Sheaf Theoretic Models of Contextuality

from Quantum Measurements to Natural Language

Author: **Daphne Wang**
Supervised by: *Mehrnoosh Sadrzadeh*

A thesis presented for the degree of
MRes in Quantum Technologies

Centre for Doctoral Training in Delivering Quantum Technologies
University College London

Abstract

It is often said that language is contextual, in the sense that the meaning of a word is highly dependent on the context it is found in. On the other hand, *contextuality* is a well-defined concept in probability theory, and in particular has been heavily studied in Quantum Mechanics, as it is considered as a major resource for the quantum advantage in Quantum Information Theory. In this project, we investigated whether natural language does exhibit some of these contextual features. We show in this report that meaning combinations are indeed contextual, and that certain phenomena can be expressed in terms of a sheaf-theoretic framework borrowed once again from quantum mechanics. This study opens the possibility of treating semantic analysis from a contextuality point of view.

Contents

Introduction

Dealing with ambiguity is a key challenge in Natural Language Processing (NLP) and does generally require a heavy amount of hand-annotated data, and large dimensional vector representations. On the other hand, quantum computing and the emergence of the field of Quantum Natural Language Processing offers some promising leads for tackling some of the biggest obstacles of classical NLP. In particular, some recent studies have shown that methods from quantum theory can be applied to natural language applications, to obtain get a better insight into peculiar phenomena, and more accurate models of ambiguity. In this project, we investigate the contextual nature of combinations of ambiguous phrases, using the framework introduced in [5, 3]. The obtained results revealed some rather unexpected properties about meaning interaction, thus introducing some encouraging data for the study of ambiguity as a contextual feature of the English language.

The study carried out in this project is based on specific examples. Indeed, the existence of some examples satisfying certain properties (here contextuality) is more important than obtaining a general statement which is true *for all* possible instances – just like in quantum mechanics, some measurements on quantum systems can be reproduced by solely statistical uncertainty, but it is the existence some sets of experiments with special statistical properties, e.g. violating Bell inequalities, which makes quantum statistics distinct from classical theories.

The first part of this report exposes the literature on both the study of contextuality in quantum systems (Chapter 1), and methods in NLP (Chapter 2). The second part, on the other hand, is entirely original. Chapters 3 and 4 are fairly independent. In Chapter 3, we will be interested in contextual features of meaning combinations using models similar to Bell experiments. In Chapter 4, we focus on peculiar sentences known as *garden-path sentences* using a framework similar to the sheaf-theoretic model usually used for quantum contextuality.

Part I

From Quantum Contextuality to Natural Language

Chapter 1

Quantum Contextuality

Early critics of quantum mechanics claimed that quantum theory was not *complete*[29], but instead was subject to unobserved *hidden variables*, as claimed that any physical theory should satisfy *local realism*. By local realism, one means that in a "complete" physical theory, the global behaviour of a system is entirely, and deterministically, determined by a set of local variables. However, the well-known *Bell theorem*[10], supported by experimental data[12], shows that a description of quantum mechanics cannot comply with local realism; assuming on the other hand that a physical theory needs to be realist to some extent, i.e. that physical quantities have a "reality" which is independent of the observer, it is the *locality* assumption that needs to be dropped.

1.1 Bell scenarios and Bell inequalities

The Bell inequalities were the first proofs of the non-existence of (local) hidden-variables in quantum mechanics. Thsection aims to introduce the basic concepts behind those inequalities, paving the way onto the formulation of *contextuality* in Section 1.2.

Given a probabilistic system defined on the set of variables Ψ, one wants to "extend" it to include extra unobserved variables (i.e. *hidden variables*), such that the observed probability distribution corresponds to the marginal distribution, averaged w.r.t. all hidden variables. For example, given measurements $A, B, C, \ldots$ with associated probability distribution P, and some outcomes $a, b, c, \ldots$ we define a compatible hidden-variable model as an extension of the original model if there exists a probability distribution Q s.t.:

$$P\left[a, b, c, \ldots | A, B, C, \ldots\right] = \int_{\Lambda} d\lambda Q(\lambda) Q\left[a, b, c \ldots | A, B, C, \ldots, \lambda\right] \tag{1.1}$$

In particular, if only *local hidden-variables* are considered, then the the probability distribution Q needs to satisfy:

$$P\left[a, b, c, \ldots | A, B, C, \ldots\right] = \int_{\Lambda} d\lambda Q(\lambda) Q\left[a \ldots | A, \lambda\right] Q\left[b \ldots | B, \lambda\right] Q\left[c \ldots | C, \lambda\right] \ldots \tag{1.2}$$

There are different types of hidden-variable models, e.g. parameter-independent models or λ-independent models(see [14, 2] for a full classification of hidden-variable models). However, the most commonly studied type, at least in the context of quantum mechanics, is the *deterministic hidden-variable model*, in which the values of the hidden-variables determines the values of the observables with *certainty*. In particular, the existence of a deterministic hidden-variable model. This implies that all observations on a given quantum system can be computed *deterministically* by local variables only.

As we know, however, it is not possible to define a deterministic hidden-variable model that is fully consistent with observations on quantum systems. This can be shown from Bell-type experiments, or *Bell scenarios*, which can be described as follows. We consider two parties A and B, which can be assumed to be spacelike separated (in order to avoid communication between them). The two parties are known to share a quantum state, which is usually entangled[1], are each free to apply local operations and measurements on their respective subsystems. In addition, if we fix the set of possible operations and measurements that each party is allowed to make, then they can record which experiment they have carried out and which outcome they have measured (see Fig. 1.1). After many repetitions of this procedure, they can meet, and compute joint probability distributions given their respective choice of experiments (i.e. given s global *measurement context*).

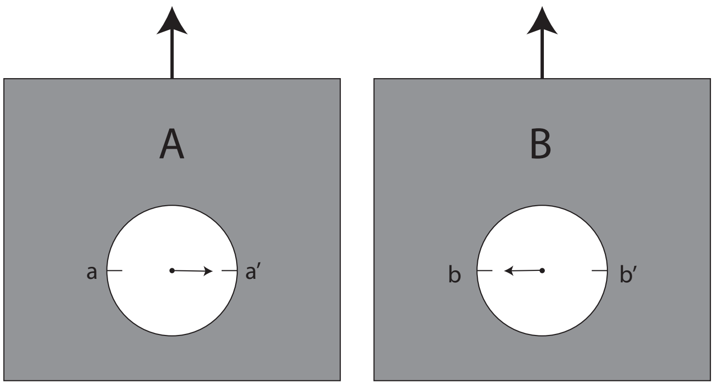


Figure 1.1: Bell experiment. This particular event corresponds to $[(a', b) \mapsto (0, 1)$.

The no-signalling property translates, in terms of the obtained statistics, as the condition that the marginal probabilities of each party, given one choice of a local measurement, do not depend on the global measurement context (otherwise, A or B might be able to infer which local measurement has been chosen by the other party, which is therefore non-local information which can instantly be obtained), formally:

$$P\left[o_a | a, b\right] = P\left[o_a | a, b'\right] \tag{1.3}$$

(and similarly $P\left[o_b | a, b\right] = P\left[o_b | a', b\right]$). The *Bell inequalities* are generally presented under the form of the *CHSH inequality*[19], which is stated as follows: given four binary measurements Q, R, S and T taking the values ± 1, then it can be shown that the expectation value of the operator $QS + RS + RT - QT$ satisfies:

$$\langle QS \rangle + \langle RS \rangle + \langle RT \rangle - \langle QT \rangle \leq 2 \tag{1.4}$$

However, if Q, R and S, T are local measurements of the state $|\Psi\rangle = \frac{1}{\sqrt{2}}\Big(|01\rangle - |10\rangle\Big)$ on local systems A and B respectively, s.t.:

$$Q = Z_{\mathsf{A}} \tag{1.5}$$
$$R = X_{\mathsf{A}} \tag{1.6}$$
$$S = -\frac{Z_{\mathsf{B}} + X_{\mathsf{B}}}{\sqrt{2}} \tag{1.7}$$
$$T = \frac{Z_{\mathsf{B}} - X_{\mathsf{B}}}{\sqrt{2}} \tag{1.8}$$

[1]Entanglement is necessary to obtain non-classical statistics.

(where $Z = +1 \cdot |0\rangle \langle 0| - 1 \cdot |1\rangle \langle 1|$ and $X = +1 \cdot \frac{1}{2}\big(|0\rangle + |1\rangle\big)\big(\langle 0| + \langle 1|\big) - 1 \cdot \frac{1}{2}\big(|0\rangle - |1\rangle\big)\big(\langle 0| - \langle 1|\big)$), then one can show that:

$$\langle QS \rangle + \langle RS \rangle + \langle RT \rangle - \langle QT \rangle = 2\sqrt{2} > 2 \tag{1.9}$$

This apparent contradiction comes from the fact that all of these measurements do not belong to the same measurement context, and in fact, the expectations values corresponds to incompatible situations.

1.2 The Kochen-Specker theorem

The Bell inequalities offers a proof by contradiction that one cannot extend the probabilistic model obtained from observations of quantum systems to a deterministic hidden-variable model. In [43], Kochen and Specker prove a stronger statement about the existence (and non-existence) of hidden-variable models.

In order to understand the so-called *Kochen-Specker theorem*, we start with the following observation: physical observables can be related to each other via functions, e.g. $\hat{B} = g(\hat{A}) = \hat{A}^2$; then measuring the value $a \in \mathbb{R}$ for the quantity $\hat{A}$ implies that the corresponding value of $\hat{B}$ would then be a^2. However, if physical quantities have a *definite value* given a specific state ψ (as it would in a deterministic hidden-variable model), this condition that can be rewritten as the existence of a function $f_{\hat{A}} : \Omega \to \mathbb{R}$ s.t. for very $\hat{B}$ s.t. $\hat{B} = g(\hat{A})$ for some function g, we need to have[43]:

$$f_{\hat{B}} = g \circ f_{\hat{A}} \tag{1.10}$$

and hence, the above condition corresponds to a necessary condition to the existence of a deterministic hidden-variable model.

1.3 Abstract approaches to contextuality

In this section, we describe a framework for characterising contextuality and the Kochen-Specker theorem introduced in the previous section, using the concept of *presheaves*.

1.3.1 Categories and presheaves

Presheaves have historically been used in topology and algebraic geometry[47], as a way of studying the *global features* of a topological space w.r.t. to its *local features*.

In order to define presheaves, we start by introducing the notion of a category. A *category* $\mathcal{C}$[46] consists of a collection of objects and arrows (or morphisms) between them, such that:

- Each object $A \in \mathcal{C}$ has an identity morphism $id_A : A \to A$;
- If $f : A \to B$ and $g : B \to C$ are arrows in the category $\mathcal{C}$, then the composition $g \circ f : A \to C$ is an arrow in $\mathcal{C}$.

In particular, we define the category **Set** for which objects are sets, and morphisms are functions between sets. A *contravariant functor*[46] between two categories $\mathcal{C}$ and $\mathcal{D}$ is a map F, denoted, $F : \mathcal{C}^{op} \to \mathcal{D}$ such that:

- For each object $A \in \mathcal{C}$, $F(A)$ in an object in $\mathcal{D}$;
- For each morphism $f : A \to B$ in $\mathcal{C}$, the morphism $Ff : F(B) \to F(A)$ is a morphism in $\mathcal{D}$ (note the reverse order of the arrow[2]);
- For each object $A \in \mathcal{C}$, the morphism $Fid_A : F(A) \to F(A)$ is the identity morphism of $F(A)$ in $\mathcal{D}$.
- For each pair of arrows $f : A \to B$ and $g : B \to C$ in $\mathcal{C}$, we have $F(g \circ f) = Ff \circ Fg$.

From this, we define a *presheaf* as a contravariant functor to the category of **Set**, i.e. $\mathcal{F} : \mathcal{C}^{op} \to \textbf{Set}$ for some category $\mathcal{C}$.

A *global section*[47] on a presheaf $\mathcal{F}$ is a function $\gamma : X \to \textbf{Set}$ such that for all $A \in \mathcal{C}$, $\gamma(A) \in \mathcal{F}(A)$, and for all $f : B \to A$ in $\mathcal{C}$:

$$\mathcal{F}f\,(\gamma(A)) = \gamma(D) \tag{1.11}$$

On the other hand, a *local section* is a function $\widetilde{\gamma} : X \to \textbf{Set}$ from a subset of objects X of $\mathcal{C}$, such that the condition (1.11) is satisfied for whenever A and B are in X.

There are several possible choices of presheaves that one can make to study quantum contextuality. We will, for most of this report consider the *distribution presheaf* $\mathcal{DE}$ which we will now introduce; however other possible choices studied in the literature will be described in Section 1.4.

1.3.2 Empirical models and presheaves

We now want to express results of quantum experiments in terms of presheaves. To do so, we consider experiments similar to Bell scenarios described in Section 1.1. The set of all possible measurement contexts, i.e. the set of all possible joint measurements will be denoted by $\mathcal{M}$, whereas the set of all possible individual measurements will be denoted by X. For example, in the case of a bipartite system $A \otimes B$, where A and B are choosing their local measurements in the set $\{a, a'\}$ and $\{b, b'\}$ respectively, we have:

$$\mathcal{M} = \{(a, b), (a, b'), (a', b), (a', b')\} \tag{1.12}$$

$$X = \{a, a', b, b'\} \tag{1.13}$$

An *event* will be defined as an assignment from a measurement context to an outcome; for example, assuming that both a and b are binary measurements taking values in $\{0, 1\}$, $[(a, b) \mapsto (0, 1)]$ will be a valid event. We then define a first presheaf, namely the *event presheaf*: $\mathcal{E} : \mathcal{P}(X)^{op} \to \textbf{Set}$, where the category $\mathcal{P}(X)$ is defined as follows:

- The objects of $\mathcal{P}(X)$ are all the possible subsets of X;
- There is a unique arrow $\subseteq : A \to B$ iff $A \subseteq B$.

We then define the set $\mathcal{E}(A)$, for $A \subseteq X$, as the set of all possible assignments from A to the set of possible outcomes O; note that one can always define a uniform set of outcomes for any set of measurements by relabelling and/or adding "extra" impossible outcomes. We also define the image of morphism under the presheaf $\mathcal{E}$ as the restriction morphism:

$$\mathcal{E}(A \subseteq B)(B) = \mathcal{E}(B)|_A \qquad \forall A \subseteq B,\ A, B \subseteq X \tag{1.14}$$

[2]A *covaraint functor*, or just *functor*, does preserve the directions of arrows.

for example, $[(a, b) \mapsto (0, 1)] \in \mathcal{E}(\{a, b\})$ will be mapped, under the restriction in a (i.e. $A = \{a\}$, $B = \{a, b\}$), to $[a \mapsto 0] \in \mathcal{E}(\{a\})$. It can easily be shown that this definition of morphisms satisfies the presheaf conditions. A *global assignment* will refer to an assignment of all measurements in X to an outcome.

We note that the concept of observations of certain events, or probabilities of events, are so far not taken into account. We define the family of probability distribution associated with all the possible joint measurements as the *empirical model* of the experiment[5]. This empirical model can be represented in the form of a table (e.g. see Fig. 1.2). We then

A	B	(0,0)	(0,1)	(1,0)	(1,1)
a	b	1/2	0	0	1/2
a	b'	3/8	1/8	1/8	3/8
a'	b	3/8	1/8	1/8	3/8
a'	b'	1/8	3/8	3/8	1/8

Figure 1.2: Empirical model associated with the measurement of the bipartite state $|\Psi\rangle = \frac{1}{\sqrt{2}}\Big(|00\rangle + |11\rangle\Big)$ with local measurements $a, b = |1\rangle\langle 1|_{A,B}$ and $a', b' = |\phi\rangle\langle\phi|_{A,B}$ where $|\phi\rangle = \frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle$.

describe empirical models in terms of a presheaf, namely $\mathcal{D}_{\mathbb{R}_+}\mathcal{E}$ where each subset A of X is associated with the set of *all* probability distributions on possible assignments of outcomes on A. An empirical model in this framework is therefore a choice of local sections from this presheaf, restricted to the measurement contexts in $\mathcal{M}$, and the observed probability distributions. In addition, a global section on this presheaf $\mathcal{D}_{\mathbb{R}_+}\mathcal{E}$ is a distribution $d \in \mathcal{D}_{\mathbb{R}_+}\mathcal{E}(X)$ which assigns a probability to *every global assignments* $s \in \mathcal{E}(X)$ so that it forms a probability distribution which is consistent with all of the observed probabilities. Hence, an experiment demonstrates *contextuality* iff the presheaf $\mathcal{D}_{\mathbb{R}_+}\mathcal{E}$ has no global section.

