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

Ondřej Rypáček
University of Oxford
Dept. of Computer Science
ondrej@cs.ox.ac.uk

Mehrnoosh Sadrzadeh
Queen Mary University of London
School of Electronic Eng. and Computer Science
mehrs@eecs.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 and instantiate it 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.

1 Introduction

Vector space 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 [**?**]. One then fixes a context window, of for instance 5 words, and computes frequencies of how many times a word has occurred in this window with other words. These frequencies are often normalised to be better representatives of rare and very common words. These models have been applied to various language processing tasks, for instance thesauri construction [**?**]. Compositional distributional models of meaning extend the vector space models from words to sentences. The categorical such models [**?**,**?**] 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 [**?**,**?**]. Nevertheless, it has been an open problem how to dealt 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 [**?**].

According to generalised quantifier theory, the meaning of a sentence with a natural language quantifier Q such as 'Q Sbj Verb' is determined by first taking the intersection of the denotation of Sbj with the denotation of subjects of the Verb, then checking if the denotation of $Q(\text{Sbj})$ is an element of this set. The denotation of Q is specified separately, for example, for $Q = \exists$, it is the set of non-empty subsets of the universe, for $Q = 2$ it is the set of subsets of the universe that have exactly 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.

In what follows, we work in the categorical compositional distributional model of [**?**]. 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. We present two concrete interpretations for this abstract setting: relations and vector spaces. The former is the basis for a truth-theoretic model and the latter works for a corpus-base model of language.

Our future work includes formalising this rather low-level treatment in the setting of categorical logic, where quantifiers are proven to be adjoints to substitution.

Lack of much structure in vector spaces (and compact closed categories in general) and in particular lack of existence of pull-backs will be obvious obstacles. WHY IT IS HARD TO DO CATEGORICAL LOGIC IN VECTOR SPACES???

CONTRIBUTION: POSSIBLE TO DO GEN QUANT IN COMPACT CLOSED CATEGORIES with Frob on some objects

DIFFERENT INSTANTAITIONS TO GET DIFFERENT TYPES OF MEANING

LEAVE THE LOGIC FOR LATER work in a fragment without logical operations such as conjunction etc

REFER TO BARONI-RAFA PAPER

We also aim to experiment with this model on corpus-based datasets and tasks.

2 Preliminaries

2.1 Corpus-Based Vector Space models of Natural Language

These models are also referred to by *distributional* models. The idea behind them mainly originates from the work of Firth [**?**] who argued that the meaning of a word depends on the context in which it often occurs. Based on this idea, computational linguists developed a vector space model for word meanings [**?**]. We denote such a model with a tuple $(V_{l[\Sigma]}, \rightarrow)$, described below:

- $l \subseteq \Sigma \times \Sigma$ is a relation on the vocabulary Σ of the language: for each $w \in \Sigma$, we have that $l(w)$ is the canonical form, otherwise known as the lemma, corresponding to the word w. Hence $l[\Sigma]$, which is the image of l on Σ is the set of lemmas of the vocabulary Σ of a language. An example of a lemma is 'kill': the canonical form of words such as 'kill', 'kills', 'killed', 'killing', etc.
- $V_{l(\Sigma)}$ is a vector space whose basis vectors are the lemmas of the language; we refer to each such basis vectors by $\overrightarrow{v}_i$.
- The meaning of each word w is represented by a vector $\overrightarrow{w} = \sum_i C_i \overrightarrow{v}_i$, obtained by counting and normalising the number of times w has occurred in the context window of any of the words in $l^{-1}(v_i)$. We refer to each such C_i by 'the degree to which w has co-occurred with $\overrightarrow{v}_i$. The length of the context window and the normalisation scheme are usually treated as parameters of the model.

Here is an example of such a model:

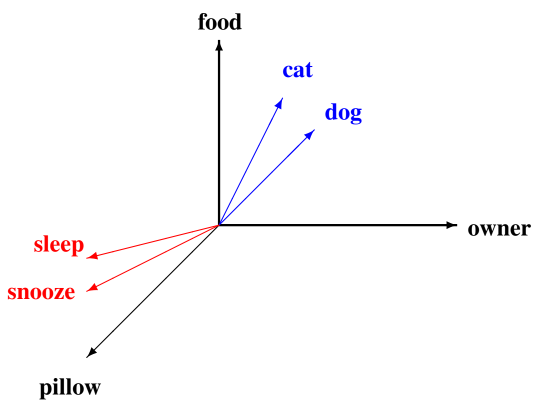


Fig. 1. An example of a vector space model of meaning

In Figure 1, for instance for $n = 5$, the coordinate of cat on food is the normalised count of the number of times cat has occurred 5 words before or after any word whose lemma is food. A normalisation scheme is to divide the number of occurrences by the total number of times a word has occurred in a corpus.

The most mainstream application of distributional models has been to deciding about the degree of similarity between meanings of words [?]. The cosine of the angle of the vector representations of words has proven to be a good measure of such similarity decisions. This is backed by the philosophy behind these models, which implies that words that have similar meanings often occur in the same context. For instance in the above example, we have a high degree of similarity between meanings of words cat and dog and less so between dog and snooze.

2.2 Generalised Quantifier Theory in Natural Language

We briefly review the theory of generalised quantifiers in natural language [?]. The definitions and concepts are slightly revised from their original presentation so that they could fit easier within the categorical setting, introduced later on. Given a vocabulary Σ and a set of lexical categories S, the words of the vocabulary are assigned lexical categories via the relation $\mathcal{X} \subseteq \Sigma \times S$, otherwise known as a lexicon or a dictionary. The expressions of the language are generated over these categories via a set of rules $\mathcal{R} \subseteq S \times S \times S$. We refer to a language so defined by the tuple $\mathcal{L}_\Sigma = (\mathcal{X}_S, \mathcal{R})$.

For the purpose of this paper, we consider the fragment of English generated over the following lexicon:

Syntactic Category	Vocabulary
NP	John, Mary, something, $\cdots$
N	cat, dog, man, $\cdots$
VP	sneeze, sleep,$\cdots$
V	love, kiss, $\cdots$
Det	a, the, some, every, each, all, no, most, few, one, two, $\cdots$

with the following phrase structure description rules:

$$\begin{aligned} \text{S} &\to \text{NP VP} \\ \text{VP} &\to \text{V NP} \\ \text{NP} &\to \text{Det N} \end{aligned}$$

A model for this language is a pair $(U, [\![\ ]\!])$, where U is a universal reference set and $[\![\ ]\!]$ is a function that assigns interpretations to the expressions of language. The interpretations of basic expressions are as follows:

$$\begin{aligned} np \in \text{NP} &\quad\Longrightarrow\quad [\![np]\!] \in U \\ n \in \text{N} &\quad\Longrightarrow\quad [\![n]\!] \subseteq U \\ vp \in \text{VP} &\quad\Longrightarrow\quad [\![vp]\!] \subseteq U \\ v \in \text{V} &\quad\Longrightarrow\quad [\![v]\!] \subseteq U \times U \\ d \in \text{Det} &\quad\Longrightarrow\quad [\![d]\!] : \mathcal{P}(U) \to \mathcal{P}\mathcal{P}(U) \end{aligned}$$

Noun phrases are interpreted as elements of the reference set, nouns and verb phrases as unary relations over it, and verbs as binary relations over it. Determiners are interpreted as *generalised quantifier*; a generalised quantifier d is a map that assigns to each $A \subseteq U$, a family of subsets of U; examples are as follows:

$$
\begin{aligned}
[\![some]\!](A) &= \{X \subseteq U \mid X \cap A \neq \emptyset\}\\
[\![\text{E}very]\!](A) &= \{X \subseteq U \mid A \subseteq X\}\\
[\![\text{no}]\!](A) &= \{X \subseteq U \mid A \cap X = \emptyset\}\\
[\![n]\!](A) &= \{X \subseteq U \mid \mid X \cap A \mid = n\}\\
[\![\text{m}ost]\!](A) &= \{X \subseteq U \mid X \text{ contains most } A\text{'s}\}\\
[\![\text{f}ew]\!](A) &= \{X \subseteq U \mid X \text{ contains few } A\text{'s}\}
\end{aligned}
$$

Note that the interpretations of determiners are not quantifiers (in the sense of first order logic) yet. They become quantifiers when applied to a set. For instance, $[\![d]\!]$ is not a quantifier, but $[\![d]\!](A)$ is. The interpretations of determiners satisfy a property referred to by *living on* or *conservativity*. This property says that a set X belongs to $[\![d]\!](A)$ iff $(X \cap A) \in [\![d]\!](A)$. This property is sometimes described by saying that 'the quantifier $[\![d]\!](A)$ lives on A'.
The interpretations of expressions generated by the rules is obtained by induction as follows:

- a noun phrase 'NP' generated by the rule 'NP $\to$ Det N'

$$
\begin{array}{rcl}
[\![\text{Det N}]\!] & = & [\![d]\!]([\![n]\!]) \quad \text{where} \quad X \in [\![d]\!]([\![n]\!]) \textbf{ iff } (X \cap [\![n]\!]) \in [\![d]\!]([\![n]\!])\\
& & \qquad\qquad\quad\ \text{for} \quad d \in Det, n \in N
\end{array}
$$

- a verb phrase 'VP' generated by the rule 'VP $\to$ V NP'

$$
[\![\text{V NP}]\!] = \{x \mid [\![np]\!]\,(\{y \mid [\![v(x,y)]\!]\})\} \qquad \text{for} \quad np \in NP, v \in V
$$

- a sentence 'S' generated by the rule 'S $\to$ NP VP'

$$
[\![\text{N}P\ VP]\!] = [\![vp]\!]([\![np]\!]) \qquad \text{for} \quad np \in NP, vp \in VP
$$

The *meaning* of a sentence is said to be *true* iff its semantic interpretation is non-empty and *false* otherwise that is we have:

Definition 1. *The meaning of a sentence in generalised quantifier theory is true iff* $[\![vp]\!]([\![np]\!]) = 1$ *and false otherwise.*

As an example, consider the meaning of a sentence with a quantified phrase at its subject position. This sentence has the form 'Det N VP' and its meaning is defined as follows:

$$
[\![\text{Det N VP}]\!] \;=\; \begin{cases} 1 & [\![vp]\!] \in [\![\text{Det N}]\!] \quad \text{for } vp \in \text{VP}\\ 0 & \text{other wise} \end{cases}
$$

By the *living on* property, the meaning of this sentence is true whenever $[\![vp]\!] \cap [\![n]\!] \in [\![\text{Det N}]\!]$. For instance, meaning of 'some cats sneeze' is true whenever $[\![\text{s}neeze]\!] \cap [\![\text{m}en]\!] \in [\![\text{some men}]\!]$. That is, whenever the set of things that sneeze and are men is a non-empty set. Similarly, the meaning of the sentence 'five men sneeze' is true whenever the set of things that sneeze and are men has five elements and so on.

