Analysing Ambiguous Nouns and Verbs with Quantum Contextuality Tools

Daphne Wang[1], Mehrnoosh Sadrzadeh[1], Samson Abramsky[2], and Víctor H. Cervantes[3]

[1] *University College London, UK*
[2] *Oxford University, UK*
[3] *University of Illinois at Urbana-Champaign, US*

Abstract

Psycholinguistic research uses eye-tracking to show that polysemous words are disambiguated differently from homonymous words, and that ambiguous verbs are disambiguated differently than ambiguous nouns. Research in Compositional Distributional Semantics uses cosine distances to show that verbs are disambiguated more efficiently in the context of their subjects and objects than when on their own. These two frameworks both focus on one ambiguous word at a time and neither considers ambiguous phrases with two (or more) ambiguous words. We borrow methods and measures from Quantum Information Theory, the framework of Contextuality-by-Default and degrees of *contextual influences*, and work with ambiguous subject-verb and verb-object phrases of English, where both the subject/object and the verb are ambiguous. We show that differences in the processing of ambiguous verbs versus ambiguous nouns, as well as between different levels of ambiguity in homonymous versus polysemous nouns and verbs can be modelled using the averages of the degrees of their contextual influences.

Keywords: *Contextuality, Ambiguity, Senses and Meanings, Quantum Mechanics*

1. Introduction

Dealing with ambiguity is the stronghold of Psycholinguistics where experiments such as eye-tracking are performed to measure the delay in committing to the interpretation of ambiguous words. This delay has been found to be related to different types of interpretations a word has. Here, we have interpretations that are *senses* and interpretations that are *meanings* (Klepousniotou, 2002). The *sense* interpretations are used for polysemous words. For instance, a word such as *newspaper* both refers to the physical object – pages of news put together– or to the content – the news that is printed on these pages. An example of object interpretation is in the phrase "I put the newspaper on the table", an example of the content interpretation is in the phrase "I learnt much from the newspaper today". The *meaning* interpretations are used for homonymous words. An example of which is the word *spring* and its meanings *metal coil* or *season after winter*.

Research in Psycholinguistics has also shown that commitment is delayed more for words with multiple senses than for words with multiple meanings (Frazier and Rayner, 1990; Pickering and Frisson, 1999, 2001a) and that the higher the degree of overlap in senses, the higher the delay in semantic commitment (Ekaterini Klepousniotou and Romero, 2008). Research in Psycholinguistics has also shown that words with different grammatical types have different effects on the semantic commitment delay. In particular, it has been shown that noun disambiguation adheres to a shorter delay, as a reader uses their immediate context to decide which interpretation to commit to, whereas verb disambiguation takes longer, as the reader often has to go over the whole sentential context and sometimes even do so a few times (Pickering and Frisson, 2001b). More precisely, frequent nouns are often immediately disambiguated using the dominance of their interpretations, whereas most verbs need large contexts and it is common that a sentence with an ambiguous verb is read and reread before any meaning is committed to.

Distributional semantics deals with the process of disambiguation using statistical measures and machine learning algorithms on the co-occurrence data collected from usages of words in large corpora of text. This line of research is motivated by early ideas of Harris (1954) and Firth (1957) that *words that often occur in the same context have similar meanings* and that *one can know a word by the company that it keeps*. In the 90's, Distributional semanticists employed a range of methods for disambiguating word meaning and notably a state of the art hierarchal tree clustering algorithm (Schütze, 1998). Nowadays, disambiguation tasks are best performed by using contextualised autoencoder embeddings such as BERT (Devlin et al., 2019).

Similar to the research in Psycholinguistics, grammatical roles of words have also been taken into account, by moving from disambiguation at the word level to disambiguation at the phrase and sentence levels. Here, the original work of Mitchell and Lapata (2008) showed that when ambiguous verbs are placed in the context of their subjects or objects, a higher degree of correlation is observed between the distances of their distributional semantic representations and similarity degrees obtained from human annotations. Subsequent work in Compositional Distributional Semantics (Coecke et al., 2010), extended this line of research from subject-verb and verb-object phrases to transitive subject-verb-object sentences (Grefenstette and Sadrzadeh, 2011; Kartsaklis and Sadrzadeh, 2016) and relative clauses (Sadrzadeh et al., 2013, 2014). In a nutshell, here the Distributional semantics of a sentence is a vector obtained from the grammatical structure of the sentence and the semantic representations of the words therein. The cosine distances between sentence vectors are computed and used as a measure for disambiguation for the verbs of the sentences. A bit more formally, if sentence S with ambiguous verb V has a larger cosine with sentence S_1 with verb interpretation V_1, we say that V disambiguates to V_1. If, on the other hand, S has a larger cosine with S_2 with verb interpretation V_2, it is said that V disambiguates to V_2.

Our research is related to both of the above frameworks which sit within computational approaches to natural language semantics. From the Psycholinguistics side, we are interested in studying the disambiguation process of words that are ambiguous and have different grammatical roles. From the Compositional Distributional side, we would like to work with phrases and their grammatical structure rather than words in isolation. The difference between our work and the Psycholinguistics research is that we work with phrases wherein both words are ambiguous, rather than focusing on one ambiguous word, either noun or verb, at a time. The difference between our work and the research in disambiguation in Compositional Distributional semantics is that we do not consider ambiguous verbs in the context of their unambiguous subjects and objects. Our goal is to work with ambiguous verbs in the context of their ambiguous subjects and objects. Similar to the lines of work done in Psycholinguistics and Compositional Distributional semantics, we would like to come up with a measurable quantity that tells us which of the words in the phrase, noun or verb, polysemous or homonymous, contributes more to the disambiguation process of the phrase. Instead of the notion of delay in semantic commitment from Psycholinguistics or differences in cosine distances used by Compositional Distributional semanticists, we will work with measures of *contextual* influences coming from research in contextual scenarios of Quantum Information Theory. We will use ambiguous polysemous and homonymous nouns and verbs identified and studied in Psycholinguistics research to form our ambiguous phrases and the original notion of *context* from Distributional semantics, i.e. the window of words before or after a word. In order to take the grammatical roles directly into account, we narrow this window to one word before or after, which in the case of an ambiguous verb becomes its ambiguous subject/object.

This paper is a proof of concept that shows the mathematical theory of CbD-contextuality, a generalisation of settings originally developed for the study of contextuality in Quantum Information Theory, can be used to provide a statistical characterisation of the effect of subjects, objects, and verbs on meaning selection in ambiguous subject-verb and verb-object phrases. We are using notions and formulae from Quantum theory to analyse linguistic phenomena, and the linguistic phenomena we study come from Psycholinguistics, a subfield of Cognitive Science. The paper thus lies *at the Intersection of natural language, Physics, and Cognitive Science*, the range of topics covered by the SemSpace series of workshops.

Which Quantum Contextuality Framework? Quantum contextuality has different interpretations in different frameworks. The well-known Bell theorem

(Bell, 1964), supported by experimental data (Hensen et al., 2015), showed that if Quantum systems need to have a "reality" independent of the observers, one should allow interactions between systems to be unrestricted spatially (non-local). In the setting of Bell, a set of inequalities were introduced that offered a proof by contradiction that one cannot extend the probabilistic models obtained from observations of Quantum systems to a deterministic hidden-variable model. In Kochen and Specker (1967), the authors provide a description of *contextuality*, now standard in Quantum Mechanics literature. The sheaf-theoretic framework of contextuality (Abramsky and Brandenburger, 2011; Abramsky et al., 2015) starts from the observation that contextuality in Quantum mechanics translates to "the impossibility of finding a global section in special presheaves[1]", i.e. a model is contextual if some of its *local* features cannot be extended *globally*.

