

Reasoning about Meaning in Natural Language with Compact Closed Categories and Frobenius Algebras ^{*}

Authors: Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, Stephen Pulman, Bob Coecke [†]

Affiliation: Department of Computer Science, University of Oxford

Address: Wolfson Building, Parks Road, Oxford, OX1 3QD

Emails: `firstname.lastname@cs.ox.ac.uk`

Abstract

Compact closed categories have found applications in modeling quantum information protocols by Abramsky-Coecke. They also provide semantics for Lambek’s pregroup algebras, applied to formalizing the grammatical structure of natural language, and are implicit in a distributional model of word meaning based on vector spaces. Specifically, in previous work Coecke-Clark-Sadrzadeh used the product category of pregroups with vector spaces and provided a distributional model of meaning for sentences. We recast this theory in terms of strongly monoidal functors and advance it via Frobenius algebras over vector spaces. The former are used to formalize topological quantum field theories by Atiyah and Baez-Dolan, and the latter are used to model classical data in quantum protocols by Coecke-Pavlovic-Vicary. The Frobenius algebras enable us to work in a single space in which meanings of words, phrases, and sentences of any structure live. Hence we can compare meanings of different language constructs and enhance the applicability of the theory. We report on experimental results on a number of language tasks and verify the theoretical predictions.

1 Introduction

Compact closed categories were first introduced by Kelly [21] in early 1970’s. Some thirty years later they found applications in quantum mechanics [1], whereby the vector space foundations of quantum mechanics were recasted in a higher order language and quantum protocols such as *teleportation* found succinct conceptual proofs. Compact closed categories are complete with regard to a pictorial calculus [21, 38]; this calculus is used to depict and reason about information flows in entangled quantum states modeled in tensor spaces, the phenomena that were considered to be mysteries of quantum mechanics and the Achilles heel of quantum logic [5]. The pictorial calculus revealed the multi-linear algebraic level needed for proving quantum information protocols and simplified the reasoning thereof to a great extent, by hiding the underlying additive vector space structure.

Most quantum protocols rely on classical, as well as quantum, data flow. In the work of [1], this classical data flow was modeled using bi-products defined over a compact closed category. However, the pictorial calculus could not extend well to bi-products, and their categorical axiomatization was not as clear as the built-in monoidal tensor of the category. Later, Frobenius algebras, originally used in group theory [14] and later widely applied to other fields of

^{*}To appear in Chubb, J., Eskandarian, A. and Harizanov, V., editors, *Logic and Algebraic Structures in Quantum Computing and Information* (2013). Cambridge University Press.

[†]Support by EPSRC grant EP/F042728/1 is acknowledged by the first two authors.

mathematics and physics such as topological quantum field theory (TQFT) [2, 24, 3], proved useful. It turned out that the operations of such algebras on vector spaces with orthonormal basis correspond to a uniform copying and deleting of the basis, a property that only holds for, hence can be used to axiomatize, classical states [9].

Compact closed categories have also found applications in two completely orthogonal areas of computational linguistics: formalizing grammar and reasoning about lexical meanings of words. The former application is through Lambek’s pregroup grammars [26], which are compact closed categories [33] and have been applied to formalizing grammars of a wide range of natural languages, for instance see [27]. The other application domain, referred to as *distributional* models of meaning, formalizes meanings of words regardless of their grammatical roles and via the context of their occurrence [13]. These models consist of vector spaces whose basis are sets of context words and whose vectors represent meanings of target words. Distributional models have been widely studied and successfully applied to a variety of language tasks [37, 28, 30] and in particular to automatic word-synonymy detection [11].

Whereas the type-logical approaches to language do not provide a convincing model of word meaning, the distributional models do not scale to meanings of phrases and sentences. The long standing challenge of combining these two models was addressed in previous work [6, 10, 34]. The solution was based on a cartesian product of the pregroup category and the category of finite dimensional vector spaces. The theoretical predictions of the model were made concrete in [15], then implemented and verified in [16]. In this article, we first recast the theoretical setting of [10] using a succinct functorial passage from a free pregroup of basic types to the category of finite dimensional vector spaces. Then, we further advance the theory and show how Frobenius algebras over vector spaces provide solutions for the problem of the concrete construction of linear maps for predicative words with complex types. As a result, we are able to compare meanings of phrases and sentences with different structures, and moreover compare these with lexical vectors of words. This enhances the domain of application of our model: we show how the theoretical predictions of the model, and in particular the Frobenius algebraic constructions, can be empirically verified by performing three experiments: the disambiguation task of [16], comparing meanings of transitive and intransitive sentences, and a new term/definition classification task.

2 Recalling Some Categorical Definitions

We start by recalling some definitions. A *monoidal category* [21] is a category \mathcal{C} with a monoidal tensor \otimes , which is associative. That is, for all objects $A, B, C \in \mathcal{C}$, we have that $A \otimes (B \otimes C) \cong (A \otimes B) \otimes C$. Moreover there exists an object $I \in \mathcal{C}$, which serves as the unit of the tensor, that is, $A \otimes I \cong A \cong I \otimes A$. These isomorphisms need to satisfy the usual coherence conditions.

A monoidal category is called *symmetric* whenever we have $A \otimes B \cong B \otimes A$, again satisfying the standard conditions. Furthermore, a monoidal category is *compact closed* whenever any object $A \in \mathcal{C}$ has a left A^l and a right adjoint A^r , that is, the following morphisms exist:

$$\begin{aligned} \epsilon_A^r: A \otimes A^r &\rightarrow I & \eta_A^r: I &\rightarrow A^r \otimes A \\ \epsilon_A^l: A^l \otimes A &\rightarrow I & \eta_A^l: I &\rightarrow A \otimes A^l \end{aligned}$$

and they satisfy the following *yanking* conditions:

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

In a *symmetric compact closed* category, the left and right adjoints collapse into one, that is we have $A^* := A^l = A^r$ and the above four equalities collapse to the following two:

$$(\epsilon_A \otimes 1_A) \circ (1_A \otimes \eta_A) = 1_A \quad (1_{A^*} \otimes \epsilon_A) \circ (\eta_A \otimes 1_{A^*}) = 1_{A^*}$$

A functor \mathcal{F} from a monoidal category \mathcal{C} to a monoidal category \mathcal{D} is a *monoidal functor* [23], whenever \mathcal{F} is a functor and moreover there exists a morphism $I \rightarrow \mathcal{F}(I)$ and the following is a natural transformation:

$$\mathcal{F}(A) \otimes \mathcal{F}(B) \rightarrow \mathcal{F}(A \otimes B)$$

satisfying the corresponding coherence conditions. A monoidal functor is *strongly monoidal* [23], whenever the above morphism and natural transformation are invertible.