As another example, consider the meaning of a sentence with a quantified phrase at its object position. This sentence has the form 'NP V Det N' and its meaning is defined as follows:

$$
[\![\text{NP V Det N}]\!] \;=\; \begin{cases} 1 & [\![np]\!] \in \{x \mid \{y \mid [\![v(x,y)]\!]\} \in [\![\text{Det N}]\!]\} \quad \text{for } v \in \text{V}, np \in \text{NP}\\ 0 & \text{otherwise} \end{cases}
$$

That is, the meaning of this sentence is true whenever $\{y \mid [\![v([\![np]\!], y)]\!]\} \in [\![\text{Det N}]\!]$, which by the *living on* property is the case whenever $\{y \mid [\![v([\![np]\!], y)]\!]\} \cap [\![n]\!] \in [\![\text{Det N}]\!]$, for n in N. For instance, meaning of 'John kissed some cats' is true whenever $\{y \mid [\![\text{kiss}([\![John]\!], y)]\!]\} \cap [\![cats]\!] \in [\![\text{some cats}]\!]$. That is, whenever, the set of things that are kissed by John and are cats is a non-empty set. Similarly, the sentence 'John kissed five cats' is true whenever the set of things that are kissed by John and are cats has five elements and so on.

2.3 Type-Logical Grammar

In this section we show how to syntactically analyse the sentences of the previous fragment in a pregroup type logic.

A pregroup algebra $P = (P, \leq, \cdot, (-)^r, (-)^l)$ is a partially ordered monoid where every element has a left and a right adjoint, that is, for every element $p \in P$, there are two elements $p^l, p^r \in P$, referred to as its left and right adjoint, and these satisfy the following four inequalities

$$p \cdot p^r \leq 1 \leq p^r \cdot p \qquad p^l \cdot p \leq 1 \leq p \cdot p^l$$

A pregroup grammar over a vocabulary Σ is a pregroup algebra $P_{\mathcal{B}}$, freely generated over a set of atomic logical types $\mathcal{B}$, together with a type-dictionary $\beta \subseteq \Sigma \times P_{\mathcal{B}}$, which assigns to each word of the vocabulary a type from the pregroup. We denote this grammar by the tuple $\mathcal{G}_{\Sigma} = (\beta, P_{\mathcal{B}})$.

For the vocabulary and the fragment of language described in the previous section, the set of atomic types are $\{m, n, s\} \in \mathcal{B}$, where m is a noun phrase, n is a noun, and s is a declarative sentence. The dictionary is as follows:

Words	Logical Types
John, Mary, something, $\cdots$	m
cat, dog, man, $\cdots$	n
sneeze, sleep,$\cdots$	$m^r \cdot s$
love, kiss, $\cdots$	$m^r \cdot s \cdot m^l$
a, the, some, every, each, all, no, most, few, one, two, $\cdots$	$m \cdot n^l$

We assign atomic types m and n to the words in the lexical categories NP and N. The type assigned to the words in VP is $m^r \cdot s$, this means that words in the lexical category VP input an argument of type m and they have to be to the right of that argument, then output a sentence of type s. The type assigned to the words in the lexicon item V is $m^r \cdot s \cdot m^l$, this means that these words input two arguments of type m and they have to be to the right of one and to the left of the other, then output a sentence. Finally, the type assigned to the words in the lexicon item Det is $m \cdot n^l$, which means that these words input an argument of the type n and they have to be to the left of that argument, then output a phrase of type m.

The grammatical reductions of the language are modelled by the partial ordering of the pregroup grammar. As an example, consider a sentence with a quantified phrase in its subject position, e.g. 'some cats sneeze', the grammatical reduction of this sentence is as follows:

$$\begin{array}{l} \text{some} \quad \text{cats} \quad \text{sneeze} \\ (m \cdot n^l) \;\; \cdot n \;\; \cdot (m^r \cdot s) \leq m \cdot 1 \cdot (m^r \cdot s) = m \cdot (m^r \cdot s) \leq 1 \cdot s = s \end{array}$$

Here, first 'some' inputs 'cats' and output a noun phrase of type m, then the verb inputs m and outputs a sentence. As another example, consider a sentence with a quantified phrase in its object position, e.g. 'John kissed some cats' the grammatical reduction of this sentence is as follows:

$$\begin{array}{ccccccc} \text{John} & \text{kissed} & \text{some} & \text{cats} & & & \\ m & \cdot(m^r \cdot s \cdot m^l) & \cdot(m \cdot n^l) & \cdot n & \leq & 1 \cdot (s \cdot m^l) \cdot m \cdot 1 = (s \cdot m^l) \cdot m \leq s \cdot 1 = s \end{array}$$

Here, again first 'some' inputs 'cats' and outputs a noun phrase, at the same time the verb inputs 'John' and outputs a verb phrase of type $s \cdot m^l$, which then inputs the m from the phrase 'some cats' and outputs a sentence.

Whereas the elements of lexical categories are syntactic, the types of a pregroup algebra have logical information in them. This information comes from the rules of the grammar of the language; in a lexical approach such information is encoded in the rules of the language. There are standard methods that transfer the lexical categories and rules of a language to logical type [**?**]. For the fragment of the language considered in this paper, we suffice to giving the following translation $t \colon S \to P_{\mathcal{B}}$ between the lexical categories and the pregroup grammar types:

Syntactic Category	Logical Type
NP	m
N	n
VP	$m^r \cdot s$
V	$m^r \cdot s \cdot m^l$
Det	$m \cdot n^l$

2.4 Category Theoretic and Diagrammatic Definitions and Axioms

This subsection briefly reviews compact closed categories and Frobenius algebras. For a formal presentation, see [**?**,**?**]. A compact closed category, $\mathcal{C}$, has objects A, B; morphisms $f \colon A \to B$; 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:

$$A \otimes A^r \xrightarrow{\epsilon^r_A} I \xrightarrow{\eta^r_A} A^r \otimes A \qquad A^l \otimes A \xrightarrow{\epsilon^l_A} I \xrightarrow{\eta^l_A} A \otimes A^l$$

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

$$(1_A \otimes \epsilon^l_A) \circ (\eta^l_A \otimes 1_A) = 1_A \qquad (\epsilon^r_A \otimes 1_A) \circ (1_A \otimes \eta^r_A) = 1_A$$
$$(\epsilon^l_A \otimes 1_A) \circ (1_{A^l} \otimes \eta^l_A) = 1_{A^l} \qquad (1_{A^r} \otimes \epsilon^r_A) \circ (\eta^r_A \otimes 1_{A^r}) = 1_{A^r}$$

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 $\mathcal{C}$:

$$\delta \colon X \to X \otimes X \qquad\qquad \iota \colon X \to I$$

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

$$\mu \colon X \otimes X \to X \qquad\qquad \zeta \colon I \to X$$

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.

Pregroup Algebras. A pregroup algebra $P = (P, \leq, \cdot, (-)^l, (-)^r)$ is a compact closed category whose objects are the elements of the set $p \in P$ are the objects of the category and the partial ordering between the elements are the morphisms. That is, for $p, q \in P$, we have that $p \to q$ is a morphism of the category iff $p \leq q$ in the partial order. The tensor product of the category is the monoid multiplication, whose unit is 1, and the adjoints of objects are the adjoints of the elements of the algebra. The epsilon and eta morpshism are thus as follows:

$$p \cdot p^r \xrightarrow{\epsilon_p^r} 1 \xrightarrow{\eta_p^r} p^r \cdot p \qquad p^l \cdot p \xrightarrow{\epsilon_p^l} 1 \xrightarrow{\eta_p^l} p \cdot p^l$$

The yanking equalities directly follow from the preroup inequalities on the adjoints. A pregroup with Frobenius structure on it becomes degenerate. To see this, suppose we have such an algebra on the object p of such a pregroup. Then the unit morphism of the internal comonoid of this algebra becomes the parietal ordering $\iota\colon p \leq 1$; taking the right adjoints of both sides of this inequality will yield $1 = 1^r \leq p^r$, and by the multiplying both sides of this with p we will obtain $p \leq p \cdot p^r$, which by adjunction results in $p \leq p \cdot p^r \leq 1$, hence we have $p \leq 1$ and also $1 \leq p$, thus p must be equal to 1. That is, assuming that we have a Frobenius algebra on an object will mean that that object is 1.

