Ambiguity in Categorical Models of Meaning

Robin Piedeleu

Balliol College

University of Oxford

A thesis submitted for the degree of

MSc in Computer Science

Trinity 2014

Abstract

Building on existing categorical accounts of natural language semantics, we propose a compositional distributional model of ambiguous meaning. Originally inspired by the high-level category theoretic language of quantum information protocols, the compositional, distributional categorical model provides a conceptually motivated procedure to compute the meaning of a sentence, given its grammatical structure and an empirical derivation of the meaning of its parts. Grammar is given a type-logical description in a compact closed category while the meaning of words is represented in a finite inner product space model. Since the category of finite-dimensional Hilbert spaces is also compact closed, the type-checking deduction process lifts to a concrete meaning-vector computation via a strong monoidal functor between the two categories. The advantage of reasoning with these structures is that grammatical composition admits an interpretation in terms of flow of meaning between words. Pushing the analogy with quantum mechanics further, we describe ambiguous words as statistical ensembles of unambiguous concepts and extend the semantics of the previous model to a category that supports probabilistic mixing. We introduce two different Frobenius algebras representing different ways of composing the meaning of words, and discuss their properties. We conclude with a range of applications to the case of definitions, including a meaning update rule that reconciles the meaning of an ambiguous word with that of its definition.

Contents

Chapter 1

Introduction

1.1 Background

Traditionally, the mathematical and computational study of natural language semantics has been tackled in conflicting ways. In particular, two contrasting approaches reflect the compositional and empirical aspects of language: the compositional type-logic approaches give priority to grammar and syntactic formalism to explain how we string words together to form sentences; the distributional approaches account for the meaning of individual words by an empirical analysis of the context in which they appear, in accordance with Firth's famous statement "You shall know a word by the company it keeps" [21]. More concretely, distributional semantics give a method to represent the meaning of words as vectors in a space whose basis is composed of relevant contextual features from a large body of text, and use the tools of linear algebra to compare them, typically with an inner product. However, such semantic models do not come with an intuitive method to compose the meaning of words and extend their interpretation to sentences. This is known as the problem of compositionality.

Recent research [12, 18, 15] provides a broader category-theoretic framework that unifies these two perspectives by successfully extending the distributional model of meaning from individual terms to sentences, thus effectively realising a compositional distributional model of meaning as first proposed in [10]. In this model, the internal logic of compact closed monoidal categories, such as Lambek Pregroups [36], allows the assignment of meaning deduced from the grammatical structure of a sentence and the meaning of its constituent parts.

More specifically, meaning is assigned via a structure preserving functor into the category of finite-dimensional Hilbert spaces, which is also a compact closed monoidal category and thus retains the same internal logic.

However, the "meaning is use" catchphrase of distributional semantics hides a more subtle reality. To us, humans, the meaning of a word can be partitioned into broad classes that organise its various possible uses. For example, the word *head* can be understood as the body part or as the leader of an organisation. Even though the distributional model of meaning captures both these senses, it seems that the representation of its meaning as a single vector is insufficient or, rather, discards precious information about the uses of the word.

Current compositional distributional models lie at the intersection of the logical and statistical aspects of language, reconciling the ambiguous and the systematic, yet there exists no theoretical account of ambiguity. How can we retain the ambiguity of word meaning in a compositional framework? Providing an answer to this question is what we set out to do.

1.2 Outline

We begin with an extensive presentation of the categorical compositional model of [12, 18, 15] from an abstract point of view while simultaneously motivating the introduction of each new structure by linguistic considerations. We gradually introduce the concept of a compact closed category and show that it is a suitable setting to model grammatical interactions in natural language. The monoidal structure allows for the juxtaposition of grammatical types while compactness (or, more generally, closedness) provides a notion of evaluation that reduces all grammatically correct sentences to the same type in a sound type-logical deductive system.

Next, we motivate the introduction of distributional models of meaning and their abstract counterpart, †-compact categories. Based on [15], we bridge the gap between syntax and semantics by constructing a strong monoidal functor between the two categorical models. This yields a procedure to compute the meaning of a sentence as a function of the meaning of its parts, in accordance with reduction rules inherited from the type grammar.

To construct concrete models, it is necessary to devise a process that copies and deletes information. To this effect we introduce Frobenius algebras and their linguistic applications in the categorical compositional model of meaning developed so far.

In parallel we express all of the above in a convenient graphical calculus that, ultimately, provides an intuitive understanding of grammatical interactions and sheds light on the flow of information between words.

Chapter 3 is the heart of this dissertation. Inspired by the categorical quantum physics literature [51, 19], we put forward abstract constructions that allow us to accommodate ambiguous meaning into the model of the preceding chapter, by extending its semantics to a category that supports probabilistic mixing: words are represented as density operators, i.e., as a statistical ensemble of their various possible meanings according to an existing distributional model; grammatical reductions are morphisms that preserve their structure.

As previously, we give a Frobenius algebra that implements the concrete composition of meaning between words. However, this algebra turns out to have a more complex structure than the previous examples. A new algebra is introduced and the composition process that each induces are compared on a few examples.

The last chapter presents applications of our model to the idea of definitions. It serves as an excuse to study some of the theoretical possibilities of a compositional model of meaning based on density operators. The role of relative clauses is discussed as well as the possibility of recovering the meaning of unknown words in a definition. At the end of the chapter we propose a rule to update the meaning of a word based on the information contained in its definition. The domain of application of this rule is more general as it opens the possibility of incremental learning in compositional distributional models of natural language semantics.

1.3 New contributions

In this dissertation we extend the functorial semantics of [15] to a new category and introduce two Frobenius algebras that implement the composition of meaning in a distributional setting, in order to account for the polysemy of words in natural language. To illustrate the expressiveness of our model, we show that some of the applications in the literature find a natural counterpart, e.g., (in)transitive sentences or relative clauses. In addition, we prove the possibility of recovering the partial meaning of unknown words when we know the meaning of a sentence in which they appear. Finally, we propose a method to update the meaning of words according to the meaning of their definitions.

We believe that the discussions on the possibilities of Frobenius algebras in Chapter 3 and the update rule of Chapter 4 open promising areas of research in compositional semantics.

Chapter 2

A compositional model of meaning

2.1 Categorical models of grammar

We wish to investigate mathematical structures that encapsulate the compositional aspect of natural language. In this chapter, we introduce the mathematical theory of categories, in order to develop a syntactical model of language. In particular, we will rely on the notion of compact closed category to capture parts of speech and the grammatical structure of sentences. In parallel, the graphical calculus associated to operations in these categories will be introduced to reason about the structure and parsing of sentences.

2.1.1 Categories capture compositionality

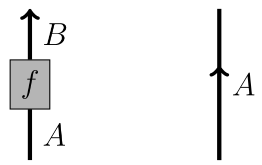
Recall that:

Definition 2.1.1. *a* category $\mathcal{C}$ *is an algebraic structure that consists of a class of* objects*, a class of* morphisms $\mathcal{C}(A,B)$ *between each ordered pair of objects* A, B*, and for every triple of objects* A*,* B *and* C*, a composition rule* $\circ : \mathcal{C}(A,B) \times \mathcal{C}(B,C) \to \mathcal{C}(A,C)$ *that is associative and satisfies a unit law, i.e., for every object* A*, there exists a distinguished morphism* 1_A *from* A *to itself, called the* identity *on* A*, such that, for* $f : B \to A$ *and* $g : A \to B$*,* $1_A \circ f = f$ *and* $g \circ 1_A = g$*.*

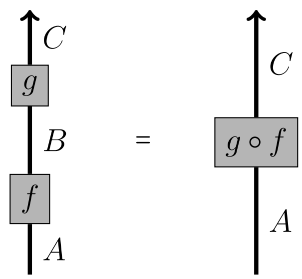
Notice that we write morphisms in a line as $f : A \to B$, $g : B \to C$. However, expressions in category theory are better pictured in diagrammatic form, in so-called *commutative diagrams* that additionally have the ability to express equations in this language (here, the composition rule is represented).

$$\begin{tikzcd}
B \arrow[r, "g"] & C \\
A \arrow[u, "f"] \arrow[ur, "g \circ f"'] &
\end{tikzcd}$$

Often, we will use a slightly different graphical notation, dual to the one above, in which morphisms are drawn as boxes with an input and output and objects as lines or wires, to be read from bottom to top (note that the identity on an object is simply drawn as a line without a box, a convention that is entirely justified by the unit law).



In this graphical depiction, composition is simply represented as connecting two boxes via matching lines, as below:



Arbitrary diagrams are essentially directed graphs: a number of vertices connected by edges.

We will also use the language of *functors* between categories. A functor corresponds to the generalisation of the notion of morphism - it is a map between categories that respects composition. Interestingly, functors between any two categories form a category in which the morphisms are called *natural transformations*. In what follows, we will assume a basic knowledge of functors, natural transformations, equivalence of categories and adjoint functors (for details and a rigorous treatment of these subjects we refer the reader to MacLane [42]).

Categories formalise the notion of *composition* of processes (morphisms) between different systems of one type (objects). Examples of categories include the category of sets and functions between them, of vector spaces and linear maps, of sets and relations, or of rings and ring homomorphisms.

In what follows, objects are to be thought of as grammatical *types* (verb, nouns, adjectives, etc.) that we can attribute to words in a sentence. The interpretation of morphisms is more subtle - we shall explore categories for which certain morphisms admit a linguistic interpretation, as an evaluation process that reduces grammatically correct sentences to a unique type.

2.1.2 Types for sentences in monoidal categories

The first step in this direction is to define the juxtaposition of objects to model the grammatical type of a juxtaposition of words in linguistics. To this effect, we introduce *monoidal categories*.

A monoidal category is a structure that allows us to compose morphisms sequentially (the ordinary way defined above) as well as horizontally. The latter is given by the existence of a functor $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$, called the *tensor* product. Thus, for a word of type A (e.g. "murderous") and another word w' of type N (e.g. "crow"), the type of the juxtaposition ww' ("murderous crow") is $A \otimes N$. Furthermore, since the tensor product is a functor, as announced, we obtain a similar way to compose morphisms - given $f_1 : A_1 \to B_1$ and $f_2 : A_2 \to B_2$ we can form $f_1 \otimes f_2 : A_1 \otimes A_2 \to B_1 \otimes B_2$. To *rigorously* capture these notions,

Definition 2.1.2. *a monoidal category is equipped with slightly more structure:*

i) a tensor product $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$

ii) that is associative*, i.e for all objects* A*,* B *and* C*, we have a natural isomorphism* $\alpha_{A,B,C} : A \otimes (B \otimes C) \to (A \otimes B) \otimes C$*;*

iii) and such that there exists a distinguished object I*, called the tensor* unit*, with two natural isomorphisms* $\rho_A : A \otimes I \to A$ *and* $\lambda_A : I \otimes A \to A$*, for all objects* A*;*

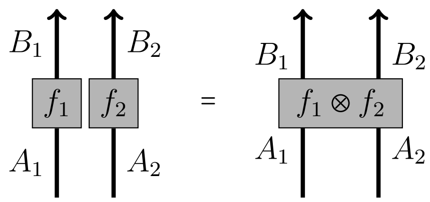
iv) subject to the following coherence conditions (see [42]):

$$\begin{tikzcd}
 & ((A \otimes B) \otimes C) \otimes D \arrow[dl, "\alpha_{A,B,C} \otimes 1_D"'] \arrow[dr, "\alpha_{(A\otimes B),C,D}"] & \\
(A \otimes (B \otimes C)) \otimes D \arrow[d, "\alpha_{A,(B\otimes C),D}"'] & & (A \otimes B) \otimes (C \otimes D) \arrow[d, "\alpha_{A,B,(C\otimes D)}"] \\
A \otimes ((B \otimes C) \otimes D) \arrow[rr, "1_A \otimes \alpha_{B,C,D}"'] & & A \otimes (B \otimes (C \otimes D))
\end{tikzcd}$$

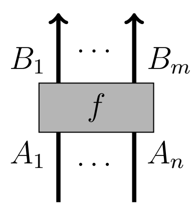
$$\begin{tikzcd}
(A \otimes I) \otimes B \arrow[rr, "\alpha_{A,I,B}"] \arrow[dr, "\rho_A \otimes 1_B"'] & & A \otimes (I \otimes B) \arrow[dl, "1_A \otimes \lambda_B"] \\
& A \otimes B &
\end{tikzcd}$$

If all the above structural isomorphisms are equalities, the category becomes *strict* monoidal. In the analysis of grammar, the categories that we will explore will all be strict monoidal. However, a category that we will encounter to provide semantics for natural language, that of (finite-dimensional) Hilbert spaces and linear maps equipped with the usual tensor product, is monoidal, yet not strict. Nonetheless, this is not a problem since, following the famous coherence theorem of MacLane [42, Theorem XI.3.1], every monoidal category is *equivalent* to a strict one. As a result, we will indifferently write $A \otimes B \otimes C$ to mean $A \otimes (B \otimes C)$ or $(A \otimes B) \otimes C$ and safely ignore the associator isomorphisms. Similarly, we will always omit ρ and λ by identifying A with either $A \otimes I$ or $I \otimes A$.

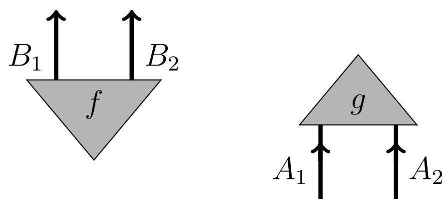
Pictorially, the tensor product of morphisms is represented by drawing their diagrams next to each other horizontally:



However, not all morphisms need to split in this way - in a monoidal category, the depiction of a generic morphism $f : A_1 \otimes \cdots \otimes A_n \to B_1 \otimes \cdots \otimes B_m$, with n and m, two (not necessarily equal) natural numbers, is represented by a box with n input and m output wires:



Specific attention is given to the unit I. The object is represented by no wire, i.e., by the empty diagram. Thus morphisms $f : I \to B_1 \otimes \cdots \otimes B_m$ and $g : A_1 \otimes \cdots \otimes A_n \to I$ are drawn as follows



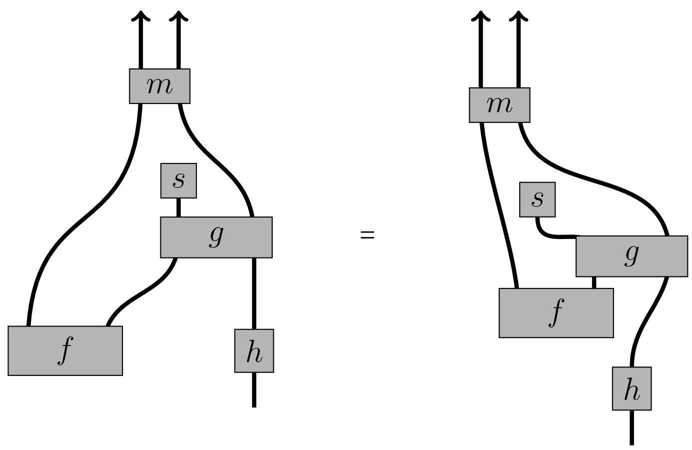
and are called *states* and *co-states* respectively.

The graphical calculus of monoidal categories is consistent and complete for the theory of monoidal categories, meaning that every equation between morphisms that can be derived from the axioms defining a monoidal category, holds if and only if it holds in the graphical language, up to *planar* graph isomorphism, .

Theorem 2.1.1. *An equation follows from the axioms of monoidal categories if and only if it can be derived, up to planar deformation in the corresponding graphical language.*

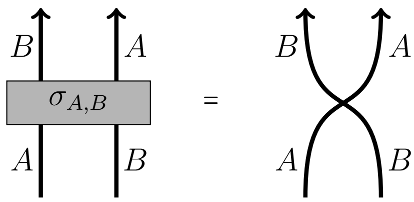
Proof. [30, Theorem 1.2] □

This theorem states, in simple terms, that we are allowed to move around boxes and bend wires as we wish - only the topology of the graph matters, i.e. the way in which boxes and wires are *connected.* However, the adjective *planar* above indicates that we may *not* cross or uncross any two wires when rearranging the graph.

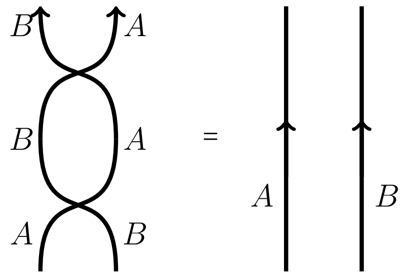


We will often omit the labels on the wires when no ambiguity can arise.

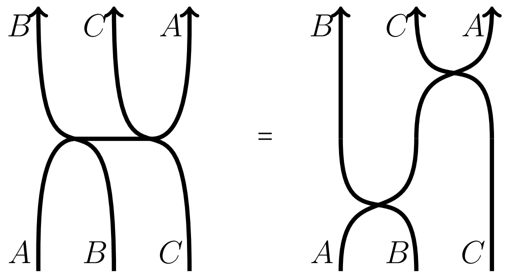
Finally, a monoidal category is called *symmetric* if we have an isomorphism $\sigma_{A,B}$: $A \otimes B \to B \otimes A$ for all objects A and B satisfying certain naturality and coherence conditions (see [42]). In the graphical calculus, symmetry is depicted by two wires crossing (and often called a *swap*):



And the coherence conditions may be drawn as



and



Similarly, Joyal and Street proved a coherence theorem for the graphical calculus of symmetric monoidal categories:

Theorem 2.1.2. *An equation follows from the axioms of symmetric monoidal categories if and only if it can be derived, up to graph isomorphism in the corresponding graphical language.*

Proof. [30, Theorem 2.3] □

In linguistics, categories that model syntax are, in general, *not* symmetric. However, the category of (finite dimensional) Hilbert spaces and linear maps, that will play an important part in the rest of this dissertation, is symmetric monoidal.

2.1.3 Evaluation in closed monoidal categories

The second step in providing a linguistic interpretation of morphisms in a category is to define a notion of evaluation that behaves as a parser for correctly typed strings of words.

Intuitively, imagine that we have a model of syntax, a category, in which all grammatically correct sentences have a single type S and all noun phrases are of type N. We want a type ${}^iV$ for intransitive verbs that reflects the fact that the juxtaposition of a noun phrase and such a verb is a correct sentence (therefore of type S). Moreover we want this to be reflected by a morphism $\mathrm{Eval} : N \otimes {}^iV \to S$ in our category. In a sense that we will make precise below, we want the verb type ${}^iV$ to behave as something that takes as input a noun phrase and outputs a sentence. Thus we want the type itself to behave as a morphism $N \to S$, in a suitable sense captured by the evaluation process:

$$N \otimes (N \to S) \xrightarrow{\mathrm{Eval}} S$$

Mathematically, these notions come to life in the structure of *closed monoidal category* that we define below.

Definition 2.1.3. *A left closed monoidal category is a monoidal category* $(\mathcal{C}, \otimes)$ *such that, for all pairs of objects* A *and* B*, there exists an object* $A \Rightarrow B$ *and a morphism* $Eval^l_{A,B} : A \otimes (A \Rightarrow B) \to B$ *satisfying the following universal property: for every morphism* $f : A \otimes C \to B$*, there exists a unique morphism* $\Pi^l(f) : C \to A \Rightarrow B$ *that makes the following diagram commute*

$$\begin{tikzcd}
A \otimes C \arrow[r, "1_A \otimes \Pi^l(f)"] \arrow[dr, "f"'] & A \otimes (A \Rightarrow B) \arrow[d, "Eval^l_{A,B}"] \\
& B
\end{tikzcd}$$

This last property gives the most general map that corresponds to the evaluation needed.

As we see, this evaluation performs a type reduction from the left. In linguistics, we will also need an evaluation that works in the other direction: an adjective, for instance, can be seen to have a type that, when paired with a noun phrase on the right, reduces to another noun phrase. This corresponds to a type $N \Leftarrow N$ and an evaluation map $\mathrm{Eval} : (N \Leftarrow N) \otimes N) \to N$.

To formalise the previous remark, we introduce the dual notion of *right* closed monoidal category: a monoidal category such that for all pairs of objects A and B, there exists an object $A \Leftarrow B$ and a morphism $\mathrm{Eval}^r_{A,B} : (A \Leftarrow B) \otimes B \to A$ satisfying a universal property that can be deduced from the previous definition.

A category that is both left and right closed monoidal is said to be *bi-closed* monoidal. In the rest of this dissertation, all categories will be bi-closed so we will simply call them closed monoidal.

In summary, the object $A \Rightarrow B$ equipped with the associated evaluation map can be seen as an internalisation, in the language of monoidal categories, of morphisms from A to B: given a morphism $f : A \to B$, we can form its (left) *name* $\ulcorner f \urcorner^l : I \to A \Rightarrow B$ that is simply $\Pi^l(f)$ in the notation above. A similar notion can be defined for the right closure of the tensor.

Now, following the progression of the previous paragraphs we would like to extend the expressiveness of the graphical calculus to closed monoidal categories. An attempt in this direction has been developed by Baez and Stay [4], but their calculus relies on placing rigidity constraints on the graphical calculus associated to a less general kind of category: a compact closed category.

Not only will compact closed categories be sufficiently expressive for the treatment of all applications in the rest of this dissertation, they will allow us to borrow a quantum physical formalism that will play a pivotal role in the next two chapters.

However, it should be noted that Coecke, Grefenstette and Sadrzadeh [15] extended the model to closed monoidal categories, departing form the compact setting in the hope of giving a compositional account of the meaning of sentences parsed by complex formal grammars such as Combinatorial Categorial Grammars (CCGs) or the Lambek-Grishin calculus.

2.1.4 A more convenient framework: compact closedness

This is the last step in our attempt to give a categorical account of syntax. In this paragraph, we will introduce a mathematical structure that further refines the categories that we explored previously, examples of which will provide concrete grammars able to parse simple natural language sentences. Furthermore, we will extend the graphical calculus of monoidal categories to account for - in purely diagrammatic form - the way in which the meaning of individual words come together to produce the meaning of sentences. Of course the concept of meaning is arbitrary as long as we do not fix the semantics of these words - this will be the task of the next sections.

