Categorical Models of Meaning: Accommodating for Lexical Ambiguity and Entailment

Daniela Ashoush

St Hugh's College

University of Oxford

A dissertation submitted for the degree of

MSc in Mathematics and Foundations of Computer Science

Trinity 2015

To my parents, Cynthia and Ramzi

Acknowledgements

With full gratitude, I would like to acknowledge the following individuals who encouraged, inspired, supported, assisted, and made sacrifices to help me complete this dissertation as well as pursue my masters degree.
In the faculty, I would like to thank Prof. Bob Coecke for all the meetings he granted me and every response I got to my queries. His ideas were crucial and insightful, and they set me on the right path in every chapter of this dissertation. I would also like to thank Oscar Cunningham for taking the time to explain to me the intricacies of finding states in $\mathbf{CPM}^2(\mathbf{Rel})$, as well as my college advisor Chris Heunen for his invaluable support and for discussing with me the iterations of the CPM-construction.
From outside the faculty, I would like to thank all my friends in Oxford for the moral support they provided me, whether it was knowingly or not. In particular I wish to extend my gratitude to David Watson who proofread parts of my dissertation and for his intelligent remarks which would have without a doubt escaped me had he not pointed them out. Moreover, I would like to thank my flatmate Hind Kraytem for helping me achieve an adequate environment at home suitable to the nature of dissertation work. I do not believe I could have found a better person to share a house with anywhere else.
From back home, I would like to deeply thank my family, Cynthia, Ramzi, and Nathalie, without whom I would not have been able to accomplish any of this. Their unconditional support and unwavering presence in all its various shapes throughout the past year have been critical in helping me find the strength and confidence to carry on and give my best day by day. In addition, I cannot forget the care and love shown to me by Téta Sou, Jean, Alexandra, Carole, and the rest. I thank them all.
Finally, a special thank you goes to Karl Dagher. His support came at the most difficult times as well as the days were everything went well. I would also like to thank him for proofreading this dissertation.

Abstract

The categorical compositional distributional model of Coecke, Sadrzadeh and Clark combines the distributional theory of meaning in terms of vectors space models, and the compositional model of meaning in terms of pregroup grammars, in a unified categorical setting. It provides a way of computing the meaning of sentences and strings of words based on the grammatical relationships between the different constituents, and the empirical meaning vectors of individual words: the grammatical reductions of pregroups are lifted to morphisms in vector spaces. This is based on the fact that pregroup grammars and vector spaces share a compact-closed structure.
In the aim of modeling a feature of language, this framework was extended to include mixed states by the means of Selinger's CPM-construction. This translates into the passage from vectors to density matrices. Two applications of this extension are modelling lexical ambiguity, and modelling entailment relationships between words.
The aim of this dissertation is to further extend the model of Coecke et al. by iterating the CPM-construction in order to accommodate for two features of language: ambiguity *and* entailment. We present an axiomatisation of the CPM^2-construction and generalise it to axiomatise the CPM^n-construction – which has the potential to accommodate for an increasing number of features. The CPM^2-construction preserves †-compact closed structure, and ensures that the grammatical reductions are carried over to the new category. We then study the structure of double-density matrices, new states introduced by the CPM^2-construction, and investigate their role in accounting for both lexical ambiguity and entailment. This framework is showed to successfully model the two features of language via a series of examples, and is equipped with independent measures of the levels of ambiguity and entailment in words.

Contents

Chapter 1

Introduction

1.1 Background and scope

Historically, the computational representation of natural language meaning has been approached in two somewhat orthogonal ways: on the one hand, *distributional semantics*, echoing Wittgenstein's saying "the meaning of a word is its use" [42], builds on the *distributional hypothesis* [23] according to which the meaning of a word is determined by its context. A widely used distributional model for word meaning is the vector space model. In this model, a set of relevant context words is chosen from a large corpus of text as a basis for a vector space. Words are then represented as vectors of co-occurrence frequencies with the different context words. While this model is quantitative and offers a way of comparing the meaning of words, it does not scale to the level of sentences, but more importantly, it fails to take into account the syntactic relations between the words of a string of words.

Compositional formal semantics, on the other hand, views words as parts of a logical expression. Following the framework of Lambek [27] and Montague [17], this model builds on Frege's principle of compositionality which states that the meaning of an expression is determined by the meanings of its constituents and the rules that combine them. However, it does not reliably provide individual word meaning.

The model of Coecke et al. [14] unifies these two approaches under a categorical setting. This model is based on the fact that pregroup grammars, used to model grammatical relationships between words, and finite-dimensional vector spaces share a compact-closed categorical structure. By the means of a strong monoidal functor between the two categories, this framework provides a way to express the grammatical reductions in the pregroup grammar as morphisms in the category of finite-dimensional vector spaces.

The framework of [14] was extended by applying Selinger's CPM-construction [41] which adjoins to the pure states used in [14] the notion of mixed states. This translates into a passage from vectors to density matrices. This extension allows to model one feature of language. Piedeleu [34] uses density matrices to model *ambiguity* in language, or more particularly, *homonymy*: two words are said to be *homonymous* if they share the same spelling and pronunciation but refer to two different concepts. In this dissertation, we will use "ambiguity" and "homonymy" interchangeably. Balkır [4] resorted to density matrices in order to model lexical entailment, or more specifically *subsumption* relationships between words: a word w_1 is said to subsume a word w_2 if its meaning generalises that of w_2. We say that w_1 is a *general* word.

The aim of this dissertation is to extend this framework even further by iterating the CPM-construction, in order to model two features of language: ambiguity and entailment.

1.2 Outline

The next three chapters introduce the background needed for this dissertation. Chapter 2 presents the category theoretical concepts encountered in the categorical framework of [14], along with their relation to semantic analysis and their graphical calculus.
Chapter 3 revolves around the categorical model of [14]: it introduces both the compositional and the distributional models of meaning in terms of compact-closed categories, and unifies these seemingly orthogonal models by means of a strong monoidal functor.
Chapter 4 describes an extension of the previous model where words are represented by density matrices. These operators embody probabilistic mixing, which allows for the accommodation of one feature of language. Density matrices are completely positive maps, a notion formalised by Selinger's CPM-construction [41].

The last four chapters develop the theory behind the suggested extension and offer applications. In Chapter 5, I reformulate some of the axioms of Coecke's axiomatisation of the CPM-construction [7] and offer diagrammatic proofs of the theorems involved. I also define and axiomatise the CPM^2-construction, which introduces maps whose structure embodies two levels of mixing. I extend this to the CPM^n-construction for arbitrary n.
In Chapter 6, I show that this new framework is adequate for representing ambiguous, general words, the grammatical relations between them, and measuring their similarity, while also accounting for relational types.
Chapter 7 investigates density matrices and the properties they satisfy in order to motivate the introduction of double-density matrices. Double-density matrices are states in $\mathbf{CPM}^2(\mathbf{C})$ which model two levels of mixing, and the properties these states satisfy are identified. In addition, this chapter offers a novel approach to the characterisation of density matrices by showing that maps satisfying the conditions of *hermiticity* and

positive-semidefiniteness have the structure of a density matrix.
Finally, Chapter 8 details the role of double-density matrices in representing ambiguous, general words, and provides applications of the new model. The sentence space chosen is **Rel**, and a characterisation of the states in $\mathbf{CPM}^2(\mathbf{Rel})$ is presented. This chapter also introduces a way of measuring independently the level of ambiguity and entailment in a word.

1.3 Contributions

- I revise the axioms proposed in Coecke's axiomatisation of the CPM-construction [7], and diagrammatise the proofs involved. I also define the CPM^2-construction and axiomatise it in terms of a *squared-environment structure*. In addition, I generalise this to the CPM^n-construction for arbitrary n.

- I show that the CPM^2-construction preserves †-compact closed structure, and define compact-closure maps in $\mathbf{CPM}^2(\mathbf{C})$. I also show that $\mathbf{CPM}^2(\mathbf{C})$ possesses †-special commutative Frobenius algebras and define the relevant maps.

- I prove that maps satisfying the conditions of hermiticity and positive-semidefiniteness possess the structure of density matrices.

- I define the notion of *double-density matrices*, states in $\mathbf{CPM}^2(\mathbf{C})$ which account for two levels of mixing, and determine the properties they satisfy.

- I demonstrate how the framework models ambiguity and entailment in language by means of concrete examples involving the categories $\mathbf{CPM}^2(\mathbf{FHilb})$ and $\mathbf{CPM}^2(\mathbf{Rel})$, and present ways to measure independently the levels of ambiguity and entailment in words.

Chapter 2

A little categorical background

The field of *category theory* formalises mathematical structures in terms of a collection of *objects* and *arrows* or *morphisms*. Category theory is of particular interest to us because it allows the study of two relevant types of connections: the connection between quantum information flow and linguistic modelling, and that between structures representing grammar – pregroup grammars – and structures representing meaning – finite dimensional Hilbert spaces. These connections are captured in the categorical framework of Coecke et al. [14], in which categories formalise the *compositionality* of natural language.

In this chapter, I aim to introduce basic notions of category theory and their relation to semantic analysis. The main focus is not to provide a tutorial on categories: for this purpose, I redirect the reader to [11]. Categories are equipped with a graphical calculus (surveyed in [40]) that will be introduced in parallel.

2.1 Basic definitions

We start by recalling the definition of a category:

Definition 2.1.1. [2] A category is an algebraic structure that comprises:

- A collection of objects ***Ob(C)***, denoted by A, B, C...
- For each pair of objects (A,B), a set **C**(A,B) of morphisms with domain A and codomain B, denoted by $f : A \to B, f \in \mathbf{C}(A, B)$
- For any triple of objects (A,B,C), a composition map:

  $$c_{A,B,C} : \mathbf{C}(A, B) \times \mathbf{C}(B, C) \to \mathbf{C}(A, C)$$

  such that $c_{A,B,C}(f, g) = g \circ f$
- For each object A, an identity morphism id_A

Elements of a category satisfy two axioms:

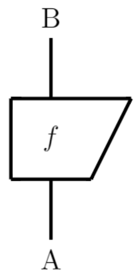
$$h \circ (g \circ f) = (h \circ g) \circ f$$
$$f \circ id_A = f = id_B \circ f$$

where the domains and codomains of the morphisms match so that the compositions are well-defined.

An interesting property, which stems from the first axiom in the above definition, is that morphisms can be composed *sequentially* and *associatively* to form new morphisms. From a linguistic perspective, objects can be thought of as grammatical types, and morphisms as "interactions" between grammatical types (these will become clearer as the next concepts unfold).

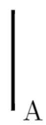
In the graphical language, morphisms are depicted by right trapezoidal boxes, with incoming and outgoing wires labelled by the corresponding objects. I use the convention whereby information flows in a bottom-top way.

$f : A \to B$



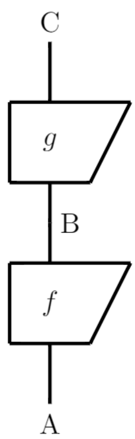
Identity is represented as a naked wire:

id_A



Finally, composition of morphisms is represented by two boxes on one wire:

$g \circ f$



2.2 Monoidal categories

A sentence is a concatenation, or juxtaposition, of words of different grammatical types, and some grammatical types are themselves juxtapositions of basic grammatical types. Monoidal categories offer a way to model this process.

Definition 2.2.1. [2] A monoidal category is a sextuple $(\mathbf{C}, \otimes, I, \alpha, l, r)$ where:

- $\mathbf{C}$ is a category
- $\otimes$: $\mathbf{C}\times\mathbf{C}\rightarrow\mathbf{C}$ is a bifunctor – a morphism of categories whose domain is a product category – called *tensor product*, which assigns to each pair of objects (A,B) a composite object A⊗B, and to each ordered pair of morphisms $(f : A \to C, g : B \to D)$, a parallel composite $f \otimes g : A \otimes B \to C \otimes D$
- I is the distinguished object of $\mathbf{C}$ called the *tensor unit*
- α, l, and r are natural isomorphisms with components:

$$\alpha_{A,B,C} : A \otimes (B \otimes C) \cong (A \otimes B) \otimes C$$

$$l_A : I \otimes A \cong A$$

$$r_A : A \otimes I \cong A$$

  such that $l_I = r_I : I \otimes I \cong I$, and satisfying the *pentagon* and *triangle* axioms [18].

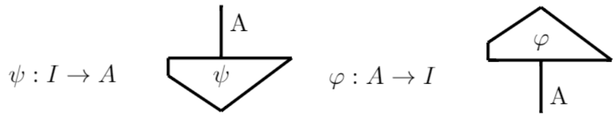
A *strict monoidal category* is a monoidal category where all the equivalences in the above definition are equalities. According to [30], any monoidal category is equivalent via a strong monoidal isomorphism to a strict monoidal category. In what follows, we will omit the natural isomorphisms.

The tensor unit gives rise to morphisms $\psi : I \to A$ called *states*, and morphisms $\varphi : A \to I$ called *effects*. Thanks to the tensor product, monoidal categories model both sequential and horizontal composition. From a linguistic perspective, this allows juxtaposition of words and grammatical types.

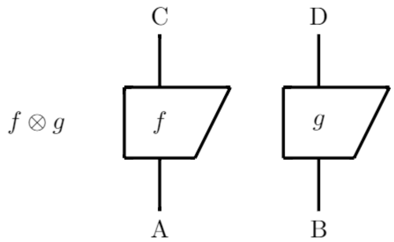
Monoidal categories admit a graphical calculus which is sound and complete, according to the following theorem by Selinger [40], originally based on a theorem by Joyal and Street [24]:

Theorem 2.2.1. *A well-formed equation between morphism terms in the language of monoidal categories follows from the axioms of monoidal categories if and only if it holds, up to planar isotopy, in the graphical language.*

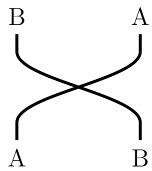
This theorem basically means that we are allowed to move boxes around, but not to cross or uncross wires. States *(resp. effects)* are represented by triangles with no input *(resp. output)* wire:



And horizontal composition is represented by putting morphisms (or wires) next to each other, ordered from left to right:



Another important notion is that of a *symmetric monoidal category*, which adjoins to the definition of a monoidal category a natural isomorphism $\sigma_{A,B} : A \otimes B \cong B \otimes A$, such that $\sigma_{B,A}^{-1} = \sigma_{A,B}$, and satisfying conditions stated in [2]. Symmetric monoidal categories also admit a sound and graphical calculus [40] where crossing and uncrossing wires is allowed, and will play an important role in manipulations involving *density matrices* in $\mathbf{CPM(C)}$ and *double-density matrices* in $\mathbf{CPM}^2\mathbf{(C)}$. Graphically, the swap map $\sigma_{A,B}$ is represented by two wires crossing:



2.3 Compact-closed categories

In the English language, a sentence is an *ordered* string of words that *interact* with each other to form meaning. For example, a transitive verb interacts with a subject on its left and an object on its right to form a sentence. These characteristics can be modelled using *compact-closed categories*:

Definition 2.3.1. [25] A *compact-closed category* is a monoidal category in which each object A has a left and right adjoint A^l and A^r, and morphisms:

$$\eta^l_A : I \to A \otimes A^l \qquad \epsilon^l_A : A^l \otimes A \to I$$
$$\eta^r_A : I \to A^r \otimes A \qquad \epsilon^r_A : A \otimes A^r \to I$$

satisfying the following *yanking equations*:

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

Furthermore, a compact-closed category is *symmetric* if $A^r = A^l := A^*$ for all A. The above four equalities collapse to two:

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

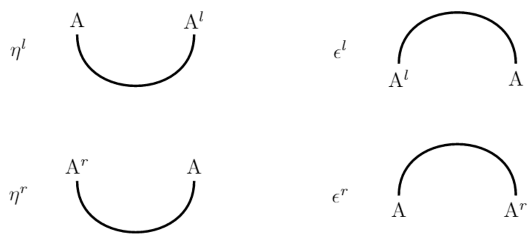
where $\epsilon_A : A^* \otimes A \to I$, and $\eta_A : I \to A^* \otimes A$.

In a compact-closed category, the left and right adjoints account for the order of words in a sentence, and the η and ϵ maps model the interactions of the different parts of a system.

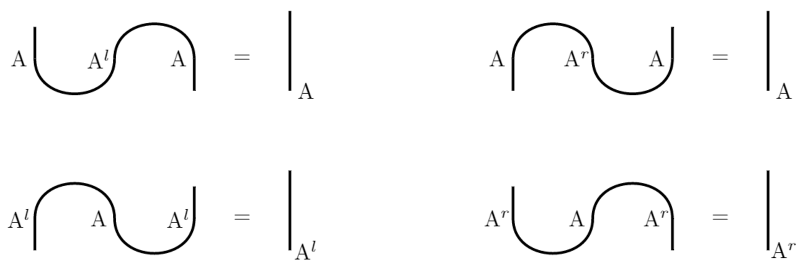
Compact-closed categories are equipped with a graphical language that is sound and complete, as expressed in [40]:

Theorem 2.3.1. *A well-formed equation between morphisms in the language of compact closed categories follows from the axioms of compact closed categories if and only if it holds, up to isomorphism of diagrams, in the graphical language*

The η and ϵ maps are depicted by *cups* and *caps* as follows:



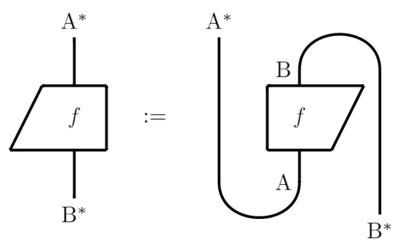
And the equations they satisfy are represented by the following, which boil down to yanking wire (hence the name "yanking equations"):



The introduction of cups and caps, also called *entanglement structure*, enables us to define three useful notions: *process-state duality*, *transposition*, and *trace*.

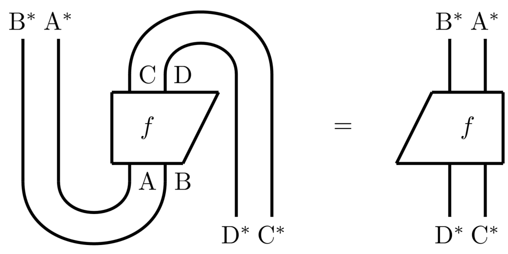
Process-state duality [10] is the concept according to which one can turn a *process* – a morphism $f : A \to B$, where $A, B \neq I$ – into a bipartite state, for example by applying a cup to the input of the process, and vice-versa, by applying a cap to one of the output wires of the bipartite state. A process f turned into a bipartite state by process-state duality is called the *name* of f, denoted $\lceil f \rceil$.

Definition 2.3.2. (Transposition) In a symmetric compact-closed category, the *transpose* of a morphism $f : A \to B$ is another morphism $f^* : B^* \to A^*$:



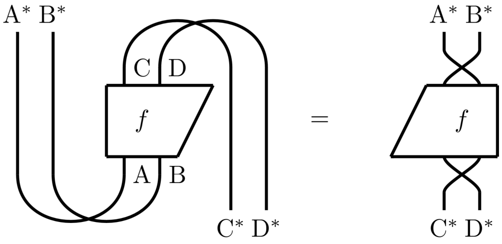
Note: the notion of transposition is also defined for non-symmetric compact-closed categories. In this case, we distinguish between left and right transpose.

Transposition of composite systems. There are two ways of defining the transpose of a composite system [10]. The first remains consistent with the 180° rotation defined above and gives the transpose of a morphism $f : A \otimes B \to C \otimes D$ by:



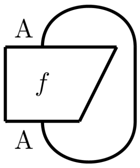
We will call this the *diagrammatic transpose* and denote it by $(\_)^*$. Notice that the cups and caps are nested.

The second definition of transposition crosses the caps and cups, which introduces a twist for the composite system:



We will call this the *algebraic transpose* and denote it by $(\ \_\ )^T$. It is in fact the one used in linear algebra. In what follows, we will consider the *diagrammatic transpose* as the default, and refer to it simply by "transpose".

Definition 2.3.3. (Trace) In a symmetric compact-closed category, the *trace* of a morphism $f : A \to A$ is the scalar:



2.4 †-compact-closed categories

So far, all the properties of language we have been seeking to formalise were compositional. One crucial property of a *distributional* nature, first introduced in [1], is that of the distance and angle between the meaning vectors of the words. To formalise this property, we introduce †-*compact-closed categories.*

Definition 2.4.1. [40] A †-*compact-closed category* $\mathbf{C}$ is a symmetric compact-closed-category with an involutive, identity-on-objects, contravariant functor $\dagger : \mathbf{C} \to \mathbf{C}$, which assigns to every morphism $f : A \to B$ its adjoint $f^\dagger : B \to A$, such that for all $f : A \to B$ and $g : B \to C$:

$$\begin{aligned} id_A^\dagger &= id_A && : A \to A \\ (g \circ f)^\dagger &= f^\dagger \circ g^\dagger && : C \to A \\ f^{\dagger\dagger} &= f && : A \to B \end{aligned}$$

The † functor of the above definition gives rise to two notions that will come in handy in the characterisation of density matrices:

Definition 2.4.2. [40] **(Unitarity, Hermiticity)** In a dagger category, a morphism $f : A \to B$ is called *unitary* if it is an isomorphism, and $f^\dagger \circ f = 1_A$ and $f \circ f^\dagger = 1_B$, i.e. $f^{-1} = f^\dagger$. A morphism $f : A \to A$ is called *self-adjoint* or *hermitian* if $f = f^\dagger$.

The † functor allows us to turn a state into an effect and an effect into a state. The notions of *inner product* of two states, and *norm* of a state, rely on this ability to turn a state into an effect and vice-versa. Before defining these two notions, we give the Dirac notation of states and effects [16]:

- A state ψ in the Dirac notation is given by $|\psi\rangle$ and called "ket"

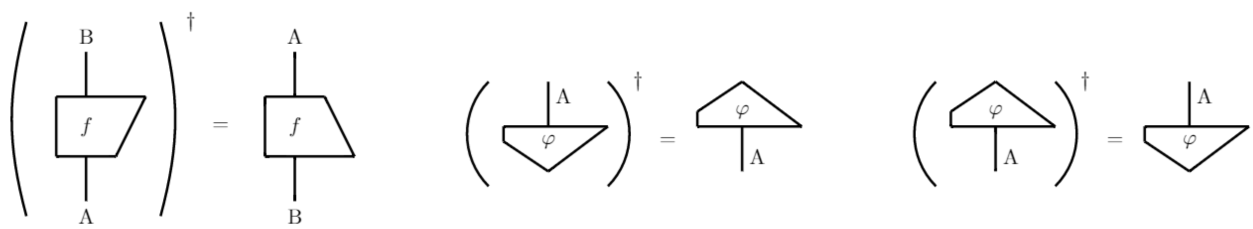
- An effect φ in the Dirac notation is given by $\langle\varphi|$ and called "bra"

Definition 2.4.3. [10] The *inner product* of two states φ and ψ is the scalar $\varphi^\dagger \circ \psi = \psi^\dagger \circ \varphi$, and is written in the Dirac notation as the "braket" $\langle\varphi|\psi\rangle$ or $\langle\psi|\varphi\rangle$.

Definition 2.4.4. [10] The *squared-norm* of a state φ is the inner product $\langle\varphi|\varphi\rangle$.

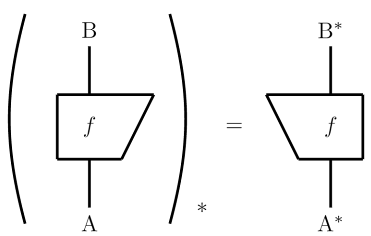
From a linguistic perspective, the inner product and the norm are essential to measure the *similarity* between two words or strings of words.

†-compact-closed categories admit a graphical language that was proved to be sound and complete in **Theorem 7.2** of [40]. Taking the adjoint of a morphism is reflecting it about the x-axis:

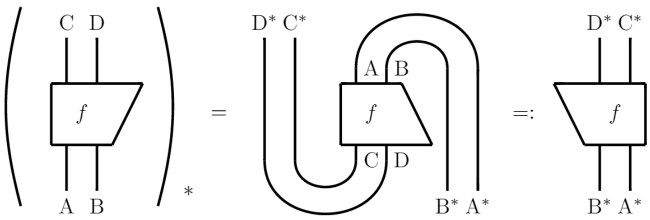


Finally, the dagger and the transpose functors give rise to a new functor which acts as follows:

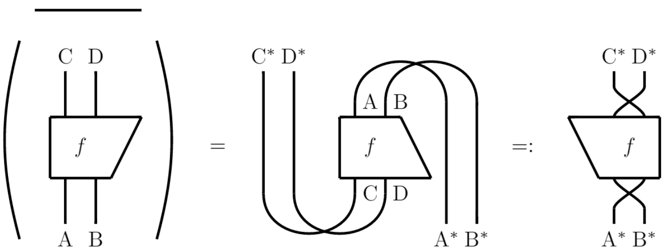
Definition 2.4.5. (Conjugation) In a †-compact-closed category, the *conjugate* of a morphism $f : A \to B$ is another morphism $f_* : A^* \to B^*$, such that $f_* = f^{*^\dagger} = f^{\dagger^*}$. The conjugate is depicted graphically by reflection about the y-axis:



Conjugate of a composite system. Two ways of defining the conjugate of a composite system emerge from the the different definitions of the transpose of a composite system. The *diagrammatic conjugate* of a composite system, denoted by $(\_)_*$, is the result of taking the diagrammatic transpose of the adjoint – or equivalently the adjoint of the diagrammatic transpose – of the system and is depicted as:



In what follows, we will consider the *diagrammatic conjugate* as the default, and refer to it simply by "conjugate". The *algebraic conjugate* of a composite system, denoted by $\overline{(\_)}$, is the result of taking the algebraic transpose of the adjoint – or equivalently the adjoint of the algebraic transpose – of the system and is depicted as:



The algebraic conjugate restores the order in the inputs and outputs by inducing a twist in the wires of the (diagrammatic) conjugate.

This chapter introduced the basic notions of category theory encountered in the framework of Coecke, Sadrzadeh and Clark [14], along with their semantic interpretation and their graphical calculus. In the next chapter, we will take a closer look at this framework.

Chapter 3

A compositional distributional categorical framework

Historically, there have been two seemingly orthogonal ways of representing language: the first builds on pregroup grammars and formalises the grammar of natural language without reasoning about the meaning of words, and the second is distributional and represents words as vectors in highly-dimensional vector spaces, without modeling syntax. The categorical model of Coecke et al. [14] seeks to unify these two approaches. This model is based on the fact that pregroup grammars and finite-dimensional vector spaces share a compact-closed categorical structure. Coecke et al. also extended their framework to model relative pronouns via the use of †-special commutative Frobenius algebras.

3.1 Pregroups as an account for compositionality

In a recent development by Lambek [28], *pregroups* are introduced as a tool for analysing the structure of syntax, using simple algebraic type reductions.

Definition 3.1.1. [5] A *pregroup algebra* is a structure $(P, \leq, \cdot, 1, (-)^l, (-)^r)$ where:

- $(P, \leq, \cdot, 1)$ is a *partially ordered monoid*
- $(-)^l$ and $(-)^r$ are unary operations on P, called the left and right adjoints, satisfying the inequalities:

$$\forall a \in P, \quad a^l \cdot a \leq 1 \leq a \cdot a^l, \quad a \cdot a^r \leq 1 \leq a^r \cdot a$$

We recall that a *partially ordered monoid* is a partially ordered set $(P, \leq, \cdot, 1)$, where P is a set of objects with partial ordering "$\leq$" , and "$\cdot$" is an associative, non-commutative monoid operation with monoid unit 1, satisfying $a \cdot 1 = a = 1 \cdot a$, for all $a \in P$. In what follows, we will omit the "$\cdot$" for the sake of simplicity, and replace "$\leq$" by "$\to$" to better illustrate type reductions.

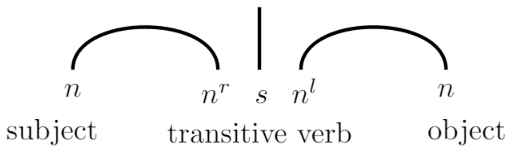
3.1.1 The role of pregroups in modelling grammar

To see how pregroups can formalise grammar of natural language, one can generate what is called a *pregroup grammar*, a pregroup algebra freely generated over a set of basic types [25]. For the purpose of this dissertation, we fix two basic grammatical types: $\{n, s\}$, where n is the grammatical type for *noun*, and s is the grammatical type for *sentence*.

A sentence is deemed *grammatical* whenever its reduction leads to the type s. Compound types are formed by adjoining and juxtaposing basic types [14]. For example, a transitive verb interacts with a subject to its left and an object to its right, to produce a valid grammatical sentence. Transitive verbs are therefore assigned the type $n^r s n^l$, and a transitive sentence reduces to a valid grammatical sentence, according to the inequalities of **Definition 3.1.1**:

$$n(n^r s n^l)n = (nn^r)s(n^l n) \to s$$

Graphically, one can represent this reduction by:



Note that this representation is analogous to the graphical calculus for compact-closed categories described in **Section 2.3**.

3.1.2 Pregroups as compact-closed categories

In [36], the authors generate a free compact-closed category from the pregroup algebra of a pregroup grammar. Let us denote this category by $\mathbf{C}_G$: the monoidal structure of $\mathbf{C}_G$ is induced by that of the pregroup algebra, each element a has left and right adjoints a^l and a^r (by **Definition 3.1.1**), and the η and ϵ maps are given by the inequalities:

$$\begin{array}{ll} \eta_a^l : 1 \leq aa^l & \epsilon_a^l : a^l a \leq 1 \\ \eta_a^r : 1 \leq a^r a & \epsilon_a^r : aa^r \leq 1 \end{array}$$

The yanking equations follow straightforwardly from the pregroup reductions.

Going back to the previous example of a transitive sentence, the reduction:

$$(nn^r)s(n^l n) \to s$$

corresponds to the map:

$$\epsilon_n^r 1_s \epsilon_n^l : nn^r s n^l n \to s$$

3.2 Finite-dimensional Hilbert spaces as an account for meaning

The model suggested in the previous section formalises the syntax of natural language, but fails to take into account the meaning of words. In this section, I introduce the *distributional* model of meaning, also called *vector space model*, which addresses this problem. Words are represented by vectors living in finite-dimensional semantic spaces, which form a †-compact-closed category.

3.2.1 Distributional semantics

Distributional semantics, first introduced by Firth [19], follows Harris's *distributional hypothesis* [23], according to which the meaning of a word is determined by its context. The basic assumption is that words appearing in a similar context must have similar meanings. The goal is to represent words as vectors in a semantic space, where the notions of angle and distance between vectors allow for *similarity* to be quantified.

Word vectors, usually normalised, live in a highly dimensional – but finite – semantic space with a fixed orthonormal basis $\{n_i\}_i$, where each n_i is a *context word* against which words to which we want to assign meaning are measured: given a word w to which we want to assign meaning, we rely on a large corpus of text to establish what is called the *relative frequency* of w with respect to each context word. Relative frequency is a measure of *co-occurrence*: it counts how many times word w occurs in the context of words n_i. The meaning vector of a word w is therefore given by:

$$\sum_i c_i n_i,$$

where c_i is the relative frequency of w with respect to context word n_i.

The vector spaces usually used in this model are *Hilbert spaces*: vector spaces that have the *inner product* structure, allowing for distance and angle measurements. As we will see in the next subsection, Hilbert spaces form a †-compact-closed category.

Models following this paradigm have been found to be very fruitful when applied to language processing tasks. [14] summarises some of these tasks, namely word sense discrimination and disambiguation [32, 39], text segmentation [6], and thesaurus extraction [21].

This model however has some important shortcomings: a problem of a distributional nature is that this model does not scale up to the level of sentences, since no corpus can reliably provide the distribution of a sentence. Another issue is that it does not take into account the ability of humans to understand new sentences: this capacity is based on a compositional mechanism whereby meaning is generated from words and their relations in a sentence.

3.2.2 FHilb as a †-compact-closed category

Finite-dimensional Hilbert spaces form a category **FHilb** that is †-compact-closed:

A monoidal category. The objects of **FHilb** are finite-dimensional vector spaces, and the morphisms are linear maps. The monoidal tensor is given by the the usual vector space tensor $\otimes$, and the tensor unit is the scalar field of the vector spaces. Theoretically, this scalar field can be chosen to be the field of complex numbers $\mathbb{C}$. We will consider $\mathbb{C}$ in the characterisation of density matrices and *double-density matrices*. However, in practice, the distributional model is obtained from real data and therefore lives in the real vector space $\mathbb{R}$. A vector, or column vector, $v \in V$ is represented by a linear map $I \to V$, where I is the scalar field, and a row vector by the linear map $V \to I$.

A symmetric compact-closed category. The vector space tensor $\otimes$ being commutative, the left and right adjoints collapse, and the adjoint of a vector space V is simply its dual vector space V^*. The inner product structure of Hilbert spaces induces an isomorphism between vector spaces and their duals: $V \cong V^*$. The η and ϵ maps are given by:

$$\eta_V : I \to V \otimes V :: 1 \mapsto \textstyle\sum_i n_i \otimes n_i$$
$$\epsilon_V : V \otimes V \to I :: v_i \otimes w_i \mapsto \langle v_i | w_i \rangle$$

The two equations are verified as follows:

$$(\epsilon_V \otimes 1_V) \circ (1_V \otimes \eta_V)(v) = (\epsilon_V \otimes 1_V)(v \otimes (\textstyle\sum_i n_i \otimes n_i)) = \sum_i \langle v | n_i \rangle \otimes n_i = v$$
$$(1_V \otimes \epsilon_V) \circ (\eta_V \otimes 1_V)(v) = (1_V \otimes \epsilon_V)((\textstyle\sum_i n_i \otimes n_i) \otimes v) = \sum_i n_i \otimes \langle v | n_i \rangle = v$$

A †-compact-closed category. The adjoint of a linear map $f : V \to W$ is the map $f^\dagger : W \to V$ satisfying $\forall v \in V, w \in W$, $\langle fv | w \rangle = \langle v | f^\dagger w \rangle$.