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 $f\colon V \to W$ 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 naturual 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:

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

Similarly, eta maps are defined as follows:

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

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

$$\begin{aligned}
\delta\colon V \to V \otimes V \quad & \text{given by} \quad \overrightarrow{v_i} \mapsto \overrightarrow{v_i} \otimes \overrightarrow{v_i} \\
\mu\colon V \otimes V \to V \quad & \text{given by} \quad \overrightarrow{v_i} \otimes \overrightarrow{v_j} \mapsto \delta_{ij} \overrightarrow{v_i} \\
\iota\colon V \to \mathbb{R} \quad & \text{given by} \quad \overrightarrow{v_i} \mapsto 1 \\
\zeta\colon \mathbb{R} \to V \quad & \text{given by} \quad 1 \mapsto \sum_i \overrightarrow{v_i}
\end{aligned}$$

Relations. Another important example of a compact closed category is Rel, the cateogry of sets and relations. Here, $\otimes$ is cartesian product with the singleton set as its unit $I = \{\star\}$, and * is identity 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:

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

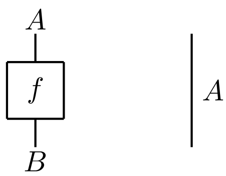
Every object in Rel has a Frobenius algebra over it given by the diagonal and codiagonal relations, as described below:

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

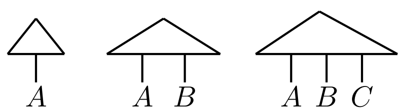
For the details of verifying that for each of the two examples above, the corresponding conditions hold see [**?**].

2.5 String Diagrams

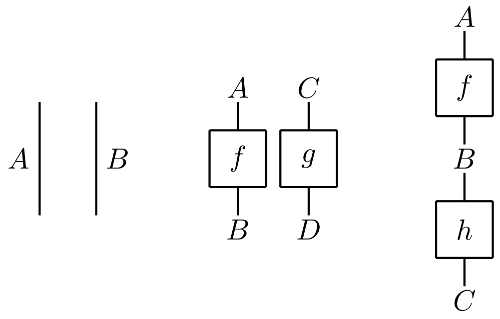
The framework of compact closed categories and Frobenius algebras comes with a complete diagrammatic calculus that visualises derivations, and which also simplifies the categorical and vector space computations. Morphisms are depicted by boxes and objects by lines, representing their identity morphisms.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



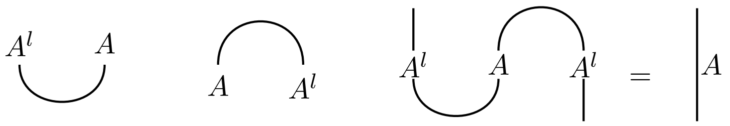
Morphisms from I to objects are depicted by triangles with strings emanating from them. In concrete categories, these morphisms represent elements within the objects. For instance, an element a in A is represented by the morphism $a : I \to A$ and depicted by a triangle with one string emanating from it. The number of strings of such triangles depict the tensor rank of the element; 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 $f \colon A \to B$, $g \colon C \to D$, and $h \colon B \to C$, are depicted as follows:

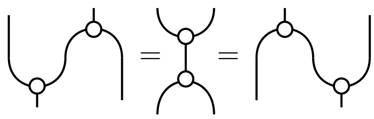


The ϵ maps are depicted by cups, η maps by caps, and yanking by their composition and straightening of the strings. For instance, the diagrams for $\epsilon^l\colon A^l \otimes A \to I$, $\eta\colon I \to A \otimes A^l$ and $(\epsilon^l \otimes 1_A) \circ (1_A \otimes \eta^l) = 1_A$ are as follows:

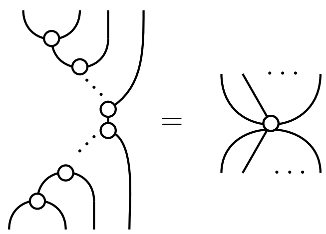


As for Frobenius algebras, the diagrams for the monoid and comonoid morphisms are as follows:

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':

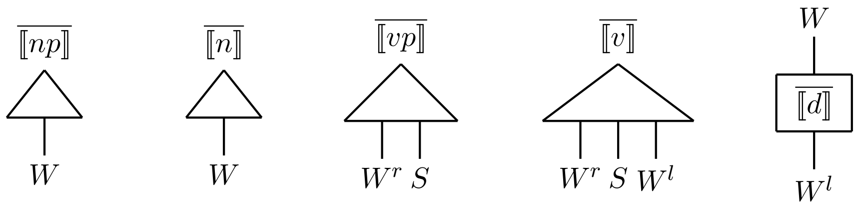


3 Abstract Compact Closed Semantics

An abstract compact closed categorical model for the language $\mathcal{L}_\Sigma = (\mathcal{X}_S, \mathcal{R})$ is a tuple $(\mathcal{C}, W, S, \overline{[\![\ ]\!]})$ where $\mathcal{C}$ is a compact closed category with two distinguished objects W and S where W has a Frobenius algebra on it and $\overline{[\![\ ]\!]}$ is a function that assigns morphisms from this category to expression of the language. The interpretations of the basic expressions are as follows:

$$\begin{aligned}
(np, \mathrm{NP}) \in \mathcal{X}_S \quad &\Longrightarrow \quad \overline{[\![np]\!]} \colon I \to W \\
(n, \mathrm{N}) \in \mathcal{X}_S \quad &\Longrightarrow \quad \overline{[\![n]\!]} \colon I \to W \\
(vp, \mathrm{VP}) \in \mathcal{X}_S \quad &\Longrightarrow \quad \overline{[\![vp]\!]} \colon I \to W^r \otimes S \\
(v, \mathrm{V}) \in \mathcal{X}_S \quad &\Longrightarrow \quad \overline{[\![v]\!]} \colon I \to W^r \otimes S \otimes W^l \\
(d, \mathrm{Det}) \in \mathcal{X}_S \quad &\Longrightarrow \quad \overline{[\![d]\!]} \colon W \to W
\end{aligned}$$

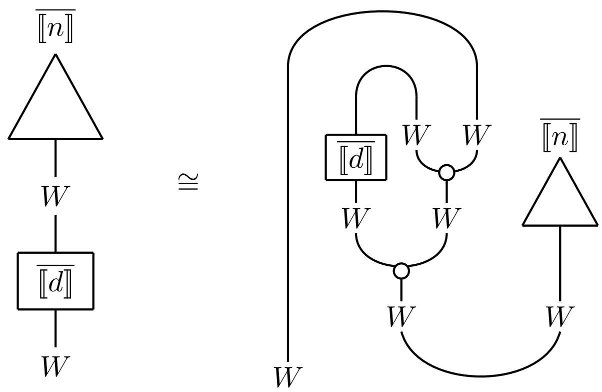
The diagrammatic semantics of the above interpretations are as follows:



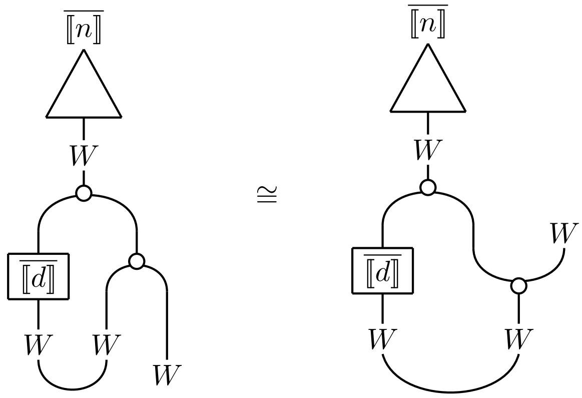
Noun phrases and nouns are elements within the object W; the former is a singleton element and the latter not necessarily so. The abstract language and its diagrammatic representation do not have means of distinguishing the two; when we instantiate these to concrete categories the difference between them becomes evident. Verb phrases are elements within the object $W^r \otimes S$; the intuition behind this representation is that in a compact closed category we have that $W^r \otimes S \cong W \to S$, where $W^r \to S = Hom(W, S)$ is an internal hom object of the category, coming from its monoidal closeness. Hence, we are modelling verb phrases as morphisms with input W and output S. Similarly, verbs are elements within the object $W^r \otimes S \otimes W^r$, equivalent to morphisms $W \otimes W \to S$ with pairs of input from W and output S.

The interpretations of the expressions generated by the rules are defined by induction as follows:

– $\overline{[\![\text{Det N}]\!]} := \overline{[\![d]\!]} \circ \overline{[\![n]\!]}$, where $\overline{[\![d]\!]} \circ \overline{[\![n]\!]} \cong (\epsilon_W \otimes 1_W) \circ (\overline{[\![d]\!]} \otimes \mu_W) \circ \sigma_W \circ \overline{[\![n]\!]}$. Diagrammatically, we have:

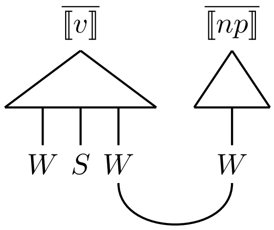


This condition expresses the ‘living on property. Intuitively, by using the axioms of compact closed categories and Frobenius algebras, the right hand side diagram above simplifies to the following left hand side diagram below, which in turn is equivalent to the right hand side diagram below:

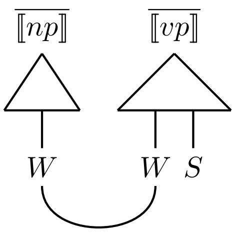


According to the diagram on the right hand side, semantics of $\overline{[\![d]\!]} \circ \overline{[\![n]\!]}$ is an element of W which is equivalent to the element obtained by making a copy (via the Frobenius map Δ_W) of the noun in W, applying the determiner to one copy and taking the intersection of the other copy (via the Frobenius map μ_W) with W.

