Composing Conversational Negation

Razin A. Shaikh*
Mathematical Institute,
University of Oxford
razin.shaikh@maths.ox.ac.uk

Lia Yeh*
Quantum Group,
University of Oxford
lia.yeh@cs.ox.ac.uk

Benjamin Rodatz*
Computer Science,
University of Oxford
benjamin.rodatz@cs.ox.ac.uk

Bob Coecke
Cambridge Quantum
bob.coecke@cambridgequantum.com

Negation in natural language does not follow boolean logic and is therefore inherently difficult to model. In particular, it takes into account the broader context of what is being negated. In previous work, we have proposed a framework for negation for words that accounts for 'worldly context'. In this paper, we extend that proposal now accounting for the compositional structure inherent in language, within the DisCoCirc framework. Concretely, first, besides worldly context, we also consider compositional textual context for negating. Then we compose the negations of single words to capture the negation of sentences. We also explore the negation of evolving meanings within compositional language circuits.

1 Introduction

Negation in language is a complicated operation. Differing views of negation in language are a recurring subject of debate amongst linguists, epistemologists, and psychologists. One view maintains that negation in language conveys denial, rather than assertion, of a proposition (the matching bias account) [9]. Another view on negation in language asserts that it is the collective notion of plausible alternatives (the contrast classes account) [24]; this goes as far back as Plato's view of not-X as *otherness*-than-X [17]. An explanation compatible with both views is that there can be different stages at which the negation is interpreted, for instance initially denying information, and later searching for alternatives [28]. In view of these accounts informed by psychology experiments, our contribution is to formulate a process to model conversational negation: from when it is first manifested, to throughout the ensuing sentences as its meaning evolves.

Recent work has conceptualized and generated experimental support for lexical entailment measures [2, 18] and logical negations [19, 29] within compositional distributional semantics of natural language processing. Yet logical negation, in isolation, is antithetical to conversational negation. Consider the sentences:

a) `This is not a planet; this is a star.`
b) `This is not a planet; this is a hamster.`

Both sentences are grammatically correct. Yet, most users of the English language will concur that there is something wrong with sentence b). Unlike sentence a), which seems reasonable without context, sentence b) must undergo a highly unusual 'contextual pressure' [15] to be

*Equal contribution

Submitted to:
ACT 2021

believable—imagine a sci-fi flick about planet-sized hamsters. The plausibility of different alternatives to a negative word naturally has a grading [24, 15]. The most reasonable alternatives are the ones similar to the negated word and are applicable across varied contexts.

This brings us to an interesting observation. Suppose we are negating the word X. We expect *not* X to be in some sense orthogonal to X [31]. We also expect that *not* X evokes alternatives like Y which are similar to X. These two ideas are in contention: transitivity of the two relations *not* $X \sim Y$ and $Y \sim X$ yields *not* $X \sim X$. In distributional approaches to natural language processing, this is a well-known problem that a word and its lexical opposite appear in similar contexts and are hard to differentiate meanings of [22].

Our next observation is that conversational negation is *operational*: a human who knows the meaning of A can infer the meaning of *not* A, without ever having to see or hear *not* A in any context. We put this to the test in [29], where we propose an operational conversational negation of a word and experimentally determine that for considering plausibility of alternatives to a negated word, its judgment positively correlates with human judgment.

In order to extend this negation of a word to negation of a sentence, a new challenge arises: the ambiguity is not only with regards to the meaning of a negated word, but also with regards to which word(s) in the negated sentence the negation is principally applied to. As an example, take the sentence "`Johnny didn't travel to Manchester by train`" [25]. Envision that `Johnny` is given emphasis—the natural conclusion is that someone else, instead of Johnny, went to Manchester by train. Correspondingly, if the emphasized word was `Manchester` or `train`, then the respective conclusions would be that Johnny went elsewhere or Johnny took another mode of transportation. Therefore, we note that the conversational negation of this sentence is arrived at from the conversational negation of its constituent words. We also see that the grammatical structure is unaltered between the non-negated and negated forms of the sentence. This is in line with the principle Kamp and Reyle employed in deciphering semantics and logic of natural language that "sentences in the scope of negation should be treated as they would be if the negation were absent" [12].

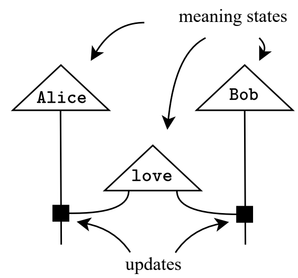
Having surveyed the intricacies at hand, we devise here a model of operational conversational negation which captures these intuitions. After reviewing the compositional DisCoCirc framework (Section 2.1) for the specific case of density matrices (Section 2.2). Then, negation is formulated to be consistent with the paradigm that the word meanings to form sentence meaning, with one or more of the words being negated (Section 3). Second, different interpretations of the negations, along with their probabilities, are encoded in this model of meaning (Section 5). Third, which interpretation was intended is made clearer by incorporating different kinds of contextual information (Sections 4 and 5). Fourth, negation can be performed on *evolving meanings* which accrue meaning within the text (Section 6).