3.3 From grammar to semantics: a functorial passage

The two models of meaning presented in the previous sections are somewhat orthogonal: one is compositional but does not account for word meaning, and one is quantitative but non-compositional. Nonetheless, $\mathbf{C}_G$ and **FHilb** have something in common: they share a compact-closed structure. The unification of these two frameworks is done by transitioning from syntax to semantics via a *strong monoidal functor* Q [35].

Definition 3.3.1. [25] [26] **(Monoidal and strongly monoidal functors)** A functor F between two monoidal categories $\mathbf{C}$ and $\mathbf{D}$ is *monoidal* if there exists a morphism $I \to F(I)$ and a natural transformation $F(A) \otimes F(B) \to F(A \otimes B)$ satisfying the corresponding coherence conditions.
A monoidal functor is said to be *strongly monoidal* or *strong monoidal* if the above morphism and natural transformation are invertible.

Proposition 3.3.1. *A strong monoidal functor F on two compact-closed categories $\mathbf{C}$ and $\mathbf{D}$ preserves the compact-closed structure, that is $F(A^l) = F(A)^l$, and $F(A^r) = F(A)^r$.*

Proof. We present the proof given in [25]: to show that $F(A^l)$ is indeed the left adjoint $F(A)^l$ of $F(A)$, we have:

$$F(A^l) \otimes F(A) \to F(A^l \otimes A) \to F(I) \to I \to F(I) \to F(A \otimes A^l) \to F(A) \otimes F(A^l)$$

The right adjoint is proved similarly. □

Let us now define the strong monoidal functor Q: $\mathbf{C}_G \to \mathbf{FHilb}$: it maps atomic grammatical types to basic vector spaces:

$$Q(n) = N \qquad Q(s) = S$$

By **Proposition 3.3.1**, $Q(t^l) = Q(t)^l$, *and* $Q(t^r) = Q(t)^r$, for every grammatical type t of $\mathbf{C}_G$. Note that since $Q(t)$ is an element of $\mathbf{FHilb}$, $Q(t^l) \cong Q(t^r) \cong Q(t)$. Furthermore, Q maps the monoidal tensor in $\mathbf{C}_G$ to the monoidal tensor in $\mathbf{FHilb}$. Therefore, juxtaposition of grammatical types in $\mathbf{C}_G$ is mapped to the tensor product of vector spaces. For example:

$$Q(nn^r) = Q(n) \otimes Q(n^r) = N \otimes N$$

Finally, Q acts on morphisms by mapping grammatical reductions in $\mathbf{C}_G$ to linear maps in $\mathbf{FHilb}$, for example:

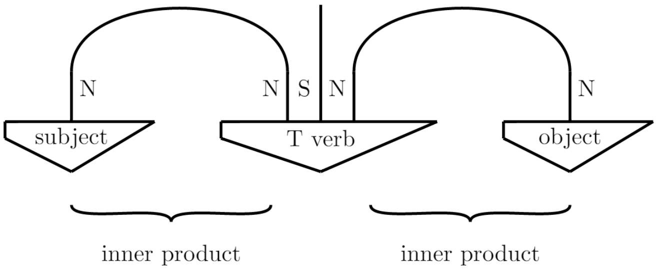
$$Q(\epsilon_n^r \cdot 1_s \cdot \epsilon_n^l : nn^r sn^l n \to s) = \epsilon_N \otimes 1_S \otimes \epsilon_N : N \otimes N \otimes S \otimes N \otimes N \to S$$

Meaning of strings of words. This definition of the strong monoidal functor Q allows to define the meaning of a sentence or string of words based on the individual meanings of the words constituting this sentence. Effectively, the grammatical reductions determine the order in which the linear maps in **FHilb** are applied.

Definition 3.3.2. Let $w_1 w_2 ... w_n$ be a string of words with types $t_1, t_2, ..., t_n$ and corresponding meaning vectors $|w_1\rangle, |w_2\rangle, ..., |w_n\rangle$. Let $\alpha : t_1 t_2 ... t_n \to x$ be a type-reduction to some grammatical type x. The meaning of $w_1 w_2 ... w_n$ is defined as:

$$|w_1 w_2 ... w_n\rangle := Q(\alpha)(|w_1\rangle \otimes |w_2\rangle \otimes ... \otimes |w_n\rangle)$$

Let us consider the example of a transitive sentence. The subject and object have basic type n, and the transitive verb $n^r s n^l$. Nouns are mapped to the basic vector space N, and the transitive verb to the tensor product space $N \otimes S \otimes N$. The pregroup reduction $n(n^r s n^l)n \to s$ corresponds to the morphism $\epsilon_n^r 1_s \epsilon_n^l : n(n^r s n^l)n \to s$ in $\mathbf{C}_G$, which is mapped to the linear map $Q(\epsilon_n^r \cdot 1_s \cdot \epsilon_n^l : n(n^r s n^l)n \to s) = \epsilon_N \otimes 1_S \otimes \epsilon_N : N \otimes N \otimes S \otimes N \otimes N \to S$ in **FHilb**. Graphically, this derivation is given by:



Let us show the explicit computation of the meaning of a transitive sentence. Let *subject*, *T verb* and *object* be defined by:

$$|subject\rangle = \sum_r c_r^{sub} |n_r\rangle$$
$$|T\ verb\rangle = \sum_{i,j,k} c_{ijk}^{verb} |n_i\rangle \otimes |s_j\rangle \otimes |n_k\rangle$$
$$|object\rangle = \sum_t c_t^{obj} |n_t\rangle$$

Then:

$$\begin{aligned}
|subject\ T\ verb\ object\rangle &= (\epsilon_N \otimes 1_S \otimes \epsilon_N)(|subject\rangle \otimes |T\ verb\rangle \otimes |object\rangle) \\
&= (\epsilon_N \otimes 1_S \otimes \epsilon_N)(\sum_r (c_r^{sub} |n_r\rangle) \otimes \sum_{i,j,k} (c_{ijk}^{verb} |n_i\rangle \otimes |s_j\rangle \otimes |n_k\rangle) \otimes \sum_t (c_t^{obj} |n_t\rangle)) \\
&= \sum_{r,i,j,k,t} c_r^{sub} c_{ijk}^{verb} c_t^{obj} \langle n_r|n_i\rangle \otimes |s_j\rangle \otimes \langle n_k|n_t\rangle \\
&= \sum_{r,i,j,k,t} c_r^{sub} c_{ijk}^{verb} c_t^{obj} \delta_{ri} \otimes |s_j\rangle \otimes \delta_{kt} \\
&= \sum_{i,j,k} c_i^{sub} c_{ijk}^{verb} c_k^{obj} |s_j\rangle
\end{aligned}$$

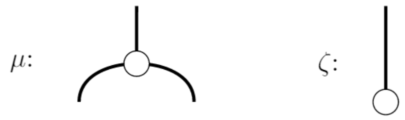
3.4 Frobenius algebras

One problem with the compositional distributional framework described so far is that some words, like relative pronouns – *who, which, that...* – cannot be modelled contextually: pronouns occur in practically any context, and so the context in which they occur cannot provide a reliable meaning. The authors of [14] extended their framework in [37] to model relative pronouns, using Frobenius algebras over vector spaces. The result is a model of relative pronouns that does not rely on co-occurrence frequencies, and that only takes into account the structural roles of the pronouns.

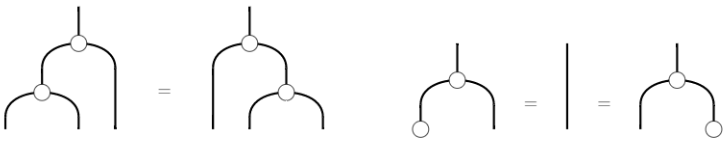
3.4.1 †-Frobenius algebras

In this subsection, we provide the basic definitions pertaining to Frobenius algebras. All definitions are taken from [10].

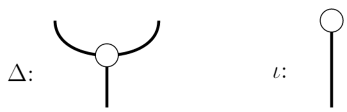
Definition 3.4.1. An *associative algebra* in a monoidal category consists of linear map $\mu : A \otimes A \to I$ and $\zeta : I \to A$, depicted by



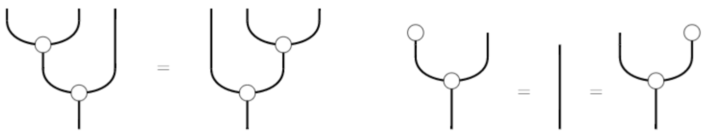
satisfying associativity and unit conditions:



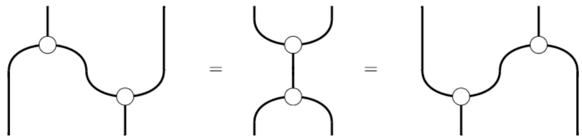
Definition 3.4.2. A *co-associative algebra* in a monoidal category consists of linear map $\Delta : I \to A \otimes A$ and $\iota : A \to I$, depicted by



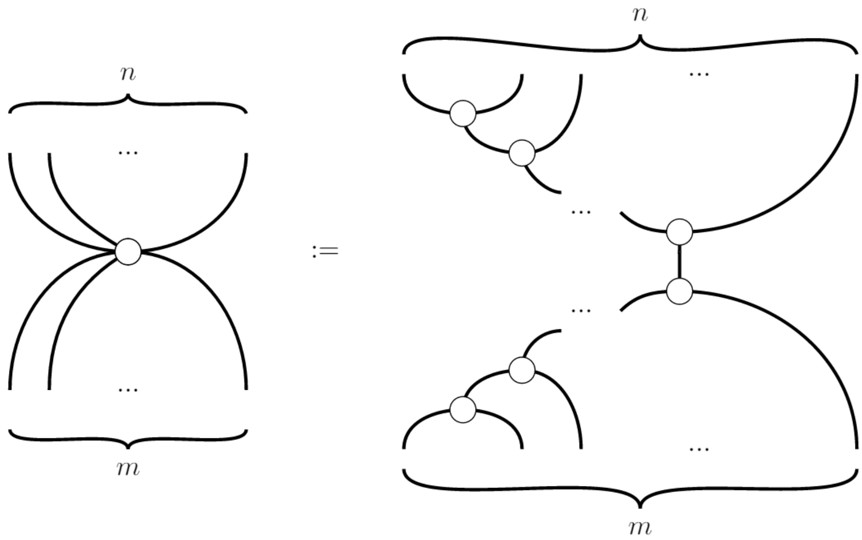
satisfying co-associativity and co-unit conditions:



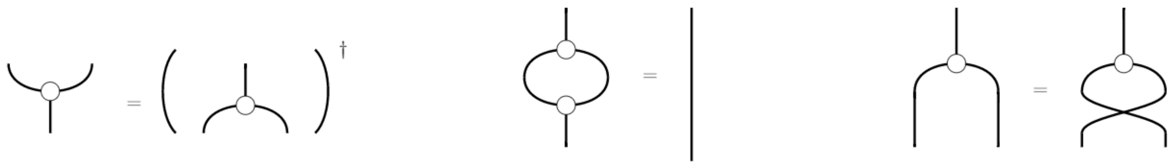
Definition 3.4.3. A *Frobenius algebra* is a quintuple $(A, \mu, \zeta, \Delta, \iota)$, such that (A, μ, ζ) is an associative algebra, (A, Δ, ι) is a co-associative algebra, and μ and Δ satisfy the *Frobenius equations*:



We define maps, called *spiders*, based on the structure provided by a Frobenius algebra:

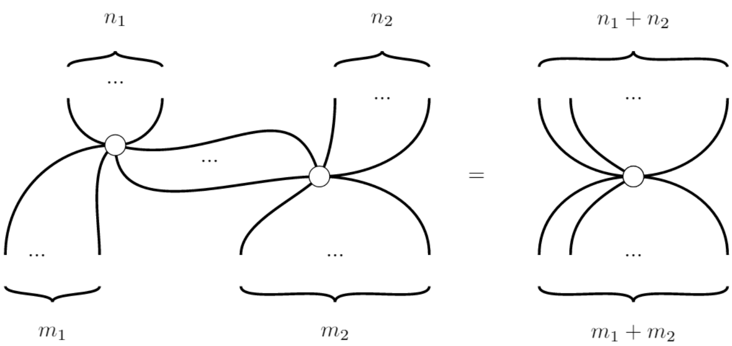


Definition 3.4.4. A †-*special commutative Frobenius algebra* is a Frobenius algebra that satisfies:



We now define the composition of spiders:

Theorem 3.4.1. (Spider fusion) *In a †-special commutative Frobenius algebra, spiders compose as:*



Proof. The proof consists in writing the spider on the left-hand side in canonical form and applying the rules of †-special commutative Frobenius algebras. The reader can consult the proof of **Theorem 8.109** in [10]. □

3.4.2 Frobenius algebras over vector spaces

In [12], the authors show that a finite-dimensional Hilbert space with an orthogonal basis has a †-commutative Frobenius algebra. Furthermore, when the basis is normalised, this †-commutative Frobenius algebra becomes special as well.

Let V be any Hilbert space, with fixed orthonormal basis $\{|i\rangle\}_i$. The maps of the †-special commutative algebra are given by:

$$\Delta :: |i\rangle \mapsto |i\rangle \otimes |i\rangle \qquad\qquad \iota :: |i\rangle \mapsto 1$$
$$\mu :: |i\rangle \otimes |j\rangle \mapsto \delta_{ij}\,|i\rangle := \begin{cases} |i\rangle & i = j \\ 0 & i \neq j \end{cases} \qquad \zeta :: 1 \mapsto \sum_i |i\rangle$$

We interpret the Δ map as *copying* information and encoding components in V into a matrix in $V \otimes V$, the ι map as *deleting*, and the μ map as *uncopying* – or *comparing* – elements: it picks out the diagonal elements of a matrix in $V \otimes V$ and returns them as a vector in V.

[37] defines subject- and object-relative pronouns in terms of the maps of the †-special commutative algebra described above:

It is clear from this definition that the relative clause (to the right of the pronoun) interacts with the head noun (to the left of the pronoun) via the relative pronoun. The relative clause is discarded, and the modified noun is returned.

In this chapter, I discussed two orthogonal models of meaning and their unification in Coecke et al.'s distributional compositional categorical model of meaning. I also discussed how this model accounts for relational types via †-special commutative Frobenius algebras. In the next chapter, I present an extension of this framework aiming to model one property of language.

Chapter 4

Accommodating for one feature of language

The framework described in the previous chapter was extended to include *density matrices*, which are generalisations of vectors. This extension allowed for one feature of language to be modelled: in [34], Piedeleu made use of density matrices to model *ambiguity* – more precisely *homonymy*,– and in [4], Balkır resorted to them in order to model *subsumption* relations, or *lexical entailment*.

In this chapter, I will justify the use of density matrices following the model in [35] by introducing *mixed states*, *completely positive maps*, along with Selinger's *CPM-construction* [41], and relate them to their linguistic interpretation in terms of ambiguity (the same analysis can be carried out for entailment). I will also suggest a way to accommodate yet another feature of language by iterating the CPM-construction.

4.1 Mixed states and density matrices

The framework we have been dealing with so far represents words as vectors in a Hilbert space, where these vectors correspond to *pure states*: a system is said to be in a *pure state* if we have complete knowledge about that system. In other words, we know precisely which state the system is in. In quantum physics, representing states as vectors in a Hilbert space has its limitations: what if we do not have complete knowledge about the state the system is in? The answer is by considering *probability distributions* over ensembles of pure states. States that are defined as such are called *mixed states*.

This situation is analogous to that of having to deal with a homonymous word. Representing a homonymous word as a convex sum of all its meanings collapses these meanings into a single vector. While this might seem to align with the distributional hypothesis, we would like to retain the ambiguity of the word in the absence of sufficient context, and allow it to collapse only when enough context is given to disambiguate the word partially or completely. It is therefore more intuitive to represent a homonymous word as a *probabilistic mixing* of its individual meanings [35].

The mathematical counterpart to a mixed state is called a *density matrix* or *density operator*. Density matrices and their characterisation will be discussed in detail in **Chapter 7**. We give the following definition of a density operator:

Definition 4.1.1. Given a set $\{|\varphi_m\rangle\}$ of pure, not necessarily orthogonal quantum states, and $\{p_m\}$ a probability distribution over them, define the density operator for this system by:

$$\rho \equiv \sum_m p_m |\varphi_m\rangle \langle\varphi_m|$$

From a linguistic perspective, the meaning of a homonymous word w is given by $\rho(w) = \sum_m p_m |w_m\rangle \langle w_m|$, where each meaning w_m has probability p_m. In the case of an unambiguous word $w^{'}$, the meaning of $w^{'}$ is given by $\rho(w^{'}) = |w^{'}\rangle \langle w^{'}|$.

Note: When accommodating for lexical entailment, the same analysis is carried out: *general words* – i.e. words that generalise the meaning of other words – are represented as a probabilistic mixing of "pure" words, and the meaning of a pure word is given by *doubling* that word.

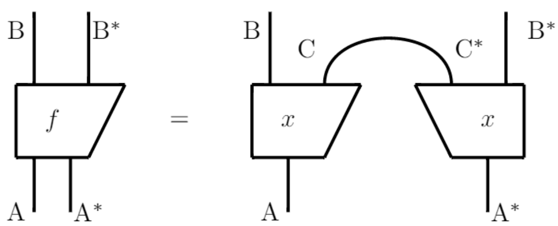
4.2 Completely positive maps and the CPM-construction

In the Hilbert space model, the morphisms are linear maps and map states to states. In the mixed setting, we need morphisms that map density matrices to density matrices. These are called *completely positive maps*.

Definition 4.2.1. Let A,B be objects in a †-compact-closed category **C**. A morphism $f : A \otimes A^* \to B \otimes B^*$ of **C** is *completely positive* or *CP* if there exists an object C of **C** and a morphism $x : A \to B \otimes C$ such that:

$$f = (1_B \otimes \epsilon_{C^*} \otimes 1_{B^*}) \circ (x \otimes x_*)$$

or graphically:



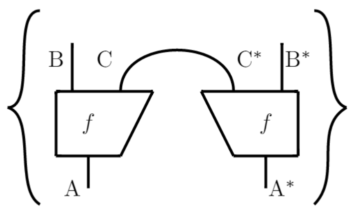
Note that Selinger proved this definition to be equivalent to his initial definition of completely positive maps in [41]. Selinger also laid down the properties of CP maps, namely that the identity for any CP maps f and g with appropriate domains and codomains, the identity $A \otimes A^* \to A \otimes A^*$, the composition $g \circ f$, the tensor $f \otimes g$ and the tensor $f \otimes f_*$ are completely positive.

We need to define a construction that introduces mixed states, preserves †-compact closure, and allows for a †-special commutative Frobenius algebra to be defined. This construction is Selinger's CPM-construction:

Definition 4.2.2. The CPM-construction [9, 41]
Given a †-compact closed category C, define a new category **CPM(C)** as follows:

(i) The objects of **CPM(C)** are the objects of **C**

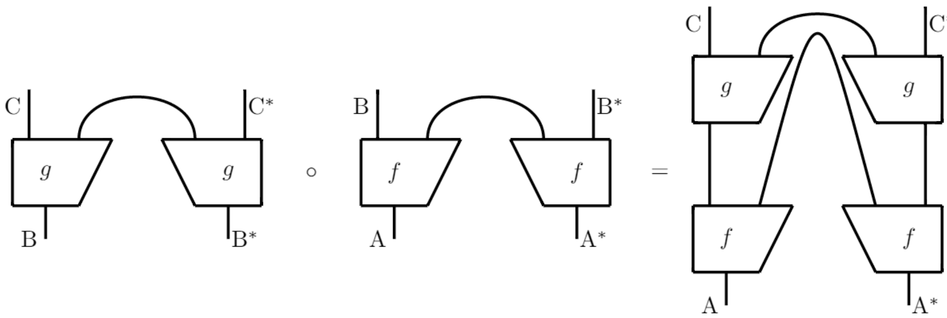
(ii) The morphisms $A \to B$ of **CPM(C)** are of the form $(1_B \otimes \epsilon_{C^*} \otimes 1_{B^*}) \circ (f \otimes f_*)$, or graphically:



where C is the *ancillary system* of $(1_B \otimes \epsilon_{C^*} \otimes 1_{B^*}) \circ (f \otimes f_*)$.

(iii) Identities are inherited from **C**

(iv) Composition is defined the usual way:



In the next section, we show that the CPM-construction preserves compact closure, and that it allows for a †-special commutative Frobenius algebra to be defined.

4.3 CPM(C) as a †-compact closed category

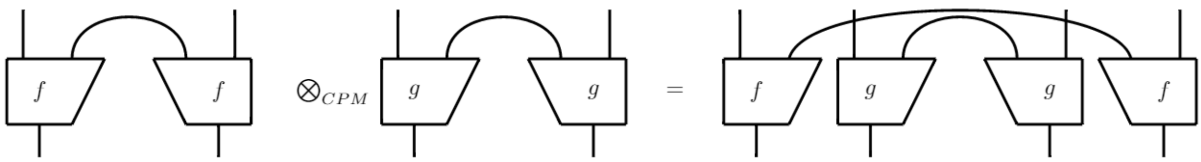
We begin this section by an important theorem stated and proved in Selinger's paper [41]:

Theorem 4.3.1. *Let **C** be a†-compact closed category. **CPM(C)** is again a †-compact closed category.*

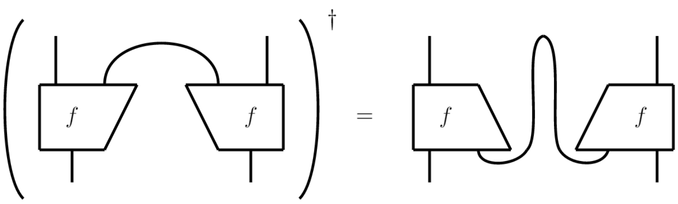
Proof. The proof of this theorem can be found in Selinger's paper [41], under **Theorem 4.20**. It is worth noting, however, that the proof uses the following: the fact that $f \otimes f_*$ is a CP map for any CP map f yields an identity-on-objects functor $F : \mathbf{C} \to \mathbf{CPM(C)}$ which maps morphisms f to $f \otimes f_*$. This functor is shown to preserve the compact structure. □

We can now extend **Definition 4.2.2** to include the following [9]:

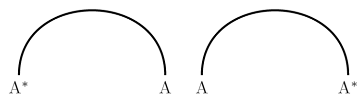
(v) The tensor unit I and the tensor product of objects are inherited from **C**, and the tensor product of morphisms is defined as follows:



(vi) The dagger is defined the usual way:



(vii) The cap $\epsilon_A : A^* \otimes_{CPM} A \to I$ is given by:

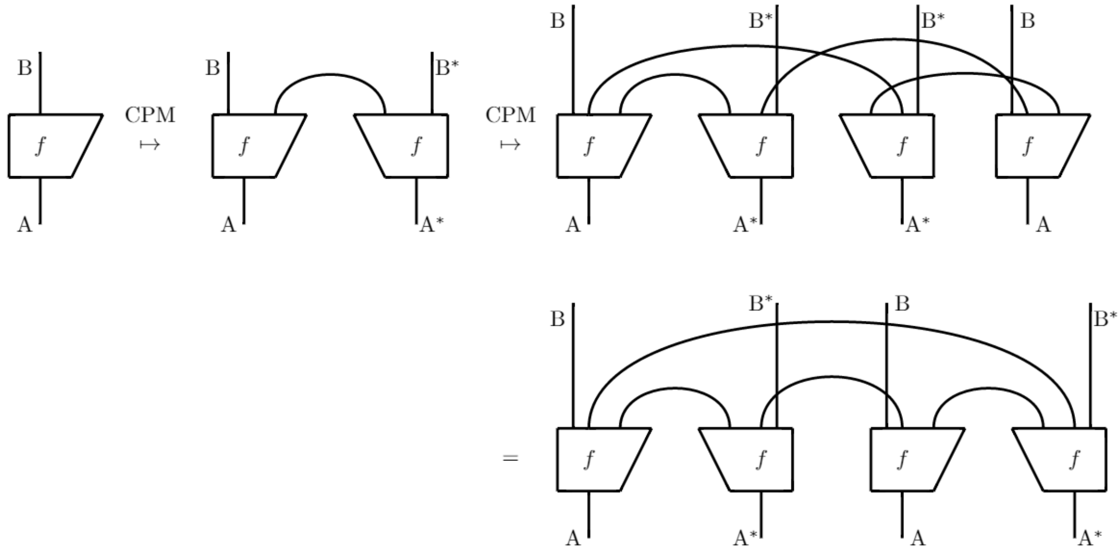


Frobenius algebra. [35] considers the *doubled version* of the Frobenius algebra in **C**, that is, the Frobenius algebra in **CPM(C)** with maps defined as the image of the Frobenius algebra maps in **C** by the functor F. Piedeleu [34] shows that the Frobenius algebra in **CPM(C)** based on a †-Frobenius algebra in **C** is indeed a †-Frobenius algebra (note that Piedeleu defines the Frobenius agebra maps in [34] in a different but ultimately equivalent way to the definition in [35]). Similarly, it can be shown that this †-Frobenius algebra, when based on a †-special commutative Frobenius algebra in **C**, is special and commutative.

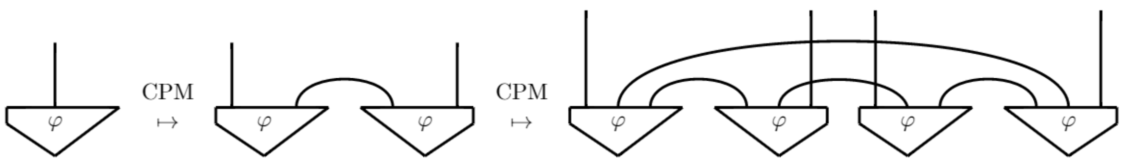
The construction defined in this chapter fulfils the required goals: the category of operator spaces and completely positive maps is a †-compact closed category and possesses a †-special commutative Frobenius algebra that accounts for copying, deleting, and comparing information. The use of density matrices allows for modelling one feature of language.

4.4 Accommodating for two features of language

Our ultimate goal is to extend the categorical framework even further to accommodate for a second feature of language. Since, by **Theorem 4.3.1**, **CPM(C)** is a †-compact closed category whenever **C** is a †-compact closed category, a solution is obtained by iterating the CPM-construction. Morphisms $f : A \to B$ in $\mathbf{CPM}^2\mathbf{(C)}$ are derived as follows:



And states are given by:



Assuming that the first application of the CPM-construction accounts for homonymy and the second for entailment, a *general, non-homonymous word* and a *non-general, homonymous word* correspond respectively to:

This chapter introduced mixed states and their mathematical counterparts, density matrices, and gave their linguistic interpretation in terms of ambiguity. The CPM-construction yields a †-compact closed category and models adequately one feature of language, and iterating the CPM-construction will allow us to account for more features.

This concludes the background needed for this dissertation. The next chapters will introduce a new framework based on *double-density matrices*, mathematical tools used to model states that have two levels of mixing. This framework will be shown to adequately model two features of language: ambiguity and entailment.

Chapter 5

Environment structures and CPM-constructions axiomatisation

The previous chapter introduced density matrices, completely positive map representing mixed states. In [7], Coecke recasts the CPM-construction as an axiomatisation of maximally mixed states.

In this chapter, I investigate and revise the methods used in Coecke's paper to axiomatise $\mathbf{CPM(C)}$, and offer an axiomatisation of $\mathbf{CPM}^2\mathbf{(C)}$, the category obtained by applying the CPM-construction to a category $\mathbf{C}$ twice, in terms of what we will call *discardings 1 and 2*. In the last section, I generalise this axiomatisation to $\mathbf{CPM}^n\mathbf{(C)}$, the category obtained by applying the CPM-construction to a category $\mathbf{C}$ n times.

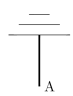
5.1 Axiomatisation of CPM(C)

In this section, I revise the axiomatisation of $\mathbf{CPM(C)}$ presented in [7], alter some of the axioms and notations to better suit higher orders of iteration of Selinger's CPM-construction, and offer a diagrammatic representation of the proofs involved.

5.1.1 Environment structure and implications

Definition 5.1.1. A $\top$- structure or *environment structure* on a †-compact-closed category C consists of:

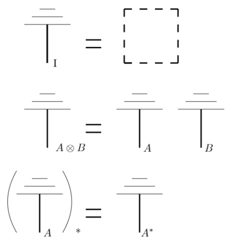
(i) a designated effect $\top_A : A \to I$ for each object A of $\mathbf{C}$, called the *maximally mixed effect* or *discarding* and depicted as:



which satisfies the following properties:

- $\top_I = 1_I$
- $\top_{A\otimes B} = \top_A \otimes \top_B$
- $(\top_A)_* = \top_{A^*}$

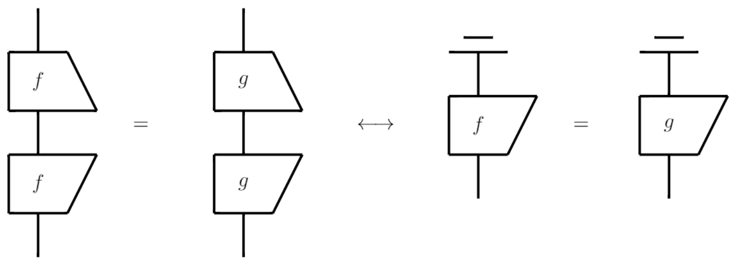
The above properties are respectively represented by the following diagrams:



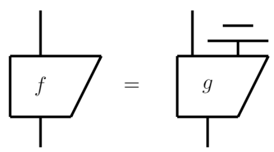
(ii) an all-objects-including sub-†-compact-closed category $\mathbf{C}_\Sigma$ of *pure morphisms*, which carries an entanglement structure, and which is such that for all morphisms f, g of $\mathbf{C}_\Sigma$:

$$f^\dagger \circ f = g^\dagger \circ g \Longleftrightarrow \top_{codom(f)} \circ f = \top_{codom(g)} \circ g \tag{5.1}$$

or graphically:



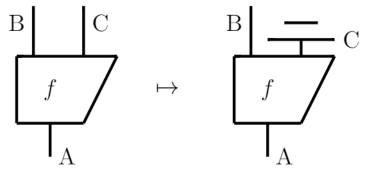
(iii) the **purifiability axiom** [13]: for every morphism $f : A \to B$ in $\mathbf{C}$, there exists a morphism $g : A \to B \otimes C$ in $\mathbf{C}_\Sigma$ such that:



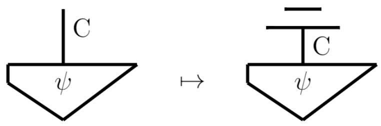
A more rigorous definition of the notion of "*purifiability*" is given in definitions **5.1.2** and **5.1.3** below.

Definition 5.1.2. In a †-compact-closed category $\mathbf{C}$ with a $\top$-structure:

- the *partial internal trace* is the map $tr^C_{A,B} : \mathbf{C}(A, B \otimes C) \to \mathbf{C}(A, B) :: f \mapsto (1_B \otimes \top_C) \circ f$, for objects A, B, C and any arrow $f : A \to B \otimes C$ in $\mathbf{C}$. Graphically:



- the *full internal trace* is the map $tr^C : \mathbf{C}(I, C) \to \mathbf{C}(I, I) :: \psi \mapsto \top_C \circ \psi$, for an object C and a state $\psi : I \to C$ in $\mathbf{C}$. Graphically:

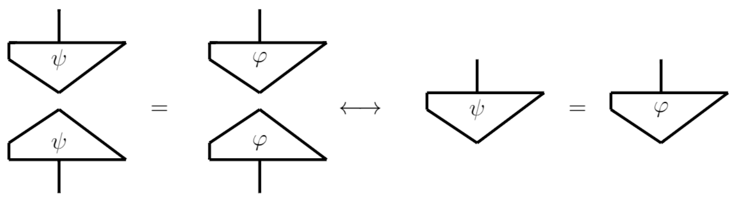


Definition 5.1.3. In a †-compact closed category $\mathbf{C}$ with an environment structure, define a *purification* of an operation $f : A \to B$ to be a pure operation $g : A \to B \otimes C$ which is such that $f = tr^C_{A,B}(g)$. f is said to be *purifiable*.