A strongly monoidal functor on two compact closed categories \mathcal{C} and \mathcal{D} preserves the compact structure, that is $\mathcal{F}(A^l) = \mathcal{F}(A)^l$ and $\mathcal{F}(A^r) = \mathcal{F}(A)^r$. To see this, consider the case of the left adjoint, for which we have the following two compositions of morphisms:

$$\begin{aligned} \mathcal{F}(A^l) \otimes \mathcal{F}(A) &\rightarrow \mathcal{F}(A^l \otimes A) \rightarrow \mathcal{F}(I) \rightarrow I \\ I &\rightarrow \mathcal{F}(I) \rightarrow \mathcal{F}(A \otimes A^l) \rightarrow \mathcal{F}(A) \otimes \mathcal{F}(A^l) \end{aligned}$$

From these, and since adjoints are unique, it follows that $\mathcal{F}(A^l)$ must be left adjoint to $\mathcal{F}(A)$. The case for the right adjoint is similar.

An example of a compact closed category is a Lambek pregroup [26], denoted by $(P, \leq, 1, \cdot, (-)^l, (-)^r)$; we refer to this category by **Preg**. This is a partially ordered monoid where each element of the partial order has a left and a right adjoint, that is we have the following inequalities, which are the partial order versions of the yanking conditions of a compact closed category:

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

An example of a pregroup is the set of all unbounded monotone functions on integers, with function composition as the monoidal tensor and the identity function as its unit. The left and right adjoints are defined using the standard definition of adjoints and in terms of the min and max operations on the integers as follows, for $f \in \mathbb{Z}^{\mathbb{Z}}$ and $m, n \in \mathbb{Z}$:

$$f^r(n) = \bigvee \{m \in \mathbb{Z} \mid f(m) \leq n\} \quad f^l(n) = \bigwedge \{m \in \mathbb{Z} \mid n \leq f(m)\}$$

An example of a symmetric compact closed category is the category of finite dimensional vector spaces and linear maps over a field (which for our purposes we take to be the set of real numbers \mathbb{R}); we refer to this category by **FVect**. The monoidal tensor is the tensor product of vector spaces whose unit is the field. The adjoint of each vector space is its dual, which, by fixing a basis $\{r_i\}_i$, becomes isomorphic to the vector space itself, that is we have $A^* \cong A$ (note that this isomorphism is not natural). The ϵ and η maps, given by the inner product and maximally entangled states or *Bell pairs*, are defined as follows:

$$\begin{aligned} \epsilon_A: A \otimes A &\rightarrow \mathbb{R} & \text{given by} & \sum_{ij} c_{ij} r_i \otimes r_j \mapsto \sum_{ij} c_{ij} \langle r_i \mid r_j \rangle \\ \eta_A: \mathbb{R} &\rightarrow A \otimes A & \text{given by} & 1 \mapsto \sum_i r_i \otimes r_i \end{aligned}$$

An example of a monoidal functor is Atiyah's definition of a topological quantum field theory (TQFT). This is a representation of category of manifolds and cobordisms **Cob** (representing, respectively possible choices of space and spacetime) over the category of finite dimensional vector spaces **FVect**. This representation is formalized using a strongly monoidal functor from **Cob** to **FVect** by Baez and Dylon [3], and assigns a vector space of states to each manifold and a linear operation to each cobordism.

3 Category Theory in Linguistics

We briefly review two orthogonal models of meaning in computational linguistics: pregroup grammars and *distributional* models of meaning, and we show how one can interpret the former in the latter using a strongly monoidal functor.

3.1 Type-Logical Pregroup Grammars

Consider the simple grammar generated by the following set of rules:

$$\begin{array}{ll}
 S \rightarrow Np Vp & itV \rightarrow \text{smile} \\
 Vp \rightarrow tV Np \mid N & tV \rightarrow \text{build} \\
 Np \rightarrow Adj Np \mid N & Adj \rightarrow \text{strong} \\
 & N \rightarrow \text{man, woman, house}
 \end{array}$$

The above rules are referred to as *generative rules*. The rules on the left describe the formation of a grammatical sentence S in terms of other non-terminals. According to these rules, a sentence is a noun phrase Np followed by a verb phrase Vp , where a verb phrase itself is a transitive verb tV followed either by a Np or a noun N , and a noun phrase is an adjective Adj followed either by a Np or a noun N . The rules on the right instantiate all but one (S) of the non-terminals to terminals. According to these, ‘smile’ is an intransitive verb, ‘build’ is a transitive verb, ‘strong’ is an adjective, and ‘man’, ‘woman’, and ‘house’ are nouns. We treat these words as lemmas and take freedom in conjugating them in our example sentences.

In a predicative approach, the non-terminals of the above grammar (except for S) are interpreted as unary or binary predicates to produce meaning for phrases and sentences. There are various options when interpreting these non-terminals: for instance, according to the first rule, we can either interpret a verb phrase as a binary predicate that inputs a noun phrase and outputs a sentence, or we can interpret a noun phrase as a binary predicate that inputs a verb phrase and outputs a sentence. We adhere to the more popular (among computational linguistics) verb-centric view and follow the former option. The types of the resulting predicates, obtained by recursively unfolding the rules, form an algebra of types, referred to as a *type-logical grammar*.

A pregroup type-logical grammar, or a pregroup grammar for short, is the pregroup freely generated over a set of basic types which, for the purpose of this paper, we take to be $\{n, s\}$. We refer to this free pregroup grammar by \mathbf{Preg}_F . Here, n is the type representing a noun phrase and s is the type representing a sentence. The complex types of this pregroup represent the predicates. For instance, $n^r \cdot s$ is the type of an intransitive verb, interpreted as a unary predicate that inputs a noun phrase and outputs a sentence. Explicit in this type is also the fact that the intransitive verb has to be on the right hand side of the noun phrase. This fact is succinctly expressed by the adjoint r of the type n . Similarly, $n^r \cdot s \cdot n^l$ is the type of a transitive verb, which is a binary predicate that inputs two noun phrases, has to be to the right of one and to the left of the other, and outputs a sentence. Finally, $n \cdot n^l$ is the type of an adjective in attributive position, a unary predicate that inputs a noun phrase and outputs another noun phrase; furthermore, it has to be to the left of its input noun phrase. These types are then assigned to the vocabulary of a language, that is to the non-terminals of the generative rules, via a relation referred to as a *type dictionary*. Our example type dictionary is as follows:

man	woman	houses	strong	smiled	built
n	n	n	$n \cdot n^l$	$n^r \cdot s$	$n^r \cdot s \cdot n^l$

Every sequence of words $w_1 w_2 \cdots w_n$ from the vocabulary has an associated *type reduction*, to which we refer to by $\alpha_{w_1 w_1 \cdots w_n}$. This type reduction represents the grammatical structure of the sequence. In a pregroup grammar, a type reduction is the result of applying the partial order, monoid, and adjunction axioms to the multiplication of the types of the words of the sequence. For example, the type reduction $\alpha_{\text{strong house}}$ associated to the sequence ‘strong house’ is computed by multiplying the types of ‘strong’ and ‘house’, that is $n \cdot n^l \cdot n$, then applying to it the adjunction and monoid axioms, hence obtaining $n \cdot n^l \cdot n \leq n$. Similarly, the type reduction of the sentence ‘strong man built houses’ is as follows:

$$\alpha_{\text{strong man built houses}} \quad : \quad n \cdot n^l \cdot n \cdot n^r \cdot s \cdot n^l \cdot n \leq n \cdot n^r \cdot s \leq s$$

In categorical terms, the type reduction is a morphism of the category \mathbf{Preg}_F , denoted by tensors and compositions of the ϵ and identity maps. For instance, the morphisms corresponding to the above adjective-noun phrase and sentence are as follows:

strong man	strong man built houses
$1_n \otimes \epsilon_n^l$	$(\epsilon_n^r \otimes 1_s) \circ (1_n \otimes \epsilon_n^l \otimes 1_{n^r \cdot s} \otimes \epsilon_n^l)$

The generative rules formalize the grammar of a natural language and their consequent type-logical grammars provide a predicative interpretation for the words with complex types. However, all the words with the same type have the same interpretation, and even worse, words with basic types are only interpreted as atomic symbols. In the next section, we will see how *distributional* models of meaning address this problem.