Computing global sections As shown in [5], given an empirical model, its global sections can be found as the solution of a linear system of equations. Indeed, by fixing an ordering of the local sections (see e.g. Fig. 1.3a for a Bell-type scenario), and an ordering for the global assignments (see Fig. 1.3b), then we define the *incidence matrix* $\widetilde{\mathbf{M}}$ as

A	B	(0,0)	(0,1)	(1,0)	(1,1)
a	b	s_1	s_2	s_3	s_4
a	b'	s_5	s_6	s_7	s_8
a'	b	s_9	s_{10}	s_{11}	s_{12}
a'	b'	s_{13}	s_{14}	s_{15}	s_{16}

(a) Enumeration of local sections for a Bell scenario.

$$t_1 = [(a, b, a', b') \mapsto (0, 0, 0, 0)]$$
$$t_2 = [(a, b, a', b') \mapsto (0, 0, 0, 1)]$$
$$t_3 = [(a, b, a', b') \mapsto (0, 0, 1, 0)]$$
$$t_4 = [(a, b, a', b') \mapsto (0, 0, 1, 1)]$$
$$\vdots$$

(b) Enumeration of global assignments for a Bell scenario.

follows[5]:

$$\widetilde{\mathbf{M}}_{i,j} = \begin{cases} 1 & \text{if } \; t_j|_C = s_i \text{where } s_i \in \mathcal{E}(C) \\ 0 & \text{otherwise} \end{cases} \tag{1.15}$$

Hence, a solution $\mathbf{x}$ of the system $\widetilde{\mathbf{M}}\mathbf{x} = \widetilde{\mathbf{v}}$, where $\widetilde{\mathbf{v}}$ is the column vector $\widetilde{\mathbf{v}} = (s_i)_i$, corresponds to coefficients x_j for which $\sum_j x_j t_j$ is exactly the observed empirical model. In addition, we want to restrict these coefficients to be within $[0, 1]$, as we want them to represent probabilities; similarly, since we also want $\sum_j x_j = 1$, so we will consider the *augmented incidence matrix* $\mathbf{M}$ (and respectively the augmented vector $\mathbf{v}$):

$$\mathbf{M} = \left(\begin{array}{ccccc} & & \widetilde{\mathbf{M}} & & \\ \hline 1 & 1 & 1 & \dots & 1 \end{array}\right) \tag{1.16}$$

$$\mathbf{v} = \left(\begin{array}{c} \widetilde{\mathbf{v}} \\ \hline 1 \end{array}\right) \tag{1.17}$$

A solution to the system $\mathbf{M}\mathbf{x} = \mathbf{v}$ in the positive real numbers therefore corresponds to a global section of the presheaf $\mathcal{D}_{\mathbb{R}_+}$. For simplicity, we will refer to the augmented incidence matrix as simply the incidence matrix for the rest of this report. The full expression for the incidence matrix can be found in Appendix A, as well as the proof of non-existence of a global section for the empirical model depicted in Fig. 1.2.

1.3.3 Logical contextuality

We now describe an even stronger type of contextuality. Indeed, instead of considering probability distributions, we here consider distributions over the Booleans $\mathbb{B} = \{0, 1\}$. From a probability distribution P over the events $e_1, e_2, \ldots \in \Omega$, its associated *Boolean distribution* $P_{\mathbb{B}}$ is given by:

$$P_{\mathbb{B}}[e_i] = \begin{cases} 1 & \text{if } P\left[e_i\right] \text{ ; i.e., } e_i \text{ is possible} \\ 0 & \text{otherwise; i.e., } e_i \text{ is impossible} \end{cases} \tag{1.18}$$

Any Boolean distribution also has to satisfy the following:

$$\bigvee_i P_{\mathbb{B}}\left[e_i\right] = 1 \tag{1.19}$$

i.e. there is at least one possible event[3]. Hence, we can reconsider the empirical models with Boolean distributions by considering the presheaf $\mathcal{D}_{\mathbb{B}}\mathcal{E}$, which associates Boolean distributions on measurement-outcome assignments. For example, the Boolean empirical model from Fig. 1.2 is given by:

A	B	(0,0)	(0,1)	(1,0)	(1,1)
a	b	1	0	0	1
a	b'	1	1	1	1
a'	b	1	1	1	1
a'	b'	1	1	1	1

(1.20)

We observe that Boolean distributions corresponds to the support of the different probability distributions. Hence, a global section on the presheaf $\mathcal{D}_{\mathbb{B}}\mathcal{E}$ corresponds to the existence of a family of global assignments compatible with the support of the probability distributions, so that each possible event is part of at least one of these global assignments. For example, in the empirical model in (1.20), the family $\{[(a, b, a', b') \mapsto (0, 0, 0, 0)], [(a, b, a', b') \mapsto (0, 0, 1, 1)], [(a, b, a', b') \mapsto (1, 1, 0, 0)], [(a, b, a', b') \mapsto (1, 1, 1, 1)],$

[3]This is the equivalent statement of $\sum_i P\left[e_i\right] = 1$, but in the Boolean ring $(\mathbb{B}, \wedge, \vee)$, where $\wedge$ and $\vee$ are respectively the *and* and *or* operations.

$[(a,b,a',b') \mapsto (0,0,1,0)]$, $[(a,b,a',b') \mapsto (1,1,0,1)]\}$ forms a global section. If an empirical model has no global sections in the Boolean distribution presheaf, then the model is said to be *possibilistically contextual*[5]. An example of a possibilistic empirical model is the so-called Hardy model, which has been introduced in [34]. The support of the obtained distribution is given by:

A	B	(0,0)	(0,1)	(1,0)	(1,1)
a	b	1	1	1	1
a	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0

(1.21)

(see [34] for the details of the proposed experiment). Indeed, the local section $[(a,b) \mapsto (0,0)]$ cannot be consistently extended to a global assignment which is consistent with the support (see the local sections in red in (1.21)).

Now, we note that if a Boolean empirical is possibilistically contextual, then it is automatically (probabilisitically[4]) contextual, since one cannot impose a probability distribution on all possible global assignments consistent with all the observations if some of these observations are not part of any global assignment. Hence, possibilistic contextuality is strictly stronger than probabilistic contextuality.

We can moreover define an even stronger degree of contextuality. If no local section of a given empirical model can be extended to a global assignment, we say that the model is *strongly contextual*. An example of such strongly contextual model is the *PR-box* introduced in [55]. The PR-box is an example of a "super-quantum" no-signalling process, which cannot be realised with quantum systems. Its distribution support is given by:

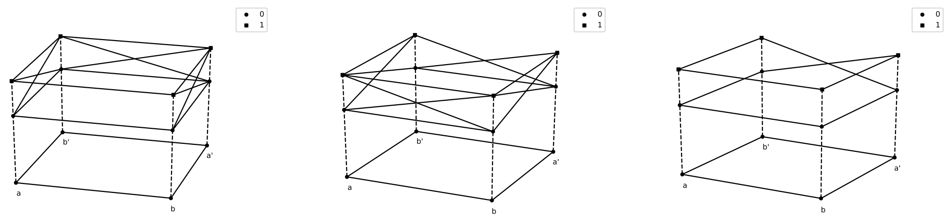
A	B	(0,0)	(0,1)	(1,0)	(1,1)
a	b	1	0	0	1
a	b'	1	0	0	1
a'	b	1	0	0	1
a'	b'	0	1	1	0

(1.22)

It can be shown to be strongly contextual since the first 3 measurement contexts shows perfect correlation between all measurements from A and B, whereas the last measurement context corresponds to anti-correlation from the outcomes of a' and b'.

Bundle diagram representation We can, for simple examples as the ones considered so far, adopt a more convenient representation of Boolean empirical models, namely *bundle diagrams*[3]. In these diagrams, we represent each of the local measurements as a vertex, and edges between every two of these vertices if the joint measurement is possible. In addition, we depict, for each individual measurements, the set of possible outcomes as a set "sitting" on top of the associated vertex, and an edge is added between two of the "outcome"-vertices if the joint measurement comes with a non-zero probability (see Fig. 1.4 for examples). In particular, global assignments can be seen in these bundle diagrams as connected loops going through exactly one outcome for each of the measurements.

[4]In order to be clear about which degree of contextuality is considered, we will describe standard contextuality as probabilistic contextuality.



(a) Bell scenario (b) Hardy model (c) PR-box

Figure 1.4: Examples of bundle diagrams.

1.4 Related work

As previously mentioned, alternative choices of presheaves can be made for the study of quantum contextuality. The main options have been studied in the context of a *topos-theoretic approach to quantum mechanics*[38, 39, 33, 16, 37].

Topos theory has originally developed as a generalisation of set theory and Boolean logic (see [32, 47] for standard references on the subject). The basic idea is that, in some circumstances, one might want to have more than the standard true/false *truth-values*. In particular, given a category $\mathcal{C}$, the associated *category of presheaves* $\mathbf{Set}^{\mathcal{C}^{op}}$ forms a topos, with specific generalised truth-values[5]. The work presented in the collection of papers [38, 39, 33, 16] corresponds to a fairly literal application of the Kochen-Specker theorem in terms of presheaves: observables are functions of each other but their values do not always respect these functional relations. The propositions considered are of the form "$A \in \Delta$" which is asking whether the physical quantity A takes values in the range Δ. These propositions have (generalised) truth values w.r.t. a *context*, which can be seen as a generalised version of the previously described measurement contexts. The non-classicality of quantum systems once again appears as the non-existence of a global section on the presheaves. The approach adopted in [37] is fairly similar to the other topos-theoretic framework, and mostly differs from it by its choice of presheaf. A full comparison of these two approaches can be found in [65].

[5] These are truth values are *sieves*; however, the definition of either sieves, or even evaluation and truth values are irrelevant for the rest of this report.

Chapter 2

Ambiguity in Natural Language

We start with a peculiar observation: even though many common words in the English language are polysemous, i.e. have more than one meaning, this does not create a considerable obstacle in the day-to-day comprehension of texts or conversations. For example, the word *charge* has 40 different senses according to the WordNet corpus[61], however, its meaning in the sentence *The bull charged* is fairly unambiguous. On the other hand, Word Sense Disambiguation has shown to be a computationally difficult task to implement, and is one of the greatest challenges of NLP.

This chapter is aimed at introducing the main methods used in NLP to model meanings of words, as well as some promising ways in which quantum methods have been used in this field.

2.1 Models of meaning

We start by describing in this section, the different common ways of modeling the semantics in NLP.

2.1.1 Distributional models of meaning

Distributional models of meanings are based on a very simple observation, which first appeared in the works of Harris[35], Firth[30] and Joos[40], namely: the meaning of a word can be inferred from its context. For example, the word *car* will be found in the context of *parking, sport, crash* . . . : the meaning of *car* therefore corresponds to its distributions according to different contexts. The idea is therefore to associate each word with a vector (a *word-vector*) in a large dimensional vector space (sometimes called the *word-space*), which stores the different occurrence frequencies w.r.t. some predefined "contexts"[1].

There are several possible ways of defining those "contexts" or word-dimensions. For example, one can define the context of a specific word of interest (a *head-word*), as being the document in which it is found (e.g. [56, 59] are such approaches), see e.g. Fig. 2.1. Other alternatives are to consider the different partS-of-speech[58, 22], i.e. for example which are the most common modifiers of the head-words and which words are more often modified by the head-word (see Fig.2.2), or *context-group discrimination* as in [60, 15], where context themselves are associated with vectors.

[1]For most of this report, the word *context* will have a more precise meaning.

$$fool = \left(\begin{array}{cccc} 36 & 58 & 1 & 4 \\ \textit{As You Like It} & \textit{Twelfth Night} & \textit{Julius Caesar} & \textit{Henry V} \end{array} \right)$$

Figure 2.1: Example of a document-classified word-vector for *fool* w.r.t. the set of Shakespeare plays *As You Like It*, *Twelfth Night*, *Julius Caesar* and *Henry V*. The data and example is taken from[41] .

Modifier	Counts	**Modified by**	Counts
motor	458	park	1517
police	375	sale	258
sport	224	crash	188
stolen	193	parking	138
	...		...

Figure 2.2: Example of counts of modifiers of / modified by *car*. The data is taken from the British National Corpus(BNC)[1].

Vector-semantics have been shown to be very useful for computing meaning similarity[23, 22, 15], thesaurus generation[63, 20, 20, 52, 21] and *Word Sense Disambiguation* (WSD). In general, WSD can be decomposed into two steps[60]: sense discrimination (i.e. detecting whether two occurrences of the same word belong to the same sense), and sense labelling (i.e. associating each occurrence of a word to its activated sense). In [60], Schütze proposes a procedure for sense discrimination using context-group disambiguation. From training data, words are associated with context vectors which store the co-occurrences of the head-word with other words; thereafter, the contexts themselves analysed so that contexts that are similar forms clusters in the word-space. Hence, the different senses of an ambiguous word will correspond to different clusters. If some test data is then input, the context of this new data is mapped into the word-space and the sense which is activated should correspond to the one which cluster is the closest. Sense labelling on the other hand is a more intricate problem, and current solutions usually make use of human annotations[44], sense-tagged corpus[28] such as the Gold Standard Corpus[64] or SemCor[48] (WordNet senses tagged corpus), although some unsupervised methods has also been proposed[66].

The main drawbacks of the distributional models is the lack of structural information that these word-vectors supply[51], e.g. *house boat* and *boat house* have the same vector representations. In addition, this representation of word meanings does not seem to allow for analogies and word association[31].

2.1.2 Symbolic representations of meaning

The distributional models described above are very convenient for computational implementations, and are particularly useful for "bag-of-words" type of applications. However, as previously mentioned, understanding of sentences and phrases with a syntactic structure is not usually possible within those models. The *symbolic representations* of meaning, however, directly exploits the grammatical structure of sentences to comprehend them.

Following from the tradition of *Montague semantics*[24], symbolic representations make use of the formal language of first order logic. These semantics are generated from a *vocabulary* (i.e. set of words) including relations and functions (i.e, tuples). For example, given the vocabulary $\{Alice, Bob, friend, like\}$, the sentence *Alice likes Bob* corresponds to the relation $likes(Alice, Bob)$ (or in other words $(Alice, Bob) \in likes$), and *Alice is a friend* would be represented by $friend(Alice)$ (or $Alice \in friend$). One can also make use of connectives (e.g. $\wedge$: "and", $\vee$:"or", $\implies$:"then", $\neg$:"not"), as well as variables, which can be used for references, and quantifiers (e.g. $\exists$: indefinite reference, $\forall$:"all"). For example the sentence *All cars are red and fast* would be represented as $\forall x \; car(x) \implies red(x) \wedge fast(x)$, and *A car moves* would be associated with $\exists y \; car(y).\; moves(y)$.

This type of semantics has been extended to treat more complex text structures, for example in Discourse Representation Theory[42, 36], which can be used for several sentences, and can deal with easy examples of anaphoric references (e.g. *John eats food. He is happy*).

Furthermore, even if first and higher-order semantics are less convenient for computational implementation, implementations from [17, 13] show some promising results for scalability of symbolic representations of language.

2.2 Quantum inspired models of meaning

Both symbolic and distributional models of meaning appear to have complementary features, and hence, attempts at combining the two approaches have been proposed (see e.g. [18]). Interestingly, these approaches have been shown to share similarities with the formalism of quantum mechanics, for instance by using tensor products and Hilbert spaces[18]. We will in this section investigate some of the models of meanings that make use of elements of quantum mechanics. In Section 2.2.1, we explore the use of density matrices in natural language models, while in Section 2.2.2, we investigate a quantum representation of *concepts*.

2.2.1 Density matrices and distributional models

We start by presenting the approach from [11], which uses an approach similar to the distributional approach, except that words are encoded within density matrices instead of vectors. The similarity between two words v, w is calculated by computing the inner product of the associated density matrices: $Tr(\rho_v \rho_w)$. The method for constructing the density matrices are using ideas from both part-of-speech parsing and document-content classification by encoding the two types of level of probability distributions as quantum superposition of in a pure state and statistical mixing respectively. In particular, the obtained experimental data from [11] showed similarity with human responses, performing better than classical models from both word similarity and word association points of view.

