W-SPIDERS

Juliette Buet

MSc in Computer Science
University of Oxford
Green Templeton College

September 2017

Abstract

The Distributional Model of Meaning is a promissing new way to represent and understand language and linguistics using tools developed for Quantum Computer Science. Representing words whose meaning depends on the sentence (such as 'WHICH', 'TO', 'BUT', etc) requires more work within this model; in particular coordination has already been studied by Dimitri Kartsaklis with ideas for representing 'AND' using GHZ-spiders, but more work is required on the subject. This thesis will focus on the representation of 'AND' and 'OR' and in particular the use of GHZ and W spiders to represent each of them respectively. The use of W-spiders to represent 'OR' comes naturally when knowing Dr Kartsaklis' idea for 'AND' as there are distributivity laws between GHZ and W spiders similar to those that exist between 'AND' and 'OR'. Some caveats regarding Dr Kartsaklis's approach have been found and fixed syntactically and a partial solution to the representation of 'OR' has been found as well in the vector space Rel.

Acknowledgements

I would like to give a very warm thank you to my advisor Bob Coecke for answering my questions throughout this project, even my last minute emails, and for always being available for me. I would also like to thank Amar Hadzihasanovic and Quanlong Wang who accepted to meet with me and answer my questions when I was looking for a way to represent W-spiders in higher dimension. Thank you as well to Robin Piedeleu who helped me understand the field as I could not the class on the Compositional Distributional Model of Meaning as part of my master.

There are also a few members of the Computer Science Department and of the Green Templeton College administrations that I would like to thank as they helped me navigate this year and more specifically this project in terms of deadline and requirements. So a big thank you to Sarah Retz-Jones, Shoshannah Holdom and Alison Franklin.

Finally, a few personal thank you to people around me who have beared with me through the ups and downs that a research project can have, and through the stress that goes with it. So thank you very much to Nicola Psaila and Jean Laurens.

Contents

Introduction

Natural language processing is an on going topic of research with various attempts at making computers better understand natural language, notably in order to have better performing software for speech recognition. While a lot of modern research focuses on machine learning and neural networks to analyse natural language, this thesis will look at a more mathematical approach. A general grammar already exists, known as the Categorical Compositional Distributional Model of Meaning, which will be presented in chapter 2. But this grammar isn't complete. The purpose of this thesis is to find ways to specifically represent the conjunction and disjunction, usually known in English by 'AND' and 'OR'.

Several recent papers ([1] [2] [3][4][5][6]) have shown that Frobenius algebras offer additional properties to the categorical grammar that have proved to be useful to represent relative pronouns and conjunction. We will further explore this direction, in particular with two kinds of Frobenius algebras generated by GHZ and W states, two types of tripartite entangled states, as presented in the first chapter.

The first two chapters of this thesis focus on background research necessary to understand current work, on ZW-Calculus and on the Categorical Compositional Distributional Model of Meaning. The third chapter starts with Kartsaklis' work on conjunction and then we present a few issues that we have found with this representation as well as potential solutions. Finally, the fourth and last chapter focuses on our work on disjunction and its representation by W-spiders. The first section of that last chapter presents Hadzihasanovic's work on generalising the W operator for Qudits, which we need in order to have models of natural language with more than a two-dimensional basis.

Chapter 1

Elements of ZW-Calculus

Multipartite quantum states and entanglement are two major elements of quantum physics. Categorical quantum mechanics (CQM) is a branch of quantum computing that combines these features with algebraic notions such as Commutative Frobenius Algebras (CFA) to better understand singular quantum effects. ZW-Calculus [7] aims at studying how two specific laws of such algebras, namely the GHZ spider and the W spider, interact with each other. The goal of this chapter is to remind the reader of a result that shows that a GHZ state and a W state, which are tripartite entangled quantum states, can be developed into Frobenius laws and used to build a complex graphical calculus. It is on such calculus that the interpretation of sentences presented in this thesis is based.

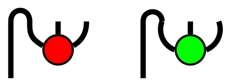
1.1 Introduction

Entanglement is a special feature that does not exist in classical physics which corresponds to a correlation in measurement of spatially seperated compound quantum systems. Entangled states can be composed of two, three or more compounds. Bipartite entangled states are used in the quantum teleportation protocol among others. This thesis will mainly look at tripartite entangled states and more specifically two kinds GHZ and W states - which, as we will show later, are different from each other. Entanglement remains an on-going subject of research, there isn't even an agreed upon classification for general multipartite entangled states for an order superior to three, notably because there is en infinite number of 4 (or more)-qubit states that are not SLOCC (stochastic local operations and classical communications) equivalent. For three qubits however, SLOCC classification is well understood and there are two non-degenerated classes that each contain the GHZ and W states respectively, and are usually represented by them.

$$|GHZ\rangle = |000\rangle + |111\rangle$$

$$|W\rangle = |100\rangle + |010\rangle + |001\rangle$$

The differences between these two states are essentially topological, relative to how they behave regarding to loops in their area. The first approach is to consider tripartite states as algebraic operations. In the graphical representation of states, this corresponds to bending one wire and turn the state into a process with one input and two outputs. This graphical representation will be explained in more detail in the next section.



Indeed, the relevant information is the arity (number of inputs + number of outputs) of the state, and not their precise position thanks to the yanking rule. This leads to the definition of Frobenius states, which lead to the creation of CFAs, a particularly well-behaved kind of algebra.

The rest of this chapter will brifely explain the graphical representation used, remind the reader of definitions and properties around Commutative Frobenius Algebras, explain the SLOCC classification and more specifically that of tripartite states, define Frobenius states, remind the reader more specifically about special and anti-special Frobenius algebras and finally explain how GHZ and W states can generate Frobenius algebras.

1.2 Graphical notations

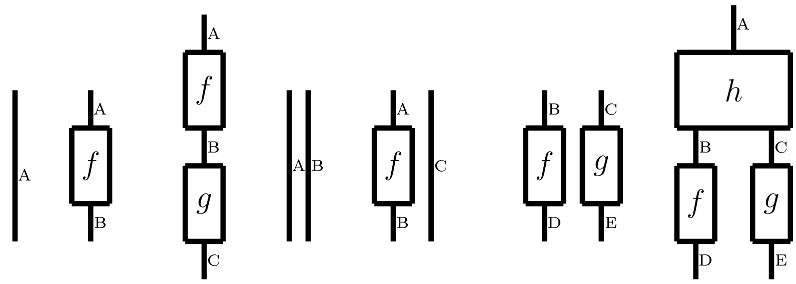
This section will quickly go over a few graphical representations of monoidal categories. Morphisms are depicted as boxes with wires on each side of the box corresponding to inputs and outpus of the morphism. When types are ambiguous, these wires will be labelled with the corresponding types.

Sequential composition is simply depicted by connecting wires (that have to correspond to the same type), and parallel composition by juxtaposing boxes side by side.

For example, the morphisms:

$$1_A \qquad f \qquad g \circ f \qquad 1_A \otimes 1_B \qquad f \otimes 1_C \qquad f \otimes g \qquad (f \otimes g) \circ h$$

are depicted as follows in a top-down fashion:

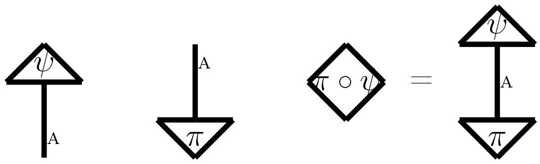


The unit object is represented by 'no wire' with a triangular shape rather than flat.

For example:

$$\psi : I \to A \qquad \pi : A \to I \qquad \pi \circ \psi : I \to I$$

are depicted as:



1.3 Commutative Frobenius Algebra

Let's start by recalling the definition of a unital algebra.

Definition 1.3.1. [7] A unital algebra is a vector space A with a multiplication $\cdot$ that verifies the following properties:

- $(\_ \cdot \_)$ is bilinear
- $(\_ \cdot \_)$ is associative: for every vectors $|u\rangle, |v\rangle, |w\rangle$:

$$|u\rangle \cdot (|v\rangle \cdot |w\rangle) = (|u\rangle \cdot |v\rangle) \cdot |w\rangle$$

- there exists a unit $|\eta\rangle$ such that for all $|u\rangle$ in A, $|\eta\rangle \cdot |u\rangle = |u\rangle \cdot |\eta\rangle = |u\rangle$

In this thesis, we will assume that A is a finite-dimensional Hilbert space H. Therefore, since $(\_ \cdot \_)$ is bilinear, there exists a unique μ such that:

$$\mu(|u\rangle \otimes |v\rangle) = (|u\rangle \cdot |v\rangle)$$

We then obtain the following definition:

Definition 1.3.2. [7] A unital algebra (H, μ, η) is a vector space H with maps $\mu : H \otimes H \to H$ and $\eta : \mathbb{C} \to H$ that verify the following equations:

- $\mu(1 \otimes \mu) = \mu(\mu \otimes 1)$
- $\mu(1 \otimes \eta) = \mu(\eta \otimes 1) = 1$

Similarly, we define a counital coalgebra on the dual space H^*. In terms of H:

Definition 1.3.3. [7] A counital coalgebra (H, δ, ϵ) is a vector space H with maps $\delta : H \to H \otimes H$ and $\epsilon : H \to \mathbb{C}$ that verify the following equations:

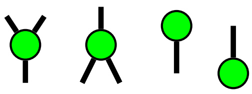
- $(1 \otimes \delta)\delta = (\delta \otimes 1)\delta$
- $(\epsilon \otimes 1)\delta = (1 \otimes \epsilon)\delta = 1$

From these definitions, we can define a Frobenius algebra:

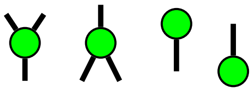
Definition 1.3.4. [7] A Frobenius algebra $\mathscr{F}$ is a vector space H with maps $\mu, \eta, \delta, \epsilon$ such that:

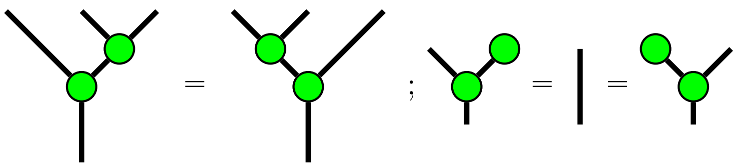
- (H, μ, η) is a unital algebra
- (H, δ, ϵ) is a counital coalgebra
- $(\mu \otimes 1)(1 \otimes \delta) = (1 \otimes \mu)(\delta \otimes 1) = \delta\mu$ (Frobenius identity)

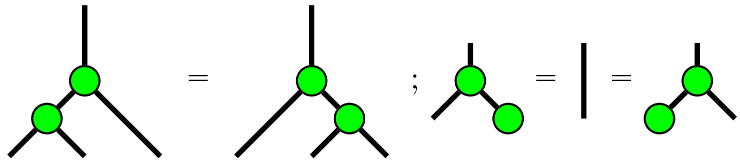
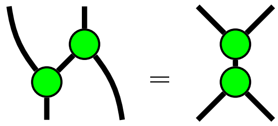
Graphically, we can depict μ ,δ ,η ,ϵ as:



Therefore, graphically the definition of the Frobenius algebra is:

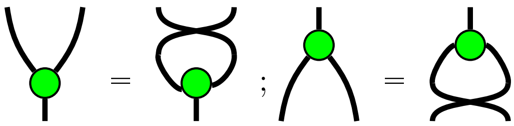
Definition 1.3.5. [7] Frobenius algebra  is a a vector space H with maps  such that:

- 

- 
- 

Let $\sigma_{A,B}$ be the swap map. Then a unital algebra (resp counital coalgebra) is commutative (resp cocoummtative) iff $\mu = \mu\sigma_{H,H}$ (resp $\delta = \sigma_{H,H}\delta$).