Definition 2.1.4. *A* compact closed category *is a (without loss of generality, strict) monoidal category* $(\mathcal{C}, \otimes)$ *with unit* I *in which each object* A *has a left and right dual, that is, two objects* A^l *and* A^r *equipped with two morphisms each, called the unit and* *the co-unit:*

$$A^l \otimes A \xrightarrow{\epsilon^l} I \xrightarrow{\eta^l} A \otimes A^l$$

$$A \otimes A^r \xrightarrow{\epsilon^r} I \xrightarrow{\eta^r} A^r \otimes A$$

such that all the following triangles commute:

$$\begin{tikzcd}
A \arrow[d, "\eta^l \otimes 1_A"'] \arrow[dr, "1_A"] & \\
A \otimes A^l \otimes A \arrow[r, "1_A \otimes \epsilon^l"'] & A
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[r, "1_A \otimes \eta^r"] \arrow[dr, "1_A"'] & A \otimes A^r \otimes A \arrow[d, "\epsilon^r \otimes 1_A"] \\
& A
\end{tikzcd}$$

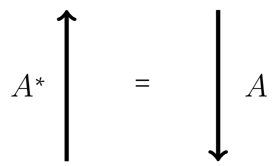
$$\begin{tikzcd}
A^l \arrow[d, "1_{A^l} \otimes \eta^l"'] \arrow[dr, "1_{A^l}"] & \\
A^l \otimes A \otimes A^l \arrow[r, "\epsilon^l \otimes 1_{A^l}"'] & A^l
\end{tikzcd}
\qquad
\begin{tikzcd}
A^r \arrow[r, "\eta^r \otimes 1_{A^r}"] \arrow[dr, "1_{A^r}"'] & A^r \otimes A \otimes A^r \arrow[d, "1_{A^r} \otimes \epsilon^r"] \\
& A^r
\end{tikzcd}$$

These last conditions are called the *yanking* equalities. A useful property states that:

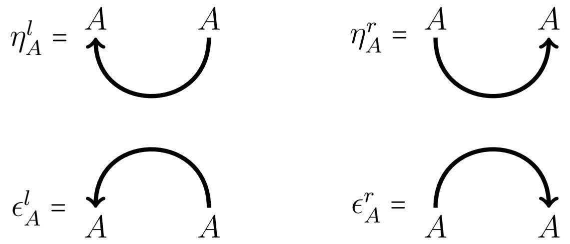
Lemma 2.1.1. *In a compact closed category, duals are unique, up to canonical isomorphism.*

Proof. If A admits another left dual B with $\eta : I \to A \otimes B$ and $\epsilon : B \otimes A \to I$, the morphism $(\epsilon_A^l \otimes 1_B) \circ (1_{A^l} \otimes \eta) : A^l \to B$ is an isomorphism with inverse $(\epsilon \otimes 1_{A^l}) \circ (1_B \otimes \eta_A^l) : B \to A^l$. The proof is carried out similarly for the right dual. ☐

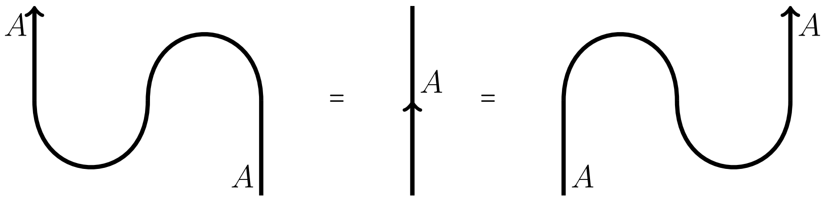
In what follows, we will sometimes write A^* in a statement that applies both to the left and right dual. This notation is consistent with that of *symmetric* compact closed categories in which, as we will see, both notions collapse to a single dual. Graphically, duals are both represented as an A-labelled wire, running from top to bottom:



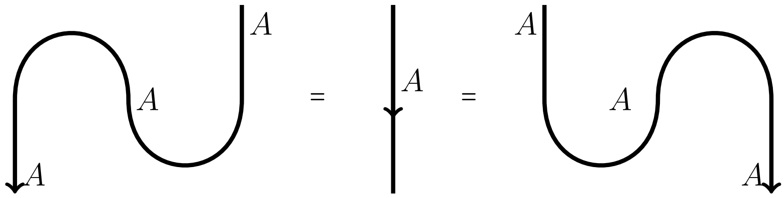
Units and co-units are pictured as directed cups and caps:



In their diagrammatic form, the yanking equations become self explanatory; it is simply a matter of pulling the wire straight:



and



Now we will define the evaluation maps that were at the heart of the previous paragraphs:

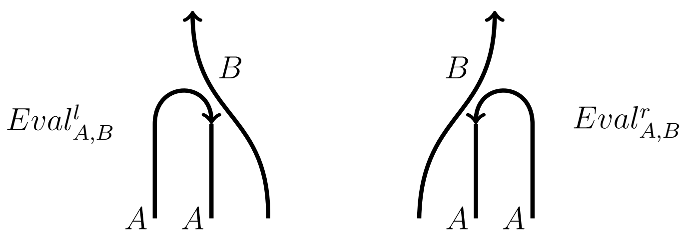
Proposition 2.1.1. *Every compact closed category is bi-closed monoidal.*

Proof. For each pair of objects A, B, the object $A \Rightarrow B$ and $B \Leftarrow A$ are defined respectively as $A^r \otimes B$ and $B \otimes A^l$ with the corresponding evaluation maps given by the following morphisms:

$$A \otimes (A^r \otimes B) \xrightarrow{\epsilon^r_A \otimes 1_B} B$$

$$(B \otimes A^l) \otimes A \xrightarrow{1_B \otimes \epsilon^l_A} A$$

Note that the parenthesis are here for clarity only - the coherence theorem for monoidal categories guarantees that the different pairings are identical, up to natural isomorphism. The following diagrams represents the evaluation maps in the graphical calculus of compact closed categories:

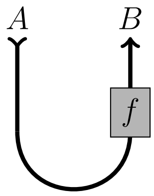


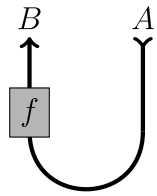
Then, we need to check that the evaluation maps as defined above satisfy the required universal property. Let C be an object of $\mathcal{C}$ and $f : A \otimes C \to B$ a morphism. Then, clearly the following diagram commutes

$$\begin{tikzcd}
A \otimes C \arrow[r, "f"] \arrow[drr, "f"'] & B \arrow[r, "\eta^r_A \otimes 1_B"] \arrow[dr, "1_B"] & A \otimes A^r \otimes B \arrow[d, "\epsilon^r \otimes 1_B"] \\
& & B
\end{tikzcd}$$

because the rightmost triangle commutes by the appropriate yanking condition. The proof for the right adjoint can be carried out similarly. □

Additionally, in a compact category, we can give a graphical representation to the names $\ulcorner f \urcorner^l : I \to A^r \otimes B$, $\ulcorner f \urcorner^r : I \to B \otimes A^l$ of a morphism $f : A \to B$

$\ulcorner f \urcorner^l = (1_{A^r} \otimes f) \circ \eta^l_{A^r} =$ 

$\ulcorner f \urcorner^r = (f \otimes 1_{A^l}) \circ \eta^r_{A^l} =$ 

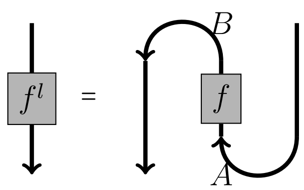
Furthermore, in a compact category, we can define dual morphisms such that the duality on objects extends to a contravariant functor.

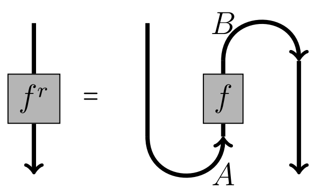
Definition 2.1.5. *In a compact closed category, the duals (also called the transposes) $f^* : B^* \to A^*$ of $f : A \to B$ are the morphisms*

$$f^l : B^l \xrightarrow{1_{B^l} \otimes \eta^l_A} B^l \otimes A \otimes A^l \xrightarrow{1_{A^l} \otimes f \otimes 1_{B^l}} B^l \otimes B \otimes A^l \xrightarrow{\epsilon^l_B \otimes 1_{A^l}} A^l$$

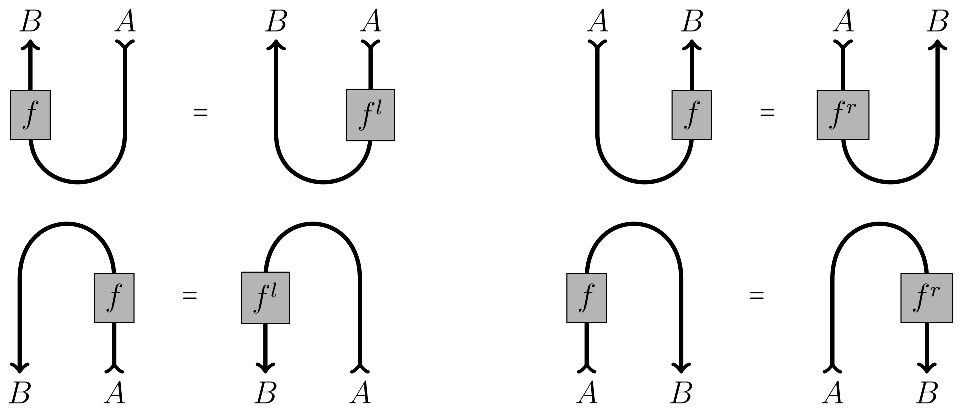
$$f^r : B^r \xrightarrow{\eta^r_A \otimes 1_{B^r}} A^r \otimes A \otimes B^r \xrightarrow{1_{A^r} \otimes f \otimes 1_{B^r}} A^r \otimes B \otimes B^r \xrightarrow{1_{A^r} \otimes \epsilon^r_B} A^r$$

In the graphical notation, the duals of a map are depicted using cups and caps to bend the wires in the other direction:

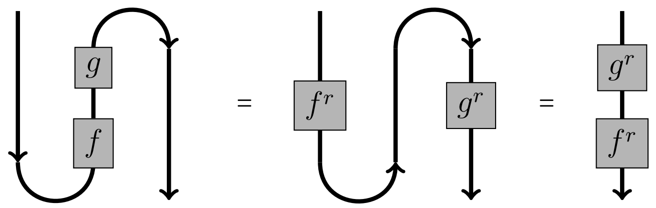




It follows that we can slide boxes along the cups and caps:



With this notation, it is clear that the mapping $f \mapsto f^*$ preserves composition, i.e. $(g \circ f)^* = f^* \circ g^*$ and that we have a functor. The proof amounts to sliding the boxes along the wires and applying the yanking equality to the remaining piece of wire; here for the right dual:



In addition, this functor preserves the monoidal structure and induces natural isomorphisms $A^{lr} \cong A \cong A^{rl}$.

Lambek pregroups To demonstrate the use of these structures in linguistics we introduce an important example of a compact closed category, due to Lambek [36]: a pregroup grammar. As an algebraic version of compact bilinear logic, it provides a compact closed simplification of his original type logical grammar that admitted a categorical interpretation in the language of closed monoidal categories.

Definition 2.1.6. *A pregroup grammar is a posetal, free compact closed category.*

Let us unravel this abstract definition. A **posetal category** is a category in which every diagram commutes. More formally, it is a category in which there is *at most* one morphism from one object to another; we further require the category to be skeletal, which, in this case, imposes that the only isomorphisms are precisely the identities on each object. As a result, a posetal category is simply a partially ordered set in which the transitivity of the partial order is induced by the composition rule of the underlying category. In such a category there is no need to label the morphisms: the usual right pointing arrow $A \to B$ is replaced by the inequality symbol $A \leq B$.

A **free category** is a category that is generated (in some precise sense) by a collection of objects and morphisms between them. Here, given a partially ordered set $\mathcal{T}$, the free compact closed category $C(\mathcal{T})$ generated by $\mathcal{T}$, contains as objects the elements of the generating set $\mathcal{T}$, their left and right duals and all tensor products thereof. The generating morphisms are precisely the morphisms $a \leq b$ of the partial order of $\mathcal{T}$, as well as the units and co-units associated to each dual, hereafter expressed as inequalities on an object a:

$$a^l \cdot a \leq 1 \leq a \cdot a^l$$

$$a \cdot a^r \leq 1 \leq a^r \cdot a$$

Traditionally, in the language of pregroups, the tensor product is written as $\cdot$, its unit as 1 and the objects in lower case. We will also adopt this convention.

The pregroup $\mathbf{Pr}(\mathcal{T})$ generated by $\mathcal{T}$ is the posetal version of $C(\mathcal{T})$, that is, the category $C(\mathcal{T})$ in which all morphisms with the same domain and codomain are identified. Note that the yanking equalities are immediately satisfied as a consequence of the partially ordered nature of the category: the only morphism from an object to itself is the identity.

From these (in)equalities, we can prove that the unit is self-dual, i.e, $1^l = 1 = 1^r$ and that duals reverse the order: for a and b such that $a \leq b$, we have $b^l \leq a^l$ and $b^r \leq a^r$. Moreover, right and left duals cancel out, $a^{l^r} = a = a^{rl}$ and dualising interacts simply with the tensor product: $(a \cdot b)^l = b^l \cdot a^l$ and $(a \cdot b)^r = b^r \cdot a^r$.

Proof. The verification of these properties can all be found in [36]. □

As stated earlier, applied to the analysis of syntax in natural language, objects of a pregroup correspond to grammatical types. Given a lexicon of words with their respective types, the tensor product denotes the juxtaposition of types according to the structure of possible strings of words. We call atomic types the generating types of the grammar; simple types the atomic types and their duals; and relational types any other type that is not simple. Any object of a pregroup can be written as a finite juxtaposition of simple types.

Given a pregroup $\mathbf{Pr}(\mathcal{T})$, a lexicon of words is a choice of morphisms of the form $\overline{word} : 1 \leq t$, where t can be any type, simple or relational. By convention, the identity $1 \leq 1$ represents the empty string. A sentence is the tensor of these morphisms: for words w_i with morphisms $\overline{w_i} : 1 \leq t_i$, we obtain the sentence $\overline{w_1 w_2 \ldots w_n} = \overline{w_1} \cdot \overline{w_2} \ldots \overline{w_n} : 1 \leq t_1 \cdot t_2 \ldots t_n$.

We assume that the atomic types of the pregroup contain a designated type s, the type of well-formed sentences. We call *reduction* any morphism $t \leq s$ that factors precisely through evaluations maps, that is, involving strictly co-units and inequalities of simple types. Additionally, we say that a sentence $1 \leq t$ is *grammatical* if there exists a reduction $t \leq s$.

Let us consider an example: "My fake plants died because I did not pretend to water them" [1]. The atomic types that we will use to parse the sentence are n, for nouns, and s, for declarative sentences. The type assignments are presented in the following table:

My	fake	plants	died	because	I
$n \cdot n^l$	$n \cdot n^l$	n	$n^r \cdot s$	$s^l \cdot s \cdot s^r$	n
did	not	pretend	to	water	them
$n^r \cdot s \cdot s^l \cdot n$	$n^r \cdot s \cdot s^r \cdot s \cdot s^l \cdot n$	$n^r \cdot s \cdot s^l$	n	$n^r \cdot s \cdot n^l$	n

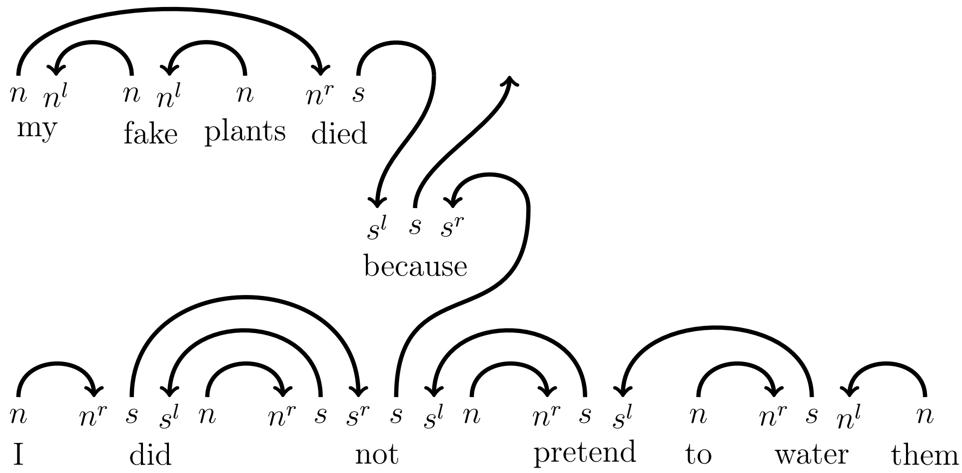
Note that the infinitive marker "to" has the type noun. It is explained by the fact that verbs have types $n^r \cdot s \cdot n^l$ for transitive verbs and $s \cdot n^l$ for intransitive verbs. Hence, the infinitive marker can be seen as eliminating the need for a subject on the left of the verb - taking the place of a noun. It is further justified by the use of the verb "pretend" in the sentence: this verb takes a clause or an infinitive verb phrase as its argument. Consider, for instance, the equivalent phrases "I closed my eyes and pretended *I* was asleep" and "I closed my eyes and pretended *to* be asleep".

The following reduction proves that the sentence "My fake plants died because I did not pretend to water them" is grammatical:

$$
\begin{aligned}
n \cdot n^l \cdot n \cdot n^l \cdot n \cdot n^r \cdot s \cdot s^l \cdot s \cdot s^r \cdot n \cdot n^r \cdot s \cdot s^l \cdot n \cdot n^r \cdot s \cdot s^r \cdot s \cdot s^l \cdot n \cdot \ldots \\
\ldots n^r \cdot s \cdot s^l \cdot n \cdot n^r \cdot s \cdot n^l \cdot n \leq \\
n \cdot 1 \cdot 1 \cdot n^r \cdot s \cdot s^l \cdot s \cdot s^r \cdot 1 \cdot s \cdot s^l \cdot 1 \cdot s \cdot s^r \cdot s \cdot s^l \cdot 1 \cdot s \cdot s^l \cdot 1 \cdot s \cdot 1 = \\
n \cdot n^r \cdot s \cdot s^l \cdot s \cdot s^r \cdot s \cdot s^l \cdot s \cdot s^r \cdot s \cdot s^l \cdot s \cdot s^l \cdot s \leq \\
1 \cdot s \cdot 1 \cdot s^r \cdot s \cdot s^l \cdot s \cdot s^r \cdot s \cdot 1 \cdot 1 = \\
1 \cdot s \cdot s^l \cdot 1 \cdot s = \\
s \cdot s^l \cdot s \leq \\
s \cdot 1 \leq s
\end{aligned}
$$

[1] A quotation by the late, great Mitch Hedberg.

It is easier to picture the reduction process using the graphical calculus introduced earlier: caps connecting dual types represent the application of the co-unit, i.e., a cancellation. Each step of the symbolic reduction above corresponds to one level of nested wires.



The graphical notation provides an intuitive appreciation of the constituencies between the different components of a sentence. Assuming that we know the meaning (however we define the notion of meaning, whether it is truth theoretic or distributional; see section 2.2) of the individual words of our example sentence, the wires connecting them can be understood as delineating a certain flow of information; they picture the mechanisms by which meaning is shared.

Obviously the choice of a wider range of atomic types can change this view. The difficulty of using categorical grammars lies in selecting the right set of types and assigning to words the right relational types, in order to engender *precisely* the grammatical sentences of the language. Here we chose simplicity for our toy example sentence. For more sophisticated approaches we refer the reader to [44] or [37].

Finally, it should be noted that the expressive power of pregroup grammars is equivalent to that of context-free grammars [8]. Attempts to move away from the compact closed setting in order to accommodate more expressive grammars are currently being investigated [15].

Symmetric compact closed categories Pregroups are an example of non-symmetric compact closed category. In linguistics, it is obvious that the order of words matter, however, in the rest of this dissertation, our model of the lexical semantics of words will be an example of a *symmetric* compact closed category. It is therefore useful to explore here how compact closedness interacts with a symmetric monoidal structure.

Proposition 2.1.2. *In a symmetric compact closed category left and right duals are naturally isomorphic.*

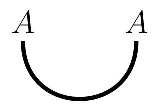
Proof. The following unit and co-unit witness the left dual structure of the right dual.

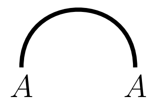
$$\eta_A : I \xrightarrow{\eta^r_A} A^r \otimes A \xrightarrow{\sigma_{A^r,A}} A \otimes A^r$$

$$\epsilon_A : A^r \otimes A \xrightarrow{\sigma_{A^r,A}} A \otimes A^r \xrightarrow{\epsilon^r_A} I$$

One can check, using the coherence of symmetric monoidal categories and the yanking equalities for η^r, ϵ^r that this unit and co-unit satisfy their own set of yanking equalities. Finally we deduce the result from the uniqueness of duals. □

In a symmetric compact closed category, we write the dual of an object A as A^*. Additionally, the graphical notation simplifies considerably: there is only one notion of cups and caps, transpose or name. The unit and co-unit maps are drawn as follows, without any orientation: the direction of the arrow is relative to what it connects.

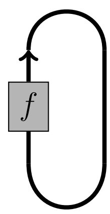
$\eta_A =$ 

$\epsilon_A =$ 

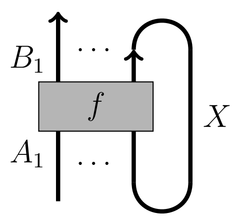
Finally, in a symmetric compact closed category, we can define a notion of *trace*: the trace of $f : A \to A$ is the morphism $\mathrm{Tr}\, f : I \to I$ defined by

$$I \xrightarrow{\eta_A} A \otimes A^* \xrightarrow{f \otimes 1_{A^*}} A \otimes A^* \xrightarrow{\epsilon_A \circ \sigma_{A,A^*}} I$$

and pictured as

$\mathrm{Tr}\, f =$ 

Similarly, for a morphism $f : A_1 \otimes \cdots \otimes A_n \otimes X \to B_1 \otimes \cdots \otimes B_m \otimes X$ we define a *partial* trace along X by

$Tr_X(f) =$ 

As another example of compact closed category, we will consider the category $\mathbf{FVect}_{\mathbb{F}}$ of finite dimensional vector spaces and linear maps over the field $\mathbb{F}$. The tensor product is the usual tensor of vector spaces whose unit is $\mathbb{F}$. This tensor is symmetric since we have a natural isomorphism $V \otimes W \cong W \otimes V$ for all vector spaces V, W satisfying the usual coherence conditions. The dual V^* of a space V is the vector space of all linear functionals on V, i.e. linear maps $V \to \mathbb{F}$ equipped with the vector space structure given by point-wise addition and multiplication by a scalar. The dual of a map $f : A \to B$ is its transpose $f^* : B^* \to A^*$ defined by $f^*(\varphi) = \varphi \circ f$; its name is its matrix $I \to A^* \otimes B$; its trace $\mathrm{Tr}\, f$ is the usual notion of the sum of the diagonal coefficients of its matrix representation. The caps are the pairing maps $V^* \otimes V \to \mathbb{F}$ defined by $\phi \otimes v = \phi(v)$, extended by linearity. The cups are the maps $\mathbb{F} \to V \otimes V^*$ given by $1 \mapsto \sum_i e_i \otimes e_i^*$ where $\{e_i\}$ is a basis of V. It looks like this map depends on a choice of coordinate but it is defined naturally as the dual notion to the co-unit, in the sense that it is the unique map that satisfies the yanking equalities (for a given co-unit). Another way to see it is by considering the isomorphism $V^* \otimes V \cong End(V)$ defined by $\phi \otimes v \mapsto \phi(-)v$. Then, the cups are the inverse image of the identity map, whose definition is clearly invariant (however, the inverse of this isomorphism cannot be written down without selecting a basis first).

Finally, the following theorem, due to Kelly and Laplaza [33], guarantees that the manipulations added to the graphical calculus of monoidal categories by the introduction of caps and cups corresponds to equations in the theory of compact closed categories:

Theorem 2.1.3. *This diagrammatic language is sound and complete for the equational theory of (symmetric) compact closed categories.*

Proof. See [33] □

The categorical approaches that purport to represent the grammatical structure of a language do not provide a model of meaning for the individual words of the same language; they offer an account of how the meaning of the parts of a sentence are pieced together to form the meaning of the sentence as a whole but cannot account for the semantics of the parts themselves. To resolve this issue, we will first describe a suitable category in which meaning can modeled. The advantage of describing meaning in a categorical setting is that this model of lexical semantics shares the compositional structure of type logical grammars (in particular, as we will see, they will both share a compact closed structure). Subsequently, an appropriate functor will bridge the gap between the categorical accounts of grammar and semantics.

2.2 A category of meaning

So far, we have seen that the compact closed structure of pregroup grammars allowed us to visualise the structure of correctly formed sentences and highlight the flow of meaning between words. Yet no rigorous account of meaning has been provided. If the graphical calculus for compact closed categories admits an intuitive interpretation in terms of information flow between the words of a sentence, what is the information that flows? The grammatical model defined previously gives a powerful mathematical account of Frege's principle of compositionality. Its strength is that it can model any form of semantic information, as long as it can be captured in a compact closed category. How meaning can be captured in such a category is the focus of this section.

2.2.1 Requirements on an abstract model of meaning

Trying to define what meaning is in language is the work of philosophers. In what follows, we will concentrate on a particular model of meaning that can be derived from one simple requirement. Obviously such a model can only give a partial answer to the deep question of meaning; an answer within a limited language game, to adopt the words of Wittgenstein [57]. The end all of computational linguistics is to program a machine that *behaves as though* it understood human language. We insist on the behavioural character of this objective to avoid the vague and unmotivated question of whether a computer can truly understand language as we do.

Independently of our definition of meaning, an essential step towards the objective stated above is for a machine to recognise when two expressions have the same meaning. This is simply a requirement of *consistency*: if we want our computer to react appropriately to any natural language expression we communicate, however we want it to react, the action we wish it to perform as a result of *understanding* this expression has to be the same for all synonymous expressions.