2 Compositional language meaning

2.1 DisCoCirc

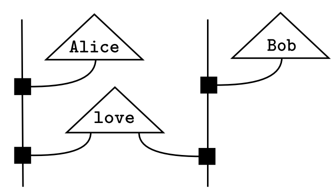
The DisCoCat framework [6] combines grammar (cf. categorial grammar [16]) and meanings (cf. vector embeddings in machine learning) within one compositional framework that enables one to compute the meaning of a sentence from the meanings of its words. To achieve this it exploits the common compact closed categorical structure, be it of vectors and linear maps, or of density matrices and CP-maps [2, 27]. The DisCoCirc framework [4] improved on DisCoCat,

(1) by enabling one to compose sentences into larger text, just as gates are composed in circuits; (2) by allowing meanings to evolve as that text evolves; (3) by specifying the sentence type as the tensored spaces of those entities that evolve. For our purposes, a DisCoCirc diagram has two key ingredients: (1) meaning states; (2) updates [5]:



(1)

For example, here we have the noun meanings `Alice` and `Bob`, which initially are separate, being updated with the verb meaning `love`. Alternatively, we can have noun-wires with open input, which we then update to being `Alice` and `Bob` respectively, and then `love`:



(2)

2.2 Positive operators: entailment, mixing and disambiguation

While the DisCoCirc framework allows for various encoding of meaning, for this paper, we have chosen the compact closed category **CPM(FHilb)**, which has the same objects as **FHilb** with morphisms being completely positive maps. The CPM construction was originally introduced in [30]. Within this category, meanings can then be represented as positive operators: linear operators, which are equal to their own conjugate transpose (Hermitian) and have non-negative eigenvalues (positive semidefinite).

In contrast to vectors, which have no inherent ordering structure [1], positive operators can be viewed as an extension of vector spaces to allow for encoding lexical entailment structure such as proposed in [2, 18]. We use these entailment measures to capture hyponomy; a word w_1 is a hyponym of w_2 if w_1 is a type of w_2; then, w_2 is a hypernym of w_1. For example, *dog* is a hyponym of *animal*, and *animal* is a hypernym of *dog*. Cohyponyms have a hypernym in common. These entailment measures are often graded and take values between 0 (no entailment) and 1 (full entailment).

Additionally, positive operators can be used to encode ambiguity—words having multiple meanings—via mixing [26, 27, 20], i.e. taking weighted sums over the different meanings. Coecke and Meichanetzidis [5] have shown how ambiguity can be disambiguated through later meaning updates.

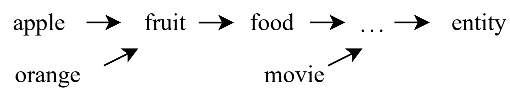


Figure 1: Example of hyponymy structure as can be found in entailment hierarchies

3 Conversational negation of words

3.1 Logical versus conversational negation

Whilst conversational negation is markedly different from logical negation, it obeys rules derived from negation in logic. Critically, the contrapositive must be sensible. Noveck et. al. gives the example that because `chocolate ice cream` fully entails `ice cream`, `John did not buy ice cream` fully entails `John did not buy chocolate ice cream` [23].

Two proposed logical negations defined on positive operators satisfy this requirement of reversing entailment. Lewis [19] proposes the operation $\neg \mathsf{X} := \mathbb{I} - \mathsf{X}$, mapping in the case of pure state to the orthogonal subspace, as Widdows and Peters did for vectors [31]. In [29], we propose another logical negation based on generalizing the matrix inverse, due to it reversing the (graded) Löwner order [2] on positive operators.

Intuitively, the logical negation of a positive operator is akin to a mixture of everything that is not that word, irrespective of their relative plausibilities. Despite this aligning with the set-theoretic notion of complement sets, this is unlike how humans perceive negation, as discussed in Section 1. Indeed, in our prior experiments on plausibility of alternatives to a negated word, we found that both of the proposed logical negations yield negative correlation with human ratings. Upon amending logical negation with the context of the negated word, these became positively correlated with human perception of negation in conversation [29].

3.2 Worldly context

Worldly context is another primary ingredient of conversational negation. It encodes the intuitive understanding of the world most readers possess. Worldly context captures the context a word tends to appear in and thus encodes the space of possible alternatives to a word. This worldly context can be utilized to restrict the results of the logical negation to the context of a word and therefore to reasonable alternatives.

To build worldly context for a given word, in [29] we propose utilizing entailment hierarchies such as displayed in Figure 1, where a directed edge indicates an entailment relation. Examples of such hierarchies are the human curated WordNet [10] and the unsupervised Hearst patterns [11].

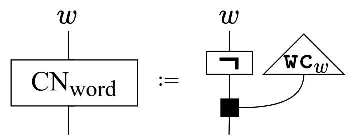
These hierarchies are used to extract the context a word normally appears in. The context of `apple` in Figure 1 is more likely to be the closer hypernyms such `fruit` and `food`, as we usually think of `apple` as a `fruit` or `food`, rather than the distant hypernym `entity`. Building on this idea, we [29] propose to construct the worldly context of a word by considering its hypernym paths and taking a weighted sum over all hypernyms. Hence, for a word w with hypernyms $h_1, \ldots, h_n$ ordered from closest to furthest, we define the worldly context $\mathtt{wc}_w$ as:

$$\llbracket \mathtt{wc}_w \rrbracket := \sum_i p_i \llbracket h_i \rrbracket \tag{3}$$

where $p_i \geq p_{i+1}$ for all i. We denote the positive operator of a word with double brackets.

3.3 Framework for conversational negation of words

In [29], conversational negation of words, written as the operation $\mathrm{CN}_{\mathrm{word}}$, is defined as



(4)

This framework can be interpreted as the following three steps.

