Mini-project
Hilary Term 2018

MSc. in Mathematics and Foundations of Computer Science

Distributional Models of Meaning

Daphne Wang

The Contextuality of a Text

Abstract

The aim of a computer is (roughly) to translate a sequence of instructions written in some kind of language to a set of mathematical operations. The goal of distributional models of meaning is to some extent to do the opposite. In other word, how can one formally represent the structure of language? In the search of a mathematical model of meaning and more generally of perception, we will investigate various ways one can place a discourse in its context.

We will approach this very broad question from two different point of views. We will first adopt a rather pragmatic approach on contextualisation, namely how to resolve ambiguity for anaphoric-type ambiguity. In the last part of this project, we will investigate perception in a more general way by borrowing some formalism from quantum information theory.

1 Introduction - The story of a DisCoCat

We can see language as having a logical structure (grammar), which is specific to a given language, and meaning which ought to be universal. These two aspects has historically been treated separately, with Lambek's grammar of one side[Lam07], and via Montague semantics[RDEWP80] or distributional approaches (e.g. [MPW16, Cla15]) on the other.

In Lambek's formalism, sentences are modelled in a *pregroup*. Recall that a pregroup $(P, \leq, \cdot, 1, \_^l, \_^r)$ is a partial order $(P, \leq)$ equipped with a monoid structure $(P, \cdot, 1)$ and left and right adjoints satisfying the reduction rules:

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

We will use graphical calculus to make these reductions more readable by writing:

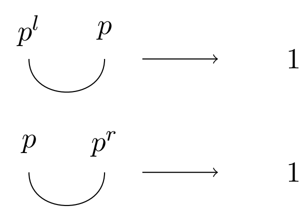
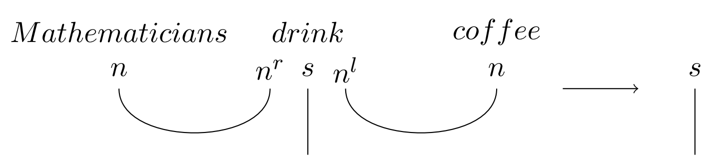


Figure 1: Reduction rules in a pregroup.

Each word on a sentence is attributed a *type* (e.g. n for nouns) such that every grammatically correct sentence reduces to a specific atomic type, the sentence type, usually denote s. Here is an example of a typed transitive sentence:



This assignment of types indeed works since the overall reduces to the sentence type. There are many more examples in [Lam07], but it is not the aim of this project to expand too much on this.

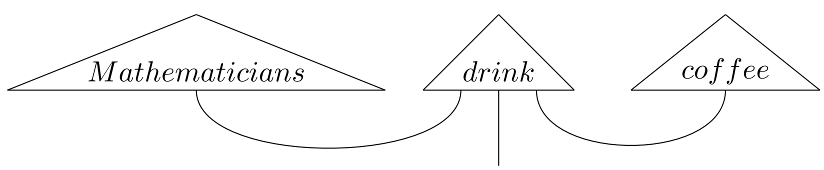
There are also several of interpretations of what "meaning", well, means. The distributional approach is based on a simple concept: similar words tend to appear in similar contexts. We can then relate words by their statistical co-occurrences[Cla15, MSC$^+$13]. For example, the word *cat*, can appear in the same contexts as *dog*, *lion* or *physics*, but will be unrelated to, say *skyscraper*. One could also see a truth-theoretic approach . The main theory on the subject is referred to as the *Montague semantics*. This, in particular, gives a way of dealing with *entailment*[RDEWP80]. For instance:

1. *The mug is on the table.*

2. *The table is in the kitchen.*

entails: *The mug is in the kitchen.*. The theory also provides a (set-theoretic) way of ruling out contradictory statements (e.g. *The mug is full and the mug is empty*) or sentences that are grammatically correct, but do not have a meaning (e.g. *I wait for a waterfall to save*).

The main idea behind *compositional distributional model of meaning* (a.k.a. DisCoCat), introduced by Coecke, Sadrzadeh and Clark in [CSC10] is to combine all of these concepts into a single theory of language: each word is associated a grammatical type coupled with its "meaning" (e.g. element of a vector space or a convex space). As in the Lambek grammar, types admit left and right adjoints so that they can reduce to a sentence type. The representation adopted is similar to the one used in [CK17] to describe quantum processes, and for instance, the previous example will be depicted as:



In terms of evaluating the meaning of a sentence, say an easy transitive sentence, we can see verbs as a relation between the the subject and the object of the sentence. For example, let's consider a universe where the only nouns are $\{Alice, Bob, Mathematicians, coffee\}$, and suppose that:

1. Alice likes Bob.

2. Alice likes coffee.

3. Mathematicians like coffee.

Hence, we can define the meaning verb *like* as:

$$|like\rangle = |Alice\rangle \langle Bob| + |Alice\rangle \langle coffee| + |Mathematicians\rangle \langle coffee|$$

using the "braket" notation, also taken from the quantum mechanics formalism. We obtain a truth-value of a sentence by computing the inner product. In this example, *Alice likes*

Bob will evaluate to 1 (i.e. true) whereas *Bob likes coffee* will evaluate to 0 (i.e. false). Similarly, predicate representation of words are adopted in [SCC14]. In this model, the proper nouns corresponds to the atomic elements of a universe $\mathcal{U}$; nouns, adjectives, non-transitive verbs, etc are represented as unary predicates (e.g. $[\![Human]\!] = \{x \in \mathcal{U} | x \text{ is a human}\}) = \{Alice, Bob, Mathematicians\}$), and transitive verbs are binary predicates (e.g. $[\![like]\!] = \{(x, y) \in \mathbb{U} | x \text{ likes } y\} = \{(Alice, Bob), (Alice, coffee), (Mathematicians, coffee)\}$). We will find this representation particularly useful in the subsequent sections.

We will in this project focus on the meaning aspect rather than the grammatical structure. We will first give a quick review of previous work on modeling meaning and cognition.

Meaning spaces Within the distributional framework, the first naive approach is store the meaning of a word in a vector(see [MSC⁺13, Cla15]). The main advantage of this method is that vector spaces are computationally convenient to implement; an example of such implementation is the `word2vec` programm based on the model introduced by Mikolov *et. al.* in [MSC⁺13]. This approach leads a very natural geometric interpretation of the lexical space, and furthermore gives fairly good results. For example by comparing the distance and direction of words on the obtained vector space, [MPW16] showed that it is possible to obtain word association such as:

- Paris is to France what Helsinki is to Finland.
- Girl is to boy what princess is to prince.

but also yielded some unexpected examples. In particular, it is difficult in this framework to differentiate antonyms since they will be very likely to appear in the same context (e.g. a sentence containing the word *love* will in most case still make sense if *love* is replaced by *hate*, even though their meanings can be considered as opposite). In addition, it might not be very convenient conceptually to represent the meaning of words by a (very high dimensional) vector. In the work of Gärdenfors (see [Gär04, Gär11]), each word is modeled as a *convex space*. Each noun is given some properties (a basis), and can be expressed as the convex combination of those properties. In this scenario, verbs can acts on objects (nouns) by modify the weight of one of its properties. This concept has also been extended in compositional distributional models of meaning in [BCG⁺17].