This simple principle entails that the category in which we model meaning needs to come equipped with some measure of similarity. As we will see in section 2.2.2, one possible way to obtain such a measure in any given closed monoidal category, is to encode degrees of similarities in the unit object, I. Then, given two expressions of the same type[1], $s : I \to T$ and $t : I \to T$, we want to obtain a map $m(s,t) : I \to I$ that quantifies the similarity between s and t. It seems that if we give ourselves a way to get a map $Ft : T \to I$ we could define $m(s,t) := Ft \circ s : I \to I$. Ideally, we want the assignment $F : t \mapsto FT$ to preserve the categorical structure, that is, we want F to be a functor. In addition the constraint $Ft : FT \to FI$ and $Ft : T \to I$ shows that F should assign each type to itself.

Note that the previous reasoning is in no way mathematically rigorous; it simply purports to motivate the content of section 2.2.2, in which we will define precisely the functor F (written †, pronounced dagger).

2.2.2 From abstract to concrete models: dagger categories

Ultimately, the concept for which we are looking is a generalisation of the scalar product: a notion of angle between vectors in a Hilbert space. The following definitions generalise this notion, first to arbitrary categories, and then to compact closed structures. This notion was first introduced by Abramsky and Coecke [2] for compact categories and the term *dagger* was coined by Selinger [51] for arbitrary categories.

Definition 2.2.1. *A dagger category, or †-category, is a category $\mathcal{C}$ equipped with an involutive, identity-on-objects functor $\dagger : \mathcal{C}^{op} \to \mathcal{C}$.*

In more concrete terms, the dagger associates to every morphism $f : A \to B$ a morphism $f : B \to A$ called its *adjoint* with $f^{\dagger\dagger} = f$ and $1_A^\dagger = 1_A$. We say that an isomorphism f is *unitary*, if its inverse f^{-1} is equal to $f^\dagger$.

Assuming that the tensor unit of $\mathcal{C}$ allows us to encode numerical information, measuring how close the transitive verbs *love* $: I \to N^l \otimes S \otimes N^r$ and *like* $: I \to N^l \otimes S \otimes N^r$ are, is simply composing *love* with *like*†:

$$I \xrightarrow{\ \ love\ \ } N^l \otimes S \otimes N^r \xrightarrow{\ \ like^\dagger\ \ } I$$

This remarkably simple definition, combined with the previous notion of compact closedness will prove to be an adequate model of meaning, according to our self-imposed constraints. First, we state the coherence conditions for which the dagger structure is compatible with a monoidal structure:

[1] At first glance, it makes little sense to compare expressions of different type, although I am sure that within the infinite diversity of language there exist two expressions that will prove this to be too simplistic.

Definition 2.2.2. *[51] A (symmetric) monoidal †-category is a (symmetric) monoidal category $\mathcal{C}$ with a dagger $\dagger : \mathcal{C} \to \mathcal{C}$ that verifies, for all $f : A \to B$ and $g : C \to D$,*

$$(f \otimes g)^\dagger = f^\dagger \otimes g^\dagger$$

and all whose coherence isomorphisms (see definition 2.1.2) - including the swap σ if the category is symmetric monoidal - are unitary.

As a result, compact categories with a dagger functor, have a simpler duality structure than more general compact categories:

Proposition 2.2.1. *In a compact closed †-category, left duals are also right duals.*

Proof. Let A be an object, A^l its left dual with unit η^l_A and co-unit ϵ^l_A. We can define $\eta^r_A := \epsilon^{l\dagger}_A$ and $\epsilon^r_A := \eta^{l\dagger}_A$ that give a unit and co-unit for a right dual of A. The yanking equalities are verified since

$$\begin{tikzcd} A \arrow[r, "\eta^l_A \otimes 1_A"] \arrow[d, "1_A \otimes (\epsilon^l_A)^\dagger"'] \arrow[dr, "1_A"] & A \otimes A^* \otimes A \arrow[d, "1_A \otimes \epsilon^l_A"] \\ A \otimes A^* \otimes A \arrow[r, "(\eta^l_A)^\dagger \otimes 1_A"'] & A \end{tikzcd} \qquad \begin{tikzcd} A^* \arrow[r, "1_{A^*} \otimes \eta^l_A"] \arrow[d, "(\epsilon^l)^\dagger_A \otimes 1_{A^*}"'] \arrow[dr, "1_{A^*}"] & A^* \otimes A \otimes A^* \arrow[d, "\epsilon^l_A \otimes 1_A"] \\ A^* \otimes A \otimes A^* \arrow[r, "1_{A^*} \otimes (\eta^l_A)^\dagger"'] & A^* \end{tikzcd}$$

commute: the lower triangles commute by the contravariance of † and the fact that it preserves the tensor product; the upper triangles are simply the yanking conditions for the left dual. □

By lemma 2.1.1, left and right duals are isomorphic in a compact closed †-category. As in symmetric monoidal categories, we write A^* for the dual of an object A and we have $A^{**} \cong A$.

However, the functors $(-)^l$ and $(-)^r$ are not naturally isomorphic. To this effect, several equivalent conditions are given in [56]. In this dissertation, we will use the stronger condition of symmetry to enforce the natural isomorphism:

Definition 2.2.3. *A †-compact category is a symmetric compact closed category such that, for all object A, $\eta_A = \epsilon^\dagger_A \circ \sigma_{A*,A}$,*

$$\begin{tikzcd}
I \arrow[r, "\epsilon_A^\dagger"] \arrow[dr, "\eta_A"'] & A^* \otimes A \arrow[d, "\sigma_{A^*,A}"] \\
 & A \otimes A^*
\end{tikzcd}$$

i.e., such that the dagger of the unit is the co-unit, up to natural isomorphism.

In particular, this coherence condition imposes that the functors $(-)^\dagger$ and $(-)^*$ commute. This can be verified by checking the commutativity of the following diagram:

$$\begin{tikzcd}
 & A^* \otimes B^* \otimes B \arrow[d, "1_{A^*} \otimes \sigma_{B^*,B}"] \\
A^* \arrow[ur, "1_{A^*} \otimes \epsilon_B^\dagger"] \arrow[r, "1_{A^*} \otimes \eta_B"'] \arrow[d, "(f^*)^\dagger = (f^\dagger)^*"'] & A^* \otimes B \otimes B^* \arrow[d, "1_{B^*} \otimes f^\dagger \otimes 1_{A^*} = (1_{B^*} \otimes f \otimes 1_{A^*})^\dagger"] \\
B^* & A^* \otimes A \otimes B^* \arrow[l, "\epsilon_A \otimes 1_{B^*}"'] \\
 & A \otimes A^* \otimes B^* \arrow[u, "\sigma_{A,A^*} \otimes 1_{B^*}"'] \arrow[ul, "\eta_A^\dagger \otimes 1_{B^*}"]
\end{tikzcd}$$

Note that we write "†-compact category", not the weaker "compact †-category" that is simply a †-category with a dagger.

Since † and $(-)^*$ are both involutive and commute, we defined a new involutive functor, called the *conjugation* functor:

Definition 2.2.4. *In a †-compact category, we set* $(-)_* := (-)^{\dagger *} = (-)^{* \dagger}$.

So far, our approach has been abstract as we have been developing a high level categorical account of meaning. To build a model from data and perform computations, we need a concrete instanciation. The next theorem, due to Selinger, lets us finally reap the result of our efforts.

Theorem 2.2.1. *Finite dimensional Hilbert spaces are complete for †-compact closed categories.*

Proof. Cf. [52] □

First, the category of finite dimensional (complex) Hilbert spaces, **FdHilb** is a †-compact closed category in which morphisms are linear maps, the dagger of a map $f : A \to B$ is its adjoint, i.e. the unique map $f^\dagger : B \to A$ such that $\langle fv|w\rangle_B = \langle v|f^\dagger w\rangle_A$ for all $v \in A$, $w \in B$, where $\langle\cdot|\cdot\rangle$ is used to denote the inner products[1] of A and B. The monoidal compact structure is inherited from the underlying vector space structure: the tensor product is the usual tensor product of vector spaces whose unit is the field[2] $\mathbb{C}$. As we have seen, contrary to the tensor of pregroups, this tensor is symmetric. Thus, the compact structure is symmetric and left and right duals collapse to a single dual.

Moreover, the inner product induces a self-dual structure on every space: the natural isomorphism $A \cong A^*$ is given by $x \mapsto \langle x|\cdot\rangle$. Self duality in **FdHilb** justifies the introduction of a new notation for vectors, called Dirac notation. One writes $|x\rangle$ for a vector and $\langle x|$ for its corresponding co-vector. Notice that this corresponds to the notion of states and co-states in a monoidal category, but sideways. In what follows we will indifferently refer to (co-)vectors as (co-)states. Hence, the pairing of a co-vector $\langle y|$ with a vector $|x\rangle$ is simply the inner product $\langle y|x\rangle$. Finally, self duality simplifies the compact structure with the following unit and co-unit on each space A (often called Bell states in the quantum physics literature):

$$\epsilon_A : A \otimes A \xrightarrow{\cong} A^* \otimes A \to \mathbb{C}$$

$$\eta_A : \mathbb{C} \to A \otimes A^* \to A \otimes A$$

where ϵ_A is simply the inner product.

Theorem 2.2.1 means that an equation follows from the axioms of dagger compact closed categories if and only if it holds in the theory of finite dimensional Hilbert spaces. In other words, **FdHilb** is the *concrete* model of meaning that we have been seeking. Indeed, this idea is not new. Representing words as vectors in a vector space with a canonical scalar product is the central theme of so-called *distributional* models in natural language processing.

2.2.3 The distributional model of meaning as a concrete model

First introduced formally by Firth [20], distributional models of meaning constitute a mathematically rigorous realisation of Wittgenstein's view: "meaning is use" [57].

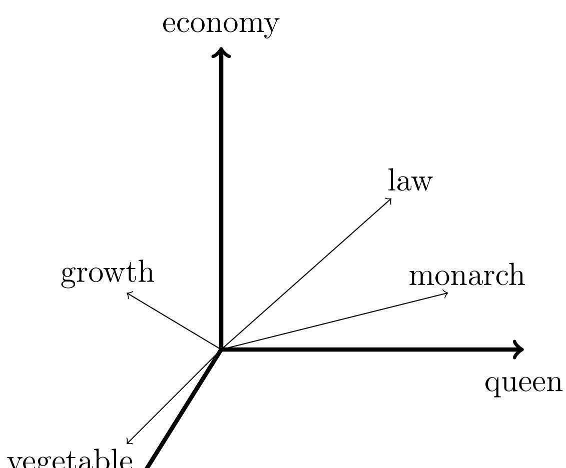
[1] A nondegenerate sesquilinear form.

[2] We will always assume the theoretical luxury of working with the algebraically closed field $\mathbb{C}$. However, in all concrete applications, the distributional model space obtained from real world data is a *real* vector space.

The prime intuition is that words appearing in similar contexts must have a similar meaning. Typically, a concrete instance is built by creating a high dimensional vector space with a fixed orthonormal basis, from a large corpus of texts like the British National Corpus. The basis is a set of designated content bearing words against which we will measure all the other words occurring in the corpus. In practice, these words can be the set of lemmatised words of the corpus, a set of manually designated words when the target application needs to be specifically tailored to a technical domain (for example medical journals) or simply the most occurring words (to the exclusion of prepositions, articles, pronouns, auxiliaries and all words whose role is purely grammatical). Furthermore, the content bearing words can be extracted from the corpus using Latent Semantic Analysis and, more precisely, the linear algebraic algorithm of single value decomposition.

Next, the coordinates of a word w in the predefined basis is obtained by counting how many times each basis words has appeared in a window of a few words that surround w. There are more complex methods to assign coordinates but they are all refinements of this simple count. For example, counts are often normalised computing the Term-Frequency divided by the Inverse-Document-Frequency. This method keeps track of how often basis words appear in a specific document and how often they appear in the entire corpus in order to weight them relatively to each other. Other methods include entropy or point-wise mutual information (for a survey of those methods, see [55]).

A vector based representation with a given basis allows us to use the canonical scalar product associated to that basis and thus obtain a measure of proximity of meaning, as studied in the abstract categorical setting of the previous section.



Typically, vectors are normalised and all that matters is the cosine of the angle between two meaning-vectors[1].

Notwithstanding its success, the distributional approach does not provide a definitive model of natural language semantics since it does not scale up to give an interpretation of sentences. Indeed, if it can account for the meaning of individual words, it fails to offer an intuitive way of composing their meaning as it ignores grammatical structure. As we have seen earlier, type logical models are orthogonal to this approach: the meaning of words is precisely their grammatical role.

Recent work (namely [9], [15] and [24]) turns to category theory to marry these two approaches and extend the distributional model of meaning from individual words to expressions, thus effectively making the compositional distributional model of meaning first proposed in [11], a reality.

The key idea is to construct a functor between a category of grammatical types and reductions and a category of meaning (here **FdHilb**), both compact closed. The closed structure is essential to obtain evaluation maps that reduce correctly typed sentences to a single type, in which sentences sit as vectors and can be compared. As pointed out in [15], this corresponds to a quantisation of natural language semantics and, as such, a generalisation of Montague style truth-theoretic semantics [43].

2.3 Functorial semantics

2.3.1 Monoidal functors

Before building the functor described in the previous paragraph, a few category theoretic considerations are in order.

Definition 2.3.1. *A monoidal functor is a functor $F : \mathcal{C} \to \mathcal{D}$ between two monoidal categories $(\mathcal{C}, \heartsuit)$ and $(\mathcal{D}, \spadesuit)$ such that*

i) there exists a natural transformation $\phi_{A,B} : F(A) \spadesuit F(B) \to F(A \heartsuit B)$

ii) and a morphism $\phi_I : I_{\mathcal{D}} \to F I_{\mathcal{C}}$.

satisfying the following commutativity conditions for all objects A, B and C in $\mathcal{C}$:

[1]Thus, we are working with the projective space induced by the vector space.

$$\begin{tikzcd}
(FA\spadesuit FB)\spadesuit FC \arrow[r] \arrow[d, "\phi_{A,B}\spadesuit 1_{FC}"'] & FA\spadesuit(FB\heartsuit FC) \arrow[r, "1_{FA}\spadesuit\phi_{B,C}"] & FA\spadesuit F(B\heartsuit C) \arrow[d, "\phi_{A,B\heartsuit C}"] \\
F(A\otimes B)\spadesuit FC \arrow[r, "\phi_{A\heartsuit B,C}"'] & F((A\heartsuit B)\heartsuit C) \arrow[r] & F(A\heartsuit(B\heartsuit C))
\end{tikzcd}$$

and

$$\begin{tikzcd}
FA\spadesuit I_{\mathcal{D}} \arrow[r, "FA\spadesuit\phi_I"] \arrow[d] & FA\spadesuit FI_{\mathcal{C}} \arrow[d, "\phi_{A,I_{\mathcal{C}}}"] \\
FA & F(A\heartsuit I_{\mathcal{C}}) \arrow[l]
\end{tikzcd}
\qquad
\begin{tikzcd}
I_{\mathcal{D}}\spadesuit FB \arrow[r, "\phi_I\spadesuit FB"] \arrow[d] & FI_{\mathcal{C}}\spadesuit FB \arrow[d, "\phi_{I_{\mathcal{C}},B}"] \\
FB & F(I_{\mathcal{C}}\heartsuit B) \arrow[l]
\end{tikzcd}$$

In the commutative diagrams above, the unlabeled morphisms correspond to structural morphisms in the monoidal categories $\mathcal{C}$ and $\mathcal{D}$. In strict monoidal categories, these morphisms collapse simply to equalities.

We say that a functor is *strongly monoidal* if the morphism ϕ_I and the natural transformation ϕ are invertible. If they are identities, the functor is strictly monoidal.

Proposition 2.3.1. *A strongly monoidal functor F between two compact closed categories $\mathcal{C}$ and $\mathcal{D}$ preserves the compact structure, that is, $F(A^l) \cong (FA)^l$ and $F(A^r) \cong (FA)^r$ for all objects A of $\mathcal{C}$.*

Proof. It is sufficient to consider the case of the left dual. The unit and co-unit pair is given by the following morphisms

$$I \xrightarrow{\phi} FI \xrightarrow{F\eta^l_A} F(A\otimes A^l) \xrightarrow{\phi^{-1}_{A,A^l}} FA\otimes F(A^l)$$

$$F(A^l)\otimes FA \xrightarrow{\phi_{A^l,A}} F(A^l\otimes A) \xrightarrow{F\epsilon^l_A} FI \xrightarrow{\phi_I^{-1}} I$$

Using the yanking equalities for the left dual of FA and the fact that the morphisms that make up the unit are inverses of those that make up the co-unit, we obtain a yanking equality for the two morphisms above.

Finally, as we have seen, duals are unique up to canonical isomorphism. □

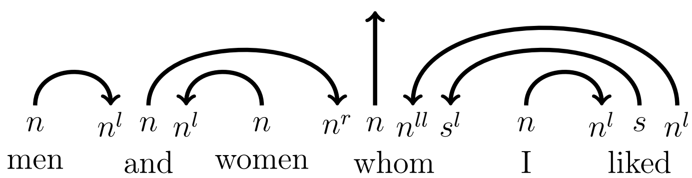
2.3.2 A quantisation functor for grammar

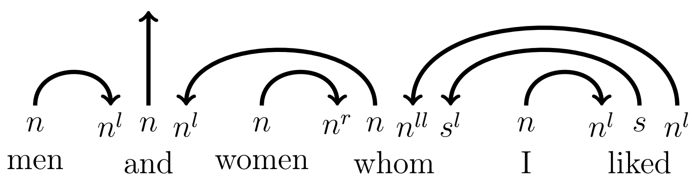
In this section, we want to construct a *strict* monoidal functor Q from the pregroup grammar $\mathbf{Pr}(T)$ to the category **FdHilb**. Since each object in **FdHilb** is its own dual we also have $Q(a^l) \cong Q(a) \cong Q(a^r)$ and because the tensor product in **FdHilb** is symmetric, we have $Q(b \cdot a) \cong Q(a \cdot b) = Q(a) \otimes Q(b)$ where the last equality holds by strictness of the functor. The partial ordering between atomic types is mapped to linear maps. However, as shown by Preller [47],

Proposition 2.3.2. *There is no strong monoidal functor from* $\mathbf{Pr}(\mathcal{T})$ *to* **FdHilb** *that maps simple types to spaces of dimension greater than one.*

Proof. Let F be such a functor and assume for contradiction that $A = Fa$ has at least two orthogonal vectors a_1 and a_2. Since $\mathbf{Pr}(\mathcal{T})$ is posetal, we have, $1_{a \cdot a^r \cdot a} = (1_a \cdot \eta_a^r) \circ (\epsilon_a^r \otimes 1_a)$ and, because F preserves the compact structure, $f := (1_A \cdot \eta_A) \circ (\epsilon_A \otimes 1_A)$ is an isomorphism in **FdHilb**. However, $f(a_1 \otimes a_2 \otimes a_1) = 0$ since ϵ_A is simply the inner product. Therefore, $\ker f$ is non empty. $\square$

As we see in the proof, the problem stems from the partially ordered nature of the category $\mathbf{Pr}(\mathcal{T})$, that identifies all morphisms with the same domain and codomain. To circumvent this issue, we need to change the domain of F to the free compact closed category $C(\mathcal{T})$. However, this change introduces a small complication into the original model of [15]. Since two morphisms between the same pair of types can be different, grammatical reductions of well-formed sentences are no longer unique. Therefore, if we want to derive the meaning of a sentence from the distributional meaning of its parts, the grammatical type of each component of the sentence does not suffice. We need to specify a reduction too. For instance, two different reductions give different meanings to the sentence "men and women whom I liked", depending on whether one associates the terms as "men and (women whom I liked)" or "(men and women) whom I liked":





This example can be found in Preller and Lambek[48] to which we refer the reader for a discussion of these ideas in the context of 2-categories. In the rest of this dissertation all examples of sentences will come with a unique unambiguous reduction.

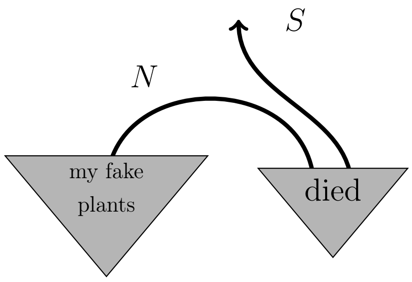
Equipped with a strict monoidal functor $F : C(\mathcal{T}) \to \mathbf{FdHilb}$, we can now define the meaning of grammatical sentences (relative to a specified reduction). Let $w_1, w_2, \ldots, w_n$ be n words, each with type t_i and associated meaning state $|w_i\rangle$ in $Q(t_i)$, for $1 \leq i \leq n$. The meaning vector of the string $w_1 w_2 \ldots w_n$, according to the reduction $\xi : t_1 \cdot t_2 \cdot \cdots \cdot t_n \to s$ is

$$|w_1 w_2 \ldots w_n\rangle := Q(\xi)\big(|w_1\rangle \otimes \cdots \otimes |w_n\rangle\big)$$

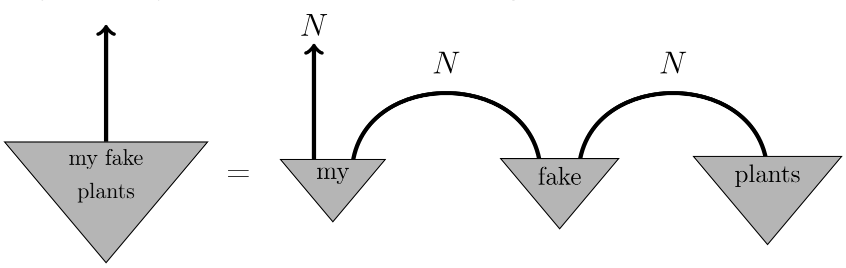
For example, consider the intransitive sentence "My fake plants died". We can assign the types $n^r \cdot s$ to the intransitive verb "die" and the noun type n to the noun phrase "my fake plants". Then the meaning of the sentence (assuming that we have already derived its meaning from its constituent parts) relative to the intuitive reduction $\iota : n \cdot n^r \cdot s \to s$, is

$$\begin{aligned} |\text{my fake plants died}\rangle &= Q(\iota)\big(|\text{my fake plants}\rangle \otimes |\text{died}\rangle\big) \\ &= (\epsilon_{Q(n)} \otimes 1_{Q(s)})\big(|\text{my fake plants}\rangle \otimes |\text{died}\rangle\big) \end{aligned}$$

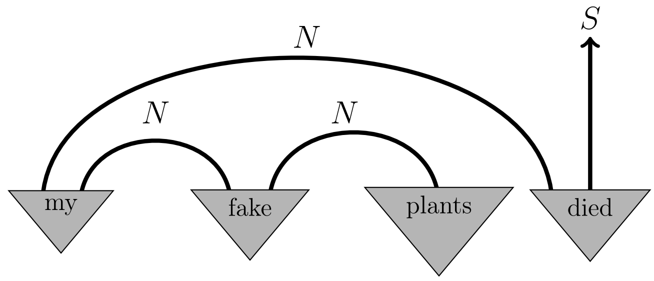
where $\epsilon_{Q(n)}$ is the co-unit of the self-dual space $Q(n)$, i.e., its scalar product. Graphically,



where $N := Q(n)$ and $S := Q(s)$. Similarly, we can compute the meaning of "my fake plants" assuming the meaning of its parts and plug it back into the diagram above. With types $my : n \cdot n^l, fake : n \cdot n^l, plants : n$ we get:



Hence, the meaning of the original sentence is represented by



Typically, in concrete instances, given a distributional model of meaning W, we assign this Hilbert space to all atomic types (in particular to the grammatical type of sentences, s); for $x \in T$

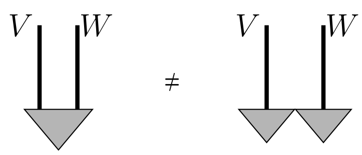
$$Q(x) = Q(s) = W$$

Thus all relational types are simply mapped to tensor products of the initial Hilbert space W.