3.2 Distributional Models of Word Meaning

Meanings of some words can be determined by their denotations. For instance, meaning of the word ‘house’ can be the set of all houses or their images; and the answer to the question ‘what is a house?’ can be provided by pointing to a house. Matters get complicated when it comes to words with complex types such as adjectives and verbs. It is not so clear what is the denotation of the adjective ‘strong’ or the verb ‘build’. The problem is resolved by adhering to a meaning-as-use model of meaning, whereby one can assign meaning to all words, regardless of their grammatical type, according to the context in which they often appear. This context-based approach to meaning is the main idea behind the *distributional* models of meaning.

First formalized by Firth in 1957 [13] and about half a century later implemented and applied to word sense disambiguation by Schütze [37], distributional models of meaning interpret words as vectors in a highly dimensional (but finite) vector space with a fixed orthonormal basis over real numbers. A basis for this vector space is a set of *target* words, which in principle can be the set of all lemmatized words of a corpus of documents or a dictionary. In practice, the basis vectors are often restricted to the few thousands most occurring words of the corpus, or a set of specialized words depending on the application domain, e.g. a music dictionary. Alternatively, they can be *topics* obtained from a dimensionality reduction algorithm such as *single value decomposition* (SVD). We refer to such a vector space with an orthonormal basis $\{w_i\}_i$, no matter how it is built, as our basic distributional vector space W ; and to \mathbf{FVect} restricted to tensor powers of W as \mathbf{FVect}_W .

In this model, to each word is associated a vector, which serves as the meaning of the word. The weights of this vector are obtained by counting how many times the word has appeared ‘close’ to a basis word, where ‘close’ is a window of n (usually equal to 5) words. This number is

usually normalized, often in the form of TF-IDF values which show how important is a specific basis word by taking into account not only the number of times it has occurred to the document, but also the number of documents in which it appears in the corpus.

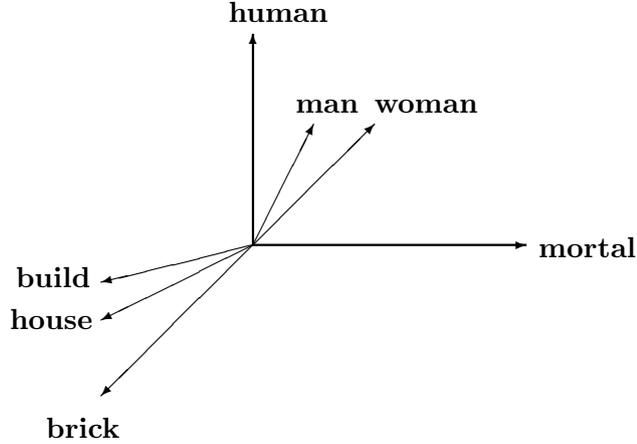


Figure 1: A Toy Distributional Model of Meaning

As an example, consider the toy vector space of Figure 1. The set {human, mortal, brick} is the basis of this vector space and the words ‘man’, ‘woman’, ‘house’ and ‘build’ each have a vector assigned to them. The words that have often appeared in the same context have a smaller angle between their vectors. For instance, ‘house’ and ‘build’ often appear close to ‘brick’, whereas ‘man’ and ‘woman’ often appear close to ‘mortal’ and ‘human’. The cosine of the angle between the word vectors has proved to be a good measure in predicting synonymy of words [11]. Despite these good predictions, the distributional models of meaning cannot serve as the definite models of natural language, as there is more to a language than the contexts of its words and these models on their own do not scale up to the interpretations of phrases and sentences. In the next section, we will see how a combination of type-logical and distributional models overcome both of their corresponding shortcomings.

3.3 Quantizing the Grammar

We provide a mapping of the free pregroup grammar \mathbf{Preg}_F to \mathbf{FVect}_W via a strongly monoidal functor \mathcal{F} . This functor assigns the basic vector space W to both of the basic types, that is, we have:

$$\mathcal{F}(n) = \mathcal{F}(s) = W$$

By functoriality, the partial orders between the basic types (for example those presented in [26]) are mapped to linear maps from W to W . The adjoints of basic types are also mapped to W , since for $x \in \{n, s\}$ we have the following, motivated by the above mentioned fact that $W^* \cong W$:

$$\mathcal{F}(x^l) = \mathcal{F}(x^r) = \mathcal{F}(x)$$

Since $W^{**} \cong W^* \cong W$, the iterated adjoint types are also mapped to W :

$$\mathcal{F}(x^{ll}) = \mathcal{F}(x^{rr}) = \mathcal{F}(x)$$

The complex types are mapped to tensor products of vector spaces, that is:

$$\mathcal{F}(n \cdot n^l) = \mathcal{F}(n^r \cdot s) = W \otimes W \quad \mathcal{F}(n^r \cdot s \cdot n^l) = W \otimes W \otimes W$$

Similarly, the type reductions are mapped to the compositions of tensor products of identity and ϵ maps of \mathbf{FVect}_W , for instance the type reduction of a transitive sentence is mapped as follows:

$$\mathcal{F}(\alpha_{\text{subj verb obj}}) = \mathcal{F}(\epsilon_n^r \otimes 1_s \otimes \epsilon_n^l) = \epsilon_W \otimes 1_W \otimes \epsilon_W: W \otimes (W \otimes W \otimes W) \otimes W \longrightarrow W$$

Now we can use the definition of [10] to provide a meaning for phrases and sentences of our grammar. The meaning of a sequence of words $w_1 w_2 \cdots w_n$ with type reduction $\alpha_{w_1 w_2 \cdots w_n}$ is:

$$\mathbf{Definition(*)} \quad \mathcal{F}(\alpha_{w_1 w_2 \cdots w_n}) (\overrightarrow{w_1} \otimes \overrightarrow{w_2} \otimes \cdots \otimes \overrightarrow{w_n})$$

As an example, take:

$$\overrightarrow{men} = \sum_i c_i^{men} \overrightarrow{w_i} \quad \overrightarrow{houses} = \sum_k c_k^{houses} \overrightarrow{w_k} \quad \overrightarrow{built} = \sum_{ijk} c_{ijk}^{built} (\overrightarrow{w_i} \otimes \overrightarrow{w_j} \otimes \overrightarrow{w_k})$$