Dealing with ambiguity Conversations on a daily basis shows that ambiguity is omnipresent in language. Moreover, we can distinguish several degrees of ambiguity:

- Unrelated words which share the same spelling (*homonymy*). For example the word *bank* could refer to the financial institution, or the edge of a river.
- Words that are related, share the same spelling but represents different things (*polysemy*). Continuing the previous example, the word *bank* can mean the abstract concept of the institution, or the physical building.
- An even more subtle type of ambiguity would be at the level of a sentence. We can for example consider a sentence where all words are well-defined and not ambiguous, but for which the meaning will depend on the previous sentences, or a more general context. For example:

Alice and Bob speak French and English.

There is no lexical ambiguity here, but are both Alice and Bob bilingual or is Alice speaking French, while Bob speaks English?

The first two points were treated in [PKCS15], while the second can to some extent be resolved by taking intonation into account[KS15].

In the case of homonymy and polysemy, one can view ambiguous words as a probabilistic mixture of two or several non-ambiguous meanings − note already the parallelism with quantum mechanics. The work presented in [PKCS15] is based upon the original work of Piedelieu[Pie14] which uses concepts of quantum mechanics to model ambiguity: the meaning of a word can then be represented as density matrices via the **CPM** construction[Sel07], i.e. doubling[CK17]. In this framework, if non-ambiguous words are considered as *bases states*, then polysemous words are depicted as a mixture of those basis states, but which is still a *pure state*, while homonymous words corresponds to the classical mixture of pure states, i.e. *mixed states*. This indeed gives rise to a much richer structure than standard probabilistic mixing. Furthermore, polysemous and homonymous words are intrinsically different[PKCS15]: polysemous words do refer to roughly the same concept, but need to be put in context to "select" one of its instantiation. We might note, as remarked by Marsden in [Mar17], that working in the category **CPM**(**Rel**) (rather than **CPM**(**FHilb**)) as in [PKCS15] can lead to anomalies, e.g. convex combination of mixed states giving a pure state. In response to that, Marsden proposed in the same paper ([Mar17]) an alternative way of dealing with ambiguity using the power-set monad (for unquantified ambiguity) and the distribution monad (for quantified ambiguity), and also introduces $\mathbf{Set}_\bullet$-enriched categories to model incomplete information so that $\perp$ refers to a "I-don't-know" element. We may also note that the **CPM**-construction is also used to describe entailment (see [PKCS15, BCLM16]), hence to describe both ambiguity and entailment, one need to apply the **CPM**-construction twice (see [PKCS15, AC16]).

From a different perspective, Kartsaklis and Sadrzadeh exposed how intonation can be modelled within the distributional compositional of language in [KS15]. The main idea is that a sentence has a *rheme* part and a *theme* part. The emphasis of the sentence is on the former, which is the part giving information; the latter just places this information in context. The intonation boundary is denote by the token $\triangleleft$ (or $\triangleright$ depending on the sentence) such that:

$$rheme \triangleleft theme$$

For example, using the example used in [KS15], we can differentiate:

1. *Mary likes* $\triangleright$ *musicals.*

2. *Mary* $\triangleleft$ *likes musicals.*

In the first case, the emphasis is put on *musicals* and hence can be an answer to the question *What does Mary like?*, while in the second case *Mary* is the main information, so this is a suitable answer to *Who likes musicals ?*.

Even though the link with ambiguity is not completely obvious here, one can see that this method can be used to resolved some ambiguous statements. For example the sentence:

The lady hit the man with an umbrella.[1]

might mean that *the lady is hitting a man who has an umbrella*, or that *she is hitting a man using an umbrella.* However, by marking the intonation boundary as:

1. *The lady hit* ▷ *the man with an umbrella.*
2. *The lady hit the man* ▷ *with an umbrella.*

the ambiguity is resolved: 1. corresponds to the statement when the man has an umbrella, and 2. is the situation where the lady has the umbrella.

1.1 Intuition and motivation

At the moment, most of the work has been done locally, in other words, within one or two sentences. Hence, the goal of this project is to have a better idea of what is happening in a more global scale. Indeed, a single sentence can have a lot of different interpretations in different contexts. As previously discussed, in many cases, this ambiguity can be resolved by putting the intonation at the right place. Now, let's consider the sentence:

I'm ready.

This may equally mean *I'm ready to go for lunch*, or *I'm ready to handle whatever is gonna happen next.* In the first interpretation, "I" is the one acting, whereas it has a more passive role in the second one. Hence, intonation here is not enough to resolve the ambiguity and only the whole context will determine the meaning of the sentence.

The first aim will then be to describe the structure of a text rather than sentences alone. Some work has been done in this direction in [CdMT18, Coe19, AS14], and we will talk about this in more details in the next section. The second idea will be to have a (rather philosophical) discussion about how a text is *perceived*, and how external parameters may interact with attribution to meaning.

As a side note, this project will use a lot of different concepts borrowed from the formulation of quantum theory, and I would like to justify and motivate this choice. One of the obvious reason is that the majority of the literature on categorical models of meaning is based on monoidal categories, which is also used to describe quantum systems[AC08], and moreover, the graphical notation used is clearly inspired from [CK17]. In addition, real life and our perception of the world has more in common with quantum mechanics than one realises and all paradoxes can have a very natural interpretation when it comes to our everyday experience:

- **Non-locality**. In quantum mechanics, if two spacelike separated parties share an entangled state, local operation made by one party changes the state of the other, this is known as non-locality. In the case of language, the truth of a sentence might be altered by something or someone outside of our reach and upon which we have no control.

[1]This example is shamelessly taken from `https://mw.lojban.org/papri/Ambiguous_sentences_in_English`.

- **Probabilistic nature and terminality of measurements**. It is said and has been proven that nature as described by quantum mechanics is not deterministic and the same state can lead to two different measurements. However, once a measurement has been made, the outcome is definite and every subsequent measurements will lead to the same result. Similarly, it is very natural to think of the future as open to many possibilities; but once a decision is made, this cannot be changed.
- **Complementarity and non-commutativity of measurements.** The uncertainty principle in quantum mechanics states that it is not possible to simultaneously measure the position and the momentum of a particle. In psychology, is it also quite clear that, for example, the process of making judgments shouldn't be commutative.
- etc.

This intuition has already been used in several areas outside of physics[Khr15] and notably led to interesting experimental results for example in psychology and cognitive science (see for example [DK12, CTF$^+$07, Khr15]).We'll give more details about this in the last section.

2 The meaning of a discourse

> There's no sense in being precise when you don't even know what you're talking about.
> *John von Neumann*

It is quite clear that a piece of text or a given discourse can only make sense when put in context. The first thing to consider is the basic example of anaphora. It is indeed very common to make references within a discourse it order to make it less complicated. Indeed, rather than saying:

John is hungry. John is going to Tesco to buy food.

It sounds more natural to say:

John is hungry. He is going to Tesco to buy food.

instead. The question now is how do we mentally make the link between He and $John$?

The formalism used in the next section is largely inspired by the work of Abramsky on contextuality (see [ASK$^+$15]).

2.1 Texts in context

Recall that we are trying to get a more *global* view of the structure of a discourse rather than (*local*) sentences alone. A standard way of doing so is by using *sheaf theory*[MM94]. Sheaf-theoretical techniques are already in use in the domain of *contextuality*, which relates to quantum mechanics non-locality [ASK$^+$15] (but not only! See for example in [Abr14] where it is also applied to the structure of databases). Let us briefly describe the method from[ASK$^+$15].