Axiom (5.1) of **Definition 5.1.1** entails an important special case: let us consider two effects $\psi, \varphi : A \to I$ in $\mathbf{C}_\Sigma$. By axiom (5.1):

$$\psi^\dagger \circ \psi = \varphi^\dagger \circ \varphi \iff \top_I \circ \psi = \top_I \circ \varphi \iff 1_I \circ \psi = 1_I \circ \varphi \iff \psi = \varphi \iff \psi^\dagger = \varphi^\dagger.$$

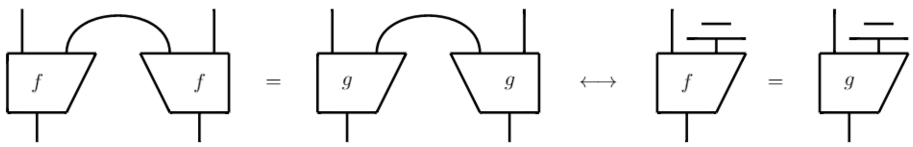
Graphically,



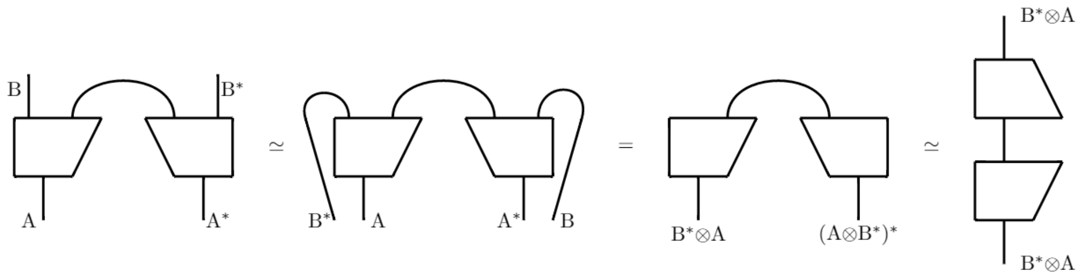
which is exactly the *preparation-state agreement axiom* [8]. The following conclusion is reached:

$$\text{axiom (5.1)} \Rightarrow \text{preparation-state agreement axiom} \tag{5.2}$$

Axiom (5.1) can also be stated as:



This stems from the fact that:



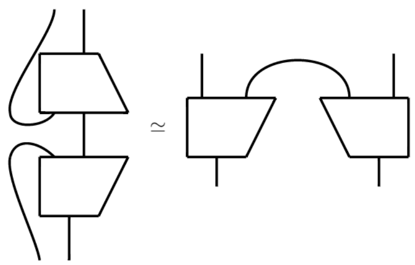
where the first and last equalities hold by properties of the entanglement structure of $\mathbf{C}_\Sigma$.

This new formulation of axiom (5.1) has an important implication stated in **Proposition 5.1.1** below. Recall first that in a †-compact-closed category $\mathbf{C}$, a morphism $f : A \to A$ of $\mathbf{C}$ is *positive* if and only if it decomposes as $f = g^\dagger \circ g$, for some morphism $g : A \to B$ of $\mathbf{C}$.

Proposition 5.1.1. *In a †-compact-closed category* $\boldsymbol{C}$ *with a* $\top$*-structure, axiom (5.1) gives rise to an isomorphism of categories*

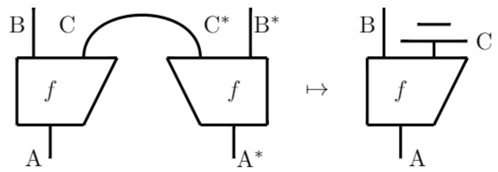
$$F : \mathbf{C}_\Sigma^{pos} \simeq \mathbf{C},$$

where $\boldsymbol{C}_\Sigma^{pos}$ *is the homset of all positive morphisms in* $\boldsymbol{C}_\Sigma$*, i.e. morphisms of the form*:



Proof. Define F as follows:

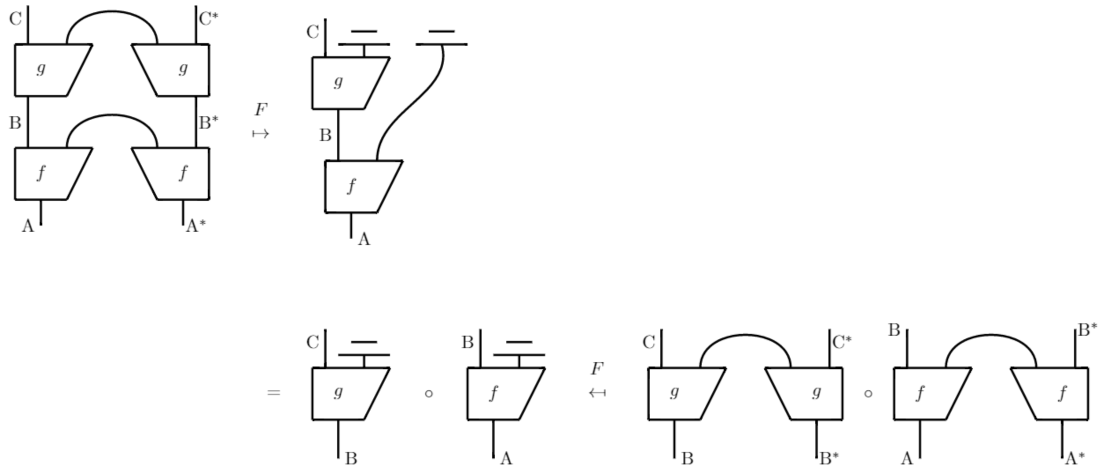
- F maps objects to themselves
- F maps morphisms $(1_B \otimes \epsilon_{C^*} \otimes 1_{B^*}) \circ (f \otimes f_*)$ in $\mathbf{C}_\Sigma^{pos}(A \otimes A^*, B \otimes B^*)$ to $(1_B \otimes \top_C) \circ f$ in $\mathbf{C}(A, B)$, or graphically:



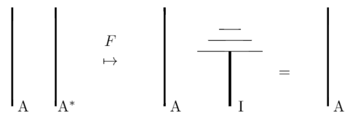
The forward direction of axiom (5.1) ensures that F assigns a unique interpretation to a given morphism in the domain category. Therefore, F is well-defined.

Functoriality is shown as follows:

- $F(g \circ f) = F(g) \circ F(f)$



- $F(id_A) = id_{F(A)} = id_A$



It remains to show that F is an isomorphism, i.e. that both the objects and the morphisms of $\mathbf{C}_\Sigma^{pos}$ and $\mathbf{C}$ are in one-to-one correspondence to each other.
The one-to-one correspondence between objects is trivially implied by the fact that F is identity-on-objects.
A one-to-one correspondence between morphisms is equivalent to showing that F is full and faithful: by the backward direction of axiom (5.1), every morphism $F_{A,B}$: $\mathbf{C}_\Sigma^{pos}(A \otimes A^*, B \otimes B^*) \to \mathbf{C}(A, B)$ is injective, which shows that F is *faithful*. By the *purifiability axiom*, every morphism f in $\mathbf{C}$ is purifiable, i.e. there exists a pure morphism g in $\mathbf{C}_\Sigma$ such that f is the result of discarding parts of the output of g. Therefore, every morphism $F_{A,B} : \mathbf{C}_\Sigma^{pos}(A \otimes A^*, B \otimes B^*) \to \mathbf{C}(A, B)$ is surjective, and F is *full*.

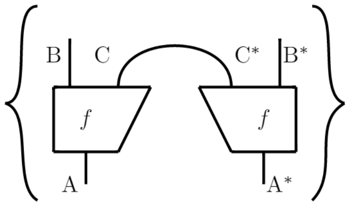
$\therefore$ axiom (1) $\Rightarrow \mathbf{C}_\Sigma^{pos} \simeq \mathbf{C}$. $\square$

5.1.2 Recovering Selinger's CPM-construction

We refer the reader the the definition of Selinger's CPM-construction (**Definition 4.2.2**) in **Chapter 4** and its extended version in **Section 4.3**. We remind the reader that given a †-compact closed category $\mathbf{C}$, the CPM-construction yields a new †-compact closed category $\mathbf{CPM(C)}$.

Theorem 5.1.1. *Let* $\mathbf{C}$ *be a* †*-compact-closed category. If* $\mathbf{C}$ *has an environment structure, then* $\mathbf{CPM(C_\Sigma)} \simeq \mathbf{C}$*, and* $\mathbf{C}_\Sigma$ *satisfies the preparation-state agreement axiom.*

Proof. The category $\mathbf{C}$ has an environment structure. Therefore, $\mathbf{C}_\Sigma$ satisfies axiom (5.1). By (5.2), $\mathbf{C}_\Sigma$ also satisfies the preparation-state agreement axiom.
Apply the CPM-construction to $\mathbf{C}_\Sigma$. The maps of $\mathbf{CPM(C_\Sigma)}$ are the maps of the form:



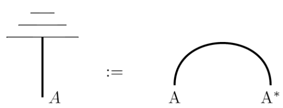
which are exactly the maps in $\mathbf{C}_\Sigma^{pos}$. By **Proposition 5.1.1**, the fact that $\mathbf{C}_\Sigma$ satisfies axiom (5.1) implies that $\mathbf{C}_\Sigma^{pos} \simeq \mathbf{C}$, and therefore that $\mathbf{C}_\Sigma^{pos}(A \otimes A^*, B \otimes B^*) \simeq \mathbf{C}(A, B)$. Therefore:

$$\mathbf{CPM(C_\Sigma)}(\mathrm{A,B}) \stackrel{\mathrm{def}}{=} \mathbf{C}_\Sigma^{pos}(A \otimes A^*, B \otimes B^*) \simeq \mathbf{C}(A, B)$$

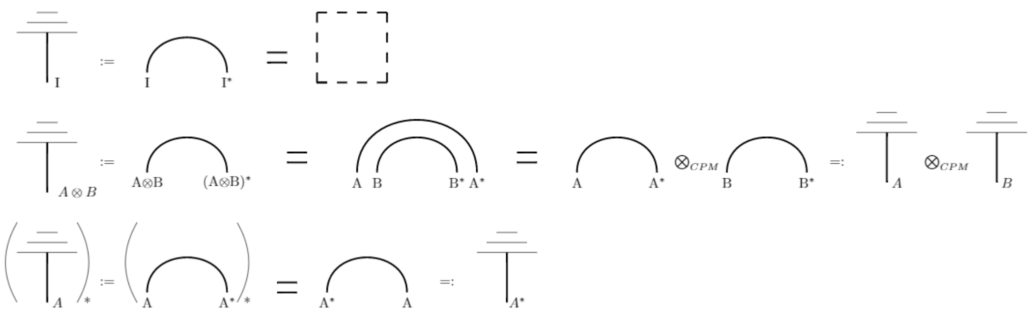
$\therefore \mathbf{CPM(C_\Sigma)} \simeq \mathbf{C}$. □

Theorem 5.1.2. *Let* $\mathbf{C}$ *be a* †*-compact-closed category with an entanglement structure, and let* $\mathbf{C}$ *satisfy the preparation state agreement axiom. Define the category* $\mathbf{CPM(C)}$ *according to* **Definition 4.2.2**. *Then* $\mathbf{CPM(C)}$ *has a* $\top$*-structure.*

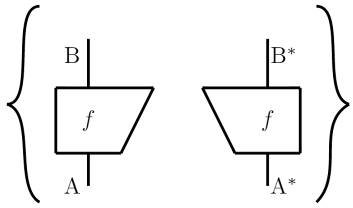
Proof. Define the maximally mixed effect as follows:



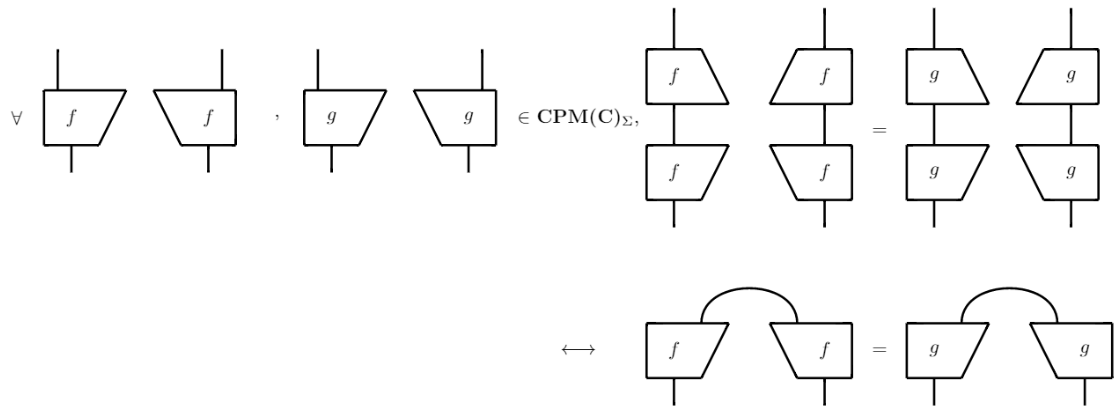
It is easy to check that $\top_A$ satisfies the required properties:



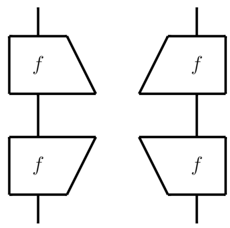
Consider now the all-objects-including sub-†-compact-closed category $\mathbf{CPM(C)}_\Sigma$ of pure morphisms, which carries an entanglement structure. The identity-on-objects embedding $F_{CPM} : \mathbf{C} \hookrightarrow \mathbf{CPM(C)}$ defined by Selinger in [41] maps pure morphisms in $\mathbf{C}$ to pure morphisms in $\mathbf{CPM(C)}$, i.e. morphisms in $\mathbf{CPM(C)}_\Sigma$. These morphisms are of the form:



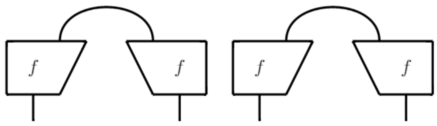
Let us show that $\mathbf{CPM(C)}_\Sigma$ satisfies axiom (5.1), graphically:



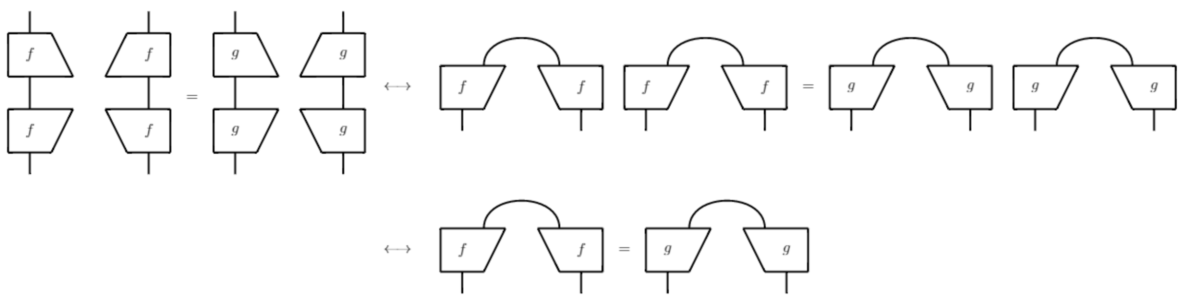
This is evident from topological manipulations:



can be stated as



by properties of the entanglement structure. Therefore,



Finally, every morphism in $\mathbf{CPM(C)}$ is purifiable, by definition of morphisms in $\mathbf{CPM(C)}$ and of the maximally mixed effect in $\mathbf{CPM(C)}$. $\square$

Corollary 5.1.1. *By theorems* **5.1.1** *and* **5.1.2**, *a †-compact-closed category* $\boldsymbol{C}$ *carrying a* $\top$*-structure coincides with* $\mathbf{CPM(C_\Sigma)} \simeq \mathbf{C}$, *and applying the CPM-construction to a †-compact-closed category* $\boldsymbol{C}$ *which satisfies the preparation-state agreement axiom induces a* $\top$*-structure on* $\boldsymbol{C}$.

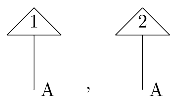
5.2 Axiomatisation of $\mathrm{CPM}^2(\mathrm{C})$

In this part, I apply and extend the findings of [7] to the *CPM*^{2}*-construction*. I start by defining the notion of *squared-environment structure* and the isomorphism of categories that stems from its axioms. I then formalise the notion of CPM^2-construction in terms of *completely squared-positive maps* and finally axiomatise it.

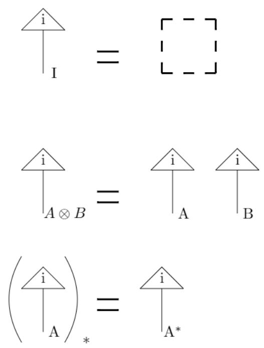
5.2.1 Squared-environment structure and implications

Definition 5.2.1. A $\top_1, \top_2$- structure or *squared-environment structure* on a †-compact-closed category C consists of:

(i) two designated effects $\top_{1,A}, \top_{2,A} : A \to I$ for each object A of $\mathbf{C}$, called respectively *discarding-1* and *discarding-2*, and depicted as:



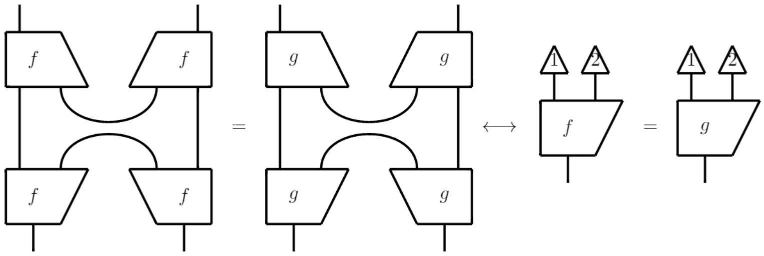
Both discarding effects satisfy the same properties as the maximally mixed state in the environment structure definition of **Part 5.1**, namely:



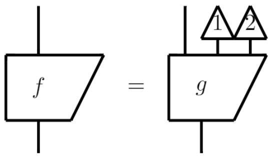
(ii) an all-objects-including sub-†-compact-closed category $\mathbf{C}_{\Sigma^2}$ of *pure morphisms*, which carries an entanglement structure, and which is such that for all morphisms f, g of $\mathbf{C}_{\Sigma^2}$ where $dom(f) = C_{f,1} \otimes C_{f,2}$, and $dom(g) = C_{g,1} \otimes C_{g,2}$:

$$\begin{aligned} &(f^\dagger \otimes f^*) \circ (1_{C_{f,1}} \otimes \eta_{C^*_{f,2}} \otimes 1_{C_{f,1}}) \circ (1_{C_{f,1}} \otimes \epsilon_{C^*_{f,2}} \otimes 1_{C_{f,1}}) \circ (f \otimes f_*) \\ =&(g^\dagger \otimes g^*) \circ (1_{C_{g,1}} \otimes \eta_{C^*_{g,2}} \otimes 1_{C_{g,1}}) \circ (1_{C_{g,1}} \otimes \epsilon_{C^*_{g,2}} \otimes 1_{C_{g,1}}) \circ (g \otimes g_*) \\ \Longleftrightarrow&(\top_{1,codom(f)} \otimes \top_{2,codom(f)}) \circ f = (\top_{1,codom(g)} \otimes \top_{2,codom(g)}) \circ g \end{aligned} \tag{5.3}$$

or graphically:



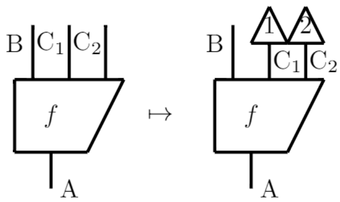
(iii) the **squared-purifiability axiom**: for every morphism $f : A \to B$ in $\mathbf{C}$, there exists a morphism $g : A \to B \otimes C_1 \otimes C_2$ in $\mathbf{C}_{\Sigma^2}$ such that:



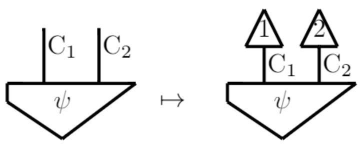
The concepts pertaining to "*squared-purifiability*" are defined below.

Definition 5.2.2. In a †-compact-closed category $\mathbf{C}$ with a squared-environment structure:

- the *squared-partial internal trace* is the map $tr^{C_1,C_2}_{A,B} : \mathbf{C}(A, B \otimes C_1 \otimes C_2) \to \mathbf{C}(A, B) :: f \mapsto (1_B \otimes \top_{1,C_1} \otimes \top_{2,C_2}) \circ f$, for objects A, B, C_1, C_2 and any arrow $f : A \to B \otimes C_1 \otimes C_2$ in $\mathbf{C}$. Graphically:



- the *squared-full internal trace* is the map $tr^{C_1,C_2} : \mathbf{C}(I, C_1 \otimes C_2) \to \mathbf{C}(I, I) :: \psi \mapsto (\top_{1,C_1} \otimes \top_{2,C_2}) \circ \psi$, for objects C_1, C_2 and a state $\psi : I \to C_1 \otimes C_2$ in $\mathbf{C}$. Graphically:

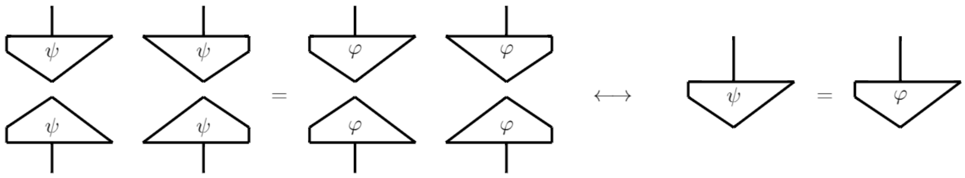


Definition 5.2.3. In a †-compact closed category $\mathbf{C}$ with a squared-environment structure, define a *squared-purification* of an operation $f : A \to B$ to be a pure operation $g : A \to B \otimes C_1 \otimes C_2$ which is such that $f = tr^{C_1,C_2}_{A,B}(g)$. f is said to be *squared-purifiable*.

As in the previous part, axiom (5.3) of **Definition 5.2.1** has two main implications: first, let us consider two effects $\psi, \varphi : A \to I$ in $\mathbf{C}_{\Sigma^2}$. By axiom (5.3):

$$\begin{aligned}
(\psi^\dagger \otimes \psi^*) \circ (\psi \otimes \psi_*) = (\varphi^\dagger \otimes \varphi^*) \circ (\varphi \otimes \varphi_*) &\iff (\top_{1,I} \otimes \top_{2,I}) \circ \psi = (\top_{1,I} \otimes \top_{2,I}) \circ \varphi \\
&\iff (1_I \otimes 1_I) \circ \psi = (1_I \otimes 1_1) \circ \varphi \\
&\iff \psi = \varphi \\
&\iff \psi^\dagger = \varphi^\dagger.
\end{aligned}$$

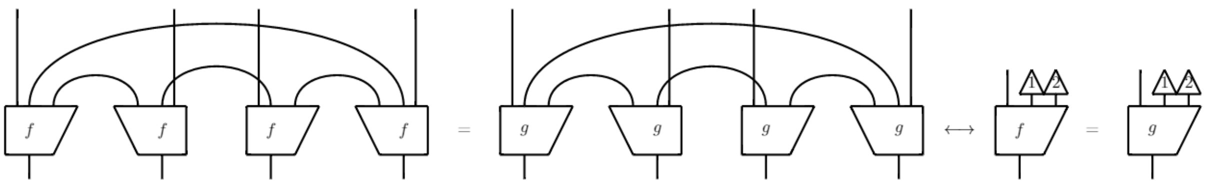
Graphically,



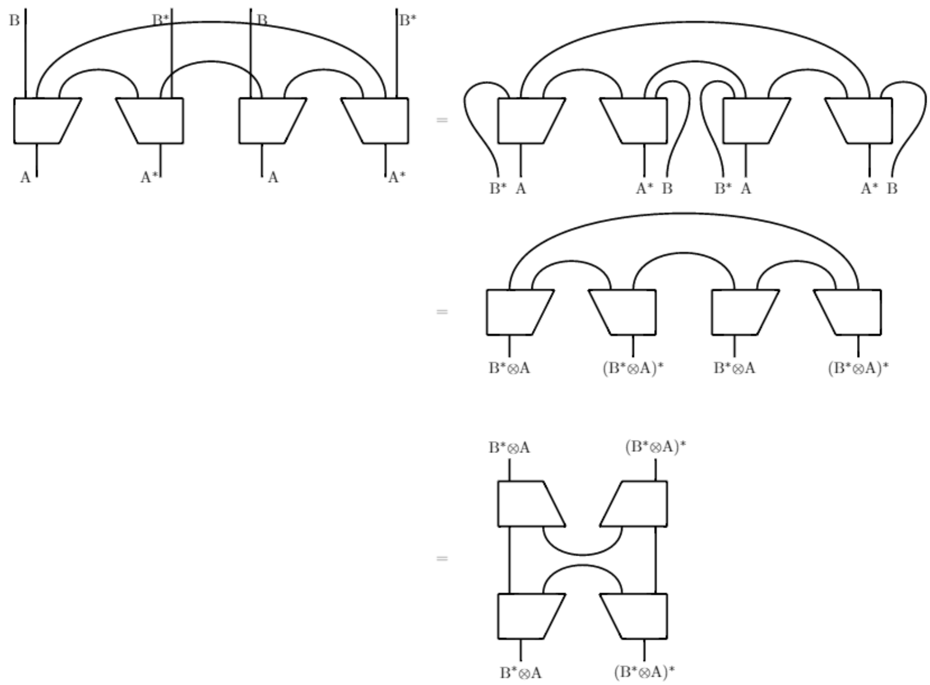
which is what we define to be the *squared-preparation-state agreement axiom.* We conclude:

$$\text{axiom (3)} \Rightarrow \text{squared-preparation-state agreement axiom} \tag{5.4}$$

Axiom (5.3) can also be stated as:



This stems from the fact that:



where the first and last equalities hold by properties of the entanglement structure of $\mathbf{C}_{\Sigma^2}$.

This reformulation of axiom (5.3) leads us to the second major implication mentioned before. But first, we introduce the notion of "*squared-positive*" maps:

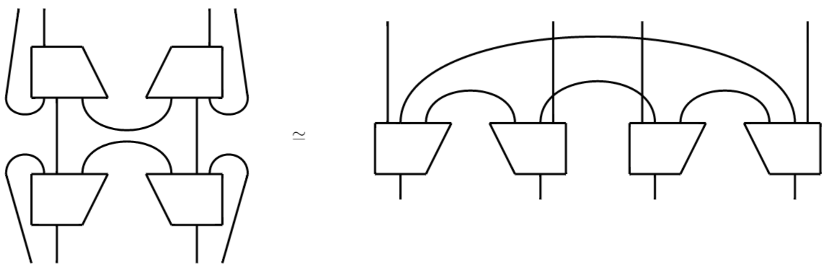
Definition 5.2.4. In a †-compact-closed category $\mathbf{C}$, a morphism $f : A \otimes A^* \to A \otimes A^*$ of $\mathbf{C}$ is *squared-positive* if and only if it decomposes as $f = (g^\dagger \otimes g^*) \circ (1_{C_{g,1}} \otimes \eta_{C^*_{g,2}} \otimes 1_{C_{g,1}}) \circ (1_{C_{g,1}} \otimes \epsilon_{C^*_{g,2}} \otimes 1_{C_{g,1}}) \circ (g \otimes g_*)$, for some morphism $g : A \to C_1 \otimes C_2$ of $\mathbf{C}$.

We now introduce the second important implication of axiom (5.3):

Proposition 5.2.1. *In a †-compact-closed category $\boldsymbol{C}$ with a squared-environment structure, axiom (5.3) gives rise to an isomorphism of categories*

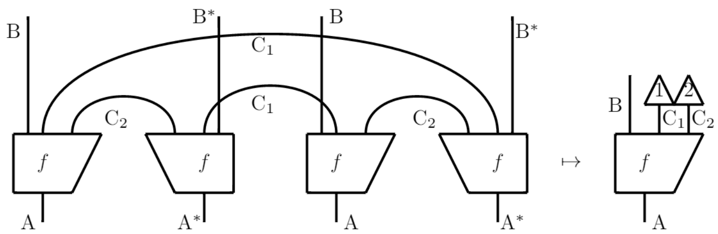
$$F_2 : \mathbf{C}_{\Sigma^2}^{pos^2} \simeq \mathbf{C},$$

where $\boldsymbol{C}_{\Sigma^2}^{pos^2}$ is the homset of all squared-positive morphisms in $\boldsymbol{C}_{\Sigma^2}$, i.e. morphisms of the form:



Proof. Define F_2 as follows:

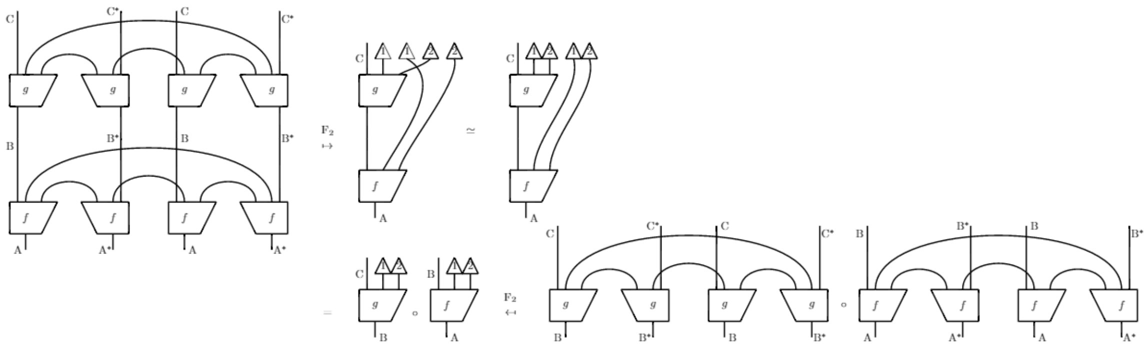
- F_2 maps objects to themselves
- F_2 maps morphisms $(1_B \otimes 1_{B^*} \otimes \epsilon_{C_1^*} \otimes 1_B \otimes 1_{B^*}) \circ (1_B \otimes \sigma_{C_1,B^*} \otimes \sigma_{B,C_1^*} \otimes 1_B) \circ (1_B \otimes 1_{C_1} \otimes 1_{B^*} \otimes \epsilon_{C_1} \otimes 1_B \otimes 1_{C_1^*} \otimes 1_{B^*}) \circ (1_B \otimes 1_{C_1} \otimes \epsilon_{C_2^*} \otimes \sigma_{C_1^*,B^*} \otimes \sigma_{B,C_1} \otimes \epsilon_{C_2^*} \otimes 1_{C_1^*} \otimes 1_{B^*}) \circ (f \otimes f_* \otimes f \otimes f_*)$ in $\mathbf{C}_{\Sigma^2}^{pos^2}(A \otimes A^* \otimes A \otimes A^*, B \otimes B^* \otimes B \otimes B^*)$ to $(1_B \otimes \top_{1,C_1} \otimes \top_{2,C_2}) \circ f$ in $\mathbf{C}(A, B)$, or graphically:



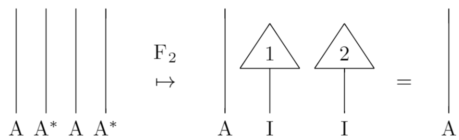
The forward direction of axiom (5.3) ensures that F_2 assigns a unique interpretation to a given morphism in the domain category. Therefore, F_2 is well-defined.

Functoriality is shown as before:

- $F_2(g \circ f) = F_2(g) \circ F_2(f)$



- $F_2(id_A) = id_{F_2(A)} = id_A$



It remains to show that F_2 is an isomorphism.
The one-to-one correspondence between objects is trivially implied by the fact that F_2 is identity-on-objects.
Fullness and faithfulness of F_2 is shown as before: by the backward direction of axiom (5.3), every morphism $F_{2_{A,B}} : \mathbf{C}^{pos^2}_{\Sigma^2}(A \otimes A^* \otimes A \otimes A^*, B \otimes B^* \otimes B \otimes B^*) \to \mathbf{C}(A, B)$ is injective, which shows that F_2 is *faithful*. By the *squared-purifiability axiom*, every morphism f in $\mathbf{C}$ is purifiable, i.e. there exists a pure morphism g in $\mathbf{C}_{\Sigma^2}$ such that f is the result of "square-discarding" parts of the output of g. Therefore, every morphism $F_{A,B} : \mathbf{C}^{pos}_{\Sigma}(A \otimes A^* \otimes A \otimes A^*, B \otimes B^* \otimes B \otimes B^*) \to \mathbf{C}(A, B)$ is surjective, and F is *full*.

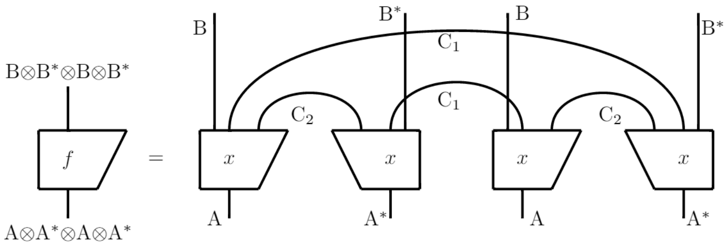
$\therefore$ axiom (3) $\Rightarrow \mathbf{C}^{pos^2}_{\Sigma^2} \simeq \mathbf{C}$. $\square$

5.2.2 $\mathbf{CP^2}$ maps and the $\mathbf{CPM^2}$-construction

Definition 5.2.5. (Completely squared-positive maps) Let A, B be objects in a †-compact-closed category $\mathbf{C}$. A morphism $f : A \otimes A^* \otimes A \otimes A^* \to B \otimes B^* \otimes B \otimes B^*$ of $\mathbf{C}$ is *completely squared-positive* or *CP*^{2} if there exist objects C_1, C_2 of $\mathbf{C}$ and a morphism $x : A \to B \otimes C_1 \otimes C_2$ such that:

$$f = (1_B \otimes 1_{B^*} \otimes \epsilon_{C_1^*} \otimes 1_B \otimes 1_{B^*}) \circ (1_B \otimes \sigma_{C_1,B^*} \otimes \sigma_{B,C_1^*} \otimes 1_B) \circ (1_B \otimes 1_{C_1} \otimes 1_{B^*} \otimes \epsilon_{C_1} \otimes 1_B \otimes 1_{C_1^*} \otimes 1_{B^*}) \circ (1_B \otimes 1_{C_1} \otimes \epsilon_{C_2^*} \otimes \sigma_{C_1^*,B^*} \otimes \sigma_{B,C_1} \otimes \epsilon_{C_2^*} \otimes 1_{C_1^*} \otimes 1_{B^*}) \circ (x \otimes x_* \otimes x \otimes x_*),$$

or graphically:

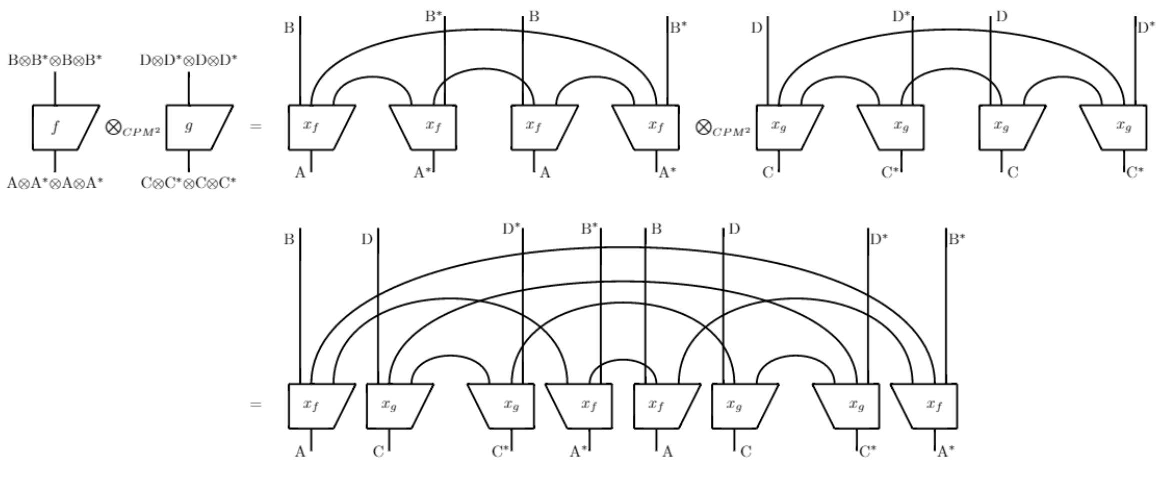


Proposition 5.2.2.