– $\overline{[\![\text{V NP}]\!]} := (1_W \otimes 1_S \otimes \epsilon_W) \circ (\overline{[\![v]\!]} \otimes \overline{[\![np]\!]})$. Diagrammatically, we have:

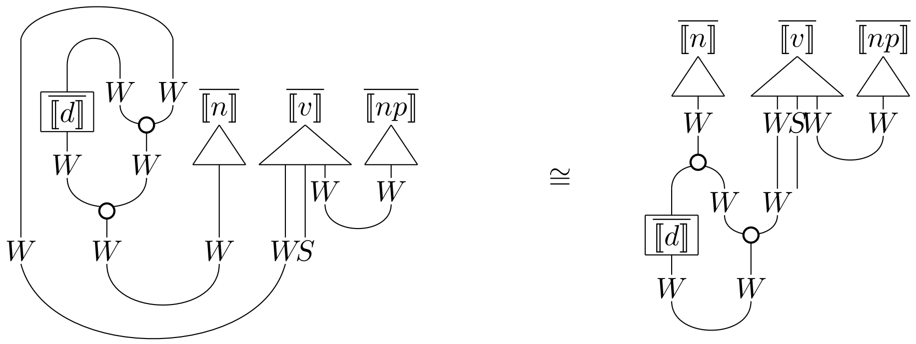


– $\overline{[\![\text{NP VP}]\!]} := (\epsilon_W \otimes 1_S) \circ (\overline{[\![np]\!]} \otimes \overline{[\![vp]\!]})$. Diagrammatically, we have:



In the abstract setting the meaning and semantic interpretation of sentences are the same: they both are represented by the object S. In the next section we show how to instantiate this setting to a concrete relational setting where meaning can be defined to be true or false. Here, we provide semantics interpretations for sentences with a quantified phrase at their subject and object position.

The interpretation of a sentence with a quantified phrase in subject position and its simplified forms are as follows:

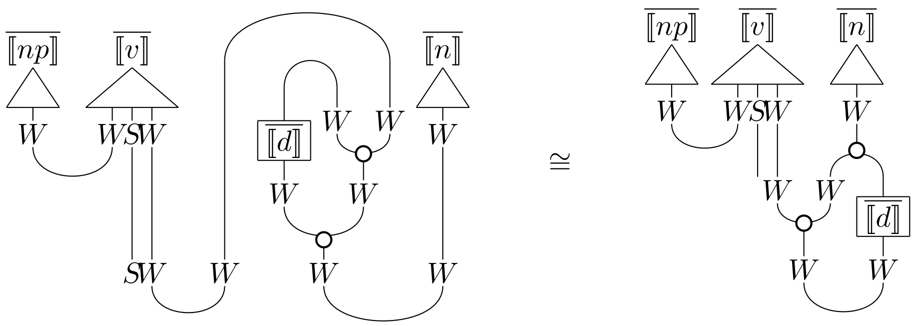


The symbolic representation of the simplified diagram above is as follows:

$$(\epsilon_W \otimes 1_S) \circ (\overline{[\![d]\!]} \otimes \mu_W \otimes 1_S) \circ (\Delta_W \otimes 1_{W \otimes S} \otimes \epsilon_W) \circ (\overline{[\![n]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![np]\!]})$$

Intuitively, the determiner 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 $\overline{[\![d]\!]}$ 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 $\overline{[\![d]\!]}$. 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.

A sentence with a quantified phrase in object position is generated by the rule 'NP V Det N'. Its diagrammatic meaning and its simplified form are as follows:

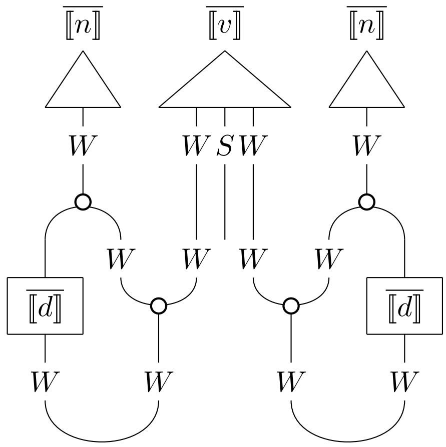


The symbolic representation of the simplified diagram above is as follows:

$$(1_S \otimes \epsilon_W) \circ (1_S \otimes \mu_W \otimes \overline{[\![d]\!]}) \circ (\epsilon_W \otimes 1_{S \otimes W} \otimes \Delta_W) \circ (\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]})$$

Intuitively, the determiner first makes a copy of the object (via the Frobenius Δ map), so now we have two copies of the object. One of these is being unified with the object argument of the verb (via the Frobenius μ map). In set-theoretic terms this is the intersection of the interpretations of object and objects-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 object of the verb and the noun belongs to the interpretation of the quantified noun.

Putting the two cases together, the semantic interpretation of a sentence with quantified phrases both at subject and at an object position has the following simplified form:



The symbolic representation of the above diagram is as follows:

$$(\epsilon_W \otimes 1_S \otimes \epsilon_W) \circ (\overline{[\![d]\!]} \otimes \otimes \mu_W \otimes 1_S \otimes \mu_W \otimes \overline{[\![d]\!]}) \circ (\Delta_W \otimes 1_{W \otimes S \otimes W} \otimes \Delta_W) \circ (\overline{[\![n]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]})$$

4 Truth Theoretic Interpretation

Given the set-theoretical model $(U, [\![\ ]\!])$ of a language $\mathcal{L}_\Sigma$, a relational instantiation of the abstract compact closed categorical interpretation is provided by the tuple $(Rel, \mathcal{P}(U), \{\star\}, \overline{[\![\ ]\!]})$, defined as follows:

- For a word with a lexical category N,NP, and VP, that is, for $s \in \{\text{N, NP, VP}\}$ and any $w \in \Sigma$ such that $(w, s) \in \mathcal{X}_\Sigma$, we have
$$\overline{[\![w]\!]} \colon \{\star\} \to \mathcal{P}(U)$$
which means $\overline{[\![w]\!]}$ is a subset of $\mathcal{P}(U)$ (morphisms in Rel are relations), that is an element of $\mathcal{P}\mathcal{P}(U)$. Note the words in VP are represented by $\overline{[\![w]\!]} \colon \{\star\} \to \mathcal{P}(U) \otimes \{\star\}$ and we have $\mathcal{P}(U) \otimes \{\star\} \cong \mathcal{P}(U)$.
- For words with lexical category V, we have
$$\overline{[\![w]\!]} \to \mathcal{P}(U) \otimes \{\star\} \otimes \mathcal{P}(U)$$
The right hand side is isomorphic to $\mathcal{P}(U) \otimes \mathcal{P}(U)$, which means $\overline{[\![w]\!]}$ is a subset of $\mathcal{P}(U) \times \mathcal{P}(U)$, that is an element of $\mathcal{P}(\mathcal{P}(U) \times \mathcal{P}(U))$.
- For a word with the lexical category Det, that is a $d \in \Sigma$ such that $(d, Det) \in \mathcal{X}$ and a $w \in \Sigma$ such that $(w, \text{N}) \in \mathcal{X}$, we have
$$\overline{[\![d]\!]} \colon \mathcal{P}(U) \to \mathcal{P}(U)$$
in Rel, i.e. a function from $\mathcal{P}(U)$ to $\mathcal{P}\mathcal{P}(U)$, namely a relational interpretation of the generalised quantifier map $[\![d]\!]$.

In this instantiation, the sentence space is the unit of the tensor, hence the meaning of a sentence with a quantified phrase at its subject position simplifies as follows:

$$\epsilon_{\mathcal{P}(U)}(\overline{[\![d]\!]} \otimes \mu_{\mathcal{P}(U)})(\Delta_{\mathcal{P}(U)} \otimes 1_{\mathcal{P}(U)} \otimes \epsilon_{\mathcal{P}(U)}) \circ (\overline{[\![n]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![np]\!]})$$

Similarly, meaning of a sentence with a quantified phrase at its object position simplifies as follows:

$$\epsilon_{\mathcal{P}(U)} \circ (\mu_{\mathcal{P}(U)} \otimes \overline{[\![d]\!]}) \circ (\epsilon_{\mathcal{P}(U)} \otimes 1_{\mathcal{P}(U)} \otimes \Delta_{\mathcal{P}(U)}) \circ (\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]})$$

Discussion about the choice of $\mathcal{P}(U)$. The reason we are working with $\mathcal{P}(U)$ and not U itself is that the interpretation of a determiner according to the abstract compact closed semantics developed in **??** is a morphism from W to W. According to the generalised quantifier theory explained in 2.2, the determiner is a function from $\mathcal{P}$ to $\mathcal{PP}(U)$, which is precisely a relation from $\mathcal{P}(U)$ to $\mathcal{P}(U)$. But as a consequence, meanings of words become a set of subsets: words with lexical category N,NP, and VP become a set of subsets of U; words with lexical category V become a set of pairs of subsets of U. Now we face a choice on how to interpret a set as a set of sets. According to the principle of indirect equality, down and up sets are obvious choices.