An important feature of the tensor representation of words is *entanglement* [31]. With fixed bases $\{|i\rangle_V\}$ and $\{|j\rangle_W\}$, a state of the tensor product $V \otimes W$ has the form

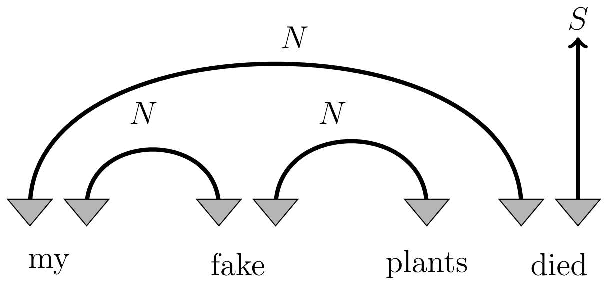
$$\sum_{ij} \tau_{ij} |i\rangle_A \otimes |j\rangle_W$$

In general, this sum cannot simply be expressed as the tensor product of a state from V and one from W; such a state is called *entangled*. The product of two states from V and W is said to be *separable*. The terminology and the mathematical formalism are borrowed from quantum physics in which the tensor product is used to denote the state of a composite system. In the graphical calculus, a separable state is the juxtaposition of two states while a general state on $V \otimes W$ is a triangle with two wires coming out:



If this equality held[1], almost all of the richness of the model would be lost: entanglement is necessary for information to flow between different words[31]. For illustration purposes, let us look at the previous example, in which all the complex types have been separated into states on atomic types:

[1]A category in which the states are all separable is called *Cartesian*. The category of sets and functions is Cartesian: a map from a set Y into a product $X_1 \times X_2$ is completely determined by how it maps elements into X_1 and X_2.



Graphically, the unit wires are disconnected from the wire that carries the output. Consequently, all grammatical interactions are destroyed and all that remains is the rightmost component of the verb multiplied by a scalar. If we are working in the projective setting, this multiplication has no effect at all on the result: no compositional exchange of information has happened and the meaning of the entire sentence is reduced to that of the right component of the verb *die*.

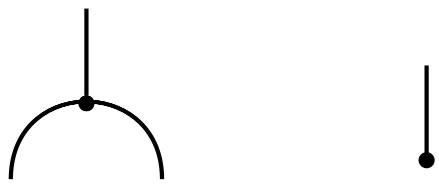
The last argument poses the question of how to construct representations of relational words as higher order tensors. We will answer this question in the next section by introducing an algebraic structure that formalises the idea of copying and deleting information in monoidal †-categories.

2.4 Relational types with Frobenius algebras

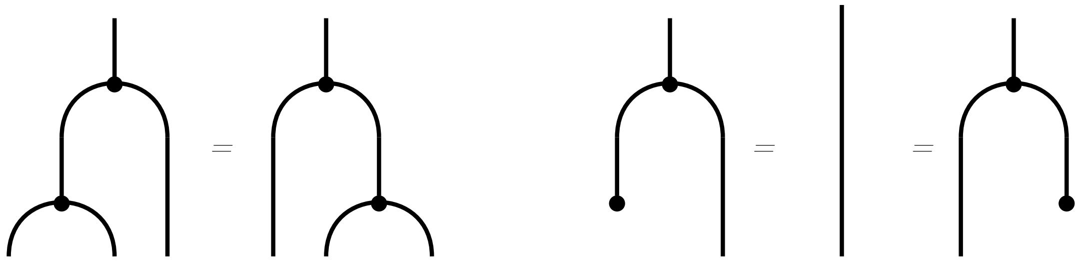
If distributional models provide a way to build meaning-vectors for words with atomic types, the question of words with relational types is more challenging. The following exposition proposes a solution, based on the work of Kartsaklis, Sadrzadeh, Pulman and Coecke [32].

2.4.1 Dagger Frobenius algebras

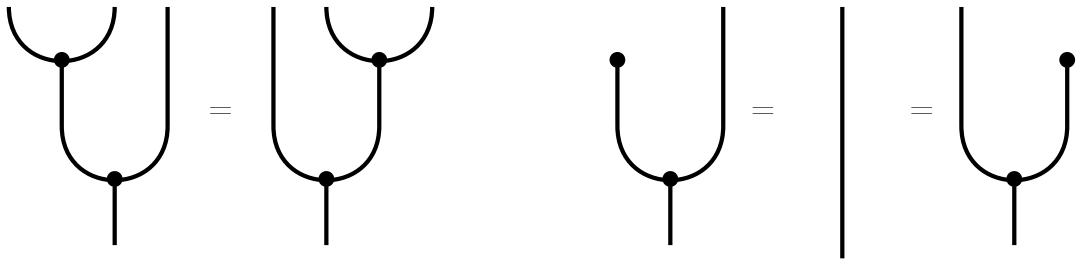
Definition 2.4.1. *In a monoidal category, a* monoid *is an object A, a* multiplication $\Delta : A \otimes A \to A$ *and a* unit $\iota : I \to A$ *depicted as*



satisfying the following associativity and unit conditions

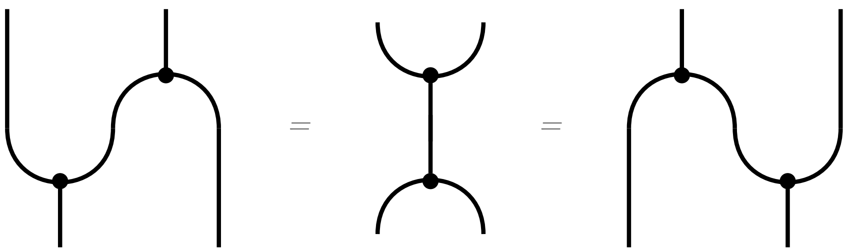


As usual these diagrams are intended to be read form bottom to top. The dual notion is that of a *co-monoid* structure on an object A, consisting of a *co-multiplication* $\nabla : A \to A \otimes A$ and a co-unit $\pi : A \to I$ which satisfy co-associativity and co-unit equations:



When an object carries both a monoid and co-monoid structure, we require these structures to be compatible in a certain sense:

Definition 2.4.2. *A* Frobenius algebra *is a choice of monoid and co-monoid for an object A such that the multiplication and co-multiplication satisfy the Frobenius condition:*

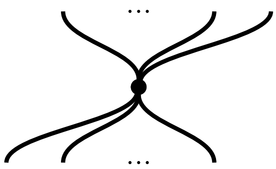


The structure that we define above was first introduced by Frobenius in different terms [22] and its equivalence with the general categorical definition that we give was first observed by Abrams in [1].

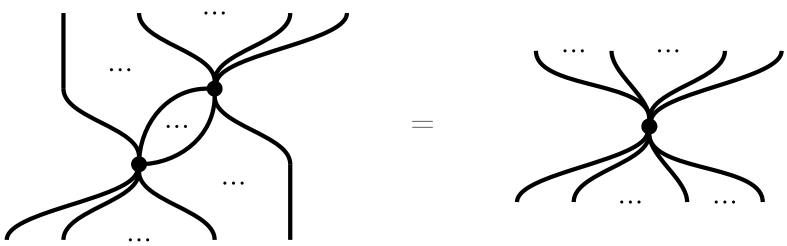
In the †-monoidal setting, the adjoint of a monoid yields a co-monoid and we naturally extend the previous definition:

Definition 2.4.3. *In a monoidal †-category, a †-Frobenius algebra is a Frobenius algebra whose co-monoid structure is adjoint to the monoid structure.*

Successive applications of the Frobenius operations admit a graphical interpretation in the form of *spiders* drawn as:



These spiders represent composition and tensoring of Frobenius operations (multiplication, co-multiplication, unit and co-unit) and only depend on the number of input and output wires [14]. The proof of this fact relies on the existence of a normal form for successive applications of Frobenius operations. As a result, spiders compose in the following fashion:

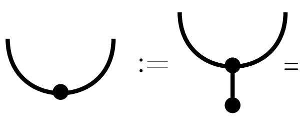
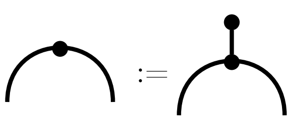


Note that the spiders on either side of the equality sign have the same number of input and output wires. As it turns out, composition of spiders captures precisely the behaviour of Frobenius algebras as defined in 2.4.2. This is what is commonly referred to as the *spider theorem* [14].

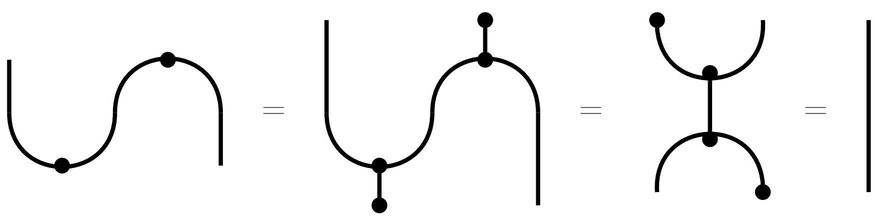
In addition, Frobenius algebras allow us to omit the direction of arrows on objects because of the following proposition.

Proposition 2.4.1. *Each Frobenius algebra induces a self-dual compact structure, i.e such that* $A^* \cong A$

Proof. The cups and caps are simply

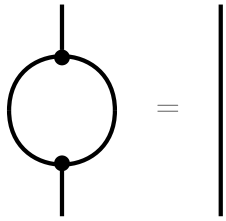
 $= \Delta^\dagger \circ \iota : I \to A \otimes A$,  $= \iota^\dagger \circ \Delta : A \otimes A \to I$

that verify the yanking equalities as a consequence of the Frobenius condition and the unit law:



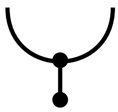
where the other equality is proven similarly. □

Furthermore, a Frobenius algebra is *normalised* if



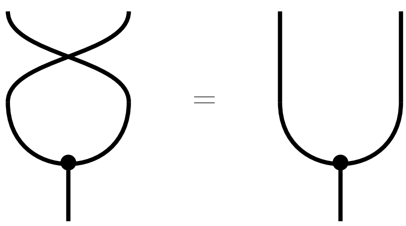
It was proven in [17] that †-Frobenius algebras in **FdHilb** are in bijective correspondence with orthogonal bases, while the normalised †-Frobenius algebras are precisely the orthonormal bases.

Concrete Frobenius algebras in FdHilb Given a finite dimensional Hilbert space H with an orthogonal basis $\{|i\rangle\}$ we can construct a †-Frobenius algebra by defining the co-monoid operations first:

 $:= |i\rangle \mapsto |i\rangle \otimes |i\rangle$  $:= |i\rangle \mapsto 1$

extended by linearity. Thus the copying map amounts to encoding faithfully the components of a vector in H as the diagonal elements of a matrix in $H \otimes H$, relative to the canonical basis. Deleting sums those components and yields a complex number. The monoid operations are their adjoints: the multiplication picks out the diagonal elements of a matrix and returns them as a vector in H; if the tensor is separable this is equivalent to the entry-wise product of vectors (sometimes called Hadamard product) relative to that basis. Note that if the tensor is not separable, the non-diagonal elements are discarded and information is lost. Finally, the unit is defined by $1 \mapsto \sum_i |i\rangle$. Recovering an orthogonal basis from a given Frobenius algebra is more involved and we refer the reader to the proof in [17].

Note that this algebra is commutative, i.e.



As announced, the co-multiplication formalises the idea of copying information, while the co-unit provides a way of deleting it. The associativity and unit law guarantee that this interpretation is consistent with our intuition: the various ways of composing the copying, merging and deleting of information are all equivalent and the order in which we perform these operations does not matter. This is the content of the spider theorem.

2.4.2 The meaning of predicates

Equipped with these notions, we can now describe a natural way to build tensor representations of relational types. The following method is due to [32]. Assume that we have a distributional model W and a strict monoidal functor Q from a compact closed categorical grammar to **FdHilb** such that all atomic types are mapped to W.

Words with relational types are treated as predicates that relate their arguments together. Furthermore, since we are in a vector space setting, the relationship between their arguments is weighted by scalars that quantify the strength of this relationship. For example, a transitive verb represents the correlation of two noun-phrases - correlation that we will represent by an entangled state on $W \otimes W$.

Experimental work [26, 25] demonstrated that higher order tensor representations for predicate types could be constructed by summing over the tensor product of their arguments weighted by the number of times that they appeared as arguments of the predicate word in the corpus. For example, the tensor representation of a few common grammatical types are exhibited below:

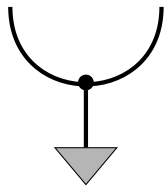
$$\begin{array}{lcr} \text{Intransitive verb} & : & \sum_i |subject_i\rangle \\ \text{Transitive verb} & : & \sum_i |subject_i\rangle \otimes |object_i\rangle \\ \text{Adjective} & : & \sum_i |noun_i\rangle \end{array}$$

where the index i counts the number of times each argument appears.

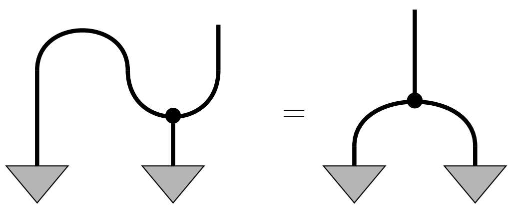
However, we notice a type mismatch immediately: the rank of these tensors is one less than the space that the functor Q assigns to their type grammatical counterpart. For example, an intransitive verb is a vector in W when it should have type $W \otimes W$: it should be a linear operator on W, taking the meaning-vector of a noun as input and outputting another vector.

This is where the copying operation of Frobenius algebras find its use. Every distributional model comes equipped with a canonical basis representing the set of context words chosen to build the model from a corpus. As we have seen, every orthonormal basis induces a special dagger Frobenius algebra that we can use to assign words to their appropriate type, as assigned by the functor Q.

Adjectives and intransitive verbs We can encode elements of W in $W \otimes W$ as



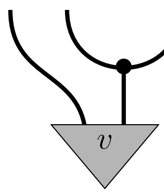
which, when applied to an argument gives



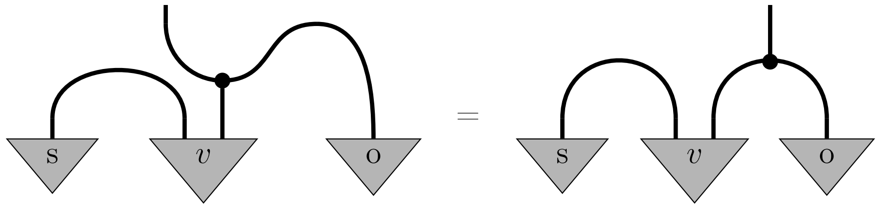
The equality is simply the application of the Frobenius condition. Here, if we presented the meaning of an intransitive verb applied to its subject, the resulting expression also provides the meaning of an adjective applied to a noun, by commutativity of the algebra.

Transitive verbs There are several ways to encode elements of $W \otimes W$ as tensors of $W \otimes W \otimes W$; we will give the encoding that provides the most accurate empirical results according to the disambiguation experiments of [32]. For different encodings, we refer the reader to the original paper.

The main idea is to copy the information about the object by applying the comultiplication on the tensor component of the object:



Applied to a subject and an object we obtain the meaning of a transitive sentence in normal form:



Intuitively, it seems natural to give more weight to information brought about by the object. For example, to disambiguate between the verb *say* and *ask* the subject does not provide much information however, a *question* is more likely to be asked than said. This intuition is validated by its success on disambiguation tasks. In a sense, the object determines the meaning of an ambiguous verb.

We conclude this chapter by quoting [32] directly:

> "An experimental future direction is a higher order evaluation of the definition classification task using an unambiguous vector space [...], 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."

This is what we set out to accomplish in the next chapter.

Chapter 3

Introducing ambiguous meaning

The premise of this chapter is that the model of distributional semantics described earlier loses essential information about the meaning of words by collapsing the different contexts in which they appear into a single vector representation. To illustrate this point, let us consider the following example: the word "queen" can be used to refer to a female monarch or to a piece on a chessboard. In a vector space model built from a corpus that contains witnesses of both these uses, the dual meaning of "queen" will be irremediably compressed to a single vector. At first glance, this is consistent with our assumptions about meaning in natural language: ultimately, the meaning of a word is all its valid uses, according to the "meaning is use" mantra of distributional semantics, and all of these uses are compiled in its vector representation. Therefore, one could argue, the meaning of "queen" as a chess piece is simply the vector space representation of its uses in a specific context. The sum of all these meaning-vectors gives the final meaning of "queen".

However, for most words, it seems possible to a human observing its use in context to classify each occurrence into a smaller set of uses: "queen" as a monarch or as chess piece, "head" as a body part, as the leader of an organisation or as brainpower, etc. This set partitions all the possible uses of the word and corresponds roughly to its different definitions available in a dictionary, attesting to the fact that the meaning of words in natural language is inherently ambiguous. Unfortunately, the process of assigning a vector representation to a word compresses partly the ambiguity of its meaning. The aim of this chapter is to provide a construction that extends distributional models so as to retain this fundamental ambiguity in the representation of each word.

3.1 Mixing and its linguistic interpretation

The previous compositional model relies on a strong monoidal functor from a compact closed category, representing grammar, to **FdHilb**. In this section, we will describe a possible new codomain category for this functor. However, before we start, we establish a few guidelines:

- our construction needs to retain a compact closed structure in order to carry the reduction maps from the category of grammar to the new category;
- we wish to be able to compare the meaning of words as in the previous model; the new category needs to come equipped with a dagger structure that implements this comparison;
- finally, we need a Frobenius algebra to merge and duplicate information in concrete models.

To achieve our goal, we will explore a quantum physical construction, originally due to Selinger [51], in the context of the categorical model of meaning developed in the previous chapter.

3.1.1 A little quantum theory

In the field of categorical quantum physics, the †-compact closed setting is an abstraction of the Hilbert space formulation of quantum theory in terms of pure states as vectors and measurements as self-adjoint operators [2, 3]. Various soundness and completeness results show that this categorical reformulation provides a rich and high-level language to model *pure* quantum information protocols. Shifting the perspective to the field of linguistics, the same formalism proposes a description of the lexical interactions of words when strung together. In both cases, the corresponding graphical calculus admits an intuitive interpretation in terms of information flow, between physical systems in one case, and words in the other.

Nevertheless, in quantum physics, the Hilbert space model is insufficient to incorporate the epistemic state of the observer in its formalism: what if one ignores the initial state of a quantum system and can only attribute a probability distribution to a set of possible states? How can one model this situation? The answer is by considering a *statistical ensemble* of pure states: for example, one may assign a $1/2$ probability that the state vector of a system is $|\psi_1\rangle$ and a $1/2$ probability that it is in state $|\psi_2\rangle$. We say that this system is in a *mixed state*.

In the Hilbert space formalism, there is no way to represent this state as a vector. In fact, any normalised sum of pure states is again a pure state (by the vector space structure). Note that the state $(\psi_1 + \psi_2)/\sqrt{2}$ is a quantum superposition and not the mathematical representation of the mixed state above.

This situation is similar to the issue that we face when trying to model ambiguity in distributional semantics. Given two different meanings of a word and their relative weights (given as probabilities), simply looking at the convex composition of the associated vectors collapses the ambiguous meaning to a single vector, thereby fusing together the two senses of the word.

The mathematical response to this problem is to move the focus away from states in a Hilbert space to a specific kind of operators on the same space. To understand the change of perspective, we need to explain how measurable or observable quantities (position, momentum, spin, energy, etc.) are described in quantum physics and how they relate to linguistics. Our exposition of quantum physical mathematical formalism is based on [35]. An observable (short for observable quantity) of a quantum system described by the finite dimensional Hilbert space $\mathcal{H}$, is, in the pure state formalism, a self-adjoint operator A on $\mathcal{H}$. The spectrum of A is the set of possible values of the observable. If our system is in state $|\psi\rangle$, the expectation value (in the probabilistic sense) of the observable A is defined as $\langle\psi|A|\psi\rangle$. The interpretation is that the probability of a physical measurement of this observable quantity giving the eigenvalue a (with associated eigenvector $|a\rangle$) is $P(a) = |\langle\psi|a\rangle|^2$. Note that this probabilistic interpretation is possible because $|\psi\rangle$ is assumed to be normalised and, since A is self-adjoint it admits a spectral decomposition in an orthonormal basis. Finally, a quick calculation gives $\langle\psi|A|\psi\rangle = \sum P(a)a$, a weighted average of all possible outcomes, where the sum is over the spectrum of A.

For mixed states, we want to obtain a similar mathematical description, consistent with the probabilistic interpretation above in the case of a statistical mixture of pure states, $|\psi_1\rangle$ with probability p_1, $|\psi_2\rangle$ with probability p_2 etc., such that $\sum p_i = 1$. In the general case, there is no state vector $|\psi\rangle$ such that $\langle\psi|A|\psi\rangle$ describes the expected value of the measurement, with the same probabilistic behavior as above. However, it turns out that there exists an positive semi-definite, self-adjoint operator ρ, with $\mathrm{Tr}(\rho) = 1$, such that the expected value of the measurement is $\mathrm{Tr}(\rho A)$. This is the content of Gleason's theorem [23]. In general, this operator can be given by $\rho = \sum p_i|\psi_i\rangle\langle\psi_i|$ where the coefficients p_i are non-negative and add up to one. Moreover, we see that $\mathrm{Tr}(\rho A) = \sum p_i\langle\psi_i|A|\psi_i\rangle$, the weighted sum of the expected values of the measurement over pure states. In this formalism, the mixed state of the previous example is formulated as a density operator $\rho = \frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2|$). Operators that describe mixed states are called *density operators.* A density operator defines a pure state precisely when it is idempotent, $\rho^2 = \rho$, that is, when it is a projection (onto a one-dimensional subspace since its trace is one).

In linguistics, the generalisation to mixed states is carried out similarly, with the following interpretation: assuming that we have a distributional model in the form of a Hilbert space M, a vector $|m\rangle$ of M represents a pure meaning. A density operator $\rho = \sum p_i |m_i\rangle\langle m_i|$ provides the meaning of a word from the meaning vectors $|m_i\rangle$. In conceptual terms, mixing is interpreted as ambiguity of meaning: a word w with meaning given by ρ can have pure meaning m_i with probability p_i. The meaning of a word represented by a general density operator is called an *ambiguous* meaning.

Now, if mixed states are density operators, we need a notion of morphism that preserves this structure, i.e., that maps states to states. In the Hilbert space model, the morphisms were simply linear maps. The corresponding notion in the mixed setting is that of *trace preserving completely positive map* (TPCP). What should such a map look like?

Let $\mathcal{M}(\mathcal{H})$ be the vector space of operators on a Hilbert space $\mathcal{H}$. First, a mapping from density operators on a space $\mathcal{H}_A$ to density operators on $\mathcal{H}_B$ should be a $i)$ linear map $\mathcal{E} : \mathcal{M}(\mathcal{H}_A) \to \mathcal{M}(\mathcal{H}_B)$. Then, it should send positive semi-definite operators to positive semi-definite operators, i.e., $ii)$ if $\rho \geq 0$, $\mathcal{E}(\rho) \geq 0$. A map satisfying $ii)$ is called *positive.* From the new definition of states, we may think that these conditions are sufficient. However, maps that satisfy conditions $i)$ and $ii)$ do not respect the monoidal structure: the tensor product of two maps satisfying $ii)$ may not satisfy $ii)$. We require a stronger positivity condition: $ii')$ a map $\mathcal{E} : \mathcal{M}(\mathcal{H}_A) \to \mathcal{M}(\mathcal{H}_B)$ is completely positive if, for every quantum system C, the map

$$1_C \otimes \mathcal{E} : \mathcal{M}(\mathcal{H}_C) \otimes \mathcal{M}(\mathcal{H}_A) \to \mathcal{M}(\mathcal{H}_C) \otimes \mathcal{M}(\mathcal{H}_B)$$

is positive, where 1_C is the identity on $\mathcal{M}(\mathcal{H}_C)$. The category of operator spaces and maps satisfying, $i)$ and $ii')$ is the structure we are looking for.