(a) The identity map $id_A : A \otimes A^* \otimes A \otimes A^* \to A \otimes A^* \otimes A \otimes A^*$ is CP^2.

(b) If $f : A \otimes A^* \otimes A \otimes A^* \to B \otimes B^* \otimes B \otimes B^*$ and $g : B \otimes B^* \otimes B \otimes B^* \to C \otimes C^* \otimes C \otimes C^*$ are CP^2*, then so is $g \circ f$.*

(c) Let $f : A \otimes A^* \otimes A \otimes A^* \to B \otimes B^* \otimes B \otimes B^*$ and $g : C \otimes C^* \otimes C \otimes C^* \to D \otimes D^* \otimes D \otimes D^*$ be CP^2*. Define the tensor product of CP*^{2} *morphisms as follows:*

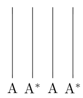


Then $f \otimes g$ is CP^2.

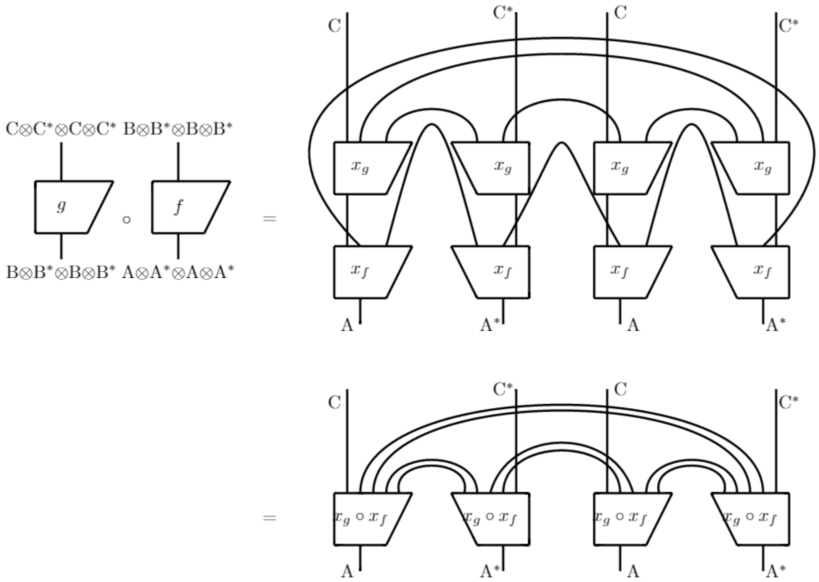
(d) If $f : A \to B$ is any morphism, then $f \otimes f_* \otimes f \otimes f_*$ is CP²

Proof. All the proofs use graphical manipulations:

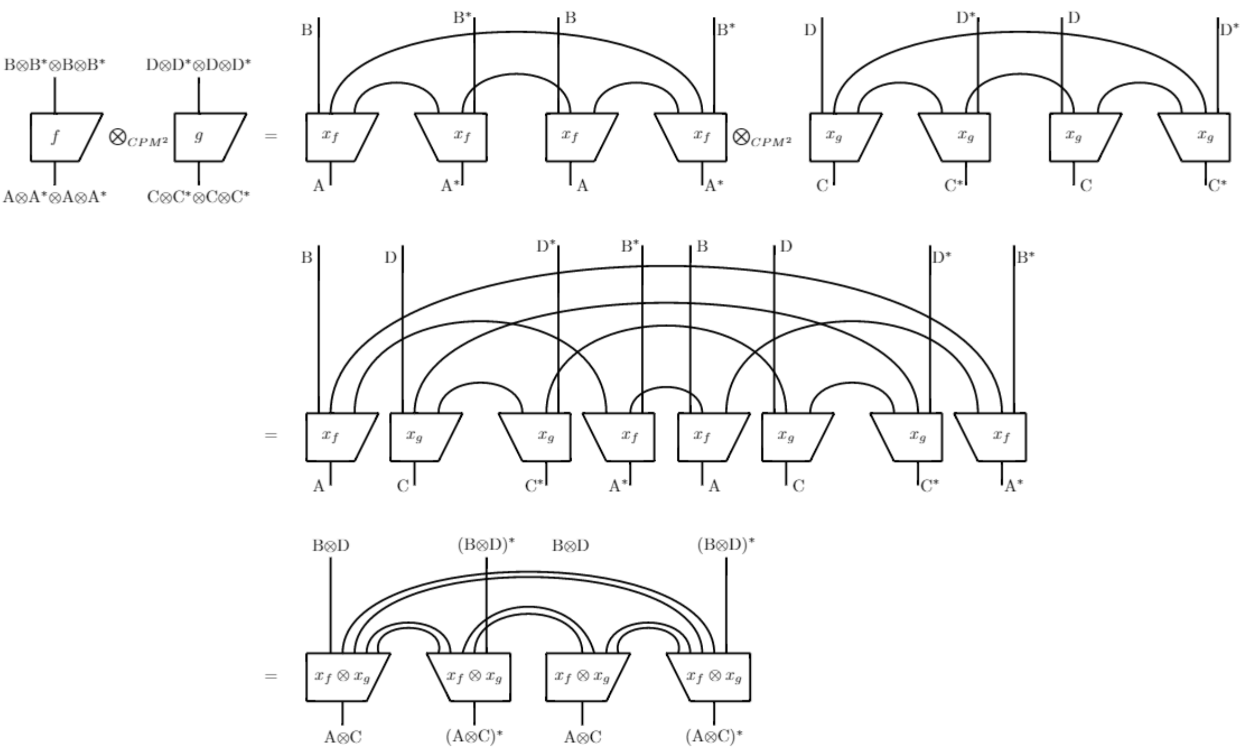
(a) Setting x to 1_A and C_1, C_2 to I in **Definition 5.2.5**, id_A is CP^2:



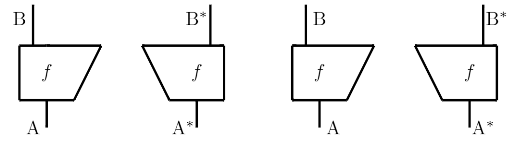
(b) Graphically,



(c) Graphically,



(d) Setting x to f and C_1, C_2 to I in **Definition 5.2.5**, $f \otimes f_* \otimes f \otimes f_*$ is CP^2:

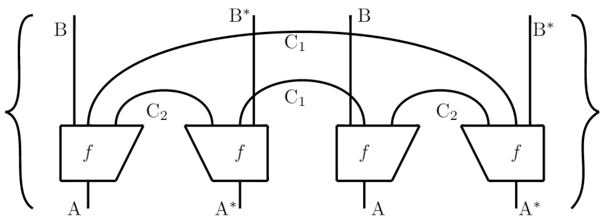


□

Definition 5.2.6. The CPM²-construction
Given a †-compact closed category C, define a new †-compact-closed category – refer to the note below – **CPM²(C)** as follows:

(i) The objects of **CPM²(C)** are the objects of **C**

(ii) The morphisms $A \to B$ of **CPM²(C)** are of the form:

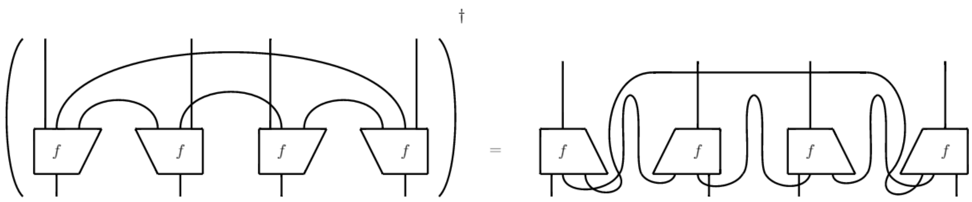


(iii) Identities are defined as in **Proposition 5.2.2***(a)*

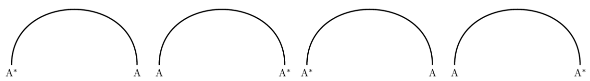
(iv) Composition is defined as in **Proposition 5.2.2***(b)*

(v) The tensor unit I and the tensor product of objects are inherited from **C**, and the tensor product of morphisms is defined as in **Proposition 5.2.2***(c)*

(vi) The dagger is defined the usual way:



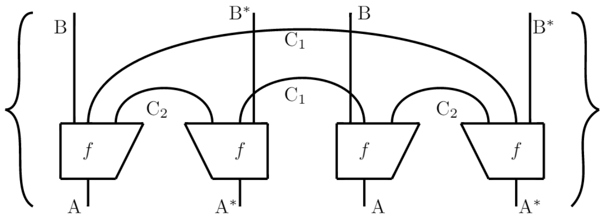
(vii) The cap $\epsilon_A : A^* \otimes_{CPM^2} A \to I$ is given by:



Note: $\mathbf{CPM^2(C)}$, where $\mathbf{C}$ is †-compact closed, is also †-compact closed since it is the result of iterating the CPM-construction on $\mathbf{CPM(C)}$, which is proved to be †-compact closed in [4]. The proof of **Theorem 6.1.1** in the next chapter offers a thorough explanation of this fact.

Theorem 5.2.1. *Let $\boldsymbol{C}$ be a †-compact-closed category. If $\boldsymbol{C}$ has a squared-environment structure, then $\boldsymbol{CPM(C_{\Sigma^2})} \simeq \boldsymbol{C}$, and $\boldsymbol{C_{\Sigma^2}}$ satisfies the squared-preparation-state agreement axiom.*

Proof. The sub-category $\mathbf{C}_{\Sigma^2}$ satisfies axiom (5.3). By (5.4) $\mathbf{C}_{\Sigma^2}$ also satisfies the squared-preparation-state agreement axiom.
Apply the CPM^2-construction to $\mathbf{C}_{\Sigma^2}$. The maps of $\mathbf{CPM(C_{\Sigma^2})}$ are the maps of the form:



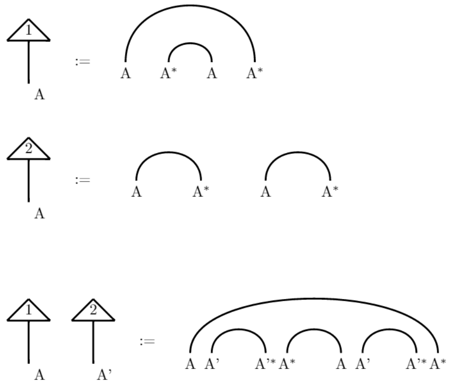
which are exactly the maps in $\mathbf{C}_{\Sigma_2}^{pos^2}$. By **Proposition 5.1.1**, $\mathbf{C}_{\Sigma^2}^{pos^2} \simeq \mathbf{C}$, and therefore $\mathbf{C}_{\Sigma^2}^{pos^2}(A \otimes A^* \otimes A \otimes A^*, B \otimes B^* \otimes B \otimes B^*) \simeq \mathbf{C}(A, B)$. Therefore:

$$\mathbf{CPM(C_{\Sigma^2})}(\mathrm{A,B}) \stackrel{\mathrm{def}}{=} \mathbf{C}_{\Sigma^2}^{pos^2}(A \otimes A^* \otimes A \otimes A^*, B \otimes B^* \otimes B \otimes B^*) \simeq \mathbf{C}(A, B)$$

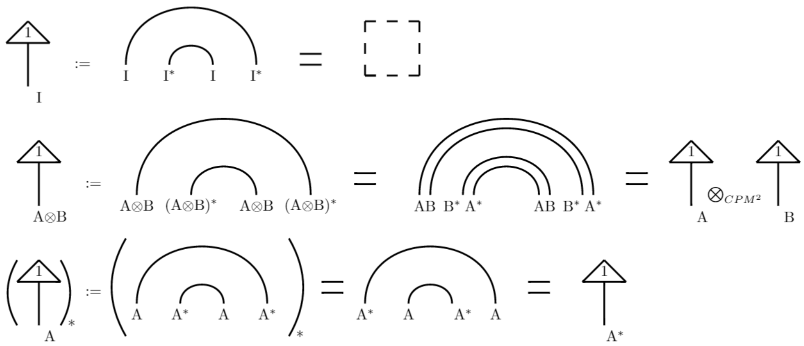
$\therefore \mathbf{CPM(C_{\Sigma^2})} \simeq \mathbf{C}$. □

Theorem 5.2.2. *Let $\boldsymbol{C}$ be a †-compact-closed category with an entanglement structure, and let $\boldsymbol{C}$ satisfy the squared-preparation state agreement axiom. Define the category $\boldsymbol{CPM^2(C)}$ according to* **Definition 5.2.6**. *Then $\boldsymbol{CPM^2(C)}$ has a squared-environment structure.*

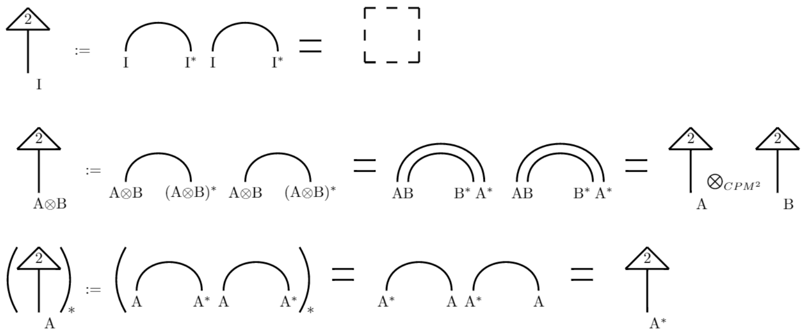
Proof. Define discardings 1 and 2 as follows:



Both $\top_{1,A}$ and $\top_{2,A}$ satisfy the required properties: first, $\top_{1,A}$:



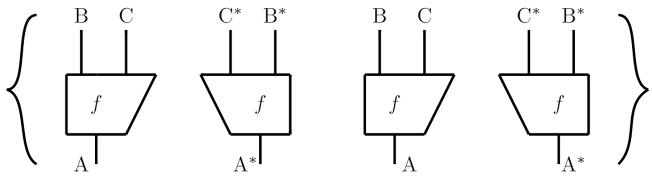
and $\top_{2,A}$:



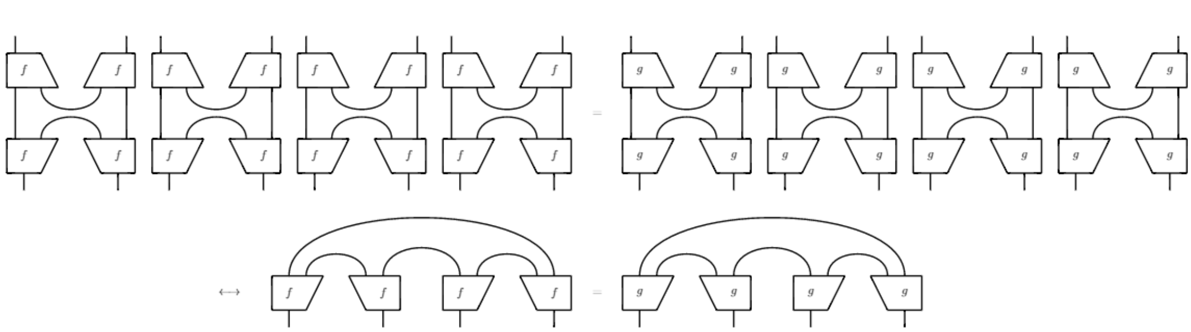
Consider now the all-objects-including sub-†-compact-closed category $\mathbf{CPM^2(C)}_{\Sigma^2}$ of pure morphisms, which carries an entanglement structure. **Proposition 5.2.2***(d)* defines an embedding

$$F_{CPM^2} : \mathbf{C} \hookrightarrow \mathbf{CPM^2(C)} :: f \mapsto f \otimes f_* \otimes f \otimes f_*$$

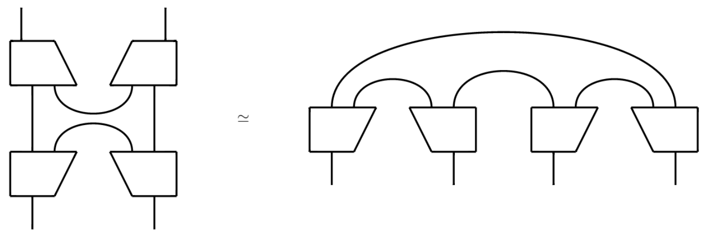
which maps pure morphisms in $\mathbf{C}$ to "squared-pure" morphisms in $\mathbf{CPM^2(C)}$, i.e. morphisms in $\mathbf{CPM^2(C)}_{\Sigma^2}$. These morphisms are of the form:



Let us show that $\mathbf{CPM(C)}_{\Sigma^2}$ satisfies axiom (5.3), graphically:



This follows immediately from the fact that:



by properties of the entanglement structure. Finally, every morphism in $\mathbf{CPM}^2(\mathbf{C})$ is squared-purifiable, by definition of morphisms in $\mathbf{CPM}^2(\mathbf{C})$ and of $\top_1, \top_2$ in $\mathbf{CPM}^2(\mathbf{C})$.

$\square$

Corollary 5.2.1. *By theorems* **5.2.1** *and* **5.2.2***, a †-compact-closed category* $\boldsymbol{C}$ *carrying a* $\top_1, \top_2$*-structure coincides with* $\mathbf{CPM}^2(\mathbf{C}_{\Sigma^2}) \simeq \mathbf{C}$*, and applying the CPM*^{2}*-construction to a †-compact-closed category* $\boldsymbol{C}$ *which satisfies the squared-preparation-state agreement axiom induces a* $\top_1, \top_2$*-structure on* $\boldsymbol{C}$*.*

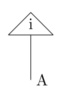
5.3 Axiomatisation of $\mathbf{CPM}^n(\mathbf{C})$

In this part, I generalise the findings of **Part 5.2** and define the notions of *n-environment structure* and *CPM*^{n}*-construction*, as well as the rules that govern them.

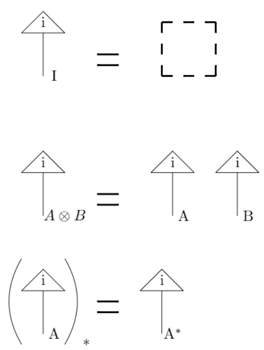
5.3.1 n-environment structure and implications

Definition 5.3.1. A $\top_1, ..., \top_n$- structure or *n-environment structure* on a †-compact-closed category C consists of:

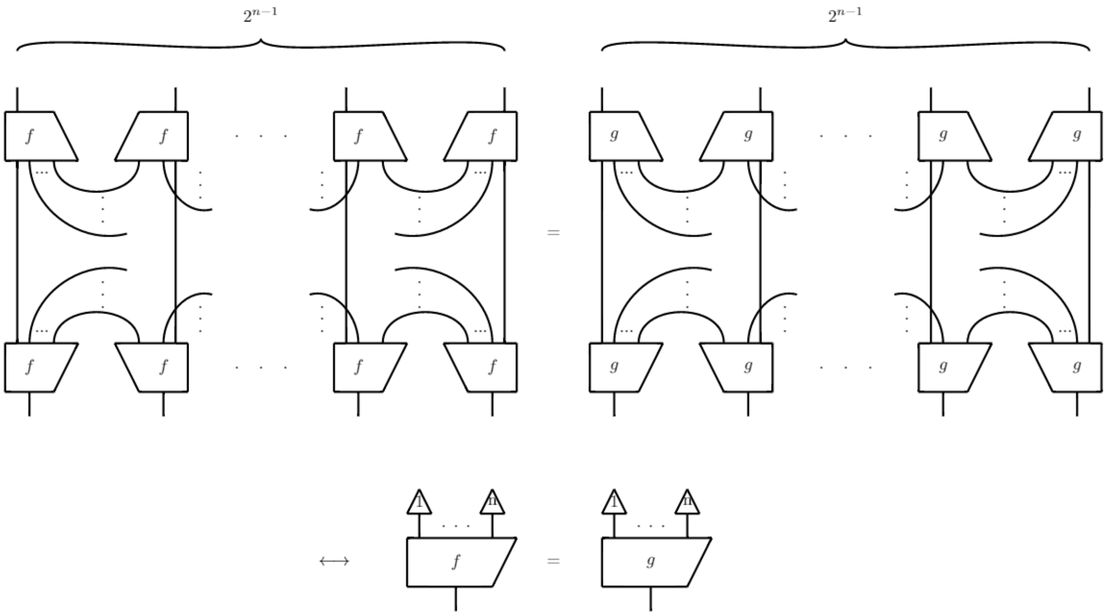
(i) n designated effects $\top_{i,A} : A \to I$, $i \in \{1, ..., n\}$ for each object A of $\mathbf{C}$, called *discarding-i*, and depicted as:



The discarding effects satisfy the same properties as the maximally mixed state in the environment structure definition of **Part 1**, namely:

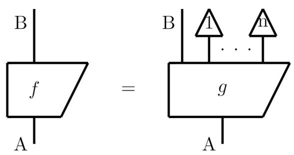


(ii) an all-objects-including sub-†-compact-closed category $\mathbf{C}_{\Sigma^n}$ of *pure morphisms*, which carries an entanglement structure, and which is such that for all morphisms f, g of $\mathbf{C}_{\Sigma^n}$ where $dom(f) = C_{f,1} \otimes ... \otimes C_{f,n}$, and $dom(g) = C_{g,1} \otimes ... \otimes C_{g,n}$:



(5.5)

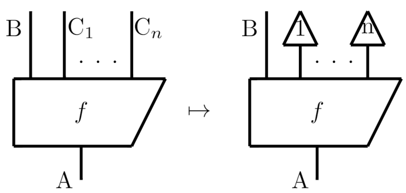
(iii) the **n-purifiability axiom**: for every morphism $f : A \to B$ in $\mathbf{C}$, there exists a morphism $g : A \to B \otimes C_1 \otimes ... \otimes C_n$ in $\mathbf{C}_{\Sigma^n}$ such that:



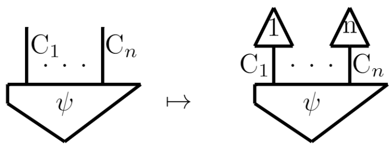
The concepts pertaining to "*n-purifiability*" are defined below.

Definition 5.3.2. In a †-compact-closed category $\mathbf{C}$ with an n-environment structure:

- the *n-partial internal trace* is the map $tr_{A,B}^{C_1,...,C_n} : \mathbf{C}(A, B \otimes C_1 \otimes ... \otimes C_n) \to \mathbf{C}(A, B) ::$ $f \mapsto (1_B \otimes \top_{1,C_1} \otimes ... \otimes \top_{n,C_n}) \circ f$, for objects $A, B, C_1, ..., C_n$ and any arrow $f :$ $A \to B \otimes C_1 \otimes ... \otimes C_n$ in $\mathbf{C}$. Graphically:



- the *n-full internal trace* is the map $tr^{C_1,...,C_n} : \mathbf{C}(I, C_1 \otimes ... \otimes C_n) \to \mathbf{C}(I, I) :: \psi \mapsto (\top_{1,C_1} \otimes ... \otimes \top_{n,C_n}) \circ \psi$, for objects $C_1, ..., C_n$ and a state $\psi : I \to C_1 \otimes ... \otimes C_n$ in $\mathbf{C}$. Graphically:



Definition 5.3.3. In a †-compact closed category $\mathbf{C}$ with an n-environment structure, define an *n-purification* of an operation $f : A \to B$ to be a pure operation $g : A \to B \otimes C_1 \otimes ... \otimes C_n$ which is such that $f = tr_{A,B}^{C_1,...C_n}(g)$. f is said to be *n-purifiable.*

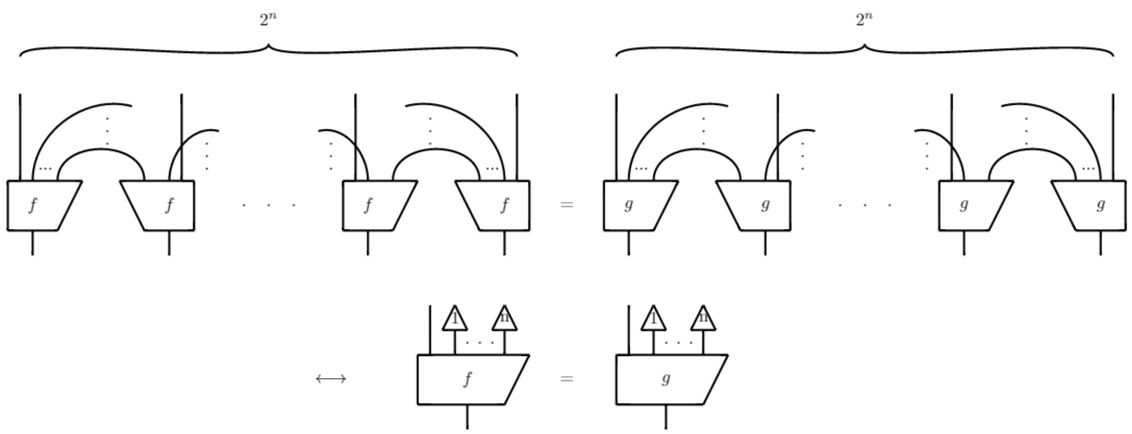
As in the previous parts, axiom (5.5) of **Definition 5.3.1** has two main implications: let us consider two effects $\psi, \varphi : A \to I$ in $\mathbf{C}_{\Sigma^n}$. We define $A^{\otimes^n}$ to be $A \otimes ... \otimes A$ n times. By axiom (5.5):

$$
\begin{aligned}
&(\psi^\dagger \otimes \psi^*)^{\otimes^{2^{n-1}}} \circ (\psi \otimes \psi_*)^{\otimes^{2^{n-1}}} = (\varphi^\dagger \otimes \varphi^*)^{\otimes^{2^{n-1}}} \circ (\varphi \otimes \varphi_*)^{\otimes^{2^{n-1}}} \\
\Longleftrightarrow &(\psi^\dagger \otimes \psi^*) \circ (\psi \otimes \psi_*) = (\varphi^\dagger \otimes \varphi^*) \circ (\varphi \otimes \varphi_*) \\
\Longleftrightarrow &(\top_{1,I} \otimes ... \otimes \top_{n,I}) \circ \psi = (\top_{1,I} \otimes ... \otimes \top_{n,I}) \circ \varphi \\
\Longleftrightarrow &(1_I \otimes ... \otimes 1_I) \circ \psi = (1_I \otimes ... \otimes 1_1) \circ \varphi \\
\Longleftrightarrow &\psi = \varphi \\
\Longleftrightarrow &\psi^\dagger = \varphi^\dagger.
\end{aligned}
$$

which is exactly what we defined to be the *squared-preparation-state agreement axiom.* We conclude:

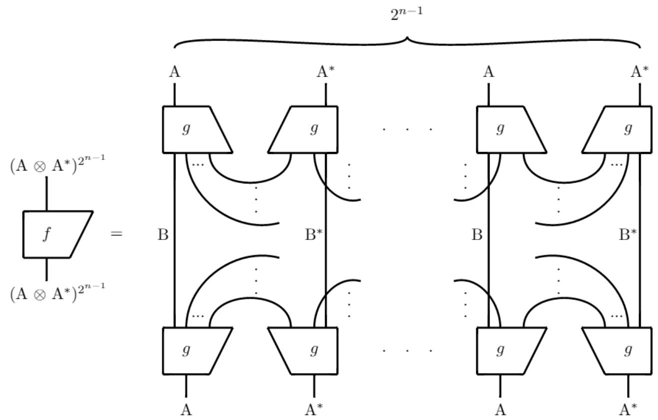
$$\text{axiom (5)} \Rightarrow \text{squared-preparation-state agreement axiom} \tag{5.6}$$

Following the same diagrammatical manipulations as before, axiom (5.5) can also be stated as:



From this reformulation of axiom (5.5), we state its second major implication. But first, we introduce the notion of "*n-positivity*":

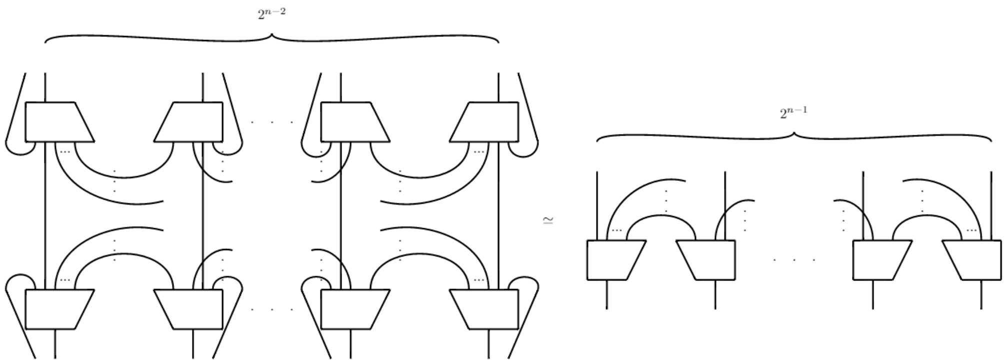
Definition 5.3.4. In a †-compact-closed category **C**, a morphism $f : (A \otimes A^*)^{\otimes 2^{n-1}} \to (A \otimes A^*)^{\otimes 2^{n-1}}$ of **C** is *n-positive* if and only if there exists a morphism $g : A \to C_1 \otimes ... \otimes C_n$ of **C** such that f decomposes as:



Proposition 5.3.1. *In a †-compact-closed category* **C** *with an n-environment structure, axiom (5.5) gives rise to an isomorphism of categories*

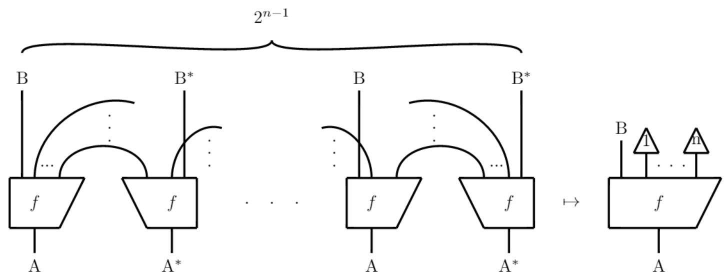
$$F_n : \mathbf{C}_{\Sigma^n}^{pos^n} \simeq \mathbf{C},$$

where $\boldsymbol{C}_{\Sigma^n}^{pos^n}$ *is the homset of all n-positive morphisms in* $\boldsymbol{C}_{\Sigma^n}$*, i.e. morphisms of the form*:



Proof. Define F_n following the definition given in the proof of **Proposition 5.2.1**:

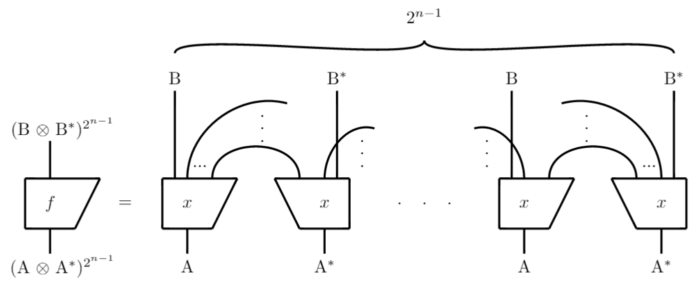
- F_n maps objects to themselves
- F_n maps morphisms in $\mathbf{C}_{\Sigma^n}^{pos^n}((A \otimes A^*)^{2^{n-1}}, (B \otimes B^*)^{2n-1})$ to maps in $\mathbf{C}(A, B)$ as follows:



Well-definedness, functoriality, and one-to-one correspondence of objects and morphisms are proved as in the proof of **Proposition 5.2.1**. □

5.3.2 $\mathbf{CP}^n$ maps and the $\mathbf{CPM}^n$-construction

Definition 3.2.1 (Completely n-positive maps) Let A, B be objects in a †-compact-closed category $\mathbf{C}$. A morphism $f : (A \otimes A^*)^{2^{n-1}} \to (B \otimes B^*)^{2^{n-1}}$ of $\mathbf{C}$ is *completely n-positive* or *CP*^{2} if there exist objects C_1, C_2 of $\mathbf{C}$ and a morphism $x : A \to B \otimes C_1 \otimes C_2$ such that:



Proposition 5.3.2.

