space of V . When the basis vectors of the vector spaces are fixed, it is further the case that we have $V^* \cong V$.

Given a basis $\{r_i\}_i$ for a vector space V , the epsilon maps are given by the inner product extended by linearity; i.e. we have:

$$\epsilon^l = \epsilon^r: V \otimes V \rightarrow \mathbb{R} \quad \text{given by} \quad \sum_{ij} c_{ij} \psi_i \otimes \phi_j \mapsto \sum_{ij} c_{ij} \langle \psi_i | \phi_j \rangle$$

Similarly, eta maps are defined as follows:

$$\eta^l = \eta^r: \mathbb{R} \rightarrow V \otimes V \quad \text{given by} \quad 1 \mapsto \sum_i r_i \otimes r_i$$

Any vector space V with a fixed basis $\{\vec{v}_i\}_i$ has a Frobenius algebra over it, explicitly given as follows, where δ_{ij} is the Kronecker delta.

$$\begin{array}{lll} \delta: V \rightarrow V \otimes V & \text{given by} & \vec{v}_i \mapsto \vec{v}_i \otimes \vec{v}_i \\ \mu: V \otimes V \rightarrow V & \text{given by} & \vec{v}_i \otimes \vec{v}_j \mapsto \delta_{ij} \vec{v}_i \\ \iota: V \rightarrow \mathbb{R} & \text{given by} & \vec{v}_i \mapsto 1 \\ \zeta: \mathbb{R} \rightarrow V & \text{given by} & 1 \mapsto \sum_i \vec{v}_i \end{array}$$

Relations. Another important example of a compact closed category is Rel , the category of sets and relations. Here, \otimes is cartesian product with the singleton set as its unit $I = \{\star\}$, and the adjoint operators are identity morphisms on objects. Closure reduces to the fact that a relation between sets $A \times B$ and C is equivalently a relation between A and $B \times C$. Given a set S with elements $s_i, s_j \in S$, the epsilon and eta maps are given as follows:

$$\begin{array}{lll} \epsilon^l = \epsilon^r: S \times S \rightarrow \{\star\} & \text{given by} & \{((s_i, s_j), \star) \mid s_i, s_j \in S, s_i = s_j\} \\ \eta^l = \eta^r: \{\star\} \rightarrow S \times S & \text{given by} & \{(\star, (s_i, s_j)) \mid s_i, s_j \in S, s_i = s_j\} \end{array}$$

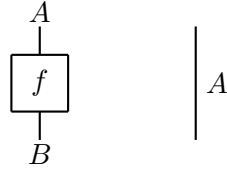
Every object in Rel has a Frobenius algebra over it given by the diagonal and codiagonal relations, as described below:

$$\begin{aligned} \delta & : S \rightarrow S \times S \quad \text{given by} \quad \{(s_i, (s_j, s_k)) \mid s_i, s_j, s_k \in S, s_i = s_j = s_k\} \\ \mu & : S \times S \rightarrow S \quad \text{given by} \quad \{(s_i, s_j), s_k \mid s_i, s_j, s_k \in S, s_i = s_j = s_k\} \\ \iota & : S \rightarrow \{\star\} \quad \text{given by} \quad \{(s_i, \star) \mid s_i \in S\} \\ \zeta & : \{\star\} \rightarrow S \quad \text{given by} \quad \{(\star, s_i) \mid s_i \in S\} \end{aligned}$$

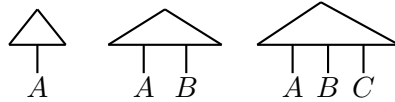
For the details of verifying that for each of the two examples above, the corresponding conditions hold see [3].