In addition, density matrices have also been used to encode different levels of ambiguity. Indeed, in [54], quantum superposition is used for different, but related, meanings of a given word (e.g. the two meanings of *bank* in *I am looking for a bank loan* and *I need to go to the bank to deposit a cheque*), while statistical mixing is used for different meanings which are unrelated, e.g. the two meanings of *bank*, this time in *bank loan* and *river bank*. This model hence has the advantage of representing different types of ambiguity in a word more accurately.

Finally, another way of using density matrices in natural language comes from *lexical entailment*[9, 45]. A word is then represented as the sum of the matrices of its *hypernyms* (i.e. words which are strictly more specific). For example.:

$$|pet\rangle\langle pet| = |pug\rangle\langle pug| + |tabby\rangle\langle tabby| + |goldfish\rangle\langle goldfish| \tag{2.1}$$

Now the entailment relation is defined as the *Lowner order*: $A \sqsubseteq B$ (A entails B), if $B - A$ is positive, for any two positive operators A and B. In the previous example, *tabby* entails *pet* since both *pug* and *goldfish* are positive operators. We can even extend this further by introducing a notion of *graded entailment.* For example, if we have:

$$|cat\rangle\langle cat| = |tabby\rangle\langle tabby| + |wildcat\rangle\langle wildcat| \tag{2.2}$$

Then, one sees that *cat* does not strictly entail *pet*, but both *cat* seems to entail *pet* "to some extent". Two possible measures of graded entailment have been proposed in [45] to account for this partial entailment.

It was moreover proposed in [8] to combine the last two uses of density matrices by constructing a "second-order" density matrix: the "ambiguity" density matrix is enriched with an entailment dimension. However, we note that it is still not clear if this construction will have any interpretation in terms of quantum systems.

2.2.2 Concept combinations and quantum states

The theory of concepts has been developed to categorize our "experience" of the world in terms of *concepts*, such as *table, pet* or even *good* or *bad.* Every concept has a number of "instances" (or *exemplars*) all of which share the some similarities. In particular, given a concept, some of its instances will be more *typical* or more representative of the concept, for example *cat* and *dog*, will be typical examplars for the concept *pet* in the way that *hedgehog* or *sheep* are not.

The *pet-fish* problem concerns how concepts combine, and in particular how typicality behaves as concepts interact. Indeed, the concept *guppy* is neither a typical examplar of *pet* nor *fish*, but is a typical examplar of the concept *pet fish.* This problem is solved using Hilbert spaces in [7] by hypothesising that a concept can be associated with a *state* when putting into a *context.*

Roughly speaking, a context of a given concept is a situation where the concept is found it. For example, *The pet is running*, is a context of *pet.*In addition, the concept, when found in a certain context is associated with a state (for example in the context e : *the pet is running*, the state p_e of the concept *pet*, may be simply *running*). This state is allowed to change when exposed to a further or different context. For example, if the context (once again of pet) changes from:

$$\textit{the pet is running} \longrightarrow \textit{the pet is running in circles}$$

the state will change from *running* to *running in circles.* In addition, it is assumed that applying the same context twice to a given state is the same as applying it once, i.e. once one collapses the state in a certain context, it won't change if the context remains the same. Furthermore, states do not always change when exposed to a new context, for example, if we consider the opposite situation:

$$\textit{the pet is running in circles} \longrightarrow \textit{the pet is running}$$

then the state *running in circle*, will not be changed when the context is changed to *running* only. We say that the state $p_{e'}$: *running in circles* is an eigenstate of the context *the pet is running.*

From this intuition, the authors embed each of the states as a vector in a Hilbert space[7]. In addition, the different contexts are associated with projectors on this Hilbert space. The pet-fish paradox is then resolve by introducing correlations when considering the combinations of two concepts: when two concepts *pet* and *fish* combine, all the states are embedded in the space $\mathcal{H}_{pet} \otimes \mathcal{H}_{fish}$. Subsequently, every state in the eigenspace of the newly formed concept will be of the form:

$$p = \sum_{u} \alpha_u \left|u\right\rangle \otimes \left|u\right\rangle \tag{2.3}$$

where u is a state in which the *pet* is a *fish* and the *fish* is a *pet* (note that these two conditions are not equivalent, the former corresponds to a state in the concept *pet* while the latter is a state in the concept *fish*). It has moreover been shown in [7] that the obtained statistics are consistent with the intuition that *goldfish* and *guppy* as typical examplars of the concept *pet-fish*.

Part II

Contextuality in Natural Language

Chapter 3

Contextuality in meaning combinations

In this chapter, we will investigate mechanisms behind meaning composition of ambiguous words. Similar to the pet-fish paradox, one can consider the two-words phrase *power plant*; each word *power* and *plant* are associated with different meanings, each coming with certain frequencies. However, when the two words combine, the meanings activated with highest frequency for the phrase, do not coincide with the most frequent meanings of neither *power* nor *plant*.

To explore this, we propose an experiment similar to Bell scenarios. We consider multiple parties, or agents, each of which will choose one measurement context from a predetermined set. Each measurement context in this set will be the "meaning-measurement" of a given word, which will interact with the meanings of other words chosen by the other parties, and will return the activated meaning according to a fixed encoding. For example the two meanings of *plant* could be encoded as: 0 : *living organism*, 1 : factory. Each individual measurement context need not be ambiguous, but only ambiguous words will be associated with a non-deterministic measurement. The interaction will also be dictated by some predetermined rules, such as in which order the words are composed, or which part-of-speech each word will correspond to. The global measurement context will be labelled by all the different choices made by the parties, and for each global measurement context, the recorded activated meaning will form a joint distribution. These distributions can then be represented in the form of an empirical model as described in Section 1.3, and analysed thereafter using the sheaf theoretic method from[5, 3]. In order to obtain a valid empirical model, all the possible combinations of words need to make some sort of sense. For example, taking two parties A and B such as A chooses a verb in the set $\{pen, see\}$ and B chooses the object of this verb within: $\{sheep, note\}$, all the possible combinations within this experiment are possible, i.e. the phrases *pen a sheep, pen a note, see a sheep* and *see a note* can be found in natural language. However, if the set of verbs is changed to $\{pen, herd\}$ we have a problem since the phrase *herd a note* does not make much sense[1].

We will furthermore restrict to 2-words combinations, i.e. two party scenarios.

Additionally, the set of ambiguous words is taken from experimental data sets from the studies[53, 62, 57, 49].

[1] This condition can be relaxed, see Section 3.1.3

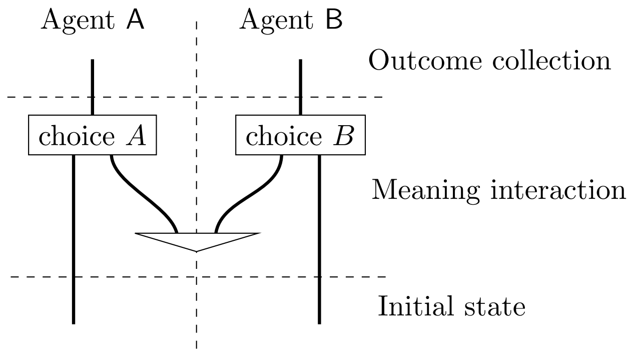


Figure 3.1: Example of a 2-words scenario. The state (triangle) represents the predefined conditions of the interaction (e.g. $verb - object$).

In Section 3.1 we consider a model similar to the possibilistic models described in Section 1.3.3. The aim of this section will be to show that meaning of ambiguous words in phrases are contextual and develop intuition about the behaviour of composition of ambiguous meanings. In Section 3.2, we investigate these properties in an empirical fashion and show that meaning combinations observed in corpora are contextual.

3.1 Logical contextuality

We will start by considering possibilisitic models, i.e. the distributions are defined in the Boolean ring $(\mathbb{B}, \wedge, \vee)$ and corresponds to the support of some probability distributions. In order to obtain a proof-of-principle, the different observable meaning combinations are tabulated using common sense; and whenever results are not obvious, alternative models are also presented.

Further examples can be found in Appendix B.

3.1.1 Semantic combinations in ambiguous contexts

We first consider a stricter model where each word is assumed to have a definite grammatical type (for example even though the word *pen* can be a verb or a noun we will restrict to either its verb meanings or its noun meanings, but not both). We will therefore restrict situation where each agent is attributed a grammatical role, and the type of phrases considered will be $verb - object$, $subject - verb$ and $adjective - noun$.

Consider a warm-up example, with only one ambiguous word, which has two clearly distinct meanings. We consider a $verb - object$ example where the two verbs $\{pen, see\}$ are interacting with objects $\{sheep, note\}$. The ambiguous context is the verb *pen*, and its two meanings can be seen as:

a. *to pen*: to write

   Example: *He penned a letter to his wife.*

b. *to pen*: to put in an enclosed space

   Example: *The farmer penned the pigs into the barn to prevent them from escaping.*

The different meanings of all the possible measurements are encoded as follows:

Encoding	Meanings of			
	pen	*see*	*sheep*	*note*
0	*write*	*see*	*sheep*	*note*
1	*enclose*	$\star$	$\star$	$\star$

Table 3.1: Encoding of meanings of *pen*, *see*, *sheep* and *note*

Where the stars corresponds to meanings which are not defined in this experiment; this means that measuring the meanings of *see*, *sheep* and *note* will return 1 with probability 0. We also note that these words actually do have more than one meaning, but most of them are irrelevant in this situation. For clarity, we only consider the following meanings:

- *to see*: to perceive by sight

  Example: *Can you see a bird in the tree?*

- *sheep*: woolly animal

  Example: *The farmer is rearing sheep for their wool.*

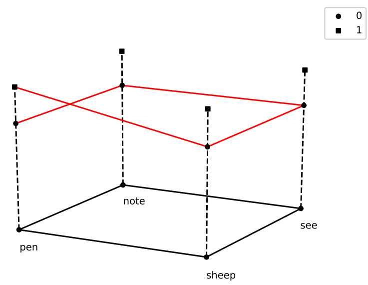
- *note*: short written record

  Example: *He made a note of the appointment*

The empirical model associated with these words combinations is depicted in Fig. 3.2, and can be understood as follows:

verb	*object*	(0,0)	(0,1)	(1,0)	(1,1)
pen	*sheep*	0	0	1	0
pen	*note*	1	0	0	0
see	*sheep*	1	0	0	0
see	*note*	1	0	0	0

(a) Empirical model



(b) Bundle diagram representation (local sections which cannot be extended are depicted in red).

Figure 3.2: Empirical model associated with the measurement contexts $\{pen, see\} \times \{see, note\}$.

- $(pen, sheep)$: One can enclose a sheep in a pen, but not "write" a sheep;
- $(pen, note)$: One can write a note, but not enclosed a note in a pen;
- For the contexts $(see, sheep)$ and $(see, note)$, only one outcome is possible (and indeed both *seeing a sheep* and *seeing a note* are sensible phrases).

We observe that this empirical model is *strongly contextual* since no local section can be extended to a global assignment.

In particular, there is nothing special about the *verb* − *object* situation, and other such examples can be found for *subject* − *verb*, *adjective* − *noun* situations as well, when considering a single ambiguous noun, verb or adjective. We will here describe a *subject* − *verb* experiment with, once again an ambiguous verb, but other examples for other situations can be found in the Appendix B.

We consider the measurement contexts $\{water, player\} \times \{dribble, fall\}$, where $\{water, player\}$ are the possible choices of subject, and $\{dribble, fall\}$ are possible choices of verbs. Here, the only ambiguous verb is the verb *dribble*, which has the possible readings:

a. *to dribble*: to drip slowly; here both pouring a liquid and drooling are contained within this meaning.

   Example: *The cold tap is only dribbling, it is driving me crazy.*

b. *to dribble*: to take the ball past an opponent (in football, hockey or basketball)

   Example: *The footballer dribbled the ball from the goal area.*

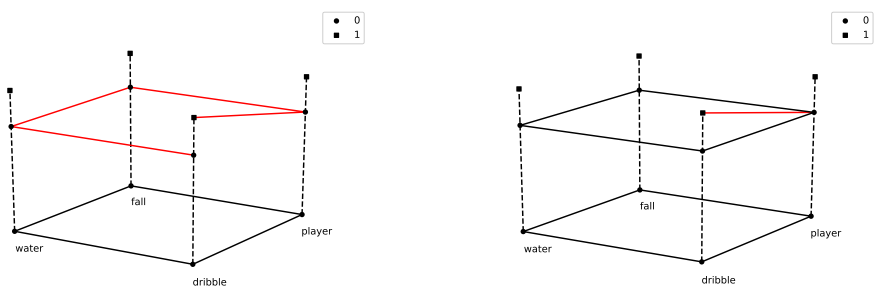
As for the previous example, the meanings are encoded as follows:

Encoding	Meanings of			
	water	*player*	*dribble*	*fall*
0	*water*	*player*	*drip*	*fall*
1	$\star$	$\star$	*ball*	$\star$

Table 3.2: Encoding of meanings of *water*, *player*, *dribble* and *fall*.

Here, some assumptions about plausibility have to be made, and affects the degree of contextuality of the model (but not the contextual nature of the measurement contexts). Indeed, we consider that the only possible readings are:
- for $(water, dribble)$: water drips;
- for $(player, dribble)$: the player dribbles the ball;
- the only possible readings of $(water, fall)$ and $(player, fall)$.

Then the empirical model is strongly contextual (see Fig. 3.3a). However, if one assumes that the reading $(player, dribble)$, understood as *the player is driveling*, is possible, then the model becomes *weakly contextual* (see Fig. 3.3b).



(a) Strong model. (b) Weak model.

Figure 3.3: Bundle diagram representations associated with the measurement contexts $\{water, player\} \times \{dribble, fall\}$. Once again, the local sections which cannot be extended are depicted in red.

We are also interested in the behaviour of such empirical models when the number of ambiguous words considered is increased. We start by comparing two ambiguous contexts, interacting with two unambiguous ones. The example presented here will be to consider the ambiguous nouns $\{coach, kid\}$ as the subjects of the (unambiguous) verbs $\{play, drive\}$. The different meanings of *coach* are:

a. *coach*: Person in charge of an athlete or a sport team

Example: *The coach was very pleased with the teams' performance.*

b. *coach*: bus

Example: *We can park the coach behind this building.*

Similarly, the considered meanings of *kid* are:

a. *kid*: Child

Example: *This is a story we tell to kids.*

b. *kid*: Young goat

Example: *The goat is taking care of its kids.*

Therefore, the different readings of each words in the empirical model are encoded as:

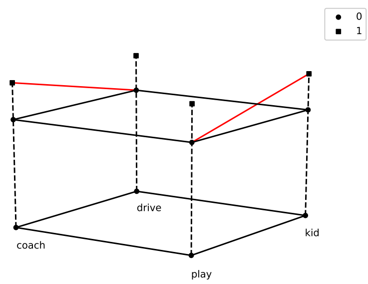
Encoding	Meanings of			
	coach	*kid*	*play*	*drive*
0	*sport*	*child*	*play*	*drive*
1	*bus*	*goat*	$\star$	$\star$

Table 3.3: Encoding of meanings of *water*, *player*, *dribble* and *fall*.

And we argue that the corresponding empirical model is as depicted in Fig. 3.4. The

subject	*verb*	(0,0)	(0,1)	(1,0)	(1,1)
coach	*play*	1	0	0	0
coach	*drive*	1	0	1	0
kid	*play*	1	0	1	0
kid	*drive*	1	0	0	0

(a) Empirical model



(b) Bundle diagram representation

Figure 3.4: Empirical model associated with the measurement contexts $\{coach, kid\} \times \{play, drive\}$.

reasoning for the example goes as follows:

- $(coach, play)$: The only sensible meaning would be the (sport) coach plays (e.g. *The coach played football professionally when he was younger*);

- $(coach, drive)$: Both meanings of *coach* can be activated here, for example *The coach drive a collection car* or *The coach*(bus) *drove through the night* are instances of each;

- $(kid, play)$: Both meanings of *kid* can also be activated in this context. For example: *the kids are playing in the playground* can activate the *child* meaning, whilst the reading *goat* of kid is possible in *the kids are playing in the field*;
- $(kid, drive)$: Only the (human) child meaning can be activated here (e.g. *You can't let kids drive, that's illegal*).