GIVE LEMMA

DEFINE WHAT DOWN AND UP SETS ARE

SAY WHY NOT UP SETS HERE

Example (I): Intransitive Verb. As a truth-theoretic example, suppose $U = \{m_1, m_2, c_1\}$, from which 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. Consider the following embedding:

$$\overline{[\![\text{men}]\!]} =\downarrow_{\neq\emptyset} \{m_1, m_2\} \qquad \overline{[\![\text{cat}]\!]} =\downarrow_{\neq\emptyset} \{c_1\} \qquad \overline{[\![\text{sneeze}]\!]} =\downarrow_{\neq\emptyset} \{m_1, c_1\}$$

So we have:

$$\overline{[\![\text{men}]\!]} = \{\{m_1\}, \{m_2\}, \{m_1, m_2\}\} \qquad \overline{[\![\text{cat}]\!]} = \{\{c_1\}\} \qquad \overline{[\![\text{sneeze}]\!]} = \{\{m_1\}, \{c_1\}, \{m_1, c_1\}\}$$

For quantifiers, we set:

$$\overline{[\![d]\!]}(\overline{[\![w]\!]}) := [\![d]\!]([\![w]\!])$$

As examples of quantified phrases we have:

$$\mathit{Some}\Big([\![\text{men}]\!]\Big) = \{\{m_1\}, \{m_2\}, \{m_1, m_2\}\} \qquad \mathit{All}\Big([\![\text{man}]\!]\Big) = \{\{m_1, m_2\}\} \qquad \mathit{No}\Big([\![\text{man}]\!]\Big) = \{\emptyset\}$$

The goal is to compute the meaning of "some men sneeze". In the first step of computation we obtain (the subscripts are dropped, they are always $\mathcal{PP}(U)$):

$$(\Delta \otimes 1)\Big(\overline{[\![\text{men}]\!]} \otimes \overline{[\![\text{sneeze}]\!]}\Big) = \\ \{(\{m_1\}, \{m_1\}), (\{m_2\}, \{m_2\}), (\{m_1, m_2\}, \{m_1, m_2\})\} \otimes \{\{m_1\}, \{c_1\}, \{m_1, c_1\}\}$$

In the second step, we obtain:

$$\Big(\overline{[\![\mathit{Some}]\!]} \otimes \mu\Big)\Big(\{(\{m_1\}, \{m_1\}), (\{m_2\}, \{m_2\}), (\{m_1, m_2\}, \{m_1, m_2\})\} \otimes \{\{m_1\}, \{c_1\}, \{m_1, c_1\}\}\Big) = \\ \overline{[\![\mathit{Some}]\!]}\Big(\{\{m_1\}, \{m_2\}, \{m_1, m_2\}\}\Big) \otimes \mu\Big(\{\{m_1\}, \{m_2\}, \{m_1, m_2\}\} \otimes \{\{m_1\}, \{c_1\}, \{m_1, c_1\}\}\Big) = \\ [\![\mathit{Some}]\!]\Big(\{m_1, m_2\}\Big) \otimes \{\{m_1\}\} = \\ \{\{m_1\}, \{m_2\}, \{m_1, m_2\}\} \otimes \{\{m_1\}\}$$

In the last step, we obtain the following:

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

Hence, the meaning of the sentence is true. For the sentence “all men sneeze”, one applies $(\overline{[\![all]\!]}\otimes\mu)$ to the result of the first step as above. The second and third steps of computation are as follows:

$$\epsilon\Big(\overline{[\![all]\!]}\Big(\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\Big)\otimes\mu\Big(\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\otimes\{\{m_1\},\{c_1\},\{m_1,c_1\}\}\Big)\Big)=$$
$$\epsilon\Big(\{\{m_1,m_2\}\}\otimes\{\{m_1\}\}\Big)=\emptyset$$

So the meaning of this sentence is false. For the sentence of ‘No men sneeze’, we have the following:

$$\epsilon\Big(\overline{[\![no]\!]}\Big(\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\Big)\otimes\mu\Big(\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\otimes\{\{m_1\},\{c_1\},\{m_1,c_1\}\}\Big)\Big)=$$
$$\epsilon\Big(\{\emptyset\}\otimes\{\{m_1\}\}\Big)=\emptyset$$

Example (II): Transitive Verb. Suppose both of the male individuals love the cat. We set:

$$\overline{[\![\text{love}]\!]}:=\downarrow_{\neq\emptyset}Dom(\text{love})\times\downarrow_{\neq\emptyset}Codom(\text{love})$$

Hence we have:

$$\overline{[\![\text{love}]\!]}:=\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\times\{\{c_1\}\}$$

Meaning of the sentence ‘Some men love cats’ is computed as follows. In the first step we compute:

$$\begin{aligned}(\sigma\otimes 1\otimes\epsilon)(\overline{[\![\text{men}]\!]}\otimes\overline{[\![\text{love}]\!]}\otimes\overline{[\![\text{cats}]\!]})&=\\ \{(\{m_1\},\{m_1\}),(\{m_2\},\{m_2\}),(\{m_1,m_2\},\{m_1,m_2\})\}\otimes\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\otimes\epsilon(\{\{c_1\}\}\otimes\{\{c_1\}\})&=\\ \{(\{m_1\},\{m_1\}),(\{m_2\},\{m_2\}),(\{m_1,m_2\},\{m_1,m_2\})\}\otimes\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\otimes\{\star\}&=\\ \{(\{m_1\},\{m_1\}),(\{m_2\},\{m_2\}),(\{m_1,m_2\},\{m_1,m_2\})\}\otimes\{\{m_1\},\{m_2\},\{m_1,m_2\}\}&\end{aligned}$$

In the second step, we apply $\overline{[\![Some]\!]}\otimes\mu$ to the above and compute:

$$\begin{aligned}\overline{[\![Some]\!]}(\{\{m_1\},\{m_2\},\{m_1,m_2\}\})\otimes\mu(\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\otimes\{\{m_1\},\{m_2\},\{m_1,m_2\}\})&=\\ [\![Some]\!](\{m_1,m_2\})\otimes\{\{m_1\},\{m_2\},\{m_1,m_2\}\}&=\\ \{\{m_1\},\{m_2\},\{m_1,m_2\}\}\otimes\{\{m_1\},\{m_2\},\{m_1,m_2\}\}&\end{aligned}$$

In the final step, we apply ϵ to the above and compute:

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

So the sentence is true. For “all men love cats”, the first step of the computation is as above. For the second and third steps we compute:

$$\epsilon\Big(\overline{[\![all]\!]}(\{\{m_1\},\{m_2\},\{m_1,m_2\}\})\otimes\mu(\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\otimes\{\{m_1\},\{m_2\},\{m_1,m_2\}\})\Big)=$$
$$\epsilon(\{\{m_1,m_2\}\}\otimes\{\{m_1\},\{m_2\},\{m_1,m_2\}\})=\{\star\}$$

So this sentence is also true. Whereas for “no men love cats” we have:

$$\epsilon\Big(\overline{[\![no]\!]}\big(\{\{m_1\},\{m_2\},\{m_1,m_2\}\}\big) \otimes \mu\big(\{\{m_1\},\{m_2\},\{m_1,m_2\}\} \otimes \{\{m_1\},\{m_2\},\{m_1,m_2\}\}\big)\Big) = \epsilon(\{\emptyset\} \otimes \{\{m_1\},\{m_2\},\{m_1,m_2\}\}) = \emptyset$$

So the meaning of this sentence is also false.

5 Justification

Recalling that, as shown in [**?**,**?**], the Frobenius μ map is the analog of set-theoretic intersection and the compact closed epsilon map is the analog of set-theoretic application, it is not hard to show that the truth-theoretic interpretation of the compact closed semantics of quantified sentences provides us with the same meaning as their generalised quantifier semantics. In what follows, we make this formal as follows.

Definition 2. *The meaning of a sentence in a concrete relational instantiation of the compact closed categorical interpretation is true iff* $\epsilon_{\mathcal{P}\mathcal{P}(U)}(\overline{[\![np]\!]} \otimes \overline{[\![vp]\!]}) = \{\star\}$ *and is false otherwise.*

Proof. That the interpretation of atoms is ??? and that the rules preserve truth?

Lemma 1. *Using the non-empty down set embedding, the meaning of a sentence with a quantified phrase at its subject position becomes equivalent to the following*

$$\{\star \mid D_k = A_i = B_j, D_k \in [\![d]\!]([\![N]\!]), A_i \in \mathcal{P}_{\neq\emptyset}([\![N]\!]), B_j \in \mathcal{P}_{\neq\emptyset}([\![VP]\!])\}$$

Proof. The non-empty down set embedding means that we have the following:

$$\overline{[\![n]\!]} = \{A_i \mid A_i \in \mathcal{P}_{\neq\emptyset}([\![n]\!])\} \qquad \overline{[\![d]\!]}\Big(\overline{[\![n]\!]}\Big) = \{D_o \mid D_o \in [\![d]\!]([\![n]\!])\}$$
$$\overline{[\![np]\!]} = \{C_l \mid C_l \in \mathcal{P}_{\neq\emptyset}([\![np]\!])\} \qquad \overline{[\![v]\!]} = \{(B_j, B_k) \mid B_j, B_k \in \mathcal{P}_{\neq\emptyset}([\![v]\!])\}$$

The meaning of a sentence with a quantified subject is computed in three steps. In the first step, we obtain:

$$(\varDelta_N \otimes 1_N)\Big(\overline{[\![\mathrm{N}]\!]} \otimes \overline{[\![\mathrm{VP}]\!]}\Big) = \{(A_i, A_i) \mid A_i \in \mathcal{P}([\![\mathrm{N}]\!])\} \otimes \{B_j \mid B_j \in \mathcal{P}_{\neq\emptyset}([\![\mathrm{VP}]\!])\}$$