All of the above systems rely on a principle called *non-signalling*, which ensures that certain laws of Quantum mechanics hold in experiments such as EPR (Einstein et al., 1935) and Bell and ensures that Quantum systems are *local*, i.e. that there is no communication between the subsystems (e.g. qubits) of. system (e.g. an entangled pair of qubits). This property is often ensured by creating a geographical separation between the two subsystems and there is no reason to assume that it should hold for natural language[2]. In fact, natural language is not the only example of a system that may be contextual but which does not satisfy the non-signalling property, psychological and behavioural experiments show that these systems also do not satisfy it. For these reasons, the setting of Contextuality by Default (CbD) (Dzhafarov and Kujala, 2016; Kujala and Dzhafarov, 2016; Dzhafarov et al., 2015b) generalises the definition of Quantum contextually to systems that are not necessarily non-signalling. In this theory, one has a set of jointly distributed measurements for each qubit and the two qubits as a whole are contextual when it is impossible to create a global joint distribution. The term (non)-signalling is therein replaced by the CbD

[1]We will not make explicit use of presheaves in this work, but roughly speaking, presheaves are structure preserving maps.

[2]In previous work (Wang et al., 2021), we go through this analogy in more detail and provide concrete reasons why non-signalling cannot hold for analysing ambiguity in natural language.

term *(in)consistent connectedness*. One can also use the more intuitive term *(non)disturbance*, which seems to be replacing the term (non)-signalling in the Quantum Mechanics literature. For the rest this paper and to remain coherent with our previous papers, we will carry on using the term (non)-signalling.

Connections to Previous and Other Work In previous work (Wang et al., 2021), we showed that meaning combinations in ambiguous phrases can in rare cases be modelled in the sheaf-theoretic framework for Quantum contextuality (Abramsky and Brandenburger, 2011), where we found one example of a case which was non-signalling and showed that this example is also possibilistically contextual. Our calculations, nonetheless, showed that a large set of other examples first explored in Wang (2020), and even the probabilistic variant of this very same possibilistic example, were all signalling. This made us make the move to the framework of Contextuality-by-Default (CbD), where we showed that some of these probabilistic examples were CbD-contextual.

Our work also adds to the connections between natural language and Quantum theory. In particular, density matrices (i.e. Quantum states) have been used in representing ambiguity in a range of papers, such as in Blacoe et al. (2013); Meyer and Lewis (2020), where connections between the Quantum notion of superposition and ambiguities hidden in representations of Distributional semantics have been studied. Further, in Piedeleu et al. (2015), the authors encode different levels of lexical ambiguity as superposition or statistical mixing, whereas in Correia et al. (2019, 2020), the authors use the same methodology to propose a way of accommodating derivational ambiguity. Quantum methods have found multiple other applications in classical NLP tasks such as language modelling (Basile and Tamburini, 2017), the contextuality and non-locality of our mental lexicon (Bruza et al., 2009), and more recently in emotion detection (Li et al., 2020). Recent work of Meichanetzidis et al. (2020) runs programs on real Quantum computers to solve small datasets of natural language tasks and brings the field of Quantum NLP one step closer to reality. Applications of Quantum methods are not limited to natural language tasks, the role of Quantum formalisms has also been explored in advancing machine learning in general, for example see the linework explored in Li et al. (2018, 2019).

Contribution and Scope In this paper, we take advantage of our previous findings, systematically form a dataset of ambiguous phrases, and analyse the properties of the disambiguation processes of these phrases using tools from CbD. As mentioned above, our dataset of ambiguous phrases is obtained by pairing ambiguous nouns with ambiguous verbs in either subject or object positions. The verbs were chosen from the dataset of Pickering and Frisson (2001b); Shutova (2010) and the nouns from Rayner and Duffy (1986); Tanenhaus et al. (1979). For each pairing, we collected co-occurrence data from the BNC (2007) and the ukWaC (Baroni et al., 2009) corpora. BNC is an open-source text corpus comprising of 100 million words, spread across documents of different nature (including press articles, fiction, transcription of spoken language, and academic publications). UKWaC is a 2 billion word corpus constructed from the Web limiting the crawl to the .uk domain. Both BNC and UKWaC are part-of-speech tagged, hence, they provide grammatical relations and the lemma forms of words, we mined the interpretation of phrases manually.

After turning the raw co-occurrences mined from the above corpora into probability distributions, we computed a quantity coming from CbD referred to as *direct influence*, which roughly speaking determines how much each word within a phrase contributes to the meaning selection for the phrase as a whole. We then looked at the specific contributions from the verbs and from the nouns, looking at their effect on the overall direct influence of the phrase, using another CbD quantity Δ. Our computations showed that on average, verbs have a larger direct influence on meaning determination of the whole phrase than nouns. We also found out that the degree of direct influence is larger for verbs with multiple meanings (homonymous verbs) than verbs with multiple senses (polysemous verbs). This finding brought us back to the Psycholinguists finding of Pickering and Frisson (2001b), which via eye-tracking had shown that the disambiguation of verbs with multiple senses needs a larger context than verbs with multiple meanings.

A similar investigation was not conclusive for nouns, however, as the homonymous and polysemous nouns of our dataset had a very small deviation from the average direct influence of all the nouns combined. So we have not been able to verify whether a similar conclusion holds for nouns, i.e. that whether a noun with two senses contributes more to the disambiguation process or a verb with two senses, or whether the equation reverts when meanings are changed to senses.

Moving to a larger dataset and using larger corpora such as Google books and news is a way to firstly provide a large scale experimentation and secondly also verifying the hypothesis for nouns. Adding causality to the sheaf theoretic framework, in the lines of recent work of Gogioso and Pinzani (2021), and analysing these models there is another avenue for future work, as is finding applications of our findings in metaphor detection and identification using the datasets of Shutova (2010).

2. Definitions and Details from Quantum Contextuality

Measuring a system in Quantum theory corresponds to assigning a value to a property of the system, e.g. that the particle was found to have position x. A system is then said to be contextual iff one cannot impose a joint probability distribution that is defined across all global measurements and which agrees with all of the observed statistics. In particular, this definition only makes sense for *non-signalling* systems. These are systems in which the probability distribution of a local measurement (marginal of the global system's distribution given all parties' measurements) is always the same, regardless of the choice of measurement of other parties. The non-signalling condition ensures that the influence of the global measurement context is due to "true contextuality" and not say, communications between parties. The first formal studies of contextuality in Quantum Systems were presented by Bell (1964), and by Kochen and Specker (1967).

Although the non-signalling approaches to contextuality provide elegant methods for quantifying and reasoning about probabilistic systems, they are too restrictive as one cannot use them to judge the contextuality of systems that are signalling. This is in particular a major problem when dealing with experimental data, as experimental setups can only approximate probabilities and are subject to noise. In such systems, observing two events with the same frequency is highly unlikely and hence will often be signalling. The mathematical framework of *Contextuality-by-Default* (CbD) (Dzhafarov and Kujala, 2016; Kujala and Dzhafarov, 2016; Dzhafarov et al., 2015b) provides tools for dealing with more general systems by extending its definition of contextuality to also include signalling systems. In the CbD framework, a system is said to be contextual if one cannot impose a globally joint probability distribution across all contexts for which the local marginals agree with *maximal probability*.

The work in Basieva et al. (2019); Cervantes and Dzhafarov (2018); Dzhafarov et al. (2015); Dzhafarov et al. (2015a) applies the CbD framework to a plethora of existing investigations on Quantum contextuality in psychological and behavioural experiments, and in fact shows that many of them are not, despite previous claims, truly contextual. Indeed, the apparent “Quantum contextuality” of these settings were all results of the signalling nature of the system.

Although a systematic study of contextuality in linguistic data has not been done before, our general line of research does come closest to the concept combination examples of Bruza et al. (2015), where the contextuality of ambiguous concepts such as “apple chip” was considered. However, as shown in Dzhafarov et al. (2015a); Dzhafarov et al. (2015), neither of the 23 examples of Bruza et al. (2015) are truly contextual within the CbD framework. In this paper, we go over examples discussed in newly accepted (and soon to appear) work (Wang et al., 2021) that presents truly contextual combinations but also analyse them using the rank-2 cyclic models of the CbD framework. In addition, we also found a relationship between the degree of ambiguity of the words in phrases and the degree of *direct influence* of the context.