1. Calculate the logical negation $\neg(\llbracket w \rrbracket)$.
2. Compute the worldly context $\llbracket \texttt{wc}_w \rrbracket$.
3. Update the meaning of $\neg(\llbracket w \rrbracket)$ by composing with $\llbracket \texttt{wc}_w \rrbracket$ to obtain $\neg(\llbracket w \rrbracket)$ $\llbracket \texttt{wc}_w \rrbracket$.

This framework is flexible to the choice of logical negation, worldly context generation and the composition operation. We studied and compared the performance of various choices of operations in [29]. While the experimental validation in [29] only focuses on nouns, the same operation is applicable to adjectives and verbs.

4 Composition with textual context for meaning evolution

In the previous section, we defined conversational negation of a word to be the process that logically negates a word and updates its meaning to reflect the worldly context associated with that word. We now embark on exploring how this conversational negation interacts with the compositional part of the compositional distributional semantics.

When a proposition is negated in language, there is often ambiguity in the meaning of that negation. A better understanding can be gained upon receiving later information in the text. This *textual context* triggers a meaning update in the reader's knowledge. As part of this later stage of interpreting the negation, we apply meaning updates in the DisCoCirc framework.

In this section, we walk through an example of how the conversational negation of an ambiguous word can be subsequently disambiguated. To justify use of positive operators in compositional distributional semantics as effective representations of both homonymous and polysemous ambiguity, we refer to recent experiments by Meyer and Lewis [20]. For further analysis and model performance for encoding and resolving ambiguity in compositional distributional semantics, we refer the reader to the extensive literature [3, 14, 13, 26, 27, 5, 20].

Consider the sentences:

a) `This is not chicken, but it is food.`
b) `This is not chicken, but it is an animal.`

The presentation of an object as `not chicken` immediately brings to the reader's mind an assortment of possibilities that `this` could be, closely associated with the concept of `chicken`. In Wordnet [21], the two most likely hypernym paths of the word `chicken` are: `chicken` the food, and `chicken` the animal. This is an instance of polysemy because both senses of the word refer to the same broader meaning: `chicken` the food was formerly `chicken` the animal. The second half of sentences a) and b) clarifies which sense was intended.

We can now answer questions like: If the next sentence states that the negated entity is `tofu`, which sense of `chicken` was intended? As a coarse-grained intuition, the positive operator encoding the meaning of `tofu` has little overlap with `animal` compared to `food`; appending `tofu`

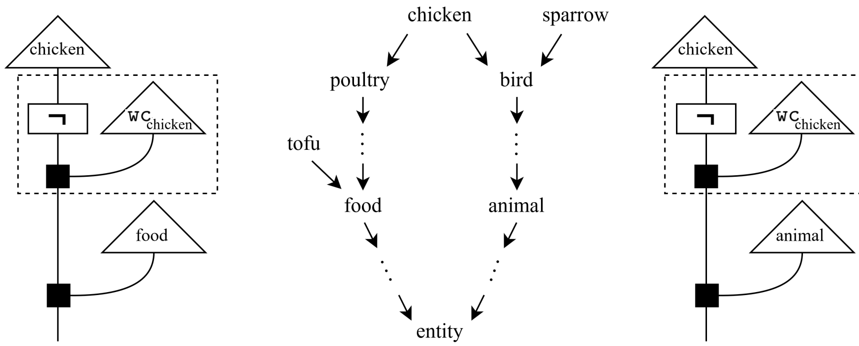


Figure 2: The worldly context of `chicken`, a mixture over the hypernym paths of `chicken` (center), encodes the ambiguity between `food` and `animal`. When pre- and post-composing the conversational negation dotted box in Figure 2 with meaning updates (left: `food`; right: `animal`), the superposition between these senses of the negation of `chicken` is preserved. Thus, we are free to utilize the disambiguation capabilities of DisCoCirc by performing meaning updates with surrounding text before and/or after our conversational negation.

has the effect of retaining `not chicken`'s overlap with `tofu`. Subsequently, in the sum over meanings of `not chicken`, this magnifies the relevance of the `food` term compared to `animal`. In contrast, if the following sentence instead indicated the negated entity is a `sparrow`, the `animal` term dominates. In sum, we have described a compositional means of disambiguating the intended meaning of a negated word. When textual context specifying or highly similar to the intended meaning is applied, this update mechanism reliably increases how much the negated word's hypernyms, cohyponyms, cohyponyms' hyponyms, etc. are entailed.

5 Conversational negation of strings of words

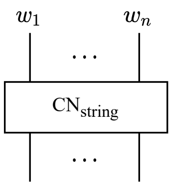
5.1 Meaning of negation of string of words

As pointed out by Oaksford and Stenning [25], the negation of more complex structures consisting of multiple words may be interpreted as the negation of a subset of the constituents. Therefore, a sentence such as "`Bob didn't drive to Oxford by car`" could be interpreted as

a) `Bob didn't drive to Oxford by car - Alice did` (negated: Bob)
b) `Bob didn't drive to Oxford by car - He carpooled` (negated: drive)
c) `Bob didn't drive to Oxford by car - He drove to London` (negated: Oxford)
d) `Bob didn't drive to Oxford by car - He drove a van` (negated: car)
e) `Bob didn't drive to Oxford by car - Alice carpooled to Oxford` (negated: Bob, drive)

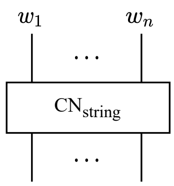
where the underline indicates which words are being negated. The last example is one of many possible cases, which negate multiple constituents. While some of these alternatives might immediately seem more plausible to the reader, the correct choice is inherently dependent on the context.