2.1 String Diagrams

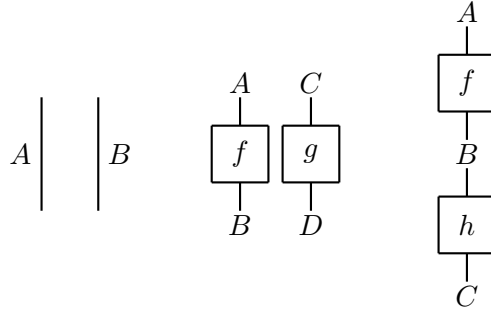
The framework of compact closed categories and Frobenius algebras comes with a complete diagrammatic calculus that visualises derivations, and which also simplifies the categorical and vector space computations. Morphisms are depicted by boxes and objects by lines, representing their identity morphisms. For instance a morphism $f: A \rightarrow B$, and an object A with the identity arrow $1_A: A \rightarrow A$, are depicted as follows:



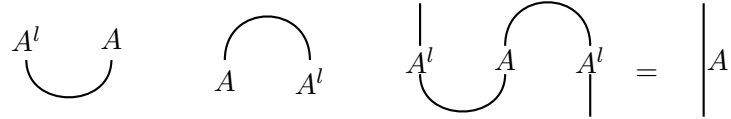
Morphisms from I to objects are depicted by triangles with strings emanating from them. In concrete categories, these morphisms represent elements within the objects. For instance, a vector a in a vector space A is represented by the morphism $a: I \rightarrow A$ and depicted by a triangle with one string. The number of strings of such triangles depict the tensor rank of the vector space; for instance, the diagrams for $a \in A$, $a' \in A \otimes B$, and $a'' \in A \otimes B \otimes C$ are as follows:



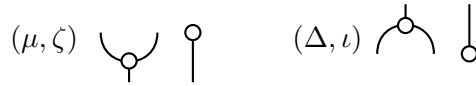
The tensor products of the objects and morphisms are depicted by juxtaposing their diagrams side by side, whereas compositions of morphisms are depicted by putting one on top of the other; for instance the object $A \otimes B$, and the morphisms $f \otimes g$ and $f \circ h$, for $f: A \rightarrow B$, $g: C \rightarrow D$, and $h: B \rightarrow C$, are depicted as follows:



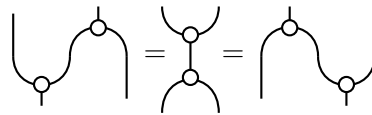
The ϵ maps are depicted by cups, η maps by caps, and yanking by their composition and straightening of the strings. For instance, the diagrams for $\epsilon^l: A^l \otimes A \rightarrow I$, $\eta: I \rightarrow A \otimes A^l$ and $(\epsilon^l \otimes 1_A) \circ (1_A \otimes \eta^l) = 1_A$ are as follows:



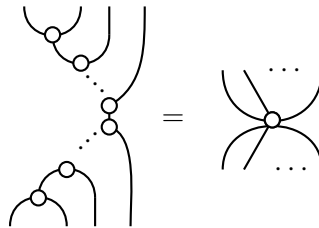
As for Frobenius algebras, the diagrams for the monoid and comonoid morphisms are as follows:



with the Frobenius condition being depicted as:

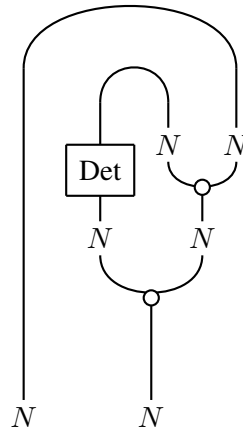


The defining axioms guarantee that any picture depicting a Frobenius computation can be reduced to a normal form that only depends on the number of input and output strings of the nodes, independent of the topology. These normal forms can be simplified to so-called ‘spiders’:



3 Diagrammatic Compact Closed Semantics

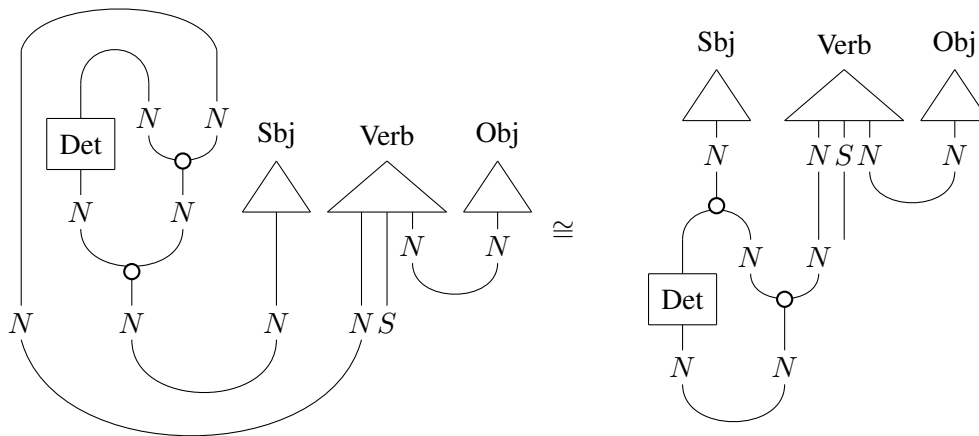
Following the terminology and notation of [2], given a quantified phrase such as “Det noun”, we call “Det” a determiner. A quantified phrase is a noun phrase which is created by the application of a determiner to a noun phrase. We suggest the following diagrammatic semantics for a determiner Det:



It corresponds to the following compact closed categorical morphism:

$$(1_N \otimes \delta_N) \circ (1_N \otimes Det \otimes \mu_N) \circ (1_N \otimes \eta_N \otimes 1_N) \circ \eta_N$$

The meaning of the sentence with a quantified phrase in a subject position and its normalised form are as follows:



Here, “Sbj”, “Obj”, and “Verb” are elements within objects N , N , and $N \otimes S \otimes N$, respectively. The symbolic representation of the normal form diagram is as follows:

$$(\epsilon_N \otimes 1_S) \circ (Det \otimes \mu_N \otimes 1_S) \circ (\delta_N \otimes 1_{N \otimes S} \otimes \epsilon_N)(Sbj \otimes Verb \otimes Obj)$$

The intuitive justification is that the determiner Det first makes a copy of the subject (via the Frobenius δ map), so now we have two copies of the subject. One of these is being unified with the subject argument of the verb (via the Frobenius μ map). In set-theoretic terms this is the intersection of the interpretations of subject and subjects-of-verb. The other copy is being inputted to the determiner map Det and will produce a modified noun based on the meaning of the determiner. The last step is the application of the unification to the output of Det . Set theoretically, this step will decide whether the intersection of the subject-of-verb and the noun belongs to the interpretation of the quantified noun.

The diagrams, morphisms, and intuitions for a quantified phrase in an object position are obtained in a similar way.

4 Truth-Theoretic Interpretation

For this part, we work in the category Rel of sets and relations. We take N to be the set of all subsets of individuals from a universal reference set U , that is $N = \mathcal{P}(U)$. A common noun in N is hence taken to be the set of all subsets of its individuals. We take S to be the one dimensional space; the origin 0 represents false and the vector $\vec{1}$ represents truth. A verb is the set of all non-empty subsets of a relation (corresponding to its predicate). For an intransitive verb, this relation is on the set $N \times S$, where each relation corresponds to a subset of N , since we have $N \times S \cong N$. For a transitive verb, it is a relation on the set $N \times S \times N \cong N \times N$.

We denote the set-theoretical semantics of a word “w” by $\llbracket w \rrbracket$ and its categorial representation as developed above by $\overline{\llbracket w \rrbracket}$. Then we have the following:

$$\begin{aligned} \overline{\llbracket \text{Sbj} \rrbracket} &= \{A_i \mid A_i \in \mathcal{P}(\llbracket \text{Sbj} \rrbracket)\} \\ \overline{\llbracket \text{Verb} \rrbracket} &= \{B_j \mid B_j \in \mathcal{P}(\llbracket \text{Verb} \rrbracket)\} \\ Det(\overline{\llbracket \text{Sbj} \rrbracket}) &= \{D_k \mid D_k \in Det(\llbracket \text{Sbj} \rrbracket)\} \end{aligned}$$

where $\llbracket \text{Sbj} \rrbracket \subseteq U$ and $\llbracket \text{Verb} \rrbracket \subseteq U$ are the set-theoretic meanings of “Sbj” and “Verb”, and $Det(S)$ is the same as in the generalised quantifier approach.