3. Contextuality-by-Default

We work with the standard version of the formalism of Contextuality-by-Default (CbD) and introduce it here, see Dzhafarov and Kujala (2016) for a more general introduction to the framework. In this setting, a *content* is a measurement, or more generally a question with a known set of answers. The *context* gathers which of these questions are asked, as well as extra information about them, e.g. their order and information about the experimental setting. Every content q_i in a context c^j then gives rise to a random variable R_i^j that takes values from the possible answers to the question q_i and returns the probabilities with which these answers are observed in the context c^j. All random variables in a given context are jointly distributed.

Our linguistic examples are analogous to the scenarios in behavioural sciences studied under the umbrella term "Question Order Effect" (Wang and Busemeyer, 2013; Dzhafarov et al., 2015; Kujala and Dzhafarov, 2016). Here a specific type of CbD system, called *cyclic systems* are modelled. A cyclic system has a rank n where each context has exactly n contents, and every content is exactly in n contexts. Since we work with pairings of verbs with only one to its left or right, in other words, as its subject or object, we work with rank 2 cyclic systems. Moreover, in such systems, all random variables are assumed to take values in $\{\pm 1\}$. Given a cyclic system, we define the quantity Δ:

$$\Delta = \sum_{i=1}^{n} \left| \left\langle R_i^{j_i} \right\rangle - \left\langle R_i^{j'_i} \right\rangle \right| \tag{1}$$

where $j_i \neq j'_i \; \forall i$ and $R_i^{j_i}, R_i^{j'_i}$ are well-defined; we write $\langle R \rangle$ for the average of the random variable R. This quantity gives an overall measure of how much do the marginals corresponding to the same contents differ from each other. Recall that the no-signalling condition is satisfied when these marginals are exactly the same in every choice of context, then the quantity Δ can be seen as a measure of the degree of signalling of the system (Basieva et al., 2019), as we have $\Delta = 0$ iff the system is non-signalling.

The probability distributions observed in each context will here be depicted in tables such as Fig. 1. The global probability distributions are on the main body of these tables, and the probability of a joint event is located in the cell at the intersection of two individual events; for example in Fig. 1 $P\left[R_1^1 = -1, R_2^1 = +1\right] = p_3$. The local probability distributions of each

	$R_1^1 = -1$	$R_1^1 = +1$	Marginal
$R_2^1 = -1$	p_1	p_2	$p_1 + p_2$
$R_2^1 = +1$	p_3	p_4	$p_3 + p_4$
Marginal	$p_1 + p_3$	$p_2 + p_4$	

Figure 1. Example of a joint distribution of random variables $R_{1,2}^1$ in a single context c_1.

	$R_1^1 = -1$	$R_1^1 = +1$	Marginal
$R_2^1 = -1$	$1/4$	0	$1/4$
$R_2^1 = +1$	0	$3/4$	$3/4$
Marginal	$1/4$	$3/4$	

(a) Context c_1

	$R_1^2 = -1$	$R_1^2 = +1$	Marginal
$R_2^2 = -1$	0	$2/5$	$2/5$
$R_2^2 = +1$	$3/5$	0	$3/5$
Marginal	$3/5$	$2/5$	

(b) Context c_2

Figure 2. Example of a cyclic system of rank 2.

individual variable, i.e. the marginals, are also depicted in these tables.

Example. Let us illustrate this with a concrete example. Suppose that we are insterested in contents q_1 and q_2 measured in two distinct contexts c_1 and c_2; with the observations depicted in Fig. 2. Similar to the models that we will consider in section 4, this example is a cyclic system of rank 2. Then the degree of signalling of the system is determined by computing the quantity Δ below:

$$\Delta = \left|\langle R_1^1 \rangle - \langle R_1^2 \rangle\right| + \left|\langle R_2^1 \rangle - \langle R_2^2 \rangle\right| \tag{2}$$

where, after the computation, we find out that $\Delta = 1$.

3.1 Contextual and Direct Influences

We want to understand and quantify the influence that the context has on ambiguous words. But first, we need to clarify what we mean by context influence.

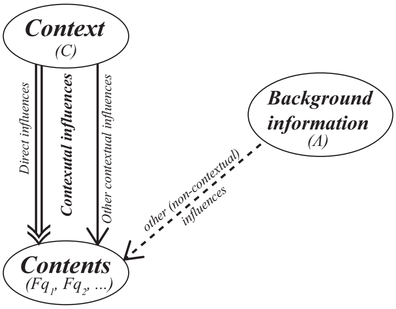
One of the ideas behind the CbD approach is to extend the notion of contextuality by allowing for the presence of *direct influences* of the context on the results of measurements. However, for systems in which changing the context results in a change of probability distribution, there is clearly some *contextual influence*. Therefore, one question is to distinguish what counts as "direct influence", and what are "truly contextual influences". In CbD, we want to minimise the disparity between the content distributions; this disparity will then be attributed to direct influences. We will refer to the minimal amount of contextual influences allowed by the observed distributions as *direct influence*, while *contextual influence* will designate any other kind of context effect, see Fig. 3a.

The above intuitions led to the development of the M-contextuality framework of Jones (2019), which also takes on ideas from the causal analysis of contextuality of Cavalcanti (2018) and the (classical) theory of causality of Pearl (2000). In Jones (2019), it was shown that every system of random variables observed in the different contexts can be expressed as a Bayesian network of the form depicted in Fig. 3b. Here, the context is treated as a single random variable C and each content q is modelled by a random variable F_q. The latter is determined by the context C and a latent variable denoted by Λ. This variable Λ models the background knowledge of the system; in our setting, this would for example include our knowledge of the world and the overall frequencies of words and their interpretations. Such a Bayesian network is referred to as a *canonical model* (Jones, 2019).

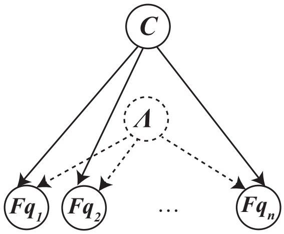
Now, given a canonical model which successfully describes the observed probabilities of a cyclic system[3], we quantify the *direct influence* of the context variable C on a given content q as:

---

[3]This definition in fact works for every pair of contexts $c_q, c'_q \in C$; but we will not make use of it in this work.



(a) Details of the different types of influences the content variables are under.



(b) Bayesian Network representation of a canonical model.

Figure 3. Contextual and Direct influences. The labels C, Fq_1, Fq_2 and Λ in both figures correspond to the random variables associated with the contexts, contents and hidden variables respectively in M-Contextuality.

$$\Delta_{c,c'}\left(F_q\right) = P\left[\Lambda \in \left\{\lambda | F_q(\lambda, c) \neq F_q(\lambda, c')\right\}\right] \tag{3}$$

Here, the λ's correspond to values that the latent random variable Λ can take.

We now have two ways of quantifying the "direct influence" of the context on a system, namely by using the "degree of signalling" Δ from CbD or by using the direct influences of the just introduced canonical models $\Delta_{c,c'}(F_q)$ for each content q. As it turns out, these quantities are intrinsically related, and the following is true:

Proposition 1. *For a cyclic system with binary random variables taking values in* $\{\pm 1\}$*, we have:*

$$\Delta = 2 \sum_q \Delta^*_{c_q, c'_q}\left(F_q\right) \tag{4}$$

where $\Delta^*_{c_q, c'_q}\left(F_q\right)$ *is the minimum direct influence of the contexts* c_q, c'_q

associated with content q across all the canonical models compatible with the observed distribution. Moreover, $\Delta^*_{c,c'}\left(F_q\right)$ *is equivalent to:*

$$\Delta^*_{c,c'}\left(F_q\right) = 1 - \sum_{v\in\{\pm 1\}} \min\left\{P[R_q^c = v], P[R_q^{c'} = v]\right\} \tag{5}$$

In order to prove Proposition 1 we need some results from CbD and M-contextuality. In the CbD framework, given a cyclic system, or more generally a system for which every content is part of exactly 2 contexts, we want to minimise the probability $P\left[S_q^c = S_q^{c'}\right] = \sum_{o\in O} P\left[S_q^c = S_q^{c'} = o\right]$ (where O is the set of possible outcomes) for a globally imposed joint distribution S across all contexts (coupling), which agrees with the observed distributions.