Given a poset $(P, \leq)$, we define a *presheaf* of P as a contravariant functor:

$$\mathcal{E} : P^{op} \to \mathbf{Set}$$

An object s of $\mathcal{E}(x)$, $x \in P$ is called a *section* over x. We say that a presheaf is a *sheaf* if it satisfies the *gluing condition*:

$$s_i|_{x_i \wedge x_j} = s_j|_{x_i \wedge x_j}$$

for every two sections $s_{i,j}$ over the elements $x_{i,j} \in P$. In the context of quantum non-locality, we define the set X as the set of all possible measurements given a particular situation. We denote by $\mathcal{M}$ the set of all measurements that can be made jointly[2],i.e. the *measurement contexts*, and O the possible outcomes of these measurement contexts. We then define the presheaf:

$$\mathcal{E} : \mathcal{P}(X) \to O^X \subseteq \mathbf{Set}$$

which is trivially sheaf when X is finite[ASK$^+$15]. And the functor[3]:

$$\begin{aligned} \mathcal{D}_{\mathbb{B}} : \mathbf{Set} &\to \mathbf{Set} \\ X &\mapsto \{\phi : X \to \mathbb{B} | \phi \text{ distribution over } (\mathbb{B}, \vee, 0, \wedge, 1)\} \end{aligned}$$

with morphisms:

$$\begin{aligned} \mathcal{D}_{\mathbb{B}}(f) : \mathcal{D}_{\mathbb{B}}(X) &\to \mathcal{D}_{\mathbb{B}}(Y) \\ \phi &\mapsto \left(y \mapsto \bigvee_{f(x)=y} \phi(x) \right) \end{aligned}$$

for every $f : X \to Y$. Note that $\mathcal{D}_{\mathbb{B}}\mathcal{E} : \mathcal{P}(X)^{op} \to \mathbf{Set}$ is a presheaf. Now, we say that a model $e \in \mathcal{D}_{\mathcal{B}}\mathcal{E}(X)$ is *contextual* if there exists a section $s \in \mathcal{E}(C)$, for some context $C \in \mathcal{M}$, which cannot be consistently extended to a *global assignment* $t \in \mathcal{E}(X)$ which is coherent for each context $\tilde{C} \in \mathcal{M}$. Note that such t is called a global assignment if there exists $C \in \mathcal{M}$ s.t.:

$$t|_C = s$$

and:

$$e(t)|_{\tilde{C}} = 1$$

for all $\tilde{C} \in \mathcal{M}$. More visually, the segment in red in Fig.2 cannot be extended to a closed path.

[2]This needs to forms a cover of the set X.

[3]This definition can be generalised by replacing $\mathbb{B}$ by any semiring $(R, +, 0, \times, 1)$. We will also describe the probabilistic case over $(\mathbb{R}_{\geq 0}, +, 0, \times, 1)$ in Section 2.3.

	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
(a,b)	1	1	1	1
(a',b)	0	1	1	1
(a',b')	1	1	1	0
(a,b')	0	1	1	1

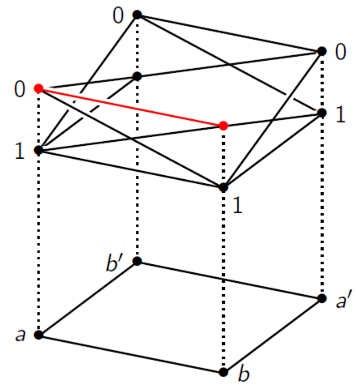
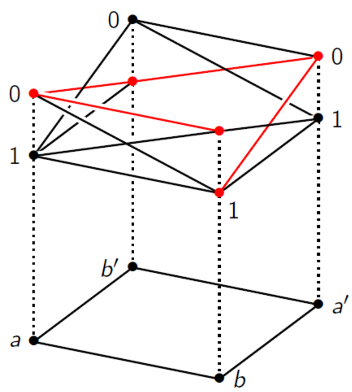
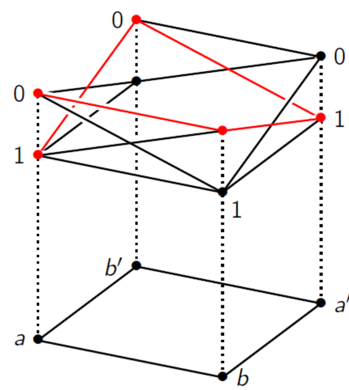


Figure 2: The Hardy model.

Indeed, one can try to extend the section (by brute force) as:





to see that any global assignment is not globally consistent.

A model is said to be *strongly contextual* if none of the local section can be extended to a global assignment, for example:

	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
(a,b)	1	0	0	1
(a',b)	1	0	0	1
(a',b')	0	1	1	0
(a,b')	1	0	0	1

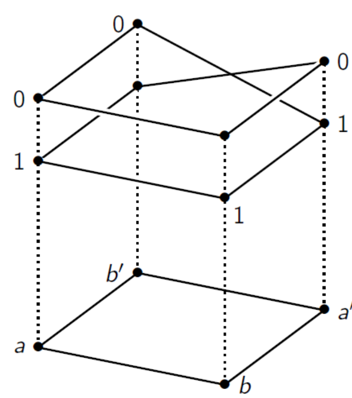


Figure 3: The PR-box.

In this example, the context (a',b') insures that no local section can be consistently extended to a global one.

So, this is nice. But why is it relevant to linguistics?

Application to texts There has been already a few attempts to obtain of way of dealing with a discourse, for example in [AS14, CdMT18]. We see in both those references that the first step towards a model of discourse is via *anaphora*, for example:

John has a cat. It's fluffy.

Here, *it* obviously refers to John's cat; but note that the sentence *It's fluffy.* alone, is not enough to determine what *it* refers to. This is a very common and intuitive figure of speech, but rather hard to model (for example, Montague semantics usually fails[AS14]). The approach used in [AS14] combines the formalism of *Discourse Representation Theory* (DRT) of [KVGR11] and Abramsky's sheaf-theoretic model of contextuality previously described. In DRT, a text is represented as a set of variables (words), and a set of conditions which dictates how these variables interact with each other; these two sets together gives the Discourse Representation Structure (DRS) of the text. Let's consider once again the simple example:

John has a cat. It's fluffy.

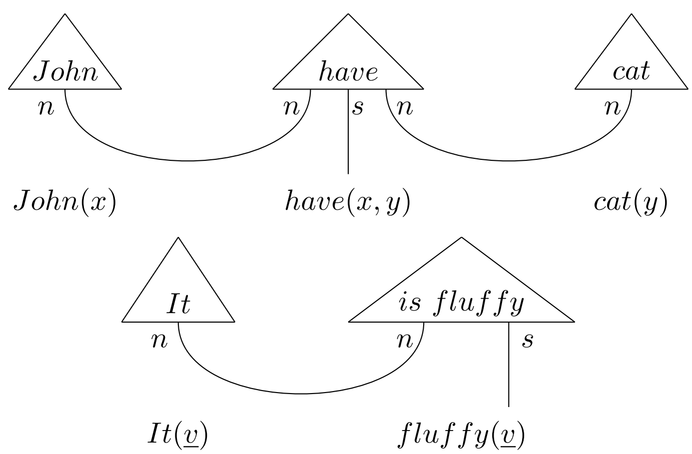
Then we decompose the two sentences as:

$$John\ has\ a\ cat \quad \mapsto \quad (\{x,y\},\{John(x), have(x,y), cat(y)\})$$

and:

$$It\ is\ fluffy \quad \mapsto \quad (\{\underline{v}\},\{It(\underline{v}), fluffy(\underline{v})\})$$

where we underline the variable(s) that needs to be assigned a meaning. We also note that we can get the conditions (number of conditions and types) from the grammatical type of the word, e.g.:



(for simplicity, we treat every type as self-adjoints, and the phrase *to be* + *adj.* as a non-transitive verb). In this procedure, we ignore the sentence type, associate each noun with a variable, and "follow the wires" to get conditions on the composite types. The DRS of the whole text is then the disjoint union of the two, i.e.:

$$(\{x,y,\underline{v}\},\{John(x), cat(y), It(\underline{v}), have(x,y), fluffy(\underline{v})\}$$

The role of sheaf theory is then to "glue" sentences together. The game is then to find a global assignment of all the "unknown" variables so that the discourse makes sense. Inspired by the predicate approach used in [SCC14], we define our universe $\mathcal{U}$ to be the set of all characters and object that have a definite identity. In the previous example, we would have:

$$\mathcal{U} = \{John, John's\ cat\}$$

We then define each unknown noun as an unary predicate satisfying certain properties. For example we would have:

$$[\![It]\!] = \{x \in \mathcal{U} | x \text{ is as object, unnamed animal, abstract thing etc...}\}$$

This is quite trivial in the previous example as it is clear that:

$$[\![It]\!] = \{John's\ cat\}$$

Then, in this case, resolving the ambiguity is trivial: the only possible assignment of the (only) unknown variable $\underline{v}$ is $\underline{v} \mapsto y$ or:

$$It \mapsto John's\ cat$$

For more complicated example, we will consider the *Čech complex* (see Fig.4) of all the unary predicates (e.g. nouns, adjectives and non-transitive verbs) as the set of all possible measurements, and the binary predicates (e.g. transitive verbs) as measurement contexts.

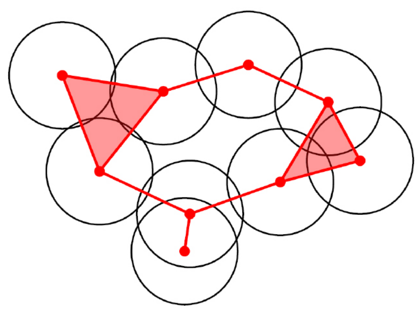


Figure 4: Example of a Čech complex.

The possible outcomes then corresponds to the known elements of $\mathcal{U}$. Let's illustrate this by a simple example:

John has a cat called Arnold. He is meowing at him.

This text has the following DRS:

$$(\{x, y, \underline{u}, \underline{v}\}, \{John(x), Arnold(y), He(\underline{u}), him(\underline{v}), Have(x, y), Meow(\underline{u}, \underline{v})\}$$

In this case, we see that both *He* and *him* can refer to both *John* and *Arnold the cat*; however, it seems reasonable to say that the verb *to meow* can only have a cat as the subject. Hence, assuming that *him* and *He* are different characters, this completely resolves the ambiguity:

	(x, x)	(x, y)	(y, x)	(y, y)
$Meow(\underline{u}, \underline{v})$	0	0	1	0

So the only possibility is $\underline{u} = y$ and $\underline{v} = x$; in other words:

$$He \mapsto Arnold$$
$$Him \mapsto John$$

However, let's consider the following example:

John has a cat called Arnold. He is happy.

In this case, we see that we need more information to resolve the ambiguity. Indeed, *He* can equally refer to *John* or *Arnold the cat*; but if we knew that *John* always wanted a cat or was sad and lonely, then we could conclude that $He \mapsto John$, or on the other hand if *Arnold* just got fed the most probable assignment would be $He \mapsto Arnold$. But in both cases, the two meanings are perfectly reasonable. Now, if we consider the following example:

John has a cat called Arnold. She's fluffy.

There is something not quite right there. Indeed, with the information that we are given, *She* doesn't seem to refer to anything (assuming obviously that *Arnold* needs to refer to a male character). This is the equivalent of the *strong contextuality* concept of [ASK$^+$15]. Note that if, borrowing the terminology of [ASK$^+$15], if a given discourse is logically contextual, but not strongly contextual (i.e. at least one global assignment exists), then we can just restrict to the consistent assignments. We can then classify the inconsistency of a given discourse as:

- A discourse is *contextual* if it is strongly contextual in the sense of [ASK$^+$15], i.e. if it is incoherent;
- If there exists a unique global assignment which makes the text coherent, then the discourse is *contextually non-ambiguous*;
- If several global assignments are possible, then the discourse is *contextually ambiguous*, and obtaining more information (i.e. more sentences) could help resolving the ambiguity.

As a concluding remark for this section, I have discovered, while writing this report, a paper by Zadrozny and Garbayo[ZG18] using similar ideas to identify contradictions and disagreement in related discourses.

2.2 Causal structure of a text

Now, let's consider the example:

(S_1) *John and his brother have exams.*
(S_2) *John is not done yet.*
(S_3) *He needs to revise.*
(S_4) *His brother has finished his.*
(S_5) *He is on holidays.*

As before, we ask for the sentences (S_3) and (S_5), who is *He*? If we isolate these two sentences:

He needs to revise.
He is on holidays.

We can see that in both cases, both *John* and *John's brother* are possible (both male humans; both having exams); so the previous formalism is not enough to resolve the ambiguity here. However, when reading the text, it is quite obvious who each *He* refers to (namely *He* is *John* in (S_3) and *John's brother* in (S_5)). We are now gonna introduce an additional structure on text, the notion of time passing.

We first note that some work has been done already in the direction of expanding the formalism of DisCoCat to the scale of a text: DisCoCirc, which has been developed in [Coe19] by Coecke. In this model, sentences are allowed to commute if they have no influence on one another. For example:

The sky is blue. The phone is ringing.

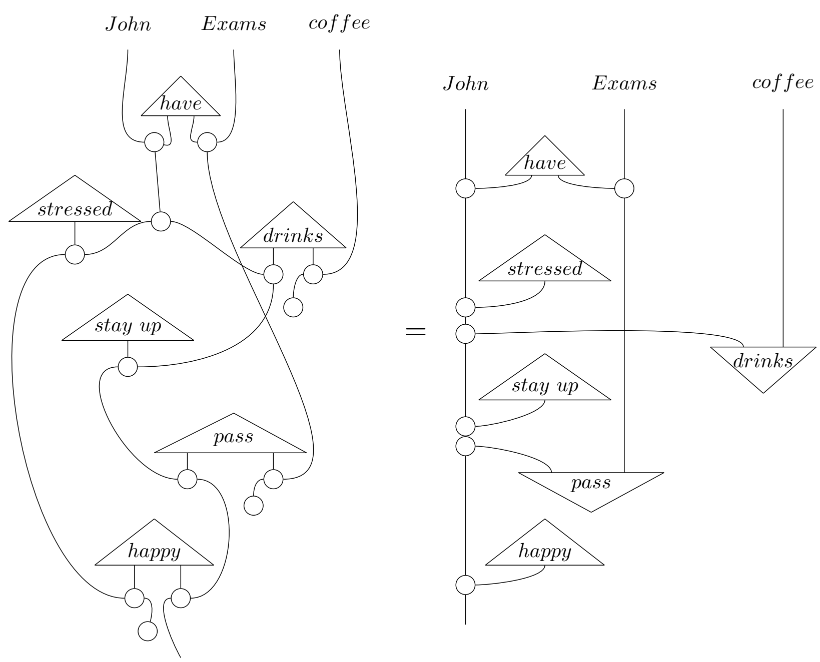
Or more importantly, they are allowed to have an influence on one another, for instance:

It's raining. I should be an umbrella.