The truth-theoretic meaning of the sentence “Det Sbj Verb” is obtained by computing the morphism developed in section 4 in category Rel . Here, since we have $S = I$, the morphisms that are applied to object S are dropped, that is we have:

$$\epsilon(Det \otimes \mu_N)(\delta_N \otimes 1_N)(\overline{\llbracket \text{Sbj} \rrbracket} \otimes \overline{\llbracket \text{Verb} \rrbracket})$$

This is computed in three steps. In the first step, we obtain:

$$(\delta_N \otimes 1_N)(\overline{\llbracket \text{Sbj} \rrbracket} \otimes \overline{\llbracket \text{Verb} \rrbracket}) = \{(A_i, A_i) \mid A_i \in \mathcal{P}(\llbracket \text{Sbj} \rrbracket)\} \otimes \{B_j \mid B_j \in \mathcal{P}(\llbracket \text{Verb} \rrbracket)\}$$

In the second step, we obtain:

$$\begin{aligned}
& (Det \otimes \mu_N) \left(\{ (A_i, A_i) \mid A_i \in \mathcal{P}(\llbracket \text{Sbj} \rrbracket) \} \otimes \{ B_j \mid B_j \in \mathcal{P}(\llbracket \text{Verb} \rrbracket) \} \right) = \\
& Det \left(\{ A_i \mid A_i \in \mathcal{P}(\llbracket \text{Sbj} \rrbracket) \} \right) \otimes \{ A_i \mid A_i = B_j, A_i \in \mathcal{P}(\llbracket \text{Sbj} \rrbracket), B_j \in \mathcal{P}(\llbracket \text{Verb} \rrbracket) \} = \\
& \{ D_k \mid D_k \in Det(\llbracket \text{Sbj} \rrbracket) \} \otimes \{ A_i \mid A_i = B_j, A_i \in \mathcal{P}(\llbracket \text{Sbj} \rrbracket), B_j \in \mathcal{P}(\llbracket \text{Verb} \rrbracket) \}
\end{aligned}$$

In the final step, we obtain:

$$\begin{aligned}
& \epsilon \left(\{ D_k \mid D_k \in Det(\llbracket \text{Sbj} \rrbracket) \} \otimes \{ A_i \mid A_i = B_j, A_i \in \mathcal{P}(\llbracket \text{Sbj} \rrbracket), B_j \in \mathcal{P}(\llbracket \text{Verb} \rrbracket) \} \right) = \\
& \{ \star \mid D_k = A_i, D_k \in Det(\llbracket \text{Sbj} \rrbracket), A_i = B_j, A_i \in \mathcal{P}(\llbracket \text{Sbj} \rrbracket), B_j \in \mathcal{P}(\llbracket \text{Verb} \rrbracket) \}
\end{aligned}$$

If the result is non-zero, the meaning of the sentence is true, else it is false.

Example. As an example, suppose we have two male individuals m_1, m_2 and a cat individual c_1 . Suppose further that the verb ‘sneeze’ applies to individuals m_1 and c_1 . Hence, we have the following interpretations for the lemmas of words “man”, “cat”, and “sneeze”:

$$\overline{\llbracket \text{men} \rrbracket} = \mathcal{P}(\{m_1, m_2\}) \quad \overline{\llbracket \text{cat} \rrbracket} = \mathcal{P}(\{c_1\}) \quad \overline{\llbracket \text{sneeze} \rrbracket} = \mathcal{P}(\{m_1, c_1\})$$

Consider the following quantified phrases and their interpretations:

$$Some(\overline{\llbracket \text{men} \rrbracket}) = \{\{m_1\}, \{m_2\}, \{m_1, m_2\}\} \quad One(\overline{\llbracket \text{man} \rrbracket}) = \{\{m_1\}, \{m_2\}\}$$

In the first step of computation of the meaning of “some men sneeze”, we obtain:

$$\{(\{\emptyset\}, \{\emptyset\}), (\{m_1\}, \{m_1\}), (\{m_2\}, \{m_2\}), (\{m_1, m_2\}, \{m_1, m_2\})\} \otimes \{\emptyset, \{m_1\}, \{c_1\}, \{m_1, c_1\}\}$$