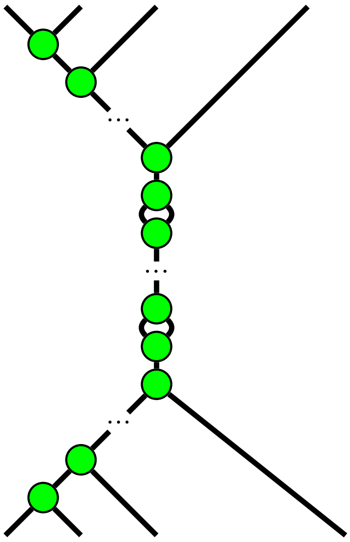
Definition 1.3.6. [7] A Frobenius algebra $(H, \mu, \delta, \eta, \epsilon)$ is called commutative iff the unital algebra (H, μ, η) is commutative and the counital coalgebra (H, δ, ϵ) is cocommutative. Graphically:



Definition 1.3.7. [7] For a commutative Frobenius algebra (CFA) $\mathscr{F} = (H, \mu, \delta, \eta, \epsilon)$, an $\mathscr{F}$-graph is a graph obtained from $1_H, \sigma_{H,H}, \mu, \delta, \eta$ and ϵ, combined with composition and tensor product. An $\mathscr{F}$-graph is said to be connected precisely whan its graphical representation is connected.

A well-known result about CFAs is that any $\mathscr{F}$-graph is entirely determined by its number of inputs, outputs and loops - where by the number of loops we mean the maximum number of edges that can be removed without disconnecting the graph. This result was proven by J. Kock in [8]. As order doesn't really matter thanks to commutativity and associativity, it is possible to transform any $\mathscr{F}$-graph in a normal form.

Theorem 1.3.8. [7] Any connected $\mathscr{F}$-graph can be put in a normal form where loops are in the middle as presented in the following diagram:

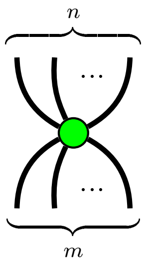
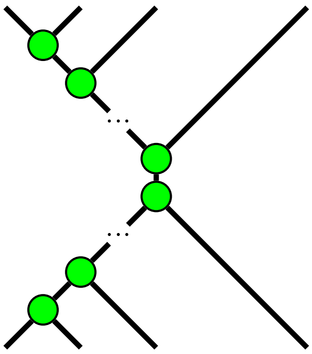


(1.1)

A particular instance of this normal form is the case where there are no loops. In this case, an $\mathscr{F}$-graph is an $\mathscr{F}$-tree and it can be represented as a spider.

Definition 1.3.9. [7] A connected $\mathscr{F}$-tree is called a spider. It is uniquely determined by its number of in-edges and out-edges.

$$S_m^n : \underbrace{H \otimes \ldots \otimes H}_{m} \rightarrow \overbrace{H \otimes \ldots \otimes H}^{n}$$

$S_m^n :=$  $=$  (1.2)

In particular for spiders without inputs or outputs, we cap off with η or ϵ.

1.4 SLOCC classification

In this section we recall some facts about SLOCC classification as it is the system in which GHZ- and W-spiders appear as canonical.

Definition 1.4.1. [7] A state $|\psi\rangle \in H_1 \otimes ... \otimes H_N$ can be converted into a state $|\phi\rangle$ through Stochastic Local Operations and Classical Communications (SLOCC) when there exists an N party protocol that succeeds with non-zero probability at turning $|\psi\rangle$ into $|\phi\rangle$, where each party p_i has access to H_i and can :

- apply any number of local (quantum) operations $O : H_i \to H_i$
- perform any amount of classical communication with the other parties

If it is the case, then we write $|\phi\rangle \preceq |\psi\rangle$.

The relation $\preceq$ forms a preorder and the resulting equivalence relation, called SLOCC-equivalence is written $\sim$. From this equivalence relation stems the notion of SLOCC class.

The following theorem on SLOCC-equivalence has been proven by W. Dür, G. Vidal and J. I. Cirac in [9].

Theorem 1.4.2. [7] Two states $|\phi\rangle, |\psi\rangle \in H_1 \otimes ... \otimes H_N$ are SLOCC-equivalent if and only if there exist invertible linear maps $L_i : H_i \to H_i$ such that :

$$|\phi\rangle = (L_1 \otimes ... \otimes L_N) |\psi\rangle .$$

We say that a state $|\phi\rangle$ is SLOCC-maximal if it is maximal with regard to $\preceq$, that is:

$$|\phi\rangle \preceq |\psi\rangle \Rightarrow |\phi\rangle \sim |\psi\rangle . \tag{1.3}$$

For two qubits, there are two SLOCC classes:

$$\begin{tikzcd} \boxed{|Bell\rangle} \arrow[d, dash, "\preceq"] \\ \boxed{|\phi\rangle \otimes |\psi\rangle} \end{tikzcd}$$

Where $|Bell\rangle$ is the state $|00\rangle + |11\rangle$ commonly known as the Bell state. Therefore, any entangled state made of two qubits can be derived from the Bell state using local invertible maps.

For three qubits, there are four SLOCC classes up to qubits permutation, but the $\preceq$ order is no longer total. The classes interact according to the following diagram:

$$\begin{tikzcd} \boxed{|GHZ\rangle} \arrow[dr, dash, "\preceq"'] & & \boxed{|W\rangle} \arrow[dl, dash, "\preceq"] \\ & \boxed{|Bell\rangle \otimes |\psi\rangle} \arrow[d, dash, "\preceq"] & \\ & \boxed{|\psi\rangle \otimes |\phi\rangle \otimes |\rho\rangle} & \end{tikzcd}$$

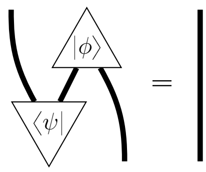
Where $|GHZ\rangle$ is the state $|000\rangle + |111\rangle$ and $|W\rangle$ is the state $|001\rangle + |010\rangle + |100\rangle$. Therefore, a maximally entangled state made of three qubits can be in one of the two classes GHZ or W.

However, for a number of qubits superior or equal to 4, there isn't a finite number of SLOCC-classes, as proven in [9], so the classification is more complicated. We will not enter in detail about such classification as it is not relevant to the rest of this thesis.

1.4.1 SLOCC-maximal entangled states

For any state $|\phi\rangle$, it can always be derived from a maxiamally entangled state and an SLOCC protocol. Graphically, SLOCC-maximality can be reformulated with the following proposition.

Theorem 1.4.3. [7] A bipartite state $|\phi\rangle : H \otimes H$ is a SLOCC-maximal entangled state iff there exists a corresponding effect $\langle\psi| : H \otimes H \to \mathbb{C}$ such that:



(1.4)

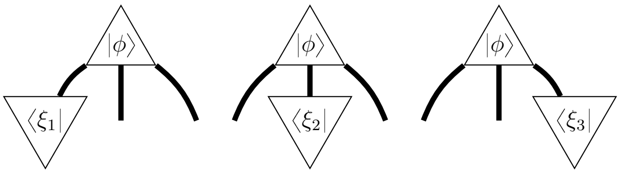
The proof of this result follows straight-forwardly by map-state duality.

1.5 Frobenius states

As has been alluded previously in this chapter, we are going to look at links between tripartite entangled states and algebraic structures. We will show that two properties are sufficient for a tripartite entangled state to generate such a structure: (a) satisfy a unit law, and (b) be associative, commutative and Frobenius. Condition (a) corresponds with being highly entangled and (b) with being highly symmetrical.

In this section, we will look at tripartite states $|\phi\rangle$ that are both highly entangled and highly symmetrical, but that also can be generalised into N-partite states that inherit this property of being able to generate an algebraic structure. First, we will introduce the notion of strong SLOCC-maximality and strong symmetry.

Definition 1.5.1. [7] A tripartite state $|\phi\rangle \in H \otimes H \otimes H$ is said to be strong SLOCC-maixmal if there exists three effects $\langle\xi_i|$ such that the three following states are all SLOCC-maximal bipartite states.

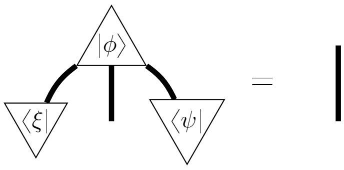


(1.5)

As proven in [7], strong SLOCC-maximality strictly implies SLOCC-maximality.

In the case where $|\phi\rangle$ is symmetric, this condition can be simplified:

Theorem 1.5.2. [7] A symmetric tripartite state $|\phi\rangle$ is strongly SLOCC-maximal iff there exists two effects $\langle\xi|$ and $\langle\psi|$ such that:



(1.6)

The GHZ state satisfies equation 1.6 by fixing $\langle\xi| = \langle+|$ and $\langle\psi| = \langle Bell|$. The W state also satisfies this equation with $\langle\xi| = \langle 0|$ and $\langle\psi| = \langle EPR| := \langle 01| + \langle 10|$.

As these are the only two SLOCC-maximal entangled tripartite states, and these are the candidates for representing coordinations in the next chapters, we are going to spend more time on them. They are both symmetric and they both have an N-partite generalisation :

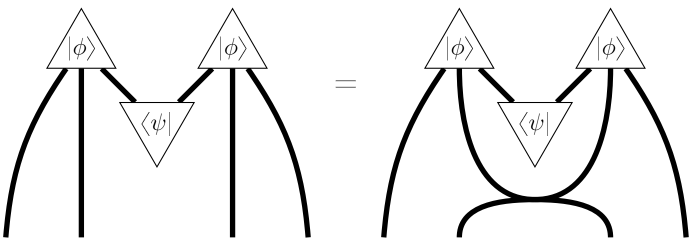
$$|GHZ_N\rangle := |00...0\rangle + |11...1\rangle \tag{1.7}$$

$$|W_N\rangle := |10...0\rangle + |010...0\rangle + ... + |00...01\rangle \tag{1.8}$$

The $N+1$ state can also be built inductively from the N state and tripartite states.

To inductively build symmetric states it suffices that the following condition holds.