In the second step, we obtain:

$$\begin{aligned}(Det \otimes \mu_N)\Big(\{(A_i, A_i) \mid A_i \in \mathcal{P}_{\neq\emptyset}([\![\mathrm{Sbj}]\!])\} \otimes \{B_j \mid B_j \in \mathcal{P}([\![\mathrm{Verb}]\!])\}\Big) &= \\ Det\Big(\{A_i \mid A_i \in \mathcal{P}_{\neq\emptyset}([\![\mathrm{Sbj}]\!])\}\Big) \otimes \{A_i \mid A_i = B_j, A_i \in \mathcal{P}([\![\mathrm{Sbj}]\!]), B_j \in \mathcal{P}([\![\mathrm{Verb}]\!])\} &= \\ \{D_k \mid D_k \in Det([\![\mathrm{N}]\!])\} \otimes \{A_i \mid A_i = B_j, A_i \in \mathcal{P}_{\neq\emptyset}([\![\mathrm{N}]\!]), B_j \in \mathcal{P}([\![\mathrm{VP}]\!])\}\end{aligned}$$

In the final step, we obtain:

$$\epsilon\Big(\{D_k \mid D_k \in Det([\![\mathrm{Sbj}]\!])\} \otimes \{A_i \mid A_i = B_j, A_i \in \mathcal{P}([\![\mathrm{Sbj}]\!]), B_j \in \mathcal{P}([\![\mathrm{Verb}]\!])\}\Big) = \{\star \mid D_k = A_i, D_k \in Det([\![\mathrm{N}]\!]), A_i = B_j, A_i \in \mathcal{P}_{\neq\emptyset}([\![\mathrm{N}]\!]), B_j \in \mathcal{P}([\![\mathrm{VP}]\!])\}$$

Definition 3. *The compact closed meaning of the sentence "Det N VP" is true in Rel if and only if*

$$\epsilon \circ (Det \otimes \mu) \circ (\sigma \otimes 1_N)\left(\overline{[\![N]\!]} \otimes \overline{[\![VP]\!]}\right) = \{\star\}$$

Proposition 1. *The compact closed meaning of a quantified sentence computed in Rel is true if and only if its generalised quantifier meaning is true.*

Proof. For the right to left direction, suppose the generalised quantifier meaning of the sentence is true, that is $[\![\text{Sbj}]\!] \cap [\![\text{Verb}]\!] \in Det([\![\text{Sbj}]\!]$ and consider the case when $Det([\![Sbj]\!]) \neq \{\emptyset\}$. We have to show that $\epsilon \circ (Det \otimes \mu) \circ (\sigma \otimes 1_N)\left(\overline{[\![\text{Sbj}]\!]} \otimes \overline{[\![\text{Verb}]\!]}\right) = \{\star\}$. For this, we need to show that there is a set G equal to $D_k = A_i = B_j$ such that G is in $Det([\![\text{Sbj}]\!])$ and $\mathcal{P}([\![\text{Sbj}]\!])$ and $\mathcal{P}([\![\text{Verb}]\!])$. Take G to be $[\![\text{Sbj}]\!] \cap [\![\text{Verb}]\!]$. Then, it is a subset of $[\![\text{Sbj}]\!]$, hence an element of $\mathcal{P}([\![\text{Sbj}]\!])$, a subset of $[\![\text{Verb}]\!]$, hence an element of $\mathcal{P}([\![\text{Verb}]\!])$, and an element of $Det([\![\text{Sbj}]\!]$.

For the left to right direction, suppose the compact closed meaning of the sentence is true in Rel, and consider the case when $Det([\![\text{S}bj]\!]) \neq \{\emptyset\}$. Then we have subsets $D_k = A_i = B_j$ such that $D_k \in Det([\![\text{S}bj]\!]), A_i \in \mathcal{P}([\![\text{S}bj]\!])$, and $B_j \in \mathcal{P}([\![\text{V}erb]\!])$. Pick an arbitrary such subset, e.g. G, such that it equal to $D_k = A_i = B_j$; we have that $G \in \mathcal{P}([\![\text{S}bj]\!])$, hence $G \subseteq [\![\text{S}bj]\!]$; and that $G \in \mathcal{P}([\![\text{V}erb]\!])$, hence $G \subseteq [\![\text{V}erb]\!]$. It thus follows that $G \subseteq [\![\text{S}bj]\!] \cap [\![\text{V}erb]\!]$. At the same time $G \in Det([\![\text{S}bj]\!])$, hence the generalised quantifier meaning is also true.

6 Vector Space Interpretation

Given the set-theoretical model $(U, [\![\ ]\!])$ of a language $\mathcal{L}_\Sigma$, a vector instantiation of the abstract compact closed categorical interpretation is provided by the tuple $(\mathcal{C}, W, S, \rightarrow)$, where W is a vector space with a basis $\{n_i\}_i$ and S is a vector space with a basis $\{s_j\}_j$. Vector meanings of words are as follows (the corresponding morphisms are obtained as before):

- For a word w with a lexical category N, NP we have

$$\overrightarrow{\text{w}} \colon I \to W \qquad \text{given by} \quad \sum_i C_i \overrightarrow{n}_i$$

- For words w with lexical category VP, we have

$$\overrightarrow{\text{w}} \colon I \to W \otimes S \qquad \text{given by} \quad \sum_{ij} C_{ij} \overrightarrow{n}_i \otimes \overrightarrow{s}_j$$

- For words w with lexical category V, we have

$$\overrightarrow{\text{w}} \colon I \to W \otimes S \otimes W \qquad \text{given by} \quad \sum_{ijk} C_{ijk} \overrightarrow{n}_i \otimes \overrightarrow{s}_j \otimes \overrightarrow{n}_k$$

- For a word d with the lexical category Det and w a word with the lexical category N, we have

$$d \colon W \to W \qquad \text{given by} \quad d(\overrightarrow{\text{w}}) = \sum_o C_o \overrightarrow{n}_o \quad \text{for} \quad \overrightarrow{\text{w}} \in W$$

Supposing $\overrightarrow{\text{w}} = \sum_i C_i \overrightarrow{n}_i$, by linearity of d, from the above it follows that $d(\overrightarrow{\text{w}}) = d(\sum_i C_i \overrightarrow{n}_i) = \sum_i C_i d(\overrightarrow{n}_i)$. If we take $d(\overrightarrow{n}_i) = \sum_t C_t^i \overrightarrow{n}_t$, we obtain that $\sum_{it} C_i C_t^i \overrightarrow{n}_t = \sum_o C_o \overrightarrow{n}_o$.

In this instantiation, the meaning of a sentence with a quantified subject is obtained by computing the following:

$$(\epsilon_W \otimes 1_S) \circ (d \otimes \mu_W \otimes 1_S) \circ (\Delta_W \otimes 1_{W \otimes S} \otimes \epsilon_W) \circ (\overrightarrow{n} \otimes \overrightarrow{v} \otimes \overrightarrow{np})$$

Setting $\overrightarrow{n} = \sum_l C_l \overrightarrow{n}_l$, $\overrightarrow{v} = \sum_{ijk} C_{ijk} \overrightarrow{n}_i \otimes \overrightarrow{s}_j \otimes \overrightarrow{n}_k$, and $\overrightarrow{np} = \sum_r C_r \overrightarrow{n}_r$, and unfolding the morphisms, in the first step of the computation we obtain the following (where δ_{rk} is 1 when $r = k$ and 0 otherwise):

$$\begin{aligned}
(\Delta_W \otimes 1_{W \otimes S} \circ \epsilon_W)\Big(\sum_l C_l \overrightarrow{n}_l \otimes \sum_{ijk} C_{ijk} \overrightarrow{n}_i \otimes \overrightarrow{s}_j \otimes \overrightarrow{n}_k \ \otimes \sum_r C_r \overrightarrow{n}_r\Big) &= \\
(\sum_l C_l \overrightarrow{n}_l \otimes \overrightarrow{n}_l) \otimes (\sum_{ijkr} C_{ijk} C_r \overrightarrow{n}_i \otimes \overrightarrow{s}_j \otimes \langle \overrightarrow{n}_k \mid \overrightarrow{n}_r \rangle) &= \\
(\sum_l C_l \overrightarrow{n}_l \otimes \overrightarrow{n}_l) \otimes (\sum_{ijkr} C_{ijk} C_r \overrightarrow{n}_i \otimes \overrightarrow{s}_j \delta_{rk})&
\end{aligned}$$

In the second step we obtain:

$$\begin{aligned}
(d \otimes \mu_W \otimes 1_S)\Big((\sum_l C_l \overrightarrow{n}_l \otimes \overrightarrow{n}_l) \otimes (\sum_{ijkr} C_{ijk} C_r \overrightarrow{n}_i \otimes \overrightarrow{s}_j \delta_{rk})\Big) &= \\
\sum_{ijkrl} C_{ijk} C_r C_l d(\overrightarrow{n_l}) \otimes \mu(\overrightarrow{n}_l \otimes \overrightarrow{n}_i) \otimes \overrightarrow{s}_j \delta_{rk} &= \\
\sum_{ijkrl} C_{ijk} C_r C_l d(\overrightarrow{n_l}) \otimes \delta_{li} \overrightarrow{n}_i \otimes \overrightarrow{s}_j \delta_{rk}&
\end{aligned}$$

The final step is as follows:

$$(\epsilon_W \otimes 1_S)\Big(\sum_{ijkrl} C_{ijk} C_r C_l d(\overrightarrow{n_l}) \otimes \delta_{li} \overrightarrow{n}_i \otimes \overrightarrow{s}_j \delta_{rk}\Big) = \sum_{ijkrl} C_{ijk} C_r C_l \langle d(\overrightarrow{n_l}) \mid \overrightarrow{n}_i \rangle \delta_{li} \overrightarrow{s}_j \delta_{rk}$$