This model is weakly contextual as the local sections $[(kid, play) \mapsto (goat, play)]$ and $[(coach, drive) \mapsto (bus, drive)]$ are not part of any global assignment.

Finally, more interesting examples arise when all individual contexts are ambiguous. We start with a $verb - object$ example. Consider the verbs $\{saw, tap\}$ with possible meanings:

1. For *saw*:
    a. *saw*: Past tense of *to see*
    b. *to saw*: To cut with a saw
2. For *tap*:
    a. *to tap*: to gently touch
    Example: *He tapped his friend on the shoulder.*
    b. *to tap*: To secretly record
    Example: *They were shocked when they discovered that all their phones were tapped.*

Note that *tap* has many more meanings (e.g. tap dance, make use of, ...), but these will not be activated in the following *object* contexts $\{beam, cabinet\}$. Now, both of *beam* and *cabinet* are ambiguous nouns, with respective meanings:

1. For *beam*:
    a. *beam*: Column of light or particles
    Example: *The light beam from the lamp illuminated the room.*
    b. *beam*: Long piece of wood, metal or concrete
    Example: *Wooden beams are characteristic of Tudor houses.*
2. For *cabinet*:
    a. *cabinet*: Group of people appointed by the head of state/political party
    Example: *The Prime Minister has appointed their new cabinet.*
    b. *cabinet*: Piece of furniture
    Example: *It's an old cabinet that I inherited from my grandparents.*

In particular, the possible readings of the different phrases are:

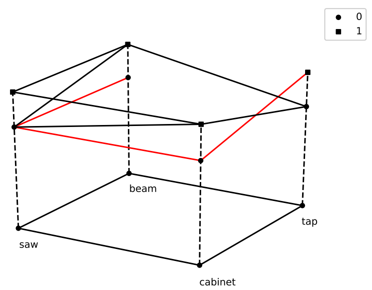
- $(saw, beam)$: See a (e.g. light) beam or a (e.g. wooden) beam or cut a (wooden) beam with a saw;
- $(saw, cabinet)$: See the members of the cabinet, see the piece of furniture or cut the piece of furniture with a saw;
- $(tap, beam)$: Touch a (e.g. wooden) beam;
- $(tap, cabinet)$: Touch the piece of furniture, or secretly record the members of the cabinet.

while the following readings are impossible:
- (*saw*, *beam*): Cut a (e.g. light) beam with a saw;
- (*saw*, *cabinet*): Cut the governing body with a saw;
- (*tap*, *beam*): Touch a (e.g. light) beam, or secretly record either a (e.g light) beam or a (e.g. wooden) beam;
- (*tap*, *cabinet*): Tap in the "touch" sense the cabinet (government) (or more generally any group of people) or secretly record furniture.

The obtained empirical model and its meaning encoding are depicted in Fig. 3.5, and is weakly contextual.

Encoding	Meanings of *saw*	*tap*	*beam*	*cabinet*
0	*see*	*touch*	*light*	*government*
1	*cut*	*record*	*wooden*	*furniture*

(a) Encoding of meanings of *saw*, *tap*, *cabinet* and *beam*.



(b) Bundle diagram representation

Figure 3.5: Empirical model associated with the measurement contexts $\{saw, tap\} \times \{beam, cabinet\}$.

A similar analysis can be done on the *subject−verb* example with contexts $\{coach, boxer\} \times \{lap, file\}$. Indeed, we consider the same meanings (and same encoding) of *coach*, and the following meaning of the other contexts:

1. For *boxer*
   a. *boxer*: Professional athlete practising boxing
   b. *boxer:* Breed of dog
2. For *lap*
   a. *to lap*: To run/move around a circular track
      Example: *The runner lapped the park fastest*
   b. *to lap*: To drink or to lick
      Example: *The dog was so thirsty it lapped up all its water in seconds.*
3. For *file*
   a. *to file*: To document or register information (in a file)
      Example: *This is not acceptable, you need to file a complaint.*
   b. *to file*: To smooth (with a file)
      Example: *After trimming her nails, she filed them.*

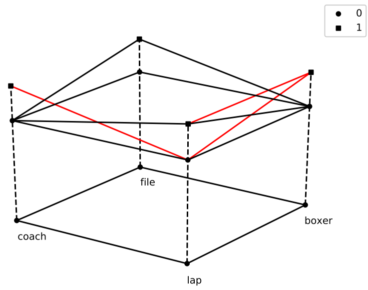
We make the following assumptions:
- Humans and dogs can both run on a track and drink;
- A bus can drive around a circular path, but cannot drink;
- Humans can document information, and smooth objects;

- Neither dogs nor buses can file information or use filing tools.

This gives rise to a weakly contextual model depicted in Fig. 3.6

Encoding	Meanings of *boxer*	*lap*	*file*
0	*boxing*	*run*	*document*
1	*dog*	*drink*	*smooth*

(a) Encoding of meanings of *boxer*, *lap* and *file*.



(b) Bundle diagram representation

Figure 3.6: Empirical model associated with the measurement contexts $\{coach, boxer\} \times \{lap, file\}$.

Finally, we will describe an *adjective* $-$ *noun* example, with measurement contexts $\{green, light\} \times \{bark, cabinet\}$. The different meanings considered are expressed as:

1. For *green*

   a. *green*: Colour
      Example: *They stared at the green door wondering whether they had the correct address*

   b. *green*: Environmental-friendly.
      Example: *The Green Party did surprisingly well in the last election.*

2. For *light*

   a. *light*: Not heavy
      Example: *He easily managed to lift the light suitcase.*

   b. *light*: Light in color, pale
      Example: *The sky was light blue yesterday.*

3. For *bark*

   a. *bark*: Tough protective covering of trees and other woody plants.

   b. *bark*: Noise made by a dog, or similar noises

and *cabinet* defined as before. Now, we once again describe the assumption made in the studied model:

- The bark of a tree can be green (colour) or environmental-friendly. Noises on the other hand cannot be either;

- Furniture and governments can also be green (colour) or environment-friendly, but members of the cabinet cannot be green in colour;

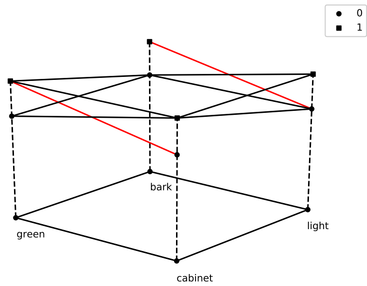
- Both the bark a tree or the bark of a dog can be “not heavy”, but only the bark of a tree can be light in colour;

- The Cabinet (government) as a whole cannot be described as *light*, but furniture can be both light-coloured and not heavy.

This also gives rise to a (weakly) contextual empirical model depicted in Fig. 3.7.

Encoding	Meanings of *green*	*light*	*bark*
0	*colour*	*not heavy*	*tree*
1	*environment*	*colour*	*noise*

(a) Encoding of meanings of *green*, *light* and *bark*.



(b) Bundle diagram representation

Figure 3.7: Empirical model associated with the measurement contexts $\{green, light\} \times \{cabinet, bark\}$.

Discussion of the results Recall that global assignments are consistent assignments across all of the measurement contexts; in this framework, a global assignment therefore means that sense attribution can be made, to some extent, logically. For example, given that $[(green, cabinet) \mapsto (colour, furniture)]$, $[(green, bark) \mapsto (colour, tree)]$ and $[(light, cabinet) \mapsto (not\ heavy, furniture)]$, the next "logical" statement $[(light, bark) \mapsto (not\ heavy, tree)]$ is indeed possible. Hence, if all local sections can be extended to a global assignment, then the empirical model can be described as the "superposition"[2] of all the global assignments.

What we have shown with these examples however, is that the meaning of individual words in composite phrases cannot be described by a any global distribution on a set of intersecting measurement contexts: it is *contextual*. For example, in the simple model in Fig. 3.2, this analysis shows that knowing how the meaning *see* interacts with both *sheep* and *note*, and how the meaning of *pen* interacts with *note* gives the wrong prediction as to which meaning of *pen* is activated when interacting with *sheep*.

We have been restricting, in this section to words of a given grammatical type, or at least meanings of a word assuming that its grammatical type is known. Now, some words can be seen as either a verb or a noun (e.g. *pen*), or as either an adjective or a noun (e.g. *light*) etc. Hence, considering words with this property can therefore lead to phrases where the meaning *and* the grammatical type can be ambiguous. For example, the phrase *train coaches* can either be a verb-phrase (i.e. *training sport coaches*) or a noun-phrase (i.e. *carriages of a train*), and, as before, different individual meanings can be activated. In the next section we will focus on these types of examples.

3.1.2 Ambiguous syntactic combinations

We relax the above model as to allow each word to occupy different parts-of-speech, but keeping the word order in which they would appear in a text, i.e. the word chosen by A

[2]Superposition is used in a very generic sense here. In this case of possibilistic models, superposition can be seen as a disjunction of propositions.

will appear before the word chosen by B in a coherent two-words phrase. We will moreover consider ambiguous words, which are ambiguous both in meaning and grammatical types. We will, as before, subsequently record the meaning activated after interaction.

We start by considering combinations of words which can both be verbs and nouns. We will study the example for which the agent A chooses contexts in the set $\{saw, leaves\}$ and B chooses its context within $\{fall, bark\}$. The readings of each of these contexts are:

1. For *saw*
   a. *saw* (verb) : Past tense of *see*
   b. *to saw* (verb): Cut with a saw
   c. *saw* (noun): Tool for sawing
2. For *leaves*
   a. *leaves* (verb) : Conjugated form of the verb *to leave*
   b. *leaves* (noun): Plural of *leaf*
3. For *fall*
   a. *to fall* (verb): To move from a higher position to a lower one
   Example: *They saw him fall from his bike.*
   b. *fall* (noun): Action of falling
   Example: *In 1989, the world witnessed the fall of the Berlin Wall.*
   c. *fall* (noun): American English for autumn
4. For *bark*
   a. *to bark* (verb): Make a sound similar to the bark of a dog
   Example: *She woke in the night to hear men barking instructions.*
   b. *bark* (noun): Tough protective covering of trees.
   c. *bark* (noun): Noise made by a dog, or similar noises

which gives rise to the following encoding:

Encoding	Meanings of			
	saw	*leaves*	*fall*	*bark*
0	*see*	*leave*	*falling*	*make noise*
1	*cut*	*leaf*	*action of falling*	*tree*
2	*tool*	$\star$	*autumn*	*noise*

Figure 3.8: Encoding of meanings of *saw*, *leaves*, *fall* and *bark*.

We now enumerate all the possible meaning combinations:

- In the case of *saw fall*: one can observe the fall of something or somebody as well as seasons passing; one cannot saw neither a fall nor a season and finally, a saw can fall;

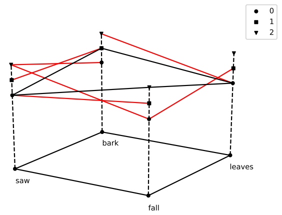
- For *saw bark*, we start from the assumption that sounds cannot be seen (but a tree bark can); similarly, sounds cannot be sawed, but a tree bark can , and a saw can bark;

- For *leaves fall*: the leaves of a tree can fall, and one can leave a season (e.g. *He leaves fall with bitterness*), and these are the only possible readings;
- Finally, in the case of *leaves bark*, only the $verb - noun$ readings are possible (since plant leaves cannot make loud noises), and both noun readings of *bark* are possible (e.g. *Dragging the bark leaves a trace on the floor*, or *the bark leaves the dog's mouth*).

This leads to the empirical model depicted in Fig. 3.9 which counts a unique global assignment.

A	B	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2
saw	$fall$	0	1	1	0	0	0	
saw	$bark$	0	1	0	0	1	0	
$leaves$	$fall$	0	0	1	1	0	0	
$leaves$	$bark$	0	1	1	0	0	0	

(a) Empirical model



(b) Bundle diagram representation

Figure 3.9: Empirical model associated with the measurement contexts $\{saw, leaves\} \times \{fall, bark\}$.

We now consider an empirical model for which the first agent chooses words which can be either an adjective or a noun. Specifically, we are interested in the possible meanings of $\{cold, light\} \times \{can, cast\}$. We then consider the following meanings:

1. For *cold*

   a. *cold* (noun) : The absence of heat
   Example: *I'm not really accustomed to the cold.*

   b. *cold* (noun): Mild viral disease
   Example: *He caught a cold, that's why he can't come.*

   c. *cold* (adj): Low temperature
   Example: *This is a very cold day.*

   d. *cold* (adj): Without human warmth
   Example: *She was very cold, she didn't smile all night.*

2. For *can*

   a. *can* (noun) : Tin container

   b. *to can* (verb): To preserve food in a can
   Example: *He cans his own sardines.*

   c. *can* (auxiliary verb): To be able to

3. For *cast*

   a. *to cast* (verb): To throw something in a particular direction
   Example: *The moon cast a dim light over the field*

   b. *to cast* (verb): To choose (actors)
   Example: *The director cast this famous actor for his new film*

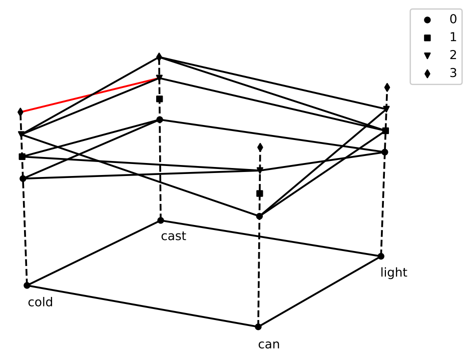
c. *cast* (noun): Group of actors
Example: *The performance from the cast was outstanding.*

d. *cast* (noun): Moulded object
Example: *I broke my arm and spent a few months in a plaster cast.*

There are many possible meaning combinations (see Fig. 3.10b), and the contextual nature of the model arises from the fact that only the *cast* of a film or a play can be *cold* in the "distant" sense.

Encoding	Meanings of			
	cold	*light*	*can*	*cast*
0	*winter*	*photons*	*tin*	*throw*
1	*illness*	*not heavy*	*preserve*	*choose*
2	*temperature*	*colour*	*able to*	*actors*
3	*distant*	$\star$	$\star$	*plaster*

(a) Encoding of meanings of *saw*, *leaves*, *fall* and *bark*.



(b) Bundle diagram representation

Figure 3.10: Empirical model associated with the measurement contexts $\{cold, light\} \times \{can, cast\}$.

Finally, we conclude this subsection by noting that syntactically ambiguous models can be studied using the stricter model of Section 3.1.1. Indeed, if the object of an auxiliary verb is a verb itself, then we can use the ambiguous verb *can*, with ambiguous objects *fish* and *spam*. In order to activate each meaning (and syntactic role) of *fish* and *spam*, we compare the effect of *can* on these objects with the unambiguous verbs *eat* and *will*. Hence, we only consider the "preserve" and "able to" meanings of *can*, the unique (verb) meanings of both *eat* and *will* as well as:

1. *Fish*:

   a. *fish* (noun) : Animal
   b. *to fish* (verb): To catch fishes
   c. *to fish* (verb): To look for something

2. *Spam*:

   a. *spam* (noun) : tinned ham

b. *to spam* (verb): To send (a lot of) unwanted emails

Now, considering the following encoding:

Encoding	Meanings of *can*	*fish*	*spam*
0	*preserve*	*animal*	*ham*
1	*able to*	*catch fish*	*email*
2	$\star$	*search*	$\star$

Figure 3.11: Encoding of meanings of *can*, *fish* and *spam*.

we obtain the following empirical model:

verb	*object*	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)	(2,2)
can	$fish$	1	0	0	0	1	1	0	0	0
can	$spam$	1	0	0	0	1	0	0	0	0
eat	$fish$	1	0	0	0	0	0	0	0	0
eat	$spam$	1	0	0	0	0	0	0	0	0
$will$	$fish$	0	1	1	0	0	0	0	0	0
$will$	$spam$	0	0	1	0	0	0	0	0	0

Figure 3.12: Empirical model associated with the measurement contexts $\{can, eat, will\} \times \{fish, spam\}$.

which is *strongly contextual*. Note that bundle diagrams for more than 2 choices of measurements from one party are considerably harder to read than for the other models considered here. However, the contextuality of the model can be shown from the observation that the meanings of *fish* and *spam* activated with the *eat* (respectively *will*) context do not extend to any local section in the support of the *will* (respectively *will*) context.