In the second step, we obtain:

$$\begin{aligned}
Some(\{\{m_1\}, \{m_2\}, \{m_1, m_2\}\}) \otimes \mu(\{\emptyset, \{m_1\}, \{m_2\}, \{m_1, m_2\}\} \otimes \{\emptyset, \{m_1\}, \{c_1\}, \{m_1, c_1\}\}) \\
= \{\{m_1\}, \{m_2\}, \{m_1, m_2\}\} \otimes \{\emptyset, \{m_1\}\}
\end{aligned}$$

In the last step, we obtain the following:

$$\epsilon(\{\{m_1\}, \{m_2\}, \{m_1, m_2\}\} \otimes \{\emptyset, \{m_1\}\}) = \{\star\}$$

Hence, the meaning of the sentence is true. For the sentence “One man sneezes”, one applies $(One \otimes \mu)$ to the result of the first step, which is as above. Hence, the second and third steps of computation are as follows:

$$\begin{aligned} \epsilon\left(\text{One}\left(\{\{m_1\}, \{m_2\}, \{m_1, m_2\}\}\right) \otimes \mu\left(\{\emptyset, \{m_1\}, \{m_2\}, \{m_1, m_2\}\} \otimes \{\emptyset, \{m_1\}, \{c_1\}, \{m_1, c_1\}\}\right)\right) \\ = \epsilon\left(\{\{m_1\}, \{m_2\}\} \otimes \{\emptyset, \{m_1\}\}\right) = \{\star\} \end{aligned}$$

Due to lack of space, we have not done the necessary expositions here, but as established in [3] and used in a similar context for relative pronouns in [12, 13], the Frobenius map μ is the analog of set-theoretic intersection and the compact closed epsilon map is the analog of set-theoretic application. Then it follows that our truth-theoretic interpretation of the compact closed semantics of quantified sentences provides us with the same truth-theoretic meaning as their generalised quantifier semantics. The proof of this claims uses the ‘living on property’ of generalised quantifiers.

5 Concrete Corpus-Based Interpretation

In a concrete vector space model, built from a corpus using distributional methods, we assume that vector meaning of the subject is $\sum_i C_i \vec{n}_i \in N$ and the linear map corresponding to the verb is $\sum_{jk} C_{jk} \vec{n}_j \otimes \vec{s}_k \in N \otimes S$. For \vec{n}_i a basis vector of N , we define the map *Det* as follows:

$$\text{Det}(\vec{n}_i) = \Phi\{\vec{w} \in N \mid d(\vec{n}_i, \vec{w}) = \alpha\} \quad (1)$$

where we have that ϕ is a linear function and α indicates how close \vec{w} is to the \vec{n}_i and depends on the quantifier expressed by *Det*.

The intuitive reading of the above is that *Det* of a word \vec{n}_i is a linear combination of all the words that are α -close to \vec{n}_i . For instance, if *Det* is ‘few’, then α is a small number (closer to 0 than to 1), indicating that we are taking the combination of vectors that are not so close to \vec{n}_i . If *Det* is ‘most’, then α will be a large number (closer to 1 than to 0), indicating that we are taking the combination of vectors that are close to \vec{n}_i . Here we are interpreting the notion of *closeness* based on the number of properties the words share. The distance α can be learnt from a corpus using a relevant task.

The underlying idea here is that the quantitative way of quantifying in set-theoretic models, which depends on the cardinality of the quantified sets, is now transformed into a geometric way of quantifying where the meaning of the quantified phrase depends on its geometric distance with other words. Hence, a quantified phrase such as ‘few cats’ returns a representative noun (obtained by taking the average of all such nouns) that is far from vector of ‘cat’ in the semantic space. This representative noun shares ‘few’ properties with ‘cat’. A quantified phrase such as ‘most cats’ returns a representative noun that is close the the vector of ‘cat’ and stands for a noun that shares ‘most’ of the properties of ‘cat’.