Lemma 1. *Given a content q and contexts* c, c' *containing q and outcome o, the maximum of* $P\left[S_q^c = S_q^{c'} = o\right]$ *for any coupling of the system is given by* $\min\left(P\left[R_q^c = o\right], P\left[R_q^{c'} = o\right]\right)$.

Proof. We need a coupling to be compatible with the observed probability distributions, i.e. that the marginals of S coincide with the original distributions. This condition means that:

$$\sum_{o'\in O} P\left[S_q^c = o, S_q^{c'} = o'\right] = P\left[R_q^c = o\right] \tag{6}$$

for each context c, c' sharing the content q, and for every value $o \in O$. In particular, this implies both of the following inequalities:

$$P\left[S_q^c = o, S_q^{c'} = o\right] \leq P\left[R_q^c = o\right] \tag{7}$$

$$P\left[S_q^c = o, S_q^{c'} = o\right] \leq P\left[R_q^{c'} = o\right] \tag{8}$$

and so:

$$P\left[S_q^c = o, S_q^{c'} = o\right] \leq \min\left(P\left[R_q^c = o\right], P\left[R_q^{c'} = o\right]\right) \tag{9}$$

In addition, given any system with content q, it is always possible to construct a coupling for which $P\left[S_q^c = S_q^{c'}\right]$ does attain its maximum (Theorem 3.3 of Dzhafarov and Kujala (2016)). The above bound is therefore saturated. □

One consequence of this is that:

$$\min P\left[S_q^c \neq S_q^{c'}\right] = 1 - \max P\left[S_q^c = S_q^{c'}\right] \tag{10}$$

We now use one of the main results about the correspondence between CbD and M-contextuality.

Proposition 2 (Proposition 8.4 of Jones (2019))**.** *Given a measurement system (i.e. context-content system with associated probability distributions), for each compatible canonical model $\mathcal{M}$, there exists a coupling S s.t.:*

$$\Delta_{c,c'}\left(F_q\right) = P\left[S_q^c \neq S_q^{c'}\right] \tag{11}$$

for every content q. Conversely, for every coupling S, there exists a canonical model $\mathcal{M}$ s.t. (11) *is satisfied.*

Corollary 1. *The minimum of direct influence given a content q and pair of contexts c, c', coincides with the minimum for $P\left[S_q^c \neq S_q^{c'}\right]$.*

We can now prove Proposition 1.

Proof of Proposition 1. By definition, we have:

$$\Delta = \sum_q \left|\left\langle R_q^{c_q}\right\rangle - \left\langle R_q^{c'_q}\right\rangle\right| \tag{12}$$

Since only binary variables are considered for this definition to make sense, each individual term of the sum is given by:

$$\begin{aligned}\left|\left\langle R_q^{c_q}\right\rangle - \left\langle R_q^{c'_q}\right\rangle\right| =& \left|P\left[R_q^{c_q} = +1\right] - P\left[R_q^{c_q} = -1\right] - P\left[R_q^{c'_q} = +1\right] + P\left[R_q^{c'_q} = -1\right]\right| \\ =& 2\left|P\left[R_q^{c_q} = +1\right] - P\left[R_q^{c'_q} = +1\right]\right| \end{aligned} \tag{13}$$

Now, let

$$m_{q-} = \min\left(P\left[R_q^{c_q} = -1\right], P\left[R_q^{c'_q} = -1\right]\right)$$

and respectively

$$m_{q+} = \min\left(P\left[R_q^{c_q} = +1\right], P\left[R_q^{c'_q} = +1\right]\right)$$

Then, each of the above terms reduces to:

$$\left|\left\langle R_q^{c_q}\right\rangle - \left\langle R_q^{c'_q}\right\rangle\right| = 2\left(1 - \left(m_{q+} + m_{q-}\right)\right) \tag{14}$$

Hence, following our previous corollary, the result follows. □

Example. The minimal degrees of direct influence for our previous example (Fig. 2), is computed as follows:

$$\Delta^*_{c_1,c_2}(F_1) = 1 - \left(\min\left\{\frac{1}{4}, \frac{3}{5}\right\} + \min\left\{\frac{3}{4}, \frac{2}{5}\right\}\right) = \frac{7}{20} \tag{15}$$

$$\Delta^*_{c_1,c_2}(F_2) = 1 - \left(\min\left\{\frac{1}{4}, \frac{2}{5}\right\} + \min\left\{\frac{3}{4}, \frac{3}{4}\right\}\right) = \frac{3}{20} \tag{16}$$

It is also easy to check that (4) indeed holds in this case as we have:

$$\Delta^*_{c_1,c_2}(F_1) + \Delta^*_{c_1,c_2}(F_2) = \frac{1}{2} = \frac{\Delta}{2} \tag{17}$$

4. Contextual Influences and Levels of Ambiguity

In natural language, words may have multiple unrelated interpretations (homonymous words), e.g. *spring* can mean *metal coil* or *the season*, or multiple related interpretations (polysemous words), e.g. *book*, which can either mean the object or the content of a book. The former is usually just referred to as *meanings*, whereas the latter is called *senses*. To avoid confusion, we follow Pickering and Frisson (2001b) and use the term *interpretation* to refer to either of these cases.

Using the CbD terminology, we take a content q_i to stand for the interpretation of a word in a phrase. A context c_j will include the phrase under consideration, i.e. the words that constitute the phrase, the grammatical structure of the phrase, i.e. how the words are combined with each other, as well as any extra information available, such as the corpus in which the frequencies were observed. We will here consider cyclic models of rank $n = 2$ where two ambiguous words can combine with each other in two different ways to form phrases. In particular, we study the differences in the behaviour of ambiguous verbs and nouns between *verb-object* and *subject-verb* cases. To do so, we consider (*verb*, *noun*) pairs such that both the verb and the noun have two different interpretations, and for which the two combinations of interest have occurred at least once in the corpus; for example, the pair (*ruin*, *bank*) should be valid as both *to ruin a bank* and *the bank ruins* make sense. Ambiguous words can have more than two different interpretations; however, to comply with the binary-variables-only setting of cyclic systems, we will work with only two of its interpretations, which will be labelled by the values $\{\pm 1\}$; even though these labels are attibuted arbitrarily to interpretations, they do not affect neither the degrees of direct influences nor the contextuality of the models.

As corpus, we have considered both the BNC (2007) and the ukWaC (Baroni et al., 2009) corpora. Our dataset was obtained by pairing each ambiguous verb of Pickering and Frisson (2001b); Shutova (2010) with a subject and an object. These were ambiguous nouns chosen from Rayner and Duffy (1986); Tanenhaus et al. (1979). This procedure led to a pair of ambiguous subject-verb and verb-object phrases, for each verb. For each such pair, we looked at the overall level of direct influence Δ associated with it, as well as the contributions from the verb and noun to Δ, which are respectively denoted as follows:

$$\Delta_v \equiv \Delta^*_{c_1,c_2}(F_{verb}) \qquad \Delta_n \equiv \Delta^*_{c_1,c_2}(F_{noun}) \tag{18}$$

The overall collected data is illustrated in Fig. 4 and has 90 entries in it. Note that not all of the subject-verb and verb-object combinations led to a valid model. For instance, some of the combinations had not occurred in either of our corpora, and hence were not included in the dataset. We also decided not to include combinations for which we did not have enough data, i.e. when the statistical uncertainty on Δ was bigger than the range of possible values. Similarly, one may also note that the degrees of direct influence Δ, Δ_v or Δ_n alone do not judge the contextuality of the model[4], but are instead measures of how much the distributions of selected interpretations vary in different contexts.