However, to constitute a compositional model of meaning, this construction needs to respect our stated goals: is the category of operator spaces and completely positive maps a †-compact closed category? What morphism plays the part of the Frobenius algebra of the previous model?

Before answering these questions, we will describe a construction that builds a similar category, not only from **FdHilb**, but from any †-compact closed category.

This construction will be realised in the graphical calculus, a language in which it will be easier to verify the properties that we need.

3.1.2 The D and CPM constructions

The category that we are going to build was originally introduced by Selinger [51] as a generalisation of the corresponding construction on Hilbert spaces. The Frobenius algebra on this structure was first described by Coecke and Spekkens in [19]. As we will see, this algebra is not a completely positive map and, therefore, requires a relaxation of the original conditions in [51].

Definition 3.1.1. *[19] Given a †-compact closed category $\mathcal{C}$ we define a category $D(\mathcal{C})$ with*

- *the same objects as $\mathcal{C}$;*
- *morphisms between objects A and B of $D(\mathcal{C})$ are morphisms $A \otimes A^* \to B \otimes B$ of $\mathcal{C}$.*
- *composition and dagger are inherited from $\mathcal{C}$ via the embedding $E : D(\mathcal{C}) \hookrightarrow \mathcal{C}$ defined by*

$$\begin{cases} A & \mapsto A \otimes A^* \quad \textit{on objects;} \\ f & \mapsto f \qquad\quad\ \textit{on morphisms.} \end{cases}$$

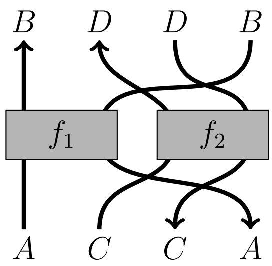
In addition, we can endow the category $D(\mathcal{C})$ of a monoidal structure by defining the tensor $\otimes_D$ by

$$A \otimes_D A = A \otimes B$$

on objects A and B, and for morphisms $f_1 : A \times A^* \to B \otimes B^*$ and $f_2 : C \times C^* \to D \otimes D^*$, by

$$f_1 \otimes_D f_2 : A \otimes C \otimes C^* \otimes A^* \xrightarrow{\cong} A \otimes A^* \otimes C \otimes C^* \xrightarrow{f_1 \otimes f_2} C \otimes C^* \otimes D \otimes D^* \xrightarrow{\cong} C \otimes D \otimes D^* \otimes C^*$$

Or graphically by,



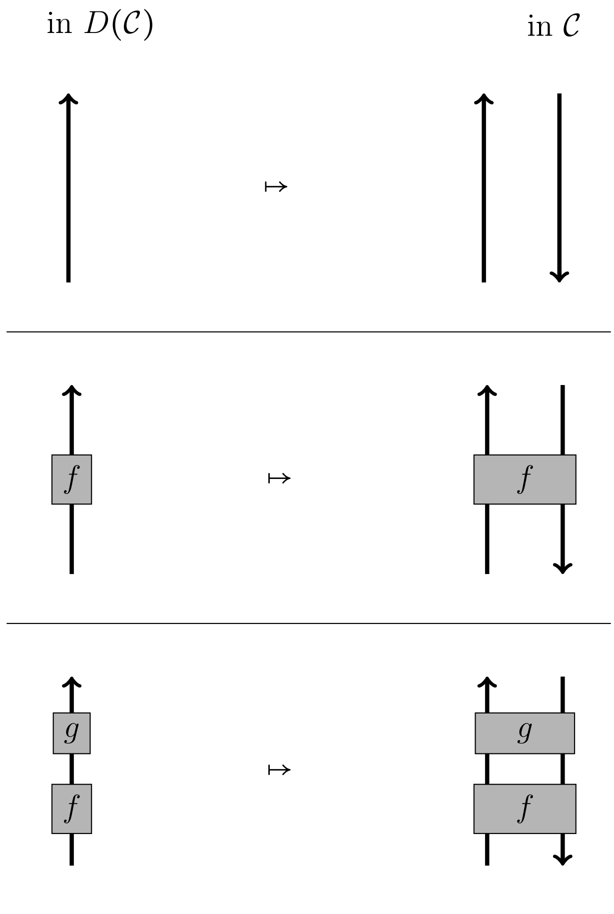
The intuitive alternative of simply juxtaposing the two morphisms as we would in $\mathcal{C}$ fails to produce a completely positive map in general, as will become clearer when we define completely positive maps in this context.

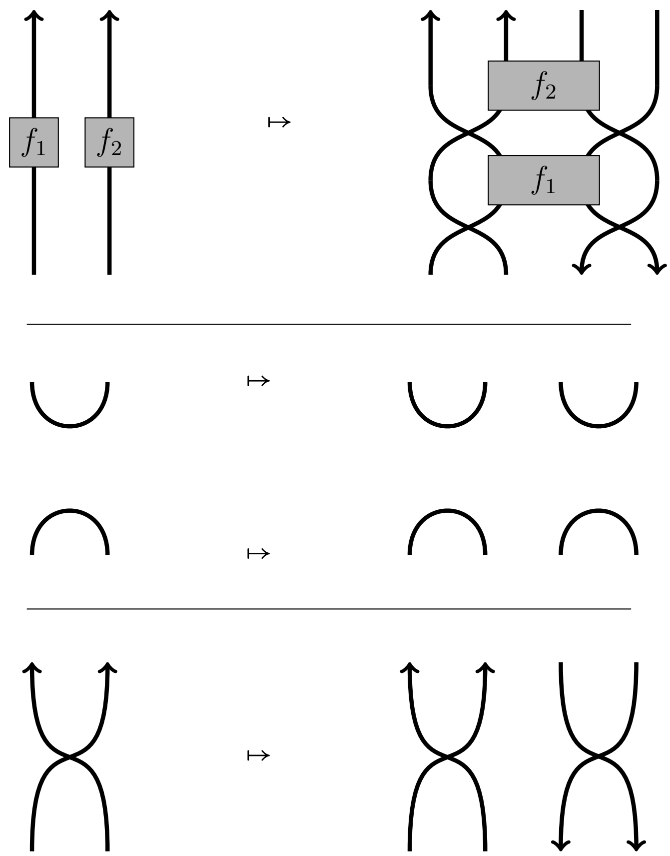
Finally, this category carries all the structure required. We refer the reader to [51] for a proof of the following:

Proposition 3.1.1. *The category $D(\mathcal{C})$ inherits a †-compact closed structure from $\mathcal{C}$ via the strict monoidal functor $M : \mathcal{C} \to D(\mathcal{C})$ defined inductively by*

$$\begin{cases} f_1 \otimes f_2 & \mapsto M(f_1) \otimes_D M(f_2) & ; \\ A & \mapsto A & \textit{on objects}; \\ f & \mapsto f \otimes f_* & \textit{on morphisms}. \end{cases}$$

For reference, we give a dictionary that translates useful diagrams from one category to the other (the objects are omitted to avoid overcrowding, and we use the symbol $\mapsto$ to denote the embedding $E : D(\mathcal{C} \to \mathcal{C})$:

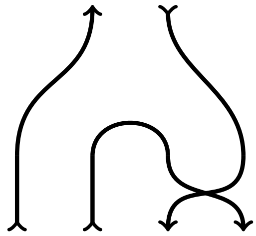




A few remarks on this table Due to the asymmetry of the $\mathcal{C}$ representation of a morphism in $D(\mathcal{C})$ we will adopt the convention of always drawing the wire for the dual object in $f : A \otimes A^* \to B \otimes B^*$, on the right of the original object. In addition, the nested structure of tensor products of morphisms imposes that we write objects from the outside-in as opposed to the usual representation from left to right. Finally, the table gives a slightly modified picture of the $\mathcal{C}$ embedding of the $D(\mathcal{C})$-tensor product. It is strictly equivalent but makes for a more symmetrical presentation of the interweaving of the wires.

Now, we define a dagger Frobenius algebra on the category $D(\mathcal{C})$. We will see that, given a suitable interpretation of $D(\mathcal{C})$ in terms of density operators, this structure constitutes a generalisation of the Frobenius algebra of the previous model.

Proposition 3.1.2. *For every object A of $D(\mathcal{C})$, the morphisms of $D(\mathcal{C})$, $\Delta : A \otimes_D A \to A$ defined by the following diagram in $\mathcal{C}$:*



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

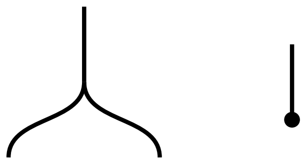
and $\iota : I \to A$ with the definition in $\mathcal{C}$

$\iota =$  $= \eta_{A^*}$

are the multiplication and unit of a dagger Frobenius algebra $\mathcal{F}_D$ - where η_A and ϵ_A are the unit, co-unit pair of A for the compact structure of $\mathcal{C}$ and σ is the natural swap isomorphism in $\mathcal{C}$.

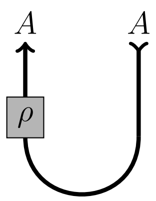
Proof. Proof of the associativity can be found in [19]. The other properties can be proven similarly. □

In $D(\mathcal{C})$, we portray the Frobenius monoid operation and the unit as



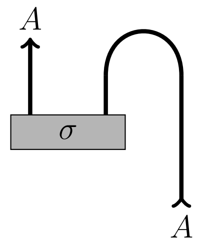
Now for the generalisations of the ideas of mixed states and (completely) positive maps from quantum theory, we refer to [51]. In what follows, we will call operator any morphism from an object to itself.

First, notice that we have an isomorphism between states of $D(\mathcal{C})$, i.e., morphisms $I \to A$ and operators on A in $\mathcal{C}$. Explicitly, the isomorphism $\mathcal{C}(A, A) \to \mathcal{C}(I, A \otimes A^*)$ is, for an operator $\rho : A \to A$,

$\rho \mapsto \ulcorner \rho \urcorner = (\rho \otimes 1_{A^*}) \circ \eta_{A^*} =$ 

whose inverse is defined by

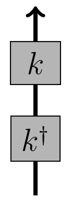
$$\sigma \mapsto \llcorner \sigma \lrcorner = 1_A \circ (\sigma \otimes \epsilon_A) =$$



for $\sigma : I \to A$ that embeds as $E(\sigma) : I \to A \otimes A^*$ in $\mathcal{C}$. By the yanking equalities, these two operations are inverses. As a result, we will indifferently write ρ for the operator on A, its name $I \to A \otimes A^*$ and its corresponding state $I \to A$ in $D(\mathcal{C})$.

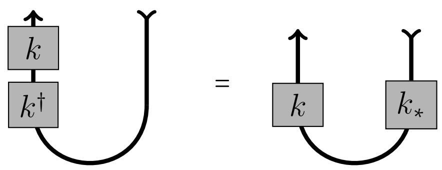
In an arbitrary †-category, a positive operator is an operator $\rho : A \to A$ that splits in the following way: there exists an object B and a morphism $k : A \to B$ such that

$$\rho = k \circ k^\dagger =$$

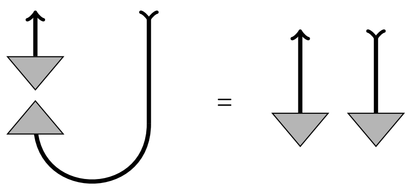


Note that positive operators are self-adjoint, i.e, $\rho^\dagger = \rho$.

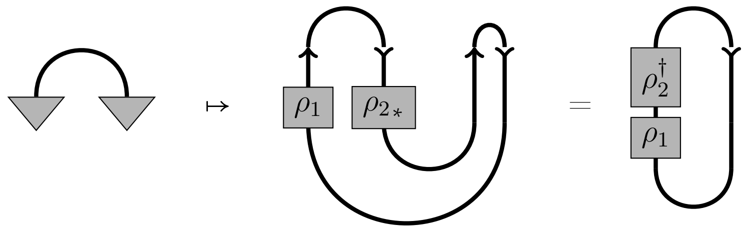
In $D(\mathcal{C})$, a mixed state is a morphism $m : I \to A$ whose embedding in $\mathcal{C}$ is the image of a positive operator by the isomorphism above, i.e., such that its embedding is the name $\ulcorner \rho \urcorner : I \to A^* \otimes A$ of a positive operator $\rho : A \to A$; graphically, its embedding has the form



A pure state is thus represented as



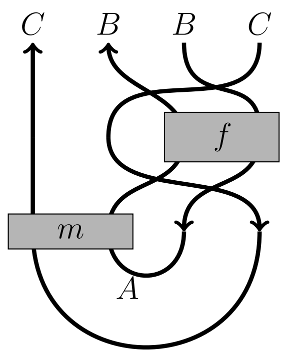
The dagger was first introduced as a generalisation of the inner product of Hilbert spaces, to derive a measure of proximity between states. With the dagger of $D(\mathcal{C})$ we retain this ability for general states (not only mixed states, for which the expression simplifies since they are self-adjoint) as evidenced by the diagram below:



$= \mathrm{Tr}(\rho_2^\dagger \rho_1)$

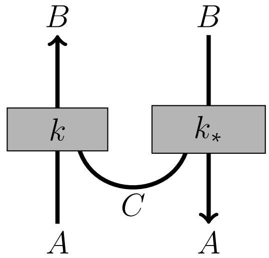
Recall our definition of a completely positive map from the previous section. In graphical terms, it can be expressed by the next definition.

Definition 3.1.2. *In $D(\mathcal{C})$, a morphism $f : A \to B$ is completely positive if, for all object C and every mixed state $m : I \to C \otimes_D A$, $(1_C \otimes_D f) \circ m$ (shown embedded in $\mathcal{C}$ below) is positive.*



In [51], this definition is proved equivalent to the existence of an object C and morphism $k : C \otimes A \to B$, in $\mathcal{C}$, such that f embeds in $\mathcal{C}$ as

$$(k \otimes k_*) \circ (1_A \otimes \eta_{C^*} \otimes 1_{A^*}) =$$

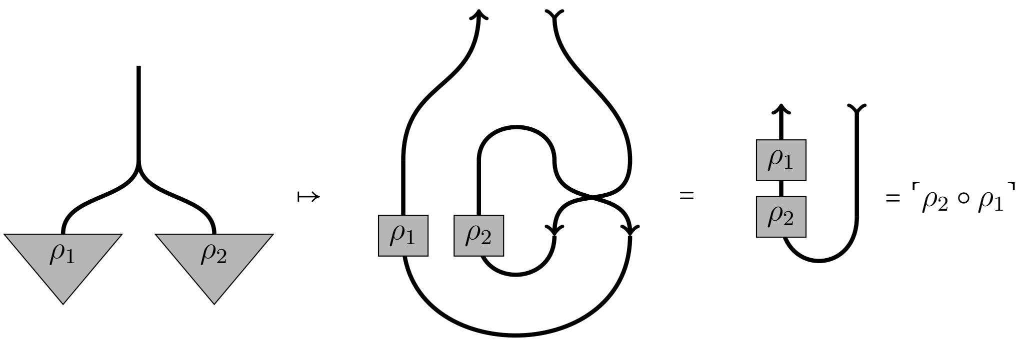


From this last representation, we easily see that the composition of two completely positive maps is completely positive. Similarly, the tensor product (of $D(\mathcal{C})$) of two completely positive maps is completely positive. Finally, for a morphism f of $\mathcal{C}$, $F(f) = f \otimes f_*$ is completely positive. Therefore, we can define:

Definition 3.1.3. *The category $CPM(\mathcal{C})$ is the subcategory of $D(\mathcal{C})$ whose objects are the same and morphisms are completely positive maps.*

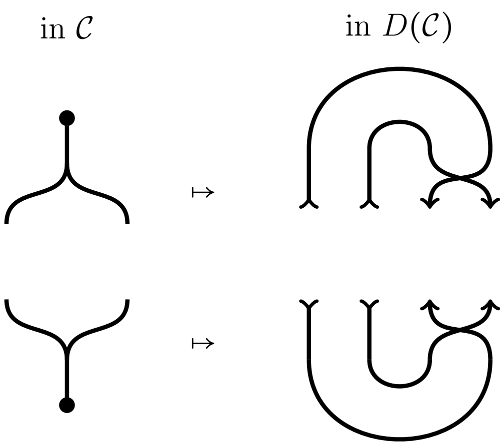
Remarks 1. The dagger Frobenius algebra is not completely positive. This is due to the asymmetry of its representation in $\mathcal{C}$ (see the next remark for a more compelling reason). Moreover, for the same reason, it is non-commutative. The asymmetry stems from the intricate structure of the tensor product $\otimes_D$, an intricacy required to preserve its complete positivity.

2. The action of the Frobenius multiplication Δ on morphisms $I \to A$ of $D(\mathcal{C})$ is particularly interesting. In fact, it implements the composition of operators of $\mathcal{C}$, in $D(\mathcal{C})$, as evidenced in the next diagram

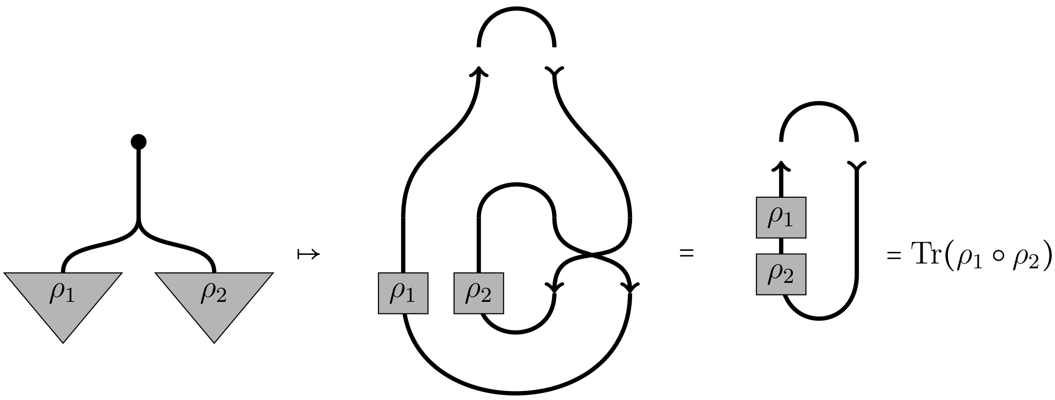


However, the composition of two positive operators ρ_1 and ρ_2 is only positive if they commute, i.e if $\rho_1 \circ \rho_2 = \rho_2 \circ \rho_1$. This is precisely why the Frobenius multiplication is not completely positive: in the general case the multiplication of two mixed states will not be a mixed state. This suggests that we could relax our notion of mixed states for linguistic applications. We will discuss this question and provide an alternative in section 3.2.2.

3. As we have already seen, every dagger Frobenius algebra induces a self-dual compact structure. In this case, the corresponding cups and caps and their embedding in $\mathcal{C}$ are shown below.



In fact, as we will see in the next section, this compact structure is the one that we use to interpret grammatical reductions in the category of meaning $D(\mathcal{C})$. Therefore, it is useful to study its action on states. If the Frobenius multiplication implements the composition of operators, looking at the diagram of the previous remark shows that the co-unit acts as taking the trace of this composition:



3.2 Compositional model of meaning: reprise

Our aim in this section is to reinterpret the compositional model of meaning of [15] as a functor from a compact closed grammar to the category $D(\mathbf{FdHilb})$. This functor is simply the composition:

$$MQ : C(\mathcal{T}) \to \mathbf{FdHilb} \to D(\mathbf{FdHilb})$$

Since M sends a Hilbert space A to an object A in $D(\mathbf{FdHilb})$, the mapping of atomic types, their duals and relational types of the grammar occur in exactly the same fashion as in the previous model.

Furthermore, note that Q is strongly monoidal and M is strictly monoidal, so the resulting functor is strongly monoidal and, in particular, preserves the compact structure. Thus, we can perform type reductions in $D(\mathbf{FdHilb})$ according to the grammatical structure dictated by the category $C(\mathcal{T})$. However, note that we are mapping grammatical types to the same objects as their duals. Therefore, applications of the co-unit in $C(\mathcal{T})$ are mapped to applications of the co-unit associated to the self-dual structure induced by the Frobenius algebra in $D(\mathcal{C})$.

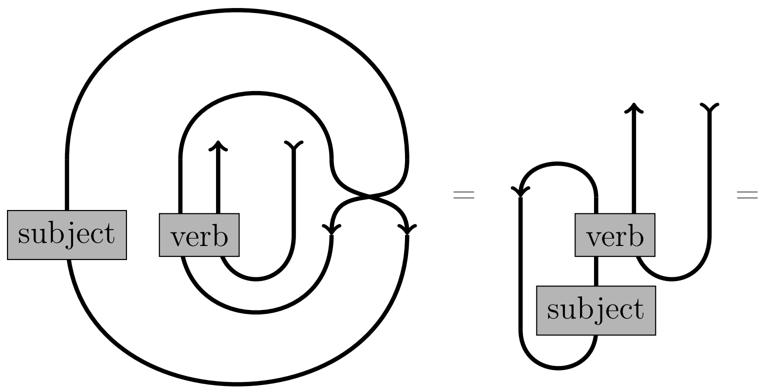
Let $w_1, w_2, \ldots, w_n$ be n words, each with type t_i and associated meaning state $\rho(w_i) : I \to MQ(t_i)$ in $D(\mathbf{FdHilb})$, for $1 \leq i \leq n$.

The meaning vector of the string $w_1 w_2 \dots w_n$, according to the reduction $\xi : t_1 \cdot t_2 \cdot \dots \cdot t_n \to s$ is

$$\rho(w_1 w_2 \dots w_n) := MQ(\xi)\big(\rho(w_1) \otimes_D \cdots \otimes_D \rho(w_n)\big)$$

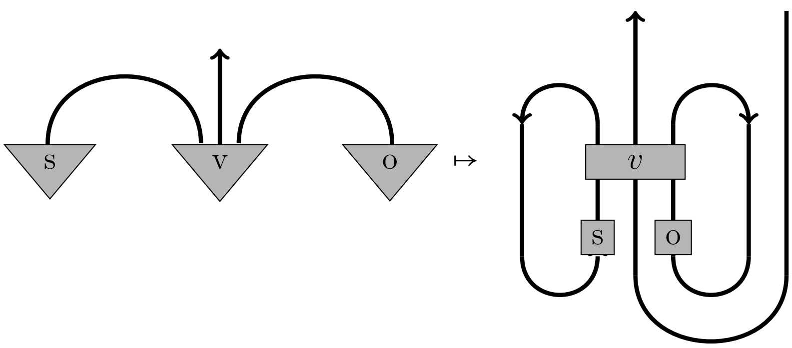
Now we give the reductions and their graphical presentations for a few examples, as in the previous chapter. Note that we will simply use standard caps and cups in $D(\mathbf{FdHilb})$ to depict the self-dual compact structure induced by the Frobenius algebra since no confusion can arise.

Intransitive sentences Given an intransitive verb (of type $n^r \cdot s$) with meaning $\rho(v)$ and a noun-phrase with meaning $\rho(n)$, we can compute the meaning of the sentence "noun-phrase verb" (relative to the usual grammatical reduction $n \cdot n^r \cdot s \to s$):



$= \ulcorner \mathrm{Tr}_N(\rho(v) \circ (\rho(n) \otimes 1_S)) \urcorner$

Transitive sentences Given a transitive verb (of type $n^r \cdot s \cdot n^l$), a subject and an object (both noun-phrases of type n), with meaning $\rho(v), \rho(s)$ and $\rho(o)$ we compute the following reduction to obtain the meaning of the sentence "subject verb object":



$= \ulcorner \mathrm{Tr}_{N,N}(\rho(v) \circ (\rho(s) \otimes 1_S \otimes \rho(o))) \urcorner$

Remark As was already noted, the Frobenius multiplication is not completely positive. As a result the meaning of well-formed strings of words may not be a mixed state. To tackle this issue, there are two possible responses:

1. *extend our interpretation* to include more general types of operators;
2. *change the model* so that the meaning of sentences is always a mixed state operator.

We will discuss three possibilities for 2. in section 3.2.2. At the moment, it is not clear what the linguistic interpretation of a general operator should be. Further research is required in that direction.