Now if we instantiate $d(\overrightarrow{n}_l)$ to $\sum_t C_t^i \overrightarrow{n}_t$, the above further simplifies to the following:

$$\sum_{ijkrl} C_{ijk} C_r C_l \langle \sum_t C_t^i \overrightarrow{n}_t \mid \overrightarrow{n}_i \rangle \delta_{li} \overrightarrow{s}_j \delta_{rk} = \sum_{ijkrlt} C_{ijk} C_r C_l C_t^i \delta_{ti} \delta_{li} \overrightarrow{s}_j \delta_{rk}$$

Similar computations provide us with the following for the meaning of a sentence with a quantified object:

$$(1_S \otimes \epsilon_W) \circ (1_S \otimes \mu_W \otimes d) \circ (\epsilon_W \otimes 1_{S \otimes W} \otimes \Delta_W)(\overrightarrow{n} \otimes \overrightarrow{v} \otimes \overrightarrow{np}) = \sum_{ijkrlt} C_{ijk} C_r C_l C_t^i \delta_{ri} \overrightarrow{s}_j \delta_{kl} \delta_{kt}$$

Intransitive Example. As an example, take W to be the two dimensional space with the basis $\{\overrightarrow{n}_1, \overrightarrow{n}_2\}$ and S to be the two dimensional space with the basis $\{\overrightarrow{s}_1, \overrightarrow{s}_2\}$. Consider an intransitive sentence with a quantified subject and the following linear expansions for its subject and verb:

$$\overrightarrow{np} := C_1 \overrightarrow{n}_1 + C_2 \overrightarrow{n}_2 \qquad \overrightarrow{vp} := C_{11} \overrightarrow{n}_1 \otimes \overrightarrow{s}_1 + C_{12} \overrightarrow{n}_1 \otimes \overrightarrow{s}_2 + C_{21} \overrightarrow{n}_2 \otimes \overrightarrow{s}_1 + C_{22} \overrightarrow{n}_2 \otimes \overrightarrow{s}_2$$

Suppose further the following for the interpretation of the determiner:

$$d(\overrightarrow{np}) = d(C_1\overrightarrow{n}_1 + C_2\overrightarrow{n}_2) = C_1 d(\overrightarrow{n}_1) + C_2 d(\overrightarrow{n}_2) = C_1'\overrightarrow{n}_1 + C_2'\overrightarrow{n}_2$$

where we further assume the following for the effect of d on each basis:

$$d(\overrightarrow{n}_1) = C_1^1\overrightarrow{n}_1 + C_2^1\overrightarrow{n}_2 \qquad d(\overrightarrow{n}_2) = C_1^2\overrightarrow{n}_1 + C_2^2\overrightarrow{n}_2$$

So we obtain the following equivalence between the application of d on words and on basis vectors:

$$d(\overrightarrow{np}) = C_1'\overrightarrow{n}_1 + C_2'\overrightarrow{n}_2 \;=\; (C_1C_1^1 + C_2C_1^2)\overrightarrow{n}_1 + (C_1C_2^1 + C_2C_2^2)\overrightarrow{n}_2$$

In the first step of the computation of the meaning vector of the sentence 'Q np vp' we have:

$$(\Delta_W \otimes 1_{W\otimes S})\Big((C_1\overrightarrow{n}_1 + C_2\overrightarrow{n}_2) \otimes (C_{11}\overrightarrow{n}_1 \otimes \overrightarrow{s}_1 + C_{12}\overrightarrow{n}_1 \otimes \overrightarrow{s}_2 + C_{21}\overrightarrow{n}_2 \otimes \overrightarrow{s}_1 + C_{22}\overrightarrow{n}_2 \otimes \overrightarrow{s}_2)\Big) =$$
$$(C_1\overrightarrow{n}_1 \otimes \overrightarrow{n}_1 + C_2\overrightarrow{n}_2 \otimes \overrightarrow{n}_2) \otimes (C_{11}\overrightarrow{n}_1 \otimes \overrightarrow{s}_1 + C_{12}\overrightarrow{n}_1 \otimes \overrightarrow{s}_2 + C_{21}\overrightarrow{n}_2 \otimes \overrightarrow{s}_1 + C_{22}\overrightarrow{n}_2 \otimes \overrightarrow{s}_2)$$

In the second step of computation, we apply $(d \otimes \mu_W \otimes 1_S)$ to the above and obtain:

$$(C_1 d(\overrightarrow{n}_1) + C_2 d(\overrightarrow{n}_2)) \otimes (C_1C_{11}\overrightarrow{n}_1 \otimes \overrightarrow{s}_1 + C_1C_{12}\overrightarrow{n}_1 \otimes \overrightarrow{s}_2 + C_2C_{21}\overrightarrow{n}_2 \otimes \overrightarrow{s}_1 + C_2C_{22}\overrightarrow{n}_2 \otimes \overrightarrow{s}_2) \qquad (*)$$

In the final step, we apply $(\epsilon_W \otimes 1_S)$ to the above and obtain:

$$C_1C_1C_{11}\langle d(\overrightarrow{n}_1) \mid \overrightarrow{n}_1\rangle \overrightarrow{s}_1 + C_1C_1C_{12}\langle d(\overrightarrow{n}_1) \mid \overrightarrow{n}_1\rangle \overrightarrow{s}_2 + C_1C_2C_{21}\langle d(\overrightarrow{n}_1) \mid \overrightarrow{n}_2\rangle \overrightarrow{s}_1 + C_1C_2C_{22}\langle d(\overrightarrow{n}_1) \mid \overrightarrow{n}_2\rangle \overrightarrow{s}_2$$
$$+$$
$$C_2C_1C_{11}\langle d(\overrightarrow{n}_2) \mid \overrightarrow{n}_1\rangle \overrightarrow{s}_1 + C_2C_1C_{12}\langle d(\overrightarrow{n}_2) \mid \overrightarrow{n}_1\rangle \overrightarrow{s}_2 + C_2C_2C_{21}\langle d(\overrightarrow{n}_2) \mid \overrightarrow{n}_2\rangle \overrightarrow{s}_1 + C_2C_2C_{22}\langle d(\overrightarrow{n}_2) \mid \overrightarrow{n}_2\rangle \overrightarrow{s}_2$$

Now, using the linear expansion of d, the above is further simplified to the following:

$$C_1C_1C_{11}C_1^1\overrightarrow{s}_1 + C_1C_1C_{12}C_1^1\overrightarrow{s}_2 + C_1C_2C_{21}C_2^1\overrightarrow{s}_1 + C_1C_2C_{22}C_2^1\overrightarrow{s}_2$$
$$+$$
$$C_2C_1C_{11}C_1^2\overrightarrow{s}_1 + C_2C_1C_{12}C_1^2\overrightarrow{s}_2 + C_2C_2C_{21}C_2^2\overrightarrow{s}_1 + C_2C_2C_{22}C_2^2\overrightarrow{s}_2$$

To obtain a more readable notation, instead of expanding, which is what we have been doing so far, let us factor things out a bit. First note that the $(*)$ above is equivalent to the following by linearity of the map d:

$$d(C_1\overrightarrow{n}_1 + C_2\overrightarrow{n}_2) \otimes (C_1C_{11}\overrightarrow{n}_1 \otimes \overrightarrow{s}_1 + C_1C_{12}\overrightarrow{n}_1 \otimes \overrightarrow{s}_2 + C_2C_{21}\overrightarrow{n}_2 \otimes \overrightarrow{s}_1 + C_2C_{22}\overrightarrow{n}_2 \otimes \overrightarrow{s}_2)$$

Then, using a matrix notation where we assume column vectors are elements of W and 2 by 2 matrices elements of $W \otimes S$, observe that the above can be written down as follows:

$$\left(\begin{pmatrix} C_1 & C_1 \\ C_2 & C_2 \end{pmatrix} \odot \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}\right) \quad \times \quad d\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} \qquad (**)$$

Here the map d is being applied to the vector meaning of the word np rather to the basis vectors of the vector space W. In the preceding computations, the map d was being applied to the basis vectors and the result of the final step of the computation was expressed in that form. A routine computations shows that the equivalence between the above assumptions on the applications of the d, on words or basis vectors, in linear expansion form has the following matrix form:

$$d\begin{pmatrix}C_1\\C_2\end{pmatrix} = \begin{pmatrix}C'_1\\C'_2\end{pmatrix} = \begin{pmatrix}C_1C_1^1 + C_2C_1^2\\C_1C_2^1 + C_2C_2^2\end{pmatrix}$$

This can then be replaces and used in the $(**)$ formula.

Transitive Example. Consider the same two dimensional W and S spaces for the case of transitive sentences and the same assumption for the vector meaning of the subject $\overrightarrow{n}$. Assume further that for the vector of the object we have $\overrightarrow{np} = C'_1\overrightarrow{n}_1 + C'_2\overrightarrow{n}_2$. In this case, because the verb is an element of a rank 3 tensor space, that is $\overrightarrow{v} \in W \otimes S \otimes W$, it is not possible to express it as a matrix in two dimensions: indeed elements of rank 3 tensor spaces are cubes rather than matrices. But opting for not expanding the corresponding tensor and just denoting the cube of the verb by $\overrightarrow{v}$, the meaning vector of the transitive sentence with a quantified subject can be expressed as follows

$$\left(\begin{pmatrix}C_1\ C_1\\C_2\ C_2\end{pmatrix} \odot \epsilon(\overrightarrow{v} \otimes \begin{pmatrix}C'_1\\C'_2\end{pmatrix}))\right) \quad \times \quad d\begin{pmatrix}C_1\\C_2\end{pmatrix}$$

The details of the tedious computations are as before, with the exception that in this case, first the verb has to be applied to its object. This is denoted by the application of the ϵ map to $\overrightarrow{v}$ and matrix form of $\overrightarrow{np}$. The rest of the computation is as before: one takes the point wise multiplication of this result with a form of diagonalisation of the vector of the subject, then applies this to the effect of the quantifier map d on the vector of the subject. The result is a sentence vectors in S.