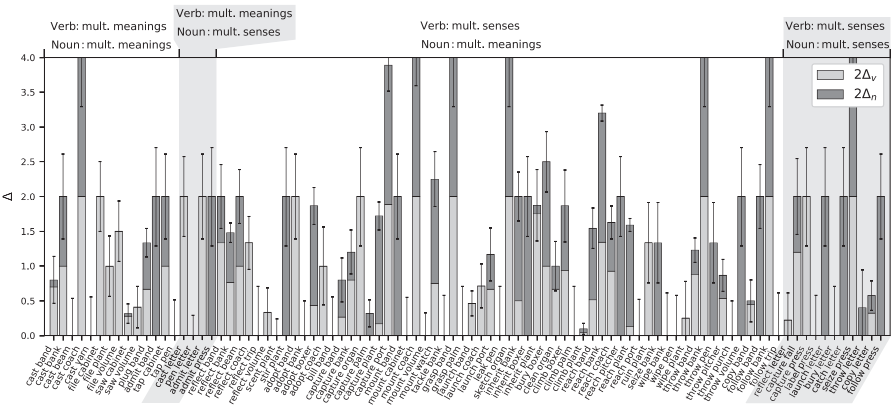


Figure 4. Direct influences for all rank-2 models considered by classes of models. The overall height of each bar represents Δ of each combination pair, and the error-bars are the corresponding uncertainty; for cases where $\Delta = 0$ (e.g. *cast beam*), only the error bar was depicted. The relative contributions from the verb ($2\Delta_v$) and the noun ($2\Delta_n$) are also depicted as respectively the bottom and top portions of each bar.

Our analysis showed that the average amount of overall direct influence Δ was similar in all classes of models, namely: verbs with multiple meanings combined with nouns with multiple meanings, verbs with multiple meanings combined with nouns with multiple senses, verbs with multiple senses combined with nouns with multiple meanings and finally verbs with multiple senses combined with verbs with multiple senses. These values are shown in Fig. 5a, and are all found to be about 1.35 (within half of the standard

[4]However, it can be shown that models with $\Delta > 2$ *cannot* be contextual within the CbD framework.

error of the mean). Therefore, the value of the quantity Δ by itself cannot be used to extract features of phrases with different levels of ambiguity; for example, it cannot be used to distinguish between senses-senses combinations and meanings-meanings combinations.

The only notable difference that we found was regarding the standard deviation of these direct influences. Indeed, our dataset showed a larger spread of the observed direct influences whenever a verb with multiple senses and/or a noun with multiple meanings were combined, see Fig. 5b. This shows that the behaviour of such combinations would be somehow more variable than for other combinations. Given the size of our dataset, however, we should be cautious with these results. A larger scale experiment is required in order to verify this claim for all possible ambiguous subject-verb, verb-object phrases.

Verbs / **Nouns**	**Meanings**	**Senses**	**Overall**
Meanings	1.24 ± 0.29	1.36 ± 0.15	1.33 ± 0.14
Senses	1.50 ± 0.43	1.38 ± 0.36	1.41 ± 0.29
Overall	1.30 ± 0.25	1.36 ± 0.14	1.35 ± 0.12

(a) Means of Δ for classes of models considered and their standard errors.

Classes of models	**SD**
Verbs with multi. meanings	1.04
Verbs with multi senses	1.20
Nouns with multi meanings	1.18
Nouns with multi senses	1.11

(b) Standard deviation (SD) of the direct influences collected for different classes of models.

Figure 5. Statistics of the total direct influences for different classes of models.

Let us now focus on the ratios of direct influences due to the verbs and the nouns. Without loss of generality, we will only consider the ratio[5] $2\Delta_v/\Delta$; we

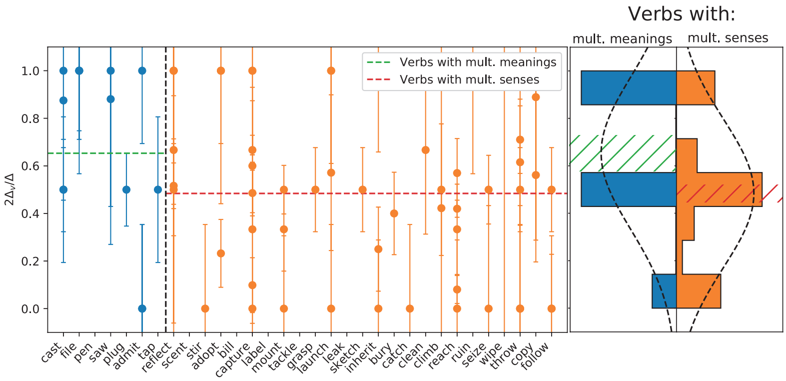
[5]The factor of 2 ensures the contributions from the verb and the noun add to 1.

can see from (4) that the quantity $2\Delta_n/\Delta$ can be obtained as $1 - 2\Delta_v/\Delta$. In addition, we ignored the data points for which $\Delta = 0$, as the ratio $2\Delta_v/\Delta$ is not defined for these cases[6].

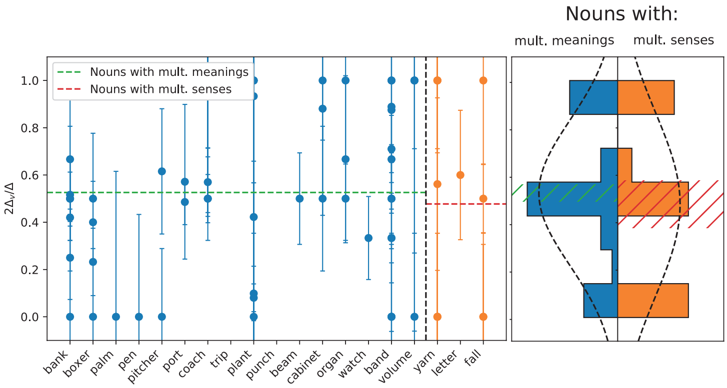
Relative contributions of the noun and the verb to Δ did not vary much in cases when the noun had multiple meanings or multiple senses; in both cases, the average contributions were about 50% within statistical error, see Fig. 6b. On the other hand, we observed a stark difference in the ratios when we looked at the level of ambiguity of the verb, see Fig. 6a. The contributions attributed to the verb are about 70% when the verb had multiple meanings, whereas they averaged at about 50% when the verb had multiple senses. We found that the average ratio Δ_v/Δ is strictly higher for verbs with multiple meanings than for verbs with multiple senses with more than 95% confidence.

The difference between processing homonymous vs polysemous nouns is mostly about their frequency effects (Frazier and Rayner, 1990). That is, whenever a homonymous noun is encountered, its meanings are activated with different thresholds depending on how common each meaning is. For instance, when the word *spring* is encountered, both of its meanings are activated, but the *season* meaning has a higher threshold than the *coil* meaning. The intended meaning is thereafter selected using a larger linguistic context (e.g. the rest of the sentence)(Dopkins et al., 1992; Duffy et al., 1988; Binder and Morris, 1995). For polysemous nouns, however, research shows that subjects originally select an *underspecified* sense with no special threshold (Frazier and Rayner, 1990; Pickering and Frisson, 1999; Frisson and Pickering, 2009), then, as was the case for homonymous nouns, the appropriate sense is selected from a larger context. In the rank-2 examples that we considered, the same noun is used in two different ambiguous phrases, but the noun has a different grammatical role in each phrase. One can then only observe context effects since we are working with the same noun and the only changes in frequency are obtained via the different grammatical roles that it is taking, i.e. subject

[6]It can be shown that these points can also be included in the analysis by fixing the ratios $2\Delta_v/\Delta = 2\Delta_n/\Delta = 50\%$ (i.e. assuming that both contents are equally "responsible" for the identical distributions), and we verified that doing so does not change the qualitative aspects of our results.



(a) Impact of the ambiguity of verbs.



(b) Impact of the ambiguity of nouns.