With this instantiation, the meaning of “Det Sbj Verb” is obtained by computing the following:

$$(\epsilon_N \otimes 1_S) \circ (Det \otimes \mu_N \otimes 1_S) \circ (\delta_N \otimes 1_{N \otimes S}) \left(\overrightarrow{\text{Sbj}} \otimes \overrightarrow{\text{Verb}} \right)$$

In the first step of computation we have:

$$(\delta_N \otimes 1_{N \otimes S}) \left(\sum_i C_i \vec{n}_i \otimes \sum_{jk} C_{jk} \vec{n}_j \otimes \vec{s}_k \right) = \left(\sum_i C_i \vec{n}_i \otimes \vec{n}_i \right) \otimes \left(\sum_{jk} C_{jk} \vec{n}_j \otimes \vec{s}_k \right)$$

In the second step we obtain:

$$\begin{aligned} (Det \otimes \mu_N \otimes 1_S) \left(\left(\sum_i C_i \vec{n}_i \otimes \vec{n}_i \right) \otimes \left(\sum_{jk} C_{jk} \vec{n}_j \otimes \vec{s}_k \right) \right) &= \sum_{ijk} C_i C_{jk} Det(\vec{n}_i) \otimes \mu(\vec{n}_i \otimes \vec{n}_j) \otimes \vec{s}_k \\ &= \sum_{ijk} C_i C_{jk} Det(\vec{n}_i) \otimes \delta_{ij} \vec{n}_i \otimes \vec{s}_k \end{aligned}$$

The final step is as follows:

$$(\epsilon_N \otimes 1_S) \left(\sum_{ijk} C_i C_{jk} Det(\vec{n}_i) \otimes \delta_{ij} \vec{n}_i \otimes \vec{s}_k \right) = \sum_{ijk} C_i C_{jk} \langle Det(\vec{n}_i) \mid \delta_{ij} \vec{n}_i \rangle \vec{s}_k$$

Example. As a distributional example, take N to be the two dimensional space with the basis $\{\vec{n}_1, \vec{n}_2\}$ and S be the two dimensional space with the basis $\{\vec{s}_1, \vec{s}_2\}$. Suppose the linear expansion of

$$\vec{N} := C_1 \vec{n}_1 + C_2 \vec{n}_2$$

and the linear expansion of

$$\vec{VP} := C_{11}(\vec{n}_1 \otimes \vec{s}_1) + C_{12}(\vec{n}_1 \otimes \vec{s}_2) + C_{21}(\vec{n}_2 \otimes \vec{s}_1) + C_{22}(\vec{n}_2 \otimes \vec{s}_2)$$

Suppose further the following for the interpretation of the determiner:

$$Det(\overrightarrow{\text{Sbj}}) = Det(C_1 \vec{n}_1 + C_2 \vec{n}_2) = Det(C_1 \vec{n}_1) + Det(C_2 \vec{n}_2) = C'_1 \vec{n}_1 + C'_2 \vec{n}_2 \quad (2)$$

Then the result of the first step of the computation of a meaning vector for the sentence ‘Q Sbj Verb’ is:

$$(C_1 \vec{n}_1 + C_2 \vec{n}_2) \otimes (C_{11} \vec{n}_1 \otimes \vec{s}_1 + C_{12} \vec{n}_1 \otimes \vec{s}_2 + C_{21} \vec{n}_2 \otimes \vec{s}_1 + C_{22} \vec{n}_2 \otimes \vec{s}_2)$$

In the second step of computation we obtain:

$$\begin{aligned} C_1 C_{11} Det(\vec{n}_1) \vec{n}_1 \otimes \vec{s}_1 + C_1 C_{12} Det(\vec{n}_1) \vec{n}_1 \otimes \vec{s}_2 + C_2 C_{21} Det(\vec{n}_2) \vec{n}_2 \otimes \vec{s}_1 + \\ C_2 C_{22} Det(\vec{n}_2) \vec{n}_2 \otimes \vec{s}_2 \end{aligned}$$