Based on this interpretation that a negation of multiple words is negating a subset of the constituents, we extend our conversational negation framework to a string of words by utilizing conversational negation of individual words (see Section 3). As the correct interpretation of which words to negate may not usually obvious, we create a mixture of all possible interpretations. Therefore, the negation of a string of n words $w_1 \otimes w_2 \otimes ... \otimes w_n$ is a weighted mixture of all the interpretations where only one word is negated, with all the interpretations where two words are negated, and so on. We have



$$:= \sum_{i=1}^{n} p_{\{w_i\}} (w_1 \otimes \cdots \otimes \mathrm{CN}_{\text{word}}(w_i) \otimes \cdots \otimes w_n) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} p_{\{w_i, w_j\}} (w_1 \otimes \cdots \otimes \mathrm{CN}_{\text{word}}(w_i) \otimes \cdots \otimes \mathrm{CN}_{\text{word}}(w_j) \otimes \cdots \otimes w_n) + \ \ldots \tag{5}$$

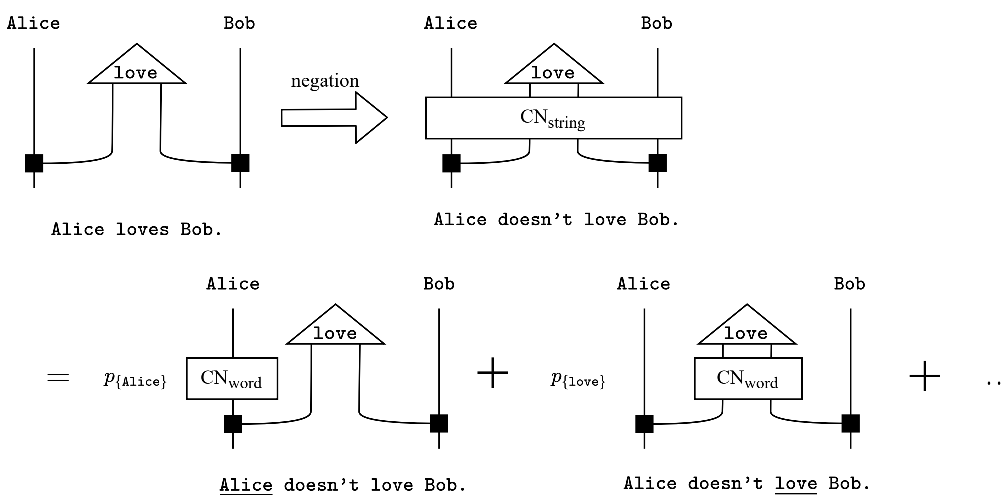
where in the overall mixture representing the negation, each interpretation has some weight. Formally, for $S = \{w_1, ..., w_n\}$ and non-empty $S' \subseteq S$ which we call the **negation set**, we get



$$:= \sum_{S' \in \mathcal{P}(S) \setminus \{\emptyset\}} p_{S'} \bigotimes_{i=1}^{n} \begin{cases} w_i & \text{if } w_i \notin S' \\ \mathrm{CN}_{\text{word}}(w_i) & \text{if } w_i \in S' \end{cases} \tag{6}$$

To apply this negation to sentences, we follow the view presented in [4], which states that in DisCoCirc, sentences can be viewed as processes updating wires. These processes are built from a combination of meaning states, interacting via updates. We propose negation of a sentence to be viewed as the same set of meaning states, first updated by the conversational negation of the words before updating the wires as if the negation was absent. This is in line with [12], who also grammatically treat sentences with negation as if they were not negated.

For example, applying this framework to the sentence "`Alice doesn't love Bob`", we get



where we sum over all possible non-empty subsets of `Alice`, `love` and `Bob` each of which is weighted by the appropriate scalar.

5.2 Deriving the weights

Therefore the main challenge becomes deriving the weights for the different interpretations of the negation. The choice of correct interpretation, and therefore the weights, is dependent on context. Context can be derived from many sources, such as the person who is speaking and their intentions. In spoken language, intonation could clarify the intent of the speaker by them emphasizing the words which are meant to be negated. Within a given text, context can be derived from the surrounding sentences; for an interpretation to be sensible it has to be similar to the surrounding sentences. Another source of context is the grammatical structure of the negated sentence itself. Given the earlier example "`Bob didn't drive to Oxford by car`" , which specifically mentions the mode of transport, intuitively the focus of negation is on this detail. Additionally, as `car` highly entails `drive`, it is more likely the more specific concept is being negated than the more general concept. If the speaker solely wanted to negate the location which Bob visited, the sentence "`Bob didn't drive to Oxford`" would be sufficient, not requiring any additional detail. The other example, `Alice doesn't love Bob`, seems more ambiguous with the grammatical structure giving no indication of the target of the negation.

Overall, no single source of context is sufficient. A combination of all contextual information available, worldly, textual, grammatical, intonation, modality, etc., is required to determine which interpretation was meant.

While the different weights for the negations sets are context dependent, some general observations can be made. Larger negation sets should tend to have smaller weights. Psychologically speaking this can be motivated by the fact that humans have limited information processing capacity, therefore having to select details to focus on [8, 25]. Considering the previous example, one would require a lot of context for the interpretation of "`Alice doesn't love Bob`" to sensibly imply "`Claire likes Dave`". Secondly, one can observe that the weight of a negation set should depend on the likelihood of its individual elements to be the target of the negation. If we know `Alice` to not be the target of the negation, then probably neither is `Alice` and `love`.