7 A Concrete Corpus Based Instantiation

So far we have not said anything about the concrete vector space content of the d maps, we just supposed they are linear, that they input a vector from W and output another in W, and given this information we computed what the abstract meaning of a sentence with a quantifier is. In this section, we provide concrete choices for these maps and present some toy experimental results.

Given a distributional model of word meaning $(V_{l[\Sigma]}, \to)$, we instantiate the vector space model as $(\mathcal{C}, V_{\mathcal{P}(l[\Sigma])}, V_{\mathcal{P}(l[\Sigma])}, \to)$. Note two things: first, the choice of a sentence space has been a bottle neck for the compositional distributional models since the start. Based on previous work [**?**], here, we are taking the two designated spaces to be the same; that is, we are taking the sentence space to be the same as the distributional space. The reason is that, as shown in [**?**] and [**?**], this provides better experimental results and the original concrete models of [**?**] can be encoded in these models using the Frobenius Δ map. Secondly, the basis vectors of this single vector space are sets of lemmas, rather than just lemmas. We denote these basis vectors by $V_{\mathcal{P}(l[\Sigma])} = \{\overrightarrow{A}_i\}_i$.

The novel concrete questions of this paper are two: how can one build concrete vectors in $V_{\mathcal{P}(l[\Sigma])}$, and are the d maps defined. With regard to the first question we would like to recall that this is not a very hard task, as the distributional spaces can already be seen as being such vector spaces. In $V_{l[\Sigma]}$, the basis is the set of lemmas, hence each basis vector $\overrightarrow{v}_i$ represents a lemma. One can immediately take the converse of the l map and replace the lemma with the set of words that have that lemma, that is with $l^{-1}[v_i]$. In this case, the vector meanings of words are as described in section 2.1. However, one can also work with sets of lemmas as basis vectors. This is common practice in advanced distributional models, where sets of lemmas are referred to by *clusters*, Here, each basis vector corresponds to a set of lemmas grouped together into clusters. The grouping criterion depends on the clustering scheme used; the most famous scheme is known as k-means. This scheme pre requires the number of clusters, that is the fixed number k; a lemma is grouped into a cluster based on its closeness to the mean vector of that cluster. A more advanced method is hierarchical agglomerative clustering (HAC) which is an iterative bottom-up approach and which does not rely on pre-setting the number of clusters.

In either case, the concrete vectors are built as follows:

- For a word w with a lexical category N, NP we have $\overrightarrow{\mathrm{w}} = \sum_i C_i \overrightarrow{A}_i$, where C_i is the weight associated to word w in the cluster A_i.
- For words w with lexical category VP, we have $\overrightarrow{\mathrm{w}} = \Delta(\sum_{ij} C_{ij} \overrightarrow{A}_i)$; following previous work [?] the term inside the Frobenius Δ is the sum of the vector meanings of the subjects of w across the corpus. This is a vector in $V_{\mathcal{P}(l[\Sigma])}$; the Δ map turns it into a matrix in $V_{\mathcal{P}(l[\Sigma])} \otimes V_{\mathcal{P}(l[\Sigma])}$.
- For words w with lexical category V, we have $\overrightarrow{\mathrm{w}} = \Delta(\sum_{ijk} C_{ijk} \overrightarrow{A}_i \otimes \overrightarrow{A}_k)$; again following previous work [?] the term inside Δ is the sum of the tensor products of vector meanings of the subjects and objects of w across the corpus; the Δ map turns this into a cube in $V_{\mathcal{P}(l[\Sigma])} \otimes V_{\mathcal{P}(l[\Sigma])} \otimes V_{\mathcal{P}(l[\Sigma])}$.
- For a word d with the lexical category Det, we have $d\colon V_{\mathcal{P}(l[\Sigma])} to V_{\mathcal{P}(l[\Sigma])}$; this map assigns to a word w with the lexical category N, a vector $d(\overrightarrow{\mathrm{w}}) = \sum_o C_o \overrightarrow{A}_o$, where C_o is the degree to which extent (e.g. the number of times) w has co-occurred with d elements of $\overrightarrow{A}_o$. Examples are as follows:

$all(w)$	C_o is the degree of statistical association between w and *all* elements of $\overrightarrow{A}_o$
$some(w)$	C_o is the degree of statistical association between w and *some* elements of $\overrightarrow{A}_o$
$no(w)$	C_o is the degree of statistical association between w and *no* elements of $\overrightarrow{A}_o$
$most(w)$	C_o is the degree of statistical association between w and *most* elements of $\overrightarrow{A}_o$
$few(w)$	C_o is the degree of statistical association between w and *few* elements of $\overrightarrow{A}_o$

7.1 Toy Experiment

Entailment in vector space models has been a tricky issue. For a long time researchers feared that it would not be possible to model it, e.g. see [?], since the most natural relationship between the two vectors was that of similarity which is two directional. So one cannot say that if $\overrightarrow{v}$ is similar to $\overrightarrow{w}$ then $\overrightarrow{v}$ entails $\overrightarrow{w}$, since this would immediately result in $\overrightarrow{w}$ also entailing $\overrightarrow{v}$. But later the work of Dagan et al and Geffet and Dagan showed changed the matter. They argued that instead of similarity, one can use *context inclusion* as a measure of entailment, this being a perfectly well one-directional relation. So one says $\overrightarrow{v}$ entails $\overrightarrow{w}$ if contexts of the word v are also context of the word w.

This notion was later more refined and *features* were used instead of contexts. Features are triples of the form ⟨ context, syntactic-relation, direction ⟩; it consists of a context, the syntactic relation between the word and the context, and a direction for this syntactic relation. An example is the feature ⟨ profit, pcomp, 0 ⟩ for the word ‘company’, it encodes the information in the phrase ‘the profit of the company’, that in this phrase ‘profit’ is the prepositional complement of ‘company’, and that profit is not modifying company. The inclusion relation was also refined and now read as follows

$$\overrightarrow{v} \vdash \overrightarrow{w} \qquad \text{iff} \qquad \text{all characteristic features of v are included in all features of w}$$

The features were harvested from the Minipar dependancy parser of Lin. Various weighting schemes such as MI and RFF were used to build vectors for words with features as their basis vectors. These measures were combined with web-based feature collection algorithms and manual judgement to produce a dataset of word pairs, where either one word entailed the other or not. Here are a few examples from the top-40 pairs:

government $\Longrightarrow$ body	bill $\Longrightarrow$ program	war $\Longrightarrow$ conflict
mortgage $\Longrightarrow$ loan	town $\Longrightarrow$ location	

So our starting point is these pairs. To these we are going to add quantifiers using their monotonicity properties and form quantified entailment pairs. Following Baroni et al., examples of these properties are as follows:

all $\vdash$ some	all $\vdash$ several	all $\vdash$ many
much $\vdash$ some	many $\vdash$ some	several $\vdash$ some
some $\nvdash$ all	several $\nvdash$ all	few $\nvdash$ all
some $\nvdash$ many	few $\nvdash$ many	

Baroni et al then use Kotlerman's *balAPinc*, which is a version of the average of precision and a support vector machine classifier SVM to evaluate the correctness of the resulting entailment pairs. We will do the same.

We provide a couple of corpus-based witnesses here. In the distributional models. A sample query from the online *Reuter News Corpus*, with at most 100 outputs per query, provides the following instantiations:

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

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 [**?**]). 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 **??** tells us that these sentences have different meanings, since their quantified subjects have different meanings. To see this, take $\overrightarrow{\text{cats}} = C_1\overrightarrow{n}_1 + C_2\overrightarrow{n}_2$, where as $few(\text{cats}) = C_1'\overrightarrow{n}_1 + C_2'\overrightarrow{n}_2$ and $most(\text{cats}) = C_1''\overrightarrow{n}_1 + C_2''\overrightarrow{n}_2$. Instantiating these in equation **??** provides us with the following three different vectors:

$$\overrightarrow{\text{cats sleep}} = C_1C_{11}\overrightarrow{s}_1 + C_1C_{12}\overrightarrow{s}_2 + C_2C_{21}\overrightarrow{s}_2 + C_2C_{22}\overrightarrow{s}_2$$
$$\overrightarrow{\text{few cats sleep}} = C_1'C_{11}\overrightarrow{s}_1 + C_1'C_{12}\overrightarrow{s}_2 + C_2'C_{21}\overrightarrow{s}_2 + C_2'C_{22}\overrightarrow{s}_2$$
$$\overrightarrow{\text{most cats sleep}} = C_1''C_{11}\overrightarrow{s}_1 + C_1''C_{12}\overrightarrow{s}_2 + C_2''C_{21}\overrightarrow{s}_2 + C_2''C_{22}\overrightarrow{s}_2$$

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:

$$\begin{cases} most(\overrightarrow{\text{dogs}}) \sim most(\overrightarrow{\text{cats}}) \sim \overrightarrow{\text{pets}} \\ \overrightarrow{\text{snooze}} \sim \overrightarrow{\text{sleep}} \end{cases} \implies \text{most cats sleep} \sim \text{most dogs snooze} \sim \text{pets sleep}$$

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:

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