3.1.3 Discussion and further remarks

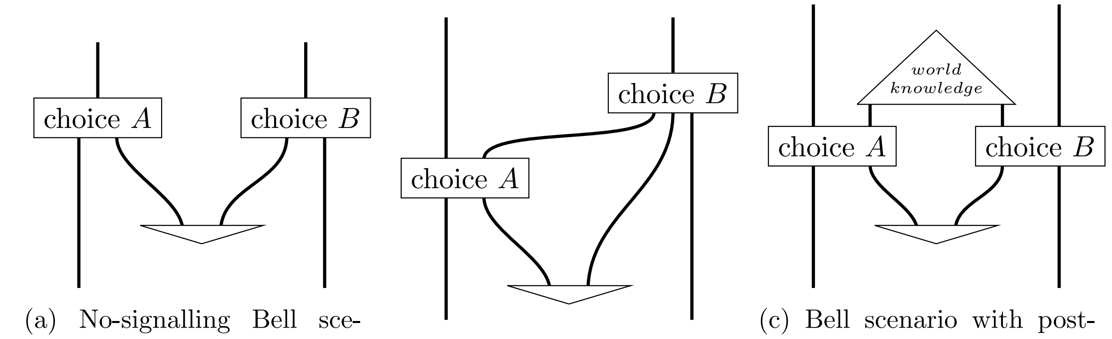
Signalling In the quantum mechanical setting, the no-signalling condition is traditionally considered as an essential property that an empirical model has to satisfy, in order to include the possibility that the two parties are space-like separated (and assuming the principle of causality). However, using the model presented in Fig. 3.2, we have:

$$(\mathit{pen}, \mathit{sheep})|_{\mathit{pen}} = [\mathit{pen} \mapsto \mathit{enclose}] \quad \neq \quad (\mathit{pen}, \mathit{note})|_{\mathit{pen}} = [\mathit{pen} \mapsto \mathit{write}] \tag{3.1}$$

Even more so, most of the possible choices of measurement contexts for A and B do not lead to a valid empirical model, for example we have seen that the following experiment

verb	*object*	(0,0)	(0,1)	(1,0)	(1,1)
pen	*sheep*	0	0	1	0
pen	*note*	1	0	0	0
herd	*sheep*	1	0	0	0
herd	*note*	0	0	0	0

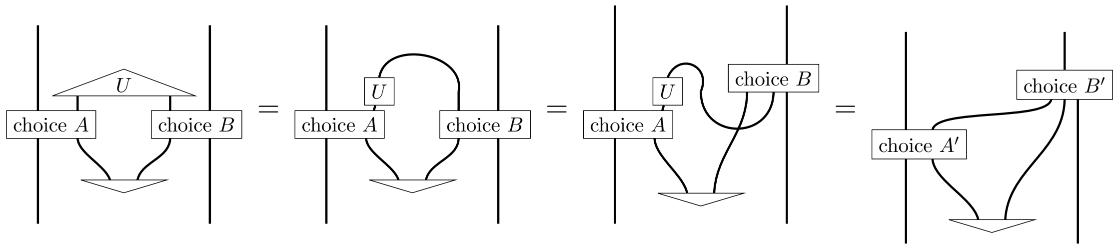
is not an empirical model, since the last row is not a distribution. All of these considerations boil down to the same principle, namely that some combinations are "rejected" because there is not a situation in our experience when something/someone herds a note or write a sheep. Hence, we believe that the apparent form of the empirical models is due to *post-selection*, which is operationally equivalent to a signalling scenario (see Fig. 3.13d).



(a) No-signalling Bell scenario

(b) Signalling Bell scenario

(c) Bell scenario with post-selection



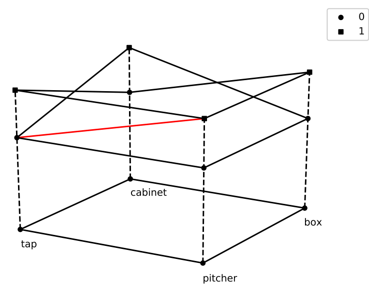
(d) Operational equivalent of post-selected and signalling systems.

On the other hand, signalling models have been studied in the context of *Contextuality-by-Default*[27, 26], which adopts the following philosophy: any (joint) distributions obtained in different measurement contexts are intrinsically *not* jointly distributed (they are all contextual "by default"), and the notion of contextuality merely reflects the fact that it might or might not be possible to *impose* a joint probability distribution on these stochastically uncorrelated distributions. Hence, from this point of view, the no-signalling property is not essential, and moreover hinders generalisations such as [25] where some measurements or properties are allowed to be "undefined" in some measurement contexts.

Finally, one may wonder if all such contextual examples in natural language which are signalling. The answer is no. Such an example is the following:

verb	*object*	(0,0)	(0,1)	(1,0)	(1,1)
tap	*pitcher*	1	1	0	1
tap	*cabinet*	0	1	1	0
box	*pitcher*	1	0	0	1
box	*cabinet*	0	1	1	0

(a) Empirical model



(b) Bundle diagram (local sections which cannot be extended are depicted in red).

In this example, the meanings of *tap* and *coach* are encoded as before, and:

Encoding	Meanings of *box*	Meanings of *pitcher*
0	*put in boxes*	*jug*
1	*fight*	*baseball player*

Figure 3.15: Encoding of meanings of *box* and *pitcher*.

Then, assuming that one can touch (tap) a single person but not a group of people (e.g. the cabinet), then the obtained model is weakly contextual and no-signalling. However, the above empirical model is no-signalling w.r.t its *possibilistic* distribution (i.e. w.r.t. to its support); however, as we will see in the analysis in Section 3.2, we expect that this model is signalling w.r.t. to its standard probabilistic representation.

Surreal and metaphorical contexts All of the above empirical models are filled w.r.t. common sense and realistic meaning. However, one may want to consider a much more general model, where all possible meanings combinations are indeed possible, and the only 0 probability events corresponds to the assignment of an "undefined" meaning to a given word. This is, in principle, not incompatible with the framework presented here, and in particular, *every choice of measurement contexts will lead to a valid empirical model*. However, if *all* possible meanings are allowed, then the empirical model cannot be possibilistically contextual. On the other hand, one may consider the *probabilisitic model* described in the next section: we would then expect that the sections which are not realistic or plausible can appear in text corpora with a low, but non-zero probability. Another experiment that one may consider is to appeal to human interpretation: similar to the experiment presented in [6], one could ask a number of people to assign a meaning to different ambiguous word combinations, including some implausible combinations, e.g. *sparkling yard*. This would, on the other hand, lead to a very different project.

Meaning selection and degree of contextuality Finally, some remarks can be made in the choices of different meanings that are considered. Indeed, while for some words, the different meanings are almost completely unrelated (e.g. *bark*), some of the different senses of the words considered are less clearly divided (e.g. *light*). More importantly, one can see that altering the ways of dividing the possible meanings changes the number of local sections which do not extend to a global assignment, and hence, can change the degree of contextuality of a model; for example, if all words are trivially associated with a single meaning, then all of these models become non-contextual. However, the choices in this report were made so that we obtain the most "coarse-grained" model for which contextuality is apparent; note that fine-graining any senses further can increase the contextuality of the model (i.e. local sections may split, hence, global assignments may not be globally consistent anymore), but not decrease the contextuality of each of the models.

Although many such possibilistically contextual examples can be found using this framework, some combinations of ambiguous words may be consistent with a superposition of global assignments. In order to further the analysis, we therefore want to consider such examples and investigate whether they are (probabilisitically contextual).

3.2 Probabilisitic contextuality

We consider the same types of models as for the previous section, but under a probabilsitic point of view. In particular, we note that we do not need to consider any of the previously discussed examples, as they were shown to be possibilisitically contextual (and recall that possibilisitc contextuality is strictly stronger than probabilistic contextuality). Hence, we will here consider empirical models which Boolean distribution presheaf has a global section.

The BNC The different probability distributions are obtained from the The British National Corpus (BNC)[1]. The BNC is an open-source text corpus comprising 100 million words, spread across documents of different nature (including press articles, fiction, transcription of spoken language, and academic publications). The BNC is part-of-speech tagged, hence, grammatical nature and the lemma form (i.e. singular and/or non-conjugated form of the word) is available. On the other hand, semantic interpretation of each of the recorded occurrences of the phrases on interest were made by hand[3].

3.2.1 Semantically ambiguous combinations

As in Section 3.1, we start with purely semantic ambiguous combinations, and fix the grammatical roles of each of the agents' choices.

The first example is a *verb−object* example with contexts in $\{saw, bore\} \times \{cabinet, beam\}$. In particular, in the case of *saw* and *bore*, we only consider the different verbs it refers to as different meanings. Hence, using the same encoding of *saw*, *beam* and *cabinet* as in Fig. 3.5a, with the following meaning of *bore*:

a. *bore*: Past tense of *to bear*

b. *to bore*: To make holes

   Example: *He bored holes on the wall.*

and adopting the encoding 0 : *bear*, 1 : *holes*. We can make some remarks on this choice of meanings under consideration. Firstly we are missing out a meaning of *to bore*, namely to make others bored, which would potentially be activated in the context of *cabinet* (i.e. governmental body); however, no such instance appeared in the BNC, and if it were going to appear in a larger corpus, this will lead to a possibilisitically contextual model, and hence automatically probabilistically contextual. In addition, the verb *to bear* is itself ambiguous (for example, the meaning of *to bear* is different in the two sentences *I can't bear the sound of a fork on a plate* or *He was bearing forks and plates from the cupboard.*); however, due to the lack of occurences of the meaning "endure" in with the present object-contexts, we decided to group all of the meanings of *to bear* together. The obtained empirical model is depicted in Fig.3.16.

In order to show the contextuality of this model, we use the fact that finding a global section on a probabilistic model is equivalent to solving the system $\mathbf{M}\mathbf{x} = \mathbf{v}$ in the positive real numbers, where $\mathbf{v}$ is the column vector associated with the empirical model (see Section 1.3), and $\mathbf{M}$ is the incidence matrix given by (1.15) (see Appendix A for the

[3]Sense-tagged corpora would have been more convenient to use here; however, these corpora are usually smaller, more restricted, and hence would not give us enough data.

verb	*object*	(0,0)	(0,1)	(1,0)	(1,1)
saw	*cabinet*	5/8	1/8	0	2/8
saw	*beam*	3/5	1/5	0	1/5
bore	*cabinet*	3/7	2/7	0	2/7
bore	*beam*	1/5	2/5	0	2/5

Figure 3.16: Empirical model associated with the measurement contexts $\{saw, bore\} \times \{cabinet, beam\}$.

full incidence matrix and empirical model vector). In particular since all elements of $\mathbf{x}$ have to be positive, then we need to have $x_9 = x_{10} = x_{11} = x_{12} = x_3 = x_4 = 0$ since:

$$x_9 + x_{10} + x_{11} + x_{12} = s_3 = 0 \tag{3.2}$$
$$x_3 + x_4 + x_{11} + x_{12} = s_{11} = 0 \tag{3.3}$$

Then, using local sections s_1 and s_9, this gives respectively:

$$x_1 + x_2 = \frac{3}{7} \tag{3.4}$$
$$x_1 + x_2 = \frac{5}{8} \tag{3.5}$$

which leads to a contradiction.

We now consider an *adjective* − *noun* model with adjective contexts *green* and *even*, and noun contexts *cabinet* and *tip*. We adopt the same meanings and encoding of *green* as described in Fig. 3.7a and of *cabinet* as above. We moreover consider the following readings of *even* and *tip*:

1. For *even*:
   a. *even* : Flat and smooth
      Example: *You can see your reflection on the even surface of the water.*
   b. *even*: Balanced, of equal size
      Example: *The school achieved an even gender balance among staff and students.*
2. For *tip*
   a. *tip*: Advice, indications of potential leads
      Example: *They couldn't solve the problem, but the teacher gave them a tip.*
   b. *tip*: The (pointed) extremity of something
      Example: *I can barely reach it from the tips of my fingers.*

The recorded frequencies are recorded in the empirical model in Fig. 3.17

The support of this model has exactly two global assignments (see Fig. 3.18), namely $[(green, even, cabinet, tip) \mapsto (colour, flat, furniture, pointed)]$ and $[(green, even, cabinet, tip) \mapsto (environment, balanced, government, advice)]$. Hence, checking whether the probabilistic model of Fig. 3.17b has a global section is exactly the same as finding $\alpha, \beta \in [0, 1]$ such that $\alpha + \beta = 1$ and the empirical model is given as:

$$\begin{aligned} &\alpha \cdot [(green, even, cabinet, tip) \mapsto (colour, flat, furniture, pointed)] \\ +\ &\beta \cdot [(green, even, cabinet, tip) \mapsto (environment, balanced, government, advice)] \end{aligned}$$

Encoding	Meanings of *even*	Meanings of *tip*
0	*flat*	*advice*
1	*balanced*	*pointed*

(a) Encoding of meanings of *even* and *tip*.

adj	*noun*	(0,0)	(0,1)	(1,0)	(1,1)
green	*cabinet*	0	1/8	7/8	0
green	*tip*	0	12/17	5/17	0
even	*cabinet*	0	1/22	21/22	0
even	*tip*	0	21/25	4/25	0

(b) Empirical model

Figure 3.17: Empirical model associated with the measurement contexts $\{green, even\} \times \{cabinet, tip\}$.

Or, in terms of tables:

$$\begin{array}{c|c|c|c} 0 & 1/8 & 7/8 & 0 \\ \hline 0 & 12/17 & 5/17 & 0 \\ \hline 0 & 1/22 & 21/22 & 0 \\ \hline 0 & 21/25 & 4/25 & 0 \end{array} = \begin{array}{c|c|c|c} 0 & \alpha & 0 & 0 \\ \hline 0 & \alpha & 0 & 0 \\ \hline 0 & \alpha & 0 & 0 \\ \hline 0 & \alpha & 0 & 0 \end{array} + \begin{array}{c|c|c|c} 0 & 0 & \beta & 0 \\ \hline 0 & 0 & \beta & 0 \\ \hline 0 & 0 & \beta & 0 \\ \hline 0 & 0 & \beta & 0 \end{array}$$

which is clearly impossible.

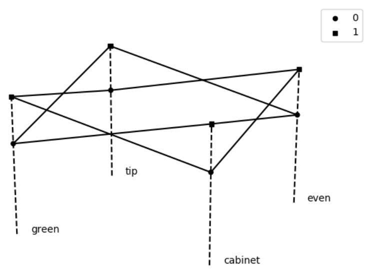


Figure 3.18: Bundle diagram for the possiblistic model associated with the measurement contexts $\{green, even\} \times \{cabinet, tip\}$.

3.2.2 Syntactically ambiguous combinations

Finally, consider an example of an empirical model with syntactically ambiguous contexts. We will consider the contexts $\{press, box\} \times \{can, leaves\}$ with the following meanings:

1. For *press*:

   a. *to press* (verb): Exert pressure upon something
   Example: *You can press Ctrl-Z to undo the previous operation.*

   b. *press* (noun): The print media publishing newspapers and magazines

   c. *press* (noun): Device used to apply pressure.
   Example: *They used to use printing presses before the invention of printers.*

2. For *box*

   a. *to box* (verb): To put in a box
   Example: *Each piece is boxed with a certificate of authenticity.*

b. *to box* (verb): To fight, to practice boxing
Example: *He used to box for England.*

c. *box* (noun): Container
Example: *They sent a box of chocolates as an apology.*

with the encoding:

Encoding	Meanings of	
	press	*box*
0	*push*	*put in boxes*
1	*media*	*fight*
2	*machine*	*container*

Figure 3.19: Encoding of meanings of *press* and *box*.

and *can* and *leaves* defined (and encoded) as in Section 3.1.2.

As in Section 3.2.1, we record the respective frequencies associated with each context from the BNC, giving the empirical model in Fig. 3.20.