Definition 1.5.3. [7] A state $|\phi\rangle$ is said to be strongly symmetric, if there exists a bipartite effect $\langle\psi|$ such that:



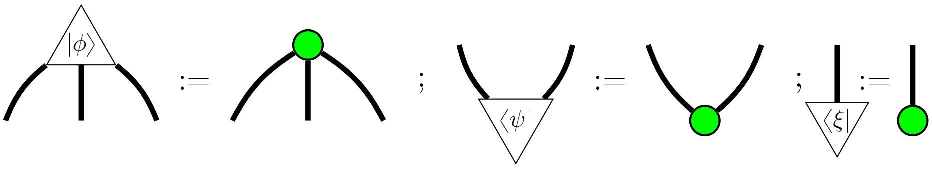
(1.9)

We have gone through all the necessary definitions and can now describe the states that have highly algebraic properties.

Definition 1.5.4. [7] A symmetric tripartite state $|\phi\rangle$ is said to be a Frobenius state if there exists two effects $\langle\xi|$ and $\langle\psi|$ such that equations 1.6 and 1.9 hold.

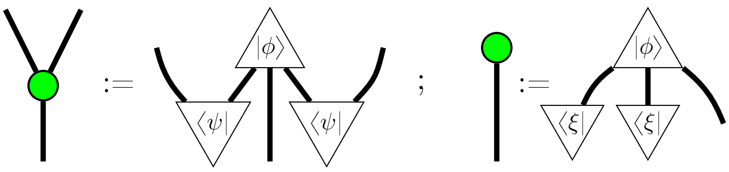
Note that the $\langle\psi|$ satisfying equations 1.6 and 1.9 must be the same effect, which is a stronger condition than having these two equations hold separately for some $\langle\psi|$ and $\langle\psi'|$.

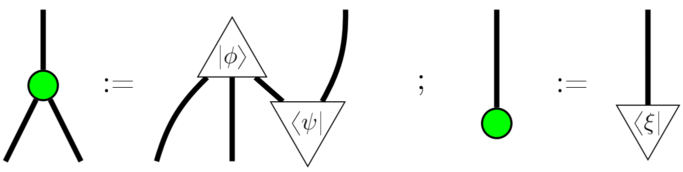
Theorem 1.5.5. [7] For any commutative Frobenius algebra  , the following is a Frobenius state with its two associated effects:



The fact that it corresponds to definition 1.5.4 is a consequence of theorem 1.3.8. Also, from a Frobenius state, it is possible to construct the corresponding Frobenius algebra.

Theorem 1.5.6. [7] For any Frobenius state $|\phi\rangle$ there exists effects $\langle\psi|$ and $\langle\xi|$ such that the following forms a commutative Frobenius algebra:



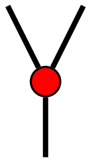
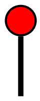


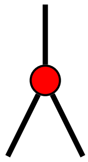
As a result, it is possible to refer to Frobenius algebras either through their usual maps $(\mu, \delta, \eta, \epsilon)$ or through the triplet $(|\phi\rangle, \langle\psi|, \langle\xi|)$. From a same state $|\phi\rangle$, it is possible to produce several Frobenius algebras by using several $\langle\xi|$. However, once, $\langle\xi|$ is fixed, $\langle\psi|$ follows immediately using equation 1.5.2 on strong SLOCC-maximality. This is analogous to the regular definition of Frobenius algebras where fixing the maps μ and ϵ completely determines the other two.

1.5.1 Frobenius algebras from $|GHZ\rangle$ and $|W\rangle$

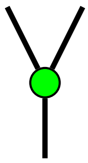
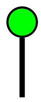
More specifically in this thesis, we are interested in the Frobenius algebras that come from the two states $|GHZ\rangle$ and $|W\rangle$.

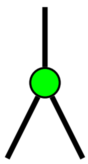
In the case of $|\phi\rangle = |GHZ\rangle$ we can fix $\langle\xi| = \sqrt{2}\langle+|$ which produces the following CFA that we shall refer to as $\mathscr{G}$:

 $= |0\rangle\langle 00| + |1\rangle\langle 11|$;  $= \sqrt{2}\,|+\rangle := |0\rangle + |1\rangle$

 $= |00\rangle\langle 0| + |11\rangle\langle 1|$;  $= \sqrt{2}\langle+| := \langle 0| + \langle 1|$ (1.10)

In the case of $|\phi\rangle = |W\rangle$ we can fix $\langle\xi| = \langle 0|$ which produces the following CFA that we shall refer to as $\mathscr{W}$:

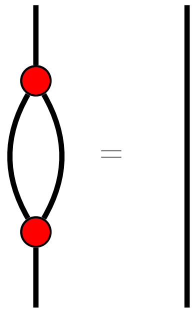
 $= |1\rangle\langle 11| + |0\rangle\langle 01| + |0\rangle\langle 10|$;  $= |1\rangle$

 $= |00\rangle\langle 0| + |01\rangle\langle 1| + |10\rangle\langle 1|$;  $= \langle 0|$ (1.11)

1.6 Special and anti-special Frobenius algebras

The normal form given in the theorem 1.3.8 suggests that loops play a key role in understanding Frobenius algebras. We will particularly look at two behaviours that these loops can have : one where they vanish and another one where they propagate outwards, disconnecting the entire graph. These behaviours are of interest because they correspond to the behaviours of CFAs produced from GHZ and W states.

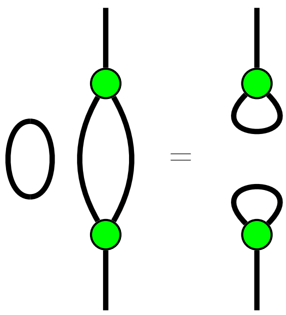
Definition 1.6.1. [7] A special commutative Frobenius algebra (SCFA) is a commutative Frobenius algebra where:



As proven in [7]:

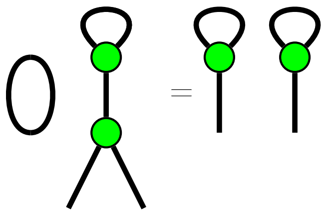
Theorem 1.6.2. [7] In a SCFA S, a connected S-graph with m inputs and n outputs is equal to the spider S_m^n.

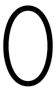
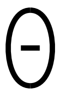
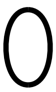
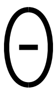
Definition 1.6.3. [7] An anti-special commutative Frobenius algebra (ACFA) is a commutative Frobenius algebra where:

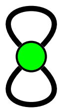


A few properties on ACFAs have been proven in [7].

Theorem 1.6.4. [7] For any ACFA, the following copying rule holds:



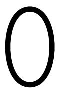
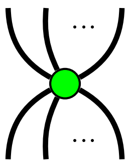
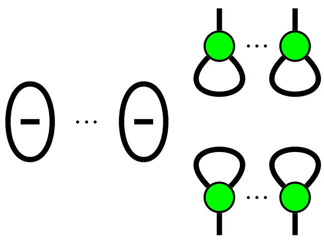
Contrary to SCFAs, scalars play a key role in characterizing ACFAs, in particular, the circle  - which can be shown to be equal to D, the dimension of the underlying Hilbert space. Assuming $D > 0$, let  $= 1/D$. Thus   $= 1$.

Theorem 1.6.5. [7] For any ACFA  $= (H, \mu, \delta, \eta, \epsilon)$, either $dim(H) = 1$ or  $= 0$

The proof is also available in [7].

From this point forward, we will assume that $dim(H) \geqslant 2$.

Theorem 1.6.6. [7] For any ACFA $\mathscr{A}$, any connected $\mathscr{A}$-graph is equal to one of the following:

- 0
- 
- 
- 

The proof can be found in [7].

1.7 GHZ and W states as commutative Frobenius algebra

As seen in the previous section, GHZ states and W states are both Frobenius states and therefore can each generate a commutative Frobenius algebra. Furthemore, for CFAs on $\mathbb{C}^2$ conditions of specialness and anti-specialness are enough to identify GHZ and W states up to SLOCC-equivalence.

As proven in [7]:

Theorem 1.7.1. [7] For any special commutative Frobenius algebra on $\mathbb{C}^2$, the induced Frobenius state is SLOCC-equivalent to $|GHZ\rangle$. Furthermore, for any tripartite state $|\phi\rangle$ that is SLOCC-equivalent to $|GHZ\rangle$, there exists $|\psi\rangle, |\xi\rangle$ such that $(|\phi\rangle, |\psi\rangle, |\xi\rangle)$ is a special commutative Frobenius algebra.

Similarly, [7] proves that:

Theorem 1.7.2. [7] For any anti-special Frobenius algebra in $\mathbb{C}^2$, the induced Frobenius state is SLOCC-equivalent to $|W\rangle$. Furthermore, for any tripartite state $|\phi\rangle$ that is SLOCC-equivalent to $|W\rangle$, there exists $|\psi\rangle, |\xi\rangle$ such that $(|\phi\rangle, |\psi\rangle, |\xi\rangle)$ is an anti-special commutative Frobenius algebra.

A direct corollary is that any SLOCC-maximal tripartite state is SLOCC-equivalent to a Frobenius state.

1.8 Summmary

In this chapter we have introduced the importance of commutative Frobenius algebras and given an alternative definition to Frobenius algebras using tripartite states. Two states in particular play a specific role, GHZ and W states, which induce respectively special and anti-special commutative Frobenius algebras. We will see later in this thesis that the Frobenius algebras created from these states can be used to represent 'OR' and 'AND' in a specific model of linguistics.

Chapter 2

Distributional Model of Meaning

In this chapter we will introduce a theory developed in [10] that aims at unifying the distributional model of meaning in terms of vector space models, and a compositional theory for grammatical types which relies on the algebra of Pregroups, developed by Lambek. The theory we are using makes it possible to compute the meaning of a well-typed sentence from the meaning of the words composing it. Using the fact that Pregroups can be viewed as a category, type reductions of Pregroups are 'lifted' to morphisms in a 'higher' category. The purpose of this lifting is also that the meanings of any two well-typed sentences all live in the same space, no matter their grammatical structure. Therefore, it is possible to compare the meanings of any two sentences by simply using the inner-product of that space, the same way meanings of words are compared in the distributional model. The mathematical structure obtained, as it is inspired from categorical quantum mechanics, uses the same diagrammatic calculus and therefore exposes how the information flows between the words in a sentence; the connexion between words and how their arrangement produces meaning becomes clear. The original categorical model developped by Coecke, Sadrzadeh and Clark uses real vector spaces but it is possible to have a Boolean value for a sentence by simply changing to the space of relations. Montague-style Booleand-valued semantics will be the one mostly used in the rest of this thesis.

2.1 Introduction

The two main theories of meaning which pre-existed [10] are a symbolic [11] and a distributional [12] ones. However these two theories are somewhat orthogonal as the advantages of one is what the other lacks and vice versa: the former is compositional but only qualitative while the latter is non-compositional but quantitative. For a discussion of these two models in Natural Language Processing see [13].