Substituting these in Definition(*), we obtain the following for the meaning of the sentence ‘men built houses’:

$$\begin{aligned} \mathcal{F}(\epsilon_n^r \otimes 1_s \otimes \epsilon_n^l) (\overrightarrow{men} \otimes \overrightarrow{built} \otimes \overrightarrow{houses}) &= \\ (\epsilon_W \otimes 1_W \otimes \epsilon_W) (\overrightarrow{men} \otimes \overrightarrow{built} \otimes \overrightarrow{houses}) &= \\ \sum_{ijk} c_{ijk}^{built} \langle \overrightarrow{men} | \overrightarrow{w_i} \rangle \langle \overrightarrow{w_k} | \overrightarrow{houses} \rangle \overrightarrow{w_j} & \end{aligned}$$

This definition ensures that the interpretations of noun phrases and sentences of any grammatical structure, for instance intransitive or transitive, will be a vector in W , hence we can measure the cosine of the angle between them and compute their synonymy. In order to determine that this measure of synonymy provides good predictions, we need to run some experiments. However, whereas we know very well how to build vectors in W for words with basic types such as ‘man’ and ‘house’, our method further requires interpretations of words with complex types to be in tensor spaces, and there is no known standard procedure to construct these. In the next section we show how the notion of a Frobenius algebra over a vector space can be of use in addressing this matter.

4 Frobenius Algebras

Frobenius algebras were originally introduced in 1903 by F. G. Frobenius in the context of proving representation theorems for group theory [14]. Since then, they have found applications in other fields of mathematics and physics, e.g. in topological quantum field theories [24] and in categorical quantum mechanics [9]. The general categorical definitions recalled below are due to Carboni and Walters [7]. Their concrete instantiations to algebras over vector spaces were developed in [9].

A Frobenius algebra over a symmetric monoidal category $(\mathcal{C}, \otimes, I)$ is a tuple $(F, \sigma, \iota, \mu, \zeta)$, where for an F object of \mathcal{C} the triple (F, σ, ι) is an associative coalgebra, that is, the following are morphisms of \mathcal{C} , satisfying the coalgebraic associativity and unit conditions:

$$\sigma: F \rightarrow F \otimes F \qquad \qquad \qquad \iota: F \rightarrow I$$

The triple (F, μ, ζ) is an associative algebra, that is, the following are morphisms of \mathcal{C} , satisfying the algebraic associativity and unit conditions:

$$\mu: F \otimes F \rightarrow F \qquad \zeta: I \rightarrow F$$

Moreover, the above σ and μ morphisms satisfy the following *Frobenius* condition:

$$(\mu \otimes 1_F) \circ (1_F \otimes \sigma) = \sigma \circ \mu = (1_F \otimes \mu) \circ (\sigma \otimes 1_F)$$

A Frobenius Algebra is *commutative* if it satisfies the following two conditions for $\delta: X \otimes Y \rightarrow Y \otimes X$, the symmetry morphism of $(\mathcal{C}, \otimes, I)$:

$$\delta \circ \sigma = \sigma \qquad \mu \circ \delta = \mu$$

Finally, a Frobenius Algebra is *isometric* or *special* if it satisfies the following condition:

$$\mu \circ \sigma = Id$$

In the category **FVect**, any vector space V with a fixed basis $\{\vec{v}_i\}_i$ has a commutative special Frobenius algebra over it, explicitly given as follows:

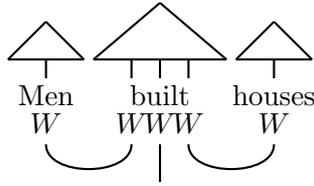
$$\begin{aligned} \sigma &:: \vec{v}_i \mapsto \vec{v}_i \otimes \vec{v}_i & \iota &:: \vec{v}_i \mapsto 1 \\ \mu &:: \vec{v}_i \otimes \vec{v}_i \mapsto \vec{v}_i & \zeta &:: 1 \mapsto \vec{v}_i \end{aligned}$$

In a Frobenius algebra over an orthonormal vector space, the coalgebra and algebra operations relate to each other via the equation $\sigma^\dagger = \mu$, where \dagger is the adjoint, equal to the transpose for vector spaces over reals.

In such Frobenius algebras, the operation σ corresponds to copying and its unit ι corresponds to deleting of the vectors. They enable one to faithfully encode vectors of W into spaces with higher tensor ranks, such as $W \otimes W, W \otimes W \otimes W, \dots$. In linear algebraic terms, for $v \in W$, we have that $\sigma(v)$ is a diagonal matrix whose diagonal elements are weights of v . The operation μ is referred to as uncopying; it loses some information when encoding a higher rank tensor into a lower rank space. In linear algebraic terms, for $z \in W \otimes W$, we have that $\mu(z)$ is a vector consisting only of the diagonal elements of z , hence losing the information encoded in the non-diagonal part.

5 Pictorial Calculi

The framework of compact closed categories comes with a complete diagrammatic calculus that allows convenient graphical representations of the derivations. We briefly introduce the fragment of this calculus that we are going to use in this paper. The objects of this fragment are the *tensors* of multi-linear algebra; that is, a vector is a rank-1 tensor, a matrix is a rank-2 tensor, and a 3d-array is a rank-3 tensor. Each tensor is represented by a triangle, whose rank can be determined by its wires. Words are represented by tensors that correspond to their meaning: subjects and objects are rank-1 tensors (vectors), adjectives and intransitive verbs are rank-2 tensors, and transitive verbs are rank-3 tensors. The ϵ maps are depicted as cups, whereas the identity morphism is a vertical straight line. The tensor products of vectors are represented by juxtaposing their corresponding triangles. For example, the meaning of a transitive sentence, following Definition(*), is depicted as follows:



Computations with Frobenius algebras can also be represented within the more general diagrammatic calculus of symmetric monoidal categories, referred to as *string* diagrams, first formalized in [20]. Specifically, the linear maps of the coalgebra and algebra are depicted by:

$$(\sigma, \iota) = \begin{array}{c} \diagup \\ \circ \\ \diagdown \end{array} \begin{array}{c} | \\ \circ \end{array} \quad (\mu, \zeta) = \begin{array}{c} \diagdown \\ \circ \\ \diagup \end{array} \begin{array}{c} | \\ \circ \end{array}$$

The *Frobenius* condition is depicted by:

$$\begin{array}{c} | \\ \circ \\ \diagdown \\ \circ \\ | \end{array} = \begin{array}{c} \diagdown \\ \circ \\ | \\ \circ \\ \diagup \end{array} = \begin{array}{c} \diagup \\ \circ \\ | \\ \circ \\ \diagdown \end{array}$$

The commutativity conditions are shown as:

$$\begin{array}{c} \diagdown \\ \circ \\ \diagup \end{array} = \begin{array}{c} \diagup \\ \circ \\ \diagdown \end{array} \quad \begin{array}{c} \diagdown \\ \circ \\ \diagup \end{array} = \begin{array}{c} \diagup \\ \circ \\ \diagdown \end{array}$$

The isometry condition is depicted by:

$$\begin{array}{c} \diagdown \\ \circ \\ \diagup \\ \circ \\ \diagdown \end{array} = \begin{array}{c} | \\ \circ \end{array}$$

Finally, the Frobenius conditions guarantee that any diagram depicting a Frobenius algebraic computation can be reduced to a normal form that only depends on the number of input and output wires of the nodes, provided that the diagram of computation is connected. This justifies depicting computations with Frobenius algebras as *spiders*, referring to the right hand side diagram below:

$$\begin{array}{c} \diagdown \\ \circ \\ \diagup \\ \dots \\ \diagdown \\ \circ \\ \diagup \end{array} = \begin{array}{c} \dots \\ \diagdown \\ \circ \\ \diagup \\ \dots \end{array}$$

For an informal introduction to compact closed categories, Frobenius algebras, and their diagrammatic calculi, see [8].