Figure 6. Relative contributions of the verb content to the overall direct influence given different levels of ambiguity for the verb or the noun. The left-hand figures correspond to the contributions of the verb-content; averages for each of the levels of ambiguity are shown with dotted lines. The right-hand figures correspond to the distributions of these data points, again for different levels of ambiguity; the 66%-confidence intervals for the means are depicted by the hatched area; the fitted normal distributions are also plotted.

vs object. These differences are observed in each example individually (i.e.

whenever $\Delta_n \neq 0$), and average out in the same manner no matter if the noun is homonymous or polysemous.

For verbs, on the other hand, frequency effects are not as clearly present as was the case for nouns, and instead, a delay in commitment to an interpretation is observed (Pickering and Frisson, 2001b). In particular, for verbs with multiple meanings, the delay mostly lasts until the arguments of the verb are seen. In our examples, we are able to observe this effect despite the short length of the phrases, since each of our phrases has at least one of the arguments of its verb, and it is this argument that changes between the two contexts. As shown above, the verb has, on average, a larger direct influence when compared to its noun argument. This degree of direct influence is considerable for verbs with multiple meanings but less so for verbs with multiple senses. Indeed as Psycholinguists show, the disambiguation of verbs with multiple senses is delayed to the end of the sentence and a larger context is needed for them (Pickering and Frisson, 2001b).

5. Conclusions and Discussion

The semantically ambiguous nature of a word gives rise to a natural interpretation of context-dependent probability distribution for each word. This distribution represents the probabilities of choosing certain interpretations of the word when making sense of a phrase. Studying the related probabilities enables us to quantify the variations of choices of interpretations as the context changes. The notion of contextuality developed in Quantum theory and the mathematical theories that come with it offer a setting that formalises and reasons about these variations, e.g. models that explain paradoxes such as Bell and CHSH (Bell, 1964; Kochen and Specker, 1967) and models that explain data coming from behavioural and psychological systems (Basieva et al., 2019; Cervantes and Dzhafarov, 2018; Dzhafarov et al., 2015; Dzhafarov et al., 2015a). In this paper, we applied these latter models to analyse natural language data.

Our work in Wang et al. (2021) showed that “true contextuality" does also arise in natural language, namely when words with more than one interpretation combine with each other in the same phrase. As the main focus of this paper, we studied the differences between the *verb-object* and *subject-verb* contexts, where both verbs and objects/subjects are ambiguous. We measured the degree of contextual influences of these phrases, using formulae from the Contextuality-by-Default and M-Contextuality frameworks. We found out that the amount of direct influence of a word within a phrase provides us with insights into the behaviour of ambiguous verbs and nouns in context. Indeed, while the variations of the distributions for nouns only seemed proportional to the overall change of distribution for the whole phrase, the same variations for verbs did depend on whether the considered interpretations of the verb had different meanings or different senses. The choice of interpretations did vary considerably more for the verbs in cases where the verb had multiple meanings. Inline with the findings of Pickering and Frisson (2001b), this can be used to argue that verbs play a more complex role in disambiguation than nouns, and that the degree of ambiguity of these verbs influences the process of interpretation selection.

Following more recent developments in the CbD framwork (Kujala and Dzhafarov (2019); Dzhafarov et al. (2020)), in previous work (Wang et al. (2021)), we computed the degree of contextuality of our contextual examples; these were $\frac{1}{3}$ for the *adopt boxer* example and $\frac{7}{30}$ for the *throw pitcher* example. One can also compute the degree of noncontextuality of the entries of our dataset and relate the findings to the linguistic features under study. Another measure worth investigating in this regard is the CbD counterpart of the *contextual fraction* of the sheaf theoretic models of contextuality, introduced in Dzhafarov (2020).

Our findings were limited by the size of the dataset and the sample size of the occurrences of each entry in corpora. We would like to undertake a larger scale experiment either by using larger corpora or via human judgements. Different types of contexts can also be considered; for example, higher rank models could be used to study the effect of the context given a syntactic structure (e.g. *verb-object*, *adjective-noun*, ...). No contextual instances of rank-4 models have been found yet, but preliminary evidence seems to suggest that the qualitative results of Section 4 can be extended to such models.

Furthermore, in line with the work done in Sadrzadeh et al. (2018); Barker and Shan (2015); Pustejovsky (1995), one can study the effect of context evolution on meaning selection. Systematic modelling of the interactions between statistics of contexts and linguistic structure is a future direction made possible by moving to sheaf theoretic semantics, e.g. see Abramsky and Brandenburger (2011); Abramsky and Sadrzadeh (2014); this constitutes work in progress. In this realm, the use of continuations (Barker and Shan, 2015) has led to the development of modular relationships between context and structure, where following research in programming language semantics, natural language computations are modelled by monads that interact with context via a set of inductive definitions. The main goal of continuation semantics of natural language has, however, always been to explain the ambiguities arising from quantification in natural language, in sentences such as "every man likes a car". Ambiguities coming from senses and meanings of polysemous and homonymous words and phrases have not been taken into account and formalised. Further, common to most other formal semantics frameworks, continuation semantics does not work with statistics of contexts coming from large scale data. We hope our line of research is one that unites linguistic structure with ambiguities and their contextual influences.

Finally, applications of this work to mainstream natural language tasks, such as automatic metaphor detection is also a future goal. Finally, the study of this paper was in English, but the methodology is applicable to other languages. Using the same methodology we can model phrases and sentences that have different word orders, for instance in verb final languages such as German or Farsi. We can also analyse languages with free word order such as Latin or Hungarian.

Acknowledgments

This work was supported by the EPSRC Centre for Doctoral Training in Delivering Quantum Technologies, grant ref. EP/S021582/1.

References

2007. The British National Corpus. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. `http://www.natcorp.ox.ac.uk/`. Version 3 (BNC XML Edition).

Abramsky, Samson, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield. 2015. Contextuality, Cohomology and Paradox. In Stephan Kreutzer, (ed.), *24th EACSL Annual Conference on Computer Science Logic (CSL 2015)*, volume 41 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 211–228. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CSL.2015.211.

Abramsky, Samson and Adam Brandenburger. 2011. The sheaf-theoretic structure of non-locality and contextuality. *New J. Phys.* 13: 113036.

Abramsky, Samson and Mehrnoosh Sadrzadeh. 2014. Semantic Unification. In *Categories and Types in Logic, Language, and Physics*, volume 8222 of *Lecture Notes in Computer Science*. Springer.

Barker, Chris and Chung-chieh Shan. 2015. *Continuations and Natural Language*. Oxford Scholarship.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43 (3): 209–226. doi: 10.1007/s10579-009-9081-4.

Basieva, Irina, Víctor H. Cervantes, Ehtibar N. Dzhafarov, and Andrei Khrennikov. 2019. True contextuality beats direct influences in human decision making. *Journal of Experimental Psychology: General* 148 (11): 1925–1937. doi:10.1037/xge0000585.

Basile, Ivano and Fabio Tamburini. 2017. Towards quantum language models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1840–1849. Copenhagen, Denmark: Association for Computational Linguistics. doi:10.18653/v1/D17-1196.

Bell, J. S. 1964. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika* 1: 195–200. doi:10.1103/PhysicsPhysiqueFizika.1.195.

Binder, K. and R. Morris. 1995. Eye movements and lexical ambiguity resolution: effects of prior encounter and discourse topic. *Journal of experimental psychology. Learning, memory, and cognition* 21 5: 1186–96.

Blacoe, William, Elham Kashefi, and Mirella Lapata. 2013. 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: Association for Computational Linguistics.

Bruza, Peter D., Kirsty Kitto, D. Nelson, and C. McEvoy. 2009. Is there something quantum-like about the human mental lexicon? *Journal of Mathematical Psychology* 53: 362–377.

Bruza, Peter D., Kirsty Kitto, Brentyn J. Ramm, and Laurianne Sitbon. 2015. A probabilistic framework for analysing the compositionality of conceptual combinations. *Journal of Mathematical Psychology* 67: 26 – 38. doi: https://doi.org/10.1016/j.jmp.2015.06.002.

