Categorical Vector Space Semantics for Lambek Calculus with a Relevant Modality (Extended Abstract)

Lachlan McPheat Mehrnoosh Sadrzadeh
University College London
London, UK
{m.sadrzadeh,l.mcpheat}@ucl.ac.uk

Hadi Wazni
Queen Mary University London
London, UK
hadi.wazni@hotmail.com

Gijs Wijnholds
Utrecht University
Utrecht, NL
g.j.wijnholds@uu.nl

We develop a categorical compositional distributional semantics for Lambek Calculus with a Relevant Modality, $!\mathbf{L}^*$, which has a limited version of the contraction and permutation rules. The categorical part of the semantics is a monoidal biclosed category with a coalgebra modality as defined on Differential Categories. We instantiate this category to finite dimensional vector spaces and linear maps via "quantisation" functors and work with three concrete interpretations of the coalgebra modality. We apply the model to construct categorical and concrete semantic interpretations for the motivating example of $!\mathbf{L}^*$: the derivation of a phrase with a parasitic gap. The effectiveness of the concrete interpretations are evaluated via a disambiguation task, on an extension of a sentence disambiguation dataset to parasitic gap phrases, using BERT, Word2Vec, and FastText vectors and Relational tensors.

1 Introduction

Distributional Semantics of natural language are semantics which model the *Distributional Hypothesis* due to Firth [11] and Harris [18] which assumes *a word is characterized by the company it keeps*. Research in Natural Language Processing (NLP) has turned to Vector Space Models (VSMs) of natural language to accurately model the distributional hypothesis. Such models date as far back as to Rubinstein and Goodenough's co-occurence matrices [35] in 1965, until today's neural machine learning methods, leading to embeddings, such as Word2Vec [40], GloVe [32], FastText [6] or BERT [10] to name a few. VSMs were used even earlier by Salton [38] for information retrieval. These models have plenty of applications, for instance thesaurus extraction tasks [9, 17], automated essay marking [23] and semantically guided information retrieval [24]. However, they lack grammatical compositionality, thus making it difficult to sensibly reason about the semantics of portions of language larger than words, such as phrases and sentences.

Somewhat orthogonally, Type Logical Grammars (TLGs) form highly compositional models of language by accurately modelling grammar, however they lack distributionality, in that such models do not accurately describe the distributional semantics of a word, only its grammatical role. Distributional Compositional Categorical Semantics (DisCoCat)[8] combines these two approaches using category theoretic methods, originally developed to model Quantum protocols. DisCoCat has proven its efficacy empirically [15, 16, 37, 43, 21, 27] and has the added utility of being a modular framework which is open to

David I. Spivak and Jamie Vicary (Eds.):
Applied Category Theory 2020 (ACT2020)
EPTCS 333, 2021, pp. 168–182, doi:10.4204/EPTCS.333.12

additions and extensions.

DisCoCat is a categorical semantics of a formal system which models natural language syntax, known as Lambek Calculus[1], denoted by **L**. The work in [20] extends Lambek calculus with a relevant modality, and denotes the resulting logic by $!\mathbf{L}^*$. As an example application domain, they use the new logic to formalise the grammatical structure of the parasitic gap phenomena in natural language.

In this paper, we first form a sound categorical semantics of $!\mathbf{L}^*$, which we call $\mathcal{C}(!\mathbf{L}^*)$. This boils down to interpreting the logical contraction of $!\mathbf{L}^*$ using comonads known as *coalgebra modalities* defined in [4]. In order to facilitate the categorical computations, we use the clasp-string calculus of [2], developed for depicting the computations of a monoidal biclosed category. To this monoidal diagrammatic diagrammatic calculus, we add the necessary new constructions for the coalgebra modality and its operations. Next, we define three candidate coalgebra modalities on the category of finite dimensional real vector spaces in order to form a sound VSM of $!\mathbf{L}^*$ in terms of structure-preserving functors $\mathcal{C}(!\mathbf{L}^*) \to \mathbf{FdVect}_{\mathbb{R}}$. We also briefly introduce a prospective diagrammatic semantics of $\mathcal{C}(!\mathbf{L}^*)$ to help visualise our derivations. We conclude this paper with an experiment to test the accuracy of the different coalgebra modalitites on $\mathbf{FdVect}_{\mathbb{R}}$. The experiment is performed using different neural word embeddings and on a disambiguation task over an extended version of dataset of [16] from transitive sentences to phrases with parasitic gaps.

This paper is an extended abstract of the full arXiv paper [25].

2 $!\mathbf{L}^*$: Lambek Calculus with a Relevant Modality

Following [20], we assume that the formulae, or types, of Lambek calculus with a Relevant Modality $!\mathbf{L}^*$ are generated by a set of atomic types At, a unary connective !, three binary connectives, $\backslash$, / and , via the following Backus-Naur Form (BNF).

$$\varphi ::= \varphi \in At \mid \emptyset \mid (\varphi,\varphi) \mid (\varphi/\varphi) \mid (\varphi\backslash\varphi) \mid !\varphi,$$

We refer to the types of $!\mathbf{L}^*$ by $\mathrm{Typ}_{!\mathbf{L}^*}$; here, $\emptyset$ denotes the empty type. An element of $\mathrm{Typ}_{!\mathbf{L}^*}$ is either atomic, made up of a modal type, or two types joined by a comma or a slash. We will use uppercase roman letters to denote arbitrary types of $!\mathbf{L}^*$, and uppercase Greek letters to denote a set of types, for example, $\Gamma = \{A_1,A_2,\ldots,A_n\} = A_1,A_2,\ldots,A_n$. It is assumed that , is associative, allowing us to omit brackets in expressions like $A_1,A_2,\ldots,A_n$.