6 Building Tensors for Words with Complex Types

The type-logical models of meaning treat words with complex types as predicates. In a matrix calculus, predicates can be modeled as matrices (or equivalently, linear maps), over the semiring of booleans. In vector spaces over reals, one can extend these 0/1 entries to real numbers and

model words with complex types as weighted predicates. These are predicates that not only tell us which instantiations of their arguments are related to each other, but also that to what extent these are related to each other. For instance, a transitive verb is a binary predicate that, in the type-logical model, tells us which noun phrases are related to other noun phrases. In a vector space model, the verb becomes a linear map that moreover tells us to what extent these noun phrases are related to each other.

Building such linear maps from a corpus turns out to be a non-trivial task. In previous work [15, 16] we argue that such a linear map can be constructed by taking the sum of the tensor products of the vectors/linear maps of its arguments. For instance, the linear map representing an n -ary predicate p with arguments a_1 to a_n is $\sum_i \bar{a}_1 \otimes \cdots \otimes \bar{a}_n$, where \bar{a}_j is the vector/linear map associated to the argument a_j and the index i counts the number of times each word a_j has appeared in the corpus as the argument of p . Following this method, the linear maps corresponding to the predicates of our simple grammar are as follows:

intransitive verb	transitive verb	adjective
$\sum_i \overrightarrow{sbj}_i$	$\sum_i (\overrightarrow{sbj}_i \otimes \overrightarrow{obj}_i)$	$\sum_i \overrightarrow{noun}_i$

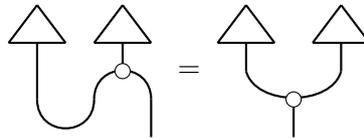
There is a problem: this method provides us with a linear map in a space whose tensor rank is one less than the rank of the space needed by Definition(*). For instance, the linear map of the transitive verb ends up being in $W \otimes W$, but we need a linear map in $W \otimes W \otimes W$. This problem is overcome by using the operations of a Frobenius algebra over vector spaces. We use the pictorial calculi of the compact closed categories and Frobenius algebras to depict the resulting linear maps and sentence vectors.

6.1 Adjectives and Intransitive Verbs

The linear maps of adjectives and intransitive verbs are elements of W . In order to encode them in $W \otimes W$, we use the Frobenius σ operation and obtain the following linear map:



For the intransitive verb, when substituted in Definition(*), that is, when applied to its subject, the above will result in the left hand side vector below, which is then normalized to the right hand side vector.

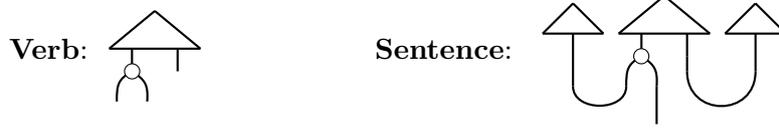


When an adjective is applied to a noun, the order of the above application is swapped: the triangle of the adjective will change place with the triangle of the subject of the intransitive verb.

6.2 Transitive Verbs

The linear map of a transitive verb is an element of $W \otimes W$; this has to be encoded in $W \otimes W \otimes W$. We face a few options here, which geometrically speaking provide us with different ways of “diagonally” placing a plane into a cube.

CpSbj The first option is to copy the “row” dimension of the linear map corresponding to the verb; this dimension encodes the information of the subjects of the verb from the corpus. In the left hand side diagram below we see how σ transforms the verb in this way. Once substituted in Definition(*), we obtain the diagram in the right hand side:



In this case, the σ map transforms the matrix of the verb as follows:

$$\sigma :: \sum_{ij} c_{ij}(\vec{n}_i \otimes \vec{n}_j) \mapsto \sum_{iij} c_{ii}(\vec{n}_i \otimes \vec{n}_i \otimes \vec{n}_j)$$

CpObj Our other option is to copy the “column” dimension of the matrix, which encodes the information about the objects of the verb from the corpus:



Now the σ -map does the following transformation:

$$\sigma :: \sum_{ij} c_{ij}(\vec{n}_i \otimes \vec{n}_j) \mapsto \sum_{ijj} c_{jj}(\vec{n}_i \otimes \vec{n}_j \otimes \vec{n}_j)$$

The diagrams above simplify the calculations involved, since they suggest a closed form formula for each case. Taking as an example the diagram of the copy-subject method, we see that: (a) the object interacts with the verb; (b) the result of this interaction serves as input for the σ map; (c) one wire of the output of σ interacts with the subject, while the other branch delivers the result. Linear algebraically, this corresponds to the computation $\sigma(\overline{verb} \times \overline{obj})^\top \times \overline{sbj}$, which expresses the fact that the meaning of a sentence is obtained by first applying the meaning of the verb to the meaning of the object, then applying the (σ version of the) result to the meaning of the subject. This computation results in the Equation 1 below:

$$\overline{sbj} \overline{verb} \overline{obj} = \overline{sbj} \odot (\overline{verb} \times \overline{obj}) \quad (1)$$

This order of application is the exact same way formalized in the generative rules of the language. On the contrary, the meaning of a transitive sentence for the copy-object results is given by Equation 2 below, which expresses the fact that the meaning of a sentence is obtained by first applying the (transposed) meaning of the verb to the meaning of the subject and then applying the result to the meaning of the object:

$$\overline{sbj} \overline{verb} \overline{obj} = \overline{obj} \odot (\overline{verb}^\top \times \overline{sbj}) \quad (2)$$

Note that equipped with the above closed forms we do not need to create or manipulate rank-3 tensors at any point of the computation, something that would cause high computational overhead.

Purely syntactically speaking, in a pregroup grammar the order of application of a transitive verb does not matter: it is applied to its subject and object in parallel. Semantically, as originated in the work of Montague [35], a transitive verb is first applied to its object and

then to its subject. In the more modern approaches to semantics via logical grammars, this order is some times based on the choice of the specific verb [19]. Our work in this paper is more inline with the latter approach, where for the specific task of disambiguating the verbs of our dataset, first applying the verb to the subject then to the object seems to provide better experimental results. According to our general theoretical setting, the linear map corresponding to the transitive verb should be a rank-3 tensor, but at the moment, apart from work in progress which tries to conjoin efforts with Machine Learning to directly build these as rank-3 tensors, we do not have the technology to do other than described in this paper. However, in the ideal case that the linear maps of words are already in the spaces allocated to them by the theory, these choices will not arise, as the compact nature of the matrix calculus implies that the application can be done in parallel in all the cases that parallel applications are prescribed by the syntax. From a linear-algebra perspective, fully populated rank-3 tensors for verbs satisfy the following equality:

$$\overrightarrow{subj} \overrightarrow{verb} \overrightarrow{obj} = (\overrightarrow{verb} \times \overrightarrow{obj})^\top \times \overrightarrow{subj} = (\overrightarrow{verb}^\top \times \overrightarrow{subj}) \times \overrightarrow{obj}$$

which shows that the order of application does not actually play a role.