3.2.1 Linguistic interpretation of operators

Our model of meaning is a relative one: meaning does not have any ontological or absolute definition; it arises from being able to compare the position of words and expressions in an inner product space. More specifically, in $D(\mathbf{FdHilb})$, the inner product of two operators ρ_1 and ρ_2 is $\mathrm{Tr}(\rho_1^\dagger \circ \rho_2)$. Therefore, we can adopt the viewpoint that, as long as we can compare them, expressions have meaning: comparing a word w_1 to another w_2 is equivalent to asking the question: is w_1 about w_2?

Let M be the Hilbert space associated to a distributional model. We can view its set of one-dimensional subspaces, represented by normalised vectors, as the set of testable atomic concepts. In this interpretation, a pure meaning operator $\rho^W = |w\rangle\langle w|$ is a projector onto the one-dimensional subspace associated to $|w\rangle$. Given another atomic concept c and its associated projector $\rho^C = |c\rangle\langle c|$, $\mathrm{Tr}((\rho^W)^\dagger \circ \rho^C)$ gives the answer to the question "is c about w?". In the original compositional model of meaning, the answer to this question was given a numerical value by taking the inner product of the two vectors. In the operator view of meaning, the answer is $\mathrm{Tr}((|w\rangle\langle w|) \circ (|c\rangle\langle c|)) = |\langle c|w\rangle|^2$. Geometrically, the angle (given by the inner product in M) that the w makes with the question's atomic concept vector reflects the probability that the answer to the question is positive (Yes).

Mixed meaning operators generalise the previous ideas to convex combinations of (orthogonal) testable concepts. Let ρ be a positive (and thus self-adjoint) operator of trace one. Its eigenvectors give a basis of the space M that correspond to testable atomic concepts. With each eigenvalue p_i of its spectrum we associate the projector onto the closed subspace generated by the corresponding eigenvector $|c_i\rangle$ (for simplicity, we assume its eigenvalues to be nondegenerate). Since the eigenvalues of ρ add up to one, they admit a probabilistic interpretation. Extending the previous remarks about projectors, computing $\mathrm{Tr}(\rho^\dagger \circ \rho^W)$ gives the weighted answer to the questions "is w about c_i with probability p_i?".

This is similar to *measurements* in quantum physics: recall that *observable* quantities are represented by self-adjoint operators, i.e., operators that we can decompose into an orthonormal set of projectors; and a state induces a probability measure on the lattice of projectors. In this setting, projectors represent properties of a system or events and self-adjoint operators are random variables whose possible values are given by their spectral decomposition. Given an observable A, and a system in state ρ, the expected value of the measurement is $\mathrm{Tr}(A\rho) = \mathrm{Tr}(A^\dagger \rho)$. Thus comparing two expressions can be seen as measuring one with respect to the other; the expected value of such a measurement quantifies how related they are to each other.

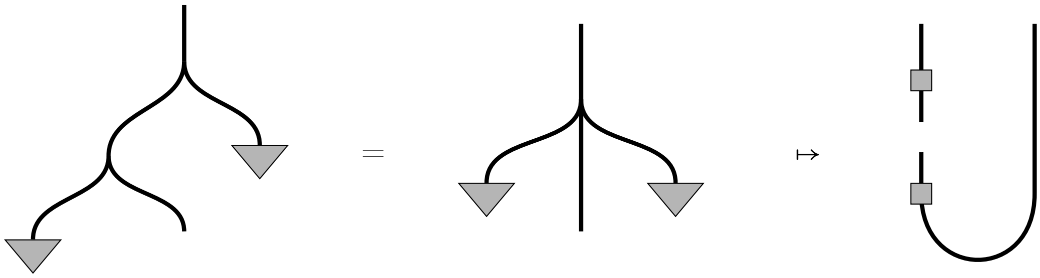
The probabilistic interpretation is justified by Gleason's theorem which guarantees that every (countably additive) probability measure μ on the lattice of closed subspaces of a Hilbert space has the form $\mu(P) = \mathrm{Tr}(P\rho)$ for some density operator ρ. Consequently, we can identify the meaning of a word with the probability measure that it induces on this lattice.

More general operators appear as a result of combining mixed states operators according to the reduction rules of our compositional model. There is no reason for such an operator to be a mixed (or pure) state itself since there is no constraint in our model that requires sentences to decompose simply into convex combinations of atomic concepts. One possible way to provide a more intuitive interpretation of a general operator is to look at its singular value decomposition. We will not explore this interpretation in this dissertation.

3.2.2 Complete positivity

To stay in the realm of positive operators, we want to modify the current model so that the meaning reductions of tensors of words to sentences is completely positive.

Drawing on the work of Leifer, Spekkens and Poulin [39, 40] we introduce the non-commutative and non-associative product of operators $\rho_1 \star \rho_2 = \rho_1^{1/2} \circ \rho_2 \circ \rho_1^{1/2}$, where $\rho^{1/2}$ is the positive square root of ρ, well defined for positive operators. In $D(\mathbf{FdHilb})$, the map $\mathcal{E}(\rho_1) : A \to A$ defined by $\rho \mapsto \rho_1^{1/2} \circ \rho \circ \rho_1^{1/2}$ is completely positive or, in other words, the product $\star$ preserves mixed states. Graphically, this map is given by the diagram

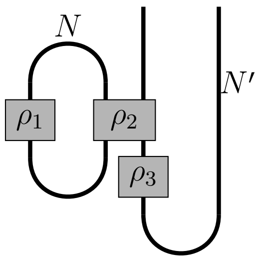


in which each gray node stands for the square root of ρ and the gap on the rightmost diagram represents where to insert the argument of the map $\mathcal{E}(\rho)$. Note that we can recover ρ from $\mathcal{E}(\rho)$ by simply applying the later to the identity operator. Following [19] we call $\mathcal{E}(\rho)$ the *modifier* of ρ.

Intuitively, we want to replace every application of the Frobenius multiplication with the $\star$ operation in order to preserve positivity. Symbolically, given a chain of composed operators on a space A, $\rho_1 \circ \cdots \circ \rho_n$, we substitute each operator by its associated modifier and apply the last map to the identity operator on A:

$$\rho_1 \circ \cdots \circ \rho_n \mapsto \mathcal{E}(\rho_1) \circ \cdots \circ \mathcal{E}(\rho_n) 1_A = \rho_1^{1/2} \cdots \circ \rho_{n-1}^{1/2} \circ \rho_n \circ \rho_{n-1}^{1/2} \circ \ldots \rho_1^{1/2}$$

Since this mapping can always be applied after computing the meaning of a sentence in terms of compositions of operators in **FdHilb**, we can now interpret diagrammatic composition of operators, in terms of $\star$ instead of $\circ$, the regular composition operator. For example, the diagram



represents the operator $\mathrm{Tr}_N(\rho_2 \star (\rho_1 \otimes 1_{N'})) \star \rho_3$. This interpretation always makes sense if we restrict ourselves to the composition of positive operators [1].

However, this operation does not interact well with pure states. A pure state is given by a projector of the form $|m\rangle\langle m|$ for some vector $|m\rangle$. The composition of a list of such projectors according to the $\star$ operation yields a multiple of the first projector of the list. This feature of the $\star$ composition limits severely the flow

[1]Note, however, that there is more than one way to write the diagram above, leading to different $\star$ products: $\mathrm{Tr}_N(\rho_2 \star (\rho_1 \otimes 1_{N'})) \star \rho_3 \neq \mathrm{Tr}_N(\rho_2 \star (\rho_1 \otimes \rho_3))$. If we adopt the convention to choose the expression with the least number of compositions inside the trace, it becomes unique.

of information between words. This is why we will not adopt this method in this dissertation. However, its links with a form of quantum Bayesian calculus make it worthy of further research (see section 4.2.2).

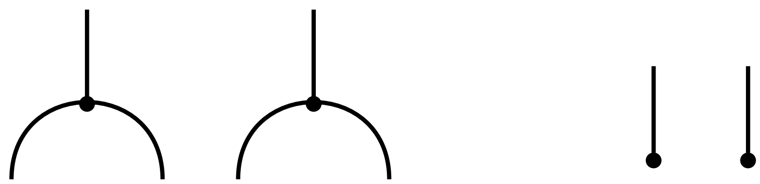
A simpler alternative is to consider $|\rho| := (\rho^\dagger \rho)^{1/2}$. We can then normalise this operator to obtain the desired density operator. This gives a simple procedure to obtain a density operator from any operator ρ and we will adopt it for lack of a better solution in the rest of this dissertation. There is a sense in which this operator is optimal: the assignment

$$\rho \mapsto |\rho|$$

is idempotent and thus a *projection* onto the subspace of self-adjoint operators.

Further research is needed to determine in what sense $|\rho|$ gives an appropriate answer and if not, what the best self-adjoint approximation of a given operator should be. One possible direction of research is to investigate notions of nearest self-adjoint operator with respect to a given norm on the space of operators. For example, the trace norm is already known to give a measure of statistical indistinguishability of quantum states [46] and seems like a good candidate for this task.

Finally, as we have seen in chapter 2, every distributional model comes with a canonical orthonormal basis and, as a result, a special dagger Frobenius algebra associated to it. The image of this algebra by the functor M is a dagger Frobenius algebra in $D(\mathbf{FdHilb})$, whose multiplication and unit are here shown embedded in **FdHilb**:

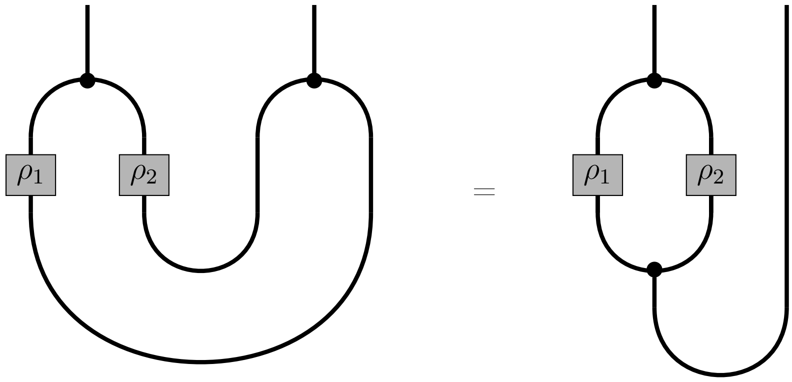


It is easy to check that this construction gives a dagger Frobenius structure on each object of $D(\mathbf{FdHilb})$. All properties are immediate consequences of the corresponding properties of the Frobenius algebra in **FdHilb**. Additionally, since the original algebra is commutative, this one is too.

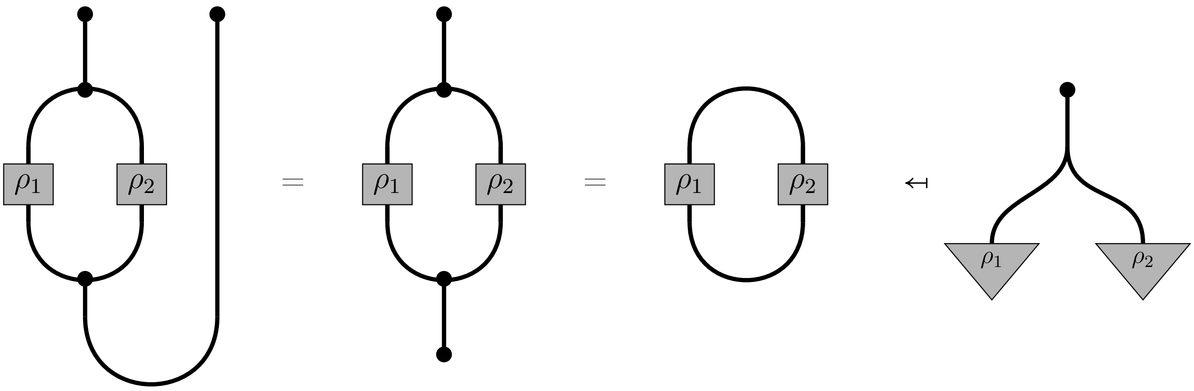
3.2.3 An alternative Frobenius algebra

Let $\mathcal{F}_C$ be the dagger Frobenius algebra that we introduced above. We can modify the meaning of sentences functor to map reductions in the category of grammar to applications of the self-dual compact structure induced by $\mathcal{F}_C$.

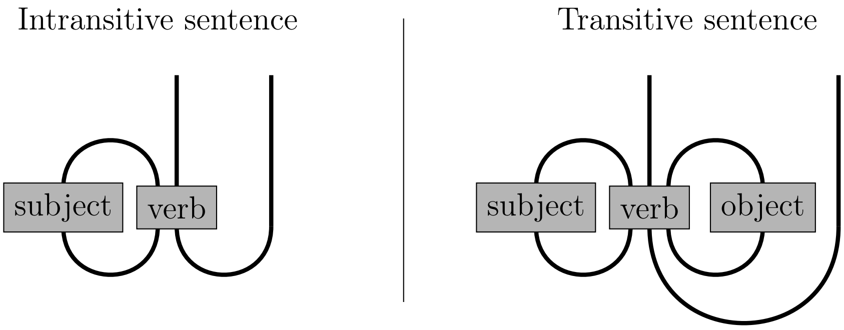
Applied to operators we get



that is clearly positive if ρ_1 and ρ_2 are. This multiplication implements the entry-wise product of matrices, sometimes called the Hadamard product. Conveniently, this allows us to stay within CPM(**FdHilb**). Interestingly, the induced self-dual compact structure is the same as that induced by $\mathcal{F}_D$: the co-unit amounts to taking the trace of the composition of two operators, as the following diagram demonstrates:



Consequently, the graphical representation of the meaning of transitive and intransitive sentences is the same as before:



However, the two structures will differ in the next section where we make use of the co-multiplications to build higher order types from simple types. When using the copying operations of each algebra the non-commutativity of $\mathcal{F}_D$ introduces significant differences in how the meaning of sentences is computed.

3.2.4 Building operators for relational types

We will now assume that we have a distributional model W and a categorial grammar $C(\mathcal{T})$ with a functor MQ that assigns the Hilbert space W to every atomic type, including s. We will also assume that we already have ambiguous meaning operators for all words with simple types. We now turn to words with relational types such as (in)transitive verbs or adjectives.

As in the original model, in concrete instantiations, higher order tensor representations of predicate words can be built by summing over their possible arguments. According to this method, the operator meaning of the predicates of our simple example sentences are:

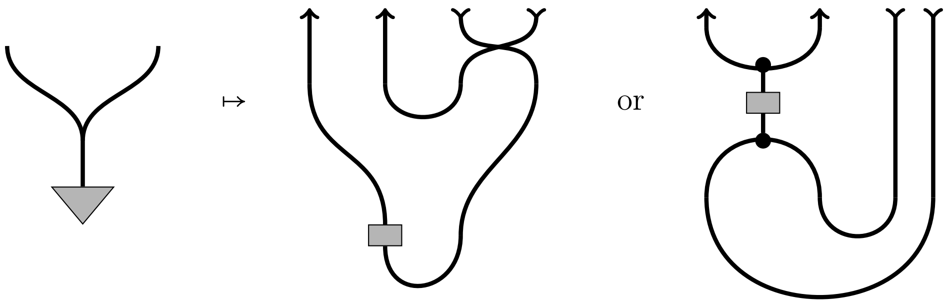
$$
\begin{array}{lcr}
\text{Intransitive verb} & : & \sum_i \rho(subject_i) \\
\text{Transitive verb} & : & \sum_i \rho(subject_i) \otimes \rho(object_i) \\
\text{Adjective} & : & \sum_i \rho(noun_i)
\end{array}
$$

where we sum over all possible arguments, typically those encountered in a corpus. In addition, to obtain density operators we need to normalise the operators above (a convex sum of density operators is a density operator).

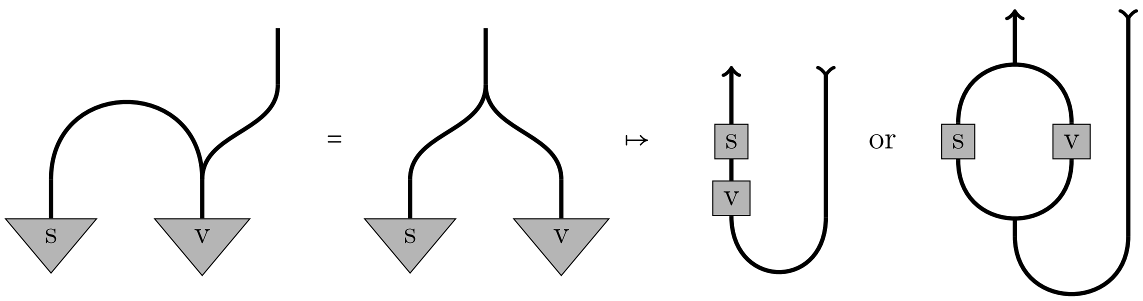
Again, we are faced with a problem: these maps are operators on a space whose tensor rank is one fewer than the rank of the image of their grammatical types in $D(\mathbf{FdHilb})$. For example, a transitive verb has type $n^r \cdot s \cdot n^l$; the image of this type by the strongly monoidal functor Q is $W \otimes W \otimes W$ but, with this method, a transitive verb is an operator on $W \otimes W$.

Following the example of the previous chapter, we will resort to the application of the usual Frobenius algebra operations to overcome this problem. The Frobenius structure on $D(\mathbf{FdHilb})$ can be interpreted as $\mathcal{F}_D$ or $\mathcal{F}_C$ and we will systematically present the results in both languages.

Adjectives and intransitive verbs We naturally encode operators on W as operators on $W \otimes W$ with the help of the Frobenius co-multiplication:

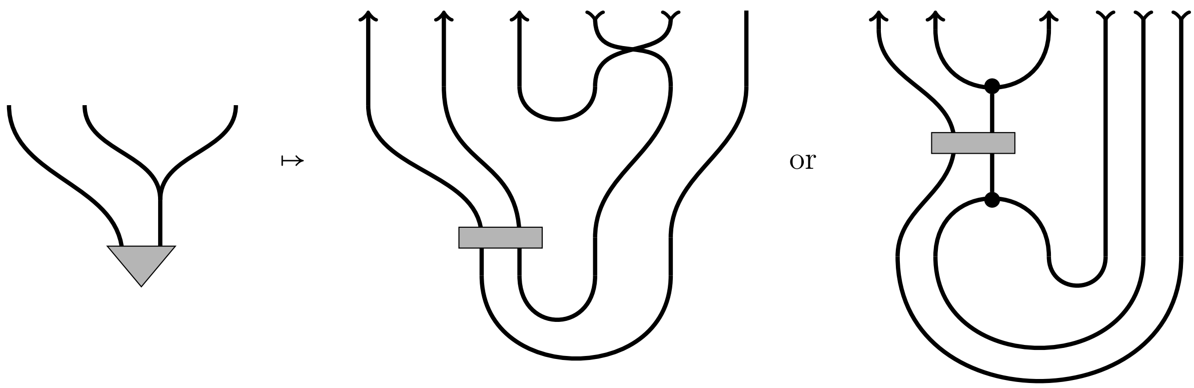


When applied to its subject, an intransitive verb outputs a sentence. Graphically, we compute its meaning as follows:

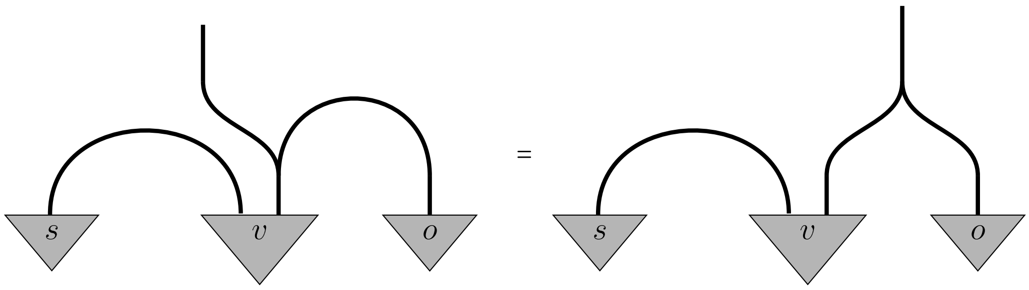


where the equality is the application of the Frobenius condition. To compute the meaning of a noun-phrase of the form "adjective noun" we simply reverse the order of application. Interestingly, the non-commutativity of the $\mathcal{F}_D$ multiplication allows us to give meaning to the transformation of a verb into an adjective, in English. For example, if we have an operator representation of the meaning of the verb *roll*, $\rho(roll)$, and the noun *stone*, $\rho(stone)$, we can compute the meaning of the intransitive sentence "the stone rolls" and that of the noun-phrase "rolling stones", as $\rho(stone) \circ \rho(roll)$ and $\rho(roll) \circ \rho(stone)$ respectively. This is an improvement over the previous model, in which the Frobenius multiplication was commutative.

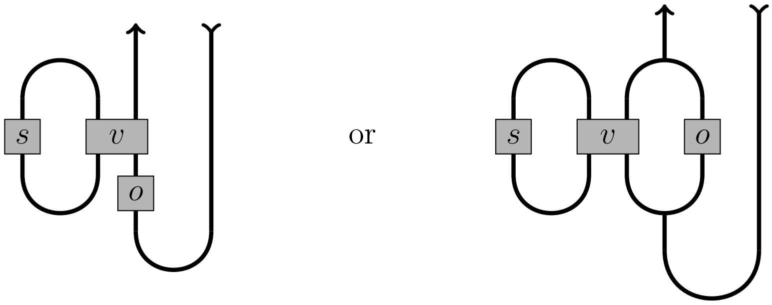
Transitive verbs There are several ways to encode an operator on $W \otimes W$ as an operator on $W \otimes W \otimes W$. Following the choice of the previous chapter, we use the co-multiplication to copy the dimension of the object and obtain the following representation for a transitive verb:



which, when applied to a subject and an object, yields the following sentence representation, shown in normal form on the right.



Here, embedded in **FdHilb**, for the algebras $\mathcal{F}_D$ and $\mathcal{F}_C$ respectively,



The verb interacts with its subject and the resulting partially traced operator object and the result is compressed on the subject side to produce an operator on the side of the object. Alternatively, the verb interacts first with its subject, the result is compressed to obtain an operator that is composed with the object. This implies that the resulting operator carries more information from the object than the subject. Conceptually, it is easy to see that in most cases, more information is carried by the object to determine the meaning of some ambiguous verb. For instance one can identify that the verb *bend* followed by *the truth* is used in its metaphorical sense but much less information is given if we know that *a child bends...*

However, it should be noted that this diagrammatic form is only one of several possibilities that give the best experimental results in the pure meaning model. Again, for a presentation of various encodings of transitive verbs with the machinery of Frobenius algebras, we refer the reader to [32] (whose examples are not restricted to the **FdHilb** model and can be easily lifted to $D(\mathbf{FdHilb})$ using either of the Frobenius algebras that we have introduced).

3.3 Compositional information flow

In this section we wish to examine the meaning of a few example sentences in order to highlight the interaction of words through composition of operators and motivate the interpretation of mixing as ambiguity.

3.3.1 Flow of ambiguity

With a model of meaning based on density operators we get a notion of entropy for free. Entropy yields a measure of the information content of each sentence or, in linguistic terms, a measure of ambiguity. For that purpose, we will use an extension of the Shannon entropy to density operators, called the Von Neumann entropy.

Given an operator $\rho = \sum_i p_i |m_i\rangle\langle m_i|$ we could define its entropy as the Shannon entropy of the probability distribution $\{p_i\}_i$ but it would not be invariant under change of basis. There is a more natural definition, based on the eigen-decomposition of ρ: the Von Neumann entropy (sometimes called quantum entropy) of ρ is defined as

$$S(\rho) = -\operatorname{Tr}(\rho \ln \rho)$$

or more simply, given the spectral decomposition of $\rho = \sum p_i |i\rangle\langle i|$, as

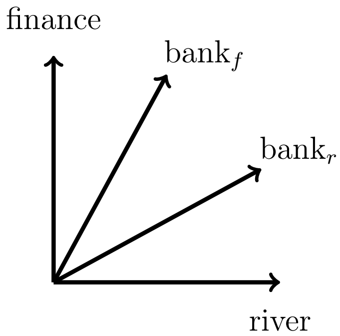
$$S(\rho) = -\sum_i p_i \ln p_i$$

where $p \ln p$ is set to zero by convention if $p = 0$, consistent with the limit of $x \ln x$ as x approaches zero. Note that, in the complex case, the logarithm of an operator exists if and only if it is invertible. In almost all concrete applications, the density operators representing the meaning of words are sparse matrices and not invertible. That is why care must be taken to restrict all notions involving inverses to an appropriate support, namely the range of the operator in question.