Cavalcanti, Eric G. 2018. Classical Causal Models for Bell and Kochen-Specker Inequality Violations Require Fine-Tuning. *Phys. Rev. X* 8: 021018. doi:10.1103/PhysRevX.8.021018.

Cervantes, Victor H. and Ehtibar N. Dzhafarov. 2018. Snow Queen is Evil and Beautiful: Experimental Evidence for Probabilistic Contextuality in Human Choices.

Coecke, Bob, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *CoRR* abs/1003.4394.

Correia, Adriana D., Michael Moortgat, and Henk T. C. Stoof. 2019. Density matrices for derivational ambiguity. *CoRR* abs/1908.07347.

Correia, Adriana D., Henk T. C. Stoof, and Michael Moortgat. 2020. Putting a spin on language: A quantum interpretation of unary connectives for linguistic applications. *CoRR* abs/2004.04128.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics. doi:10.18653/v1/N19-1423.

Dopkins, Stephen, Robin K Morris, and Keith Rayner. 1992. Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. *Journal of Memory and Language* 31 (4): 461–476. doi:https://doi.org/10.1016/0749-596X(92)90023-Q.

Duffy, Susan A, Robin K Morris, and Keith Rayner. 1988. Lexical ambiguity and fixation times in reading. *Journal of Memory and Language* 27 (4): 429–446. doi:https://doi.org/10.1016/0749-596X(88)90066-6.

Dzhafarov, E. N., Ru Zhang, and Janne Kujala. 2015. Is there contextuality in behavioural and social systems? *Philosophical Transactions of the Royal Society of London Series A* 374 (2058): 20150099. doi:10.1098/rsta.2015.0099.

Dzhafarov, Ehtibar, Ru Zhang, Víctor Cervantes, and Janne Kujala. 2015a. On Contextuality in Behavioral Data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374. doi:10.1098/rsta.2015.0234.

Dzhafarov, Ehtibar N. 2020. The contextuality-by-default view of the sheaf-theoretic approach to contextuality. In *Samson Abramsky on Logic and Structure in Computer Science and Beyond*, Outstanding Contributions to Logic. Springer. To appear.

Dzhafarov, Ehtibar N. and Janne V. Kujala. 2016. Context–content systems of random variables: The Contextuality-by-Default theory. *Journal of Mathematical Psychology* 74: 11 – 33. doi:https://doi.org/10.1016/j.jmp.2016.04.010. Foundations of Probability Theory in Psychology and Beyond.

Dzhafarov, Ehtibar N., Janne V. Kujala, and Víctor H. Cervantes. 2020. Contextuality and noncontextuality measures and generalized bell inequalities for cyclic systems. *Physical Review A* 101 (4). doi:10.1103/physreva.101.042119.

Dzhafarov, Ehtibar N., Janne V. Kujala, and Jan-Åke Larsson. 2015b. Contextuality in Three Types of Quantum-Mechanical Systems. *Foundations of Physics* 45 (7): 762–782. doi:10.1007/s10701-015-9882-9.

Einstein, A., B. Podolsky, and N. Rosen. 1935. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.* 47: 777–780. doi:10.1103/PhysRev.47.777.

Ekaterini Klepousniotou, Debra Titone and Carolina Romero. 2008. Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34: 1534–1543.

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

Frazier, Lyn and Keith Rayner. 1990. Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language* 29 (2): 181–200. doi:https://doi.org/10.1016/0749-596X(90)90071-72.

Frisson, Steven and Martin Pickering. 2009. Semantic Underspecification in Language Processing. *Language and Linguistics Compass* 3: 111–127. doi:10.1111/j.1749-818X.2008.00104.x.

Gogioso, Stefano and Nicola Pinzani. 2021. The sheaf-theoretic structure of definite causality.

Grefenstette, Edward and Mehrnoosh Sadrzadeh. 2011. Experimenting with transitive verbs in a discocat. *CoRR* abs/1107.3119.

Harris, Zellig S. 1954. Distributional structure. *WORD* 10 (2-3): 146–162. doi:10.1080/00437956.1954.11659520.

Hensen, B., H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W.

Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. 2015. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* 526 (7575): 682–686. doi:10.1038/nature15759.

Jones, Matt. 2019. Relating causal and probabilistic approaches to contextuality. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 377 (2157): 20190133. doi:10.1098/rsta.2019.0133.

Kartsaklis, Dimitri and Mehrnoosh Sadrzadeh. 2016. A compositional distributional inclusion hypothesis. pages 116–133. doi:10.1007/978-3-662-53826-5_8.

Klepousniotou, Ekaterini. 2002. The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language* 81 (1): 205–223. doi:https://doi.org/10.1006/brln.2001.2518.

Kochen, S. and E. P. Specker. 1967. The Problem of Hidden Variables in Quantum Mechanics. *Journal of Mathematics and Mechanics* 17 (1): 59–87.

Kujala, Janne and Ehtibar Dzhafarov. 2016. Proof of a Conjecture on Contextuality in Cyclic Systems with Binary Variables. *Foundations of Physics* 46. doi:10.1007/s10701-015-9964-8.

Kujala, Janne V. and Ehtibar N. Dzhafarov. 2019. Measures of contextuality and non-contextuality. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 377 (2157): 20190149. doi:10.1098/rsta.2019.0149.

Li, Qiuchi, Dimitris Gkoumas, Alessandro Sordoni, Jian-Yun Nie, and Massimo Melucci. 2020. Quantum-inspired neural network for conversational emotion recognition. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*.

Li, Qiuchi, Sagar Uprety, Benyou Wang, and Dawei Song. 2018. Quantum-Inspired Complex Word Embedding. In *Proceedings of The Third Workshop on Representation Learning for NLP*, pages 50–57. Melbourne, Australia: Association for Computational Linguistics. doi:10.18653/v1/W18-3006.

Li, Qiuchi, Benyou Wang, and Massimo Melucci. 2019. CNM: An Interpretable Complex-valued Network for Matching. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4139–4148. Minneapolis, Minnesota: Association for Computational Linguistics. doi:10.18653/v1/N19-1420.

Meichanetzidis, Konstantinos, Stefano Gogioso, Giovanni De Felice, Nicolò Chiappori, Alexis Toumi, and Bob Coecke. 2020. Quantum Natural Language Processing on Near-Term Quantum Computers.

Meyer, Francois and Martha Lewis. 2020. Modelling lexical ambiguity with density matrices. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 276–290. Online: Association for Computational Linguistics. doi:10.18653/v1/2020.conll-1.21.

Mitchell, Jeff and Mirella Lapata. 2008. Vector-based models of semantic composition. In *Proceedings of ACL-08: HLT*, pages 236–244. Columbus, Ohio: Association for Computational Linguistics.

Pearl, Judea. 2000. *Causality*. Cambridge University Press.

Pickering, Martin and Steven Frisson. 1999. The processing of metonymy: evidence from eye movements. *Journal of Experimental Psychology : Learning , Memory, and Cognition* 25: 1347– 65.

———. 2001a. Obtaining a figurative interpretation of a word: support for underspecification. *Metaphor and Symbol* 16: 149–171.

———. 2001b. Processing Ambiguous Verbs: Evidence from Eye Movements. *Journal of experimental psychology. Learning, memory, and cognition* 27: 556–73. doi:10.1037/0278-7393.27.2.556.

Piedeleu, Robin, Dimitri Kartsaklis, Bob Coecke, and Mehrnoosh Sadrzadeh. 2015. Open system categorical quantum semantics in natural language processing. *CoRR* abs/1502.00831.

Pustejovsky, James. 1995. *The Generative Lexicon*. The MIT Press.

Rayner, Keith and Susan A. Duffy. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14 (3): 191–201. doi:10.3758/BF03197692.

