# A Frobenius Model of Information Structure
# in Categorical Compositional Distributional Semantics

**Dimitri Kartsaklis**  **Mehrnoosh Sadrzadeh**

Queen Mary University of London
School of Electronic Engineering and Computer Science
Mile End Road, London E1 4NS, UK
{d.kartsaklis;m.sadrzadeh}@qmul.ac.uk

## Abstract

The categorical compositional distributional model of Coecke et al. (2010) provides a linguistically motivated procedure for computing the meaning of a sentence as a function of the distributional meaning of the words therein. The theoretical framework allows for reasoning about compositional aspects of language and offers structural ways of studying the underlying relationships. While the model so far has been applied on the level of syntactic structures, a sentence can bring extra information conveyed in utterances via intonational means. In the current paper we extend the framework in order to accommodate this additional information, using Frobenius algebraic structures canonically induced over the basis of finite-dimensional vector spaces. We detail the theory, provide truth-theoretic and distributional semantics for meanings of intonationally-marked utterances, and present justifications and extensive examples.

## 1 Introduction

Distributional models of meaning, in which a word is represented as a high dimensional vector of contextual statistics in a metric space, provide a convincing framework for lexical semantics that has been found useful in a number of natural language processing tasks (Schütze, 1998; Landauer and Dumais, 1997; Manning et al., 2008). Despite their success at the word level, the underlying hypothesis of these approaches does not naturally scale up to phrases or sentences due to the infinite capacity of language to produce new meanings from a finite vocabulary and a set of grammar rules.

Coecke et al. (2010) provide a solution to the problem by noticing that the category of finite-dimensional vector spaces and linear maps is homomorphic to a grammar expressed as a pregroup (Lambek, 2008); specifically, both share compact closed structure (Kelly, 1972). In practice this means that any grammatical derivation based on the type-logical identities of the individual words in a sentence can be translated to a (multi-)linear map which, when applied on the vectorial representations of the words therein, results in a sentence vector. The grammatical type of a word determines the vector space in which this word lives. Taking nouns to be simple vectors in a basic vector space $N$, an adjective, for example, becomes a linear map $N \to N$, or equivalently, a matrix in $N \otimes N$; furthermore, a transitive verb is a bi-linear map $N \otimes N \to S$, living in $N \otimes S \otimes N$. Composition takes the form of tensor contraction, which is a generalization of matrix multiplication to higher order tensors.

In general, the model resembles a quantitative linear-algebraic version of the formal semantics approach (Montague, 1970), in the sense that syntax strictly guides the semantic composition. Interestingly, syntax seems to co-exist with a distinct structural layer, the purpose of which is to optimize the message that an utterance conveys. This aspect is known as *information structure*, and at the phrase or sentence level is expressed as a distinction between a theme part (information that is generally agreed to be known to both of the interlocutors) and a rheme part—information that is new for the addressee. The exact relation that holds between syntactical and information structure is an interesting and controversial topic. For example, a theme does not have to comprise a valid grammatical constituent

in the strict sense of the term, as it is evident in the following example:

(1) Q: Do you need anything?
A: [I would like]$_T$ [some tea]$_R$

The distinction between a theme and rheme is denoted by the presence of a boundary that can be expressed by phonological, morphological or even syntactical means, depending on the language. Furthermore, the presence of such boundaries suggest the existence of a distinct composition operator related to information structure and different than the one that would be normally used for syntax.

In this paper we extend the categorical model of Coecke et al. (2010) in a way to accommodate an information structure layer of composition. In order to achieve this, we model intonational boundaries (the devices for defining information structure in English) by using the multiplication part of the Frobenius algebra that is canonically induced over any vector space with fixed basis, in order to endow equal contribution of the theme and rheme on the vectorial representation of a sentence, thus putting emphasis on the appropriate part. The resulting model can be seen as containing two types of composition operators: the usual tensor contraction for accommodating syntax, and the Frobenius multiplication for accommodating information structure. We discuss the implications in terms of the resulting vectorial representations for phrases and sentences, and provide connections with existing models from the current literature of compositional distributional semantics. Various examples demonstrate the potential of the model.

## 2 Categorical compositional distributional semantics

The categorical model of Coecke et al. (2010) assigns semantic representations to phrases and sentences of language, based on their grammatical structure and the semantics of individual words. In its most abstract form, this model can be expressed in terms of a structure-preserving passage between grammar and meaning:

$$\mathcal{F} : \text{Grammar} \to \text{Meaning}$$

Given a sequence of words $w_1 \cdots w_n$, its categorical meaning is defined to be:

$$[\![w_1 \cdots w_n]\!] := \mathcal{F}(\alpha)([\![w_1]\!], \cdots, [\![w_n]\!]) \qquad (1)$$

Here, $\alpha$ is derived from the grammatical relationships amongst the words in the sequence. This notion can be formalised in a coherent way, if both the grammar and the meaning are expressed in a high level logical structure, referred to by *compact closure*. Lambek's pregroup algebras (Lambek, 2008) and vector space distributional semantics are examples of compact closed structures. Stipulating that the grammar is expressed in a pregroup algebra and that the meaning of words are vectors constructed using the distributional hypothesis (Harris, 1968), Eq. 1 gets a more concrete form:[1]

$$\overrightarrow{w_1 \cdots w_n} := \mathcal{F}(\alpha)(\overrightarrow{w}_1 \otimes \cdots \otimes \overrightarrow{w}_n) \qquad (2)$$

In the proceeding subsections we make these notions precise and provide intuitions and examples.

### 2.1 Pregroup grammars

A pregroup grammar is a pregroup algebra, linked to the vocabulary of a language via the notion of a type dictionary. We define these structures below.

A pregroup algebra is a partially ordered monoid where each element has a left and a right adjoint. It is denoted by a tuple $(P, \leq, \cdot, 1, (-)^l, (-)^r)$, where $(P, \leq)$ is a partially ordered set, and $\cdot$ is a monoid multiplication with 1 as its unit. For each element $p \in P$ there are $p^l, p^r \in P$, referred to by $p$'s *left* and *right* adjoints, satisfying the following four inequalities:

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

When a pregroup algebra is generated over a base set $\mathcal{B}$, it is denoted by $P(\mathcal{B})$. Given the vocabulary of a language $\Sigma$ and a set of its basic grammatical types $\mathcal{B}$, a pregroup grammar is a relation $D \subseteq \Sigma \times P(\mathcal{B})$ that assigns grammatical types from the pregroup algebra $P(\mathcal{B})$ to the words of the vocabulary $\Sigma$. Such a pregroup grammar is denoted by $P(\mathcal{B}, \Sigma)$.