For a pure meaning operator ρ, $\rho^2 = \rho$ and the entropy is equal to zero. This is consistent with the idea that such an operator represents an unambiguous word or phrase.

How does the entropy evolve when composing words to form sentences? This question is very hard to answer precisely in full generality. Nonetheless, it is interesting to examine the interaction of a ambiguous word with a pure meaning word to build intuition - for instance the particular interaction of an ambiguous verb or adjective with an unambiguous noun. In fact, since density operators are convex sums of pure operators, all interactions are convex combinations of this simple form of word composition. In addition, the disambiguations of polysemous verbs is one of the key NLP tasks on which the previous compositional models were tested and thus constitutes an interesting case study. Finally, the analysis of this particular case provides an striking point of comparison of the two Frobenius structures.

Example Consider the noun *bank*: it can be used to refer to a financial institution or to the sloping ground beside a river. For simplicity, we will represent the distributional meaning of *bank* in a two-dimensional Hilbert space, with basis words $finance$ and $river$, as a mixture of two pure meaning operators bank_f and bank_r. Let $\rho(\text{bank}) = p|\text{bank}_r\rangle\langle\text{bank}_r| + (1-p)|\text{bank}_f\rangle\langle\text{bank}_f|$, for some $0 < p < 1$.



Now, suppose we want to compute the meaning of the expression "river bank". We did not give any specific rule to deal with the juxtaposition of two nouns to form a noun-phrase but we can consider that the word *river* plays the role of an adjective and use the construction of section 3.2.4.

a) For the algebra $\mathcal{F}_D$ the meaning of the expression "river bank" is given by the composition of $\rho(\text{bank})$ with the projection onto the horizontal axis

$$\sigma = |r\rangle\langle r|\rho(\text{bank}) = p\langle r|\text{bank}_r\rangle|r\rangle\langle\text{bank}_r| + (1-p)\langle r|\text{bank}_f\rangle|r\rangle\langle\text{bank}_f|$$

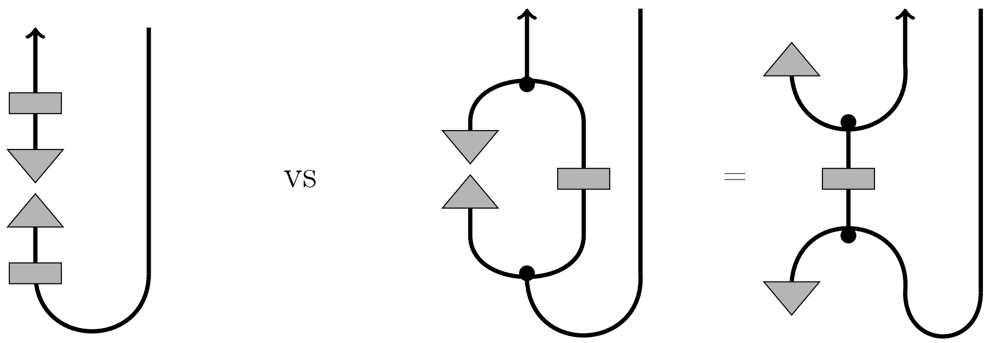
We see that, since $\langle r|\text{bank}_f\rangle < \langle r|\text{bank}_r\rangle$ the new weighing is skewed in favor of the bank_r meaning.

Furthermore, the density operator associated to σ (in the sense of section 3.2.2), $||\sigma||^{-1}(\sigma^\dagger\sigma)^{1/2}$ is also a projector and thus, represents a pure state, whose entropy vanishes: the composition of an ambiguous word with a unambiguous word gives as a result an unambiguous meaning to the resulting expressions.

The same mechanism is at play in disambiguating the meaning of ambiguous intransitive verbs because the last result holds in generality: given a projection onto a one-dimensional subspace $|w\rangle\langle w|$ and a density operator ρ, $|w\rangle\langle w|\rho$ is a (not necessarily orthogonal) projection. A quick calculation shows that its associated density operator is the projector onto the one-dimensional space spanned by $\rho|w\rangle$: $(|w\rangle\langle w|\rho)^\dagger(|w\rangle\langle w|\rho) = \rho|w\rangle\langle w|w\rangle\langle w|\rho = \rho|w\rangle\langle w|\rho$, who is an orthogonal projector once normalised. In a sense, the meaning of the pure word determined that of the ambiguous word.

b) For the algebra $\mathcal{F}_C$ the meaning of "river bank" is the Hadarmad product of ρ(bank) and the projection onto the horizontal axis, which yields this same projection again. This is a special degenerate case that occurs when the unambiguous word is one of the basis elements that induces the Frobenius algebra. In the non degenerate case where *river* is not a basis element we do not necessarily get a projector so that the ambiguity is maintained.

Perhaps the comparison is easier to understand graphically. We can represent the meaning according to both $\mathcal{F}_D$ and $\mathcal{F}_C$ as follows



It is clear that the flow of ambiguity is stopped by the pure state in the first diagram. In the second picture, ambiguity flows.

3.3.2 Finding the right structure

If the algebra $\mathcal{F}_D$ seems better suited to represent the meaning of the example phrase above, it is not always the case. There may be combinations of words for which we want the ambiguity of the polysemous word to dominate and induce an ambiguous meaning of the overall phrase.

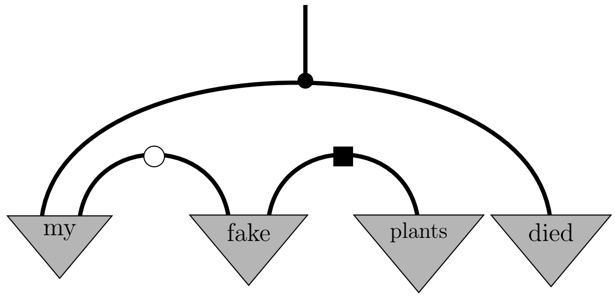
Moreover, the algebras $\mathcal{F}_D$ and $\mathcal{F}_C$ are not the only possible Frobenius structures on $D(\mathbf{FdHilb})$. For instance, we can see that every word induces a normalised Frobenius algebra via its spectral decomposition: it is the Frobenius structure whose co-multiplication duplicates the information relative to its orthonormal basis of eigenvectors. This gives an interesting alternative to compute the meaning of the previous class of examples: assume that we have an ambiguous term whose meaning is given by $\rho = \sum_i r_i |i\rangle\langle i|$ for a spectral decomposition $\{|i\rangle\}$, and a pure meaning word whose associated projector is $|w\rangle\langle w|$. Composing their meaning according to the dagger Frobenius algebra induced by the orthonormal basis $\{|i\rangle\}$ yields

$$\sum_i r_i |\langle w|i\rangle|^2 |i\rangle\langle i|$$

This sum will be skewed towards the meaning of ρ that is closest to that of $|w\rangle$, as we would expect of such a composition; it does not determine the meaning fully as in the compositional semantics of the algebra $\mathcal{F}_D$ but provides a lesser degree of interaction that may be suitable in many situations in which some degree of ambiguity needs to be preserved.

There are many more. In fact, Coecke, Heunen and Kissinger [16] introduced the category $\mathrm{CP}^*(\mathcal{C})$ of dagger Frobenius algebras (with some technical conditions) and completely positive maps, over an arbitrary †-compact category $\mathcal{C}$, in order to study the interaction of classical and quantum systems in a single categorical setting: classical systems are precisely the commutative algebras and completely positive maps are quantum channels, that is, physically realisable processes between systems. Interestingly, in accordance with the content of the no-broadcasting theorem for quantum systems (see [5]) the multiplication of a commutative algebra is a completely positive morphism (e.g $\mathcal{F}_C$) while the multiplication of a non-commutative algebra is not (e.g. $\mathcal{F}_D$).

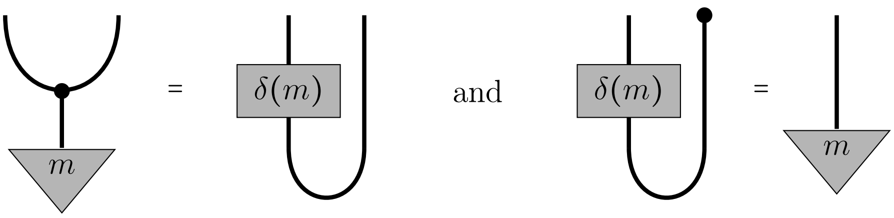
With linguistic applications in mind, this suggests various ways of composing the meaning of words each corresponding to a specific Frobenius algebra operation. Conceptually, this idea makes sense since a verb does not compose with its subject in the same way that an adjective composes with the noun phrase to which it applies. The various ways of composing words may also offer a theoretical base for the introduction of logic in distributional models of natural language.



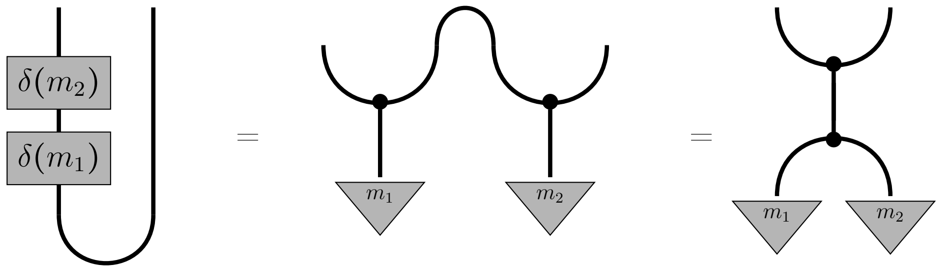
3.3.3 Recovering unambiguous meaning

We are now going to show that we can simulate the pure state compositional model of meaning within the new density operator framework. Therefore, the latter can be seen as an extension of the former. Note that this result justifies the application of the point-wise product in the Frobenius algebra of the previous model as a particular case of the composition of operators. These results apply to the algebra $\mathcal{F}_D$ - in the case of $\mathcal{F}_C$, we can easily recover the previous model by considering exclusively pure state operators: in graphical terms this amounts to drawing the diagrams of the previous chapter twice, reflected on the horizontal axis.

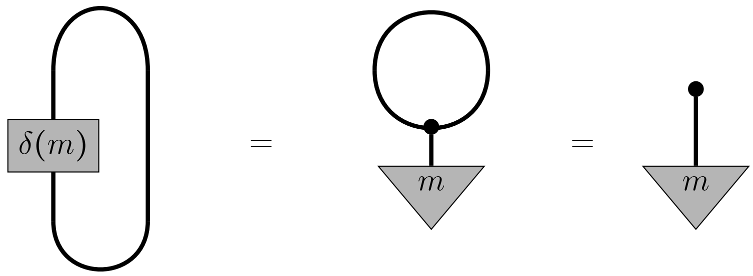
Assume that we have a distributional model represented by a Hilbert space W and a canonical basis $|i\rangle$. There is a natural way to associate a density operator to a vector $|m\rangle$: if $\sum_i m_i |i\rangle$ is the decomposition of $|m\rangle$ in the canonical basis, the operator $\delta(m) := \sum_i m_i |i\rangle\langle i|$ is a diagonal matrix with respect to this basis. For the rest of this section we will call *diagonal* every operator (or, equivalently, its state in $D(\mathbf{FdHilb})$) that can be represented by a diagonal matrix in the basis $|i\rangle$. Conversely, given a diagonal operator on W of the form $\sum_i w_i |i\rangle\langle i|$ we let $|w\rangle = \sum_i w_i |i\rangle$ be its associated vector. Graphically, in **FdHilb**



We will first prove that the Frobenius algebra in $D(\mathbf{FdHilb})$ acts on diagonal operators as the Frobenius algebra associated to the canonical basis acts on vectors in W. Since the Frobenius multiplication in $D(\mathbf{FdHilb})$ implements the composition of operator on each object, for diagonal operators it simply multiplies all the entries, which is precisely the action of the Frobenius multiplication on W. Graphically, in **FdHilb**



Now, for the co-unit, we have



Since all reductions make use of the self-dual compact structure induced by the Frobenius algebra, it is clear that the reduction of a tensor product of diagonal operators on W behaves exactly like the corresponding reduction of the tensor product of vectors of W.

3.3.4 Where does ambiguity come from?

To conclude this section we discuss briefly how to build ambiguous meaning operators in a concrete distributional model. The question was answered partly for relational types assuming that we had at our disposal meaning operators for words with simple types. But, how do we obtain the meaning operators for words with simple types? If our original idea was to represent ambiguous meaning as a mixed state operator, we need to define where the ambiguity arises. If a word can be used with at least two different meanings, how do we recognise each meaning in a text? What constitutes a witness of the use of this word in different senses? In other words [58]:

> "Given that context is a key factor in resolving polysemy, the central question in the theory of polysemy is still that of what aspects of word meanings are predefined and invariant across multiple contexts, versus what other aspects are indeterminate and only realized in context."

There are various ways in which one can build ambiguity into a distributional model. For different applications, the questions above admit different answers. For example, ambiguity is useful to integrate the predictions of different models. Assume that we obtain two distributional models built from two (not necessarily disjoint) corpora of texts. Let W and W' be their associated Hilbert spaces. We can form the space $W \oplus W'/\sim$, where $\sim$ is the equivalence relation that identifies basis vectors representing the same words. As a result of this construction a word can have two meanings induced by its representation in W and W': for a word m, let $|m\rangle : \mathbb{C} \to W$ (resp. $|m'\rangle : \mathbb{C} \to W'$ be its meaning in W (resp. W'); we have the canonical map $|\widetilde{w}\rangle : \mathbb{C} \to W \to W \oplus W'/\sim$ (resp. $|\widetilde{w'}\rangle : \mathbb{C} \to W' \to W \oplus W'/\sim$) and can assign the operator $\rho^m = p|\widetilde{w}\rangle\langle\widetilde{w}| + (1-p)|\widetilde{w'}\rangle\langle\widetilde{w'}|$ to m. The probability p reflects a particular weighing of both models according to relevant criteria (length of the corpus, target application domain, etc.). This method generalises to more than two different models by taking convex sums of pure meaning operators.

Ambiguity can be found within a single model too. Looking at the distributional data of a polysemous word can reveal semantic clusters. For instance, the two senses of the word *bank*, as a financial institution and as the slope beside a river, are far removed from one another. In vector space semantics, [53] gives metric notions that characterise polysemous words according to their geometric properties. There have been a few other attempts at using machine learning clustering techniques to group the uses of words and extract their different meanings [34][6], but the classification of polysemous words remains a difficult problem.

Chapter 4

An application: the meaning of definitions

To demonstrate some of the theoretical possibilities of a compositional model of meaning based on density matrices, we will study several properties of definitions. A definition is "a statement of the exact meaning of a word" according to the Oxford English dictionary. In a distributional model, the meaning of a word is not derived from its definition but corresponds to all its uses. How can we reconcile these two notions?

4.1 Defining definitions

As a first step in this direction we need to extend our compositional model to support definitions. When looking quickly at a few entries from the Oxford dictionary, we notice that definitions of nouns often consist of 1) a simple and general class term and 2) a clause that specifies where the defined word lies in that class. For instance, a "queen" can be defined as a "woman who rules an independent state". Intuitively, many definitions have the form: "noun-phrase pronoun relative-clause". In this example, the larger class is that of women, and the queens are those women who rule an independent state. We witness a similar possible form in the next example: a "portrait" is a "painting, drawing or photograph of a person". Here, the class is the conjunction of paintings, drawings and photographs that the preposition "of" restricts it to those of a person.

Therefore, in order to give a compositional account of definitions we need to assign meaning to relative pronouns and the preposition "of". In fact, we cannot simply extract the distributional meaning of those words: they appear in such a variety of contexts that their distributional interpretation does not capture any essential information about their use. They are noise words. This is because their use is predominantly grammatical; they relate the information of the relative clause to that of the noun phrase.

Recent work by Clark, Coecke and Sadrzadeh [49, 50] provides semantic interpretations of relative pronouns in a categorical compositional model of meaning, in terms of Frobenius algebras and their graphical calculus. We can adapt these results to the operator model of meaning developed in the previous chapter. As usual, diagrams will help us to visualise the information flow between words and, in this context, picture how relative pronouns combine, duplicate and discard information to output the meaning of a definition.

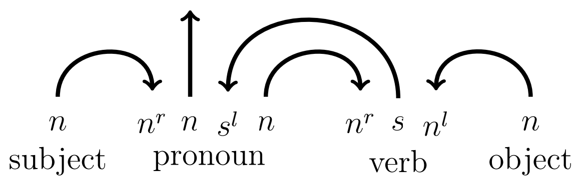
4.1.1 Relative clauses

We start with a grammatical analysis of the types of relative pronouns and the reduction of clauses that contain them. Relative clauses can have two forms: an active and a passive voice. The phrase "The country that the queen rules" is an example of an active use of *that* while "The queen who ruled the country" is an example of the passive form. Let $\mathcal{T}$ be the usual set of grammatical types n and s and $C(\mathcal{T})$ the free compact closed category in which we model grammar. The types of relative pronouns are

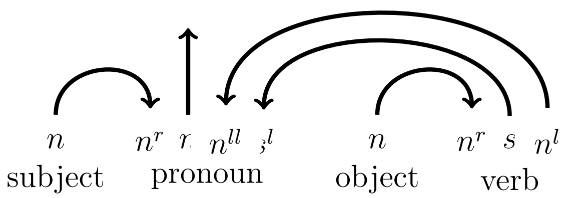
$$\begin{aligned} \text{Active:} \quad & n^r \cdot n \cdot s^l \cdot n \\ \text{Passive:} \quad & n^r \cdot n \cdot n^{ll} \cdot s^l \end{aligned}$$

resulting in the reductions

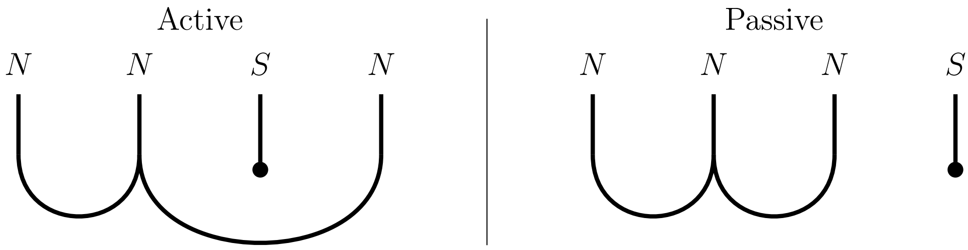
Active:



Passive:

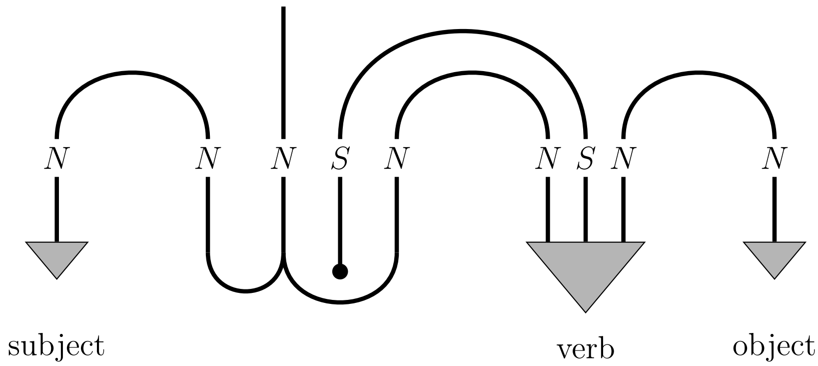


Secondly, we fix a semantics functor $MQ : C(\mathcal{T}) \to D(\mathbf{FdHilb})$. Using the operations of the Frobenius algebra in $D(\mathbf{FdHilb})$ we define relative pronouns as follows:

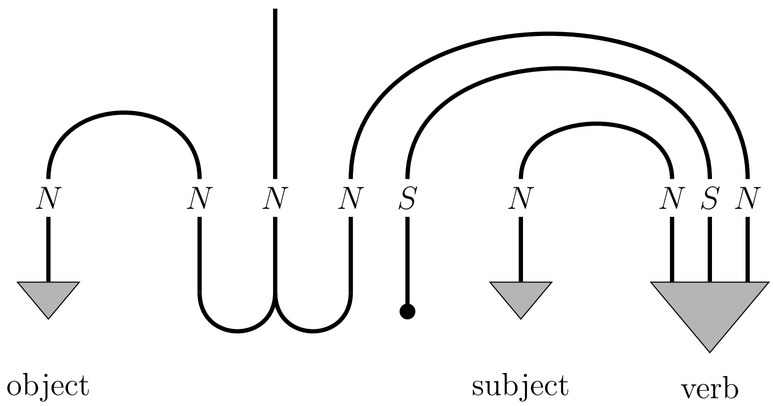


where $N \coloneqq MQ(n)$ and $S \coloneqq MQ(s)$.

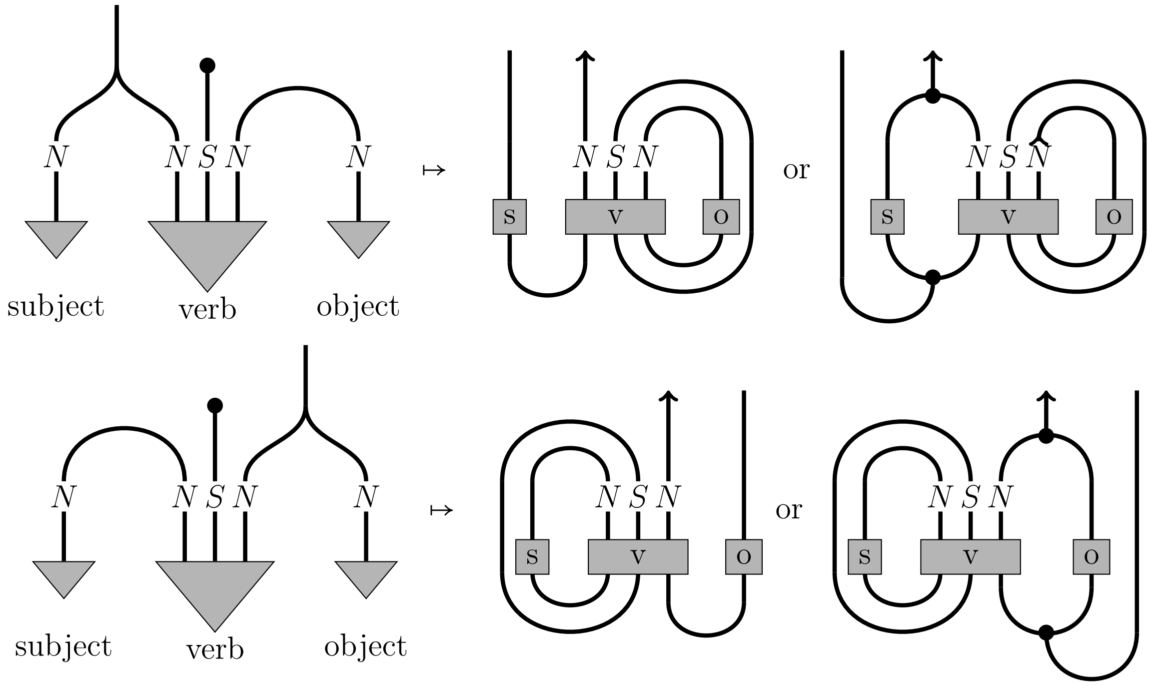
In order to understand this definition, let us depict how a relative pronoun relates the relative clause and the head noun-phrase together. First in the active form:



Then in the passive form:



By applying the rewriting rules for †-Frobenius algebras, the yanking equalities (i.e. pulling all the wires) and the commutativity of the tensor product in $D(\mathbf{FdHilb})$, these two diagrams can be reduced to the following normal forms (whose embedding in **FdHilb** is given in the language of $\mathcal{F}_D$ and $\mathcal{F}_C$):

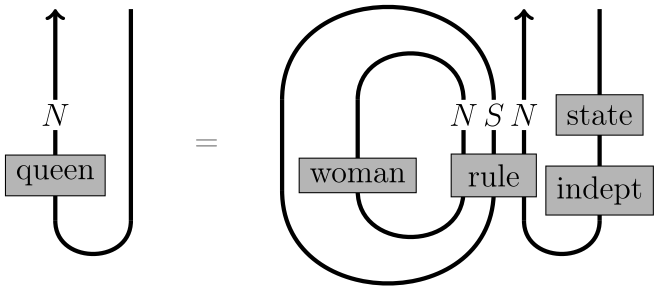


These pictures give us a better idea of the information flow that occurs through a relative pronoun: in a sentence of the type "noun-phrase relative-pronoun relative-clause", the information of the relative clause is computed by modifying the verb with the rest of the clause and by discarding the transitive sentence information that is not needed. Finally, the modified verb acts on (i.e., is composed with) the noun-phrase.