Since Det is a linear map, the above is equal to the following:

$$Det(C_1 \vec{n}_1)(C_{11} \vec{n}_1 \otimes \vec{s}_1 + C_{12} \vec{n}_1 \otimes \vec{s}_2) + Det(C_2 \vec{n}_2)(C_{21} \vec{n}_2 \otimes \vec{s}_1 + C_{22} \vec{n}_2 \otimes \vec{s}_2)$$

According to the expansion of the assumption in equation 2, the above is equivalent to the following:

$$(C'_1 \vec{n}_1)(C_{11} \vec{n}_1 \otimes \vec{s}_1 + C_{12} \vec{n}_1 \otimes \vec{s}_2) + (C'_2 \vec{n}_2)(C_{21} \vec{n}_2 \otimes \vec{s}_1 + C_{22} \vec{n}_2 \otimes \vec{s}_2)$$

Substituting this in the last step of the computation provide us with the following vector in S :

$$C'_1 C_{11} \vec{s}_1 + C'_1 C_{12} \vec{s}_2 + C'_2 C_{21} \vec{s}_1 + C'_2 C_{22} \vec{s}_2 \quad (3)$$

Small Corpus-Based Witness. A large scale experimentation for this model constitutes work in progress. For the sake of providing intuitions for the above symbolic constructions, we provide a couple of corpus-based witnesses here. In the distributional models, the most natural instantiation of the distance d in equation 1 is the co-occurrence distance. For a noun ' n ' and determiners ' few ' and ' $most$ ', we define these as generally as follows:

$$\begin{aligned} few(n) &= Avg\{\text{nouns that share few properties with } n\} \\ most(n) &= Avg\{\text{nouns that share most properties with } n\} \end{aligned}$$

For the purpose of this toy-example, the above can be instantiated in the simplest possible way as follows:

$$\begin{aligned} few(n) &= Avg\{\text{nouns that co-occurred with } n \text{ few times}\} \\ most(n) &= Avg\{\text{nouns that co-occurred with } n \text{ most times}\} \end{aligned}$$

In this case, a sample query from the online *Reuter News Corpus*, with at most 100 outputs per query, provides the following instantiations:

$$\begin{aligned} few(dogs) &= Avg\{\text{bike, drum, snails}\} \\ most(dogs) &= Avg\{\text{cats, pets, birds, puppies}\} \\ few(cats) &= Avg\{\text{fluid, needle, care}\} \\ most(cats) &= Avg\{\text{dogs, birds, rats, feces}\} \end{aligned}$$

A cosine-based similarity measure over this corpus results in the fact that any of the words in the ' $most(n)$ ' set are more similar to ' n ' than any of the words in the ' $few(n)$ ' set. This is indeed because the words in the former set are geometrically closer to ' n ' than the words in the latter set, since they have co-occurred with them more. This is the first advantage of our model over a distributional model, where words such as ' few ' and ' $most$ ' are treated as noise and hence meanings of phrase such as ' few cats', ' $most$ cats', and ' $cats$ ' become identical (and similarly for any other noun). Moreover, in our setting we can establish that ' $most$ cats' and ' $most$ dogs' have similar meanings, because of the over lap of their determiner sets. A larger corpus and a more thorough statistical analysis will let us achieve more, that for instance, ' few cats' and ' few dogs' also have similar meanings.

At the level of sentence meanings, compositional distributional models do not interpret determiners (e.g. see the model of [11]). As a result, meanings of sentences such as ‘cats sleep’, ‘most cats sleep’ and ‘few cats sleep’ will become identical; meanings of sentences ‘most cats sleep’ and ‘few dogs snooze’ become very close, since ‘cats’ and ‘dogs’ often occur in the same context and so do ‘sleep’ and ‘snooze’. In our setting, equation 3 tells us that these sentences have different meanings, since their quantified subjects have different meanings. To see this, take $\overrightarrow{\text{cats sleep}} = C_1 \overrightarrow{n}_1 + C_2 \overrightarrow{n}_2$, where as $\overrightarrow{\text{few(cats)}} = C'_1 \overrightarrow{n}_1 + C'_2 \overrightarrow{n}_2$ and $\overrightarrow{\text{most(cats)}} = C''_1 \overrightarrow{n}_1 + C''_2 \overrightarrow{n}_2$. Instantiating these in equation 3 provides us with the following three different vectors:

$$\begin{aligned} \overrightarrow{\text{cats sleep}} &= C_1 C_{11} \overrightarrow{s}_1 + C_1 C_{12} \overrightarrow{s}_2 + C_2 C_{21} \overrightarrow{s}_2 + C_2 C_{22} \overrightarrow{s}_2 \\ \overrightarrow{\text{few cats sleep}} &= C'_1 C_{11} \overrightarrow{s}_1 + C'_1 C_{12} \overrightarrow{s}_2 + C'_2 C_{21} \overrightarrow{s}_2 + C'_2 C_{22} \overrightarrow{s}_2 \\ \overrightarrow{\text{most cats sleep}} &= C''_1 C_{11} \overrightarrow{s}_1 + C''_1 C_{12} \overrightarrow{s}_2 + C''_2 C_{21} \overrightarrow{s}_2 + C''_2 C_{22} \overrightarrow{s}_2 \end{aligned}$$

On the other hand, we have that ‘most cats sleep’ and ‘most dogs snooze’ have close meanings, one which is close to ‘pets sleep’. This is because, their quantified subjects and their verbs have similar meanings, that is we have:

$$\left\{ \begin{array}{l} \overrightarrow{\text{most(dogs)}} \sim \overrightarrow{\text{most(cats)}} \sim \overrightarrow{\text{pets}} \\ \overrightarrow{\text{snooze}} \sim \overrightarrow{\text{sleep}} \end{array} \right. \implies \overrightarrow{\text{most cats sleep}} \sim \overrightarrow{\text{most dogs snooze}} \sim \overrightarrow{\text{pets sleep}}$$

At the same time, ‘few cats sleep’ and ‘most dogs snooze’ have a less-close meaning, since their quantified subjects have different meanings, that is:

$$\overrightarrow{\text{most(dogs)}} \not\sim \overrightarrow{\text{few(cats)}} \implies \overrightarrow{\text{most dogs snooze}} \not\sim \overrightarrow{\text{few cats sleep}}$$

References

- [1] M. Baroni and R. Zamparelli. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Conference on Empirical Methods in Natural Language Processing (EMNLP-10)*, Cambridge, MA, 2010.
- [2] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2):159–219, 1981.
- [3] B. Coecke and E. Paquette. Introducing categories to the practicing physicist. In B. Coecke, editor, *New Structures for Physics*, volume 813 of *Lecture Notes in Physics*, pages 167–271. Springer, 2008.
- [4] B. Coecke, M. Sadrzadeh, and S. Clark. Mathematical Foundations for Distributed Compositional Model of Meaning. *Lambek Festschrift. Linguistic Analysis*, 36:345–384, 2010.
- [5] J. Curran. *From Distributional to Semantic Similarity*. PhD thesis, School of Informatics, University of Edinburgh, 2004.
- [6] J.R. Firth. A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis*. 1957.
- [7] E. Grefenstette and M. Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1394–1404, 2011.

- [8] E. Grefenstette and M. Sadrzadeh. Prior disambiguation of word tensors for constructing sentence vectors. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, page 15901601, 2013.
- [9] G.M. Kelly and M.L. Laplaza. Coherence for compact closed categories. *Journal of Pure and Applied Algebra*, 19(0):193 – 213, 1980.
- [10] Anders Kock. Strong functors and monoidal monads. *Archiv der Mathematik*, 23:113–120, 1972.
- [11] Jeff Mitchell and Mirella Lapata. Composition in distributional models of semantics. *Cognitive Science*, 34:1388–1439, 2010.
- [12] Clark S. Sadrzadeh M. and C. Coecke. Forbenius anatomy of word meanings i: subject and object relative pronouns. *Journal of Logic and Computation*, 23:1293–1317, 2013.
- [13] Clark S. Sadrzadeh M. and C. Coecke. Forbenius anatomy of word meanings 2: possessive relative pronouns. *Journal of Logic and Computation*, doi: 10.1093/logcom/exu027, 2014.