MixCpDI We can also use a mixture of σ and μ maps. There are three reasonable options here, all of which start by applying two σ 's to the two wires of the linear map of the verb (that is, one for each of the dimensions). Then one can either apply a ι to one of the copies of the first wire, or a ι to one of the copies of the second wire. These two options are depicted as follows:

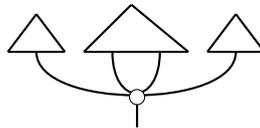


The first diagram has the same normal form as the copy-subject option, and the second one has the same normal form as the copy-object option.

Finally, one can apply a μ to one wire from each of the copied wires of the verb, the result of which is depicted in the following left hand side diagram. When substituted in Definition(*), we obtain the following right hand side diagram for the meaning of the transitive sentence:



The normal form of the diagram of the sentence is obtained by collapsing the three dots and yanking the corresponding wires, resulting in the following diagram:



Linear algebraically, the spider form of the verb is equivalent to $\sum_i (\overrightarrow{sbj}_i \odot \overrightarrow{obj}_i)$. A verb obtained in this way will only relate the properties of its subjects and objects on identical bases and there will be no interaction of properties across bases. For instance, for a certain verb v , this construction will result in a vector that only encodes to what extent v has related subjects and objects with property \overrightarrow{w}_1 , and has no information about to what extent v has related subjects with property \overrightarrow{w}_1 to objects with property \overrightarrow{w}_2 . The closed form of the above diagram is:

$$\overrightarrow{sbj} \odot \left(\sum_i (\overrightarrow{sbj}_i \odot \overrightarrow{obj}_i) \right) \odot \overrightarrow{obj}$$

Model	Description
CPSBJ	Copy subject on relational matrices
CPOBJ	Copy object on relational matrices
MIXCPDL	Diagonalize on relational matrices
KRON	Diagonalize on direct matrices
MULTP	Multiplicative model

Table 1: The models.

7.1 Disambiguation

We first test our models on a disambiguation task, which is an extension of Schütze’s original disambiguation task from words to sentences. A dataset for this task was originally developed in [29] for intransitive sentences, and later extended to transitive sentences in [16]; we use the latter. The goal is to assess how well a model can discriminate between the different senses of an ambiguous verb, given the context (subject and object) of that verb. The entries of this dataset consist of a target verb, a subject, an object, and a landmark verb used for the comparison. One such entry for example is, “write, pupil, name, spell”. A good model should be able to understand that the sentence “pupil write name” is closer to the sentence “pupil spell name” than, for example, to “pupil publish name”. On the other hand, given the context “writer, book” these results should be reversed. The evaluation of this experiment is performed by calculating Spearman’s ρ , which measures the degree of correlation of the cosine distance with the judgements of 25 human evaluators, who has been asked to assess the similarity of each pair of sentences using a scale from 1 to 7. As our baseline we use a simple additive model (ADDTV), where the meaning of a transitive sentence is computed as the addition of the relevant context vectors.

Results The results of this experiment are shown in Table 2, indicating that the most successful model for this task is the copy-object model. The better performance of this model against the copy-subject approach provides us some insights about the role of subjects and objects in disambiguating our verbs. By copying the dimension associated with the object, the compression of the original sentence matrix, as this was calculated in [16], takes place along the dimension of subjects (rows), meaning that the resulting vector will bring much more information from the objects than the subjects (this is also suggested by Equation 2). Hence, the fact that this vector performs better than the one of the copy-subject method provides an indication that the object of some ambiguous verbs (which turns out to be the case for our dataset) can be more important for disambiguating that verb than the subject. Intuitively, we can imagine that the crucial factor to disambiguate between the verbs “write”, “publish” and “spell” is more the object than the subject: a book or a paper can be both published and written, but a letter or a shopping list can only be written. Similarly, a word can be spelled and written, but a book can only be written. The subject in all these cases is not so important.

The copy-object model is followed closely by the (KRON) and the MULTP models. The similar performance of these two is not a surprise, given their almost identical nature. Finally, the bad performance of the model (MIXCPDL) that is obtained by the application of the uncopied μ map conforms to the predictions of the theory, as these were expressed in Section 6.

7.2 Comparing Transitive and Intransitive Sentences

In this section we will examine the potential of the above approach in practice, in the context of an experiment aiming to compare transitive and intransitive sentences. In order to do that, we

Model	High	Low	ρ
ADDTV	0.90	0.90	0.050
MULTP	0.67	0.60	0.163
MIXCPDL	0.75	0.77	0.000
KRON	0.31	0.21	0.168
CPsBJ	0.95	0.95	0.143
CPoBJ	0.89	0.90	0.172
UpperBound	4.80	2.49	0.620

Table 2: Disambiguation results. High and Low refer to average scores for high and low landmarks, respectively. UpperBound refers to agreement between annotators.

use the dataset of the previous verb disambiguation task (see detailed description in Section 7.1) to conduct the following simple experiment: We create intransitive versions of all the transitive sentences from target verbs and their high and low landmarks by dropping the object; then, we compare each transitive sentence coming from the target verbs with all the other intransitive sentences, expecting that the highest similarity would come from its own intransitive version, the next higher similarity would come from the intransitive version that uses the corresponding high landmark verb, and so on. To present a concrete example, consider the entry “write, pupil, name, spell, publish”. Our transitive sentence here is s_{tr} = “pupil write name”; the intransitive version of this is s_{in} = “pupil write”. We also create intransitive versions using the high and the low landmarks, getting s_{hi} = “pupil spell” and s_{lo} = “pupil publish”. If the similarity between two sentences s_1 and s_2 is given by $sim(s_1, s_2)$, we would expect that:

$$sim(s_{tr}, s_{it}) > sim(s_{tr}, s_{hi}) > sim(s_{tr}, s_{lo}) > sim(s_{tr}, s_u)$$

where s_u represents an unrelated intransitive version coming from a target verb different than the one of s_{tr} . The results of this experiment are shown in Table 3 below, for 100 target verbs.

Case	Errors	%
$sim(s_{tr}, s_{hi}) > sim(s_{tr}, s_{it})$	7 of 93	7.5
$sim(s_{tr}, s_{lo}) > sim(s_{tr}, s_{it})$	6 of 93	5.6
$sim(s_{tr}, s_u) > sim(s_{tr}, s_{it})$	36 of 9900	0.4

Table 3: Results of the comparison between transitive and intransitive sentences.

The outcome follows indeed our expectations for this task. We see, for example, that the highest error rate comes from cases where the intransitive sentence of the high-landmark verb is closer to a transitive sentence than the intransitive version coming from the sentence itself (first row of the table). Since the meaning of a target verb and the high-landmark verb were specifically selected to be very similar given the context (subject and object), this is naturally the most error-prone category. The seven misclassified cases are presented in Table 4, where the similarity of the involved intransitive versions is apparent.