As an example, suppose $\mathcal{B} = \{n, s\}$, where $n$ stands for a well-formed noun phrase and $s$ for a well-formed sentence. Suppose further that $\Sigma = \{\text{Mary, snores, likes, musicals}\}$. The pregroup dictionary consists of the following set:

$$\{(\text{Mary}, n), (\text{snores}, n^r s), (\text{likes}, n^r s n^l), (\text{musicals}, n)\}$$

---

[1] One can translate the types from other type logics, such as the syntactic calculus and CCG to pregroups and carry on with the same calculations. There is also recent work that directly assigns vector semantics to CCG (Lewis and Steedman, 2013).

One says that a sequence of words $w_1 w_2 \cdots w_n$ for $w_i \in \Sigma$ forms a grammatical sentence, according to a pregroup grammar $P(\mathcal{B}, \Sigma)$, whenever we have:

$$t_1 \cdot t_2 \cdot \ldots \cdot t_n \leq s$$

for $(w_i, t_i) \in D$. The above inequality is often referred to by *grammatical reduction*. For example, 'Mary likes musicals' is a grammatical sentence, since we have the following reduction:

$$n \cdot n^r \cdot s \cdot n^l \cdot n \leq 1 \cdot s \cdot 1 = s \qquad (3)$$

## 2.2 Distributional models

The only piece of information provided by a derivation like the one in Eq. 3 is whether the sentence in question is well-formed or not. Furthermore, we are unable to distinguish between words of the same type. Distributional models of meaning provide a solution to these problems by following the *distributional hypothesis* (Harris, 1968), which states that semantically similar words must appear in similar contexts. Hence, the semantic representation of a word can be given in terms of its distributional behaviour in a large corpus of text. In its simplest form, a word vector is comprised by numbers that show how many times the target word co-occurs with every other word in a selected subset of the vocabulary (usually the most frequent content-bearing words). This allows the representation of words as points in some high dimensional space, where semantic relatedness can be measured (usually by cosine distance) and evaluated. For a concise introduction to distributional models, see (Turney and Pantel, 2010). We will now proceed to show how the quantitative approach of distributional models can be combined with the compositional model of Section 2.1 into a unified account.

## 2.3 Categorical generalization

The theory of categories generalises algebraic constructions to categorical ones (Mac Lane, 1971). Herein, instead of sets, functions or relations, one has objects $A, B$ and morphisms $f: A \rightarrow B$. The generalised binary operation over these is referred to by a product. Posing different conditions on the objects, morphisms, or the product results in different kinds of categories. A *monoidal category* has a product with a unit $I$, that is $A \otimes I \cong I \otimes A \cong A$. These categories are generalisations of partially ordered monoids: elements of the partial order become objects of the category and the partial orderings between them become morphisms. Furthermore, *compact closed categories* are generalisations of pregroups, where the adjunction inequalities correspond to the following $\epsilon$ and $\eta$ morphisms:

$$\epsilon^r: A \otimes A^r \rightarrow I \qquad \eta^r: I \rightarrow A^r \otimes A$$
$$\epsilon^l: A^l \otimes A \rightarrow I \qquad \eta^l: I \rightarrow A \otimes A^l$$

These maps needs to adhere to four axioms, referred to as *yanking equations*, which ensure that all relevant diagrams commute.

The importance of the theory of categories for this paper is that finite-dimensional vector spaces and linear maps also form a compact closed category, denoted by **FVect**. Herein, objects are vector spaces, morphisms are linear maps, and the product is the tensor product between vector spaces whose unit is the scalar field of the vector spaces, in our case, real numbers ($\mathbb{R}$). In the presence of a fixed basis (which is the case we are interested in) the adjoints become identity, that is we have $V^r \cong V^l \cong V$, for a vector space $V$ spanned by $\{\overrightarrow{v}_i\}_i$. As a result the four $\epsilon$ and $\eta$ maps reduce to two:

$$\epsilon: V \otimes V \rightarrow \mathbb{R} \qquad \eta: \mathbb{R} \rightarrow V \otimes V$$

The $\epsilon$ map takes the inner product of two vectors and the $\eta$ map produces a diagonal matrix. The fact that both pregroup algebras and vector spaces form compact closed categories allows us to develop a structure preserving passage between the two mathematical structures, thus enabling us to bridge the grammatical structure to distributional semantics.

## 2.4 From grammar to distributions

A structure preserving passage from grammatical structures (in the form of a pregroup grammar) to semantics (in the form of vector spaces) is given by a map denoted as follows:

$$\mathcal{F}: P(\Sigma, \mathcal{B}) \rightarrow \textbf{FVect} \qquad (4)$$

This is a strongly monoidal passage, which means that it has the following compositional properties for juxtapositions of types in a pregroup grammar:

$$\mathcal{F}(1) = \mathbb{R} \qquad (5)$$
$$\mathcal{F}(p \cdot q) = \mathcal{F}(p) \otimes \mathcal{F}(q) \qquad (6)$$
$$\mathcal{F}(p^r) = \mathcal{F}(p^l) = \mathcal{F}(p) \qquad (7)$$
$$\mathcal{F}(p \leq q) = \mathcal{F}(p) \rightarrow \mathcal{F}(q) \qquad (8)$$

On the level of basic types we assign a vector space to each basic type, that is, $\mathcal{F}(n) = N$ and $\mathcal{F}(s) = S$. As a result of the above assignments, words that have simple types, for example noun phrases, will become vectors in vector space $N$. Words that are functions of one argument become matrices, e.g. intransitive verbs with type $n^r \cdot s$ are elements of $N \otimes S$; and words that are functions of two arguments, e.g. transitive verbs with type $n^r \cdot s \cdot n^l$, become tensors of order 3, living in $N \otimes S \otimes N$ for the specific case. The grammatical reductions are translated to compositions of morphisms, and in particular $\epsilon$-maps.

$$
\begin{aligned}
\overrightarrow{\text{Mary snores}} &= \mathcal{F}(n \cdot n^r \cdot s)(\overrightarrow{\text{Mary}} \otimes \overrightarrow{\text{snores}}) \\
&= (\epsilon_N^r \otimes 1_S)(\overrightarrow{\text{Mary}} \otimes \overrightarrow{\text{snores}})
\end{aligned}
$$

A simple computation shows that the above is equal to $\overrightarrow{\text{Mary}} \times \overrightarrow{\text{snores}}$; similarly, for the meaning of a transitive sentence we obtain:

$$
\overrightarrow{\text{Mary likes musicals}} = \overrightarrow{\text{Mary}} \times \overrightarrow{\text{likes}} \times \overrightarrow{\text{musicals}}
$$

Note that tensor contraction (in spaces with fixed basis) is associative, so there is no need to keep track of brackets in the above. The situation is similar to pregroups, where the monoid multiplication is again associative.