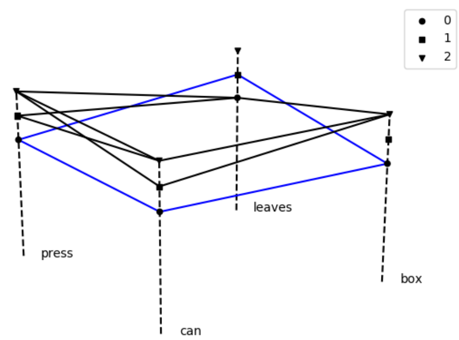
A	B	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)	(2,2)
press	*can*	2/25	0	0	0	0	41/50	0	1/50	2/25
press	*leaves*	0	6/13	0	5/13	0	0	2/13	0	0
box	*can*	7/74	0	0	0	0	0	0	1/74	33/37
box	*leaves*	0	2/3	0	0	0	0	1/3	0	0

Figure 3.20: Empirical model associated with $\{press, box\} \times \{can, leaves\}$.

As for the previous example, one can show that this model is probabilistically contextual by noting that the only global assignment of the support which includes $[press \mapsto push]$ is $[(press, box, can, leaves) \mapsto (push, put\ in\ boxes, tin, leaf)]$ (see Fig. 3.21). This implies that all the probabilities from the support global assignment have to be equal in order to have a global assignment in the possbilistic model as well; since this is not the case, we conclude that the probabilistic model is indeed contextual.

A	B	(0,0)	(0,1)	(0,2)	...
press	*can*	1	0	0	...
press	*leaves*	0	1	0	...
box	*can*	1	0	0	...
box	*leaves*	0	1	0	...

(a) Support of the model



(b) Bundle diagram

Figure 3.21: Possibilistic analysis of the empirical model associated with $\{press, box\} \times \{can, leaves\}$. The global assignment of interest is depicted in blue.

3.2.3 Discussion of the results

The fact that all of these models are (probabilisitically) contextual shows that the obtained probability distributions cannot be described as the marginals of a global probability distribution over all possible global assignments. That is, the measurement context which is considered plays an important role in how the meaning probabilities are assigned. In future work, we would like to get a quantitative measure of the contextuality of the different empirical models presented here, using for example using the contextual fraction[4] or the "degree of contextuality" from the Contextuality-by-default framework[25].

Signalling Recall that most of the models described in 3.1 were *signalling* (in the Boolean distributions). Hence, one can ask whether these probabilistic models are also signalling (note that their Boolean distributions are no-signalling). As we did in Section 3.1, we consider the marginal probabilities from, say Fig.3.16, and observe that:

$$
\begin{aligned}
(saw, cabinet)|_{saw} &= \left(\frac{5}{8} + \frac{1}{8}\right)[saw \mapsto see] + \frac{1}{4}[saw \mapsto cut] \\
&= \frac{3}{4}[saw \mapsto see] + \frac{1}{4}[saw \mapsto cut] \qquad (3.6)\\
(saw, beam)|_{saw} &= \left(\frac{3}{5} + \frac{1}{5}\right)[saw \mapsto see] + \frac{1}{5}[saw \mapsto cut] \\
&= \frac{4}{5} + \frac{1}{5}[saw \mapsto see] + \frac{1}{5}[saw \mapsto cut] \qquad (3.7)
\end{aligned}
$$

i.e. $(saw, cabinet)|_{saw} \neq (saw, beam)|_{saw}$, and the model is again signalling. Similar analysis can be made for the other probabilistic models.

In this chapter, we have studied the meaning behaviour of combinations of ambiguous words. In particular, we have shown that meaning combinations in natural language exhibits features which are similar to *contextuality* as defined in the context of quantum mechanics.

Chapter 4

Syntactic models and Garden-path sentences

In this chapter, we will move away from Bell-type experiments, and study a way of using presheaves in disambiguitating sentences that are syntactically and semantically ambiguous. Indeed, the model considered here will not make use of any agent, and measurement contexts will correspond to different subphrases of a given sentence. In particular, we will focus on so-called *garden-path sentences*. Garden-path sentences are sentences which, although grammatically correct, we lead the reader to adopt a likely interpretation (according to its experience), which will turn out to be incorrect; for example *I convinced her children are noisy*, or *The man who hunts ducks out on weekends*.

Recall that presheaves are used to study the transition from local behaviour to global behaviour. With this analysis, we expect to model this effect using presheaves, by simulating how different possible meanings are selected to construct bigger and bigger phrases. We start by constructing all the possible choices of interpretations that can be made as one reads the sentence, and then identify the only one that is grammatically consistent. We then add probabilities to the framework as to "rank" the possible choices at each stage to then conclude that the choices with higher probabilities do not lead to the correct interpretation. We will consider the following examples of garden-path sentences:

- *Complex houses students*
- *The horse raced fell*
- *The old man the boat*

Sentences to topological spaces For each word in the sentence, we associate a vertex (0-simplex). Then, higher order simplices are added as follows: we add a $n-1$-simplex ($n-1$-dimensional generalised triangle) across n words if the n-word combination can form a n-words phrase that can be found in natural language, or even, if they are grammatically sound. For example:

- If a sentence contains the words *old* and *man*, an edge (1-simplex) will join the vertices representing *old* and *man*;

- If a sentence contains the words *the*, *old* and *man*, a triangle (2-simplex) will be joining *the*, *old* and *man*;

- If a sentence contains the words *the*, *very*, *old* and *man*, the respective vertices will for a tetrahedron (3-simplex);

- etc.

Note that a $n+1$-words (grammatical) sentence should be represented by a n-simplex. Hence, since n-simplices with $n > 4$ are hard to represent on the plane, we will simplify some of the diagrams (but the reasoning also generalises to higher dimensions), in particular, only the critical simplices will be represented, i.e. simplices which plays a direct role in the comprehension of garden path sentences. For example, for the sentence *the old man the boat* we will consider the simplicial complex in Fig.4.1.

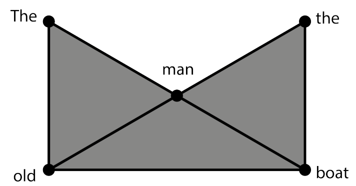


Figure 4.1: *The old man the boat.* The simplices that are ignored are *The old man boat*, *old man the boat* and the overall sentence *the old man the boat.*

Constructing a presheaf For each word, we associate its set of meanings. For each of the larger simplices, we select the different possible meanings of each word which can be activated in the corresponding phrase (see Fig. 4.2).

Here, a global assignment will represent a consistent way of assigning meaning to all the words s.t. the global sentence is grammatically correct. Hence, by definition of a garden path sentence, such a global assignment will exist. However, we expect that, when adding probabilistic weights, the simplices with the "highest" probabilities will not form an overall simplex, i.e. is not grammatically consistent.

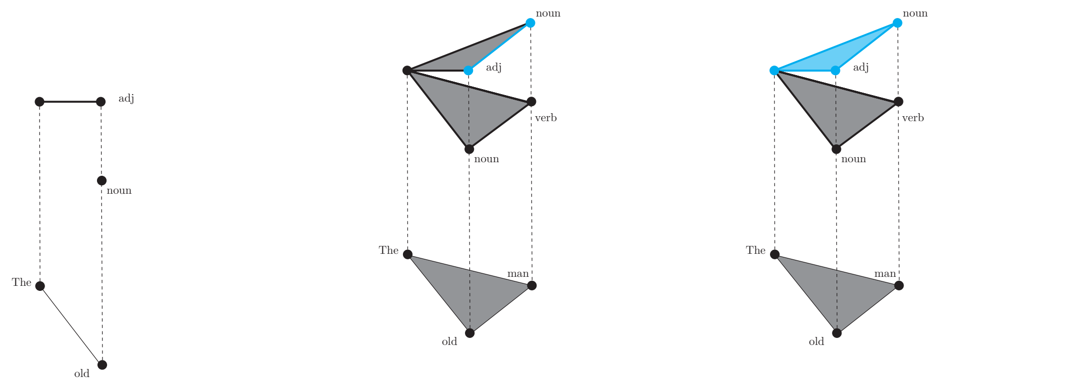
4.1 Qualitative and logical analysis of garden-path sentences

In this section, we will be interested in the global coherence of the sentences, the peculiar features of garden-path sentences will be investigated in the next section.

We start with the easiest three-words garden-path sentence *Complex houses students.* Here, the ambiguity comes from the words *complex*, which can be a noun or an adjective (or compound-noun, but it is easier to count this as an adjective here), and *houses* which can be a noun or a verb. In particular, the compound *houses students* can only be seen as a *verb-noun* phrase, hence, the ambiguity mainly comes from the phrase *complex houses*, which can be:

- Houses in a complex (*adjective-noun*);
- The complex is hosting (*noun-verb*);
- Complicated houses (*adjective-noun*).

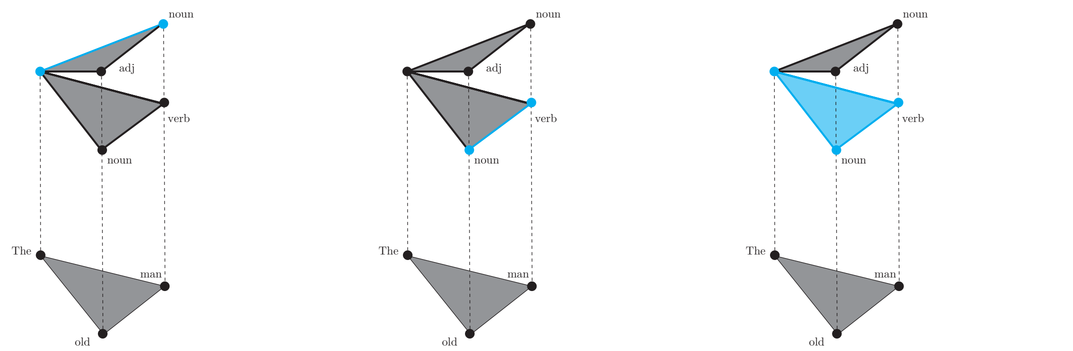
The latter is a very unusual meaning, which was not found in the BNC, and we will, therefore, decide to ignore it. In particular, we note that the most common meanings of *complex* and *houses* do not give rise to a very likely meaning composition, which echos the analysis carried out in Chapters 3. In this example (see Fig. 4.3), the only correct interpretation, which is coherent as a whole sentence, is: *the complex provides accommodation for students.*



(a) The man

(b) Old (adjective) man (noun)

(c) The old (adjective) man(noun)



(d) The old(noun)

(e) Old(noun) man(verb)

(f) The old(noun) man(verb)

Figure 4.2: Bundle diagram of the phrase *the old man* and meaning of all the simplices (the simplices of interest are depicted in blue).

We now want to do the same analysis for the sentence *The horse raced fell.* In this example, all the words are unambiguous apart from *raced* which can be either an adjective or a verb. The possible meanings of the phrase *The horse raced* are the following:

- *The horse was running in a competition* or *The horse was running* – we will consider those both as the same;

- *The horse that was raced* (e.g. *in a competition*) or *The horse was overtaken* – as for the other syntactical choice, we will assume those two meanings to be equivalent.

The corresponding bundle diagram is shown in Fig. 4.4.

We finally present a more intricate example, *The old man the boat.* As previously mentioned, this example is too complex to represent in full, but a simpler complex can be used (and the remaining simplices can be extended in the only possible way). There are two sources of ambiguity in this sentence, namely *old* (either adjective or noun) and *man* (either noun or verb). Accordingly, the phrases *The old man* and *old man boat* have two possible meanings which we will discuss:

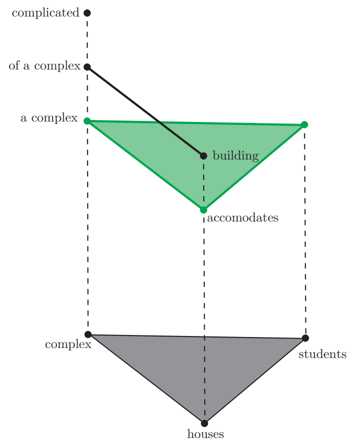


Figure 4.3: Bundle diagram associated with the sentence *Complex houses students*. The global assignment *The complex accommodates students* is depicted in green.

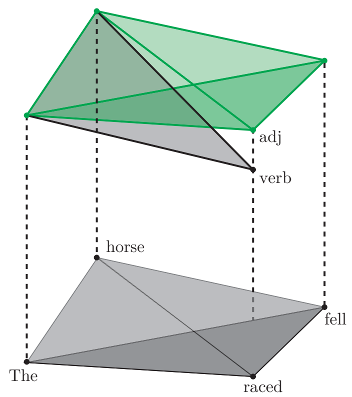
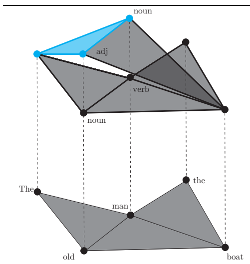
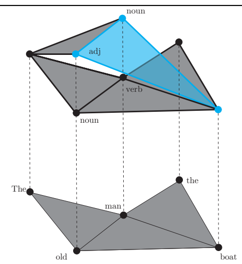
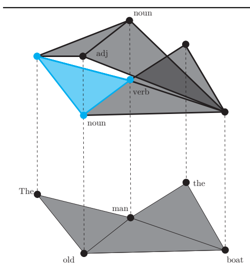
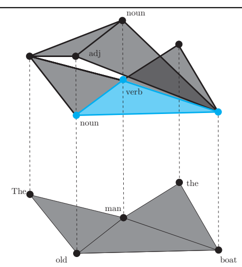


Figure 4.4: Bundle diagram of *The horse raced fell*. The highlighted complex corresponds to the unique global assignment associated to this sentence, and the only correct reading is *The horse that was raced fell*.

The old man		*old man boat*	
Local section	Meaning	Local section	Meaning
	The aged person		Boat for old people
	The “old” is taking care of		the “old” is taking care of the boat

The only reading for which *the old man the boat* is coherent as a sentence is *The "old" is taking care of the boat.*

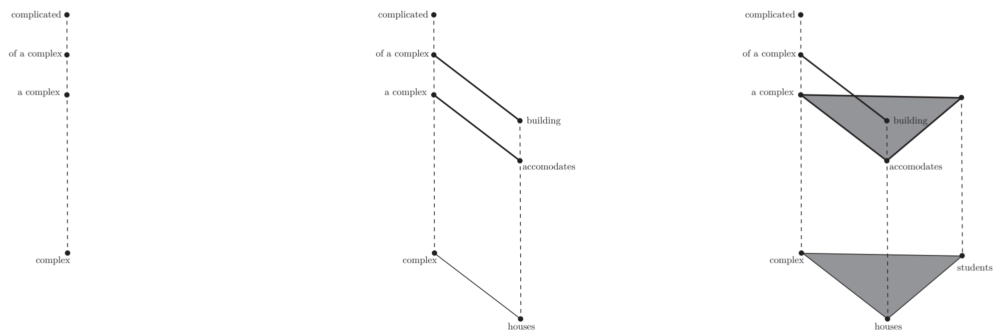
4.2 Probabilities in global assignments

In order to understand the peculiar nature of garden-path sentences, we decompose the considered sentences in terms of the different chronological "stages", for example in the sentence *Complex houses students*, we consider the stages *Complex*, *Complex houses* and finally *Complex houses students*, and compute the different degrees of likeliness of the competing meanings from data collected from the BNC.