5.2.1 Determining weights using entailment

As mentioned earlier, one possible source of context can be the surrounding text. In a text which solely talks about `Alice` and `Bob`, the sentence "`Alice doesn't love Bob`" probably intends to negate the word `love`, therefore asserting that `Alice` feels emotions other than love for `Bob`. Building on this intuition, we propose to use entailment measures to derive the weights for the different interpretations of a negation. If the given interpretation of negation entails the surrounding sentences to a high degree, then the interpretation is consistent with the surrounding text and hence, it is likely to be the intended meaning of the sentence.

We thus compare each possible interpretation of the negation with the surrounding sentences, where sentences closer in the text have more influence towards the final weighting than sentences that are further away. Let us consider the following, simplified scenario of a negation, followed immediately by the clarification with both sentences of the same grammatical structure:

`This is not red wine`
`This is white wine`

Here, we colour code the sentences to simplify differentiating them throughout the example. For the negation, which is in the first sentence, we have to determine the weights for the negation set {`red`}, {`wine`} and {`red, wine`}, denoted by $p_{\{\mathtt{red}\}}$,$p_{\{\mathtt{wine}\}}$ and $p_{\{\mathtt{red,\ wine}\}}$ respectively. Given the immediate clarification in the following sentence, as a human reader, we know that the intended negation was of `red`, without modifying the `wine`.

To mathematically come to the same conclusion, we calculate the entailment of the different interpretations of the negation with the follow-up sentence. We thus calculate the entailment of the three different interpretations (negation sets) of `not red wine` with `white wine`.

Relying on the simplicity of the example, we compare the two sentences word by word, i.e. adjectives and nouns individually and then take the product of the results. We consider

- `not` <u>`red`</u> `wine` $\sqsubseteq$ `white wine` - We first calculate the entailment of $\mathrm{CN}_{\text{word}}$(`red`) with `white` which is medium as something that is not red could have many other colors, including white. The entailment of `wine` with `wine` is maximal as a wine is indeed a wine. Therefore the overall score of this interpretation is high.
  Overall entailment: high
- `not red` <u>`wine`</u> $\sqsubseteq$ `white wine` - This interpretation has a medium entailment between $\mathrm{CN}_{\text{word}}$(`wine`) with `wine` due to the fact that the conversational negation of a word is similar to said word as observed by [22, 25]. Yet this interpretation has a low entailment between `red` with `white` - something being red does not entail that it is white.
  Overall entailment: low
- `not` <u>`red`</u> <u>`wine`</u> $\sqsubseteq$ `white wine` - This interpretation has a medium entailment between $\mathrm{CN}_{\text{word}}$(`red`) with `white` and a medium entailment between $\mathrm{CN}_{\text{word}}$(`wine`) with `wine`. Therefore the overall score is medium.
  Overall entailment: moderate

Comparing the three interpretations, the first option has the highest score, matching our intuition of being the correct choice.

While this entailment method, presented here, relies on the sentences having identical grammatical structure to compare the sentences word by word, we can also directly compare entailment between two sentences. For this to be robust, we need conjunctions which interact well with the entailment measures to guarantee

$$\mathsf{A}_1 \sqsubseteq_k \mathsf{B}_1, \mathsf{A}_2 \sqsubseteq_{k'} \mathsf{B}_2 \iff \mathsf{A}_1 \mathbin{\blacktriangledown} \mathsf{A}_2 \sqsubseteq_{k \cdot k'} \mathsf{B}_1 \mathbin{\blacktriangledown} \mathsf{B}_2 \tag{7}$$

De las Cuevas et. al. [7] prove this property for certain combinations of conjunctions (called Compr) and entailment measures when $k = k' = 1$. However, further exploration is required to derive weights using entailment measures from any arbitrary sentences. Yet the example presented here provides compelling arguments for using entailment to derive weights for the negation sets.

6 Conversational negation of evolving meanings

6.1 Negating evolving meanings

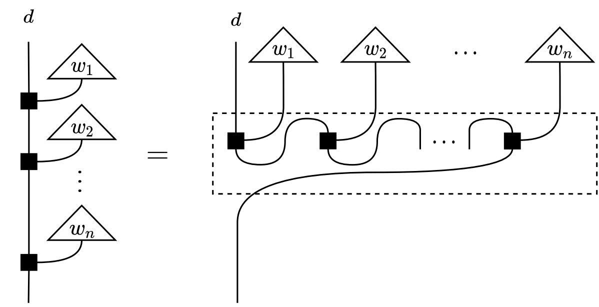
One of the key features of DisCoCirc is that it allows the meaning of entities to evolve as text evolves. The meanings are updated when the wires of the entities are composed with some meaning states. In DisCoCirc, texts that have the same meaning result in the same updates on the wires, even if they contain different sentences. For example, a text containing the following two sentences

```
Bob is a scientist.
Bob is an alcoholic.
```

results in the same circuit as the text containing the single sentence:

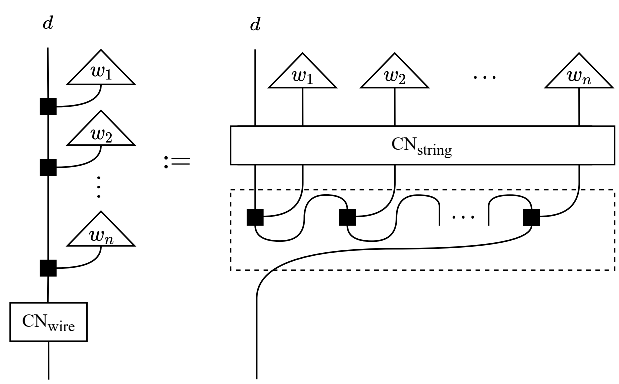
```
Bob who is a scientist is an alcoholic.
```

This motivates us to think of all the meaning updates to a wire as a single large sentence. If we have a sequences of updates to a wire, we can morph the wire using the snake equation to derive a single meaning update process that updates the wire with a sequence of word meanings. Hence, for a wire d whose meaning evolves through updates by words $w_1, \cdots, w_n$, we have



(8)

Now, to perform conversational negation, we can simply apply the framework for conversational negation of a string of words described in Section 5. Therefore, we define conversational negation of a dynamic evolving entity as:



(9)

We call the dynamic entity d at the initial stage when no meaning updates have been performed on it as d_I. Then equation 9 can be written as

$$\mathrm{CN_{wire}}\left(d_I \blacksquare\, w_1 \blacksquare \cdots \blacksquare\, w_n\right) = \blacksquare_{(n \text{ times})} \circ \mathrm{CN_{string}}(d_I, w_1, \cdots, w_n) \tag{10}$$

$$= \blacksquare_{(n \text{ times})} \circ \left[\sum_{S' \in \mathcal{P}(S) \setminus \{\emptyset\}} p_{S'} \bigotimes_{i=0}^{n} \begin{cases} w_i & \text{if } w_i \notin S' \\ \mathrm{CN_{word}}(w_i) & \text{if } w_i \in S' \end{cases} \right] \tag{11}$$

where $S = \{d_I = w_0, w_1, \cdots, w_n\}$. Here, $\mathrm{CN_{word}}$ for $w_1, \cdots, w_n$ can be calculated using the method from Section 3. Now we will calculate $\mathrm{CN_{word}}$ for d_I. Since d_I contains no information at the initial stage, it is equal to the maximally mixed state $\mathbb{I}$. Then the logical negation of d_I is

$$\neg(\llbracket d_I \rrbracket) = \mathbb{I} - \mathbb{I} = \llbracket 0 \rrbracket$$

resulting in the zero matrix. Therefore, for conversational negation of d_I, we get

$$\mathrm{CN}_{\mathrm{word}}\,(\llbracket d_I \rrbracket) \;=\; \neg(\llbracket d_I \rrbracket) \mathbin{\blacksquare} \llbracket \mathtt{wc}_{d_I} \rrbracket \;=\; \llbracket 0 \rrbracket \mathbin{\blacksquare} \llbracket \mathtt{wc}_{d_I} \rrbracket \;=\; \llbracket 0 \rrbracket$$

Thus, for all negation sets containing d_I, the tensor product becomes 0, eliminating all such terms from the summation. This means that the conversational negation of a dynamic entity with evolving meaning can be derived just from the meaning states used to update its meaning.

6.2 Example

Consider the following text where the meaning of the words evolves as the text evolves:

`Alice is a human.`	`Alice is an archaeologist.`
`Bob is a human.`	`Bob is a biologist.`
`Claire is a human.`	`Claire is a pianist.`
`Daisy is a dog.`	`Daisy is a pet.`

Suppose we want to perform the conversational negation of `Alice`'s wire and evaluate how much it entails the wires of `Bob`, `Claire` and `Daisy`. Based on the given text, it is reasonable to expect that someone who is not `Alice` (a human archaeologist) is more likely to be `Bob` (a human biologist) than `Claire` (a human pianist). In fact, someone who is not `Alice` is still more likely to be `Claire` (a human pianist) than `Daisy` (a pet dog). Now we will analyze if the conversational negation presented in Section 6 reflects this intuition.

When we apply the conversational negation on `Alice`'s wire, we get a mixture containing all possible negation sets along with their weights (Equation 11). These negation sets are nonempty subsets of {`Alice`, `human`, `archaeologist`}. The weights of negation sets can be—for instance—determined based on surrounding text as discussed in Section 5.2.1.

However, to explore the maximum entailment that can be achieved from $\mathrm{CN}_{\mathrm{wire}}$(`Alice`) to each remaining wire, we only consider the most appropriate negation sets of $\mathrm{CN}_{\mathrm{wire}}$(`Alice`).

- `Bob` - Since `Bob` is a human biologist, the best negation set of $\mathrm{CN}_{\mathrm{wire}}$(`Alice`) for `Bob` is {`Alice`, `archaeologist`}. The table below shows the entailment between this negation set of $\mathrm{CN}_{\mathrm{wire}}$(`Alice`) and `Bob`'s wire. From the table, it is clear that $\mathrm{CN}_{\mathrm{wire}}$(`Alice`) **highly** entails `Bob`.

$\mathrm{CN}_{\mathrm{wire}}$(`Alice`)	`Bob`'s wire	Entailment
$\mathrm{CN}_{\mathrm{word}}$(`Alice`)	`Bob`	medium
`human`	`human`	1 (max)
$\mathrm{CN}_{\mathrm{word}}$(`archaeologist`)	`biologist`	high

- `Claire` - Similar to `Bob`, the best negation set for `Claire` is {`Alice`, `archaeologist`}. As shown in table below, $\mathrm{CN}_{\mathrm{wire}}$(`Alice`) **moderately** entails `Claire`.