## 2.5 Frobenius algebras

Compact closed categories on their own do not have much structure: there is a binary operation and the maps $\epsilon$ and $\eta$. The expressive power of these categories can be increased using Frobenius algebras. We define these below.

Given a compact closed category $\mathcal{C}$, an object $X \in \mathcal{C}$ has a Frobenius structure on it if there exist the following morphisms:

$$
\begin{aligned}
\Delta &\colon X \to X \otimes X & \iota &\colon X \to I \\
\mu &\colon X \otimes X \to X & \zeta &\colon I \to X
\end{aligned}
$$

These have to satisfy certain conditions, the most important to us being the *Frobenius condition*:

$$
(\mu \otimes 1_X) \circ (1_X \otimes \Delta) = \Delta \circ \mu = (1_X \otimes \mu) \circ (\Delta \otimes 1_X)
$$

Vector spaces with fixed basis do have such structures over them, generally referred to by *copying* and *merging*. For $\overrightarrow{v} \in V, \overline{w} \in V \otimes V$, we have that

$\Delta(\overrightarrow{v}) \in V \otimes V$ is a diagonal matrix whose diagonal elements are weights of $\overrightarrow{v}$, and $\mu(\overline{w}) \in V$ is a vector consisting only of the diagonal elements of $\overline{w}$.

These structures have been used in previous work to encode lower dimensional verb matrices into higher dimensional tensors (Kartsaklis et al., 2012; Kartsaklis et al., 2014) and to pass the information around sentences with relative clauses by copying and merging (Sadrzadeh et al., 2013; Sadrzadeh et al., 2014).

## 2.6 Graphical calculus

In the presence of higher order tensor product spaces, calculations can become quite complex. The formalism of compact closed categories and Frobenius structures is complete with regard to a graphical calculus (Selinger, 2011) that simplifies the computations to a great extend. We briefly overview the main components of this language.

Objects are depicted by lines and morphisms by boxes. Tensor products between objects and morphisms are given by juxtaposition of their diagrams, while composition of morphisms amounts to connecting outputs to inputs. Examples are as follows:
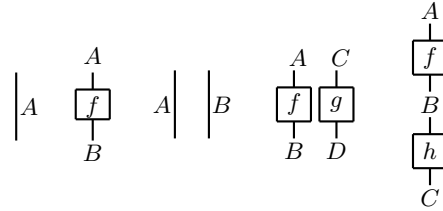


The $\epsilon$ maps are depicted by cups, $\eta$ maps by caps, and yanking by their composition and straightening of the strings. For instance:



The diagrams corresponding to the Frobenius morphisms are as follows:

$$
(\mu, \zeta) = \quad\quad\quad (\Delta, \iota) =
$$



with the Frobenius condition being depicted as:



The defining axioms guarantee that any picture of a Frobenius computation can be reduced to a normal form (so-called a "spider") that only depends on the number of input and output strings of the nodes:
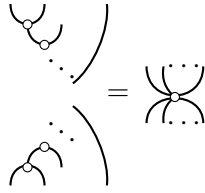
Elements within the objects (for the case of vector spaces, vectors) are depicted by morphisms from the unit. These are shown by triangles with a number of strings emanating from them. The number of strings denotes the order of the tensor; for instance, the diagrams for $\vec{v} \in V$, $\overline{v'} \in V \otimes W$, and $\overline{v''} \in V \otimes W \otimes Z$ are as follows:
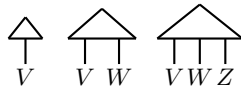
## 3 Information structure and intonation

The term *information structure* collectively refers to techniques that aim to enhance the communication between two interlocutors in order to optimize the conveyed message for the benefit of the addressee (Chafe, 1976). One such technique, for example, is to emphasize a particular part of the utterance that is important for the listener by changing the spoken pitch:

(2)  Q: What does Mary like?
     A: Mary likes MUSICALS

The emphasis imposes a specific information structure to the uttered sentence, essentially splitting it in two parts: The part in upper-case above is what Steedman (2000) calls *rheme*—the information that the speaker wishes to make common ground for the listener; the rest of the sentence, i.e. what the listener already knows, is called *theme*. The question in (2) puts the listener in a specific attentional state, in the context of which an answer such as:

(3)  A: #MARY likes musicals

will be *infelicitous*, that is, not compatible with that state.

The distinction between theme (or topic) and rheme (or comment) has great significance from an information structure point view, since it defines a generic shape for the sentence that directly reflects the attentional needs of the addressee. A further dimension that can be found in both rheme and theme distinguishes between the *focus*, that is, the specific word that receives most of the intonational emphasis, and the background, which consists of the rest of the words in the specific text segment. Note that in contrast to rheme/theme distinction, focus and background seem to operate at the lexical level.[2]

Furthermore, we should point out that although the examples we use in this paper are mainly based on question/answer dialogues, this is not by any means the only case where the presence of a specific information structure can be useful. For example, consider the dialogue:

(4)  –I think Mary likes jazz.
     –Mary likes MUSICALS.

Information structure can be expressed in different ways that may vary from language to language. In English, for example, the means for defining information structure is *intonation*: variations of spoken pitch, the purpose of which is to emphasize parts of the utterance that might be important for the conveyed message, as we saw above in our examples. However, in other languages such as Japanese or Cantonese, the intonational boundaries can be also specifically marked by morphological devices, e.g. special particles (Féry and Krifka, 2008). Finally, the position of a text segment in a sentence can also be an indication of its information-structural role. In English, for example, themes tend to appear at the beginning of a clause.

In this paper we concentrate on the sentence-level distinction between rheme and theme.

## 4 Grammar and intonation

The presence of a distinct layer of information structure that seems to co-exist with the grammatical structure of a sentence, poses the interesting question regarding the exact relationship that holds between those different structural aspects. For example, although the text segment "Mary likes" forms a perfectly acceptable theme, most linguists would agree that it does not also comprise a valid grammatical constituent. In spite of this claim, though, it is interesting to note that a number of categorial grammars, including Combinatory Categorial Grammar (CCG) (Steedman, 2001), treat text segments like the above as possible syntactic constituents. Consider the following ditransitive sentence:

---

[2]Actually many authors use the term *focus* as a synonym for *rheme*; the definitions we give in this paper follow (Steedman, 2000).