The original paper was inspired by [14] for using tensor spaces and pair vectors with their grammatical types. Pregroups have been used recently for natural language but the reason they are particularly interesting in our context is that they share a common structure with vector spaces and tensor products. Both can be viewed as categories and these categories are compact closed. Therefore a compact closed category combining both the meanings of words as vectors in vector spaces and their grammatical roles as types in Pregroups will be used to compute the meaning of sentences. And such a category can be found by taking the tensor product of vector spaces and a Pregroup, thus creating (meaning, type) pairs.

As we have seen in the previous chapter, categories and their graphical representation are very convenient for type-checking, the reduction in the product cateogory will verify the grammatical correctness and assign a vector in the vector space of meaning to each sentence. The theory obtained is therefore compositional and the meaning is built from the vectors of meaning of all the words.

If we restrict vectors to range over $\mathbb{B} = \{0, 1\}$, a Montague-style Boolean-valued semantics emerges. Sentences are therefore assigned the meaning either true or false rather than a more nuanced one on a real vector space.This simply correponds to using the category of relations rather than the category of vector spaces as described in [15].

2.2 Two 'camps' within computational linguistics

We briefly present the two domains of computational lingistics which provide the linguistic backgroun for this chapter, literature will be given if the reader is interested in more details.

2.2.1 Vector space models of meaning

The key idea behind the distributional model [12], is that words appear in a certain context, and that context is key in determining the meaning of a word. Formally, a context can have several definitions but the simplest one is just an n word window around a word. Two nouns that are always used as subjects of the same verbs, qualified by the same adjectives or objects of the same actions will, intuitively, be quite similar, for example 'cat' and 'dog'.

They both sleep, run or walk. They can both be small, big or furry. And they can also be bought, cleaned and stroked.

So the meaning of a word will be represented in a vector space that has a very high dimension. The orthogonal basis of this vector space will be context words. Once these words have been chosen, it only remains to count how many times each of these words appear in the same context as a word in a given text or group of texts. A big advantage of this representation is that it is easy to measure distance between words as they all live in the same vector space. Some computational models built along these lines have used up to tens of thousands of basis vectors and bodies of text of up to a billion words. The meanings obtained therefore can be very nuanced and cover a large range of usages. Some thesauri using this method have been built, in particular, in [16], the top 10 most similar nouns to 'introduction' are: 'launch', 'implementation', 'advent', 'addition', 'adoption', 'arrival', 'absence', 'inclusion' and 'creation'.

As explained in [10] There are a few advantages of vector-based representations over hand-built ontologies:

- "they are created objectively and automatically from text;
- they allow representation of gradations of meaning;
- they relate well to experimental evidence indicating that the human cognitive system is sensitive to distributional information (see [17] and [18])."

2.2.2 Algebra of Pregroups as a type-categorial logic

This paragraph will provide a very breif overview of Pregroups, for more details please refer to [19], [20], [21], [22].

Definition 2.2.1. [10] A partially ordered monoid $(P, \leqslant, \cdot, 1)$ is a partially ordered set, equipped with a monoid multiplication $(\_ \cdot \_)$ with unit 1; where for $p, q, r \in P$, if $p \leqslant q$ then we have $r \cdot p \leqslant r \cdot q$ and $p \cdot r \leqslant q \cdot r$.

Definition 2.2.2. [10] A Pregroup $(P, \leqslant, \cdot, 1, (-)^l, (-)^r)$ is a partially ordered monoid whose each element $p \in P$ has a left-adjoint p^l and a right adjoint p^r, i.e. the following hold:

$$p^l \cdot p \leqslant 1 \leqslant p.p^l \text{ and } p \cdot p^r \leqslant 1 \leqslant p^r \cdot p$$

Some properties of interest in a Pregroup are [10]:

- Adjoints are unique.

- Adjoints are order reversing: $p \leqslant q \Rightarrow q^r \leqslant p^r$ and $q^l \leqslant p^l$.
- The unit is self adjoint: $1^l = 1 = 1^r$.
- Multiplication is self adjoint: $(p \cdot q)^r = q^r \cdot p^r$ and $(p \cdot q)^l = q^l \cdot p^l$.
- Opposite adjoints annihilate each other: $(p^l)^r = p = (p^r)^l$.
- Same adjoints iterate: $p^{ll}p^l \leqslant 1 \leqslant p^r p^{rr}, p^{lll}p^{ll} \leqslant 1 \leqslant p^{rr}p^{rrr}, \ldots$

The way Pregroups formalise the grammar of natural languages is by first fixing a set of grammatical roles and a partial ordering between them and then generating a Pregroup with these types. (It has been proven that such a Pregroup exists). The only types of sentences that will be used in this thesis are affirmative and negative (i.e., which contain a negation) transitive sentences in English, and the coordination of such sentences.

Using the same notation as [10], we fix the following basic types:

n: noun	s: declarative statement
j: infinitive of the verb	σ: glueing type

From these basic types, it is possible to form compound types using adjoints and juxtaposition. A type (either basic or compound) is assigned to each word of the dictionary. Just like some verbs can both be transitive or intransitive, the type of a word may depend on its usage. The type s is the type to which a sentence has to reduce to be considered grammatical. The reduction process has been shown to be decidable. For more clarity about the order of the process, $\rightarrow$ will be used instea of $\leqslant$, and for simplicity the $\cdot$ between juxtaposed types will be dropped.

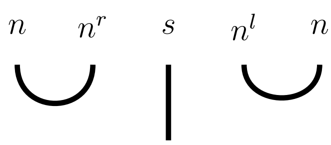
One basic example is given in [10]:

$$\begin{array}{ccc} \text{John} & \text{likes} & \text{Mary} \\ n & (n^r s n^l) & n \end{array}$$

It is grammatical because of the following reduction:

$$n(n^r s n^l)n \rightarrow 1 s n^l n \rightarrow 1 s 1 \rightarrow s$$

Graphically, this can be represented as:



The use of glueing types appears in the negation of the sentence and comes from the type suggested for "does" and "not" in [23]. [10] preferred this typing system as it allows for the information to flow and be acted upon in the sentence and as such assits in constructing the meaning of the whole sentence.

2.3 Modeling a language in a concrete category

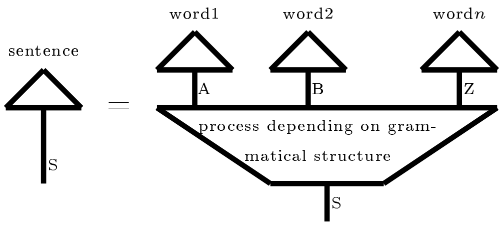
As already explained, we will use a category-theoretic model of language, which fits in the theory developed in the previous chapter. Using categories brings the following advantages:

- If we want to give quantitative meaning to a sentence that is more nuanced that simply true or false, it requires a mathematical structure that can store this additional information while keeping the compositional structure. Even if Montague semantics will be the approach mainly used in this thesis, the next step in the research is naturally to look in vectorial spaces, and the consequences in such a space will often be given.
- As stated earlier, we consider compact closed categories. Their structural morphisms will be the basic building blocks to construct the morphism that takes as an input the meaning of words and outputs the meaning of the sentence formed by the juxtaposition of these words.
- Lifting to categories makes it possible to reason on the grammatical structure of different sentences. We are able to obtain more information than just whether a sentence is grammatically correct or not, but how it is grammatically correct, reason about ambiguities in grammatical sentences and how these ambiguities gives rise to different meaning interpretations.

2.3.1 The 'from-meaning-of-words-to-meaning-of-a-sentence' process

Monoidal categories are a good tool to represent processes between system of varying types. Our process here is the one that takes the meaning of words as an input (the number of words can vary) and outputs the meaning of the sentence as an output, if the sentence is well-typed (that is, gramatically correct).

Diagrammatically it corresponds to the following process:



where the triangles represent meanings, both of words and sentences.

The juxtaposition (word 1)...(word n) is the sentence itself, which has type $A\otimes ... \otimes Z$. This becomes a sentence of type S through the grammatical structure. The concrete manner in which grammatical structure performs this role will be explained below.

2.3.2 Compact closed categories

Definition 2.3.1. [10] A monoidal category is compact closed if for each object A there are also objects A^r and A^l, and morphisms

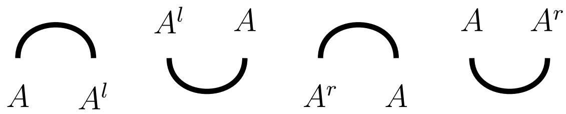
$$\eta^l : I \to A \otimes A^l, \epsilon^l : A^l \otimes A \to I, \eta^r : I \to A^r \otimes A, \epsilon^r : A \otimes A^r \to I$$

which satisfy:

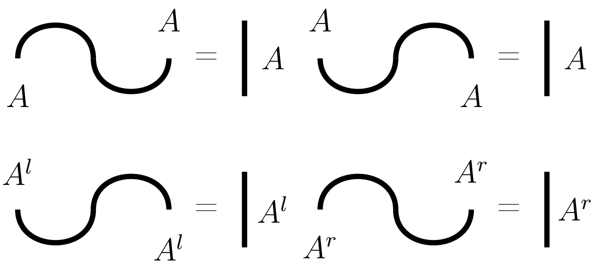
$$\begin{array}{cc} (1_A \otimes \epsilon^l) \circ (\eta^l \otimes 1_A) = 1_A & (\epsilon^r \otimes 1_A) \circ (1_A \otimes \eta^r) = 1_A \\ (\epsilon^l \otimes 1_{A^l}) \circ (1_{A^l} \otimes \eta^l) = 1_{A^l} & (1_{A^r} \otimes \epsilon^r) \circ (\eta^r \otimes 1_{A^r}) = 1_{A^r} \end{array}$$

A compact closed category is somewhat orthogonal to a Cartesian category in the sense that a large triangle $A \otimes B$ cannot be decomposed in two smaller triangle A and B (which is the case in a cartesian category). On the contrary, this connectedness encodes meaning as it allows interaction between the words in the sentence.

Graphically, the morphisms $\eta^l, \epsilon^l, \eta^r, \epsilon^r$ can be depicted as (read in top-down fashion):



The axioms of compact closure simplify to "yanking" equations:



FVECT is a compact-closed category with $V^l = V^r = V^* = V$. Let $\{\vec{e_i}\}_i$ be a base of V, then:

$$\eta^l = \eta^r : \mathbb{R} \to V \otimes V :: 1 \mapsto \sum_i \vec{e_i} \otimes \vec{e_i}$$

and

$$\epsilon^l = \epsilon^r : V \otimes V \to \mathbb{R} :: \sum_{ij} c_i j \vec{v_i} \otimes \vec{w_j} \mapsto \sum_{ij} c_{ij} \langle \vec{v_i} | \vec{w_j} \rangle.$$

A Pregroup is a a posetal category. We take $[p \leqslant q]$ to be the singleton $\{p \leqslant q\}$ whenever $p \leqslant q$ and empty otherwise. Any equational statement between morphisms in posetal categories is trivially satisfied, since there is at most one morphism between two objects. A Pregroup is a compact closed category for:

$$\begin{array}{ll} \eta^l = [1 \leqslant p \cdot p^l] & \epsilon^l = [p^l \cdot p \leqslant 1] \\ \eta^r = [1 \leqslant p^r \cdot p] & \epsilon^r = [p \cdot p^r \leqslant 1] \end{array}$$

For complete explanation on how FVECT and Pregroups as compact closed category, the reader can refer to [10].