We first note that the initial stage is always a single word without context, so we can assume that the word is in its "ground state", i.e. all possible meanings are equally probable. Hence, we start the analysis of the sentence *Complex houses students* at the stage *Complex houses*, where the two competing meanings are "houses in a complex" and "the complex accomodates". At this stage, the possiblistic distribution presheaf associated with *Complex houses* is depicted (as a bundle diagram) in Fig. 4.5b. We note that this presheaf has two global assignments and *a global section.* In particular, there is also a global section in the probabilistic distribution presheaf, which is is given by:

$$d_1(x) = \begin{cases} \frac{2}{3} & \text{if } x = [(complex, houses) \mapsto (adjective, noun)] \\ \frac{1}{3} & \text{if } x = [(complex, houses) \mapsto (noun, verb)] \end{cases} \tag{4.1}$$

where these probabilities are obtained from observed frequencies in the BNC. Now, we note that the global assignment, at the stage *Complex houses*, with highest probability does not extend to a global assignment in the next stage: *Complex houses students* (see Fig. 4.5). Hence, if these observed frequency are representative of the "degree of comitte-



(a) *Complex* (b) *Complex houses* (c) *Complex houses students*

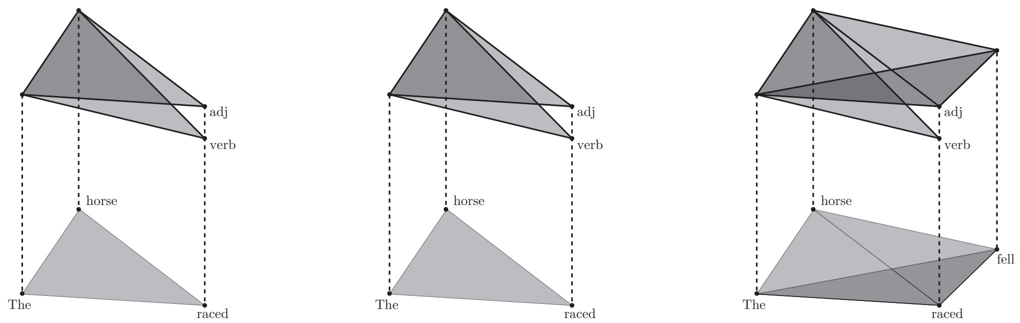
Figure 4.5: Decomposition of the sentence *Complex houses students* in its different stages.

ment" that one associate with a given reading, then this explains the apparent difficulty of the comprehension of the sentence since one has to switch the rankings of plausibility of the different meanings of the phrase *Complex houses* at the stage *Complex houses students.*

A similar observation can be made for the garden-path sentence *The horse raced fell.* Indeed, the phrases are all unambiguous up to the stage *The horse raced.* At this stage, one sees that there are two global global assignments, leading to a global section in the probabilistic distribution presheaf:

$$d_2(x) = \begin{cases} \frac{71}{90} \simeq 0.79 & \text{if } x = [raced \mapsto verb] \\ \frac{19}{90} \simeq 0.21 & \text{if } x = [raced \mapsto adjective] \end{cases} \tag{4.2}$$

(note that *the* and *horse* are unambiguous, i.e. has a deterministic outcome under this model, and doesn't add any information to the global assignments). As for the previous example, one see that the global assignment with the highest probability at the stage *The horse raced* does not extend to a global assignment in the final stage *The horse raced fell*, which implies that, once again, the reader has to revise its choice of most likely reading at a later stage.



(a) *The horse* (b) *The horse raced* (c) *The horse raced fell*

Figure 4.6: Decomposition of the sentence *The horse raced fell* in its different stages.

Finally, in the example *The old man the boat*, we observe a progression as depicted in Fig.4.8. The turning point occurs at the stage *The old man* (Fig. 4.7b), where one can find the following global section (in the probabilstic distribution presheaf):

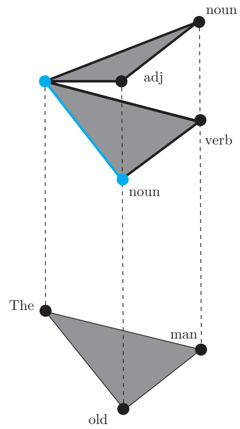
$$d_3(x) = \begin{cases} \frac{3559}{3561} \simeq 0.9994 & \text{if } x = [(old, man) \mapsto (adjective, noun)] \\ \frac{2}{3561} \simeq 0.0006 & \text{if } x = [(old, man) \mapsto (noun, verb)] \end{cases} \tag{4.3}$$

where once again, the meaning which is not part of the global assignment at the final stage comes up with (a much) higher probability. We also make a couple of additional remarks on this example. At the stage, *The old*, there is only one possible global assignment, namely with *old* being a noun; even if this is fairly unlikely to appear in natural language, if the phrase *The old* appears without context, the only possible meaning of *the whole phrase* available to a reader is indeed to consider *old* as a noun. In addition, at the final stage appears an another local section compatible with *old man* being an adjective-noun phrase, which is however, not part of a global assignment. We decided to include this local section since a reader might consider an incomplete or grammatically incorrect meaning instead of the odd meaning presented. Indeed, these two local sections appeared in our dataset with respective frequencies:

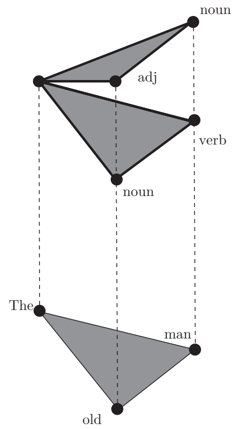
$$P([(old, man, boat) \mapsto (adjective, noun, noun)]) = \frac{3}{4} \tag{4.4}$$

$$P([(old, man, boat) \mapsto (noun, verb, noun)]) = \frac{1}{4} \tag{4.5}$$

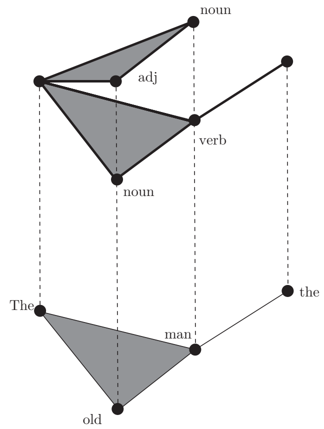
where the only occurrence of the latter was indeed the garden-path sentence considered here.



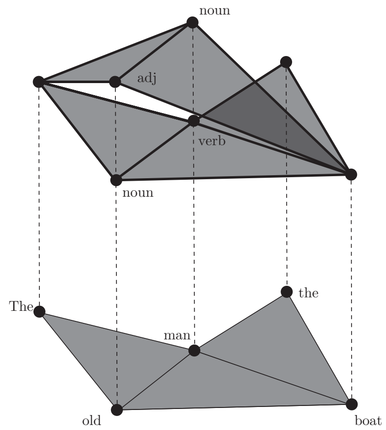
(a) *The old*



(b) *The old man*



(c) *The old man the*



(d) *The old man the boat*

Figure 4.7: Decomposition of the sentence *The horse raced fell* in its different stages.

Figure 4.8: Decomposition of the sentence *The old man the boat* in its different stages.

4.3 Discussion and future research directions

We interpret the obtained results as follows. The possibilistic analysis of the different garden-path sentences showed that all models were compatible with a unique global assignment; therefore, there is a *unique* way of reading those sentences which is compatible with the structure of English grammar. We also note that if there were more than one global assignment associated, this would mean that the sentence has an ambiguous meaning, and hence, that extra context is needed in order to fully disambiguate the sentence (for example the sentence *I can fish* would have three different global assignment, corresponding to *I am able to catch fishes*, *I am able to search for information* and *I put fishes in tin cans*). In addition, there were, in each possibilistic model, at least one local section which does not extend to a global assignment. This shows that there are some interpretations of subphrases which do not lead to a consistent meaning when considering the whole sentence. Finally, the probabilistic analysis unveiled that these globally inconsistent phrases come with higher probabilities, and hence the sentences appear "ungrammatical" or incorrect at first sight.

Future research directions The model presented in this chapter is fairly basic and highly made up of intuition, hence, we would want to study this model in more details in future work. In particular, it has a clear intuitionistic flavour, so one promising way of extending this research would be to use for example topos theory[32, 47], where the different stages of truth will be the chronological stages and so on. In addition, this bears similarities with the concept of filtrations from topological data analysis[50], so might want to investigate that as well.

A more long-term goal would be to embed this analysis into a general framework that could be used for disambiguation in a discourse. This would be of particular interest when considering ambiguous sentences, i.e. sentence with more than one global assignment.

Conclusion

In this report, we have investigated some of the contextual features of natural language, using the sheaf-theoretic formalism introduced in [5, 3]. In particular, we have shown that two-words phrases that are ambiguous do exhibit contextuality. However, the analogy with quantum measurement breaks when considering the marginal probability distribution: it does appear that these “local” probability distributions are distinguishable for different “measurement contexts”, which shows that some global interaction also needs to happen when meaning interact. Furthermore, we have shown that meaning combination also gives rise to contextual models when considering the combinations of words in sentences. This is effect is particularly magnified in sentences in which low-probability readings somehow carries the most prominent roles.

This project gives rise to promising avenues for future research. For example, some evidence suggests that the framework described in Chapter 3 for meaning combinations can be adapted into a model similar to the quantum-like concept combination model proposed by Aerts and Gabora (see [6, 7] and Section 2.2.2), where words would take the place of concepts, and meaning the place of exemplars. In addition, we might consider extending the quantitive analysis of this model by quantifying the degree of contextuality (e.g. using [4, 25]) and considering larger text corpora in order to have more accurate frequencies. As for the model presented in Chapter 4, an interesting continuation of this study might be to allow for references by identifying some of the simplicies of separate phrases. This would in particular be appealing in the context of anaphora resolution.

Bibliography

[1] The british national corpus. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium, `http://www.natcorp.ox.ac.uk/`, 2007.

[2] Samson Abramsky. Relational hidden variables and non-locality. *Studia Logica*, 101(2):411–452, Mar 2013.

[3] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield. Contextuality, cohomology and paradox, 2015.

[4] Samson Abramsky, Rui Soares Barbosa, and Shane Mansfield. Contextual fraction as a measure of contextuality. *Physical Review Letters*, 119(5), Aug 2017.

[5] Samson Abramsky and Adam Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036, Nov 2011.

[6] Diederik Aerts and Liane Gabora. A theory of concepts and their combinations I. *Kybernetes*, 34(1/2):167–191, Jan 2005.

[7] Diederik Aerts and Liane Gabora. A theory of concepts and their combinations II: A Hilbert space representation. *Kybernetes*, 34(1/2):192–221, Jan 2005.

[8] Daniela Ashoush and Bob Coecke. Dual density operators and natural language meaning. In *SLPCS@QPL*, 2016.

[9] Desislava Bankova, Bob Coecke, Martha Lewis, and Daniel Marsden. Graded entailment for compositional distributional semantics, 2016.

[10] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1:195–200, Nov 1964.

[11] William Blacoe, Elham Kashefi, and Mirella Lapata. A quantum-theoretic approach to distributional semantics. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 847–857, Atlanta, Georgia, June 2013. Association for Computational Linguistics.

[12] D. Bohm and Y. Aharonov. Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky. *Phys. Rev.*, 108:1070–1076, Nov 1957.

[13] Johan Bos. Towards wide-coverage semantic interpretation.

[14] Adam Brandenburger and Noson Yanofsky. A classification of hidden-variable properties. *Journal of Physics A: Mathematical and Theoretical*, 41(42):425302, sep 2008.

[15] Curt Burgess and Kevin Lund. Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12(2-3):177–210, 1997.

[16] J. Butterfield and C. J. Isham. A Topos Perspective on the Kochen-Specker Theorem: IV. Interval Valuations, 2001.

[17] Stephen Clark and James R. Curran. Wide-coverage efficient statistical parsing with ccg and log-linear models. *Computational Linguistics*, 33(4):493–552, 2007.

[18] Stephen Clark and S. Pulman. Combining symbolic and distributional models of meaning. In *AAAI Spring Symposium: Quantum Interaction*, 2007.

[19] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[20] C. J. Crouch. A cluster-based approach to thesaurus construction. In *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '88, page 309–320, New York, NY, USA, 1988. Association for Computing Machinery.

[21] James R Curran and Marc Moens. Improvements in automatic thesaurus extraction. In *Proceedings of the ACL-02 workshop on Unsupervised lexical acquisition*, pages 59–66, 2002.

[22] Ido Dagan, Shaul Marcus, and Shaul Markovitch. Contextual word similarity and estimation from sparse data. In *31st Annual Meeting of the Association for Computational Linguistics*, pages 164–171, Columbus, Ohio, USA, June 1993. Association for Computational Linguistics.

[23] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

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

[25] Ehtibar N. Dzhafarov. Replacing nothing with something special: Contextuality-by-default and dummy measurements. *Quantum Foundations, Probability and Information*, page 39–44, 2018.

[26] Ehtibar N. Dzhafarov and Janne V. Kujala. Probabilistic foundations of contextuality. *Fortschritte der Physik*, 65(6-8):1600040, 2017.

[27] Ehtibar N. Dzhafarov, Janne V. Kujala, and Victor H. Cervantes. Contextuality-by-default: A brief overview of ideas, concepts, and terminology, 2015.

[28] Philip Edmonds and Scott Cotton. Senseval-2: Overview. In *The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems*, SENSEVAL '01, page 1–5, USA, 2001. Association for Computational Linguistics.

[29] A. Einstein, B. Podolsky, and N. Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.*, 47:777–780, May 1935.

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

[31] Robert French and Christophe Labiouse. Four problems with extracting human semantics from large text corpora. 07 2002.

[32] Robert Goldblatt. *Topoi: the categorial analysis of logic.* Elsevier, 2014.

[33] J. Hamilton, C. J. Isham, and J. Butterfield. A Topos Perspective on the Kochen-Specker Theorem: III. Von Neumann Algebras as the Base Category, 1999.

[34] Lucien Hardy. Nonlocality for two particles without inequalities for almost all entangled states. *Phys. Rev. Lett.*, 71:1665–1668, Sep 1993.

[35] Zellig S. Harris. Distributional structure. *WORD*, 10(2-3):146–162, 1954.

[36] Irene Heim. *File Change Semantics and the Familiarity Theory of Definiteness*, pages 223 – 248. 01 2008.

[37] Chris Heunen, Nicolaas P. Landsman, Bas Spitters, and Sander Wolters. The gelfand spectrum of a noncommutative c*-algebra: A topos-theoretic approach. *Journal of the Australian Mathematical Society*, 90(1):39–52, Feb 2011.

[38] C. J. Isham and J. Butterfield. A topos perspective on the Kochen-Specker theorem: I. Quantum States as Generalized Valuations, 1998.

[39] C. J. Isham and J. Butterfield. A topos perspective on the Kochen-Specker theorem: II. Conceptual Aspects, and Classical Analogues, 1998.

[40] Martin Joos. Description of language design. *The Journal of the Acoustical Society of America*, 22(6):701–707, 1950.

[41] Daniel Jurafsky and James Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, volume 2. 02 2008.

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

[43] S. Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17(1):59–87, 1967.

[44] Claudia Leacock, Geoffrey Towell, and Ellen Voorhees. Corpus-based statistical sense resolution. In *Proceedings of the Workshop on Human Language Technology*, HLT '93, page 260–265, USA, 1993. Association for Computational Linguistics.

[45] Martha Lewis. Compositional hyponymy with positive operators. In *RANLP*, 2019.

[46] Saunders Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 2013.

[47] Saunders MacLane and Ieke Moerdijk. *Sheaves in geometry and logic: A first introduction to topos theory.* Springer Science & Business Media, 2012.

[48] George Miller, Claudia Leacock, Randee Tengi, and Ross Bunker. A semantic concordance. pages 303–308, 01 1993.

[49] Allison Mullaly, Christina Gagné, Thomas Spalding, and Kristan Marchak. Examining ambiguous adjectives in adjective-noun phrases: Evidence for representation as a shared core-meaning with sense specialization. *The Mental Lexicon*, 5:87–114, 06 2010.

[50] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1), Aug 2017.

[51] Sebastian Padó and Mirella Lapata. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199, 2007.

[52] Patrick Pantel and Dekang Lin. Discovering word senses from text. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, page 613–619, New York, NY, USA, 2002. Association for Computing Machinery.

[53] Martin Pickering and Steven Frisson. Processing ambiguous verbs: Evidence from eye movements. *Journal of experimental psychology. Learning, memory, and cognition*, 27:556–73, 03 2001.

[54] Robin Piedeleu, Dimitri Kartsaklis, Bob Coecke, and Mehrnoosh Sadrzadeh. Open system categorical quantum semantics in natural language processing, 2015.

[55] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, March 1994.

[56] Yonggang Qiu and Hans-Peter Frei. Concept based query expansion. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '93, page 160–169, New York, NY, USA, 1993. Association for Computing Machinery.

[57] K Rayner and SA Duffy. Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Memory & cognition*, 14(3):191—201, May 1986.

[58] Gerda Ruge. Experiments on linguistically-based term associations. *Information Processing & Management*, 28(3):317 – 332, 1992.

[59] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513 – 523, 1988.

[60] Hinrich Schütze. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123, 1998.

[61] Dagobert Soergel. Wordnet. an electronic lexical database. 10 1998.

[62] Michael K. Tanenhaus, James M. Leiman, and Mark S. Seidenberg. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, 18(4):427 – 440, 1979.

[63] Yorick Wilks, Dan Fass, Cheng-Ming Guo, James E. McDonald, Tony Plate, and Brian M. Slator. Providing machine tractable dictionary tools. *Machine Translation*, 5(2):99–154, 1990.