(5)  John gave Mary a flower

In CCG, this sentence has a number of different syntactic derivations, two of them are the following:

$$
\begin{array}{ccccc}
\text{John} & \text{gave} & \text{Mary} & \text{a flower} \\
\hline
\text{NP} & \text{((S\backslash NP)/NP)/NP} & \text{NP} & \text{NP} \\
\end{array}
$$

$$
\cfrac{
  \cfrac{
    \cfrac{\text{((S\backslash NP)/NP)/NP} \quad \text{NP}}{\text{(S\backslash NP)/NP}} >
    \quad \text{NP}
  }{\text{S\backslash NP}} >
}{\text{S}} < \qquad (9)
$$

$$
\cfrac{
  \cfrac{
    \cfrac{\text{NP}}{\text{S/(S\backslash NP)}} >\mathbf{T}
    \quad
    \cfrac{\text{((S\backslash NP)/NP)/NP} \quad \text{NP}}{\text{(S\backslash NP)/NP}} >
  }{\text{S/NP}} >\mathbf{B}
  \quad \text{NP}
}{\text{S}} > \qquad (10)
$$

Note that (9) proceeds by first composing the part corresponding to the verb phrase ("gave Mary a flower"); later, in the final step, the verb phrase is composed with the subject 'John'. The situation is reversed for (10), where the use of type-raising and composition rules of CCG allow the construction of the fragment "John gave Mary" as valid grammatical text constituent, which is later combined with the direct object of the sentence ('a flower'). Steedman (2000) argues that this form of different syntactic derivations that one can get even for very simple sentences when using CCG (some times referred to with the somewhat belittling term "spurious readings"), actually serve to reflect variations in information structure. Each one of the above derivations subsumes a different intonational pattern, distinguishing the rheme from the theme when the sentence is used for answering different questions: (9) answers to "Who gave Mary a flower?", whereas (10) to "What did John give to Mary?".

In other words, the claim here is that (a) surface structure and information structure coincide; and (b) the role of information structure is to provide a particular interpretation of the surface structure. Let us define this important idea in a precise way, since it will be the cornerstone of the model presented in this paper:

**Postulate 4.1** *Intonational boundaries in an utterance determine the intended syntactic structure.*

In our grammatical formalism, pregroup grammars, variations in a grammatical derivation similar to above are only implicitly assumed, since the order of composition remains unspecified. This fact is apparent in the pregroup derivation of the example sentence, where both (9) and (10) are subsumed into the following *reduction diagram*:

$$
\begin{array}{ccccc}
\text{John} & \text{gave} & \text{Mary} & \text{a flower} \\
n & n^r s\, n^l n^l & n & n
\end{array} \qquad (11)
$$

Furthermore, it is directly reflected in our semantic space through the functorial passage, via the fact that tensor contraction is associative:

$$
\begin{aligned}
\overrightarrow{\text{John}} \times (\overrightarrow{\text{gave}} \times \overrightarrow{\text{Mary}} \times \overrightarrow{\text{flower}}) \;&=\; \\
(\overrightarrow{\text{John}} \times \overrightarrow{\text{gave}} \times \overrightarrow{\text{Mary}}) \times \overrightarrow{\text{flower}} &
\end{aligned} \qquad (12)
$$

Eq. 12 constitutes a natural manifestation of the *principle of combinatory transparency* (Steedman, 2001): no matter in what order the various text constituents are combined, the semantic representation assigned to the sentence is always the same; in other words, information structure should not affect semantic conditions. Note, however, that even in the strict setting of formal semantics this is not always the case. Consider the behaviour of the following sentence under the presence of the focus-sensitive particle 'only':

(6)  a.  John only gave Mary A FLOWER
     b.  John only gave MARY a flower

The use of different intonational focus clearly changes the semantic value of the sentence: (6a) is true if the only thing that John gave to Mary was a flower (but he might have given things to other girls as well), while (6b) is true if the only person who got a flower from John was Mary.

In the more relaxed and quantitative setting of a compositional distributional model of meaning, the idea of having vectorial representations of words and sentences that reflect intonational patterns seems even more legitimate. This concept is aligned with the distributional nature of such models: given a text corpus containing information structure annotations (of any kind), we would assume that the co-occurrence vector of a word under focus (say, $\overrightarrow{\text{BOOK}}$) would slightly differ from that of the vector representing the normal use of the word ($\overrightarrow{\text{book}}$).[3] Furthermore, we would expect that, after the composition, this difference would be also reflected in

---

[3]In the trivial case, this would be true by the presence of intonational markers in the immediate context of a word under focus, as opposed to its normal use.

the vector representing the meaning of the entire sentence. From the next section we start working towards imposing this behaviour on the categorical model of Coecke et al. (2010).

# 5   Intonation in pregroups

Traditionally, a notational system describing intonation consists of markings that indicate pitch accents and boundaries. Using the notation of Pierrehumbert and Hirschberg (1990), for example, we get the following for our example sentence:[4],[5]

(7)   [MARY]$_R$   [likes MUSICALS]$_T$
       H* L             L+H* LH%

The prosody starts with a sharp pitch accent (H*) that puts the focus on 'Mary', and continues with a rapid fall to low pitch (L boundary) that signifies a transition from rheme to theme. Within theme now, the focus goes to 'musicals' which gets the less rapidly rising pitch L+H*, whereas the boundary LH% expresses a rising continuation that marks the end of theme. In the case that theme precedes the rheme, we have the following pattern:

(8)   [MARY likes]$_T$   [MUSICALS]$_R$
       L+H* LH%           H* LL%

As mentioned earlier, this paper mainly addresses the rheme/theme aspect of information structure, which is directly related to boundary markings. We start by representing an intonational boundary using a special token $\triangleright$, for which the following relations hold:

$$\textit{theme} \triangleright \textit{rheme} \quad \text{or} \quad \textit{rheme} \triangleleft \textit{theme} \quad (13)$$

Naturally, $\triangleright$ is equivalent to LH%, while $\triangleleft$ corresponds to L in the Pierrehumbert and Hirschberg (1990) notation. It is very important to emphasize at this point that the above introduced tokens are far from an ad-hoc means for achieving a goal. Recall from our discussion in Section 3 that while in English the means of imposing information structure is purely phonological, this is not necessarily the case for other languages. As a concrete case, in Buli (a Gur language spoken in Ghana), the rheme is preceded by a *focus marker*, which again can be interpreted as an information-structural boundary since it

---

[4]Example taken from (Steedman, 2000).