2.3.3 Categories representing both grammar and meaning

We have seen earlier in this chapter that vector spaces can be used to assign meaning to words in a language, and Pregroups can be used to assign grammatical structure to sentences. Both can be consiered as compact closed categories. A mathematical structure that unifies both of these aspects of language is the product FVECT $\times$ P which naturally has projectors into FVECT and P. For more details on this construction see [10].

2.3.4 Meaning of a sentence as a morphism in FVECT$\times$ P

Definition 2.3.2. [10] We refer to an object (W, p) of FVECT $\times$ P as a meaning space. This consists of a vector space W in which the meaning of a word $\vec{w} \in W$ lives and the grammatical type p of the word.

Definition 2.3.3. [10] We define the vector $\overrightarrow{w_1 ... w_n}$ of the meaning of a string of words $w_1 ... w_n$ to be

$$\overrightarrow{w_1 ... w_n} := f(\vec{w_1} \otimes ... \otimes \vec{w_n})$$

where for (W_i, p_i) meaning space of the word w_i, te linear map f is built by substituting each p_i in $[p_1 ... p_n \leqslant x]$ with W_i.

For a wide range of example on this model, please refer to [10].

2.4 Relations vs Vectors for Montague-style semantics

So far, the category in which the meaning of words existed was FVECT and from this we built matrices with real numbers as entries. If, instead of taking matrices on $(\mathbb{R}, +, \times)$ we take matrices on the semiring $(\mathcal{R}, +, \times)$ then we also obtain a compact closae category.

In particular, it is possible to use $(\mathbb{B}, \vee, \wedge)$ in which case we obtain a category isomorphic to FREL of finite sets and relations. This category works as follow: let X be a set $\{x_i | 1 \leqslant i \leqslant |X|\}$ and Y be a set $\{y_j | 1 \leqslant j \leqslant |Y|\}$, the relation $r \subseteq X \times Y$ can be represented by an $|X| \times |Y|$ matrix where there is a 1 on the i^{th} line and j^{th} column iff $(x_i, y_j) \in r$ and a 0 otherwise. It is possible to compose two relations r and s if they share a space, that is $r \subseteq X \times Y$ and $s \subseteq Y \times Z$ and the relation $s \circ r$ is

$$\{(x, z) | \exists y \in Y : (x, y) \in r, (y, z) \in s\}.$$

The composition induces matrix multiplication.

In no way in the previous construction was the choice of FVECT important, so it can simply be replaced by FREL and the previous construction still works and we can use the category FREL $\times P$ to represent both meaning and grammar.

2.5 Summary

This chapter shows a linguistic model that combines the advantages of being both distributional and compositional. However, more work needs to be done. In particular representing some logical words such as 'not', 'and', 'or', 'if then'. The first model that comes to mind for 'or' and 'and', which is vector sum and product, does not correspond well with the logical functions as they are not fully distributive. The rest of this thesis will focus more specifically on 'and' and 'or'.

Chapter 3

Perfecting the representation of 'AND'

A main open problem with categorical compositional distributional semantics is the representation of words that do not carry meaning of their own from a distributional perspective, such as determiners, prepositions, relative pronouns, coordinators, etc. This thesis deals with the topic of coordination and more specifically in this chapter, on the representation of 'AND'.

Dimitri Kartsaklis has set the first stone on the work on coordination and more specificly on 'AND' [1]. He has found that 'AND' can be represented by GHZ-spiders. While this representation works well in many cases, it also has a few caveats. In a first part we will briefly present his work before introducing a couple possible solutions for perfecting this representation.

3.1 Coordination in Categorical Compositional Distributional Semantis

In the past, researchers have found that Frobenius algebras are useful in language and can be used to represent linguistic aspects that compact closed categories do not manage to grasp ([2] [3] [4] [5] [6]).

Frobenius multiplication enforces the two inputs to contribute equally to the result, which is exactly what coordination is all about. And the Frobenius co-multiplication allows to duplicate information, which can be necessary for some compound grammatical types.

3.1.1 Frobenius algebra over FVECT

Kartsaklis' paper introduces a Frobenius algebra over FVECT without going through GHZ or W states. The algebra $(\text{FVECT}, \mu, \delta, \eta, \epsilon)$ with a basis $\{\vec{v_i}\}$ is as follows:

$$\mu :: \vec{v_i} \otimes \vec{v_j} \mapsto \delta_{ij}\vec{v_i} := \begin{cases} \vec{v_i} & i = j \\ \vec{0} & i \neq j \end{cases} \qquad \eta :: 1 \mapsto \sum_i \vec{v_i}$$
$$\delta :: \vec{v_i} \mapsto \vec{v_i} \otimes \vec{v_i} \qquad \epsilon :: \vec{v_i} \mapsto 1$$

But this Frobenius algebra corresponds to the GHZ algebra as it is a special Frobenius algebra.

3.1.2 Linguistic uses of the Frobenius operators

These Frobenius operations offer transformations that did not previously exist in compact closed categories. For example the μ composition corresponds to element-wise multiplication and avoids the transformational effect of regular composition in a compact closed category (for example, an intransitive verb is a map that transforms a noun into a sentence). This interaction is used for example by Sadrzadeh et al [2] [3] on nouns modified by relative clauses (e.g. "The man who likes Mary"). The construction presented for the relative pronoun uses μ-composition of the noun (here 'man') with the verb phrase (here 'likes Mary').

This effect is even clearer when using FREL rather than FVECT where it acts as an intersection of the involved sets or relations.

The Frobenius co-multiplication has also been used in previous papers for linguistic, for example in [4].

3.1.3 Coordination in Categorical Compositional Distributional Semantics

Merging and copying information seems to be two useful tools when trying to model coordination. Distributivity is also something that is to be expected.

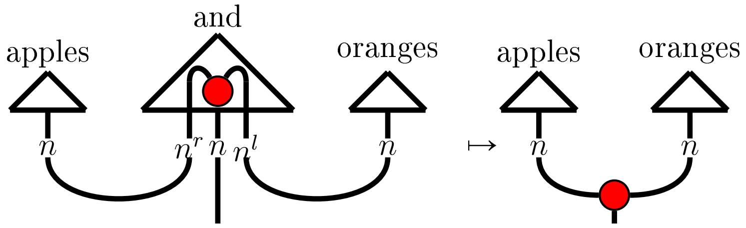
In the rest of this thesis, we will consider that conjunction happens between two conjuncts of the same type and produces a result that is again of that specific type. This correspons to X CONJ $X \to X$. Therefore 'and' is assigned the pregroup type $x^r \cdot x \cdot x^l$, where x can either be an atomic or compound type.

For the detailed construction of how μ-composition can be used instead of ε-composition, the reader is advised to refer to [10]. Here we will simply admit that the coordinator morphism can be expressed as follows:

$$\overline{conj_X} : I \xrightarrow{\eta^r_X \otimes \eta^l_X} X^r \otimes X \otimes X \otimes X^l \xrightarrow{1_{x^r} \otimes \mu_X \otimes 1_{X^l}} X^r \otimes X \otimes X^l \tag{3.1}$$

Coordinationg atomic types

This is a priori the simplest case. We will focus on conjuncts that are either vectors (such as nouns) or that can be reduced to a vector (such as noun phrases). Using equation 3.1, coordination over noun phrases looks like:



(3.2)

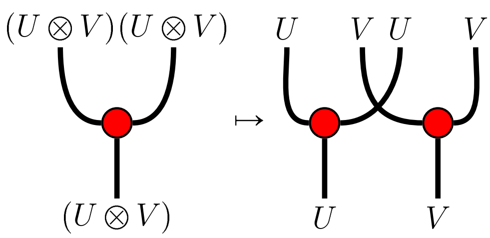
The case of coordinating sentences is very similar and an example can be found in [10].

Coordinationg compound types

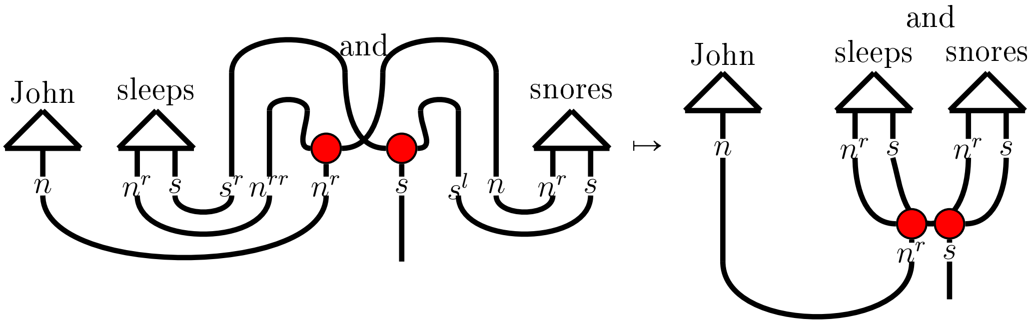
The previous paragraph uses the merging operator of the Frobenius algebra but not the copying one, the latter will be necessary over compound types.

We recall the following operation on a compound object $U \otimes V$:

Luckily, Frobenius operators also coherently lift to compound object along the following graphical representation:



We can therefore apply equation 3.1 to compound structure, for example with the sentence below:



(3.3)

We can observe the following actions:

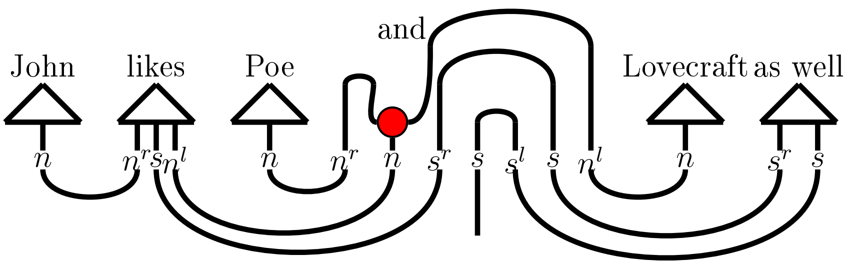
- The subject of both verbs 'John' is copied at the n^r input of the conjunction;
- the coordinator interacts on the left with the verb 'sleeps' and on the right with the verb 'snores';
- the merging operator of the frobenius algebra turns the s carried by each verb into the s that forms the final sentence, thus returning a single vector of a well-typed sentence.

This generalises to compounds types of order higher than 2.

Non-standard forms of coordination

[10] considers the following sentence : "John likes Poe, and Lovecraft as well", which expands into "John likes Poe, and John likes Lovecraft as well". The meaning of this sentence should be the same as "John likes Poe and Lovecraft".