The six cases of the second category (where an intransitive sentence from a low-landmark gives higher similarity than the normal intransitive version) are quite similar, since in many cases dropping the object leads to semantically identical expressions. For the transitive sentence “tribunal try crime”, for example, the low-landmark intransitive version “tribunal test” has almost identical meaning with the normal intransitive version “tribunal try”, so it is easier to be “mistakenly” selected by the model as the one closest to the original transitive sentence.

s_{tr}	s_{in}	s_{hi}
people run round	people run	people move
boy meet girl	boy meet	boy visit
management accept responsibility	management accept	management bear
patient accept treatment	patient accept	patient bear
table draw eye	table draw	table attract
claimant draw benefit	claimant draw	claimant attract
tribunal try crime	tribunal try	tribunal judge

Table 4: Errors in the first category of comparisons.

Finally, the model performs really well for cases when an unrelated intransitive sentence is compared with a transitive one, with only a 0.4% error rate. Here many of the misclassifications can also be attributed to the increased ambiguity of the involved verbs when the object is absent. For example, the similarity between “man draw sword” and “man draw” is considered smaller than the similarity of the first sentence with “man write”. Although this is an obvious error, we should acknowledge that the two intransitive sentences, “man draw” and “man write”, are not so different semantically, so the error was not completely unjustified.

7.3 Definition Classification

The ability of reliably comparing the meaning of single words with larger textual fragments, e.g. phrases or even sentences, can be an invaluable tool for many challenging NLP tasks, such as definition classification, sentiment analysis, or even the simple everyday search on the internet. In this task we examine the extend to which our models can correctly match a number of terms (single words) with a number of definitions. Our dataset consists of 112 terms (72 nouns and 40 verbs) extracted from a junior dictionary together with their main definition. For each term we added two more definitions, either by using entries of WordNet for the term or by simple paraphrase of the main definition using a thesaurus, getting a total of three definitions per term. In all cases a definition for a noun-term is a noun phrase, whereas the definitions for the verb-terms consist of verb phrases. A sample of the dataset entries can be found in Table 5.

Term	Main definition	Alternative def. 1	Alternative def. 2
blaze	large strong fire	huge potent flame	substantial heat
husband	married man	partner of a woman	male spouse
foal	young horse	adolescent stallion	juvenile mare
horror	great fear	intense fright	disturbing feeling
apologise	say sorry	express regret or sadness	acknowledge shortcoming or failing
embark	get on a ship	enter boat or vessel	commence trip
vandalize	break things	cause damage	produce destruction
regret	be sad or sorry	feel remorse	express dissatisfaction

Table 5: Sample of the dataset for the term/definition comparison task (noun-terms in the top part, verb-terms in the bottom part).

We approach this evaluation problem as a classification task, where terms have the role of classes. First, we calculate the distance between each definition and every term in the dataset. The definition is “classified” to the term that gives the higher similarity. Due to the nature the dataset, this task did not require human annotation and we evaluate the results by calculating

separate F1-scores for each term, and getting their average as an overall score for the whole model. The results are presented in Table 6.

Results Since this experiment includes verb phrases, where the subject is missing, we construct our verb vectors by summing over all context vectors of objects with which the verb appears in the corpus; that is, we use $\overrightarrow{verb} = \sum_i \overrightarrow{obj}_i$. This is referred to as the relational model (RELTN), and is compared with the multiplicative model. Additive model serves again as our baseline. We evaluate separately the performance on the noun terms and the performance on the verb terms, since a mixing of the two sets would be inconsistent.

The relational model delivers again the best performance, although the difference from the multiplicative model is small. All models perform better on the verb terms than the noun part of the dataset, yet in general F-scores tend to be low. This is natural, since the challenge that this task poses to a machine is great, and F-score considers anything but the perfect result (every definition assigned to the correct term) as unacceptable.

An error analysis shows that for the noun-term set the relational model returns the correct main definition in 25 of the 72 cases, whereas in 47 cases (65%) the correct definition is in the top-five list for that term (Table 7). The multiplicative model performs similarly, and better for the verb-term set. For this experiment we also calculated Mean Reciprocal Ranks values, which again were very close for the two models. Furthermore, some of the “misclassified” cases can also be considered as somehow “correct”. For example, the definition we originally assigned to the term ‘jacket’ was ‘short coat’; however, the system “preferred” the definition ‘waterproof cover’, which is also correct. Some interesting other cases are presented in Table 8.

8 Conclusion and Future Work

In summary, after a brief review of the definitions of compact closed categories and monoidal functors, we ramified their applications to natural language syntax and semantics. We recasted the categorical setting of [10], which was based on the product of the category of a pregroup type-logic with the category of finite dimensional vector spaces, in terms of a monoidal functor from the former to the latter. This passage is similar to the vector space representation of category of manifolds in a topological quantum field theory. We showed how the operations of Frobenius algebras over vector spaces provide a concrete instantiation of this setting and used their pictorial calculus to simplify the multi-linear algebraic computations. This instantiation resulted in meanings of all sentences living in a basic vector space W , hence we became able to compare their meanings with one another and also with meanings of single words and phrases. We developed experiments based on this instantiation and evaluated the predictions of our model in a number of sentence and phrase synonymy tasks.

We conclude that the concrete setting of this paper provides a robust and scalable base for an implementation of the framework of [10], ready for further experimentation and applications. It overcomes the shortcomings of our first implementation [16], whose main problem was that the

Model	Nouns			Verbs		
	P	R	F1	P	R	F1
ADDTV	0.21	0.17	0.16	0.28	0.25	0.23
MULTP	0.21	0.22	0.19	0.31	0.30	0.26
RELTN	0.22	0.24	0.21	0.32	0.28	0.27

Table 6: Results of the term/definition comparison task.

	Rank	MULTP		RELTN	
		Count	%	Count	%
Nouns	1	26	36.1	25	34.7
	2-5	20	27.8	22	30.6
	6-10	11	15.3	5	6.9
	11-72	15	20.8	20	27.8
Verbs	1	15	37.5	8	20.0
	2-5	10	25.0	13	32.5
	6-10	6	15.0	4	10.0
	11-40	9	22.5	15	37.5

Table 7: Results of the term/definition comparison task based on the rank of the main definition.

Term	Original definition	Assigned definition
rod	fishing stick	round handle
jacket	short coat	waterproof cover
mud	wet soil	wet ground
labyrinth	complicated maze	burial chamber

Table 8: A sample of ambiguous cases where the model assigned a different definition than the original.

vector space representation of the atomic type s was taken to be the tensor space $W \otimes W$ (for a transitive sentence), hence the logical and the concrete types did not match. As a consequence, sentences with nested structures such as ‘Mary saw John reading a book’ could not be assigned a meaning; furthermore, meanings of phrases and sentences with different grammatical structure lived in different vector spaces, so a direct comparison of them was not possible.