[5]From now on we explicitly mark themes and rhemes in our examples for clarity.

separates the theme from rheme; this is shown in the following example (Fiedler et al., 2006):

(9)   Q: What did the woman eat?
       A: ɔ̀   ŋɔ̀b   **kà**   **túé**
          3sg  eat   (FM)   beans

To formalize this mixing of syntactical and information structure in the context of a pregroup grammar, we add the two boundary markers to the vocabulary and introduce two new atomic types:

$$\text{Theme: } \theta \qquad \text{Rheme: } \rho \qquad (14)$$

An intonation pregroup grammar then will have the following form:

$$P(\Sigma \cup \{\triangleright, \triangleleft\}, \{n, s, \theta, \rho\})$$

For the case of a simple transitive sentence, we get the following boundary types, based on the fact that now the boundary (and not the verb) becomes the head of our sentence:

$$\triangleright : \theta^r \cdot s \cdot \rho^l \qquad \triangleleft : \rho^r \cdot s \cdot \theta^l \qquad (15)$$

The type dictionary changes accordingly: a transitive verb such as 'like' will be assigned two more types $n^r \cdot \theta$ and $\theta \cdot n^l$ depending whether it produces a left-hand theme or a right-hand theme in the sentence; similarly, nouns will be assigned the extra type $\rho$. For the two cases of Eq. 13, we obtain the following derivations:





After transferring this to **FVect** via our functor in Eq. 4, and extending its action on atomic types by defining $\mathcal{F}(\theta) = \Theta$ and $\mathcal{F}(\rho) = P$, we get the obvious semantic counterpart:



There are some important observations based on the derivation in (18) above. Firstly, our simple sentence now is given in terms of a theme and a rheme,

as required, both of which contribute equally to its construction. Additionally, note that our verb is not any more a function of two arguments (of a subject and an object) as in the canonical case, but of a single noun: it takes as input a subject in order to return a theme. Hence, in contrast to a typical case of a transitive verb, the semantic representation of which requires a tensor of order 3, in this case the corresponding linear map takes the form $N \to \Theta$, which can be canonically represented by a *matrix $N \otimes \Theta$*.

The question of how to properly model intonation in compositional distributional semantics is evidently epitomized in choosing an appropriate form for the tensor of the $\triangleright$ token in (18). In order to provide an answer to this, we first need to examine the concepts of rheme and theme from a semantic point of view.

## 6   A semantic truth-theoretic argument

We use as an example the following simple case:

(10)   Q: Who does John like?
        A: [John likes]$_T$ [MARY]$_R$

From an extensional point of view, the semantic value of the theme can be seen as a set of *alternative* options (Rooth, 1992), each one of which may be used as a response to the given question:

$$[\![\text{John (might) like}]\!] = \{x | \text{John (might) like } x\}$$

As a consequence, the role of the rheme now is to *restrict* the set of alternatives to a specific choice (Steedman, 2000). Note that this action of restricting the available choices is responsibility of the intonational boundary; indeed, the boundary can be seen as a binary operator that performs the merging of the theme with the rheme, restricting the alternatives set of the theme to a specific response:

(11)   [John likes]$_T$ $\triangleright$ [MARY]$_R$ :=
        $\triangleright$([John likes]$_T$,[MARY]$_R$)

This is what Diagram (18) shows; in our multilinear setting, the boundary becomes a bi-linear map $\Theta \otimes P \to S$ that performs the required "restriction". Now, what is the most appropriate way to model this operation in the extensional setting discussed above? Note that by simply checking if rheme is contained in the alternatives set is not sufficient; this would return *true* or *false* as an answer to a question that expects a person. A more appropriate choice then

is to model the boundary by using set intersection: we take the meaning of rheme to be a singleton that contains the answer, and the meaning of the sentence to be the intersection of rheme with theme:

$$\{\text{Mary}\} \cap \{x | \text{John (might) like } x\} \qquad (19)$$

The answer will be again the singleton $\{\text{Mary}\}$ if Mary is included in the set of people who John potentially likes, and the empty set otherwise. Thus we have achieved our goal: the theme set has been restricted according to the provided response. We generalize this argument to an arbitrary pair of rheme and theme (with $S_{\text{theme}}$ denoting theme's corresponding alternative set) as follows:

$$\triangleright(\text{rheme}, \text{theme}) = \{\text{rheme}\} \cap S_{\text{theme}}$$
$$\begin{cases} \text{rheme} & \text{if rheme} \in S_{\text{theme}} \\ \varnothing & o.w. \end{cases} \qquad (20)$$

### 6.1   From sets to vector spaces

We transfer the above reasoning to vector spaces, by encoding sets and relations in vectorial forms. The vectorial form of a set is a vector space (let it be $N = \{n_i\}_i$) whose basis vectors are the elements of the set. For the sake of demonstration (and this will become clear as the section reads on), we define our sentence space to be a one dimensional space where the origin denotes falsity and everything else denotes truth. One can take this to be a dimension in any vector space; here we take it to be in $N$ and denote its basis vector with a basis vector of $N$. Furthermore, a binary relation such as $likes(x, y)$ can be represented as an adjacency matrix $\mathbf{W}$ in which $W_{ij}$ is 1 if the pair $(i, j)$ is contained in the relation and 0 otherwise. Note that this matrix is isomorphic to a tensor in $N \otimes S \otimes N$, since our sentence space is one-dimensional.