Additional wiring is necessary to represent this ellipsis, known as stripping. The following diagram shows how the result is achieved:



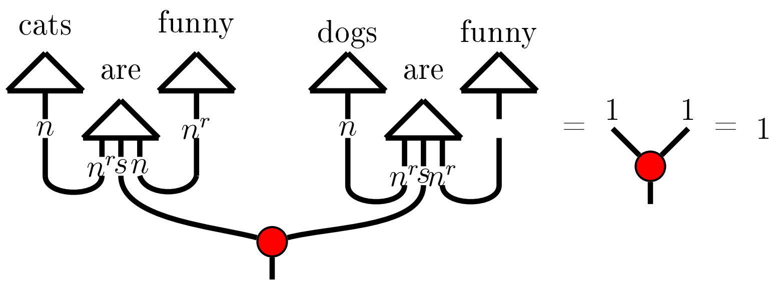
(3.4)

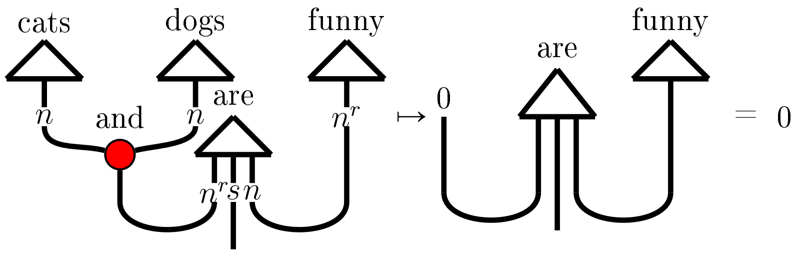
3.2 Some caveats in Kartsaklis' model

During my research, I have come across a few issues with the representation of 'AND' with GHZ-spiders.

An issue appears, for example, in a Montague-style toy model where 'cats' would be represented by $|0\rangle$, 'dogs' by $|1\rangle$ and where the adjective funny qualified both these animals - therefore funny would be represented by $|0\rangle + |1\rangle$.

Kartsaklis's representation of 'AND' works in the sentence 'cats are funny and dogs are funny' but no longer works in the sentence 'cats and dogs are funny'. Indeed, the first sentence simply corresponds to applying a logical 'AND' to two booleans evaluated as 'true' (each half of the sentence on each side of the end already has the type S of the sentence); while the second one performs the 'AND' on two vectors that represent two nouns, and these vectors being orthogonal, the result is simply the boolean 'false'.





(3.5)

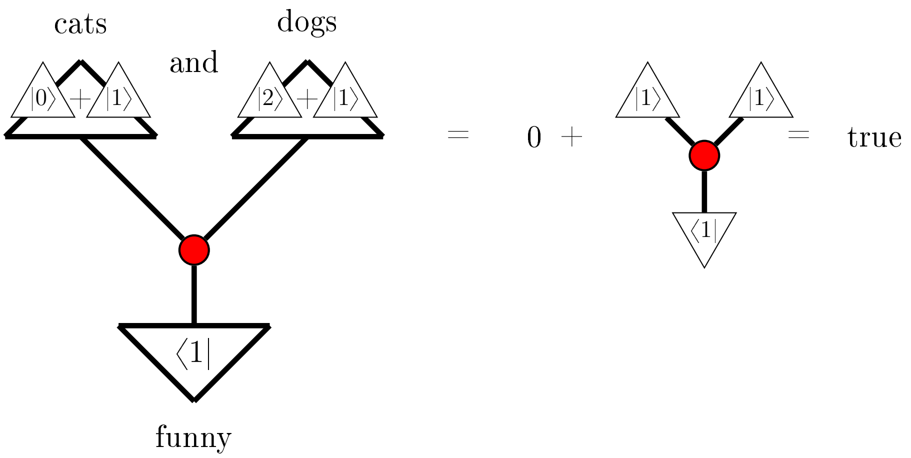
However, a first reader would find that both these sentences are typically considered to have the same meaning in English.

In both cases, we are in the first situation described in Kartsaklis' paper where the merging operator of the Frobenius algebra is used. Merging two sentences that are evaluated as "true" easily corresponds to true, however merging two nouns that are orthogonal results in the empty set in the category Rel and therefore the sentence is evaluated as false.

There are three possible responses to this problem.

3.2.1 First fix by non-orthogonality

The first idea is that these two vectors should not be orthogonal, indeed 'cats' and 'dogs' are pretty similar - they are both 'funny' - and a simple way to cope with this problem is to represent 'cats' by $|cat\backslash dog\rangle + |cat \cap dog\rangle$, similarly 'dogs' by $|dog\backslash cat\rangle + |dog \cap cat\rangle$, and 'funny' would then be $|cat \cap dog\rangle$. This idea works and the sentence 'cats and dogs are funny' with this representation holds the truth value 'true' as expected. In the following diagram, for readability, $|0\rangle = |cat\backslash dog\rangle$, $|1\rangle = |cat \cap dog\rangle$ and $|2\rangle = |dog\backslash cat\rangle$.



(3.6)

However, this does not fix the fact that two sentences that should intuitively have the same meaning - 'cats are funny and dogs are funny' and 'cats and dogs are funny' - don't behave the same way in a toy model.

This is not necessarily a problem as the use of 'AND' in these two sentences is somewhat different: while the first one is just the logical AND between two propositions that are true, the second one has a behaviour closer to 'OR' as it has the same meaning as 'take either a cat or a dog, it is funny'. Contrary to the case presented in Katarski's paper, the Frobenius operator that we expect here should be a union and not an intersection. The 'AND' in the second sentence does not work, and possibly should not work, as there are no animals that are both cats and dogs. By representing 'cats' and 'dogs' orthogonally, what this sentence is saying is: 'take an animal that is both a cat and a dog, it is funny', or 'take an animal that is in the intersection of the set of cats and the set of dogs, it is funny'. This sentence is not true, as this would be a Frankenstein murder operation, which is far from being funny.

What this example says is that there are cases where 'AND' in English does not behave like an intersection, and therefore GHZ spiders may not be the best representation.

This issue becomes more complicated when coordinating adjectives. If we consider the toy model where 'loyal' is represented by $|0\rangle$, 'furry' by $|1\rangle$ and 'dogs' by $|0\rangle + |1\rangle$ to mean that dogs are both loyal and furry. Then the sentence 'dogs are loyal and dogs are furry' evaluates as true if we use a GHZ spider for 'and', but the sentence 'dogs are loyal and furry' evaluates as false. It is exactly the same issue as the problem before except this time the coordination happens on the object and on a compound type n^r. This is the exact same issue as above and we could argue that 'loyal' and 'furry' should not be orthogonal because they both qualify dogs.

However, in a real model in FVECT that uses real numbers rather than booleans, it is possible to imagine two adjectives that would never be used together except for one very odd object - for example 'metallic' and 'smooth' which can never qualify the same object except for an 'alien spacecraft'. That is, with a vector basis of say a thousand vectors, the element-wise multiplication would be 0 except for one coefficient:

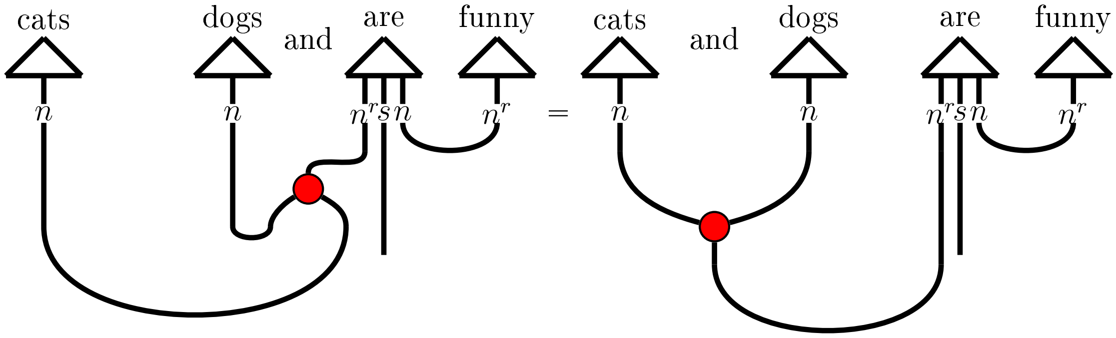
$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 \cdot b_1 \\ a_2 \cdot b_2 \\ \vdots \\ a_n \cdot b_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a_n \cdot b_n \end{pmatrix}$$

Therefore these two adjectives would be almost orthogonal and as a result the sentence 'alien spacecrafts are metallic and smooth' would have a truth value very close to 'false'(as most of the coefficients are 0, the meaning of the sentence vector would almost be entirely 0 as well) despite being entirely true. In a more complex Boolean model, this would however work as the intersection of 'metallic' and 'smooth' woul be exactly 'alien spacecraft' (as it is the only word qualified by both in our model), and the sentence would then still be true.

3.2.2 Second fix by distributivity

The second idea is that both the sentence 'dogs and cats are funny' and 'dogs are funny and cats are funny' should have the same meaning and behave in the same way no matter the model. Therefore, what we want to do is to use the copying operator of the Frobenius algebra, copy 'are funny' and test each of the two copies with 'dogs' and 'cats'.

The problem is that the copying operator of the Frobenius algebra only perfectly copies elements of the basis. If theoretically we are considering a different approach, mathematically this still corresponds to the same operation:

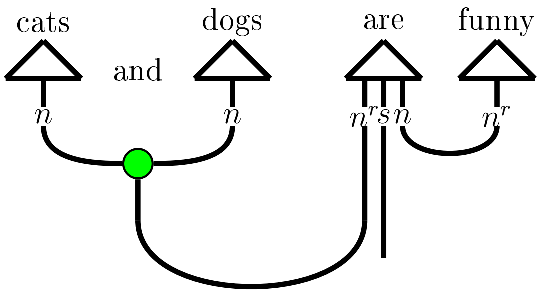


(3.7)

The possible solutions are therefore: limit the use to a base vector for the corresponding adjective 'funny', or expand the sentence into 'cats are funny and dogs are funny' before analysing it, though I have not explored possibilities of doing it computationally.

3.2.3 Third fix by considering union and intersection

As we have seen before, the main issue so far is that this 'AND' seems to correspond to a union more than an intersection. And as we will see in the next chapter, a good representation for 'OR', and for the union is the W-spider, or the multiplication in a anti-special Frobenius algebra. In this case, using a W-spider to represent 'AND' does give the result 'true' without modifying the orthogonality of 'cats' and 'dogs' or the grammatical structure of the sentence.



(3.8)

When we compute the result considering the verb 'to be' as the identity, we obtain:

$$\langle 010 + 011 | 001 + 010 + 100 \rangle = 1.$$

The solution could therefore be to see whether the conjunction, no matter what the English word used, correponds to a union or an intersection and use the appropriate spider in each case. Again, we have not explored the possibility for a computer to solve this problem.

3.3 Summary

Frobenius algebras are useful to represent conjunctions and in particular GHZ spiders can be used to represent intersectionality, which in English usually translates into 'AND'. However, not all use cases of 'AND' correspond to an intersection and there are, therefore, a few problems with using the GHZ spider to blindly represent 'AND'. Three approaches to fix this issue have been offered in this chapter.