An experimental future direction is a higher order evaluation of the definition classification task using an unambiguous vector space along the lines of [37], where each word is associated with one or more *sense* vectors. A model like this will avoid encoding different meanings of words in one vector, and will help us separate the two distinct tasks of composition and disambiguation that currently are interwoven in a single step.

From the theoretical perspective, one direction is to start from type-logical grammars that are more expressive than pregroups. In recent work [36] we have shown how the functorial passage to **FVect** can be extended from a pregroup algebra to the Lambek Calculus [25], which has a monoidal (rather than compact) structure. It remains to show how this passage can be extended to more expressive versions of Lambek calculi, such as Lambek-Grishin algebras [31], calculi of discontinuity [32], or abstract categorical grammars [18]. More specifically, one wants to investigate if the new operations and axioms of these extensions are preserved by the vector space semantic functor.

Another venue to explore is meanings of logical words, where any context-dependant method fails to succeed. The abstract operations of Frobenius algebras have been used to model a limited class of logical operations on vector spaces [9], hence these might turn out to be promising in this aspect too.

References

- [1] S. Abramsky and B. Coecke. 2004. *A categorical semantics of quantum protocols*. Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 415–425.

arXiv:quant-ph/0402130.

- [2] M. Atiyah. 1989. *Topological quantum field theories*. Publications Mathématique de l’Institut des Hautes Études Scientifiques **68**, 175–186.
- [3] J. C. Baez and J. Dolan. 1995. *Higher-dimensional algebra and topological quantum field theory*. Journal of Mathematical Physics **36**, 6073D0-6105.
- [4] M. Baroni and R. Zamparelli. 2010. *Nouns are vectors, adjectives are matrices*. Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP).
- [5] G. Birkhoff and J. von Neumann. 1936. *The logic of quantum mechanics*. Annals of Mathematics **35**, 823–843.
- [6] S. Clark and S. Pulman. 2007. *Combining Symbolic and Distributional Models of Meaning*. Proceedings of AAAI Spring Symposium on Quantum Interaction. AAAI Press.
- [7] A. Carboni and R. F. C. Walters. 1987. *Cartesian bicategories I*. Journal of Pure and Applied Algebra **49**.
- [8] B. Coecke, and E. Paquette. 2011. *Categories for the Practicing Physicist. New Structures for Physics*, 167–271. B. Coecke (ed.). Lecture Notes in Physics **813**. Springer.
- [9] B. Coecke, D. Pavlovic and J. Vicary. 2008. *A new description of orthogonal bases*. Mathematical Structures in Computer Science **1**.
- [10] B. Coecke, M. Sadrzadeh and S. Clark. 2010. *Mathematical Foundations for Distributed Compositional Model of Meaning*. Lambek Festschrift. Linguistic Analysis **36**, 345–384. J. van Benthem, M. Moortgat and W. Buszkowski (eds.).
- [11] J. Curran. 2004. *From Distributional to Semantic Similarity*. PhD Thesis, University of Edinburgh.
- [12] K. Erk and S. Padó. 2004. *A Structured Vector Space Model for Word Meaning in Context*. Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), 897–906.
- [13] J. R. Firth. 1957. *A synopsis of linguistic theory 1930-1955*. Studies in Linguistic Analysis.
- [14] F. G. Frobenius. 1903. *Theorie der hyperkomplexen Größen*. Preussische Akademie der Wissenschaften Berlin: Sitzungsberichte der Preußischen Akademie der Wissenschaften zu Berlin. Reichsdr.
- [15] E. Grefenstette, M. Sadrzadeh, S. Clark, B. Coecke, S. Pulman. 2011. *Concrete Compositional Sentence Spaces for a Compositional Distributional Model of Meaning*. International Conference on Computational Semantics (IWCS’11).
- [16] E. Grefenstette and M. Sadrzadeh. 2011. *Experimental Support for a Categorical Compositional Distributional Model of Meaning*. Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP).
- [17] E. Grefenstette and M. Sadrzadeh. 2011. *Experimenting with Transitive Verbs in a DisCo-Cat*. Proceedings of Workshop on Geometrical models of natural language semantics (GEMS).
- [18] P. de Groot. 2001. *Towards Abstract Categorical Grammars*. Proceedings of Association for Computational Linguistic, 148–155.

- [19] P. de Groote and F. Lamarche. 2002. *Classical non-associative Lambek calculus*. *Studia Logica* **71**, 355–388.
- [20] A. Joyal and R. Street. 1991. *The Geometry of Tensor Calculus I*. *Advances in Mathematics*, Vol. 88.
- [21] G. M. Kelly. 1972. *Many-variable functorial calculus (I)*. *Lecture Notes in Mathematics, Coherence in Categories* **281**, 66–105. G. M. Kelly, M. Laplaza, G. Lewis and S. MacLane (eds.).
- [22] G. M. Kelly and M. L. Laplaza. 1980. *Coherence for compact closed categories*. *Journal of Pure and Applied Algebra* **19**, 193–213.
- [23] A. Kock. 1972. *Strong functors and monoidal monads*. *Archiv der Mathematik* **23**, 113–120. Springer.
- [24] J. Kock. 2003. *Frobenius Algebras and 2D Topological Quantum Field Theories*. London Mathematical Society Student Texts. Cambridge University Press.
- [25] J. Lambek. 1958. *The mathematics of sentence structure*, *American Mathematics Monthly* **65**, 154–170.
- [26] J. Lambek. 2008. *From Word to Sentence*. Polimetrica, Milan.
- [27] J. Lambek and C. Casadio. 2006. *Computational algebraic approaches to natural language*. Polimetrica, Milan.
- [28] T. Landauer, and S. Dumais. 1997. *A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge*. *Psychological review*.
- [29] J. Mitchell and M. Lapata. 2008. *Vector-based models of semantic composition*. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, 236–244.
- [30] C. D. Manning, P. Raghavan, and H. Schütze. 2008. *Introduction to information retrieval*. Cambridge University Press.
- [31] M. Moortgat. 2009. *Symmetric categorial grammar*. *Journal of Philosophical Logic*, 681–710.
- [32] G. Morrill. 1995. *Discontinuity in Categorical Grammar*. *Linguistics and Philosophy* **18**, 175–219.
- [33] A. Preller and J. Lambek. 2007. *Free compact 2-categories*. *Mathematical Structures in Computer Science* **17**, 309–340.
- [34] A. Preller and M. Sadrzadeh. 2011. *Bell States and Negative Sentences in the Distributed Model of Meaning*. *Electronic Notes in Theoretical Computer Science, Proceedings of the 6th QPL Workshop on Quantum Physics and Logic* **270**, 141–153.
- [35] R. Montague. 1974. *English as a formal language*. *Formal Philosophy*, 189–223.
- [36] M. Sadrzadeh, E. Grefenstette and B. Coecke. 2012. *Lambek vs. Lambek: Vector Space Semantics and String Diagrams for Lambek Calculus*. *Annals of Pure and Applied Logic*, accepted for publication.

- [37] H. Schütze. 1998. *Automatic Word Sense Discrimination*. Computational Linguistics **24**, 97–123.
- [38] P. Selinger, 2011. *A survey of graphical languages for monoidal categories*. New Structures for Physics, 275–337. B. Coecke (ed.). Springer-Verlag.