Let us apply categorical composition to compute a vectorial representation for the theme of our sentence, "John likes".

$$(\epsilon_N^r \otimes 1_S) \left( \overrightarrow{n_3} \otimes \left( \sum_{ij} W_{ij} \overrightarrow{n_i} \otimes \overrightarrow{n_j} \right) \right) \; =$$

$$\sum_{ij} W_{ij} \langle \overrightarrow{n_3} | \overrightarrow{n_i} \rangle \overrightarrow{n_j} = \sum_{ij} W_{ij} \delta_{3i} \overrightarrow{n_j} = \sum_{j} W_{3j} \overrightarrow{n_j} \quad (21)$$

Hence the vectorial representation of "John likes" becomes indeed the subset of all individuals who might be liked by the person denoted by vector $\overrightarrow{n_3}$, and can be seen as the semantic value of the theme

of our sentence. The next step is to compose this theme with the rheme 'Mary'; in other words, we must decide an appropriate type of composition for our intonational boundary. Let us first try again standard categorical composition:

$$(1_S \otimes \epsilon_N^l)\left(\left(\sum_j W_{3j}\overrightarrow{n_j}\right) \otimes \overrightarrow{n_1}\right) =$$

$$\sum_j W_{3j}\langle n_j|n_1\rangle = \sum_j W_{3j}\delta_{j1} = W_{31} \qquad (22)$$

Note that this corresponds to a set membership test; the result is 1 if Mary is included in the set of alternative responses and 0 otherwise. However, as noted before, in information structure terms a more appropriate operation would be to take the intersection of the singleton {Mary} with the set of alternatives. Interestingly, set intersection now corresponds to element-wise vector multiplication (in this work denoted by symbol $\odot$) and the vector space equivalent of Eq. 19 becomes:

$$\left(\sum_j W_{3j}\overrightarrow{n_j}\right) \odot \overrightarrow{n_1} = \begin{cases} \overrightarrow{n_1} & \text{if } W_{31} = 1 \\ \overrightarrow{0} & o.w. \end{cases} \qquad (23)$$

The result is now 'Mary', if Mary is included in the set of valid answers, and the zero vector otherwise. The fact that the meaning of our sentence becomes an element of the noun space demonstrates clearly that, in information structure terms, there is a necessity for a shared vector space between sentences and nouns (or noun phrases)—a direct consequence of the fact that now the meaning of a sentence is mainly focused on a specific noun or noun phrase therein. Furthermore, since a sentence is now expressed as a merging of a theme and a rheme, it is also required that $\Theta = S = P$ (and equal to what we took to be $N$ in the preceding). In the next section we encode the above reasoning in the abstract form of compact closed categories and then present an instantiation in vector spaces.

## 7  Intonation in compact closed categories with Frobenius structure

The point with regard to shared spaces is accomplished by the following types assignment:

$$\mathcal{F}(x) = W \qquad \forall x \in \{n, s, \theta, \rho\} \qquad (24)$$

As a consequence of the above, the vector spaces assigned to transitive verbs are computed as follows:

$$\mathcal{F}(n^r \cdot \theta) = \mathcal{F}(\theta \cdot n^l) = W \otimes W$$

Furthermore, boundaries are assigned to the following vector space:

$$\mathcal{F}(\theta^r \cdot s \cdot \rho^l) = \mathcal{F}(\rho^r \cdot s \cdot \theta^l) = W \otimes W \otimes W$$

We have now arrived at a central point of this paper. As the semantic representation of a boundary, we assign the following morphism:

$$(1_W \otimes \mu_W \otimes 1_W) \circ (\eta_W \otimes \eta_W) \qquad (25)$$

Note that the above is indeed an element in $W \otimes W \otimes W$:

$$\triangleright, \triangleleft : I \cong I \otimes I \xrightarrow{\eta_W \otimes \eta_W} W \otimes W \otimes W \otimes W \quad (26)$$
$$\xrightarrow{1_W \otimes \mu_W \otimes 1_W} W \otimes W \otimes W$$

The reasoning behind our assignment will become clear in a moment. For now, we proceed to a formal definition:

**Definition 7.1** *The meaning vector of a sentence expressed in information structure terms is given by:*

$$(1_W \otimes \mu_W \otimes 1_W) \circ (\eta_W \otimes \eta_W)(\overrightarrow{theme} \otimes \overrightarrow{rheme}) \quad (27)$$

*when theme precedes the rheme, or as follows in the opposite case.*

$$(1_W \otimes \mu_W \otimes 1_W) \circ (\eta_W \otimes \eta_W)(\overrightarrow{rheme} \otimes \overrightarrow{theme}) \quad (28)$$

These vectors are depicted as follows:



$$(29)$$



$$(30)$$

Note that the normal forms at the right-hand side of the diagrams above are direct applications of the Frobenius condition. Furthermore, either the theme or rheme here might correspond to large text constituents, i.e. phrases or even sentences. In this case, the proposed framework guarantees that an appropriate vector will be created for them based on categorical composition.

70

## 8 Vector space instantiation

Our justification for using the semantic form of Eq. 25 for the boundary comes from the fact that it produces normal forms as below:

$$\mu(\overrightarrow{\text{theme}} \otimes \overrightarrow{\text{rheme}}) \quad \mu(\overrightarrow{\text{rheme}} \otimes \overrightarrow{\text{theme}}) \qquad (31)$$

This is exactly how element-wise vector multiplication is defined from a categorical perspective:

$$\mu(\overrightarrow{v_1} \otimes \overrightarrow{v_2}) = \overrightarrow{v_1} \odot \overrightarrow{v_2} = \qquad (32)$$

As a result, the linear algebraic instantiations of Definition 7.1 become as follows:

$$\overrightarrow{\text{rheme}} \odot \overrightarrow{\text{theme}} \qquad \overrightarrow{\text{theme}} \odot \overrightarrow{\text{rheme}}$$

We stress again the fact that rheme and theme can have complex structures, and their vector meanings will reflect this strutter. For simple transitive sentences[6] of the form "subject verb ▷ object" or "subject ◁ verb object", we get linear algebraic meanings as follows:

$$(\overrightarrow{\text{subj}} \times \overrightarrow{\text{verb}}) \odot \overrightarrow{\text{obj}} \qquad \overrightarrow{\text{subj}} \odot (\overrightarrow{\text{verb}} \times \overrightarrow{\text{obj}})$$

As an example of a composed theme, consider:

$$(33)$$

A vector is computed for the theme 'Mary likes' according to the rules of the grammar, and then this vector is element-wise multiplied with the vector of the rheme (which, in this example, is just the distributional vector of the word).

## 9 Interpretation

The transition from the set-theoretical framework to high dimensional real vector spaces poses the question what is the role of element-wise vector multiplication in the latter setting. Compositional models based on element-wise vector addition or multiplication are usually referred to as *vector mixture*

models—a term that emphasizes on the *equal contribution* of each word to the final result, which produces a kind of average of the input vectors. Note that this behaviour stands in direct contrast with the categorical compositional approach, in which the type-logical identities of words strictly depend on their grammatical role. Due to their simplicity, vector mixture models have been studied extensively (Mitchell and Lapata, 2008), demonstrating steady and reasonably good performance in a number of tasks.