Chapter 4

Representing 'OR' using W-spiders

In this chapter we will focus on the representation of disjunction in coordination, which is usually represented by 'OR' in English, using W-spiders. In a first part we will look at generalising W-spiers for qudit using the work of Amar Hadzihasanovic [24] so that it is possible to work with toy models that have a vector basis bigger than 2. We will then show that W-spiders, with fixes inspired from the previous section, can be used to represent 'OR'.

4.1 W-spiders in dimensions higher than 2

Hadzihasanovic argues in his DPhil thesis that the W gate for qubits can be viewed as a fermion oscillator. When generalising to a dimension $d \in \mathbb{N}$, the fermion becomes an abelian anyon [25], that is a particle that, when exchanged with another particle, receives a phase $q := e^{\frac{2i\pi}{d}}$. The corresponding unitary operator on $\mathbb{C}^d \otimes \mathbb{C}^d$ is:

$$x : |k\rangle \otimes |j\rangle \mapsto q^{jk} |j\rangle \otimes |k\rangle, \qquad j, k = 0, ..., d-1$$

with inverse:

$$x^{\dagger} : |k\rangle \otimes |j\rangle \mapsto q^{-jk} |j\rangle \otimes |k\rangle, \qquad j, k = 0, ..., d-1$$

In the case where $d = 2$, this corresponds to the crossing in ZW calculus.

Hadzihasanovic chooses to keep the canonical self-duality maps:

$$\eta := \sum_{j=0}^{d-1} |j\rangle \otimes |j\rangle, \qquad \epsilon := \sum_{j=0}^{d-1} \langle j| \otimes \langle j|.$$

Some notions of q-arithmetic are needed to define the algebra obtained.

Definition 4.1.1. [24] For all $q \in \mathbb{C}$, $n \in \mathbb{N}$, the q-integer $[n]_q$ is defined by

$$[n]_q := \sum_{k=0}^{n-1} q^k.$$

The q-factorial of n is then defined by:

$$[n]_q! := \prod_{k=1}^{n} [k]_q,$$

and, for $k = 0, ..., n$ the q-binomial coefficient is

$$\binom{k}{n}_q := \frac{[n]_q!}{[k]_q![n-k]_q!}.$$

When $q = 1$ we obtain the regular integer, factorial and binomial coefficient. When $q = e^{\frac{2i\pi}{d}}$ then we have $[d]_q = 0$, hence $[n]_q! = 0$ for all $n \geqslant d$.

In particular, the q-binomial coefficients satisfy the q-Vandermonde identities:

$$\binom{k}{n}_q = \sum_{i=0}^{k} q^{(j-i)(k-i)} \binom{i}{j}_q \binom{n-j}{k-i}_q.$$

Fixing again $q = e^{\frac{2i\pi}{d}}$, Hadzihasanovic defines the W map as:

$$w : |n\rangle \mapsto \sum_{k=0}^{n} \binom{k}{n}_q^{\frac{1}{2}} |k\rangle \otimes |n-k\rangle, \qquad n = 0, ..., d-1$$

With the discard map : $v : |0\rangle \mapsto 1, |n\rangle \mapsto 0$ for $n = 1, ..., d-1$.

Theorem 4.1.2. [24] The comonoid (w, v) forms a bialgebra with its transpose monoid, and x as a braiding.

The proof can also be found in [24]. Hadzihasanovic even proves universality for Qudits of his extended form of ZW-calculus.

As we will be working in *Rel* rather than real vectors in the rest of this chapter, we will use the following map for W:

$$w : |n\rangle \mapsto \sum_{k=0}^{n} |k\rangle \otimes |n-k\rangle, \qquad n = 0, ..., d-1$$

.

4.2 W-spiders can represent 'OR' in any dimension

In this section we will use the representation of W-spiders presented in the previous section to check on a few examples whether W-spiders would be a good match for representing 'OR'. Intuitively, as GHZ-spiders can represent 'AND'; because ZW-calculus contains rational arithmetic with the GHZ map acting as multiplication and the W map acting as addition [26]; and because of the distributivity between the GHZ and W maps, it seems natural to think that W spiders would be a good match. As seen in the previous chapter, when 'AND' corresponds to a union, the W map actually gives a more coherent result than the GHZ map when computing the truth value of the sentence.

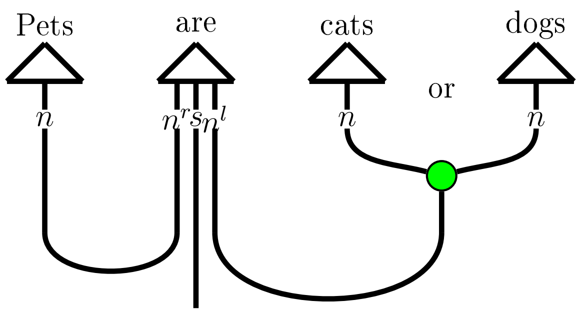
4.2.1 Dimension 2

The dimension 2 is the simplest one to work with as the W map's definition on Qubits is commonly known and accepted. Coordination works well but on some example, it seems to be more coincidental than inherently linked to the structure.

Coordinating atomic types

The model we use here is: cats are $|0\rangle$, dogs are $|1\rangle$ and pets are either cats or dogs and therefore $|0\rangle + |1\rangle$.

The sentence 'pets are cats or dogs', considering the verb 'to be' as the indentity and 'OR' as the W operator, evaluates as 'true', which is what we expected. When doing the maths though, it seems however that this result is quite coincidental and would not hold true in higher dimension.

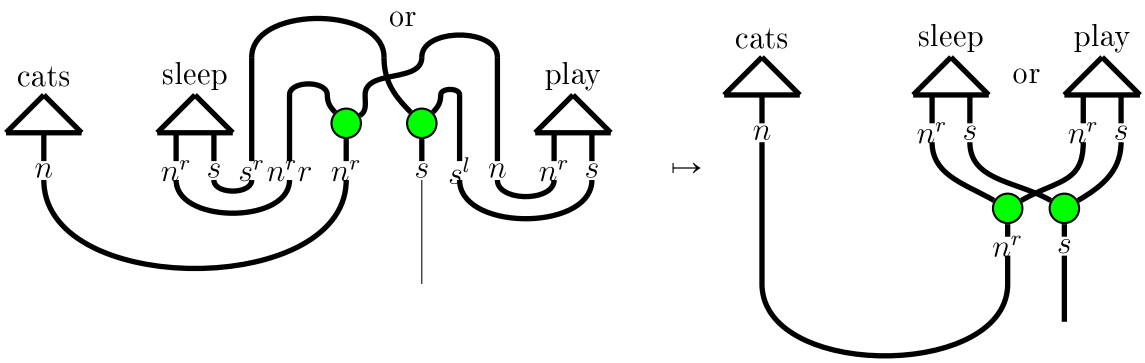


Coordinating compound types

As W states are also Frobenius state, the proof presented in chapter 2 that the Frobenius operators coherently lift to compound objects also holds for the W map. So for an intransitive verb, we also have:

$$\begin{aligned}\overline{disj}_{n^r\otimes s} &= (1_{(n^r\otimes)^r} \otimes w_{n^r\otimes s} \otimes 1_{(n^r\otimes s)^l}) \circ (w^r_{n^r\otimes s} \otimes w^l_{n^r\otimes s}) \\ &= (1_{s^r} \otimes 1_{n^{rr}} \otimes w_{n^r} \otimes w_s \otimes 1_{s^l} \otimes 1_n) \circ (1_{s^r} \otimes w^r_{n^r} \otimes x_{n^r,s} \otimes w^l_s \otimes 1_n) \circ (w^r_s \otimes w^r_n)\end{aligned}$$

We still consider our basis $|cat\rangle = |0\rangle$ and $|dog\rangle = |1\rangle$. Let's say that 'sleep' is an activity that only cats do while 'play' is an activity that both cats and dogs do. Therefore sleep is going to be represented by the vector $\begin{pmatrix}1\\0\end{pmatrix}$ and play is going to be represented by the vector $\begin{pmatrix}1\\1\end{pmatrix}$. The sentence 'cats sleep or play' holds true when computing the maths.



The sentence 'dogs sleep or play' holds true as well. Logically it is true as they do do either one of these activities but this is not necessarily the meaning that an English speaker would want using natural language.

We have not looked at non-standard forms of coordination.

4.2.2 Higher dimension and non-orthogonality fix

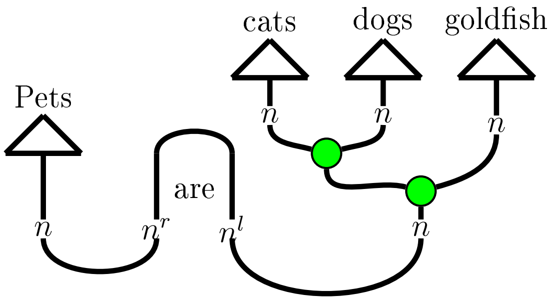
In dimension two, we were quite limited in the complexity of the sentences we could write. By using Hadzihasanovic's adapted representation of the W operator, we can test our intuition on more complex sentences.

Coordinating atomic compounds

Let's slightly extend our first example to dimension 3 with cats being represented by $|0\rangle$, dogs by $|1\rangle$ and goldfish by $|2\rangle$. Pets are this time going to be $|0\rangle + |1\rangle + |2\rangle$.

This time the sentence 'pets are cats or dogs or goldfish' evaluates as false (the order in which we evaluate the 'or' doesn't matter thanks to associativity). So our feeling that it was only coincidental in dimension 2 is verified.

We will try to use the same non-orthogonality technique of chapter 3 where pet is considered to be $|cats \cap dogs \cap goldfish\rangle = |0\rangle$, cats are going to be represented by $|cats\backslash(dogs \cup goldfish)\rangle + |cats \cap dogs \cap goldfish\rangle = |1\rangle + |0\rangle$, dogs by $|dogs\backslash(cats \cup goldfish)\rangle + |cats \cap dog \cap goldfish\rangle = |2\rangle + |0\rangle$ and goldfish by $|goldfish\backslash(cats \cup dogs)\rangle + |cats \cap dogs \cap goldfish\rangle = |3\rangle + |0\rangle$. The four kinds of vectors are going to be considered the basis vectors.



With this technique the vector $|0\rangle$ for pet is copied onto goldfish, cats and dogs and then tested against the vectors for cats, dogs and goldfish. This time the truth value of the sentence is true. We notice however that if we replaced the disjunction with conjunction (that is using a GHZ map instead of a W map), the sentence would also be true.

This result can easily be generalised to a coordination of n elements. If we keep on appointing $|0\rangle$ for 'pets', the maths are fairly straightforward as $w(|0\rangle) = |0\rangle \otimes |0\rangle$ (any other order of the basis vector would lead to the same result, the maths would simply be a little more complicated to do). As long as $|0\rangle$ is part of the description of each animal, the sentence will hold as true.