(a) The identity map $id_A : (A \otimes A^*)^{2^{n-1}} \to (A \otimes A^*)^{2^{n-1}}$ is CP^n.

(b) If $f : (A \otimes A^*)^{2^{n-1}} \to (B \otimes B^*)^{2^{n-1}}$ and $g : (B \otimes B^*)^{2^{n-1}} \to (C \otimes C^*)$ are CP^n, then so is their composition, $g \circ f$, defined the usual way.

(c) Let $f : (A \otimes A^*)^{2^{n-1}} \to (B \otimes B^*)$ and $g : (C \otimes C^*) \to (D \otimes D^*)$ be CP^n. Define the tensor product of CP^n morphisms as it was defined for CP^2 morphisms. Then $f \otimes g$ is CP^n.

(d) If $f : A \to B$ is any morphism, then $(f \otimes f_*)^{2^{n-1}}$ is CP^n

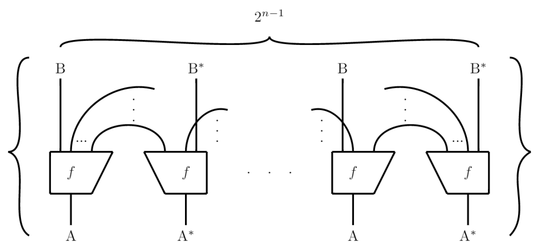
Proof. All the proofs use the same graphical manipulations as the proofs of **Proposition 5.2.2**. □

Definition 5.3.5. The $\mathbf{CPM}^n$-construction

Given a †-compact closed category C, define a new †-compact-closed category $\mathbf{CPM}^n\mathbf{(C)}$ as follows:

(i) The objects of $\mathbf{CPM}^n\mathbf{(C)}$ are the objects of $\mathbf{C}$

(ii) The morphisms $A \to B$ of $\mathbf{CPM}^n\mathbf{(C)}$ are of the form:



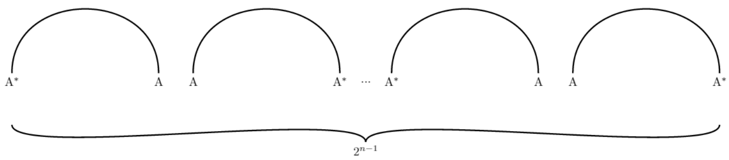
(iii) Identities are defined as in **Proposition 5.3.2***(a)*

(iv) Composition is defined the usual way.

(v) The tensor unit I and the tensor product of objects are inherited from $\mathbf{C}$, and the tensor product of morphisms is defined as in **Proposition 5.3.2***(c)*

(vi) The dagger is defined the usual way.

(vii) The cap $\epsilon_A : A^* \otimes_{CPM^n} A \to I$ is given by:



Note: $\mathbf{CPM}^n\mathbf{(C)}$, where $\mathbf{C}$ is †-compact closed, is also †-compact closed since it is the result of iterating the CPM-construction on a †-compact closed category.

Theorem 5.3.1. *Let $\boldsymbol{C}$ be a †-compact-closed category. If $\boldsymbol{C}$ has an n-environment structure, then $\boldsymbol{CPM(C_{\Sigma^n})} \simeq \boldsymbol{C}$, and $\boldsymbol{C}_{\Sigma^n}$ satisfies the squared-preparation-state agreement axiom.*

Proof. The proof uses **Proposition 5.1.1** the same way the proof of **Theorem 5.2.1** uses **Proposition 5.2.1**. □

Theorem 5.3.2. *Let $\boldsymbol{C}$ be a †-compact-closed category with an entanglement structure, and let $\boldsymbol{C}$ satisfy the squared-preparation state agreement axiom. Define the category $\boldsymbol{CPM^n(C)}$ according to* **Definition 5.3.5**. *Then $\boldsymbol{CPM^n(C)}$ has an n-environment structure.*

Proof. Define the n discardings intuitively by following the model of the proof of **Theorem 5.2.2**. Checking that every $\top_i$ satisfies the required properties follows easily from graphical manipulations.

Consider now the all-objects-including sub-†-compact-closed category $\mathbf{CPM}^n\mathbf{(C)}_{\Sigma^n}$ of pure morphisms, which carries an entanglement structure. **Proposition 5.3.2***(d)* defines an embedding

$$F_{CPM^n} : \mathbf{C} \hookrightarrow \mathbf{CPM}^n\mathbf{(C)} :: f \mapsto (f \otimes f_*)^{2^{n-1}}$$

which maps pure morphisms in $\mathbf{C}$ to "n-pure" morphisms in $\mathbf{CPM}^n\mathbf{(C)}$, i.e. morphisms in $\mathbf{CPM}^n\mathbf{(C)}_{\Sigma^n}$. We show that $\mathbf{CPM(C)}_{\Sigma n}$ satisfies axiom (5.5) the same way $\mathbf{CPM(C)}_{\Sigma^2}$ was showed to satisfy axiom (5.3) in the proof of **Theorem 5.2.2**.
Finally, every morphism in $\mathbf{CPM}^n\mathbf{(C)}$ is n-purifiable, by definition of morphisms in $\mathbf{CPM}^n\mathbf{(C)}$ and of $\top_i$, $i \in \{1, ..., n\}$ in $\mathbf{CPM}^n\mathbf{(C)}$. □

Corollary 5.3.1. *By theorems* **5.3.1** *and* **5.3.2***, a †-compact-closed category* ***C*** *carrying an n-environment structure coincides with* $\mathbf{CPM}^n\mathbf{(C}_{\Sigma^n}\mathbf{)} \simeq \mathbf{C}$*, and applying the CPM^n-construction to a †-compact-closed category* ***C*** *which satisfies the squared-preparation-state agreement axiom induces an n-environment structure on* ***C***.

The methods in [7] were therefore revised and adapted in order to axiomatise the CPM^2-construction. These methods were also generalised to allow the axiomatisation of the CPM^n-construction. It is worthy to point out that one of the notable uses of $\mathbf{CPM}^n\mathbf{(C)}$ is accommodating an increasing number of features of language.

In the next chapter, we will dive into the details of the compact closure of $\mathbf{CPM}^2\mathbf{(C)}$, more precisely that of $\mathbf{CPM}^2\mathbf{(FHilb)}$, and investigate the Frobenius algebras of these categories.

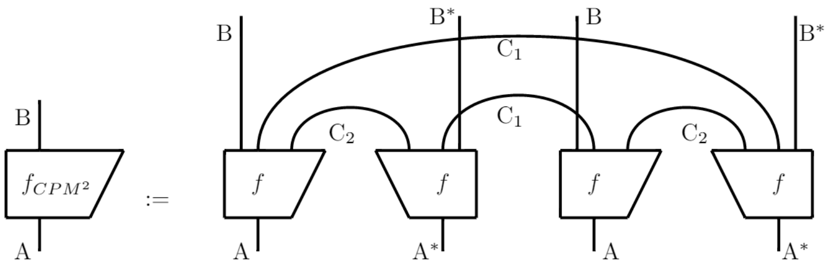
Chapter 6

Further analysis of the CPM^2-construction framework

In the previous chapter, I offered a definition and axiomatisation of the CPM^2-construction. Taking a closer look at the CPM^2-construction framework, grammatical reductions should be carried over to the new category. For this, the CPM^2-construction needs to preserve compact closure. It also needs to have a dagger structure in order to have a measure of similarity between words and between sentences. Finally, it needs to be equipped with a †-special commutative Frobenius algebra so as to account for relational types.

6.1 $\mathrm{CPM}^2(\mathrm{C})$ as a †-compact closed category

We denote by F_{CPM^2} the identity-on-objects functor with domain category $\mathbf{C}$ and codomain category $\mathbf{CPM^2(C)}$ which maps morphisms $f : A \to B$ of $\mathbf{C}$ to morphisms $f_{CPM^2} : A \to B$ of $\mathbf{CPM^2(C)}$, where f_{CPM^2} is defined as follows:



C_1 and C_2 are the ancillary systems of f_{CPM^2}. F_{CPM^2} being essentially the result of applying the CPM-construction twice, C_1 comes from the first application of the CPM-construction and is called the *first ancillary system*, and C_2 comes from the second application and is called the *second ancillary system*.

Note: When there is absolutely no confusion, we will denote f_{CPM^2} simply by f.

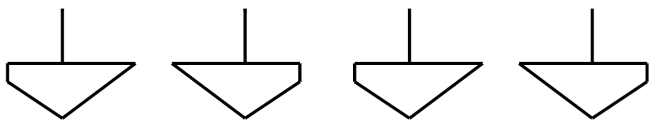
Theorem 6.1.1. *Let $\mathbf{C}$ be a †-compact closed category. The functor $F_{CPM^2} : \mathbf{C} \to$ $\mathbf{CPM^2(C)}$ preserves the †-compact closed structure.*

Proof. **Theorem 4.3.1** gives rise to a †-compact-closed-structure-preserving functor F_{CPM}: $\mathbf{C} \to \mathbf{CPM(C)}$. The functor F_{CPM^2} is therefore given by the successive application of F_{CPM}, that is: $F_{CPM^2} = F_{CPM}F_{CPM}$. By **Theorem 4.3.1**, given a †-compact closed category $\mathbf{C}$, the category $\mathbf{CPM(C)}$ is †-compact closed. By the second application of **Theorem 4.3.1**, the category $\mathbf{CPM^2(C)}$ is †-compact closed. □

We now proceed with a few definitions pertaining to purifiability, which echo the ones given in **Section 5.2**. We also give the linguistic interpretation of these concepts.

Definition 6.1.1. (1-, 2-, and 1,2-pure morphisms) Let $f : A \to B$ be a morphism in $\mathbf{CPM^2(C)}$. f is *1-pure* if $C_1 = I$; f is *2-pure* if $C_2 = I$; and f is *1,2-pure* if $C_1 = C_2 = I$.

From a linguistic perspective, if we suppose that the first application of the CPM-construction accounts for ambiguity, and that the second application accounts for entailment, then a 1-pure state corresponds to an unambiguous, general word, a 2-pure state corresponds to an ambiguous, non-general word, and a 1,2-pure state corresponds to an unambiguous, non-general word. The graphical representations of 1- and 2-pure states are given in **Section 4.4**, and the representation of a 1,2-pure state is given below:



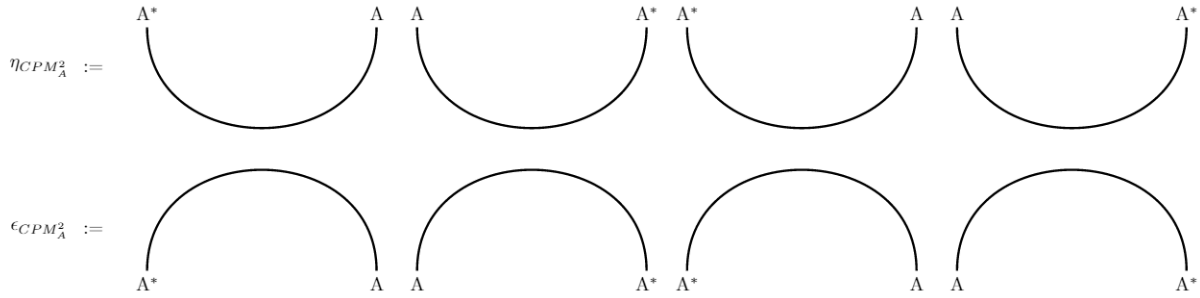
Let us delve into the †-compact closed structure of $\mathbf{CPM^2(C)}$. The objects of $\mathbf{CPM^2(C)}$ are the objects of $\mathbf{C}$, and its morphisms are CP^2 maps. The identities, composition, and tensor unit and product $\otimes_{CPM^2}$ are defined in **Proposition 5.2.2**, and the dagger is defined in **Definition 5.2.6**.

It remains to define the compact-closure maps η and ϵ. Recall that a †-compact closed category is symmetric, among other things, and therefore $\eta^l = \eta^r = \eta$, and $\epsilon^l = \epsilon^r = \epsilon$. The compact-closure maps in $\mathbf{CPM^2(C)}$ are defined by taking the image of the original compact-closure maps in $\mathbf{C}$ by F_{CPM^2}:

$$\eta_{CPM^2_A} : I \to A^* \otimes_{CPM^2} A := \eta_A \otimes \eta_{A^*} \otimes \eta_A \otimes \eta_{A^*}$$

$$\epsilon_{CPM^2_A} : A^* \otimes_{CPM^2} A \to I := \epsilon_A \otimes \epsilon_{A^*} \otimes \epsilon_A \otimes \epsilon_{A^*}$$

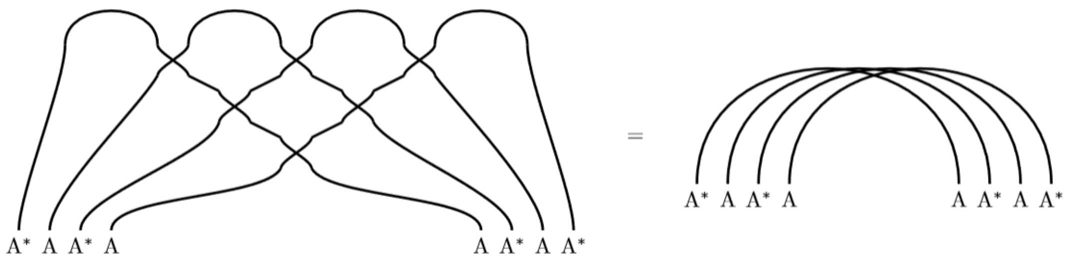
Graphically:



Note that since any monoidal category is equivalent to a strict one,

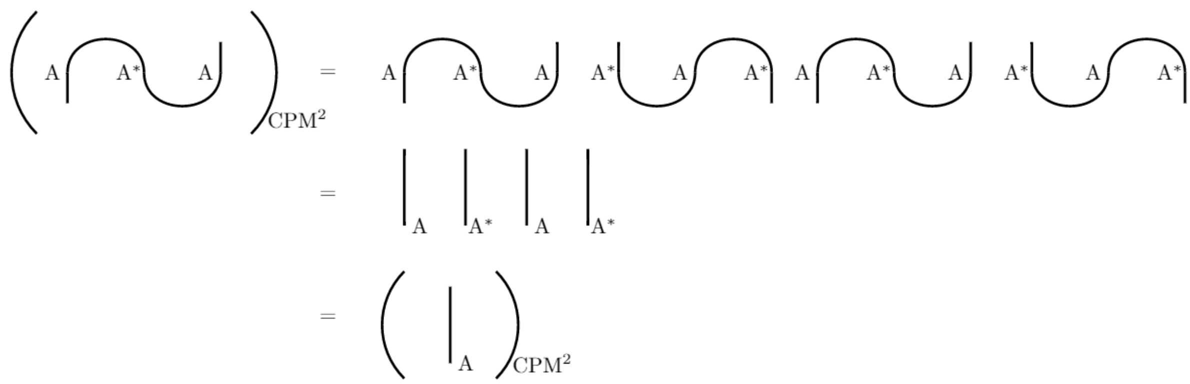
$$(A^* \otimes A) \otimes (A \otimes A^*) \otimes (A^* \otimes A) \otimes (A \otimes A^*) \simeq (A^* \otimes A \otimes A^* \otimes A) \otimes (A \otimes A^* \otimes A \otimes A^*)$$

and we can represent the cap as:



This representation of the cap is more convenient in diagrams corresponding to grammatical reductions in $\mathbf{CPM^2(FHilb)}$. An equivalent representation of the cup in $\mathbf{CPM^2(C)}$ can be found similarly.

These compact-closure maps do satisfy the required yanking equations. The derivation for one of the equations is shown below. The second can be proved similarly:



Maps in $\mathbf{CPM^2(FHilb)}$. Words with two features of language – homonymy and entailment – are represented as states in the †-compact closed category $\mathbf{CPM^2(FHilb)}$. As in

FHilb, $V^* \cong V$. The concrete compact-closure maps, given by the first representation, are defined as follows:

$$\eta_{CPM_V^2} : I \to V^{\otimes^8} :: 1 \mapsto (\sum_i n_i \otimes n_i) \otimes (\sum_j n_j \otimes n_j) \otimes (\sum_k n_k \otimes n_k) \otimes (\sum_l n_l \otimes n_l)$$
$$\epsilon_{CPM_V^2} : V^{\otimes^8} \to I :: (v_i \otimes w_i) \otimes (v_j \otimes w_j) \otimes (v_k \otimes w_k) \otimes (v_l \otimes w_l) \mapsto \langle v_i|w_i\rangle \langle v_j|w_j\rangle \langle v_k|w_k\rangle \langle v_l|w_l\rangle$$

When given by the equivalent representation, the η and ϵ maps are defined as follows:

$$\eta_{CPM_V^2} : I \to V^{\otimes^8} :: 1 \mapsto \sum_{i,j,k,l} (n_i \otimes n_j \otimes n_k \otimes n_l) \otimes (n_i \otimes n_j \otimes n_k \otimes n_l)$$
$$\epsilon_{CPM_V^2} : V^{\otimes^8} \to I :: (v_i \otimes v_j \otimes v_k \otimes v_l) \otimes (w_i \otimes w_j \otimes w_k \otimes w_l) \mapsto \langle v_i|w_i\rangle \langle v_j|w_j\rangle \langle v_k|w_k\rangle \langle v_l|w_l\rangle$$

The yanking equation satisfaction is easy to show and follows straightforwardly from the derivations in **Subsection 3.2.2**.

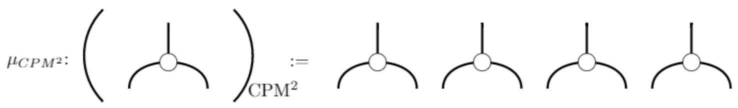
Note: When representing maps in **CPM(C)**, we often replace the "doubled" wires by a single thick wire, and "doubled" boxes by a single box with thick sides. We could do the same for maps in **CPM**^{2}**(C)**, by resorting to "extra thick" wires and boxes. However, with higher and higher levels of CPM-construction iterations, this representation quickly becomes bulky and impractical. This is why I chose to represent maps in **CPM**^{2}**(C)** between parentheses and with the subscript "CPM^2". Maps that are not confined between these special parentheses belong to the original category **C**.

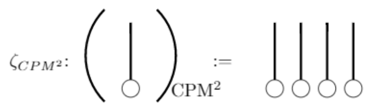
6.2 Frobenius algebras in CPM^2(C)

Let (μ, ζ) be an associative algebra in **C**, and (Δ, ι) a coassociative coalgebra in **C**. We define the maps:

$$\mu_{CPM^2} = F_{CPM^2}(\mu) \qquad \zeta_{CPM^2} = F_{CPM^2}(\zeta)$$

represented graphically by:

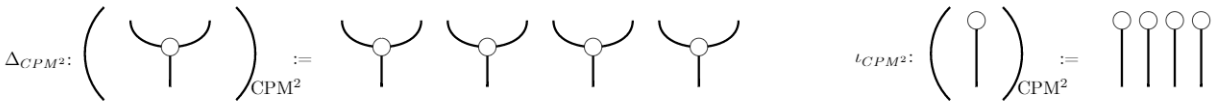




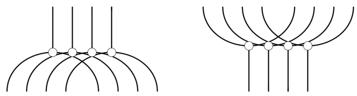
and the maps:

$$\Delta_{CPM^2} = F_{CPM^2}(\Delta) \qquad \iota_{CPM^2} = F_{CPM^2}(\iota)$$

represented graphically by:

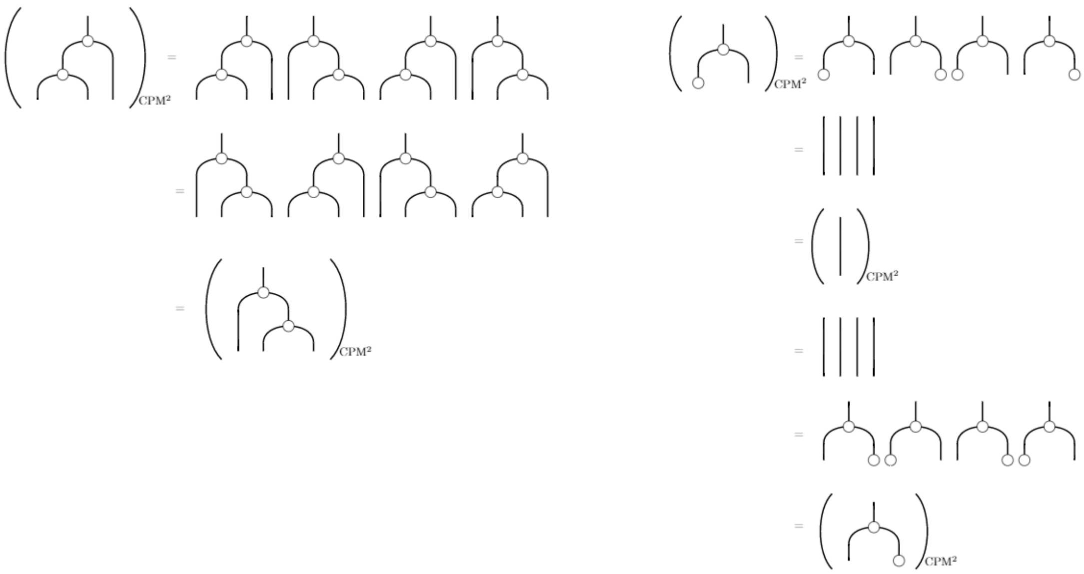


Note that μ_{CPM^2} and Δ_{CPM^2} have equivalent representations that are more practical in the representation of grammatical reductions, respectively:



Proposition 6.2.1. $(\mu_{CPM^2}, \zeta_{CPM^2})$ *is an associative algebra in* ***CPM²(C)***.

Proof. The proof is given graphically by:

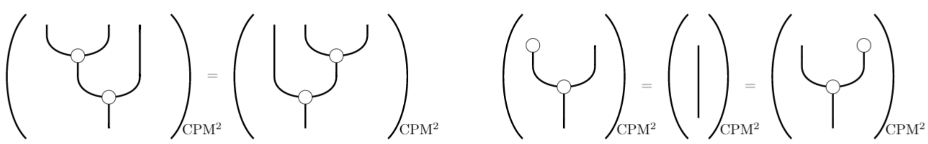


□

The proofs of the propositions below all follow straightforwardly from the equations that hold in **C**, and follow the model of the proof above.

Proposition 6.2.2. $(\Delta_{CPM^2}, \iota_{CPM^2})$ *is a coassociative coalgebra in* ***CPM²(C)***.

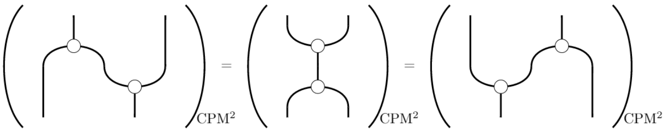
Proof. The following follows from the fact that (Δ, ι) is a coassociative coalgebra in **C**:



□

Proposition 6.2.3. *Let* $(\mu, \zeta, \Delta, \iota)$ *be a Frobenius algebra in* ***C***. *Then,* $(\mu_{CPM^2}, \zeta_{CPM^2}, \Delta_{CPM^2}, \iota_{CPM^2})$ *is a Frobenius algebra in* ***CPM²(C)***.

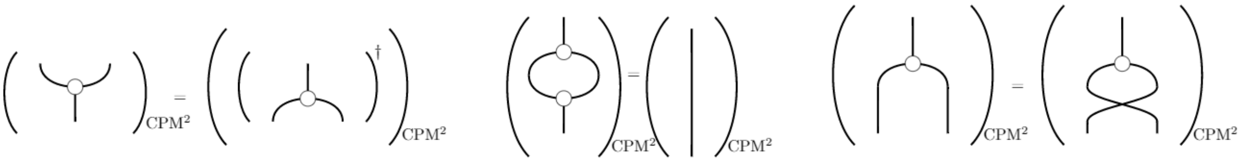
Proof. The following follows from the fact that $(\mu, \zeta, \Delta, \iota)$ is a Frobenius algebra in **C**:



□

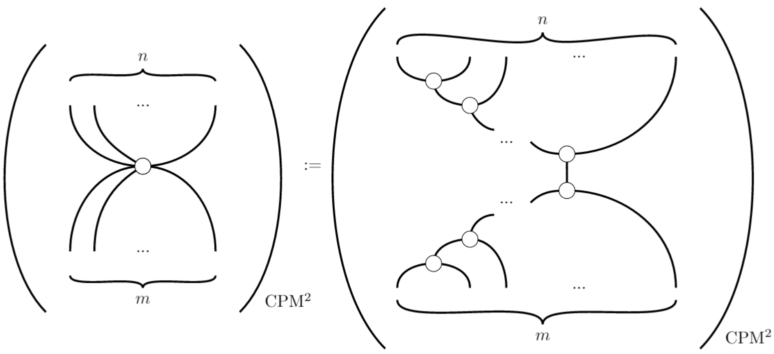
Proposition 6.2.4. *Let $(\mu, \zeta, \Delta, \iota)$ be a †-special commutative Frobenius algebra in* ***C****. Then, $(\mu_{CPM^2}, \zeta_{CPM^2}, \Delta_{CPM^2}, \iota_{CPM^2})$ is a †-special commutative Frobenius algebra in* ***CPM²(C)***.

Proof. The following follows from the fact that $(\mu, \zeta, \Delta, \iota)$ is a †-special commutative Frobenius algebra in **C**:

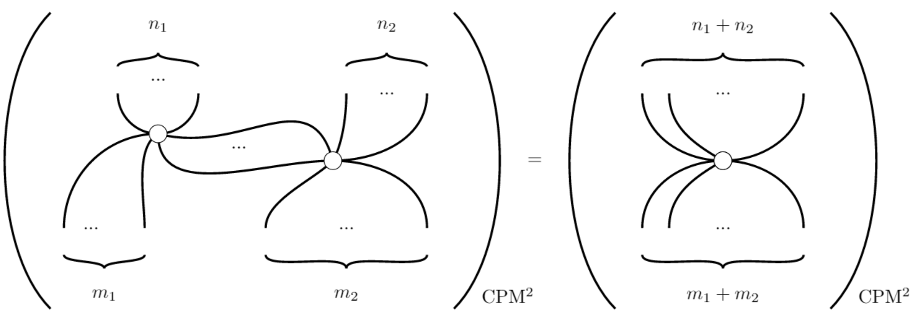


□

We define the spiders in $\mathbf{CPM}^2(\mathbf{C})$ based on the structure provided by a Frobenius algebra:



These spiders compose in the expected way:



†-special commutative Frobenius algebras in $\mathbf{CPM}^2(\mathbf{FHilb})$. Let V be any Hilbert space, with fixed orthonormal basis $\{|i\rangle\}_i$. The maps of the †-special commutative algebra in $\mathbf{CPM}^2(\mathbf{C})$ are given by:

$$
\begin{aligned}
\Delta_{CPM^2} &:: |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes |l\rangle \mapsto (|i\rangle \otimes |i\rangle) \otimes (|j\rangle \otimes |j\rangle) \otimes (|k\rangle \otimes |k\rangle) \otimes (|l\rangle \otimes |l\rangle) \\
\iota_{CPM^2} &:: |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes |l\rangle \mapsto 1 \\
\mu_{CPM^2} &:: (|i_1\rangle \otimes |j_1\rangle) \otimes (|i_2\rangle \otimes |j_2\rangle) \otimes (|i_3\rangle \otimes |j_3\rangle) \otimes (|i_4\rangle \otimes |j_4\rangle) \\
&\quad \mapsto \delta_{i_1 j_1}\delta_{i_2 j_2}\delta_{i_3 j_3}\delta_{i_4 j_4} |i_1\rangle \otimes |i_2\rangle \otimes |i_3\rangle \otimes |i_4\rangle \\
\zeta_{CPM^2} &:: 1 \mapsto \sum_i |i\rangle \otimes \sum_j |j\rangle \otimes \sum_k |k\rangle \otimes \sum_l |l\rangle
\end{aligned}
$$

We note that $|i\rangle$, $|j\rangle$, $|k\rangle$, and $|l\rangle$ are vectors in the same basis.

The definitions of Δ_{CPM^2} and μ_{CPM^2} in the equivalent representation are given by:

$$
\begin{aligned}
\Delta_{CPM^2} &:: |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes |l\rangle \mapsto (|i\rangle \otimes |j\rangle \otimes |k\rangle \otimes |l\rangle) \otimes (|i\rangle \otimes |j\rangle \otimes |k\rangle \otimes |l\rangle) \\
\mu_{CPM^2} &:: (|i_1\rangle \otimes |i_2\rangle \otimes |i_3\rangle \otimes |i_4\rangle) \otimes (|j_1\rangle \otimes |j_2\rangle \otimes |j_3\rangle \otimes |j_4\rangle) \\
&\quad \mapsto \delta_{i_1 j_1}\delta_{i_2 j_2}\delta_{i_3 j_3}\delta_{i_4 j_4} |i_1\rangle \otimes |i_2\rangle \otimes |i_3\rangle \otimes |i_4\rangle
\end{aligned}
$$

Thus, the CPM^2-construction framework preserves †-compact closed structure, and possesses †-special commutative Frobenius algebras derived from the †-special commutative Frobenius algebras in $\mathbf{C}$. This framework is therefore adequate for representing ambiguous and general words, the grammatical relations between them, and measuring their similarity, while also accounting for relational types.

In the next chapter, we will look deeper into the characterisation of density matrices and their counterparts in $\mathbf{CPM}^2(\mathbf{C})$.

Chapter 7

Density matrices and double-density matrices

Chapters **5** and **6** introduced and axiomatised the CPM^2-construction, and showed that this framework is adequate for our purposes. This chapter investigates closely density matrices, their properties and their characterisation, and introduces the notion of *double-density matrices*, states of $\mathbf{CPM}^2(\mathbf{C})$ that will be at the center of modelling two features of language in the next chapter.

7.1 Characterisation of density matrices

Chapter 4 introduces density operators, mathematical tools used to express *mixed states*, or probability distributions over ensembles of pure states.

Definition 7.1.1. Given a set $\{|\varphi_m\rangle\}$ of pure, not necessarily orthogonal quantum states, and $\{p_m\}$ a probability distribution over them, define the density operator for this system by:

$$\rho \equiv \sum_m p_m |\varphi_m\rangle \langle\varphi_m|$$

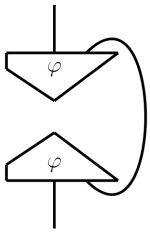
These operators, whose formalism was first introduced in 1927 independently by von Neumann [33] and Landau [29], present useful characteristics [22]:

- They are self-adjoint, or *Hermitian*: $\rho = \rho^\dagger$
- They are positive-semidefinite: $\forall |\psi\rangle, \langle\psi| \rho |\psi\rangle \geq 0$
- Their trace is one.

Density matrices $\rho : A \to A$ are represented by morphisms

$$(1_A \otimes \epsilon_{C^*}) \circ (\varphi \otimes 1_{C^*}) \circ (\varphi^\dagger \otimes 1_{C^*}) \circ (1_A \otimes \eta_{C^*}) \tag{7.1}$$

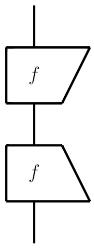
where $\varphi : I \to A \otimes C$ is a bipartite state, or graphically as follows:



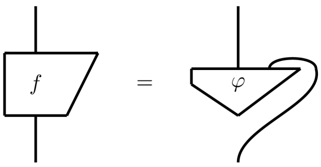
They can equivalently be represented by morphisms

$$f \circ f^{\dagger} \tag{7.2}$$

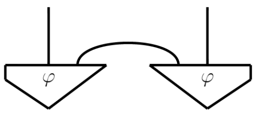
graphically:



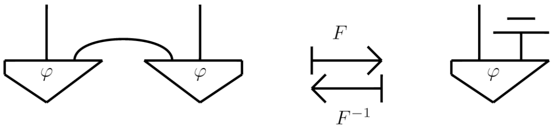
by taking



The *name* of a density matrix, denoted by $\lceil\rho\rceil$, is obtained by process-state duality. The resulting $\otimes$-positive bipartite state is represented graphically by:



When there is absolutely no confusion, we will refer to names of density matrices simply as density matrices. Notice that:

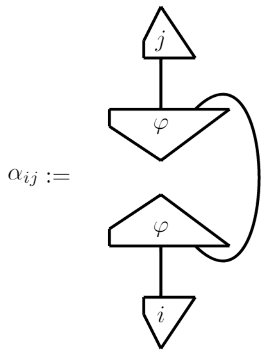


where $F : \mathbf{C}_{\Sigma}^{pos} \simeq \mathbf{C}$ is the isomorphism of categories defined in **Proposition 5.1.1**.

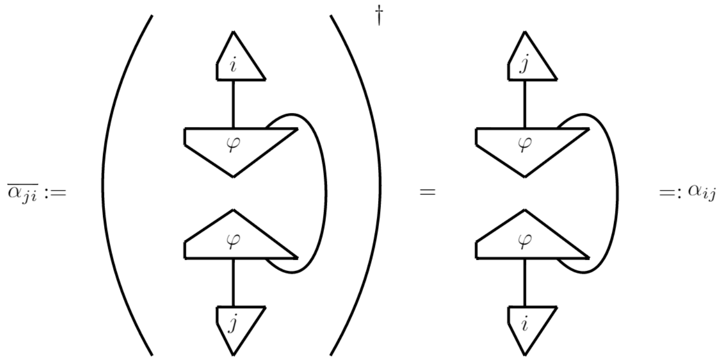
The main aim of this section is to show that any map of the form (7.1) – or equivalently (7.2) – satisfies the conditions of hermiticity and positive-semidefiniteness. The next section will deal with the reverse direction.