The significance of the Frobenius operators for our model (as opposed to some other form of combinatory mechanism) is that their concrete manifestation in a vector space setting imposes exactly this vector mixture behaviour, in the form of element-wise vector multiplication. In other words, the result is a combination of two compositional approaches, vector mixtures and categorical models, in a unified framework: while categorical composition is still applied to compute vectorial representations for a theme and a rheme, the two parts contribute equally to the final result via element-wise multiplication imposed by the Frobenius operators. This puts the necessary focus on the appropriate part of the sentence, reflecting the variation in meaning intended by the intonational pattern.

To what extent the notion of a rheme as a means for restricting the theme applies in **FVect**? Note that, from a geometric perspective, element-wise vector multiplication acts as a scaling of the basis; for example, $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \odot \left(\begin{smallmatrix} 2.0 \\ 0.5 \end{smallmatrix}\right)$ transforms the vector space in which the first vector lives so that the units on the $x$-axis are doubled and the units on the $y$-axis are halved.[7] Furthermore, a zero value in one vector would completely eliminate the corresponding component in the other. Hence, the concept of restricting the theme has now taken a new quantitative form, generalizing appropriately our initial intuition (motivated by set intersection) to the multi-dimensional, real-valued setting of **FVect**.

## 10 Relation to previous work

How does the above derivations correlate to the premises of the original framework, in which 'likes' is a transitive verb with type $n^r \cdot s \cdot n^l$? Note that another application of the Frobenius condition on the normal form of Diagram (33) will give us:

---

[6]We provide more complicated examples later in Sect. 11.

[7]Of course we can think of a similar scaling taking place on the two axes of the second vector by factors $x$ and $y$.

Mary likes musicals Mary likes musicals likes

$$\quad (34)$$

In other words, the semantic representation of word 'likes' can be still regarded as a bi-linear map, faithfully encoded in a tensor of order 3, as required by the framework. In this case, the tensor of 'likes' in **FVect** is seen as created by applying the morphism $1_W \otimes \Delta_W$ on a matrix representing the verb 'likes'. The limitation, of course, is that now the middle wire carrying the result (the sentence vector space) cannot be any more differentiated from the two argument wires (the noun vector spaces), since it is produced by copying one of them.

Note that these are the Frobenius models of Kartsaklis et al. (2014), referred to as Copy-Subject and Copy-Object, and originally used as a means for faithfully encoding a verb matrix to a tensor of order 3, thus restoring the functorial relation between the semantic representation and the grammatical type. The present theory[8] offers an alternative more complete account that goes far beyond providing a convenient way to expand a matrix to a cube.

## 11 Covering complex intonational patterns

So far we examined simple cases of intonation, in which our sentence consisted of a single rheme and a theme. In this section we turn our attention to some more interesting examples.

### 11.1 Multiple rhemes

We will first examine the case of a sentence with more than one rhemes. Imagine the following question/answer dialogue:

(12) Q: Who likes whom?
A: [JOHN]$_R$ [likes]$_T$ [MARY]$_R$

In our pregroup notation, this introduces two distinct intonational boundaries in the sentence. The derivation takes the following form:

$$\text{John} \quad \triangleleft \quad \text{likes} \quad \triangleright \quad \text{Mary}$$
$$\rho \quad \rho^r \; s \; \theta^l \quad \theta \; \theta \quad \theta^r \; s \; \rho^l \quad \rho \qquad (35)$$

Note that the type of 'likes' now becomes $\theta \cdot \theta$; in other words, the theme is not any more a function (no adjoint is present in the type), but a *higher order* atomic entity. This is directly reflected in **FVect** where we get:

$$\text{John} \quad \triangleleft \quad \text{likes} \quad \triangleright \quad \text{Mary} \quad \text{John likes Mary}$$

$$\quad (36)$$

The result of this computation is now a matrix and not a vector. Indeed, if we follow the linear algebraic calculations we get:

$$(\mu_W \otimes \mu_W)(\overrightarrow{\text{John}} \otimes \overline{\text{likes}} \otimes \overrightarrow{\text{mary}}) \quad = \qquad (37)$$
$$(\overrightarrow{\text{John}} \otimes \overrightarrow{\text{Mary}}) \odot \overline{\text{likes}}$$

The behaviour above follows the premises of the proposed model: Since our theme is a matrix, the calculations follow naturally, producing another matrix as the rheme (the tensor product of the two individual rhemes) that *restricts* as required the theme via element-wise multiplication. Note that this means that a sentence with one rheme would not be comparable with a sentence with two rhemes, since it would live in a different space. That is again not surprising: the shape of theme defines the shape of the sentence vector space, and only themes of the same order can be compared to each other.

### 11.2 Relational words as rhemes

We have conveniently avoided to discuss until now the case in which the rheme is not a noun phrase, but a relational word as below:

(13) Q: How does John feel about Mary?
A: [John]$_T$ [LIKES]$_R$ [Mary]$_T$

In pregroups we model such a situation by the following derivation:

$$\text{John} \quad \triangleright \quad \text{likes} \quad \triangleleft \quad \text{Mary}$$
$$\theta \quad \theta^r \; s \; \rho^l \quad \rho \; \rho \quad \rho^r \; s \; \theta^l \quad \theta \qquad (38)$$

Note that this time the verb becomes a higher order rheme, getting the type $\rho \cdot \rho$. However, when this is transferred to **FVect** the symmetry of the category and the commutativity of the Frobenius algebra means that the vector of the sentence becomes equal to that of Example (12). In general, problems due to commutativity of the Frobenius operators can be resolved if one moves to non-commutative versions

---

[8]An early account of which also appears in the doctoral thesis of the first author (Kartsaklis, 2015).

of Frobenius algebras. Piedeleu et al. (2015) explore such constructions in the context of language by elevating the categorical model of Coecke et al. (2010) to an open quantum system setting, in which words are represented as mixed states.

## 11.3 Nested rhemes

Consider the following case:

(14)   What was the book Mary wrote about?
       [Mary wrote a book about]$_\textsf{T}$ [ART]$_\textsf{R}$

The interesting point here is that the intonational boundary is placed in a position that constitutes a glaring violation of the grammatical structure, which in the normal case has the following form:

$$\begin{array}{cccccc} \text{Mary} & \text{wrote} & \text{a book} & \text{about} & \text{art} \\ n & n^r\,sn^l & n & n^l n n^r & n \end{array} \qquad (39)$$

For cases like these we should recall that our framework is entirely built on the assumption of Postulate 4.1: in the context of information structure, intonational boundaries *determine* the intended syntactical structure. For our case, we get:

$$\begin{array}{ccccccc} \text{Mary} & \text{wrote} & \text{a book} & \text{about} & \rhd & \text{art} \\ n & n^r\,\theta\,n^l & n & n^r n & \theta^r\,s\,\rho^l & \rho \end{array} \qquad (40)$$

The linear-algebraic result follows trivially as the usual element-wise composition of the theme with the rheme.