However, if one of the animals is not a pet (for example 'elephants' being $|4\rangle$) then the sentence can evaluate as either true or false according to the numbering of the vector basis. For example, with this numbering, the sentence 'pets are cats or elephants' evaluates as false, however, if pets is $|1\rangle$, cats is $|2\rangle + |1\rangle$ and elephants is $|0\rangle$ (with dogs being $|3\rangle + |1\rangle$ and goldfish being $|4\rangle + |1\rangle$) then the sentence is true. This is obviously a problematic ambiguity. One possible fix is to take the sum of all possible results with all the possible numbering of the basis vectors, the operation would have a very costly complexity but would give a deterministic result. In our case, that would be 'true', thus, like in the previous subsection on dimension 2, having an evaluation for 'OR' that is rather closer to the logical 'OR' than to that commonly used in natural language.

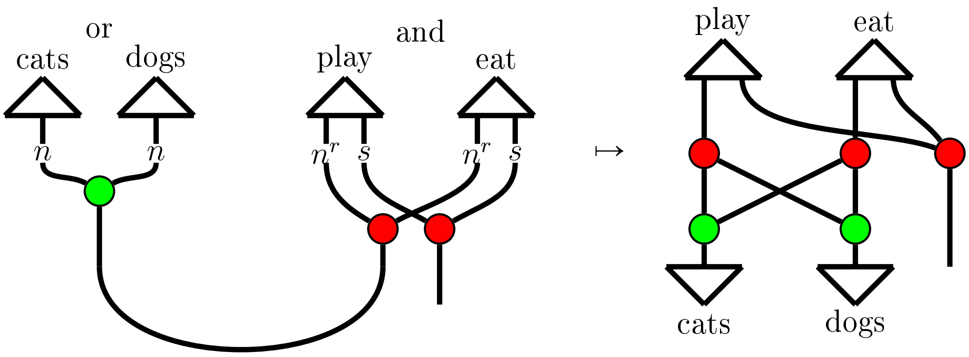
Coordinating compound types

We are again going to test this case with a disjunction of intransitive verbs. We still have our cats $|0\rangle$, dogs $|1\rangle$ and goldfish $|2\rangle$. Cats sleep, play and eat, dogs play and eat and goldfish eat and swim. So sleep is represented by $\begin{pmatrix}1\\0\\0\end{pmatrix}$, play by $\begin{pmatrix}1\\1\\0\end{pmatrix}$, eat by $\begin{pmatrix}1\\1\\1\end{pmatrix}$ and swim by $\begin{pmatrix}0\\0\\1\end{pmatrix}$. As expected the three sentences 'cats sleep or play or eat', 'dogs play or eat' and 'goldfish eat or swim' hold true, and these hold true no matter the ordering of the vector basis. These would also hold true if a conjunction instead of a disjunction was used.

However, if we introduce the verb 'bark', represented by $\begin{pmatrix}0\\1\\0\end{pmatrix}$ because only dogs bark, then 'cats or dogs bark' holds as true but 'cats or dogs sleep' holds as false. Again, this is because the truth value of these sentences depends on the numbering of the vector basis. The same fix as in the previous paragraph - taking the sum of all possible meanings - also works and gives the same result - both these sentences would then be true.

4.2.3 Distributivity of the conjunction and the disjunction

One major advantage of using the pair GHZ and W spiders to represent conjunction and disjunction is that there is a good distributivity of one over the other. Let's take for example the sentence 'cats or dogs play and eat'. Then according to [27] the following is true:

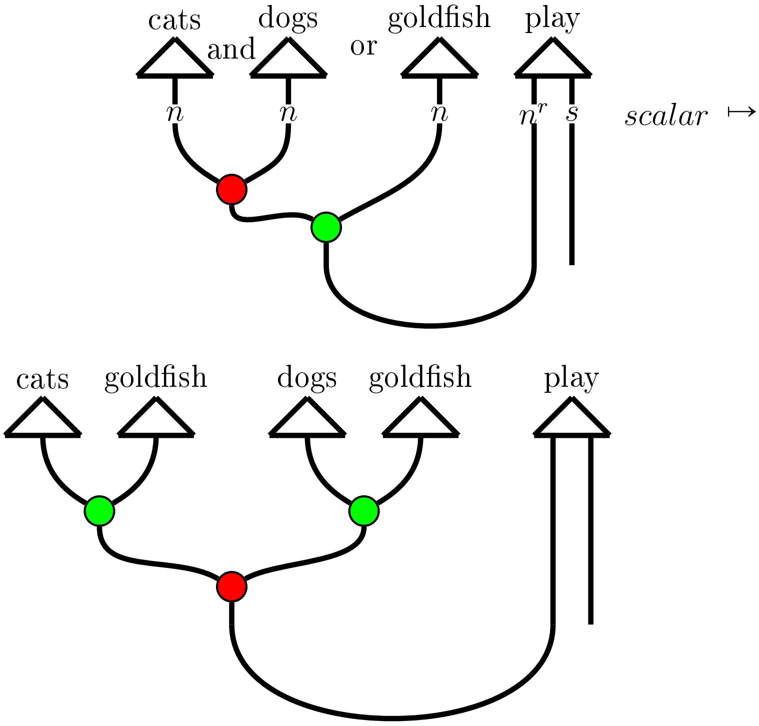


(4.1)

As we can see, this transformation makes sure that both verbs are distributed onto both subjects and vice versa.

Another example of distributivity can be found in the sentence 'cats and dogs or goldfish play' which could be bracketed in two ways, either as '(cats and dogs) or goldfish play' or 'cats and (dogs or goldfish) play'. We will stick to the former but both could be studied similarly.

In this case distributivity works as follows according to [26]:



(4.2)

A scalar appears but would only be significant in a study with vectorial spaces on real numbers and not in Rel. Other than that, the distributivity is exactly what we would expect should happen in this situation.

4.3 Summary

In summary, W spiders seem to be a good lead to represent disjunction but there still are a few issues that need to be explored. In particular, it is necessary to explore more in details the difference in usage of 'and' and 'or' in the natural language as they are more subtle than what pure logic tells us.

Conclusion and Future Work

As we have seen in this thesis, Frobenius algebras offer a promising solution for representing conjunction and disjunction in the Categorical Compositional Distributional Model of Meaning. However, representing the natural language usage of 'OR' and 'AND' can be quite ambiguous and more work is needed on how to identify the subtle differences between these usages. Indeed, in some cases it seems that W Frobenius algebras would be better suited to represent 'AND' and vice versa.

More work is also needed on the representation of 'OR' and the generalisation of W spiders to Qudits as changing the order of the basis vector can sometimes radically change the meaning of the sentence, which is very unfortunate. A possible solution would be to take the sum of all the meanings of the sentence by permutating the numbering of the basis vector, which would solve this lack of determinism.

But the properties between GHZ and W spiders make the distributivity of 'OR' and 'AND' with one another correspond very well to what would actually be expected. The conclusion of this work is therefore quite nuanced with clear advances but also a few problems.

Bibliography

[1] Dimitri Kartsaklis. Coordination in categorical compositional distributional semantics. *Electronic Proceedings in Theoretical Computer Science*, August 2016.

[2] M. Sadrzadeh S. Clark B. Coecke. The frobenius anatomy of word meanings i: subject and object relative pronouns. *Journal of Logic and Computation*, 2013.

[3] M. Sadrzadeh S. Clark B. Coecke. The frobenius anatomy of word meanings ii: Possessive relative pronouns. *Journal of Logic and Computation*, 2014.

[4] Dimitri Kartsaklis Mehmoosh Sadrzadeh Stephen Pulman. A unified sentence space for categorical distributional-compositional semantics: Theory and experiments. *Proceedings of the 24th International Conference on Computational Linguistics*, 2012.

[5] Dimitri Kartsaklis and Mehrnoosh Sadrzadeh. A frobenius model of information structure in categorical compositional distributional semantics. *Proceedings of the 14th Meeting on the Mathematics of Language*, 2015.

[6] Bob Coecke and Martha Lewis. *Quantum Interaction: 9th International Conference, QI 2015, Filzbach, Switzerland, July 15-17, 2015, Revised Selected Papers*, chapter A Compositional Explanation of the 'Pet Fish' Phenomenon, pages 345 – 384. Springer International Publishing, 2016.

[7] Bob Coecke and Aleks Kissinger. The compositional structure of multipartite quantum entanglement. *ICALP 2010: Automata, Languages and Programming*, pages 297–308, July 2010.

[8] J. Kock. *Frobenius Algebras and 2D Topological Quantum Field Theories.* Cambridge University Press, 2003.

[9] G. Vidal W. Dür and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev.*, November 2000.

[10] Bob Coecke Mehrnoosh Sadrzadeh Stephen Clark. Mathematical foundations for a compositional distributional model of meaning. *Lambek Festschirft*, 2010.

[11] D.R. Dowty R.E. Wall S. Peters. *Introduction to Montague Semantics.* D. Reidel Publishing Company, 1981.

[12] H. Schuetze. Automatic word sense discrimination. *Computational Linguistics*, 1998.

[13] G.Gazdar. *Paradigm merger in natural language processing*, volume Computig Tomorrow: Future Research Directions in Computer Science, pages 88–109. Cambridge University Press, 1996.

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

[15] B. Coecke. Introducing categories to the practicing physicist. *What is category theory? Advanced Studies in Mathematics and Logic*, voloume 30:45–74, 2006.

[16] James R. Curran. *From Distributional to Semantic Similarity.* Univerisyt of Edinburhg, 2004.

[17] J.R. Saffran E.L Newport R.N. Asling. Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 1999.

[18] D.P. Spence and K.C. Owens. Lexial co-occurence and association strength. *Journal of Psycholinguistic Research*, 1990.

[19] J. Lambek. Type grammar revisited. *Logical Aspects of Computational Linguistics*, 1999.

[20] J. Lambek. Iterated galois connections in arithmetics and linguistics. *Galois Connections and Applications, Mathematics and its Applications*, 2004.

[21] J. Lambek. From word to sentence. *Polimetrica*, 2008.

[22] W. Buszkowski. Lambek grammars based on pregroups. *Logical Aspects of Computational Linguistics*, 2001.

[23] A. Preller. Towards discours representation via pregroup grammars. *JoLLI*, 2007.

[24] Amar Hadzihasanovic. *The algebra of entanglement and the geometry of composition*, chapter 5 part 3. Department of Computer Science, Universit of Oxford, 2017.

[25] P. Panangaden and É. Paquette. *New Structures for Physics*, chapter A categorical presentation of quantum computation with anyons, pages 983–1025. Springer Berlin Heidelberg, 2010.

[26] Bob Coecke Aleks Kissinger Alex Merry Shibdas Roy. The ghz/w-calculus contains rational airthmetic. *Electronic Proceedings in Theoretical Computer Science*, 2011.

[27] Amar Hadzihasanovic. A diagrammatic axiomatisation for qubit entanglement. *LICS '15 Proceedings of the 2015 30th Annual ACM/IEEE Symposium on Logic in Computer Science*, 2015.