Proposition 7.1.1. *An operator ρ of the form (7.1) is Hermitian.*

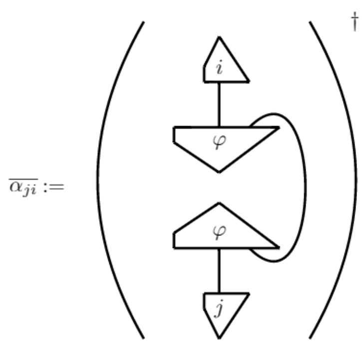
Proof. A Hermitian matrix satisfies the property that $\forall i, j,\ \alpha_{ij} = \overline{\alpha_{ji}}$ where α_{ij} is the entry in the ith row and jth column, and $\overline{\alpha_{ji}}$ is the complex conjugate of α_{ji}. Graphically, the entry α_{ij} of ρ is given by:



We prove graphically that $\alpha_{ij} = \overline{\alpha_{ji}}$:



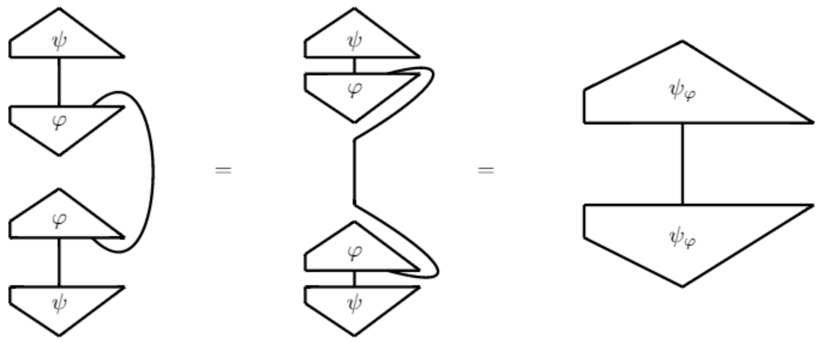
where



since the adjoint of an operator is its conjugate transpose. $\square$

Proposition 7.1.2. *An operator ρ of the form (7.1) is positive-semidefinite.*

Proof. Consider the state $\psi : I \to A$. We show graphically that $\langle\psi|\,\rho\,|\psi\rangle \geq 0$:



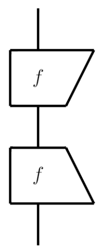
And since $\langle\psi_\varphi|\psi_\varphi\rangle$ is a non-negative scalar $\forall\psi_\varphi$, $\langle\psi|\,\rho\,|\psi\rangle \geq 0$. □

7.2 Representation of positive-semidefinite Hermitian Operators

In this section, we will show that any positive-semidefinite operator with the property

$$\forall i, j, \alpha_{ij} = \overline{\alpha_{ji}}$$

can be represented as:



First, let us lay down useful properties of Hermitian operators.

Proposition 7.2.1. *Let ρ be a Hermitian operator over a finite dimensional space. The eigenvalues of ρ are real.*

Proof. Let $\vec{x}$ be an eigenvector with eigenvalue λ. Without loss of generality, $\vec{x}$ can be rescaled to have length one. Then:

$$\begin{aligned}
\lambda &= \lambda \langle \vec{x} | \vec{x} \rangle && (\langle \vec{x} | \vec{x} \rangle = 1) \\
&= \langle \vec{x} | \lambda \vec{x} \rangle && \text{by linearity} \\
&= \langle \vec{x} | \rho \vec{x} \rangle && (\rho \vec{x} = \lambda \vec{x}) \\
&= \langle \rho^{\dagger} \vec{x} | \vec{x} \rangle && \text{by definition of adjoints} \\
&= \langle \rho \vec{x} | \vec{x} \rangle && \text{by hermiticity of } \rho \\
&= \langle \lambda \vec{x} | \vec{x} \rangle && (\rho \vec{x} = \lambda \vec{x}) \\
&= \overline{\lambda} \langle \vec{x} | \vec{x} \rangle && \text{by anti-linearity} \\
&= \overline{\lambda} && (\langle \vec{x} | \vec{x} \rangle = 1)
\end{aligned}$$

And $\lambda = \overline{\lambda} \iff \lambda \in \mathbb{R}$. □

Proposition 7.2.2. *Let ρ be a Hermitian operator over a finite dimensional space. Eigenvectors of ρ corresponding to different eigenvalues are orthogonal.*

Proof. Let $\vec{x}$ and $\vec{y}$ be eigenvectors of ρ corresponding respectively to eigenvalues λ_1 and λ_2, such that $\lambda_1 \neq \lambda_2$.

$$\begin{aligned}
\lambda_1 \langle \vec{x} | \vec{y} \rangle &= \langle \overline{\lambda_1} \vec{x} | \vec{y} \rangle && \text{(by anti-linearity)} \\
&= \langle \lambda_1 \vec{x} | \vec{y} \rangle && \lambda \in \mathbb{R} \text{ by } \textbf{Proposition 7.2.1} \\
&= \langle \rho \vec{x} | \vec{y} \rangle && (\rho \vec{x} = \lambda_1 \vec{x}) \\
&= \langle \rho^{\dagger} \vec{x} | \vec{y} \rangle && \text{by hermiticity of } \rho \\
&= \langle \vec{x} | \rho \vec{y} \rangle && \text{by definition of adjoints} \\
&= \langle \vec{x} | \lambda_2 \vec{y} \rangle && (\rho \vec{y} = \lambda_2 \vec{y}) \\
&= \lambda_2 \langle \vec{x} | \vec{y} \rangle && \text{by linearity}
\end{aligned}$$

Therefore, $\lambda_1 \langle \vec{x} | \vec{y} \rangle = \lambda_2 \langle \vec{x} | \vec{y} \rangle \implies (\lambda_1 - \lambda_2) \langle \vec{x} | \vec{y} \rangle = 0 \implies \langle \vec{x} | \vec{y} \rangle = 0$ (since $\lambda_1 \neq \lambda_2$). □

Similar proofs of propositions **7.2.1** and **7.2.2** can be found on page 182 of [3], and on page 195 of [38].

The following is a useful property of positive-semidefinite Hermitian operators:

Proposition 7.2.3. *The eigenvalues of a positive-semidefinite Hermitian operator are real and non-negative.*

Proof. Let ρ be a positive-semidefinite Hermitian operator. By **Proposition 7.2.1**, the eigenvalues of ρ are real.
Let $|x\rangle$ be an eigenvector of ρ corresponding to some eigenvalue λ. By positive-semidefiniteness:

$$\langle x | \rho | x \rangle \geq 0$$

This is equivalent to:

$$\langle x|\,\lambda\,|x\rangle \geq 0 \iff \langle x|x\rangle\,\lambda \geq 0 \iff \lambda \geq 0$$

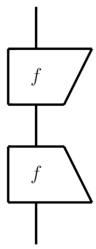
where the last equivalence holds because $\langle x|x\rangle$ is positive. $\square$

We now prove the main theorem of this section:

Theorem 7.2.1. *Let ρ be a positive-semidefinite operator over a finite-dimensional space, with the property*

$$\forall i, j, \alpha_{ij} = \overline{\alpha_{ji}}.$$

Then, there exists a morphism f such that $\rho = f \circ f^\dagger$, or graphically:



Proof. ρ is a Hermitian operator. Therefore, it is diagonalisable (a proof of this statement can be found on page 197 of [38]): there exists an invertible matrix U and a diagonal matrix D such that $U^{-1}\rho U = D$.

We claim that U is the matrix formed by the eigenvectors of ρ, and that the entries of D are the eigenvalues of ρ. Let us write $U = (\vec{c_1}, \vec{c_2}, ..., \vec{c_n})$, where $\{\vec{c_i}\}_i$ are the column vectors of U, and denote by λ_i the entry in the i^{th} row and i^{th} column of D. Then:

$$U^{-1}\rho U = D \iff \rho U = UD \iff \rho\vec{c_i} = \lambda_i\vec{c_i}, i \in \{1, 2, ..., n\}$$

The column vectors of U are therefore the right eigenvectors of ρ, and the entries of D are the corresponding eigenvalues of ρ. By a similar argument, the row vectors of U^{-1} are the left eigenvectors of ρ.

By **Proposition 7.2.2**, $\{c_i\}_i$ are orthogonal. We can always pick an orthonormal basis of eigenvectors. U is then a matrix whose columns are orthonormal, and is therefore unitary, i.e.

$$UU^\dagger = I, \quad U^\dagger U = I, \quad U^{-1} = U^\dagger$$

D is the diagonal matrix whose entries are the eigenvalues of ρ. By **Proposition 7.2.3**, the entries of D are real and non-negative. We denote by $D^{1/2}$ the square root of D. The entry in the i^{th} row and i^{th} column of $D^{1/2}$ is $\sqrt{\lambda_i}$. $D^{1/2}$ being a diagonal matrix with real non-negative entries, it is easy to see that it is Hermitian.

We now claim that $(UD^{1/2}U^{-1})$ is a root of ρ. Recall that R is a root of ρ iff $RR = \rho$.

$$\begin{aligned}(UD^{1/2}U^{-1})(UD^{1/2}U^{-1}) &= UD^{1/2}(U^{-1}U)D^{1/2}U^{-1} \\ &= UD^{1/2}D^{1/2}U^{-1} \\ &= UDU^{-1} \\ &= \rho\end{aligned}$$

Therefore, $(UD^{1/2}U^{-1})$ is a root of ρ.

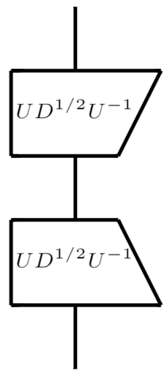
The last step is to show that $(UD^{1/2}U^{-1})$ is Hermitian. In fact,

$$(UD^{1/2}U^{-1})^\dagger = (U^{-1})^\dagger(D^{1/2})^\dagger U^\dagger = UD^{1/2}U^{-1}$$

where the last equality holds by hermiticity of $D^{1/2}$ and because $U^{-1} = U^\dagger$. Therefore,

$$\rho = (UD^{1/2}U^{-1})(UD^{1/2}U^{-1})^\dagger$$

and

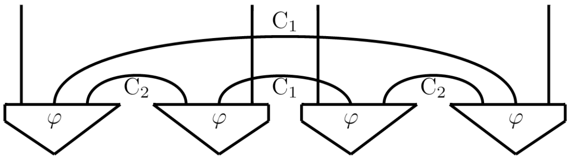


thus concluding our proof. $\square$

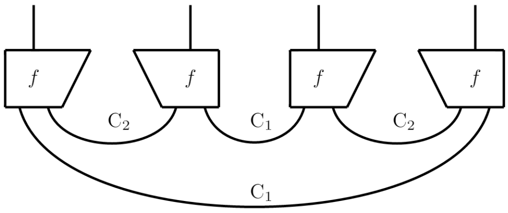
7.3 Characterisation of double-density matrices

Just as density matrices are the mathematical counterpart to mixed states, we define *double-density matrices* to be the mathematical tool used to represent *doubly mixed states*, where a doubly mixed state embodies a *two-level probability distribution* over an ensemble of 1,2-pure states. By two-level probability distribution, we mean that there are two levels of mixing, one corresponding to the first ancillary system, and one corresponding to the second.

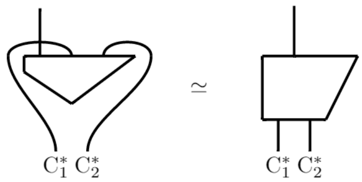
The name of a density matrix is a state in **CPM(C)**. We define a *double-density matrix* to be a state in $\mathbf{CPM}^2(\mathbf{C})$, with graphical representation:



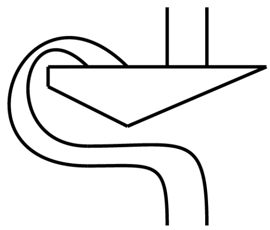
or equivalently:



using the following isomorphism:



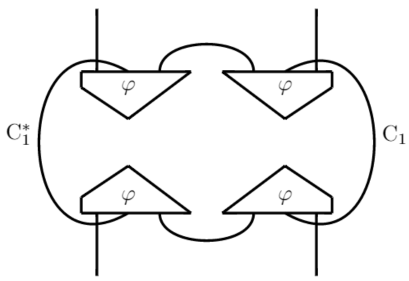
We define two transformations on a double-density matrix that enable us to recover two density matrices. *Density matrix-1* is the result of the following transformation:



and is the morphism $\rho_1 : A \otimes A^* \to A \otimes A^*$ defined by:

$$(\epsilon_{C_1} \otimes 1_A \otimes 1_{A^*} \otimes \epsilon_{C_1}) \circ (1_{C_1^*} \otimes \sigma_{A,C_1} \otimes \epsilon_{C_2^*} \otimes \sigma_{C_1^*,A^*} \otimes 1_{C_1}) \circ (1_{C_1^*} \otimes \varphi \otimes \varphi_* \otimes 1_{C_1})$$
$$\circ (1_{C_1^*} \otimes \varphi^\dagger \otimes \varphi^* \otimes 1_{C_1}) \circ (1_{C_1^*} \otimes \sigma_{C_1,A} \otimes \eta_{C_2^*} \otimes \sigma_{A^*,C_1^*} \otimes 1_{C_1}) \circ (\eta_{C_1} \otimes 1_A \otimes 1_A^* \otimes \eta_{C_1}) \tag{7.3}$$

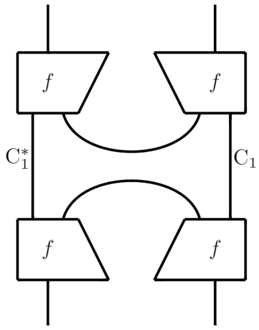
where $\varphi : I \to A \otimes C_1 \otimes C_2$ is a tripartite state, or graphically as follows:



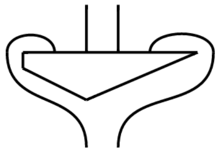
Density matrix-1 can equivalently be represented by a morphism

$$(f \otimes f_*) \circ (1_{C_1^*} \otimes \eta_{C_2} \otimes 1_{C_1}) \circ (1_{C_1^*} \otimes \epsilon_{C_2} \otimes 1_{C_1}) \otimes (f^\dagger \otimes f^*) \tag{7.4}$$

where f is a morphism with domain $C_1^* \otimes C_2^*$ and codomain A, or graphically:



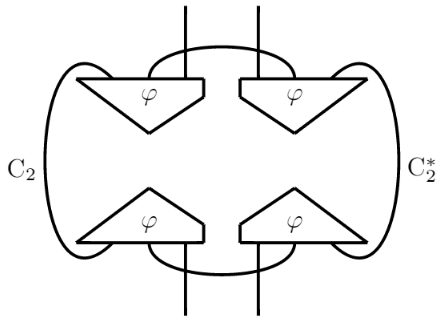
Density matrix-2 is the result of the following transformation:



and is the morphism $\rho_2 : A^* \otimes A \to A^* \otimes A$ defined by:

$$\begin{gathered}(1_{A^*} \otimes \epsilon_{C_1} \otimes 1_A) \circ (\epsilon_{C_2^*} \otimes \sigma_{C_1^*,A^*} \otimes \sigma_{A,C_1} \otimes \epsilon_{C_2^*}) \circ (1_{C_2} \otimes \varphi_* \otimes \varphi \otimes 1_{C_2^*}) \\ \circ(1_{C_2} \otimes \varphi^* \otimes \varphi^\dagger \otimes 1_{C_2^*}) \circ (\eta_{C_2^*} \otimes \sigma_{A^*,C_1^*} \otimes \sigma_{C_1,A} \otimes \eta_{C_2^*}) \circ (1_{A^*} \otimes \eta_{C_1} \otimes 1_A)\end{gathered} \tag{7.5}$$

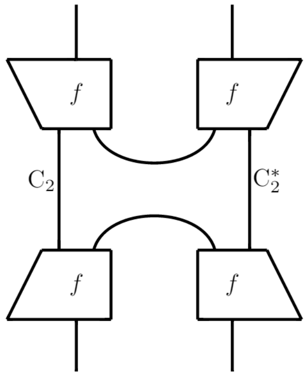
noindent where $\varphi : I \to A \otimes C_1 \otimes C_2$ is a tripartite state, or graphically as follows:



Density matrix-2 can equivalently be represented by a morphism

$$(f \otimes f_*) \circ (1_{C_2} \otimes \eta_{C_1^*} \otimes 1_{C_2^*}) \circ (1_{C_2} \otimes \epsilon_{C_1^*} \otimes 1_{C_2^*}) \otimes (f^\dagger \otimes f^*) \tag{7.6}$$

where f is a morphism with domain $C_1^* \otimes C_2^*$ and codomain A, or graphically:

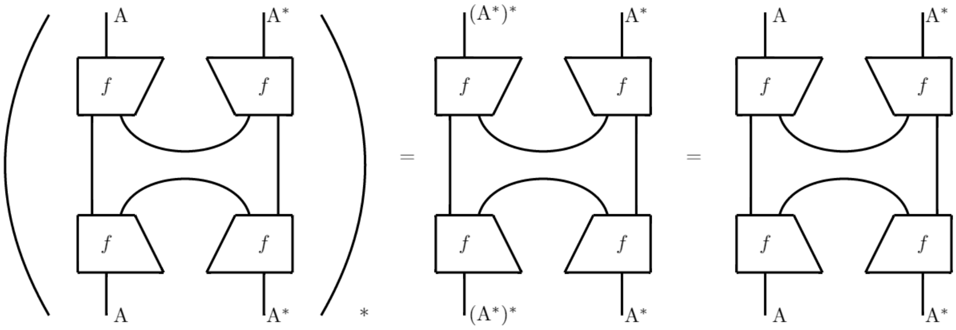


Morphisms of the form (7.3) – equivalently (7.4) – and (7.5) – equivalently (7.6) – are clearly density matrices. Therefore, they satisfy the conditions of hermiticity and positive-semidefiniteness. We say that a double-density matrix is *doubly-Hermitian* and *doubly-positive-semidefinite*. These two notions are defined below:

Definition 7.3.1. (double-hermiticity). A morphism $\varphi : I \to A \otimes A^* \otimes A \otimes A^*$ is *doubly-Hermitian* if the morphisms $\varphi_1 : A \otimes A^* \to A \otimes A^*$, and $\varphi_2 : A^* \otimes A \to A^* \otimes A$, where φ_1 and φ_2 are the result of applying transformations 1 and 2 on φ, are Hermitian.

Definition 7.3.2. (double-positive-semidefiniteness). A morphism $\varphi : I \to A \otimes A^* \otimes A \otimes A^*$ is *doubly-positive-semidefinite* if the morphisms $\varphi_1 : A \otimes A^* \to A \otimes A^*$, and $\varphi_2 : A^* \otimes A \to A^* \otimes A$, where φ_1 and φ_2 are the result of applying transformations 1 and 2 on φ, are positive-semidefinite.

Density matrices-1 and -2 also satisfy another property: they are *diagrammatically self-conjugate*. In fact,



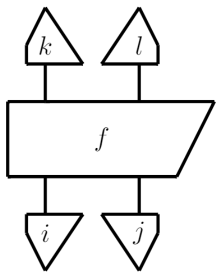
Self-conjugation of density matrix-2 is shown similarly. Double-density matrices satisfy the condition of *double-self-conjugation*, defined below:

Definition 7.3.3. (double-self-conjugation). A morphism $\varphi : I \to A \otimes A^* \otimes A \otimes A^*$ is *doubly-self-conjugate* if the morphisms $\varphi_1 : A \otimes A^* \to A \otimes A^*$, and $\varphi_2 : A^* \otimes A \to A^* \otimes A$, where φ_1 and φ_2 are the result of applying transformations 1 and 2 on φ, are (diagrammatically) self-conjugate.

Let us now characterise the entries of density matrices-1 and -2. We consider a morphism $f : A \otimes B \to C \otimes D$, where A, B, C, and D are spanned by $\mathcal{B}_A = \{a_i\}_i$, $\mathcal{B}_B = \{b_i\}_i$, $\mathcal{B}_C = \{c_i\}_i$, and $\mathcal{B}_D = \{d_i\}_i$. For simplicity, we will assume that $\mathcal{B}_A$, $\mathcal{B}_B$, $\mathcal{B}_C$ and $\mathcal{B}_D$ have cardinality n. The matrix corresponding to the morphism f is an $n^2 \times n^2$ matrix with entries:

$$\begin{pmatrix}
f_{11}^{11} & \cdots & f_{11}^{1n} & f_{11}^{21} & \cdots & f_{11}^{2n} & \cdots & f_{11}^{n1} & \cdots & f_{11}^{nn} \\
\vdots & \ddots & & & & & & & & \vdots \\
f_{1n}^{11} & & \ddots & & & & & & & f_{1n}^{nn} \\
f_{21}^{11} & & & \ddots & & & & & & f_{21}^{nn} \\
\vdots & & & & \ddots & & & & & \vdots \\
f_{2n}^{11} & & & & & \ddots & & & & f_{2n}^{nn} \\
\vdots & & & & & & \ddots & & & \vdots \\
f_{n1}^{11} & & & & & & & \ddots & & f_{n1}^{nn} \\
\vdots & & & & & & & & \ddots & \vdots \\
f_{nn}^{11} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & f_{nn}^{nn}
\end{pmatrix}$$

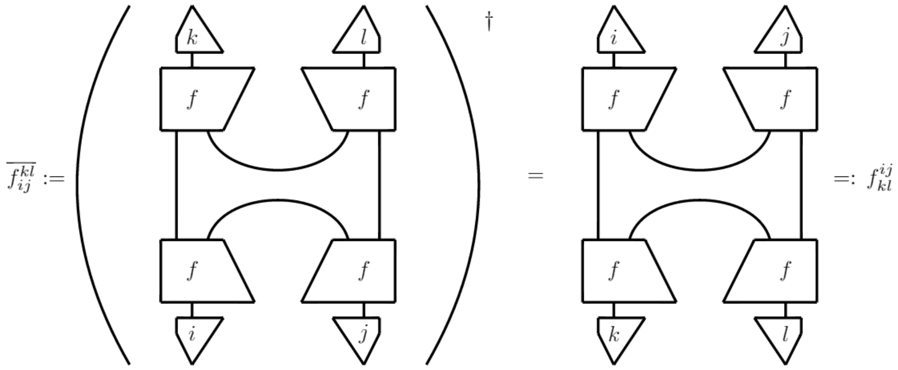
where f_{ij}^{kl} is the entry in the $((i-1)n+j)^{th}$ row and $((k-1)n+l)^{th}$ column. In fact, the algebraic representation of this matrix is $\sum\limits_{i,j,k,l} f_{ij}^{kl} (a_i \otimes c_k) \otimes (b_j \otimes d_l)$. Note that the diagrammatic representation of f_{ij}^{kl} for $f : A \otimes A^* \to A \otimes A^*$ is given by:



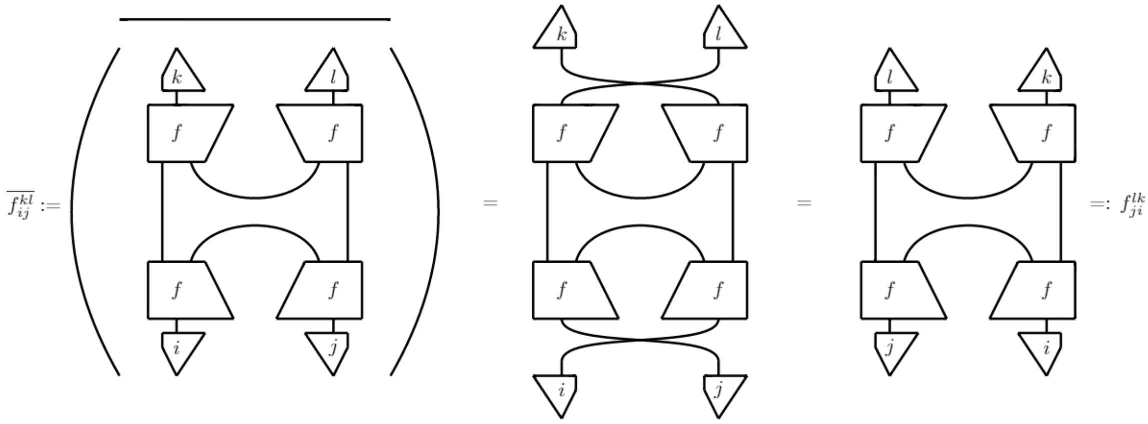
Proposition 7.3.1. *The entries of density matrices-1 and -2 satisfy*

$$\overline{f_{ij}^{kl}} = f_{kl}^{ij} \qquad\qquad \overline{f_{ij}^{kl}} = f_{ji}^{lk}$$

Proof. The first equality follows immediately from the hermiticity of density matrices-1 and 2. In fact:



The algebraic conjugate of a matrix A is the matrix whose entries are the conjugates of entries in A. Therefore:



The equalities for density matrix-2 are proved similarly. □

In this chapter, we explored density matrices even further, gave their graphical representation as (7.1) – equivalently (7.2) –, and showed that morphisms of the form (7.1) satisfy hermiticity and positive-semidefiniteness. We also showed that morphisms satisfying hermiticity and positive-semidefiniteness could be written as morphisms of the form (7.1). We then introduced the notion of double-density matrices, quadripartite states whose structure encloses two density matrices: density matrices-1 and -2. We defined the properties satisfied by density matrices-1 and -2 and double-density matrices, and characterised the entries of density matrices-1 and -2.

The next chapter details the role of double-density matrices in representing homonymous, general words.

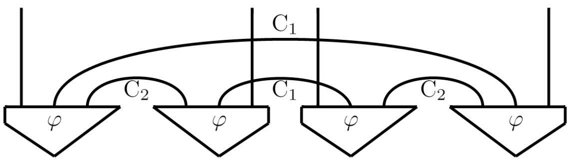
Chapter 8

Introducing ambiguity and entailment in formal semantics

This chapter investigates the algebraic structure of double-density matrices and gives a detailed account of the way they model two features of language: ambiguity and entailment. We demonstrate that double-density matrices conserve ambiguity and/or entailment when context is lacking, and collapse gradually when more and more context is provided. Furthermore, double-density matrices come equipped with two measures of entropy: linguistically, one measures the level of ambiguity of the word, and the other one the level of entailment.

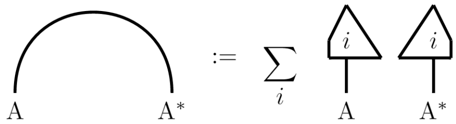
8.1 Taking a closer look at homonymous, general words

As seen in the previous chapters, a homonymous – or simply *ambiguous* – general word w is represented by a CP^2 state or *double-density matrix* $\rho^{(2)}(w)$:

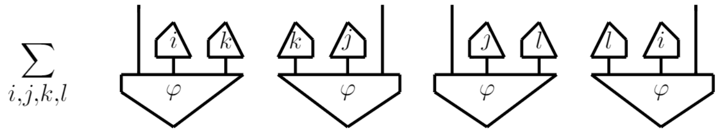


[10] gives an interpretation of the cap in terms of basis vectors of an orthonormal basis – or ONB –: given a cap $\epsilon_A : A^* \otimes A \to I$ over A, and $\{|i\rangle\}_i$ an orthonormal basis of A, the cap is interpreted as $\sum_i \langle i|_* \otimes \langle i|$, where $\langle i|_*$ denotes the conjugate of $\langle i|$.

The caps in CP^2 maps are of the form $\epsilon_{A^*} : A \otimes A^* \to I$. Following the model of [10], their interpretation is therefore $\sum_i \langle i| \otimes \langle i|_*$, graphically:



We interpret double-density matrices in terms of basis vectors of two ONBs, $\mathcal{B}_1$ and $\mathcal{B}_2$, where $\mathcal{B}_1$ is an ONB for the first ancillary system C_1, and $\mathcal{B}_2$ is an ONB for the second ancillary system C_2:



We notice that $|i\rangle$ and $|j\rangle$ are basis vectors of $\mathcal{B}_1$, and $|k\rangle$ and $|l\rangle$ are basis vectors of $\mathcal{B}_2$. This means that indices i and j vary over the same interval $\mathcal{I}_1$, and indices k and l vary over the same interval $\mathcal{I}_2$.

The algebraic representation of the above diagram is given by:

$$\rho^{(2)}(w) = \sum_{\substack{i,j \in \mathcal{I}_1 \\ k,l \in \mathcal{I}_2}} p_{i,k}p_{j,k}p_{j,l}p_{i,l} \, |\varphi_{i,k}\rangle \otimes |\varphi_{j,k}\rangle_* \otimes |\varphi_{j,l}\rangle \otimes |\varphi_{i,l}\rangle_* \tag{8.1}$$

where $p_{x,y}$ is the probability term of $|\varphi_{x,y}\rangle$.

Let us now take another approach and adopt a linguistic point of view. We consider an ambiguous, general word w, with N *unambiguous meanings* $a_1, a_2, ..., a_N$. Let w entail M words $e_1, e_2, ..., e_M$. We will refer to these words as *subsumed words*.

To each unambiguous meaning a_i corresponds a set $\mathcal{E}_i$ of subsumed words. We illustrate this idea by the following example: let "Beirut" be an ambiguous, general word, with unambiguous meanings "Beirut city" and "Beirut band". The city of Beirut has different neighbourhoods n_1, n_2, ... ,n_α, and the band Beirut has band members $m_1, m_2, ..., m_\beta$. The set $\{n_1, n_2, ..., n_\alpha, m_1, m_2, ..., m_\beta\}$ is the set of words subsumed by "Beirut", and the subsets $\{n_1,\ n_2,\ ...\ ,n_\alpha\}$ and $\{m_1, m_2, ..., m_\beta\}$ correspond respectively to unambiguous meanings "Beirut city" and "Beirut band".

To formalise this idea:

to unambiguous meaning a_1 corresponds the set $\{e_{1,1}, e_{1,2}, ..., e_{1,\alpha}\} = \mathcal{E}_1$,
to unambiguous meaning a_2 corresponds the set $\{e_{2,1}, e_{2,2}, ..., e_{2,\beta}\} = \mathcal{E}_2$,

.

.

.

to unambiguous meaning a_N corresponds the set $\{e_{N,1}, e_{N,2}, ..., e_{N,\omega}\} = \mathcal{E}_N$.

The disjoint union of the sets $\mathcal{E}_i$'s is equal to the set of all subsumed words $\{e_1, e_2, ..., e_M\}$. Note that a subsumed word belongs to exactly one of the $\mathcal{E}_i$'s.

The words $\{a_i\}_i$ serve as "dummy words" and do not correspond to vectors in the noun space. The words $\{e_{x,y}\}$, however, *do* correspond to vectors in the noun space, and the indices x and y respectively account for ambiguity and entailment. It is clear that x varies over the set $\{1, 2, ..., N\}$, but the interval over which y varies is not as obvious. Looking back at expression (8.1), each $|\varphi_{x,y}\rangle$ corresponds to subsumed word $e_{x,y}$, indices i and j account for ambiguity, and indices k and l account for entailment. Therefore, x varies over $\mathcal{I}_1 = \{1, 2, ..., N\}$, and y varies over $\mathcal{I}_2$. In order for all subsumed words to be taken into account in the expression of w, y has to vary over the interval $\mathcal{I}_2 = \{1, 2, ...max\}$, where max denotes the maximum of $\{\alpha, \beta, ..., \omega\}$, or in other words, the maximal cardinality of the $\mathcal{E}_i$'s.

This definition of $\mathcal{I}_2$ poses a new challenge: if there exists a set $\mathcal{E}_x$ of cardinality less than max, the algebraic representation of w will have vectors $|\varphi_{x,t}\rangle$, where t is greater than the cardinality of $\mathcal{E}_x$, and hence $|\varphi_{x,t}\rangle$ will not correspond to any of the words subsumed by w. Let us illustrate this with an example: let w have two unambiguous meanings, and let $\mathcal{E}_1 = \{e_{1,1}, e_{1,2}\}$, $\mathcal{E}_2 = \{e_{2,1}, e_{2,2}, e_{2,3}\}$. The algebraic representation of w has a vector $|\varphi_{1,3}\rangle$ that does not represent any of the words subsumed by w.

One way of dealing with this problem is the following: every set $\mathcal{E}_x = \{e_{x,1}, e_{x,2}, ..., e_{x,v}\}$, where the cardinality v is less than max, is extended to $\{e_{x,1}, e_{x,2}, ..., e_{x,v}, e_{x,v+1}, ..., e_{x,max}\}$, where the words $e_{x,v+1}, ..., e_{x,max}$ are repetitions of some of the words in $\{e_{x,1}, e_{x,2}, ..., e_{x,v}\}$. It is important to note that the addition of these words should preserve original probability distributions. In our earlier example, we extend $\mathcal{E}_1$ to include a word $e_{1,3}$ which is a repetition of $e_{1,1}$. $e_{1,1}$ occurs originally with probability $p_{1,1}$, we could set $p_{1,3}$ and the new value of $p_{1,1}$ to half its original value. In the examples I give below, I will consider simple cases where the cardinalities of the $\mathcal{E}_i$'s are originally all the same.

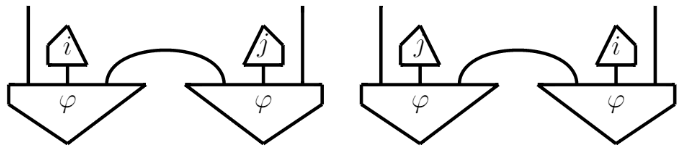
Note: This is not the only solution to this problem, but we leave this for future work.