The main idea of the paper is that sentences are not states, but rather *processes* acting on the protagonists and objects; verbs and relationships are then represented as connectives. For example:

(S_1) *John has exams.*
(S_2) *He is stressed.*
(S_3) *He drinks coffee.*
(S_4) *He stays up all night.*
(S_5) *He passes his exams.*
(S_6) *Now, John is happy.*

is presented as the following circuit:



Also note that the "spiders" are connected to the subject of each verb. Now we see from the LHS that by combining each sentences in blocks, this gives the structure of a Directed Acyclic Graph (DAG) representing the *causal structure* of the text (see Fig.5).

We can then use this structure and apply it to refine the construction defined in the previous section.

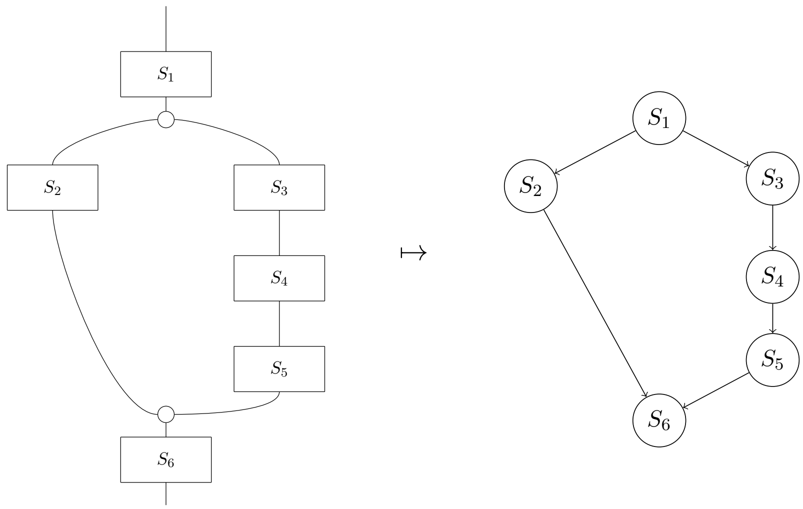
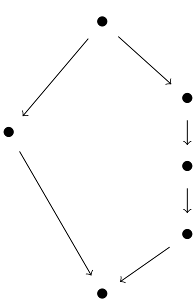


Figure 5: DAG structure of a text.

Let $\mathcal{T}$ be the causal structure associated with a text, for example:



where all the nodes are sentences. Then we note that $\mathcal{T}$ is a (finite dimensional) poset. Hence, we can define the presheaf $\mathcal{F}$ over $\mathcal{T}$ as:

$$\mathcal{F} : \mathcal{T}^{op} \to \mathfrak{S} \subseteq \mathbf{Set}$$

where $\mathfrak{S}$ is the set of sentences of a given a text, satisfying the standard presheaf conditions (note that since $\mathcal{T}$ is finite, then this is trivially a sheaf). As for a general presheaf, we need the $\_^{op}$ since any sentence needs to be coherent with other sentences which have influence *on* them; and hence, this corresponds going backwards in time to "check" consistency with previous events or facts. Hence, we will then be interested in the existence of global sections over the presheaf $\mathcal{D}_{\mathbb{B}}\mathcal{E}\left(\mathcal{F}(\mathcal{T})\right)$. Now, recall the example:

(S_1) *John and his brother have exams.*

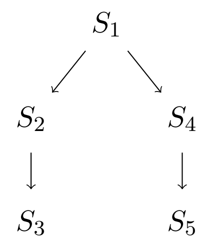
(S_2) *John is not done yet.*

(S_3) *He needs to revise.*

(S_4) *His brother has finished his.*

(S_5) *He is on holidays.*

Then, the corresponding causal order is:



Hence, we see that this indeed resolves the ambiguity. Indeed, since (S_3) needs to be compatible with (S_2), then He in (S_3) needs to refer to $John$; and similarly, compatibility of (S_5) w.r.t (S_4) implies that the second *He* corresponds to *John's brother*.

2.3 Probabilistic contextuality

So far, we have only considered *possible* situations, i.e. even though some assignments might sound absurd, we cannot rule them out since there might be a context for which the interpretation is realisable. For example:

John is listening to a song on his phone. He likes it.

Now obviously it is more *probable* that *John likes the song* rather than *John likes his phone*; even though both are possible. Therefore, it would be of interest to add some kind a probability notion into this framework.

Probabilistic meaning As briefly mentioned before, we could, instead of working over the booleans, work over any semiring. In particular, to get back a probabilistic interpretation, we could, and will, work over the $(\mathbb{R}_{\geq 0}, +, 0, \times, 1)$ semiring. This approach has been briefly described in [AS14], along with some empirical results. We'll adopt a more naive approach here: let's assume that in general, people are 70% likely to say that they like a song, rather than their phone. Hence, writing the previous example DRS as:

$$(\{x, y, z, \underline{u}, \underline{v}\}, \{John(x), song(y), phone(z), He(\underline{u}), it(\underline{v}), listen(x, y), like(\underline{u}, \underline{v})\})$$

	(x,x)	(x,y)	(x,z)	(y,x)	(y,y)	(y,z)	(z,x)	(z,y)	(z,x)
like	0	0.7	0.3	0	0	0	0	0	0

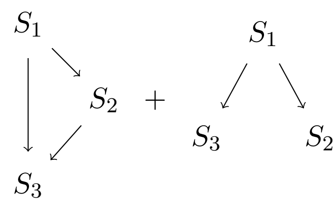
Probabilistic time So far we have also assumed that causal orders were definite. But is this always the case? Let's consider the example:

(S_1) *John and his friend are walking on the street.*
(S_2) *He gets distracted by a pigeon.*
(S_3) *He trips over a rock.*

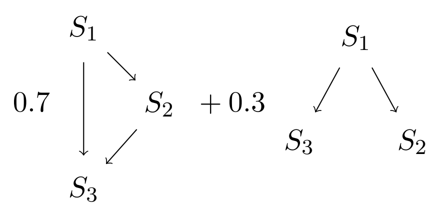
Here, there is an ambiguity on whether each He corresponds to $John$ or *his friend*. How does the causal order look like? Well, it can either one of:

(a) (b)

In the case of (a), the same person gets distracted: the character trips because of the pigeon and because he is walking on the street. In the second case, *He* is S_2 and S_3 refer to different people; the two sentences S_2 and S_3 do not have any influence on each other but both depend on the fact that *John and his friend* are on the street. Hence, we would like to write the causal order as:



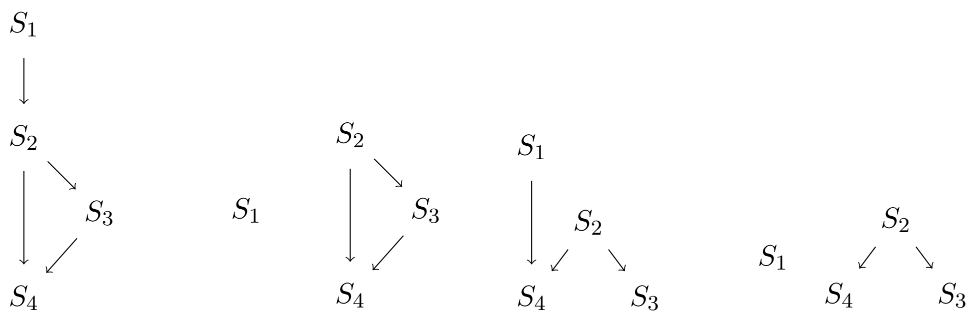
Furthermore, these two causal order are not born equal; clearly it is more likely that the same person, whether it be *John* or *his friend*, trips over the rock because of the pigeon. Hence we can add probabilities on top of the superposition as for example:



Interaction of probabilities The next obvious question is how does those two "add-on" probabilities interact. In the case when the causal order, or each possible causal order, makes a text completely unambiguous, this is pretty straight forward. Let's consider the example:

(S_1) *John is clumsy.*
(S_2) *John and his friend are walking on the street.*
(S_3) *He gets distracted by a pigeon.*
(S_4) *He trips over a rock.*

Hence, this gives rise to the following possible causal order:



(a) (b) (c) (d)

Which completely resolve the two ambiguities:

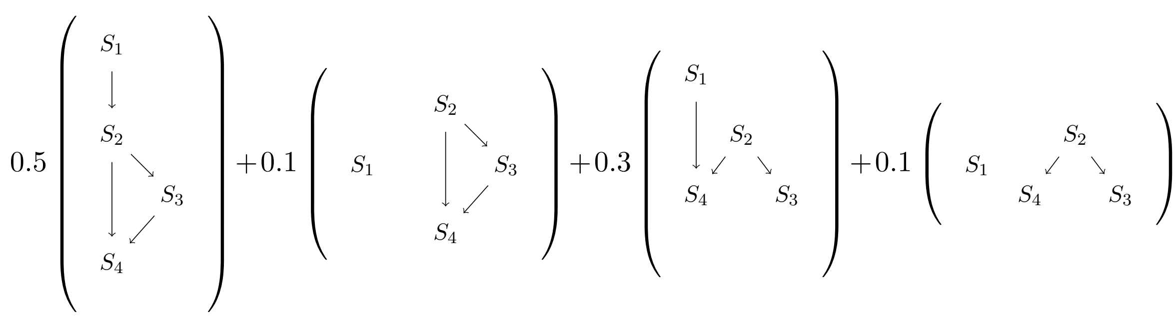
(a) $[S_3 : He \mapsto John;\ S_4 : \quad He \mapsto John]$

(b) $[S_3 : He \mapsto friend;\ S_4 : \quad He \mapsto friend]$

(c) $[S_3 : He \mapsto friend;\ S_4 : \quad He \mapsto John]$

(d) $[S_3 : He \mapsto John;\ S_4 : \quad He \mapsto friend]$

Now, if we write the overall (indefinite) causal order as:



We then obtain the table of probabilities as the standard union of events:

	John	*friend*
Distracted	0.5+0.1=0.6	0.1+0.3=0.4
Trips	0.5+0.3=0.8	0.1+0.1=0.2

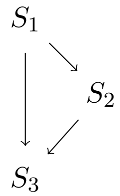
This example is easy because given a definite causal order, each unknown comes with probability of either 0 or 1. What about a more general case, for example:

(S_1) *John and his friend are walking on the street.*

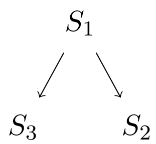
(S_2) *He gets distracted by a pigeon.*

(S_3) *He trips over a rock.*

Here, there is ambiguity in both the causal order and the assignments with each definite causal order. *But*, we do have extra conditions, namely:



$$\Rightarrow (S_2 : He, S_3 : He) \mapsto (John, John) \vee (friend, friend)$$



$$\Rightarrow (S_2 : He, S_3 : He) \mapsto (John, friend) \vee (friend, John)$$

Note that this is weirdly similar to the Bells states[CK17]:

$$B_0 = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$
$$B_1 = \frac{|10\rangle + |01\rangle}{\sqrt{2}}$$

i.e. imposing causal structures creates entanglement of sentences.

From this example, we see that it would make more sense to consider contextual ambiguity as the *quantum* superposition of references, and the causal structure ambiguity as *classical* superposition. In this case, the previous example can be represented as the mixed state:

$$|\psi\rangle = |\alpha\rangle \otimes (0.7\,|B_0\rangle + 0.3\,|B_1\rangle)$$

where $|\alpha\rangle$ is the non-ambiguous part of the text.

All of this is useful for, say a text, where everything and everyone lives in a self-contained universe; in this universe everything is pretty much pre-determined by whoever is writing the text. In order to deal with conversation or a more interactive type of discourse, in particular how a text is perceived, we need add the past and the current "state" of the person reading the text, or listening to a conversation. We will explore this aspect of this topic in the next section.

3 A word on perception

> For, after all, how do we know that two and two make four? Or that the force of gravity works? Or that the past is unchangeable? If both the past and the external world exist only in the mind, and if the mind itself is controllable – what then?
> *1984 - G. Orwell*

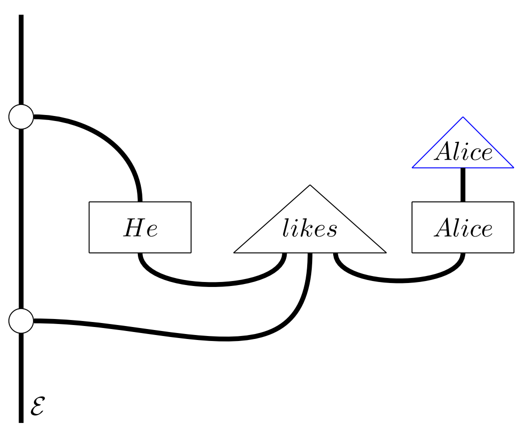
We have already moved from understanding a sentence to understanding a collection of sentences. To get a grasp of perception, we want to move to one level up. Why is this important? We see already from the discussion about probabilistic contextuality that it is very subjective to "attribute" weights to either meaning or causal structures. The whole concept of *meaning* is itself very subjective and individual-dependent. This section is going to be rather informal compared to the previous one; its aim is simply to give a more metaphysical discussion about perception.

3.1 Are brains quantum open systems?

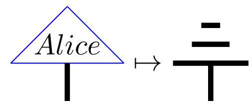
The context is essential in our idea of what's happening or more generally our perception of the world. Reading the same text, people will have different pictures in their head, and for example, a film adaptation will never be the same as picturing a book in our heads. This arguably involves more than just memories, for example the mood of the person receiving a discourse, and its physical environment - a discourse may not have the same impact from a cold and dark place rather than from a sunny tropical island.

Furthermore, as more and more information is given, the environment itself becomes more and more precise. Hence it would make sense to store all information that we know *within* the environment.