## 11.4 Rheme in the middle of sentence

In many cases a noun phrase can serve as the rheme while being placed in the middle of the sentence, splitting the theme into two parts:

(15)   Did Mary write an essay about art?
       [Mary wrote]$_\textsf{T}$ [A BOOK]$_\textsf{R}$ [about art]$_\textsf{T}$

In these cases, the left-hand intonational boundary gets the type $\theta^r \cdot \rho \cdot \rho^l$, as below:

$$\begin{array}{cccccccc} \text{Mary} & \text{wrote} & \rhd & \text{a book} & \lhd & \text{about} & \text{art} \\ n & n^r\,\theta & \theta^r\,\rho\,\rho^l & \rho & \rho^r\,s\,\theta^l & \theta\,n^r & n \end{array} \qquad (41)$$

In other words, a new rheme is produced that is used as input to the right-hand intonational boundary. In **Fvect** we get the following interaction:

$$\begin{array}{cccccc} \text{Mary} & \text{wrote} & \text{a book} & \text{about} & \text{art} \end{array} \qquad (42)$$

By application of the spider equality (Section 2.6) we get the normal form below, which computes a meaning for the sentence as the element-wise multiplication of the vectors composed for the two themes with the vector of the rheme:

$$\begin{array}{cccccc} \text{Mary} & \text{wrote} & \text{a book} & \text{about} & \text{art} \end{array} \qquad (43)$$

## 12   Conclusion and future work

The present paper provides a first account of intonation and information structure for the emerging field of categorical compositional distributional semantics. In a more generic level, it lays the groundwork for a model capable of accommodating two different types of composition over a distributional setting. An experimental evaluation is deferred for the future, preferably in the context of a question-answering task. There is also a lot of interesting work to be done on the theory side. At the current stage, for example, the semantic value of intonational boundaries is given by direct assignment of a specific morphism—a common practice in the past for the relevant literature. A future direction, then, more aligned with the categorical nature of the model, would be to embed the appropriate translation into the functorial passage itself. This challenging goal requires novel theoretical contributions that will elevate the concept of a pregroup grammar to a new entity equipped with Frobenius structure.

Finally, the categorical compositional model of Piedeleu et al. (2015) is very relevant to our interests, since it can accommodate a variety of non-commutative Frobenius algebras the linguistic intuition of which in relation to this work remains to be explored.

## References

Wallace L. Chafe. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li, editor, *Subject and Topic*. Academic Press, New York.

B. Coecke, M. Sadrzadeh, and S. Clark. 2010. Mathematical Foundations for a Compositional Distributional Model of Meaning. Lambek Festschrift. *Linguistic Analysis*, 36:345–384.

Caroline Féry and Manfred Krifka. 2008. Information Structure: Notional distinctions, ways of expression. In Piet van Sterkenburg, editor, *Unity and diversity of languages*, pages 123–135. John Benjamins Publishing.

Ines Fiedler, Katharina Hartmann, Brigitte Reineke, Anne Schwarz, and Malte Zimmermann. 2006. Subject focus in West African languages. In *International Conference of Information Structure*, Potsdam.

Z. Harris. 1968. *Mathematical Structures of Language*. Wiley.

Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Stephen Pulman. 2012. A unified sentence space for categorical distributional-compositional semantics: Theory and experiments. In *Proceedings of 24th International Conference on Computational Linguistics (COLING 2012): Posters*, pages 549–558, Mumbai, India, December. The COLING 2012 Organizing Committee.

Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, Stephen Pulman, and Bob Coecke. 2014. Reasoning about meaning in natural language with compact closed categories and Frobenius algebras. *arXiv preprint arXiv:1401.5980*.

Dimitri Kartsaklis. 2015. *Compositional Distributional Semantics with Compact Closed Categories and Frobenius Algebras*. Ph.D. thesis, University of Oxford.

G Maxwell Kelly. 1972. Many-variable functorial calculus (I). In G.M. Kelly, M. Laplaza, G. Lewis, and S. MacLane, editors, *Coherence in categories*, pages 66–105. Springer.

J. Lambek. 2008. *From Word to Sentence*. Polimetrica, Milan.

T. Landauer and S. Dumais. 1997. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquision, Induction, and Representation of Knowledge. *Psychological Review*.

Mike Lewis and Mark Steedman. 2013. Combined distributional and logical semantics. *Transactions of the Association for Computational Linguistics*, 1:179–192.

Saunders Mac Lane. 1971. *Categories for the Working Mathematician*. Number 5 in Graduate Texts in Mathematics. Springer-Verlag.

C.D. Manning, P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press.

J. Mitchell and M. Lapata. 2008. Vector-based Models of Semantic Composition. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pages 236–244.

Richard Montague. 1970. English as a formal language. In *Linguaggi nella Società e nella Tecnica*, pages 189–224. Edizioni di Comunità, Milan.

Robin Piedeleu, Dimitri Kartsaklis, Bob Coecke, and Mehrnoosh Sadrzadeh. 2015. Open System Categorical Quantum Semantics in Natural Language Processing. In *Proceedings of the 6th Conference on Algebra and Coalgebra in Computer Science (CALCO)*, Nijmegen, Netherlands, June.

Janet Pierrehumbert and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in communication*, pages 271–312. MIT Press, Cambridge MA.

Mats Rooth. 1992. A theory of focus interpretation. *Natural language semantics*, 1(1):75–116.

M. Sadrzadeh, S. Clark, and B. Coecke. 2013. The Frobenius anatomy of word meanings I: subject and object relative pronouns. *Journal of Logic and Computation*, Advance Access, October.

M. Sadrzadeh, S. Clark, and B. Coecke. 2014. The Frobenius anatomy of word meanings II: Possessive relative pronouns. *Journal of Logic and Computation*, June.

H. Schütze. 1998. Automatic Word Sense Discrimination. *Computational Linguistics*, 24:97–123.

Peter Selinger. 2011. A survey of graphical languages for monoidal categories. In Bob Coecke, editor, *New structures for physics*, pages 289–355. Springer.

Mark Steedman. 2000. Information structure and the syntax-phonology interface. *Linguistic inquiry*, 31(4):649–689.

Mark Steedman. 2001. *The Syntactic Process*. MIT Press.

Peter D Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1):141–188.