Two types of sums. Two different levels of mixing occur in a double-density matrix: one accounts for ambiguity, and one for entailment.
In order to better illustrate this point, I accompany the explanation with the following example: consider the ambiguous, general word $w =$"Beirut", with unambiguous meanings $a_1 =$"Beirut city" and $a_2 =$"Beirut band". The city of Beirut has neighbourhoods $e_{1,1} =$"Ashrafieh", that we will denote by "A", and $e_{1,2} =$"Monot", that we will denote by "M", while the band has members $e_{2,1} =$"Zach", denoted by "Z", and $e_{2,2} =$"Paul",

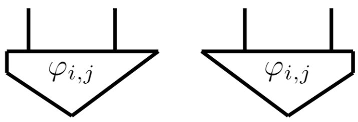
denoted by "P". The algebraic representation of "Beirut" has 16 terms, listed in the table below. Note that the dirac notation and tensors are assumed for clarity:

$i=1$	$j=1$	$k=1$	$l=1$	AA_*AA_*
			$l=2$	AA_*MM_*
		$k=2$	$l=1$	MM_*AA_*
			$l=2$	MM_*MM_*
	$j=2$	$k=1$	$l=1$	AZ_*ZA_*
			$l=2$	AZ_*PM_*
		$k=2$	$l=1$	MP_*ZA_*
			$l=2$	MP_*PM_*
$i=2$	$j=1$	$k=1$	$l=1$	ZA_*AZ_*
			$l=2$	ZA_*MP_*
		$k=2$	$l=1$	PM_*AZ_*
			$l=2$	PM_*MP_*
	$j=2$	$k=1$	$l=1$	ZZ_*ZZ_*
			$l=2$	ZZ_*PP_*
		$k=2$	$l=1$	PP_*ZZ_*
			$l=2$	PP_*PP_*

The terms in this table come from two levels of mixing: in fact, let us consider the representation of the double-density matrix where the caps over C_1 are represented by a sum over basis vectors:

$\sum_{i,j\in\mathcal{I}_1}$


Notice that this representation is equivalent to:

$\sum_{i,j\in\mathcal{I}_1}$


Each bipartite state $|\varphi_{i,j}\rangle$ is of the form

$$|\alpha_{i,1}\rangle \otimes |\beta_{j,1}\rangle_* + |\alpha_{i,2}\rangle \otimes |\beta_{j,2}\rangle_* + ... + |\alpha_{i,M}\rangle \otimes |\beta_{j,M}\rangle_*$$

and every term $|\varphi_{i,j}\rangle \otimes |\varphi_{i,j}\rangle_*$ is of the form

$$(|\alpha_{i,1}\rangle \otimes |\beta_{j,1}\rangle_* + |\alpha_{i,2}\rangle \otimes |\beta_{j,2}\rangle_* + ... + |\alpha_{i,M}\rangle \otimes |\beta_{j,M}\rangle_*)$$
$$\otimes(|\beta_{j,1}\rangle \otimes |\alpha_{i,1}\rangle_* + |\beta_{j,2}\rangle \otimes |\alpha_{i,2}\rangle_* + ... + |\beta_{j,M}\rangle \otimes |\alpha_{i,M}\rangle_*)$$

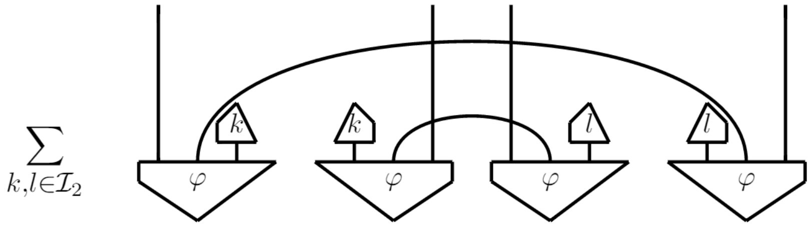
Here, the summation inside the parentheses corresponds to *entailment mixing*, because the second ancillary system models entailment. The outer summation over indices i and j corresponds to *ambiguity mixing*, because the first ancillary system models ambiguity.

In order to illustrate these two levels of mixing, let us go back to our example:

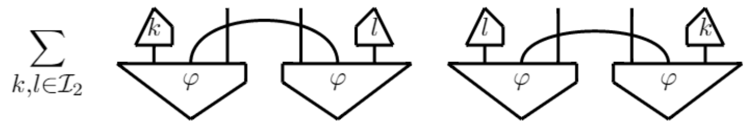
- for i=1, j=1, $|\varphi_{1,1}\rangle \otimes |\varphi_{1,1}\rangle_* = (AA_* + MM_*)(AA_* + MM_*) = AA_*AA_* + AA_*MM_* + MM_*AA_* + MM_*MM_*$
- for i=1, j=2, $|\varphi_{1,2}\rangle \otimes |\varphi_{1,2}\rangle_* = (AZ_* + MP_*)(ZA_* + PM_*) = AZ_*ZA_* + AZ_*PM_* + MP_*ZA_* + MP_*PM_*$
- for i=2, j=1, $|\varphi_{2,1}\rangle \otimes |\varphi_{2,1}\rangle_* = (ZA_* + PM_*)(AZ_* + MP_*) = ZA_*AZ_* + ZA_*PM_* + PM_*AZ_* + PM_*MP_*$
- for i=2, j=2, $|\varphi_{2,2}\rangle \otimes |\varphi_{2,2}\rangle_* = (ZZ_* + PP_*)(ZZ_* + PP_*) = ZZ_*ZZ_* + ZZ_*PP_* + PP_*ZZ_* + PP_*PP_*$

At this stage, the "+" between the parentheses denotes entailment mixing. Adding all of these terms together – i.e. adding ambiguity mixing –, we recover the 16 terms in the table above.

Another approach to this two-level mixing is to swap the levels, and consider the diagram where the caps over C_2 are represented by a sum over basis vectors:



By dragging the first component to the far right, we obtain:



The same analysis can now be carried out. Here, the "+" inside the parentheses will correspond to *ambiguity* mixing, and the outer summation corresponds to *entailment* mixing. Note that after computing all the terms, the last component in each should be returned to the first position, in order to retrieve the correct order. This shows that **the order of mixing does not matter**: whether we perform entailment mixing first or second, the representation of w remains unchanged.

8.2 An application

The examples in this chapter will deal with the sentence space $S = \{\perp, \top\}$, where $\perp$ stands for *"false"* and $\top$ for *"true"*. This means that, as in Montague semantics, the meaning of a sentence boils down to its truth value.

In **Rel**, the meaning of a sentence can be either $|\perp\rangle$, $|\top\rangle$, or $|\perp\rangle + |\top\rangle$, where the superposition of $|\perp\rangle$ and $|\top\rangle$ stems from lack of sufficient context. We recall that **Rel** is the †-compact-closed category with the class of sets as objects, binary relations $R \subseteq X \times Y$ as morphisms $X \to Y$, and relational composition as composition of morphisms. The tensor product is given by the Cartesian product of sets/relations, the tensor unit by the singleton set $\{\star\}$, and the dagger is defined by $R^\dagger := \{(y, x) \mid (x, y) \in \mathbf{R}\}$ [20].

[35] presents a toy model of linguistic ambiguity in **CPM(Rel)**, thereby adding a dimension interpreted as ambiguity. A sentence in **CPM(Rel)** can take four possible values: $|\perp\rangle \otimes |\perp\rangle$, $|\top\rangle \otimes |\top\rangle$, $(|\perp\rangle + |\top\rangle) \otimes (|\perp\rangle + |\top\rangle)$, and $|\perp\rangle \otimes |\perp\rangle + |\top\rangle \otimes |\top\rangle$. Here, the first three states correspond to pure states, and the fourth one is a mixed state, representing ambiguity. Note that there are five states in **CPM(Rel)**, the fifth one being $\emptyset$, which also corresponds to a pure state.

In the examples below, I introduce yet another dimension, that of entailment. This requires us to work in the category $\mathbf{CPM^2(Rel)}$.

8.2.1 States in $\mathbf{CPM^2(Rel)}$

In [31], Marsden provides a graph theoretic perspective of **CPM(Rel)** and devises a way to characterise the states of **CPM(Rel)**. Building on this approach, Cunningham [15] devised a four-step way of characterising the states of $\mathbf{CPM^2(Rel)}$ diagrammatically:

Algorithm for the diagrammatic characterisation of the states of $\mathbf{CPM}^2(\mathbf{Rel})$:

Starting with a set S, we built a graph $\mathcal{G}$ corresponding to a state of $\mathbf{CPM}^2(\mathbf{Rel})$ as follows:

1. Select a subset of the elements of S to be **vertices** of $\mathcal{G}$
2. Out of the set of possible edges, select a subset of edges between the different vertices and label them as *blue edges*. All self-edges (edges from a vertex to itself), must be selected.
3. Out of the set of possible edges, select a subset of edges between the different vertices and label them as *red edges*. All self-edges must be selected.
4. Out of the set of possible *alternating squares* (cycles of four edges of alternating colours), select a subset of squares and label them as *purple squares*. All self-squares, i.e. alternating squares whose two *blue* edges are the same **or** whose two *red* edges are the same), must be selected.

In order to find the corresponding state in $\mathbf{CPM}^2(\mathbf{Rel})$, we follow this nomenclature process: for every alternating square in $\mathcal{G}$:

- Pick one of the red edges to be the starting edge
- Pick a direction – clockwise or counterclockwise: every edge of the alternating square in now a directed edge (h, t), where h is the head of the edge, and t its tail.
- Go through each of the four edges in the direction chosen, naming the head of every edge. The result is an ordered quadruple of vertices.

Note 1: An alternating square can have more than one description: the nomenclature process can start with any one of the red edges, and can alternate through edges either clockwise or counterclockwise ($\mathcal{G}$ is an undirected graph). All possible descriptions of every alternating square in $\mathcal{G}$ make up the state in $\mathbf{CPM}^2(\mathbf{Rel})$ that $\mathcal{G}$ corresponds to.

Note 2: An ordered quadruple (A, B, C, D) corresponds to $|A\rangle \otimes |B\rangle \otimes |C\rangle \otimes |D\rangle$ in $\mathbf{CPM}^2(\mathbf{Rel})$. For simplicity, the bra, ket, and tensor notations will be assumed, and we write $ABCD$.

We now apply this technique to find the graphs corresponding to the different states of $\mathbf{CPM}^2(\mathbf{Rel})$. Starting with a set $S = \{\bot, \top\}$, step 1 of the algorithm leads to three possible cases:

- $\mathcal{G}$ has no vertices: $\mathcal{G}$ is the empty diagram, and corresponds to the empty state $\emptyset$.

- $\mathcal{G}$ has one vertex: this case has two subcases:

    - The following diagram, corresponding to the state $\bot\bot\bot\bot$:

    - The following diagram, corresponding to the state $\top\top\top\top$:

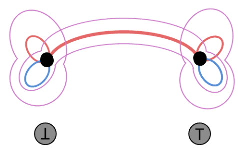
  Note that all edges are self-edges, and all alternating squares are self-squares.

- $\mathcal{G}$ has two vertices: the application of steps 2 and 3 of the algorithm leads to three subcases:

    - No edges other than self-edges are labelled. This is represented by the following diagram and corresponds to the state $\bot\bot\bot\bot + \top\top\top\top$:

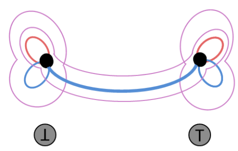
      Here, all alternating squares are self-squares.

    - One "non-self" edge is labelled *red*. This is represented by the following diagram and corresponds to the state $\bot\bot\bot\bot + \top\top\top\top + \bot\top\top\bot + \top\bot\bot\top$:



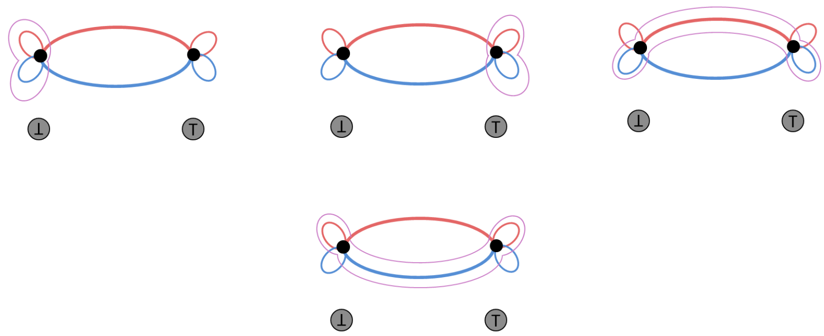
      Here also, all alternating squares are self-squares.

    - One "non-self" edge is labelled *blue*. This is represented by the following diagram and corresponds to the state $\bot\bot\bot\bot + \top\top\top\top + \bot\bot\top\top + \top\top\bot\bot$:



As in the previous cases, all alternating squares are self-squares.

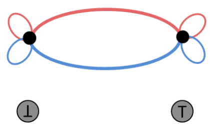
- One "non-self" edge is labelled *red* and another one *blue*. This case has four self-squares, shown independently below



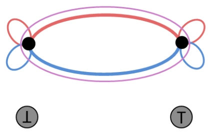
and corresponding respectively to states $\perp\perp\perp\perp$, $\top\top\top\top$, $\perp\top\top\perp + \top\perp\perp\top$, and $\perp\perp\top\top + \top\top\perp\perp$. These self-squares will be assumed in the following diagrams and will not be represented, for the sake of clarity.

There are three alternating squares that are not self-squares. This subcase has therefore $2^3 = 8$ *subsubcases*:

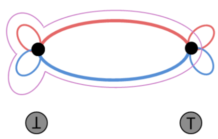
- The diagram below, corresponding to the state $\perp\perp\perp\perp + \top\top\top\top + \perp\top\top\perp + \top\perp\perp\top + \perp\perp\top\top + \top\top\perp\perp$:



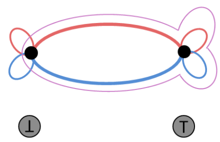
- The diagram below, corresponding to the state $\perp\perp\perp\perp + \top\top\top\top + \perp\top\top\perp + \top\perp\perp\top + \perp\perp\top\top + \top\top\perp\perp + \perp\top\perp\top + \top\perp\top\perp$:



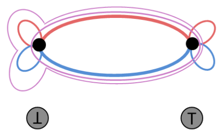
- The diagram below, corresponding to the state $\perp\perp\perp\perp + \top\top\top\top + \perp\top\top\perp + \top\perp\perp\top + \perp\perp\top\top + \top\top\perp\perp + \perp\perp\perp\top + \perp\perp\top\perp + \perp\top\perp\perp + \top\perp\perp\perp$:



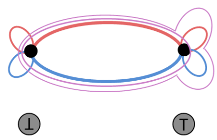
- The diagram below, corresponding to the state $\bot\bot\bot\bot + \top\top\top\top + \bot\top\top\bot + \top\bot\bot\top + \bot\bot\top\top + \top\top\bot\bot + \bot\top\top\top + \top\bot\top\top + \top\top\bot\top + \top\top\top\bot$:



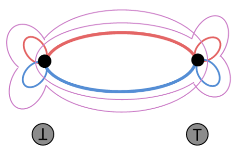
- The diagram below, corresponding to the state $\bot\bot\bot\bot + \top\top\top\top + \bot\top\top\bot + \top\bot\bot\top + \bot\bot\top\top + \top\top\bot\bot + \bot\top\bot\top + \top\bot\top\bot + \bot\bot\bot\top + \bot\bot\top\bot + \bot\top\bot\bot + \top\bot\bot\bot$:



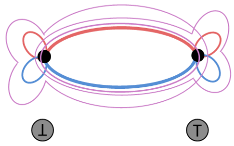
- The diagram below, corresponding to the state $\bot\bot\bot\bot + \top\top\top\top + \bot\top\top\bot + \top\bot\bot\top + \bot\bot\top\top + \top\top\bot\bot + \bot\top\bot\top + \top\bot\top\bot + \bot\top\top\top + \top\bot\top\top + \top\top\bot\top + \top\top\top\bot$:



- The diagram below, corresponding to the state $\bot\bot\bot\bot + \top\top\top\top + \bot\top\top\bot + \top\bot\bot\top + \bot\bot\top\top + \top\top\bot\bot + \bot\bot\bot\top + \bot\bot\top\bot + \bot\top\bot\bot + \top\bot\bot\bot + \bot\top\top\top + \top\bot\top\top + \top\top\bot\top + \top\top\top\bot$:



- The diagram below, corresponding to the state $\bot\bot\bot\bot + \top\top\top\top + \bot\top\top\bot + \top\bot\bot\top + \bot\bot\top\top + \top\top\bot\bot + \bot\top\bot\top + \top\bot\top\bot + \bot\bot\bot\top + \bot\bot\top\bot + \bot\top\bot\bot + \top\bot\bot\bot + \bot\top\top\top + \top\bot\top\top + \top\top\bot\top + \top\top\top\bot$:



States of $\mathbf{CPM}^2(\mathbf{Rel})$ are summarised in the table below. We will look at the linguistic interpretation of some of these states in our examples.

State	*Description*
1	∅
2	⊥⊥⊥⊥
3	⊤⊤⊤⊤
4	⊥⊥⊥⊥ + ⊤⊤⊤⊤
5	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤
6	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥
7	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥
8	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥ + ⊥⊤⊥⊤ + ⊤⊥⊤⊥
9	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥ + ⊥⊥⊥⊤ + ⊥⊥⊤⊥ + ⊥⊤⊥⊥ + ⊤⊥⊥⊥
10	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥ + ⊥⊤⊤⊤ + ⊤⊥⊤⊤ + ⊤⊤⊥⊤ + ⊤⊤⊤⊥
11	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥ + ⊥⊤⊥⊤ + ⊤⊥⊤⊥ + ⊥⊥⊥⊤ + ⊥⊥⊤⊥ + ⊥⊤⊥⊥ + ⊤⊥⊥⊥
12	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥ + ⊥⊤⊥⊤ + ⊤⊥⊤⊥ + ⊥⊤⊤⊤ + ⊤⊥⊤⊤ + ⊤⊤⊥⊤ + ⊤⊤⊤⊥
13	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥ + ⊥⊥⊥⊤ + ⊥⊥⊤⊥ + ⊥⊤⊥⊥ + ⊤⊥⊥⊥ + ⊥⊤⊤⊤ + ⊤⊥⊤⊤ + ⊤⊤⊥⊤ + ⊤⊤⊤⊥
14	⊥⊥⊥⊥ + ⊤⊤⊤⊤ + ⊥⊤⊤⊥ + ⊤⊥⊥⊤ + ⊥⊥⊤⊤ + ⊤⊤⊥⊥ + ⊥⊤⊥⊤ + ⊤⊥⊤⊥ + ⊥⊥⊥⊤ + ⊥⊥⊤⊥ + ⊥⊤⊥⊥ + ⊤⊥⊥⊥ + ⊥⊤⊤⊤ + ⊤⊥⊤⊤ + ⊤⊤⊥⊤ + ⊤⊤⊤⊥

8.2.2 A practical example

In all concrete applications, the distributional model space obtained from real world data is a real vector space. Therefore, all vectors in the noun space have real entries, and a vector $|v\rangle$ is equal to its conjugate $|v\rangle_*$. In what follows, we will drop the lower star notation.

We will consider the ambiguous, general word "Beirut", with unambiguous meanings "Beirut city" and "Beirut band", each with respective set of subsumed words {Ashrafieh, Monot} and {Zach, Paul}. As earlier, we will denote "Ashrafieh", "Monot", "Zach" and "Paul" by A, M, Z, and P. Our examples will also involve the noun phrases "Lebanese national day", "top single", and "Zach's birthday", which we will denote by L, S, and B.

Notice that the vectors corresponding to L, S, and B live in the noun space. In fact, by grammatical reductions, we can show that the grammatical type of each of these noun phrases is n. We give here one example: in the English language, the grammatical type of an adjective is given by nn^l. The noun phrase "Lebanese national day" has type $(nn^l)(nn^l)n$, which reduces as follows:

$$(nn^l)(nn^l)n \to n(n^l n)(n^l n) \to n$$

For the purposes of this example, we will take all subsumed words, as well as L, S, and B to be our set of context words. The set $\{A, M, Z, P, L, S, B\}$ forms an ONB. The words A, M, Z, P, L, S, and B are unambiguous, non-general words, and are represented by 1,2-pure states in $\mathbf{CPM^2(FHilb)}$. Therefore:

$$\rho^{(2)}(A) = |A\rangle \otimes |A\rangle \otimes |A\rangle \otimes |A\rangle$$

which we will denote also by $|AAAA\rangle$ or $|A\rangle^{\otimes^4}$. The others are defined similarly.

The representation of "Beirut" was derived in the previous section. Recall that:

$$\begin{aligned}\rho^{(2)}(\mathit{Beirut}) = &|AAAA\rangle + |AAMM\rangle + |MMAA\rangle + |MMMM\rangle \\ &+ |AZZA\rangle + |AZPM\rangle + |MPZA\rangle + |MPPM\rangle \\ &+ |ZAAZ\rangle + |ZAMP\rangle + |PMAZ\rangle + |PMMP\rangle \\ &+ |ZZZZ\rangle + |ZZPP\rangle + |PPZZ\rangle + |PPPP\rangle\end{aligned}$$

So $\rho^{(2)}(\mathit{Beirut})$ is of the form $\sum_m c_m^b \, |b_m^1 b_m^2 b_m^3 b_m^4\rangle$. Notice that there are no probability terms in the above derivation. This is because, for the purposes of the examples of this section, it is only important for us to know whether a term is present in the expression of a homonymous, general word or not. The probability terms are superfluous and are therefore omitted. In the next section, which deals with measuring the level of ambiguity and generality of a word, these terms will play an important role, and we will define them then.

At this point, we introduce a basic yet intuitive measure of entailment: a word w_1 is subsumed by a word w_2 if $\rho^{(2)}(w_1)$ is contained in the expression of $\rho^{(2)}(w_2)$. Building on this measure, we can clearly see that A, M, Z, and P are subsumed by "Beirut", since $|A\rangle^{\otimes^4}$, $|M\rangle^{\otimes^4}$, $|Z\rangle^{\otimes^4}$, and $|P\rangle^{\otimes^4}$ appear in the expression of $\rho^{(2)}(\mathit{Beirut})$. A more precise measure of entailment is given in the next section.

We now define the three verbs we will use in our examples: the transitive verbs *celebrate* and *play-in* are of the form

$$\sum_{i,j,k} c_{ijk} \, |subject_i\rangle \otimes |x_j\rangle \otimes |object_k\rangle$$

where $|x\rangle \in \{|\bot\rangle, |\top\rangle\}$, and are defined by:

$$\begin{aligned}|celebrate\rangle =& |A\rangle \otimes |\top\rangle \otimes |L\rangle + |M\rangle \otimes |\top\rangle \otimes |L\rangle + |Z\rangle \otimes |\top\rangle \otimes |B\rangle \\ &+ |P\rangle \otimes |\top\rangle \otimes |B\rangle + |Z\rangle \otimes |\bot\rangle \otimes |S\rangle + |P\rangle \otimes |\bot\rangle \otimes |S\rangle\end{aligned}$$

$$|play\text{-}in\rangle = |Z\rangle \otimes |\top\rangle \otimes |A\rangle + |P\rangle \otimes |\top\rangle \otimes |A\rangle$$

The intransitive verb *perform* is of the form

$$\sum_{i,j} c_{ij} |subject_i\rangle \otimes |x_j\rangle$$

where $|x\rangle \in \{|\bot\rangle, |\top\rangle\}$, and is defined by:

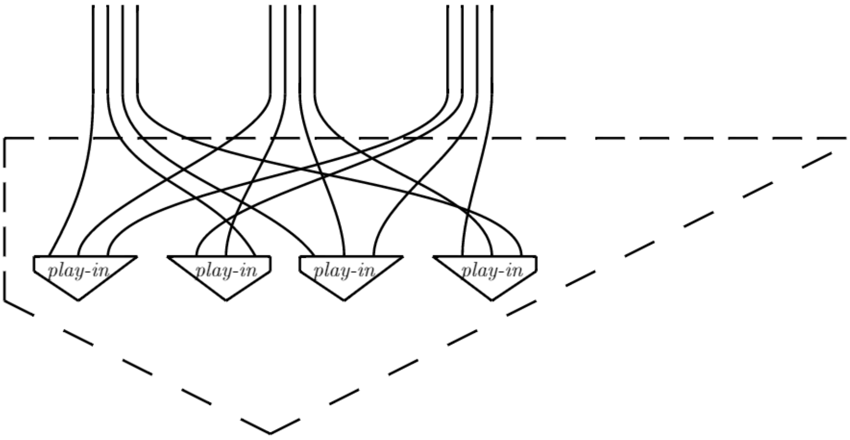
$$|perform\rangle = |Z\rangle \otimes |\top\rangle + |P\rangle \otimes |\top\rangle$$

Since these verbs are unambiguous and non-general, their double-density matrix representation is computed in a way similar to that of unambiguous, non-general nouns. We show here the computation of $\rho^{(2)}(play\text{-}in)$. The ket and tensor notations are assumed for in the middle steps for clarity.

$$\begin{aligned}\rho^{(2)}(play\text{-}in) =&(Z\top A + P\top A)(Z\top A + P\top A)(Z\top A + P\top A)(Z\top A + P\top A)\\ =&(ZZ\top\top AA + ZP\top\top AA + PZ\top\top AA + PP\top\top AA)\\ &(ZZ\top\top AA + ZP\top\top AA + PZ\top\top AA + PP\top\top AA)\\ =&|ZZZZ\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |ZZZP\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |ZZPZ\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4}\\ &+ |ZZPP\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |ZPZZ\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |ZPZP\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4}\\ &+ |ZPPZ\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |ZPPP\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |PZZZ\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4}\\ &+ |PZZP\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |PZPZ\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |PZPP\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4}\\ &+ |PPZZ\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |PPZP\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4} + |PPPZ\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4}\\ &+ |PPPP\rangle \otimes |\top\rangle^{\otimes 4} \otimes |A\rangle^{\otimes 4}\end{aligned}$$

This corresponds graphically to:

$\rho^{(2)}(\textit{play-in}) :=$



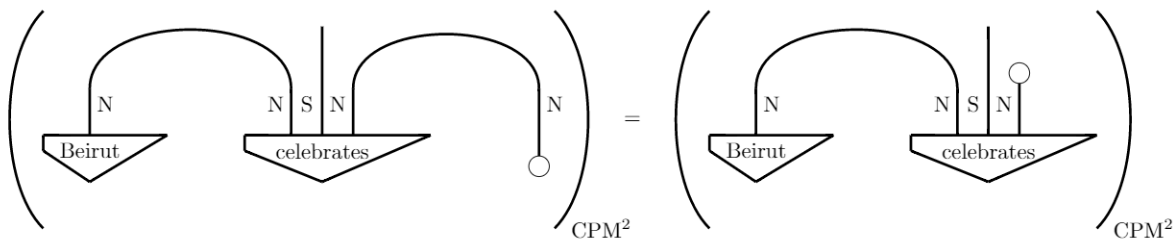
So the double-density matrix of a transitive verb has the form

$$\sum_{i,j,k} c_{ijk} \, |s_i^1 s_i^2 s_i^3 s_i^4\rangle \otimes |x_j^1 x_j^2 x_j^3 x_j^4\rangle \otimes |o_k^1 o_k^2 o_k^3 o_k^4\rangle$$

$\rho^{(2)}(\textit{celebrate})$ and $\rho^{(2)}(\textit{perform})$ are computed similarly. In what follows, we will explain thoroughly how the meaning of the sentences is computed, but we will not explicitly show all the terms involved. The word $\rho^{(2)}(\textit{celebrate})$, for example, is composed of 1296 terms.

Let us now demonstrate that our framework accommodates indeed for both ambiguity and entailment relationships:

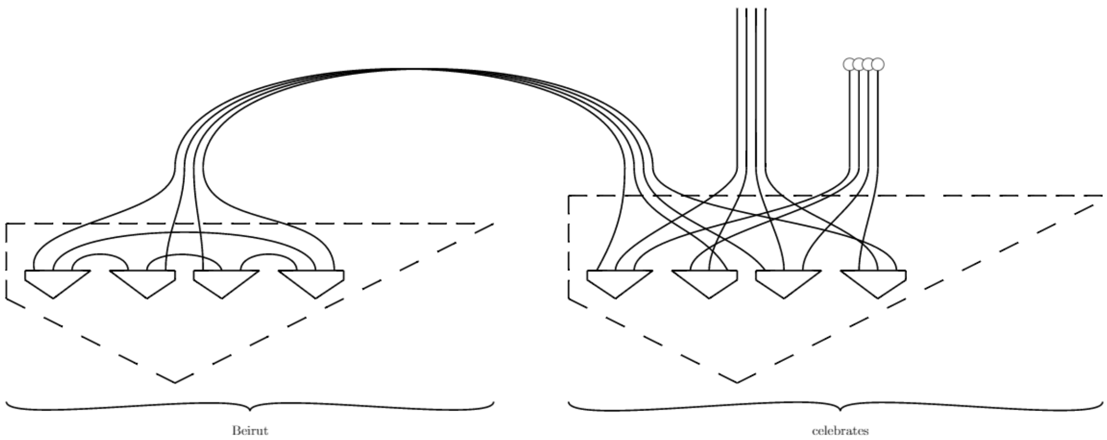
"Beirut celebrates". In this sentence, the word "Beirut" is still ambiguous, and all the neighbourhoods and the band members remain subsumed. Based on the definition of *celebrate*, we expect this sentence to be neither true nor false, but to be tending towards being true. Let us verify that all of the above expectations are satisfied by our model.



Here, the ι map deletes the object of *celebrate*. The meaning of this sentence is given by:

$$(\epsilon_{CPM_N^2} \otimes 1_{CPM_S^2} \otimes \iota_{CPM_N^2})(\rho^{(2)}(\textit{Beirut}) \otimes \rho^{(2)}(\textit{celebrate}))$$

where the maps in $\mathbf{CPM}^2\mathbf{(C)}$ work as defined in **Chapter 6**. We unfold the CPM^2 representation of the sentence in order to better visualise the interactions between the different components.



The ι map, as explained before, deletes the third component of $\rho^{(2)}(celebrate)$. "Beirut" and the subject component of *celebrate* interact via the ϵ map: the ϵ map takes the inner product of $\rho^{(2)}(Beirut)$ and the first component of $\rho^{(2)}(celebrate)$. This effectively picks out the components of $\rho^{(2)}(celebrate)$ and outputs terms of the form $|x_1\rangle\otimes|x_2\rangle\otimes|x_3\rangle\otimes|x_4\rangle$, where $|x_1\rangle, |x_2\rangle, |x_3\rangle, |x_4\rangle \in \{|\bot\rangle, |\top\rangle\}$. To formalise this:

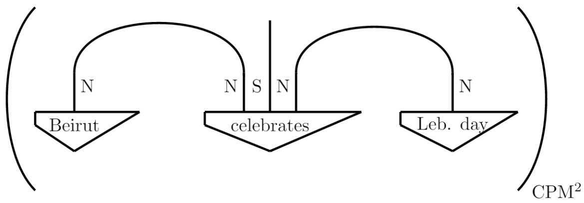
$$
\begin{aligned}
&(\epsilon_{CPM_N^2} \otimes 1_{CPM_S^2} \otimes \iota_{CPM_N^2})(\rho^{(2)}(Beirut) \otimes \rho^{(2)}(celebrate)) \\
=&(\epsilon_{CPM_N^2} \otimes 1_{CPM_S^2} \otimes \iota_{CPM_N^2})(\sum_m c_m^b\, |b_m^1 b_m^2 b_m^3 b_m^4\rangle \otimes \sum_{i,j,k} c_{ijk}\, |s_i^1 s_i^2 s_i^3 s_i^4\rangle \otimes |x_j^1 x_j^2 x_j^3 x_j^4\rangle \otimes |o_k^1 o_k^2 o_k^3 o_k^4\rangle) \\
=& \sum_{m,i,j,k} c_m^b c_{ijk} \epsilon_{CPM_N^2}(|b_m^1 b_m^2 b_m^3 b_m^4\rangle \otimes |s_i^1 s_i^2 s_i^3 s_i^4\rangle) \otimes 1_{CPM_S^2}(|x_j^1 x_j^2 x_j^3 x_j^4\rangle) \otimes \iota_{CPM_N^2}(|o_k^1 o_k^2 o_k^3 o_k^4\rangle) \\
=& \sum_{m,i,j,k} c_m^b c_{ijk}\, \langle b_m^1|s_i^1\rangle \langle b_m^2|s_i^2\rangle \langle b_m^3|s_i^3\rangle \langle b_m^4|s_i^4\rangle\, (|x_j^1\rangle \otimes |x_j^2\rangle \otimes |x_j^3\rangle \otimes |x_j^4\rangle) \\
=& \sum_{m,i,j,k} c_m^b c_{ijk}\, \langle b_m^1 b_m^2 b_m^3 b_m^4|s_i^1 s_i^2 s_i^3 s_i^4\rangle\, (|x_j^1 x_j^2 x_j^3 x_j^4\rangle)
\end{aligned}
$$

We now give the meaning of the sentence "Beirut celebrates". The ket and tensor notations are assumed, for clarity.