The idea is that there is a dual interaction between the discourse and what we previously called the environment, i.e. our internal state, even at the level of one sentence. When listening to a reference (e.g. see anaphora is the previous section), the mind will search through the memory to obtain the "pointer" to the reference (either know or the combination of all possible references that we know of). An similarly, after the listening to the sentence, then we have obtained new information, which will update our internal state. We then depict the said environment as an ancilla:

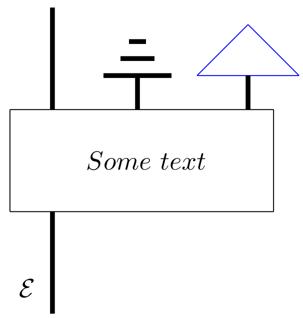


This shows an example of a general sentence, with one reference that needs to be "asked" to the environment (which we call $\mathcal{E}$), and one new piece of information in blue. Here, we assume, that there is only one *Alice* in the world so that this name is completely unambiguous. If however there has not been a reference to any *Alice* before and that we have no idea who this refers to, we can replace it by a complete uncertainty, e.g.:

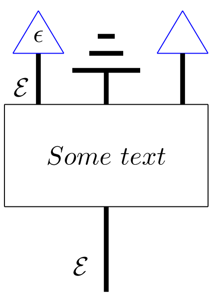


Also note that even if we are using the notation of [CK17], time is going *down*.

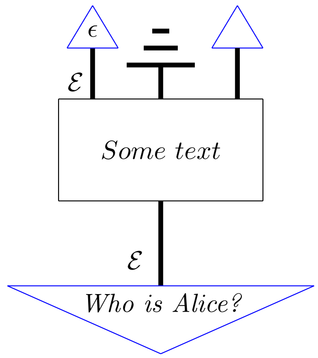
Hence a whole discourse depicted as a process will be of the general form:



If we want to impose a initial internal state ϵ (including for example the past and mood of the individual), we then obtain:



This is very similar (pictorially and conceptually) to the concept of quantum open systems as used in quantum information. In quantum information theory, open systems are used to obtain a more realistic model of quantum processes, by introducing possible external effects which can affect the process in question. This includes for example the presence of noise, the possibility of coupling within the environment or unintended measurements. Here however, the environment is not merely a side effect but actually what we should be interested in. In particular, since all the information is stored within this ancilla, it will make sense to obtain a truth-value of a given sentence, or the answer to a question by *measuring* the environment. For example:



The analogy with quantum measurement is here quite useful since we can model different types of questions as different types of measurement. We can then imagine, for example, see *yes-no* questions, or questions with a very specific answer (e.g. *Who does John like?*) as basis state measurements, whereas questions such as *Who is Alice?* can be a projection onto a subspace, and hence will be similar to *von Neumann measurements*.

Note that we are using here "double wires" here. This choice was made for two different reasons, that are not completely unrelated.

The first one comes from the fact that every word that requires the environment is inherently ambiguous from the previous discussion, so doubling the wires ensures that this is still coherent with the standard DisCoCat formalism. The second reason comes from the theory of Khrennikov *et. al.* (see [Khr15, CTF$^+$07, ST92, AOK11]) which states that *mental states* are essentially quantum-like states.

3.2 Time and space matters

Motivated by the experimental results from a psychological experiment (see [ST92]), Khrennikov developed a quantum-like model of decision making, for example in the context of the *Prisoner's Dilemma* (see [AOK11]). This game can be described by the "reward table":

(A, B)	Cooperates	Betray
Cooperates	(c, c)	(a, d)
Betray	(d, a)	(b, b)

Figure 8: The Prisoner's Dilemma - pay-off table.

where the rewards a, b, c, d satisfies $a < b < c < d$. This game is known to have a *rational* solution that maximises the pay-off for each player individually, namely, to betray. It has however been shown that obtained statistical data[ST92] violates classical probability models, but is on the other hand similar to the violation of Bell inequalities. This is attributed to *non-sequential reasoning* [ST92] and *irrational behaviour*[AOK11]. In [AOK11], the mental state of the player (in the Prisoner's Dilemma) is modelled as the tensor product of a prediction state, i.e. what he thinks the other player is thinking, and his own alternative state in which he estimates his reward in either case. Both of these states are represented as quantum states, and are allowed to interfere during the decision making process. The decision itself is seen as the measurement of the final mental state.

In modelling natural languages, it is also seems reasonable to account for some similar type of irrational behaviour. Let's recall the previous example:

John is listening to a song on his phone. He likes it.

We can have a similar interpretation here. Considering the "measurement" as determining which of the two meaning is the right one. *John likes the song* is the obvious rational choice, but cannot exclude the other possibility. Hence, it is *possible* to "project" onto the less probable choice, which will appear as irrational.

Alongside, some experimental data on psychological experiments in recognition of ambiguous figures[CTF+07] (see Fig.9) or on public opinions[Moo02] exhibit features similar to the probabilities obtained in quantum experiments[Khr15], i.e. violates the formula of total probability:

$$P(B) = \sum_k P(A_k)P(B|A_k)$$

for any partition $\{A_k\}$, in the sense that order matters in opinion and perception.

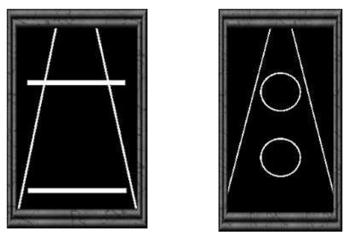


Figure 9: Ambiguous figures.

Note that the results explained in [Khr15] are similar to the concept of contextuality discussed on the previous section.

This has a direct application to the previous measurement analogy used to the retrieve information from an internal state subject to a discourse. Indeed, we would expect these "questions" to be non-commutative: the answer to a given question could depend on previously asked questions and their answers. Similarly, the physical context should have an influence on attributing coefficients for contextually ambiguous terms or in the case of indefinite causal orders. For example:

(S_1) *John is clumsy.*
(S_2) *John and his friend are walking on the street.*
(S_3) *He gets distracted by a pigeon.*
(S_4) *He trips over a rock.*

in comparison with:

(S_1) *John is clumsy.*
[*Some long and irrelevant text*]
(S_2) *John and his friend are walking on the street.*
(S_3) *He gets distracted by a pigeon.*
(S_4) *He trips over a rock.*

The structure of the causal order shouldn't change. However, it seems reasonable to assume that the further away the sentence S_1 is compared to the rest of the text, the less relevant it becomes; and hence, the weightings for the causal structures for which S_1 has influence on other sentences should decrease.

3.3 Tomography

To conclude this section, we carry on the analogy with quantum systems and ask ourselves, can we get a *master equation* of the way the mind works? In quantum information, the standard way of determining the nature of a state or a process is *quantum tomography*[NC01]. This is done by applying a series of (selected) measurements in order to obtain a more and more precise description of the state or process (e.g. black box) in question[NC01]. This may be applicable in situations where experiences can be duplicated, for example by playing a video game of through "interactive" books or films[4]. However in the case of understanding the mind, this raises some issues. Indeed, the act of *measuring*, i.e. observing, makes changes to the system itself[BS08], which is in this case, also the observer. Hence, this interpretation of perception induces an interesting question: how can one get a better understanding of perception if getting any information in turn alters perception itself?

4 Conclusion

We have, in this report, investigated various ways of putting a discourse within a context. Sheaf theory seems to provide a convenient mathematical framework to deal with anaphoras and references within a text. In addition, this method is compatible with causal structures, which not only resolves remaining ambiguity in some cases but also loops back to the quantum-like description of ambiguity in the literature by inducing a notion of entanglement of sentences. Keeping the analogy with quantum mechanics, we explored the advantages of interpreting perception as a process in a open quantum system.

Further directions There are a few directions that would, in my opinion, be interesting to explore in relation with the work done here.

- **Link with Dynamic Epistemic Logic**
  The vision of the world that each individual have is built on its personal experience. Hence, using intuitionistic and constructive logic is very sensible in the sense that we get a model that is closer to everyday life than standard logic. For example, a baby wouldn't know whether the Earth is flat or round; and as we grow up, we assess truth of statements by our knowledge and past experiences.