[64] Lars Wissler, Mohammed Almashraee, Dagmar Monett, and Adrian Paschke. The gold standard in corpus annotation. 06 2014.

[65] Sander Wolters. A Comparison of Two Topos-Theoretic Approaches to Quantum Theory, 2010.

[66] Deniz Yuret and Mehmet Yatbaz. The noisy channel mode for unsupervised word sense disambiguation. *Computational Linguistics*, 36:111–127, 03 2010.

Appendix A

The incidence matrix

The (augmented) incidence matrix for a bipartite Bell scenario (with two choices of binary measurements for each party) is given by:

$$\mathbf{M} = \begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix} \tag{A.1}$$

Hence, given an empirical model as depicted in Fig. 1.3a, the systems that needs to be solved is:

$$\begin{pmatrix}
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
1&0&1&0&1&0&1&0&0&0&0&0&0&0&0&0\\
0&1&0&1&0&1&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&1&0&1&0&1&0\\
0&0&0&0&0&0&0&0&0&1&0&1&0&1&0&1\\
1&1&0&0&0&0&0&0&1&1&0&0&0&0&0&0\\
0&0&0&0&1&1&0&0&0&0&0&0&1&1&0&0\\
0&0&1&1&0&0&0&0&0&0&1&1&0&0&0&0\\
0&0&0&0&0&0&1&1&0&0&0&0&0&0&1&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1
\end{pmatrix}
\begin{pmatrix} x_1\\x_2\\x_3\\x_4\\x_5\\x_6\\x_7\\x_8\\x_9\\x_{10}\\x_{11}\\x_{12}\\x_{13}\\x_{14}\\x_{15}\\x_{16}\end{pmatrix}
=
\begin{pmatrix} s_1\\s_2\\s_3\\s_4\\s_5\\s_6\\s_7\\s_8\\s_9\\s_{10}\\s_{11}\\s_{12}\\s_{13}\\s_{14}\\s_{15}\\s_{16}\\1\end{pmatrix} \tag{A.2}$$

with $x_i \in [0,1]$ for all i. For example, taking the Bell scenario depicted in Fig. 1.2, this gives:

$$\begin{pmatrix}
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
1&0&1&0&1&0&1&0&0&0&0&0&0&0&0&0\\
0&1&0&1&0&1&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&1&0&1&0&1&0\\
0&0&0&0&0&0&0&0&0&1&0&1&0&1&0&1\\
1&1&0&0&0&0&0&0&1&1&0&0&0&0&0&0\\
0&0&0&0&1&1&0&0&0&0&0&0&1&1&0&0\\
0&0&1&1&0&0&0&0&0&0&1&1&0&0&0&0\\
0&0&0&0&0&0&1&1&0&0&0&0&0&0&1&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1
\end{pmatrix}
\begin{pmatrix} x_1\\x_2\\x_3\\x_4\\x_5\\x_6\\x_7\\x_8\\x_9\\x_{10}\\x_{11}\\x_{12}\\x_{13}\\x_{14}\\x_{15}\\x_{16}\end{pmatrix}
=
\begin{pmatrix} 1/2\\0\\0\\1/2\\3/8\\1/8\\1/8\\3/8\\3/8\\1/8\\1/8\\3/8\\1/8\\3/8\\3/8\\1/8\\1\end{pmatrix} \tag{A.3}$$

In particular, the rows highlighted in red imply that:

$$x_1 + x_2 + x_3 + x_4 = 1/2 \tag{A.4}$$
$$x_2 + x_4 + x_6 + x_8 = 1/8 \tag{A.5}$$
$$x_3 + x_4 + x_{11} + x_{12} = 1/8 \tag{A.6}$$
$$x_1 + x_5 + x_9 + x_{13} = 1/8 \tag{A.7}$$

Now, (A.5) + (A.6) + (A.7) leads to:

$$x_1 + x_2 + x_3 + 2x_4 + x_5 + x_6 + x_8 + x_9 + +x_{11} + x_{12} + x_{13} = 3/8 \tag{A.8}$$

and in addition we also need to have (A.4) $\leq$ (A.5) + (A.6) + (A.7), which leads to a contradiction since $1/2 > 3/8$

Appendix B

Contextual examples: meaning combinations

B.0.1 Ambiguous verb in unambiguous context

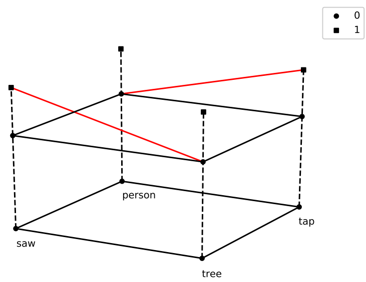
We will here consider some *verb-object* and *subject-verb* examples.

Object context

Let's consider an example of two ambiguous verbs (*saw* and *tap*) with unambiguous objects:

verb	*object*	(0,0)	(0,1)	(1,0)	(1,1)
saw	*tree*	1	0	1	0
saw	*person*	1	0	0	0
tap	*tree*	1	0	0	0
tap	*person*	1	0	1	0

(a) Empirical model



(b) Bundle diagram (local sections which cannot be extended are depicted in red).

The meanings are encoded as follows (ignoring all other meanings[1]):

Outome	*saw*	*tap*
0	*see*	*touch*
1	*cut*	*record*

The unambiguous meanings are once again sent to 0. The reasoning is:

- You can both saw and see a tree; you can both touch and record a human being;

[1]The verb *tap* has other meanings, e.g. *humans have always tapped natural resources.* or *They know how to tap (dance).* But those meanings are incompatible with the objects presented here, so the model is not *per sei* incomplete.

- You can't saw a human being (debatable I suppose, but even if this is possible, it doesn't change the contextuality of the model, as long as you can't record a tree; these cases would become less debatable when co-occurrence probabilities are collected from corpora); and you can't record a tree.

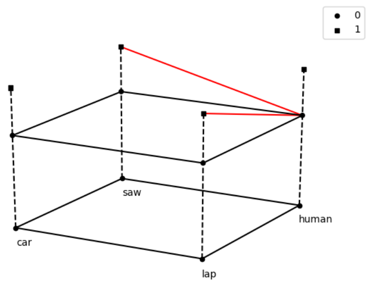
This model is weakly contextual.

B.0.2 Subject context

With two ambiguous verbs as well, we have for e.g.:

subject	*verb*	(0,0)	(0,1)	(1,0)	(1,1)
car	*lap*	1	0	0	0
car	*saw*	1	0	0	0
human	*lap*	1	1	0	0
human	*saw*	1	1	0	0

(a) Empirical model



(b) Bundle diagram (local sections which cannot be extended are depicted in red).

with the same encoding of all the words as in the main text (Section 3.1).

B.0.3 Ambiguous noun in umambiguous context

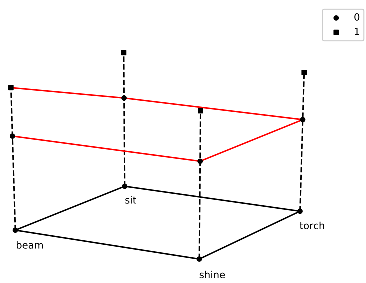
We now consider with ambiguous nouns in contexts. The results are similar to the ones in Section B.0.1, and we will interested in the following situations: $verb - noun$, $noun - verb$ and $adjective - noun$.

Acted on context

For the first example, we consider the noun *beam* which we associate the meanings: $[beam : light \mapsto 0, beam : stick \mapsto 1]$ (ignoring other meanings), then we can consider the following:

noun	*verb*	(0,0)	(0,1)	(1,0)	(1,1)
beam	*shine*	1	0	0	0
beam	*sit*	0	0	1	0
torch	*shine*	1	0	0	0
torch	*sit*	1	0	0	0

(a) Empirical model



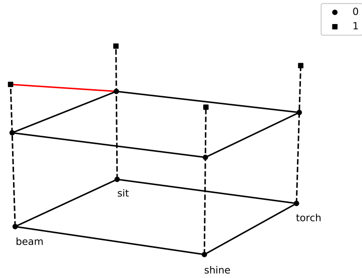
(b) Bundle diagram (local sections which cannot be extended are depicted in red).

Figure B.3: Stronger model

which is strongly contextual. If we moreover assume that a (light) beam can sit, then we have the weaker model:

noun	*verb*	(0,0)	(0,1)	(1,0)	(1,1)
beam	*shine*	1	0	0	0
beam	*sit*	1	0	1	0
torch	*shine*	1	0	0	0
torch	*sit*	1	0	0	0

(a) Empirical model



(b) Bundle diagram (local sections which cannot be extended are depicted in red).

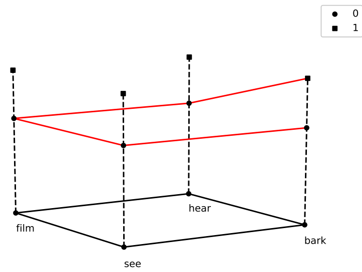
Figure B.4: Weaker model

Acting on context

Assuming that we cannot hear the *bark* of a tree, and we cannot see the *bark* of a dog, the following empirical model is strongly contextual:

verb	*noun*	(0,0)	(0,1)	(1,0)	(1,1)
see	*bark*	1	0	0	0
see	*film*	1	0	0	0
hear	*bark*	0	1	0	0
hear	*film*	1	0	0	0

(a) Empirical model



(b) Bundle diagram (local sections which cannot be extended are depicted in red).

where for *bark*: $[bark : tree \mapsto 0, bark : dog \mapsto 1]$.

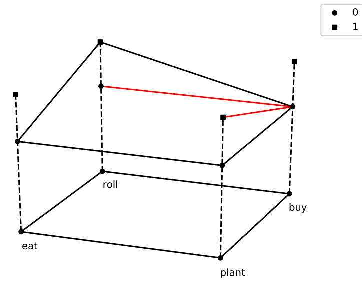
Similarly, consider the following meanings of *plant* and *roll*:

Outcome	*plant*	*roll*
0	*vegetable*	*bread*
1	*factory*	*tube*

the following empirical model is weakly contextual:

$verb$	$noun$	(0,0)	(0,1)	(1,0)	(1,1)
eat	$plant$	1	0	0	0
eat	$roll$	0	1	0	0
buy	$plant$	1	1	0	0
buy	$roll$	1	1	0	0

(a) Empirical model



(b) Bundle diagram (local sections which cannot be extended are depicted in red).

Modifying context

We first note that here we only consider the individual meanings of each of the words, not the meaning of the global phrase; for example, the meanings of the phrase *red pencil*, "pencil that writes in red" or "pencil which is red" are considered to be equivalent. We start by considering the noun *perch*:

- A fish perch can be alive; so does a human

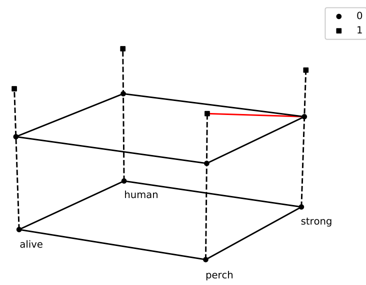
- A perch (rod) can be strong; once again so does a human, and I suppose so does a fish

- A perch (rod) cannot be alive

Then, the following empirical model is weakly contextual:

$modifier$	$noun$	(0,0)	(0,1)	(1,0)	(1,1)
$alive$	$perch$	1	0	0	0
$alive$	$human$	1	0	0	0
$strong$	$perch$	1	1	0	0
$strong$	$human$	1	0	0	0

(a) Empirical model



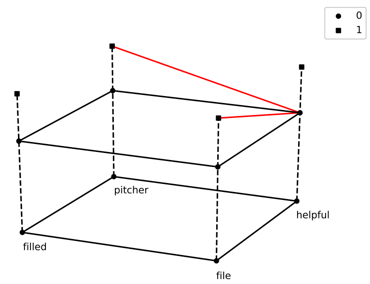
(b) Bundle diagram (local sections which cannot be extended are depicted in red).

where $[perch : fish \mapsto 0, perch : rod \mapsto 1]$.

Now consider the two ambiguous nouns *file*, here only document (0) or tool (1), and *pitcher*, jug (0) or baseball player (1). Then assuming that the a tool or a baseball player cannot be filled, while a document or a jug can:

$modifier$	$noun$	(0,0)	(0,1)	(1,0)	(1,1)
$filled$	$file$	1	0	0	0
$filled$	$pitcher$	1	0	0	0
$helpful$	$file$	1	1	0	0
$helpful$	$pitcher$	1	1	0	0

(a) Empirical model



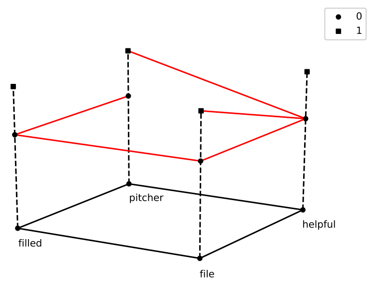
(b) Bundle diagram (local sections which cannot be extended are depicted in red).

Figure B.8: Weaker model

If moreover, we assume that a jug cannot be helpful (i.e. it is unlikely to describe a jug as helpful), then the model becomes strongly contextual:

$modifier$	$noun$	(0,0)	(0,1)	(1,0)	(1,1)
$filled$	$file$	1	0	0	0
$filled$	$pitcher$	1	0	0	0
$helpful$	$file$	1	1	0	0
$helpful$	$pitcher$	0	1	0	0

(a) Empirical model



(b) Bundle diagram (local sections which cannot be extended are depicted in red).

Figure B.9: Stronger model

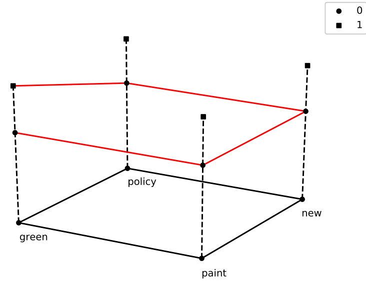
B.0.4 Ambiguous adjectives in unambiguous context

As for the previous section, only the meaning of the adjectives individually are considered here. In addition, the only situation considered is $adjective - noun$.

Let's consider the adjective *green* (colour: 0, environment-friendly: 1). It gives rise to the following strongly contextual example:

adjective	*noun*	(0,0)	(0,1)	(1,0)	(1,0)
green	*paint*	1	0	0	0
green	*policy*	0	0	1	0
new	*paint*	1	0	0	0
new	*policy*	1	0	0	0

(a) Empirical model

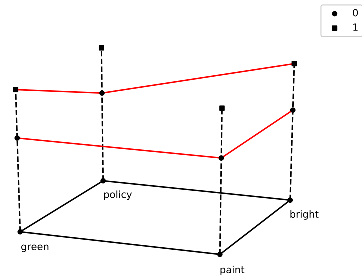


(b) Bundle diagram (local sections which cannot be extended are depicted in red).

Note that if the paint can also be made environmental-friendly the model remains contextual (but weakly contextual). We can also slightly modify this example by substituting *new* with the ambiguous adjective *bright* ($[bright : light\ colour \mapsto 0, bright : clever \mapsto 1]$), we get:

adjective	*noun*	(0,0)	(0,1)	(1,0)	(1,1)
green	*paint*	1	0	0	0
green	*policy*	0	0	1	0
bright	*paint*	1	0	0	0
bright	*policy*	0	0	1	0

(a) Empirical model



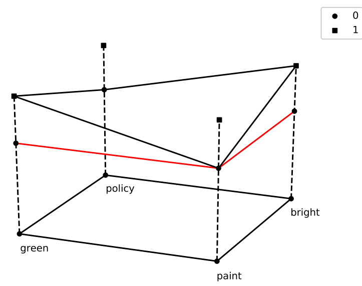
(b) Bundle diagram (local sections which cannot be extended are depicted in red).

Figure B.11: Stronger model

which is also strongly contextual. If we now loosen the assumptions that paint can be clever or environment-friendly, the model becomes weakly contextual:

adjective	*noun*	(0,0)	(0,1)	(1,0)	(1,1)
green	*paint*	1	0	1	0
green	*policy*	0	0	1	0
bright	*paint*	1	0	1	0
bright	*policy*	0	0	1	0

(a) Empirical model



(b) Bundle diagram (local sections which cannot be extended are depicted in red).

Figure B.12: Weaker model