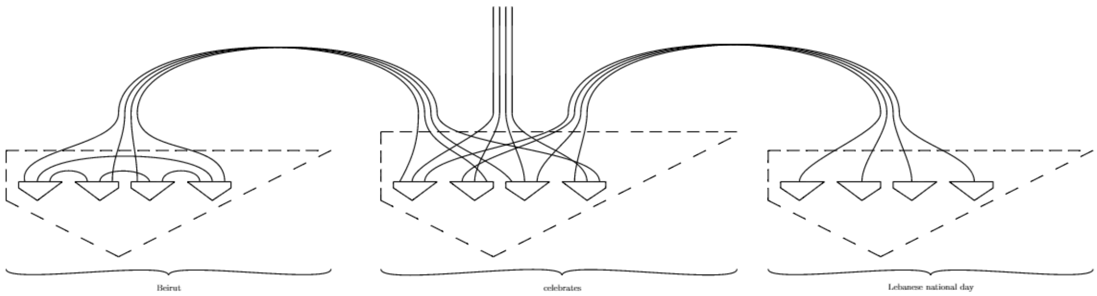
$$
\begin{aligned}
&(\epsilon_{CPM_N^2} \otimes 1_{CPM_S^2} \otimes \iota_{CPM_N^2})(\rho^{(2)}(Beirut) \otimes \rho^{(2)}(celebrate)) \\
=&(\langle AAAA|AAAA\rangle + \langle AAMM|AAMM\rangle + \langle MMAA|MMAA\rangle + \langle MMMM|MMMM\rangle)\top^{\otimes 4} \\
+&(\langle ZAAZ|ZAAZ\rangle + \langle ZAMP|ZAMP\rangle + \langle PMAZ|PMAZ\rangle + \langle PMMP|PMMP\rangle) \\
&(\top+\bot)\top\top(\top+\bot) \\
+&(\langle AZZA|AZZA\rangle + \langle AZPM|AZPM\rangle + \langle MPZA|MPZA\rangle + \langle MPPM|MPPM\rangle) \\
&\top(\top+\bot)(\top+\bot)\top \\
+&(\langle ZZZZ|ZZZZ\rangle + \langle ZZPP|ZZPP\rangle + \langle PPZZ|PPZZ\rangle + \langle PPPP|PPPP\rangle) \\
&(\top+\bot)(\top+\bot)(\top+\bot)(\top+\bot) \\
=&4(\top^{\otimes 4} + (\top+\bot)\top\top(\top+\bot) + \top(\top+\bot)(\top+\bot)\top + (\top+\bot)(\top+\bot)(\top+\bot)(\top+\bot))
\end{aligned}
$$

The middle step in this derivation shows clearly the impact of each of the components of "Beirut" in the meaning of this sentence, which indicates that "Beirut" conserved its ambiguity. It is also clear from this middle step that all the words subsumed by "Beirut" had an impact in the meaning of this sentence. Finally, the meaning of the sentence is neither true nor false, but we can see that it is dominantly true. This confirms our expectations.

"Beirut celebrates Lebanese national day". In this sentence, the meaning of "Beirut" is completely disambiguated and subsumes both neighbourhoods. Furthermore, based on the way *celebrate* is defined, this sentence is true. Let us verify that our expectations are satisfied:



When we unfold the CPM^2 representation of this diagram, we get:



The meaning of this sentence is given by:

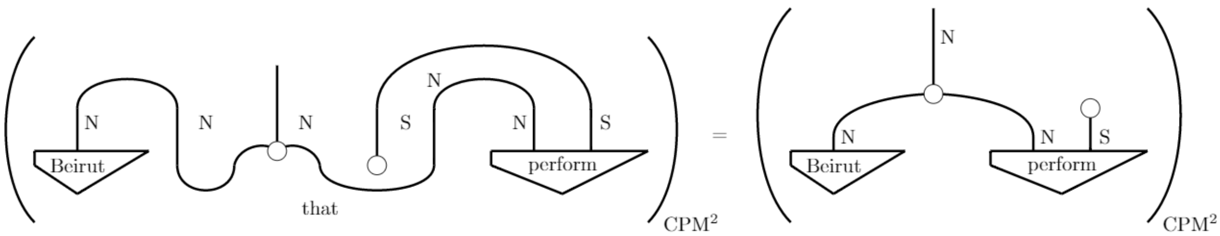
$$
\begin{aligned}
&(\epsilon_{CPM_N^2} \otimes 1_{CPM_S^2} \otimes \epsilon_{CPM_N^2})(\rho^{(2)}(Beirut) \otimes \rho^{(2)}(celebrate) \otimes \rho^{(2)}(Lebanese\ national\ day))\\
=&(\epsilon_{CPM_N^2} \otimes 1_{CPM_S^2} \otimes \iota_{CPM_N^2})(\sum_m c_m^b \left|b_m^1 b_m^2 b_m^3 b_m^4\right\rangle \otimes \sum_{i,j,k} c_{ijk} \left|s_i^1 s_i^2 s_i^3 s_i^4\right\rangle \otimes \left|x_j^1 x_j^2 x_j^3 x_j^4\right\rangle \otimes \left|o_k^1 o_k^2 o_k^3 o_k^4\right\rangle\\
&\otimes |LLLL\rangle)\\
=&\sum_{m,i,j,k} c_m^b c_{ijk} \epsilon_{CPM_N^2}(\left|b_m^1 b_m^2 b_m^3 b_m^4\right\rangle \otimes \left|s_i^1 s_i^2 s_i^3 s_i^4\right\rangle) \otimes 1_{CPM_S^2}(\left|x_j^1 x_j^2 x_j^3 x_j^4\right\rangle) \otimes \epsilon_{CPM_N^2}(\left|o_k^1 o_k^2 o_k^3 o_k^4\right\rangle \otimes |LLLL\rangle)\\
=&\sum_{m,i,j,k} c_m^b c_{ijk} \left\langle b_m^1 | s_i^1\right\rangle \left\langle b_m^2 | s_i^2\right\rangle \left\langle b_m^3 | s_i^3\right\rangle \left\langle b_m^4 | s_i^4\right\rangle \left\langle o_k^1 | L\right\rangle \left\langle o_k^2 | L\right\rangle \left\langle o_k^3 | L\right\rangle \left\langle o_k^4 | L\right\rangle (\left|x_j^1\right\rangle \otimes \left|x_j^2\right\rangle \otimes \left|x_j^3\right\rangle \otimes \left|x_j^4\right\rangle)\\
=&\sum_{m,i,j,k} c_m^b c_{ijk} \left\langle b_m^1 b_m^2 b_m^3 b_m^4 | s_i^1 s_i^2 s_i^3 s_i^4\right\rangle \left\langle o_k^1 o_k^2 o_k^3 o_k^4 | LLLL\right\rangle (\left|x_j^1 x_j^2 x_j^3 x_j^4\right\rangle)
\end{aligned}
$$

Explicitly:

$$
\begin{aligned}
&(\epsilon_{CPM_N^2} \otimes 1_{CPM_S^2} \otimes \epsilon_{CPM_N^2})(\rho^{(2)}(Beirut) \otimes \rho^{(2)}(celebrate) \otimes \rho^{(2)}(Lebanese\ national\ day)) \\
=&(\langle AAAA|AAAA\rangle + \langle AAMM|AAMM\rangle + \langle MMAA|MMAA\rangle + \langle MMMM|MMMM\rangle) \\
&\top^{\otimes^4} \langle LLLL|LLLL\rangle \\
=&4\top^{\otimes^4}
\end{aligned}
$$

The derivation shows clearly that the only terms that impact the meaning of this sentence are the ones related to the city of Beirut, and that all neighbourhoods of Beirut are represented in this sentence. Furthermore, this sentence is true, as expected.

"Beirut that performs". We expect this noun phrase to completely disambiguate the meaning of Beirut, while still subsuming all members.



The meaning of this noun phrase is given by the following derivation:

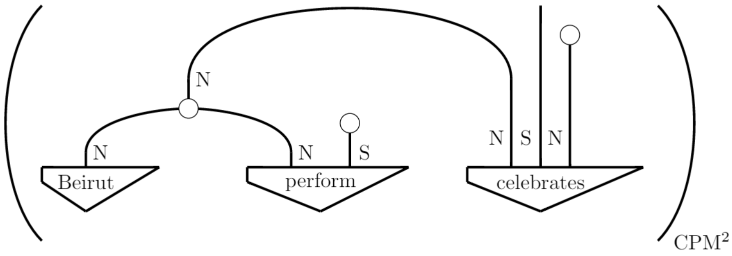
$$
\begin{aligned}
&(\Delta_{CPM_N^2} \otimes \iota_{CPM_S^2})(\rho^{(2)}(Beirut) \otimes \rho^{(2)}(perform)) \\
=&(\Delta_{CPM_N^2} \otimes \iota_{CPM_S^2})(\sum_m c_m^b \, |b_m^1 b_m^2 b_m^3 b_m^4\rangle \otimes \sum_{i,j} c_{ij} \, |s_i^1 s_i^2 s_i^3 s_i^4\rangle \otimes |x_j^1 x_j^2 x_j^3 x_j^4\rangle \\
=&\sum_{m,i,j} c_m^b c_{ij} \Delta_{CPM_N^2}(|b_m^1 b_m^2 b_m^3 b_m^4\rangle \otimes |s_i^1 s_i^2 s_i^3 s_i^4\rangle) \otimes \iota_{CPM_S^2}(|x_j^1 x_j^2 x_j^3 x_j^4\rangle) \\
=&\sum_{m,i,j} c_m^b c_{ij} \delta_{im}^1 \delta_{im}^2 \delta_{im}^3 \delta_{im}^4 \, |s_i^1\rangle \otimes |s_i^2\rangle \otimes |s_i^3\rangle \otimes |s_i^4\rangle \\
=&\sum_{i,j} c_i^b c_{ij} \, |s_i^1 s_i^2 s_i^3 s_i^4\rangle
\end{aligned}
$$

The meaning of "Beirut that performs" is therefore:

$$(\Delta_{CPM_N^2} \otimes \iota_{CPM_S^2})(\rho^{(2)}(Beirut) \otimes \rho^{(2)}(perform)) = ZZZZ + ZZPP + PPZZ + PPPP$$

This clearly shows that "Beirut that performs" disambiguates the meaning of "Beirut" and subsumes all members.

"Beirut that performs celebrates". We expect this sentence to disambiguate the meaning of "Beirut", but to neither be true nor false according to the definition of *celebrate*, since the band celebrates Zach's birthday but not winning "top single".



We treat this as an intransitive sentence with subject "Beirut that performs". The derivation for the meaning of this sentence is similar to the derivation for "Beirut celebrates", and the meaning of "Beirut that performs celebrate" is given by:

$$\begin{aligned}
&(\langle ZZZZ|ZZZZ\rangle + \langle ZZPP|ZZPP\rangle + \langle PPZZ|PPZZ\rangle + \langle PPPP|PPPP\rangle)\\
&(\top+\bot)(\top+\bot)(\top+\bot)(\top+\bot)\\
=&4(\top+\bot)(\top+\bot)(\top+\bot)(\top+\bot)
\end{aligned}$$

We see clearly that the only terms that impact the meaning of the sentence are the terms representing the members, and the resulting sentence is neither true nor false. We can say that this sentence is *unambiguously true and false.*

"Beirut plays in Beirut". In this sentence, the first occurrence of "Beirut" refers to the band and subsumes all band members, and the second occurrence of "Beirut" refers to the city, and subsumes only "Ashrafieh", since the band cannot play in two places at the same time.

The derivation for the meaning of this sentence works like the one for "Beirut celebrates Lebanese national day". Explicitly:

$$\begin{aligned}
&(\epsilon_{CPM^2_N} \otimes 1_{CPM^2_S} \otimes \epsilon_{CPM^2_N})(\rho^{(2)}(\mathit{Beirut}) \otimes \rho^{(2)}(\mathit{play\text{-}in}) \otimes \rho^{(2)}(\mathit{Beirut}))\\
=&(\langle ZZZZ|ZZZZ\rangle + \langle ZZPP|ZZPP\rangle + \langle PPZZ|PPZZ\rangle + \langle PPPP|PPPP\rangle)\\
&\top^{\otimes^4} \langle AAAA|AAAA\rangle\\
=&4\top^{\otimes^4}
\end{aligned}$$

The middle step of this computation shows clearly that the terms related to the band members interact with the subject component of *play-in*, and that neighbourhood "Ashrafieh" interacts with the object component of the verb.

Note: One could compute the meaning of the noun phrases "Beirut who plays in Beirut" and "Beirut that Beirut plays in" to recover respectively only the terms related to the band members and only the term representing Ashrafieh.

"Zach performs" versus "Beirut performs". This example will enable us to better see the subsumption relationships. To make this more interesting, we redefine *perform* to be:

$$|perform\rangle = |Z\rangle \otimes |\top\rangle + |P\rangle \otimes (\frac{1}{2}|\top\rangle + \frac{1}{2}|\bot\rangle)$$

This means that Zach performs all the time, and Paul performs half of the time.

When we compute the meaning of "Zach performs" using the methods described above, we obtain $\langle ZZZZ|ZZZZ\rangle\, \top^{\otimes^4} = \top^{\otimes^4}$. This means that it is always true that Zach performs, as expected by the definition of *perform*.
The computation of "Beirut performs" yields

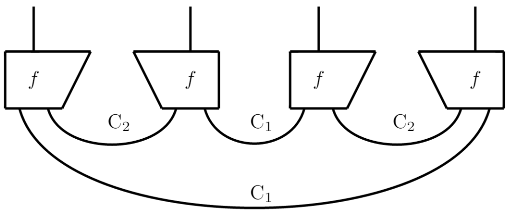
$$\begin{aligned}&\langle ZZZZ|ZZZZ\rangle\, \top^{\otimes^4} + \langle ZZPP|ZZPP\rangle\, \top\top(\frac{1}{2}\top + \frac{1}{2}\bot)(\frac{1}{2}\top + \frac{1}{2}\bot)\\ &+ \langle PPZZ|PPZZ\rangle\, (\frac{1}{2}\top + \frac{1}{2}\bot)(\frac{1}{2}\top + \frac{1}{2}\bot)\top\top\\ &+ \langle PPPP|PPPP\rangle\, (\frac{1}{2}\top + \frac{1}{2}\bot)(\frac{1}{2}\top + \frac{1}{2}\bot)(\frac{1}{2}\top + \frac{1}{2}\bot)(\frac{1}{2}\top + \frac{1}{2}\bot)\end{aligned}$$

which suggests that it is not always true that all members of the band perform all the time, as expected.

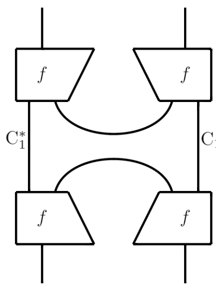
8.3 Measuring ambiguity and entailment

In this section, we provide a measure of the levels of ambiguity and entailment of a word.

Recall from the previous chapter that the structure of a double-density matrix $\rho^{(2)}$ encloses two density matrices: applying transformation-1 to:

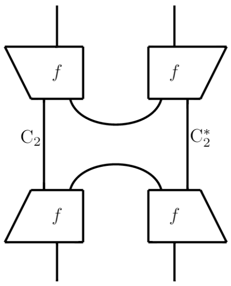


yields density matrix-1:



Recall that C_1 is the ancillary system which results from the first application of the CPM-construction, and that the first application of the CPM-construction is the one that accounts for *ambiguity*.

Applying transformation-2 to the double-density matrix yields density matrix-2:



Recall that C_2 is the ancillary system which results from the second application of the CPM-construction, and that the second application of the CPM-construction is the one that accounts for *entailment*.

Density matrices come equipped with a measure of entropy, or in other words, a measure of *mixedness*. In linguistic terms, density matrix-1 is equipped with a measure of ambiguity, and density matrix-2 is equipped with a measure of entailment. This measure is the von Neumann entropy.

Definition 8.3.1. (Von Neumann entropy). [43] Given a density operator ρ, define the *von Neumann entropy* $S(\rho)$ of ρ by

$$S(\rho) = -tr(\rho \log(\rho))$$

or, given the spectral decomposition of ρ as $\lambda_i |i\rangle \langle i|$, by

$$S(\rho) = -\sum_i \lambda_i \log(\lambda_i)$$

Note that the logarithm in the above definition is typically taken to be base 2, and we use the conventions $0 \log 0 = 0$ and $x \log 0 = -\infty$ for $x > 0$.

The von Neumann entropy $S(\rho)$ is equal to 0 if and only if ρ is pure, and it is maximal and equal to $\log D$ for a maximally mixed state, where D is the dimension of the Hilbert space.

Let us now compute the levels of ambiguity and entailment of "Beirut". For the purposes of this example, we can restrict the set of context words to be {Ashrafieh, Monot, Zach, Paul}. If $\{e_1, e_2, e_3, e_4\}$ is an orthonormal basis for the noun space N, where

$e_1 = (1 \quad 0 \quad 0 \quad 0)^T$, $e_2 = (0 \quad 1 \quad 0 \quad 0)^T$, $e_3 = (0 \quad 0 \quad 1 \quad 0)^T$, and $e_4 = (0 \quad 0 \quad 0 \quad 1)^T$, we set $A = e_1$, $M = e_2$, $Z = e_3$, and $P = e_4$.

$$\begin{aligned}
\rho^{(2)}(Beirut) = &p_1 |AAAA\rangle + p_2 |AAMM\rangle + p_3 |MMAA\rangle + p_4 |MMMM\rangle \\
&+p_5 |AZZA\rangle + p_6 |AZPM\rangle + p_7 |MPZA\rangle + p_8 |MPPM\rangle \\
&+p_9 |ZAAZ\rangle + p_{10} |ZAMP\rangle + p_{11} |PMAZ\rangle + p_{12} |PMMP\rangle \\
&+p_{13} |ZZZZ\rangle + p_{14} |ZZPP\rangle + p_{15} |PPZZ\rangle + p_{16} |PPPP\rangle
\end{aligned}$$

where the p_i's are the probability terms associated to the different components. We denote by $\rho_1(w)$ and $\rho_2(w)$ density matrices-1 and -2 associated to word w. $\rho_1(Beirut)$ is composed of the following terms:

$$\begin{aligned}
\rho_1(Beirut) = &p_1 |AA\rangle \langle AA| + p_2 |MM\rangle \langle AA| + p_3 |AA\rangle \langle MM| + p_4 |MM\rangle \langle MM| \\
&+p_5 |ZA\rangle \langle ZA| + p_6 |PM\rangle \langle ZA| + p_7 |ZA\rangle \langle PM| + p_8 |PM\rangle \langle PM| \\
&+p_9 |AZ\rangle \langle AZ| + p_{10} |MP\rangle \langle AZ| + p_{11} |AZ\rangle \langle PM| + p_{12} |MP\rangle \langle MP| \\
&+p_{13} |ZZ\rangle \langle ZZ| + p_{14} |PP\rangle \langle ZZ| + p_{15} |ZZ\rangle \langle PP| + p_{16} |PP\rangle \langle PP|
\end{aligned}$$

and $\rho_2(Beirut)$ is composed of the following terms:

$$\begin{aligned}
\rho_2(Beirut) = &p_1 |AA\rangle \langle AA| + p_2 |AM\rangle \langle AM| + p_3 |MA\rangle \langle MA| + p_4 |MM\rangle \langle MM| \\
&+p_5 |ZZ\rangle \langle AA| + p_6 |ZP\rangle \langle AM| + p_7 |PZ\rangle \langle MA| + p_8 |PP\rangle \langle MM| \\
&+p_9 |AA\rangle \langle ZZ| + p_{10} |AM\rangle \langle ZP| + p_{11} |MA\rangle \langle PZ| + p_{12} |MM\rangle \langle PP| \\
&+p_{13} |ZZ\rangle \langle ZZ| + p_{14} |ZP\rangle \langle ZP| + p_{15} |PZ\rangle \langle PZ| + p_{16} |PP\rangle \langle PP|
\end{aligned}$$

The measurements in this section are computed using Matlab. We start by computing the ambiguity and entailment measurements of an unambiguous, non-general word. Let us consider the word "Ashrafieh". $\rho^{(2)}(Ashrafieh) = |AAAA\rangle$, and $\rho_1(Ashrafieh) = \rho_2(Ashrafieh) = |AA\rangle \langle AA|$. As expected, we obtain $S(\rho_1(Ashrafieh)) = S(\rho_2(Ashrafieh)) = 0$, because $1 \log 1 = 0$. The same goes for all unambiguous, non-general words.

At this point, it is important to stress that since ρ_1 and ρ_2 are density matrices, their trace should be equal to one. It is therefore crucial to multiply the entries of ρ_1 and ρ_2 respectively by $\frac{1}{tr(\rho_1)}$ and $\frac{1}{tr(\rho_2)}$, in order to make sure their trace is one: we normalise the matrices by their trace.

In what follows, we will deal with different probability terms: p_{city} and p_{band} are the probabilities related to ambiguity, where p_{city} is the probability that "Beirut" refers to

"Beirut city", and p_{band} is the probability and "Beirut" refers to "Beirut band". Naturally, $p_{city} = 1 - p_{band}$. p_{ash} and p_{mon} are the probabilities associated to neighbourhoods Ashrafieh and Monot, where p_{ash} is the extent to which the city of Beirut is represented by Ashrafieh, and p_{mon} is the extent to which the city is represented by Monot. $p_{ash} = 1 - p_{mon}$. We define p_{zac} and p_{pau} in the same way: p_{zac} is the extent to which the band "Beirut" is represented by member Zach, p_{pau} is the extent to which the band is represented by member Paul, and $p_{zac} = 1 - p_{pau}$. The probability terms p_A, p_M, p_Z, and p_P are therefore defined by $p_A = p_{city}p_{ash}$, $p_M = p_{city}p_{mon}$, $p_Z = p_{band}p_{zac}$, and $p_P = p_{band}p_{pau}$. Finally, each p_i in the expressions above is of the form $p_i^1 p_i^2 p_i^3 p_i^4$, where $p_i^1, p_i^2, p_i^3, p_i^4 \in \{p_A, p_M, p_Z, p_P\}$.

Measures of ambiguity. In these examples, we will set $p_{ash} = p_{mon} = p_{zac} = p_{pau} = \frac{1}{2}$, and vary p_{city} and p_{band}.
For $p_{city} = p_{band} = \frac{1}{2}$, $S(\rho_1(Beirut)) = 2$.
For $p_{city} = \frac{2}{3}, p_{band} = \frac{1}{3}$, we expect this number to decrease, because the state is less mixed. In fact, $S(\rho_1(Beirut)) = 1.4439$.

Measures of entailment. In these examples, we will set $p_{city} = p_{band} = \frac{1}{2}$, and vary p_{ash}, p_{mon}, p_{zac} and p_{pau}.
For $p_{ash} = p_{mon} = p_{zac} = p_{pau} = \frac{1}{2}$, $S(\rho_2(Beirut)) = 2$.
For $p_{ash} = p_{zac} = \frac{2}{3}, p_{mon} = p_{pau} = \frac{1}{3}$, we expect this number to decrease, because the state is less mixed. In fact, $S(\rho_2(Beirut))$ is an imaginary number with real part 1.4439 and negligible imaginary part.

The imaginary part comes from the fact that ρ_1 is diagonalisable but not necessarily diagonal: $\rho_1 = VDV^{-1}$, where V is the matrix of eigenvectors and D is the matrix of eigenvalues of ρ_1. $\log \rho_1$ is given by $V \log(D) V^{-1}$. The eigenvectors of ρ_1 are not necessarily real, and $\log \rho_1$ could therefore have complex diagonal entries, thus leading to a complex measurement.
In order to compare measurements, we could either disregard the imaginary part, or define a partial ordering $z_1 \prec z_2$ if and only if $|z_1| < |z_2|$.

Alternatively, we could circumvent the problem of complex numbers by defining yet another measure of entropy: $tr(\rho \odot \rho)$ for normalised ρ, where $\odot$ denotes the Hadamard – or point-wise – product of matrices. $tr(\rho \odot \rho)$ varies between $\frac{1}{D}$ for a maximally mixed state, where D is the dimension of the Hilbert space, and 1 for pure states.

Going back to the measures of entailment, where $p_{ash} = p_{mon} = p_{zac} = p_{pau} = \frac{1}{2}$:
For $p_{ash} = p_{mon} = p_{zac} = p_{pau} = \frac{1}{2}$, $tr(\rho_2(Beirut) \odot \rho_2(Beirut)) = 0.125$.

For $p_{ash} = p_{zac} = \frac{2}{3}, p_{mon} = p_{pau} = \frac{1}{3}$, we expect this number to increase, because the state is less mixed. In fact, $tr(\rho_2(Beirut) \odot \rho_2(Beirut)) = 0.2312$.

Thus we have demonstrated that double-density matrices reliably model ambiguity and entailment in language: words conserve their ambiguity and the extent of their subsumption when context is lacking, and they gradually collapse – i.e. lose their ambiguity or their generality – as more and more context is provided. We have also provided ways to measure levels of ambiguity and entailment in double-density matrices, using density matrices-1 and -2.

Chapter 9

Conclusion and future work

9.1 Summary

In this dissertation, I proposed a further extension to the model of [14] to include *double-density matrices*. Double-density matrices are states in $\mathbf{CPM}^2(\mathbf{C})$, the category resulting from the application of the CPM-construction on $\mathbf{C}$ twice.

The CPM^2-construction was defined and axiomatised in terms of a *squared-environment structure*, following the revised model of [7]. These methods were also generalised to allow the axiomatisation of the CPM^n-construction. The CPM^2-construction framework was showed to preserve †-compact closed structure, which enables the modelling of grammatical relations, and allows for measures of similarity between words and strings of words. Furthermore, $\mathbf{CPM}^2(\mathbf{C})$ possesses †-special commutative Frobenius algebras derived from the †-special commutative Frobenius algebras in $\mathbf{C}$, thus accounting for the modelling of relational types. Theoretically, this framework is adequate in modelling two features of language.

Double-density matrices were investigated in detail: their structure encloses two density matrices, density matrices-1 and -2. These density matrices not only satisfy hermiticity and positive-semidefiniteness, but they are also diagrammatically self-conjugate. The properties satisfied by double-density matrices were then identified, and the entries of density matrices-1 and -2 characterised.

Finally, we demonstrated how double-density matrices account for lexical ambiguity and entailment, and showed that the proposed framework is successful in modelling these features in language by the means of concrete examples. We also provided ways of measuring independently the levels of ambiguity and entailment of words.

9.2 Future work

Many opportunities for future research are available, some of which I discussed in the body of my dissertation. I list here some of the directions for future research:

- In **Chapter 7**, I showed that every morphism satisfying the hermiticity and positive-semidefiniteness conditions have the structure of density matrices. One aspect worth developing – and that I could not investigate due to time limitations – is the following: if the morphisms given by applying transformations 1 and 2 on a state $\varphi : I \to A \otimes A^* \otimes A \otimes A^*$ are hermitian, positive-semidefinite, and diagrammatically self-conjugate, are these conditions enough to show that the original state φ has the structure of a double-density matrix?

- In **Chapter 8**, in order to deal with the problem of having terms in the expression of an ambiguous, general word w that do not correspond to any of the words subsumed by w, I suggested to set all the $\mathcal{E}_i$'s to have maximal cardinality. Other ways of dealing with this issue could be researched, with the investigations geared towards improving the computational efficiency of this solution.

- I suggested extending the $\mathcal{E}_i$'s to have maximal cardinality by repeating some of the subsumed words are revising the probability terms so that the original distribution remains intact. Other ways of extending these sets could be researched, for example by resorting to null vectors.

- It could also be particularly interesting to give the linguistic interpretation of the states of $\mathbf{CPM}^2(\mathbf{Rel})$. I offered the interpretation of some of them in the examples of **Chapter 8**, but did not extend my analysis to all states due to time limitations.

- **Chapter 5** offers an axiomatisation of the CPM^n-construction and suggests that higher orders of iteration account for higher orders of mixing, which could lead to modelling an increasing number of features of language. The next step could be to model three features of language and investigate maps in $\mathbf{CPM}^3(\mathbf{C})$.

Bibliography

[1] 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*. IEEE. 2004, pp. 415–425.

[2] Samson Abramsky and Nikos Tzevelekos. “Introduction to categories and categorical logic”. In: *New structures for physics*. Springer, 2011, pp. 3–94.

[3] G. Donald Allen. *Lectures on Linear Algebra and Matrices*. Texas A&M University, 2003.

[4] Esma Balkır. “Using density matrices in a compositional distributional model of meaning”. PhD thesis. Master’s thesis, University of Oxford, 2014.

[5] Wojciech Buszkowski. “Lambek grammars based on pregroups”. In: *Logical aspects of computational linguistics*. Springer, 2001, pp. 95–109.

[6] Freddy YY Choi, Peter Wiemer-Hastings, and Johanna Moore. “Latent semantic analysis for text segmentation”. In: *In Proceedings of EMNLP*. Citeseer. 2001.

[7] Bob Coecke. “Axiomatic description of mixed states from Selinger’s CPM-construction”. In: *Electronic Notes in Theoretical Computer Science* 210 (2008), pp. 3–13.

[8] Bob Coecke. “De-linearizing linearity: projective quantum axiomatics from strong compact closure”. In: *Electronic Notes in Theoretical Computer Science* 170 (2007), pp. 49–72.

[9] Bob Coecke and Chris Heunen. “Pictures of complete positivity in arbitrary dimension”. In: *arXiv preprint arXiv:1110.3055* (2011).

[10] Bob Coecke and Aleks Kissinger. “Picturing quantum processes”. In: *Book of Abstracts*. 2014, p. 4.

[11] Bob Coecke and Éric Oliver Paquette. “Categories for the practising physicist”. In: *New Structures for Physics*. Springer, 2011, pp. 173–286.

[12] Bob Coecke, Dusko Pavlovic, and Jamie Vicary. “A new description of orthogonal bases”. In: *arXiv preprint arXiv:0810.0812* (2008).

[13] Bob Coecke and Simon Perdrix. “Environment and classical channels in categorical quantum mechanics”. In: *Computer Science Logic*. Springer. 2010, pp. 230–244.

[14] Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. “Mathematical foundations for a compositional distributional model of meaning”. In: *arXiv preprint arXiv:1003.4394* (2010).

[15] Oscar Cunningham. privately handed notes. 2015.

[16] Paul Adrien Maurice Dirac. *The principles of quantum mechanics*. 27. Oxford university press, 1981.

[17] David R Dowty, Robert Wall, and Stanley Peters. *Introduction to Montague semantics*. Vol. 11. Springer Science & Business Media, 2012.

[18] Pavel Etingof et al. “Tensor categories”. In: *Lecture Note for the MIT Course* 18 (2009).

[19] John Rupert Firth et al. *Studies in linguistic analysis*. Blackwell, 1962.

[20] Stefano Gogioso. “A Bestiary of Sets and Relations”. In: *arXiv preprint arXiv:1506.05025* (2015).

[21] Gregory Grefenstette. *Explorations in automatic thesaurus discovery*. Vol. 278. Springer Science & Business Media, 2012.

[22] Brian C. Hall. *Quantum Theory for Mathematicians*. Springer, 2013.

[23] Zellig Sabbettai Harris. “Mathematical structures of language”. In: (1968).

[24] André Joyal and Ross Street. “The geometry of tensor calculus, I”. In: *Advances in Mathematics* 88.1 (1991), pp. 55–112.

[25] Dimitri Kartsaklis et al. “Reasoning about meaning in natural language with compact closed categories and frobenius algebras”. In: *Logic and Algebraic Structures in Quantum Computing and Information, Association for Symbolic Logic Lecture Notes in Logic. Cambridge University Press.(To appear)* (2013).

[26] Anders Kock. “Strong functors and monoidal monads”. In: *Archiv der Mathematik* 23.1 (1972), pp. 113–120.

[27] Joachim Lambek. “The mathematics of sentence structure”. In: *American mathematical monthly* (1958), pp. 154–170.

[28] Joachim Lambek. “Type grammar revisited”. In: *Logical aspects of computational linguistics*. Springer, 1999, pp. 1–27.

[29] Lev. D. Landau. “The damping problem in wave mechanics”. In: *Z. Phys* (1927), pp. 45–430.

[30] Saunders Mac Lane. *Categories for the working mathematician*. Vol. 5. Springer Science & Business Media, 1978.

[31] Daniel Marsden. “A Graph Theoretic Perspective on CPM (Rel)”. In: *arXiv preprint arXiv:1504.07003* (2015).

[32] Diana McCarthy et al. “Finding predominant word senses in untagged text”. In: *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics. 2004, p. 279.

[33] J. von Neumann. “Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik”. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1927 (1927), pp. 245–272.

[34] Robin Piedeleu. “Ambiguity in Categorical Models of Meaning”. PhD thesis. Master’s thesis, University of Oxford, 2014.

[35] Robin Piedeleu et al. “Open System Categorical Quantum Semantics in Natural Language Processing”. In: *arXiv preprint arXiv:1502.00831* (2015).

[36] Anne Preller and Joachim Lambek. “Free compact 2-categories”. In: *Mathematical Structures in Computer Science* 17.02 (2007), pp. 309–340.

[37] Mehrnoosh Sadrzadeh, Stephen Clark, and Bob Coecke. “The Frobenius anatomy of word meanings I: subject and object relative pronouns”. In: *Journal of Logic and Computation* 23.6 (2013), pp. 1293–1317.

[38] Lorenzo Sadun. *Applied Linear Algebra: The Decoupling Principle*. United States of America: American Mathematical Society, 2008.

[39] Hinrich Schütze. “Automatic word sense discrimination”. In: *Computational linguistics* 24.1 (1998), pp. 97–123.

[40] Peter Selinger. “A survey of graphical languages for monoidal categories”. In: *New structures for physics*. Springer, 2011, pp. 289–355.

[41] Peter Selinger. “Dagger compact closed categories and completely positive maps”. In: *Electronic Notes in Theoretical Computer Science* 170 (2007), pp. 139–163.

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

[43] Karol Zyczkowski and Ingemar Bengtsson. “An introduction to quantum entanglement: a geometric approach”. In: *arXiv preprint quant-ph/0606228* (2006).