Sadrzadeh, Mehrnoosh, Stephen Clark, and Bob Coecke. 2013. The Frobenius anatomy of word meanings I: subject and object relative pronouns. *Journal of Logic and Computation* 23 (6): 1293–1317. doi:10.1093/logcom/ext044.

———. 2014. The Frobenius anatomy of word meanings II: possessive relative pronouns *. *Journal of Logic and Computation* 26 (2): 785–815. doi:10.1093/logcom/exu027.

Sadrzadeh, Mehrnoosh, Matthew Purver, Julian Hough, and Ruth Kempson. 2018. Exploring semantic incrementality with dynamic syntax and vector space semantics. In *Proceedings of the 22nd Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. Aix-en-Provence, France: SEMDIAL.

Schütze, Hinrich. 1998. Automatic Word Sense Discrimination. *Computational Linguistics* 24 (1): 97–123.

Shutova, Ekaterina. 2010. Automatic metaphor interpretation as a paraphrasing task. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1029–1037.

Tanenhaus, Michael K., James M. Leiman, and Mark S. Seidenberg. 1979. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior* 18 (4): 427 – 440. doi:https://doi.org/10.1016/S0022-5371(79)90237-8.

Wang, Daphne. 2020. *Sheaf Theoretic Models of Contextuality from Quantum Measurements to Natural Language*. Master's thesis, MRes in Quantum Technologies, Centre for Doctoral Training in Delivering Quantum Technologies University College London.

Wang, Daphne, Mehrnoosh Sadrzadeh, Samson Abramsky, and Víctor H. Cervantes. 2021. On the quantum-like contextuality of ambiguous phrases. In *Proceedings of Workshop on Semantic Spaces at the Intersection of NLP, Physics, and Cognitive Science (SemSpace2021) colocated with 14th International Conference on Computational Semantics (IWCS 2021)*. ACL Anthology.

Wang, Zheng and Jerome R. Busemeyer. 2013. A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science* 5 (4): 689–710. doi:10.1111/tops.12040.

A. Complete Dataset

(Verb, Noun)	Δ	err(Δ)	$2\Delta_v$
(*admit*, *band*)	2.00	1.41	0.00
(*cast*, *band*)	0.80	0.67	0.70
(*cast*, *bank*)	2.00	1.22	1.00
(*cast*, *beam*)	0.00	1.07	0.00
(*cast*, *coach*)	4.00	1.41	2.00
(*cast*, *yarn*)	0.00	1.12	0.00
(*file*, *cabinet*)	2.00	1.01	2.00
(*file*, *plant*)	1.00	0.87	1.00
(*file*, *volume*)	1.50	0.87	1.50
(*plug*, *band*)	1.33	0.41	0.67
(*saw*, *cabinet*)	0.32	0.29	0.28
(*saw*, *volume*)	0.41	0.60	0.41
(*tap*, *cabinet*)	2.00	1.22	1.00
(*tap*, *pen*)	0.00	1.03	0.00

Figure 7. Verb with mult. meanings - Nouns with mult. meanings

(Verb, Noun)	Δ	err(Δ)	$2\Delta_v$
(*admit*, *letter*)	2.00	1.22	2.00
(*admit*, *press*)	2.00	1.41	0.00
(*cast*, *letter*)	2.00	1.15	2.00
(*pen*, *letter*)	0.00	0.59	0.00

Figure 8. Verb with mult. meanings - Nouns with mult. senses

(Verb, Noun)	Δ	err(Δ)	$2\Delta_v$
(*bury*, *letter*)	0.00	1.41	0.00
(*capture*, *fall*)	2.00	1.10	1.20
(*capture*, *press*)	2.00	1.41	2.00
(*catch*, *letter*)	2.00	1.41	0.00
(*copy*, *letter*)	0.58	0.42	0.32
(*follow*, *press*)	2.00	1.22	0.00
(*label*, *press*)	0.00	1.15	0.00
(*launch*, *letter*)	2.00	1.41	0.00
(*reflect*, *letter*)	0.22	0.78	0.22
(*seize*, *press*)	4.00	1.15	2.00
(*throw*, *letter*)	0.40	1.10	0.00

Figure 9. Verbs with mut. senses - Nouns with mult. senses

(Verb, Noun)	Δ	err(Δ)	$2\Delta_v$
(*adopt*, *band*)	2.00	1.22	2.00
(*adopt*, *bank*)	0.00	1.00	0.00
(*adopt*, *boxer*)	1.87	0.53	0.43
(*adopt*, *coach*)	1.00	1.12	1.00
(*bill*, *band*)	0.00	1.12	0.00
(*bury*, *boxer*)	2.50	0.87	1.00
(*capture*, *band*)	0.80	0.63	0.27
(*capture*, *bank*)	1.20	0.63	0.80
(*capture*, *organ*)	2.00	1.41	2.00
(*capture*, *palm*)	0.32	0.39	0.00
(*capture*, *plant*)	1.72	0.40	0.17
(*capture*, *port*)	3.89	0.75	1.89
(*clean*, *organ*)	1.00	0.71	0.67
(*climb*, *boxer*)	1.87	1.03	0.93
(*climb*, *palm*)	0.00	1.41	0.00
(*climb*, *plant*)	0.10	0.16	0.04
(*copy*, *band*)	0.50	0.60	0.44
(*follow*, *band*)	2.00	0.91	0.00
(*follow*, *bank*)	4.00	1.41	2.00
(*follow*, *trip*)	0.00	1.22	0.00
(*grasp*, *band*)	4.00	1.41	2.00
(*grasp*, *palm*)	0.00	1.41	0.00
(*inherit*, *bank*)	2.00	0.71	0.50
(*inherit*, *boxer*)	2.00	1.15	0.00
(*inherit*, *plant*)	1.88	1.03	1.75
(*launch*, *band*)	0.46	0.36	0.46
(*launch*, *coach*)	0.71	0.63	0.71
(*launch*, *port*)	1.17	0.76	0.67
(*leak*, *pen*)	0.00	1.41	0.00
(*mount*, *band*)	2.00	1.22	0.00
(*mount*, *cabinet*)	0.00	1.10	0.00
(*mount*, *coach*)	4.00	0.82	2.00
(*mount*, *volume*)	0.00	0.65	0.00
(*mount*, *watch*)	2.25	0.79	0.75
(*reach*, *band*)	1.54	0.59	0.51
(*reach*, *bank*)	3.20	0.23	1.34
(*reach*, *coach*)	1.63	0.47	0.93
(*reach*, *pitcher*)	2.00	1.15	0.00
(*reach*, *plant*)	1.59	0.19	0.13
(*reach*, *port*)	0.00	1.04	0.00

(Verb, Noun)	Δ	err(Δ)	$2\Delta_v$
(*reflect*, *band*)	2.00	0.91	1.33
(*reflect*, *bank*)	1.48	0.28	0.76
(*reflect*, *beam*)	2.00	0.77	1.00
(*reflect*, *coach*)	1.33	0.76	1.33
(*reflect*, *trip*)	0.00	1.41	0.00
(*reflect*, *volume*)	0.33	0.71	0.33
(*ruin*, *plant*)	1.33	1.15	1.33
(*scent*, *plant*)	0.00	0.49	0.00
(*seize*, *bank*)	1.33	1.15	0.00
(*sketch*, *organ*)	4.00	1.41	2.00
(*stir*, *plant*)	2.00	1.41	0.00
(*tackle*, *bank*)	0.00	1.15	0.00
(*throw*, *band*)	1.23	0.35	0.87
(*throw*, *bank*)	4.00	1.41	2.00
(*throw*, *pen*)	1.33	1.15	0.00
(*throw*, *pitcher*)	0.87	0.46	0.53
(*throw*, *punch*)	0.00	1.00	0.00
(*throw*, *volume*)	2.00	1.41	0.00
(*wipe*, *bank*)	0.00	1.22	0.00
(*wipe*, *pen*)	0.00	1.15	0.00
(*wipe*, *plant*)	0.25	1.06	0.25

Figure 10. Verb with mult. senses - Noun with mult. meanings