  The formalism of Dynamic Epistemic Logic[BS17, BCS05] seems to be a very natural model for doing this. In particular, this gives a way of dealing with public announcement or public refutation (if something is commonly said to be true or false, then an individual will accept it as true), failure (learning by failed experiments), and belief revision (it is reasonable to say that people change their mind). Then, one can represent life in a general point of view as a gigantic epistemic program.

- **Graded entailment and weighted causal structures**
  It should be possible to get an even more precise description of a given context by

[4] This idea is entirely due to Vincent Wang during the Distributional Models of Meaning meetings.

introducing *weights* on the causal structures introduced in Section 2. It would be interesting to see if the formalism of graded entailment developed in [BCLM16] can be adapted to this model described in this project.

References

[Abr14] Samson Abramsky. Contextual Semantics: From Quantum Mechanics to Logic, Databases, Constraints, and Complexity. *arXiv e-prints*, page arXiv:1406.7386, Jun 2014.

[AC08] Samson Abramsky and Bob Coecke. Categorical quantum mechanics. *arXiv e-prints*, page arXiv:0808.1023, Aug 2008.

[AC16] Daniela Ashoush and Bob Coecke. Dual density operators and natural language meaning. *arXiv preprint arXiv:1608.01401*, 2016.

[AOK11] Masanari Asano, Masanori Ohya, and Andrei Khrennikov. Quantum-like model for decision making process in two players game. *Foundations of Physics*, 41(3):538–548, Mar 2011.

[AS14] Samson Abramsky and Mehrnoosh Sadrzadeh. Semantic Unification A sheaf theoretic approach to natural language. *arXiv e-prints*, page arXiv:1403.3351, Mar 2014.

[ASK+15] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield. Contextuality, Cohomology and Paradox. *arXiv e-prints*, page arXiv:1502.03097, Feb 2015.

[BCG+17] Joe Bolt, Bob Coecke, Fabrizio Genovese, Martha Lewis, Dan Marsden, and Robin Piedeleu. Interacting Conceptual Spaces I : Grammatical Composition of Concepts. *arXiv e-prints*, page arXiv:1703.08314, March 2017.

[BCLM16] Desislava Bankova, Bob Coecke, Martha Lewis, and Daniel Marsden. Graded Entailment for Compositional Distributional Semantics. *arXiv e-prints*, page arXiv:1601.04908, Jan 2016.

[BCS05] Alexandru Baltag, Bob Coecke, and Mehrnoosh Sadrzadeh. Algebra and sequent calculus for epistemic actions. *Electronic Notes in Theoretical Computer Science*, 126:27–52, 2005.

[BS08] Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. *Investigative Ophthalmology & Visual Science - INVEST OPHTHALMOL VISUAL SCI*, 3, 01 2008.

[BS17] Alexandru Baltag and Sonja Smets. Modeling correlated information change: from conditional beliefs to quantum conditionals. *Soft Computing*, 21(6):1523–1535, 2017.

[CdMT18] Bob Coecke, Giovanni de Felice, Dan Marsden, and Alexis Toumi. Towards Compositional Distributional Discourse Analysis. *arXiv e-prints*, page arXiv:1811.03277, Nov 2018.

[CK17] Bob Coecke and Aleks Kissinger. *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press, 2017.

[Cla15] Stephen Clark. *Vector Space Models of Lexical Meaning*, chapter 16, pages 493–522. John Wiley & Sons, Ltd, 2015.

[Coe19] Bob Coecke. The mathematics of text structure. 2019.

[CSC10] Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. Mathematical Foundations for a Compositional Distributional Model of Meaning. *arXiv e-prints*, page arXiv:1003.4394, March 2010.

[CTF+07] Elio Conte, Orlando Todarello, Antonio Federici, Francesco Vitiello, Michele Lopane, Andrei Khrennikov, and Joseph P. Zbilut. Some remarks on an experiment suggesting quantum-like behavior of cognitive entities and formulation of an abstract quantum mechanical formalism to describe cognitive entity and its dynamics. *Chaos, Solitons & Fractals*, 31(5):1076 – 1088, 2007.

[DK12] Ehtibar N. Dzhafarov and Janne V. Kujala. Quantum Entanglement and the Issue of Selective Influences in Psychology: An Overview. *arXiv e-prints*, page arXiv:1209.0041, Aug 2012.

[Gär04] Peter Gärdenfors. *Conceptual spaces: The geometry of thought*. MIT press, 2004.

[Gär11] Peter Gärdenfors. Semantics based on conceptual spaces. pages 1–11, 01 2011.

[Khr15] Andrei Khrennikov. Quantum-like modeling of cognition. *Frontiers in Physics*, 3:77, 2015.

[KS15] Dimitri Kartsaklis and Mehrnoosh Sadrzadeh. A frobenius model of information structure in categorical compositional distributional semantics. In *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015)*, pages 62–74. Association for Computational Linguistics, 2015.

[KVGR11] Hans Kamp, Josef Van Genabith, and Uwe Reyle. Discourse representation theory. In *Handbook of philosophical logic*, pages 125–394. Springer, 2011.

[Lam07] J. Lambek. From word to sentence: A pregroup analysis of the object pronoun who(m). *J. of Logic, Lang. and Inf.*, 16(3):303–323, July 2007.

[Mar17] Dan Marsden. Ambiguity and Incomplete Information in Categorical Models of Language. *arXiv e-prints*, page arXiv:1701.00660, Jan 2017.

[MM94] Saunders MacLane and Ieke Moerdijk. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. 1994.

[Moo02] David W Moore. Measuring new types of question-order effects: Additive and subtractive. *The Public Opinion Quarterly*, 66(1):80–91, 2002.

[MPW16] Stephen McGregor, Matthew Purver, and Geraint Wiggins. Words, Concepts, and the Geometry of Analogy. *arXiv e-prints*, page arXiv:1608.01403, August 2016.

[MSC+13] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pages 3111–3119, USA, 2013. Curran Associates Inc.

[NC01] Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information*. 2001.

[Pie14] Robin Piedeleu. *Ambiguity in categorical models of meaning*. PhD thesis, 2014.

[PKCS15] Robin Piedeleu, Dimitri Kartsaklis, Bob Coecke, and Mehrnoosh Sadrzadeh. Open System Categorical Quantum Semantics in Natural Language Processing. *arXiv e-prints*, page arXiv:1502.00831, February 2015.

[RDEWP80] David R. Dowty, Robert E. Wall, and Stanley Peters. Introduction to montague semantics. 11, 01 1980.

[SCC14] Mehrnoosh Sadrzadeh, Stephen Clark, and Bob Coecke. The Frobenius anatomy of word meanings I: subject and object relative pronouns. *arXiv e-prints*, page arXiv:1404.5278, April 2014.

[Sel07] Peter Selinger. Dagger compact closed categories and completely positive maps: (extended abstract). *Electronic Notes in Theoretical Computer Science*, 170:139 – 163, 2007. Proceedings of the 3rd International Workshop on Quantum Programming Languages (QPL 2005).

[ST92] Eldar Shafir and Amos Tversky. Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, 24(4):449 – 474, 1992.

[ZG18] Wlodek Zadrozny and Luciana Garbayo. A Sheaf Model of Contradictions and Disagreements. Preliminary Report and Discussion. *arXiv e-prints*, page arXiv:1801.09036, Jan 2018.