$\mathrm{CN}_{\mathrm{wire}}$(`Alice`)	`Claire`'s wire	Entailment
$\mathrm{CN}_{\mathrm{word}}$(`Alice`)	`Claire`	medium
`human`	`human`	1 (max)
$\mathrm{CN}_{\mathrm{word}}$(`archaeologist`)	`pianist`	medium

- `Daisy` - For `Daisy` the pet dog, the best negation set of $\mathrm{CN}_{\mathrm{wire}}$(`Alice`) is {`Alice`, `human`, `archaeologist`}. From the table below, $\mathrm{CN}_{\mathrm{wire}}$(`Alice`) only **slightly** entails `Daisy`.

$\mathrm{CN}_{\mathrm{wire}}$(`Alice`)	`Daisy`'s wire	Entailment
$\mathrm{CN}_{\mathrm{word}}$(`Alice`)	`Daisy`	medium
$\mathrm{CN}_{\mathrm{word}}$(`human`)	`dog`	medium
$\mathrm{CN}_{\mathrm{word}}$(`archaeologist`)	`pet`	low

Therefore, in our proposed framework, someone who is not `Alice` has the highest chance to be (from most to least likely): `Bob`, `Claire` and `Daisy`, which indeed lines up with the human intuition. Yet the final result of the negation depends on the weights of the negation sets, which are determined by the context. Hence, if for some reason, the negation set {`Alice, human`} has been determined to be the correct interpretation, then $\mathrm{CN}_{\mathrm{wire}}$(`Alice`) might be more closely related to `Daisy` than `Claire` after all.

7 Future work

The intuitions presented in this paper are based on anecdotal evidence gathered utilizing the implementation done for the introduction of conversational negation of words [29]. To empirically validate the framework for conversational negation of strings of words and of evolving meaning ($\mathrm{CN}_{\mathrm{string}}$ and $\mathrm{CN}_{\mathrm{wire}}$), experiments should be devised to (1) explore context determination from surrounding text through entailment and to (2) evaluate the results of the conversational negation. With the basic intuition for deriving the weights from surrounding sentences being solely presented for grammatically identical sentences, further work needs to be done to generalize this process. Additionally, other sources of context, such as grammar, should be explored.

A limit of the current framework for conversational negation of words is that in some cases, the alternatives invoked by the negation of words are not similar to the original word but rather antipodal, i.e. `Alice is not happy` is more likely to assert that she is `sad` instead of `joyful`. This is a limit of the current implementation of the conversational negation of words and independent of the interpretation of negation of multiple words presented here.

The framework for conversational negation of wires is well-defined on meaning updates of single wires, yet it does not tackle entangled wires. As entanglement is a key feature of DisCoCirc, this warrants further exploration.

Finally, a grander challenge is to formalize a mathematical model of the logic underlying conversational NOT, AND, and OR. This requires investigating the extent to which boolean logic holds in a setting known to not follow boolean logic. A long-term goal would be to extend the conversational negation process to a *conversational logic* process, compatible with compositional distributional semantics, particularly its properties with regards to entailment.

References

[1] Esma Balkir, Mehrnoosh Sadrzadeh & Bob Coecke (2016): *Distributional Sentence Entailment Using Density Matrices*. In Mohammad Taghi Hajiaghayi & Mohammad Reza Mousavi, editors: *Topics in Theoretical Computer Science*, Springer International Publishing, Cham, pp. 1–22. Available at `https://doi.org/10.1007/978-3-319-28678-5_1`.

[2] Dea Bankova, Bob Coecke, Martha Lewis & Dan Marsden (2019): *Graded hyponymy for compositional distributional semantics*. *Journal of Language Modelling* 6(2), p. 225–260. Available at `https://doi.org/10.15398/jlm.v6i2.230`.

[3] William Blacoe, Elham Kashefi & 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*, Association for Computational Linguistics, Atlanta, Georgia, pp. 847–857. Available at `https://www.aclweb.org/anthology/N13-1105`.

[4] Bob Coecke (2020): *The Mathematics of Text Structure*. Available at `https://arxiv.org/abs/1904.03478`.

[5] Bob Coecke & Konstantinos Meichanetzidis (2020): *Meaning updating of density matrices*. *FLAP* 7, pp. 745–770. Available at `https://arxiv.org/abs/2001.00862`.

[6] Bob Coecke, Mehrnoosh Sadrzadeh & Stephen Clark (2010): *Mathematical foundations for a compositional distributional model of meaning*. In J. van Benthem, M. Moortgat & W. Buszkowski, editors: *A Festschrift for Jim Lambek*, *Linguistic Analysis* 36, pp. 345–384. Available at `https://arxiv.org/abs/1003.4394`. Arxiv:1003.4394.

[7] Gemma De las Cuevas, Andreas Klinger, Martha Lewis & Tim Netzer (2020): *Cats climb entails mammals move: preserving hyponymy in compositional distributional semantics*. In: *Proceedings of SEMSPACE 2020*. Available at `https://arxiv.org/abs/2005.14134`.

[8] Jonathan St B. T. Evans (1989): *Bias in human reasoning: Causes and consequences.* Lawrence Erlbaum Associates, Inc.

[9] Jonathan St B. T. Evans, John Clibbens & Benjamin Rood (1996): *The Role of Implicit and Explicit Negation in Conditional Reasoning Bias*. *Journal of Memory and Language* 35, pp. 392–409.