4.1.2 Meaning of unknown words

Here, we will be interested in recovering the meaning of an unknown word contained in a definition assuming that we already know the meaning of the defined word. For instance, with the definition "a *queen* is a woman who rules an independent state", assuming we already have a meaning-vector assigned to *queen* in a distributional model space W, what can we infer about *woman*, knowing the meaning of all the other words intervening in the definition? In this setting, a definition will be a statement of the form "*word* is $definition$" where we assume that the verb *is* signifies an equality between the meaning of *word* and $definition$. If $definition$ is a function of an unknown parameter word x, we write $word = definition(x)$.

In general, the possibility of solving the equation $word = definition(x)$ is a matter of solving a system of linear equations. We can study a simple example involving a relative clause to understand the general situation. Again, consider the definition "a queen is a woman who rules an independent state" in which we want to solve for the word *state*.



Above is the graphical representation of the meaning of the definition according to the rules introduced previously (with the added rule for adjectives that was introduced in the previous chapter). Or, in symbolic form,

$$\mathrm{Tr}_{N,S}\big(\rho(rule)(\rho(woman)\otimes 1_S \otimes 1_N\big)\rho(indept)\rho(state) = \rho(queen) \tag{4.1}$$

It is clear that, if the operators $\rho(indent)$ and $Tr_{N,S}(\rho(rule)(\rho(woman)\otimes 1_S \otimes 1_N))$ are invertible, we can compute the meaning of *state*. This is simply an equation of the form:

$$\rho x = \sigma$$

where ρ, x and σ are operators on the same space. However, in practice the operators associated to words in a distributional model are not invertible; their range is a small subspace of a high-dimensional vector space. Nonetheless we can compute an approximate solution on an appropriate support using the Moore-Penrose pseudoinverse.

Assume now that we wish to compute the meaning of *woman*, knowing the meaning of all the other words. If $\rho(indept)\rho(state)$ is invertible, we get an equation of the form

$$\mathrm{Tr}_N(\rho(x\otimes 1_{N'})) = \sigma \tag{4.2}$$

Rewriting ρ_1 as $\sum_i \rho_i^l \otimes \rho_i^r$, where the ρ_i^l (resp. ρ_i^r) are independent linear operators on N (resp. N'), we get the equation

$$\sum_i \mathrm{Tr}(\rho_i^l x)\rho_i^r = \sigma$$

This is a linear equation for which there is a solution if and only if σ is a linear combination of the ρ_i^r. Again, this rarely happens in practice, but we may calculate the least square approximation of the solution.

All definitional equations involve successive resolution of equations of the form 4.1 or 4.2 - equation 4.1 corresponding to an application of the co-unit, and equation 4.2 to the application of the Frobenius multiplication. Since tracing out and composing are the only operations that we apply to obtain the meaning of a sentence from its individual parts, this is all we need to recover the meaning of a word.

4.2 Updating meaning

In a definition "*word* is *expression*" one can interpret the verb *to be*, not as an equality of the two meanings but as updating our information about the *word*. Thus, the definition " a *queen* is a woman who rules an independent state" does not compel us to change the meaning of *queen* to that of "woman who rules independent state" but more subtly to update the meaning of *queen* if it did not take that piece of information into account.

We are going to provide a procedure to update the meaning of a word based on the information content of its definition.

4.2.1 Compatibility of two different meanings

First, we need to check that the information that the definition provides is compatible with what we already know about the defined *word*.

Fix a distributional model space W and two operators ρ^w, ρ^d, representing the meaning of a word w and its definition d. We have seen that the meaning of a word can be identified with the probability measure it induces on the lattice of subspaces (or equivalently, projectors) of W. The probability measure associated to ρ^w quantifies our state of belief about the meaning of the word w. For a projector P, $\mathrm{Tr}(P\rho^w)$ is the probability that the meaning of w is related to that of the atomic concepts in the support of P, that is, the set of atomic concepts whose span is the range of P. The lattice of projectors of W constitutes the set of events or properties that an expression can satisfy. A property is *almost certain* if the probability of its associated projector P is one, i.e $\mathrm{Tr}(P\rho) = 1$. Abusing notation, we will also say that a projector is almost certain relative to some density operator.

In this context, what we are looking for is a criterion of compatibility of two probability distributions. While there is no definitive answer to this vague problem, we will justify a simple qualitative rule introduced in [7], providing an answer in the special case of the compatibility of two density operators.

Our compatibility criterion should reflect the objective that we have, namely to obtain a new meaning operator for w that takes into account the information provided by d. Our criterion will be a simple consistency requirement: we want to obtain a meaning assignment for which almost certain properties of w and d are equally almost certain. Formally,

Definition 4.2.1. *We say that d is a compatible definition of w if there exists a density operator ρ such that, for all projectors P, if* $\mathrm{Tr}(P\rho) = 1$ *then* $\mathrm{Tr}(P\rho^w) = 1 = \mathrm{Tr}(P\rho^d)$.

Linguistically, it is a very weak yet, essential requirement: we want to update the meaning of w to include the certain information provided by its definition. The updated meaning should at least satisfy the almost certain properties of d. If such a state does not even exist, it seems sensible to judge the definition incompatible with the information about w in our current model since we cannot update w with the information of d.

Now, recall that the order on the lattice of projectors of W is defined as

$$P \leq Q \quad \text{if and only if} \quad QP = P$$

In this lattice, the meet $P \wedge Q$ is the projector onto the intersection of the ranges of P and Q; the join $P \vee Q$ is the projector onto the smallest closed subspace of W containing the union of the ranges of P and Q. We now prove two useful lemmas.

Lemma 4.2.1. *Let R be the projector onto the range of ρ. For a projector P,* $\mathrm{Tr}(P\rho) = 1$ *if and only if $R \leq P$.*

Proof. Let $\rho = \sum_i \rho_i |i\rangle\langle i|$ be the eigen-decomposition of ρ. If $\mathrm{Tr}(P\rho) = 1$ we have $\sum_i \rho_i \langle i|P|i\rangle = 1$ but, since $0 \leq \langle i|P|i\rangle \leq 1$, $0 \leq \rho_i \leq 1$, for all i we deduce $\langle i|P|i\rangle = 1$. Because $\langle i|i\rangle = 1$, we get $P|i\rangle = |i\rangle$ and, finally, since $R = \sum_i |i\rangle\langle i|$, $PR = R$ or $R \leq P$.

For the converse, if $R \leq P$, $\mathrm{Tr}(P\rho) = \mathrm{Tr}(PR\rho) = \mathrm{Tr}(R\rho) = \mathrm{Tr}(\rho) = 1$, because $R\rho = \rho$. □

Lemma 4.2.2. *For density operator ρ and projectors P and Q,* $\mathrm{Tr}((P \wedge Q)\rho) = 1$ *if and only if* $\mathrm{Tr}(P\rho) = 1 = \mathrm{Tr}(Q\rho)$.

Proof. Let R be the projector onto the range of ρ. By lemma 4.2.1, $\mathrm{Tr}(P\rho) = 1 = \mathrm{Tr}(Q\rho)$ if and only if $R \leq P$ and $R \leq Q$, that is, if and only if $R \leq P \wedge Q$. This last condition is equivalent to $\mathrm{Tr}((P \wedge Q)\rho) = 1$. □

Consequently, this lemma yields a necessary condition for the existence of a density operator that satisfies satisfying the condition of definition 4.2.1. Let R_w and R_d be the projectors onto the range of ρ_w and ρ_d respectively.

Proposition 4.2.1. *Definition d is compatible with w if and only if there exists a density operator ρ for whom $R_w \wedge R_d$ is almost certain, i.e such that* $\mathrm{Tr}(R_w \wedge R_d)\rho = 1$. *In particular, if $R_w \wedge R_d$ is the null projector, d is not a compatible definition of w.*

This condition was introduced in [7] and proven again in a more general Bayesian setting in [41].

4.2.2 Update rule

Now we wish to prove the existence of an operator that satisfies the condition of definition 4.2.1. In light of the previous proposition, we assume that $R_w \wedge R_d$ is non zero.

First, observe that if $R_w \wedge R_d = R_w$ there is nothing to do: all the almost certain properties of d are almost certain properties of w and there is no need to update the meaning operator ρ_w. Linguistically, this is the case when the definition is too general and provides no information about w. For example, " a queen is a monarch" is likely to provide very little information and require no updating of the meaning of *queen*.

In the general case, we want to update ρ_w to its best approximation in the subspace of density operators for which $R_w \wedge R_d$ is almost certain. To achieve this we need to define precisely the notion of approximation with which we wish to operate. The metric that we choose is the Hilbert-Schmidt distance. It has been argued that it provides an operational measure of indistinguishability of quantum states [38], in the sense of measuring how close the probabilistic predictions of two quantum states are relative to a complete set of mutually complementary observables. Therefore, the Hilbert-Schmidt distance can be interpreted as an information distance. Furthermore, the use of this specific distance will be justified by proposition 4.2.3.

Recall the notation $M \star N = M^{1/2} N M^{1/2}$.

Proposition 4.2.2. *For a projector P, the map $\rho \mapsto P \star \rho$ is a projection onto the space of positive operators for which P is almost certain.*

Proof. See Herbut [27]. □

Therefore, the updated state we are looking for in the case of a word w and its (compatible) definition w is, up to a normalisation factor,

$$\hat{\rho}^w = (R_w \wedge R_d)\rho^w(R_w \wedge R_d)$$

This formulation corresponds to the state update that occurs after a quantum measurement[1]. Thus, we can see definitions as performing a measurement on the word they define. For consistency, we need to check that the probabilities of all the properties that imply P are invariant by the mapping $\rho \mapsto P \star \rho$. In fact, it turns out that the only operator that has this property is $P \star \rho$, by the following proposition.

[1] Called Lüders' rule.

Proposition 4.2.3. *Let P be a projector and ρ, ρ' two density operators. We write $P^{\downarrow} = \{Q | Q \leq P\}$. We have*

$$\mathrm{Tr}(Q\rho) = \mathrm{Tr}(Q\rho'), \text{ for all } Q \in P^{\downarrow}, \quad \text{if and only if} \quad \rho' = P\rho P$$

Proof. See the proof of the related claim in Herbut [28, Lemma 4]. □

Thus, we have presented a sound and conceptually motivated rule to update the meaning of a word given its definition. Of course, this rule can be applied to any statement about the word and provides a general implementation of a learning algorithm in a compositional distributional model, allowing incremental knowledge acquisition. This is equivalent to a feedback mechanism: the compositional possibilities built on top of an existing distributional model can serve to improve the latter.

Chapter 5

Conclusion and future work

In this dissertation, we have extended the categorical compositional model of [12, 18, 15] to account for the compositional aspects of ambiguity and polysemy in natural language. We have found that a quantum physical mixing construction on the †-compact category of finite dimensional Hilbert spaces finds a natural linguistic interpretation and that the associated diagrammatic language provides an intuitive understanding of the flow of information between words of a sentence.

There are many directions of future research, some of which were suggested earlier and discussed in context in the main development.

Although conceptually motivated and based on an existing successful model of meaning, the real test of validity of the model that we propose will be experimental. Evaluating compositional models of meaning is not trivial and depends on the target applications. To test the quality of the compositional process the most obvious tasks involve comparing the meaning of sets of sentences with properties that witness certain features of language. For instance, a concrete task on which our model could and should be tested is the disambiguation of polysemous verbs: given a set of verbs with more than one identified meaning, the sentences in which they occur should provide enough information to disambiguate them. For a description of this task, see [15, Section 5.2]. In real data applications the complexity and efficiency of our model needs to be examined closely. At first glance it requires a quadratic increase in resources from the pure state model since the basic types are now represented by matrices and not vectors. Optimisation techniques need to be investigated to reduce the overhead.

Empirical evidence is also necessary to determine which Frobenius algebra yields the best results and to which task each is the most suited. In parallel we would like to undertake the more theoretical endeavour of investigating the relationship of the

CP* construction with our linguistic model. Considering a broader range of Frobenius structures may lead to more flexible ways of composing meaning in our model.

A weakness of distributional models is the meaning of words that serve a purely logical role, like the logical connective *then* or the negation. These are sometimes called noise words because their omnipresence in language dilutes their meaning into a useless distributional representation. Density operators support a form of logic whose distributional and compositional properties could be examined. Again, we may equally overcome these limitations through some insightful application of different Frobenius structures. In fact, the link between the types of logical connectives and Frobenius algebras was already drawn by Hines [29].

While we briefly mentioned entropy as a measure of ambiguity, its linguistic interpretation could be exploited further. Since it is based on a measure of semantic proximity, the current model gives good predictions about the synonymy between words and expressions. It is possible that the greater expressiveness of density operators allows us to extend those predictions to more complex semantic relations such as meronymy or hyponymy. For instance, common measures of relative information such as the Kullback-Leibler divergence could be useful in unraveling the distributional counterpart of these more intricate lexical hierarchies. In addition, density operators can be (partially) ordered in various ways [1] and it is worthy to study which of these orders conceals useful linguistic information.

It should also be noted that the D and CPM constructions can be applied to any †-compact categories and, in particular to $D(\mathbf{FdHilb})$ or $CPM(\mathbf{FdHilb})$ themselves. Applying them twice could accommodate a second level of mixing to describe information content more explicitly while keeping the current layer intact.

Furthermore, we would like to extend the compositional model to more expressive categorial grammars such as Combinatorial Categorial Grammars [54] and Lambek-Grishin algebras [45], more directly applicable to large corpora.

Finally, the results of section 4.2 need not be restricted to the case of definitions. They can be seen in the broader context of two agents communicating and updating their states of knowledge with the information that they share. If meaning is use, it cannot be detached from communication: language is fundamentally a dialogue between at least two people and meaning arises not in isolation but in this interaction. The work of Coecke and Spekkens [19] building on Leifer and Poulin [40] provides a framework in which quantum states, seen as epistemic states, can be updated according to generalised Bayesian principles. In this setting, quantum dynamics is

[1] For one such order, see [13].

understood as a belief revision and propagation mechanism. The connections with the dynamics of ordinary language dialogue constitute an avenue for future research.

Bibliography

[1] Lowell Abrams. Two-dimensional topological quantum field theories and frobenius algebras. *Journal of Knot Theory and its Ramifications*, 5(05):569–587, 1996.

[2] Samson Abramsky and Bob Coecke. A categorical semantics of quantum protocols. In *Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on*, pages 415–425. IEEE, 2004.

[3] Samson Abramsky and Bob Coecke. Categorical quantum mechanics. *Handbook of quantum logic and quantum structures: quantum logic*, pages 261–324, 2008.

[4] John Baez and Mike Stay. *Physics, topology, logic and computation: a Rosetta Stone*. Springer, 2011.

[5] Howard Barnum, Carlton M Caves, Christopher A Fuchs, Richard Jozsa, and Benjamin Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76(15):2818, 1996.

[6] Gemma Boleda, Sabine Schulte im Walde, and Toni Badia. Modeling regular polysemy: A study on the semantic classification of catalan adjectives. *Computational Linguistics*, 38(3):575–616, 2012.

[7] Todd A Brun. How much state assignments can differ. *arXiv preprint quant-ph/0208088*, 2002.

[8] Wojciech Buszkowski. Lambek grammars based on pregroups. In *Logical aspects of computational linguistics*, pages 95–109. Springer, 2001.

[9] S. Clark, B. Coecke, and M. Sadrzadeh. A compositional distributional model of meaning. *Proceedings of the Second Symposium on Quantum Interaction*, pages 133–140, 2008.

[10] S. Clark and S. Pulman. Combining symbolic and distributional models of meaning. *Proceedings of AAAI Spring Symposium on Quantum Interaction*, 2007.

[11] S. Clark and S. Pulman. Combining symbolic and distributional models of meaning. *Proceedings of AAAI Spring Symposium on Quantum Interaction*, 2007.

[12] Stephen Clark, Bob Coecke, and Mehrnoosh Sadrzadeh. A compositional distributional model of meaning. In *Proceedings of the Second Quantum Interaction Symposium (QI-2008)*, pages 133–140, 2008.

[13] Bob Coecke. Entropic geometry from logic. *Electronic Notes in Theoretical Computer Science*, 83:39–53, 2013.

[14] Bob Coecke and Ross Duncan. Interacting quantum observables. In *Automata, Languages and Programming*, pages 298–310. Springer, 2008.

[15] Bob Coecke, Edward Grefenstette, and Mehrnoosh Sadrzadeh. Lambek vs. lambek: Functorial vector space semantics and string diagrams for lambek calculus. *Annals of Pure and Applied Logic*, 164(11):1079–1100, 2013.

[16] Bob Coecke, Chris Heunen, and Aleks Kissinger. Categories of quantum and classical channels. *arXiv preprint arXiv:1305.3821*, 2013.

[17] Bob Coecke, Dusko Pavlovic, and Jamie Vicary. A new description of orthogonal bases. *arXiv preprint arXiv:0810.0812*, 2008.

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

[19] Bob Coecke and Robert W Spekkens. Picturing classical and quantum bayesian inference. *Synthese*, 186(3):651–696, 2012.

[20] J. R. Firth. A synopsis of linguistic theory 1930-55. 1952-59:1–32, 1957.

[21] J.R. Firth. A synopsis of linguistic theory 1930-1955. *Studies in linguistic analysis*, pages 1–32, 1957.

[22] Georg Frobenius. *Theorie der hyperkomplexen Grössen II.* Königliche Akademie der Wissenschaften, 1903.

[23] Andrew M Gleason. Measures on the closed subspaces of a hilbert space. *Journal of mathematics and mechanics*, 6(6):885–893, 1957.

[24] Edward Grefenstette. *Category-Theoretic Quantitative Compositional Distributional Models of Natural Language Semantics*. PhD thesis, June 2013.

[25] Edward Grefenstette and Mehrnoosh Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404. Association for Computational Linguistics, 2011.

[26] Edward Grefenstette, Mehrnoosh Sadrzadeh, Stephen Clark, Bob Coecke, and Stephen Pulman. Concrete sentence spaces for compositional distributional models of meaning. In *Computing Meaning*, pages 71–86. Springer, 2014.

[27] F Herbut. Derivation of the change of state in measurement from the concept of minimal measurement. *Annals of Physics*, 55(2):271–300, 1969.

[28] F Herbut. On compatibility and improvement of different quantum state assignments. *Journal of Physics A: Mathematical and General*, 37(19):5243, 2004.

[29] Peter Hines. Types and forgetfulness in categorical linguistics and quantum mechanics. *CoRR*, abs/1303.3170, 2013.

[30] André Joyal and Ross Street. The geometry of tensor calculus, i. *Advances in Mathematics*, 88(1):55–112, 1991.

[31] Dimitri Kartsaklis and Mehrnoosh Sadrzadeh. A study of entanglement in a categorical framework of natural language. *arXiv preprint arXiv:1405.2874*, 2014.

[32] Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, Stephen Pulman, and Bob Coecke. Reasoning about meaning in natural language with compact closed categories and frobenius algebras. *Logic and Algebraic Structures in Quantum Computing and Information, Association for Symbolic Logic Lecture Notes in Logic. Cambridge University Press.(To appear)*, 2013.

[33] Gregory M Kelly and Miguel L Laplaza. Coherence for compact closed categories. *Journal of Pure and Applied Algebra*, 19:193–213, 1980.

[34] Anna Korhonen, Yuval Krymolowski, and Zvika Marx. Clustering polysemic subcategorization frame distributions semantically. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 64–71. Association for Computational Linguistics, 2003.

[35] Greg Kuperberg. A concise introduction to quantum probability, quantum mechanics, and quantum computation. *unpublished notes*, 2005.

[36] Joachim Lambek. Type grammar revisited. In *Logical aspects of computational linguistics*, pages 1–27. Springer, 1999.

[37] Joachim Lambek. *From Word to Sentence: a computational algebraic approach to grammar*. Polimetrica sas, 2008.

[38] Jinhyoung Lee, MS Kim, and Časlav Brukner. Operationally invariant measure of the distance between quantum states by complementary measurements. *Physical review letters*, 91(8):087902, 2003.

[39] Matthew S Leifer and Robert W Spekkens. Towards a formulation of quantum theory as a causally neutral theory of bayesian inference. *Physical Review A*, 88(5):052130, 2013.

[40] MS Leifer and David Poulin. Quantum graphical models and belief propagation. *Annals of Physics*, 323(8):1899–1946, 2008.

[41] MS Leifer and Robert W Spekkens. A bayesian approach to compatibility, improvement, and pooling of quantum states. *arXiv preprint arXiv:1110.1085*, 2011.

[42] Saunders Mac Lane. *Categories for the working mathematician*, volume 5. springer, 1998.

[43] Richard Montague. *English as a formal language*. Ed. di Comunità, 1970.

[44] Michael Moortgat. *Categorial investigations: Logical and linguistic aspects of the Lambek calculus*. Number 9. Walter de Gruyter, 1988.

[45] Michael Moortgat. Symmetric categorial grammar. *Journal of Philosophical Logic*, 38(6):681–710, 2009.

[46] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[47] Anne Preller. From logical to distributional models. In *QPL: International Workshop on Quantum Physics and Logic-EPTCS: Electronic Proceedings in Theoretical Computer Science*, 2014.

[48] Anne Preller and Joachim Lambek. Free compact 2-categories. *Mathematical Structures in Computer Science*, 17(02):309–340, 2007.

[49] Mehrnoosh Sadrzadeh, Stephen Clark, and Bob Coecke. The frobenius anatomy of word meanings i: subject and object relative pronouns. *Journal of Logic and Computation*, page ext044, 2013.

[50] Mehrnoosh Sadrzadeh, Stephen Clark, and Bob Coecke. The frobenius anatomy of word meanings ii: possessive relative pronouns. *Journal of Logic and Computation*, page exu027, 2014.

[51] Peter Selinger. Dagger compact closed categories and completely positive maps. *Electronic Notes in Theoretical Computer Science*, 170:139–163, 2007.

[52] Peter Selinger. Finite dimensional hilbert spaces are complete for dagger compact closed categories. *Electronic Notes in Theoretical Computer Science*, 270(1):113–119, 2011.

[53] Sylvia Springorum, Sabine Schulte im Walde, and Jason Utt. Detecting polysemy in hard and soft cluster analyses of german preposition vector spaces. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pages 632–640, 2013.

[54] Mark Steedman. *The syntactic process*, volume 35. MIT Press, 2000.

[55] Peter D Turney, Patrick Pantel, et al. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1):141–188, 2010.

[56] Jamie Vicary. Categorical formulation of finite-dimensional quantum algebras. *Communications in Mathematical Physics*, 304(3):765–796, 2011.

[57] Ludwig Wittgenstein. *Philosophical investigations.* John Wiley & Sons, 2010.

[58] Vera I Zabotkinaa and Elena L Boyarskayab. Sense disambiguation in polysemous words: cognitive perspective. *Psychology in Russia. State of the Art*, 6(3), 2013.