[10] Christiane Fellbaum (1998): *WordNet: An Electronic Lexical Database*. Available at `https://wordnet.princeton.edu/`.

[11] Marti A. Hearst (1992): *Automatic acquisition of hyponyms from large text corpora*. In: *Coling 1992 volume 2: The 15th international conference on computational linguistics*.

[12] Hans Kamp & Uwe Reyle (2013): *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*. 42, Springer Science & Business Media.

[13] Dimitri Kartsaklis & Mehrnoosh Sadrzadeh (2013): *Prior Disambiguation of Word Tensors for Constructing Sentence Vectors*. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Seattle, Washington, USA, pp. 1590–1601. Available at `https://www.aclweb.org/anthology/D13-1166`.

[14] Dimitri Kartsaklis, Mehrnoosh Sadrzadeh & Stephen Pulman (2013): *Separating Disambiguation from Composition in Distributional Semantics*. In: *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, Association for Computational Linguistics, Sofia, Bulgaria, pp. 114–123. Available at `https://www.aclweb.org/anthology/W13-3513`.

[15] Germán Kruszewski, Denis Paperno, Raffaella Bernardi & Marco Baroni (2016): *There is no logical negation here, but there are alternatives: Modeling conversational negation with distributional semantics*. *Computational Linguistics* 42(4), pp. 637–660. Available at `http://dx.doi.org/10.1162/COLI_a_00262`.

[16] Joachim Lambek (2007): *From Word to Sentence: A Pregroup Analysis of the Object Pronoun Who(m)*. Journal of Logic, Language, and Information 16(3), pp. 303–323. Available at `https://doi.org/10.1007/s10849-006-9035-9`.

[17] Edward N. Lee (1972): *Plato on Negation and Not-Being in the Sophist*. The Philosophical Review 81(3), pp. 267–304. Available at `http://www.jstor.org/stable/2184327`.

[18] Martha Lewis (2019): *Compositional Hyponymy with Positive Operators*. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), INCOMA Ltd., Varna, Bulgaria, pp. 638–647, doi:10.26615/978-954-452-056-4_075. Available at `http://dx.doi.org/10.26615/978-954-452-056-4_075`.

[19] Martha Lewis (2020): *Towards logical negation for compositional distributional semantics*. IfCoLoG Journal of Logics and their Applications 7(3). Available at `https://arxiv.org/abs/2005.04929`.

[20] Francois Meyer & Martha Lewis (2020): *Modelling Lexical Ambiguity with Density Matrices*. In: Proceedings of the 24th Conference on Computational Natural Language Learning, Association for Computational Linguistics, Online, pp. 276–290, doi:10.18653/v1/2020.conll-1.21. Available at `https://www.aclweb.org/anthology/2020.conll-1.21`.

[21] George A. Miller (1995): *WordNet: A Lexical Database for English*. Commun. ACM 38(11), p. 39–41, doi:10.1145/219717.219748. Available at `https://doi.org/10.1145/219717.219748`.

[22] Saif M. Mohammad, Bonnie J. Dorr, Graeme Hirst & Peter D. Turney (2013): *Computing Lexical Contrast*. Computational Linguistics 39(3), pp. 555–590, doi:10.1162/COLI_a_00143. Available at `https://doi.org/10.1162/COLI_a_00143`.

[23] Noveck, Chierchia, Chevaux, Guelminger & Sylvestre (2002): *Linguistic-pragmatic factors in interpreting disjunctions*. Thinking and Reasoning 8, pp. 297–326.

[24] Mike Oaksford (2002): *Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning*. Thinking & Reasoning 8(2), pp. 135–151, doi:10.1080/13546780143000170. Available at `https://doi.org/10.1080/13546780143000170`.

[25] Mike Oaksford & Keith Stenning (1992): *Reasoning with conditionals containing negated constituents*. Journal of Experimental Psychology: Learning, Memory, and Cognition 18(4), p. 835. Available at `https://doi.org/10.1037/0278-7393.18.4.835`.

[26] Robin Piedeleu (2014): *Ambiguity in Categorical Models of Meaning*. Master's thesis, University of Oxford.

[27] Robin Piedeleu, Dimitri Kartsaklis, Bob Coecke & Mehrnoosh Sadrzadeh (2015): *Open System Categorical Quantum Semantics in Natural Language Processing*.

[28] Jérôme Prado & Ira Noveck (2006): *How reaction time measures elucidate the matching bias and the way negations are processed*. Thinking & Reasoning - THINK REASONING 12, pp. 309–328, doi:10.1080/13546780500371241.

[29] Benjamin Rodatz, Razin A. Shaikh & Lia Yeh (2021): *Conversational Negation using Worldly Context in Compositional Distributional Semantics*. In: To appear in Proceedings of SEMSPACE 2021. Available at `https://arxiv.org/abs/2105.05748`.

[30] Peter Selinger (2007): *Dagger Compact Closed Categories and Completely Positive Maps: (Extended Abstract)*. Electronic Notes in Theoretical Computer Science 170, pp. 139–163, doi:https://doi.org/10.1016/j.entcs.2006.12.018. Available at `https://www.sciencedirect.com/science/article/pii/S1571066107000606`. Proceedings of the 3rd International Workshop on Quantum Programming Languages (QPL 2005).

[31] Dominic Widdows & Stanley Peters (2003): *Word vectors and quantum logic: Experiments with negation and disjunction*. Mathematics of language 8(141-154). Available at `https://www.puttypeg.net/papers/quantum-senses.pdf`.