A **sequent** of $!\mathbf{L}^*$ is a pair of an ordered set of types and a type, denoted by $\Gamma \vdash A$. The derivations of $!\mathbf{L}^*$ are generated by the set of axioms and rules presented in table 1. The logic $!\mathbf{L}^*$ extends Lambek Calculus **L** by endowing it with a modality denoted by !, inspired by the ! modality of Linear Logic, to enable the structure rule of contraction in a controlled way, although here it is introduced on a non-symmetric monoidal category but is introduced with an extra structure allowing the !-ed types to commute over other types. So what $!\mathbf{L}^*$ adds to **L** is the $(!L),(!R)$ rules, the (perm) rules, and the (contr) rule.

[1]There is a parallel pregroup syntax which gives you the same semantics, as discussed in [5]

$$\frac{}{A \vdash A}$$

$$\frac{\Gamma \vdash A \quad \Delta_1, B, \Delta_2 \vdash C}{\Delta_1, B/A, \Gamma, \Delta_2 \vdash C}\,(/L) \qquad \frac{\Gamma, A \vdash B}{\Gamma \vdash B/A}\,(/R) \qquad \frac{\Gamma \vdash A \quad \Delta_1, B, \Delta_2 \vdash C}{\Delta_1, \Gamma, A\backslash B, \Delta_2 \vdash C}\,(\backslash L) \qquad \frac{A, \Gamma \vdash B}{\Gamma \vdash A\backslash B}\,(\backslash R)$$

$$\frac{\Gamma_1, A, \Gamma_2 \vdash C}{\Gamma_1, !A, \Gamma_2 \vdash C}\,(!L) \qquad \frac{!A_1, \ldots, !A_n \vdash B}{!A_1, \ldots, !A_n \vdash !B}\,(!R) \qquad \frac{\Delta_1, !A, \Gamma, \Delta_2 \vdash C}{\Delta_1, \Gamma, !A, \Delta_2 \vdash C}\,(\mathrm{perm}_1) \qquad \frac{\Delta_1, \Gamma, !A, \Delta_2 \vdash C}{\Delta_1, !A, \Gamma, \Delta_2 \vdash C}\,(\mathrm{perm}_2)$$

$$\frac{\Delta_1, !A, !A, \Delta_2 \vdash C}{\Delta_1, !A, \Delta_2 \vdash C}\,(\mathrm{contr})$$

Table 1: Rules of $!\mathbf{L}^*$.

3 Categorical Semantics for $!\mathbf{L}^*$

We associate $!\mathbf{L}^*$ with a category $\mathcal{C}(!\mathbf{L}^*)$, with $\mathrm{Typ}_{!\mathbf{L}^*}$ as objects, and derivable sequents of $!\mathbf{L}^*$ as morphisms whose domains are the formulae on the left of the turnstile and codomains the formulae on the right. The category $\mathcal{C}(!\mathbf{L}^*)$ is monoidal biclosed[2]. The connectives , and $\backslash, /$ in $!\mathbf{L}^*$ are associated with the monoidal structure on $\mathcal{C}(!\mathbf{L}^*)$, where , is the monoidal product, with the empty type as its unit and $\backslash, /$ are associated with the two internal hom functors with respect to ,, as presented in [39]. The connective ! of $!\mathbf{L}^*$ is a *coalgebra* modality, as defined for *Differential Categories* in [4], with the difference that our underlying category is not necessarily symmetric monoidal, but we ask for a restricted symmetry with regards to ! and that ! be a lax monoial functor. In Differential Categories ! does not necessarily have a monoidal property, i.e. it is not a strict, lax, or strong monoidal functor, but there are examples of Differential Categories where strong monoidality holds. We elaborate on these notions via the following definition.

Definition 1. *The category $\mathcal{C}(!\mathbf{L}^*)$ has types of $!\mathbf{L}^*$, i.e. elements of $\mathrm{Typ}_{\mathbf{L}}$, as objects, derivable sequents of $!\mathbf{L}^*$ as morphisms, together with the following structures:*

- *A **monoidal product** $\otimes\colon \mathcal{C}(!\mathbf{L}^*) \times \mathcal{C}(!\mathbf{L}^*) \to \mathcal{C}(!\mathbf{L}^*)$, with a unit I.*
- ***Internal hom-functors** $\Rightarrow\colon \mathcal{C}(!\mathbf{L}^*)^{\mathrm{op}} \times \mathcal{C}(!\mathbf{L}^*) \to \mathcal{C}(!\mathbf{L}^*)$, $\Leftarrow\colon \mathcal{C}(!\mathbf{L}^*) \times \mathcal{C}(!\mathbf{L}^*)^{\mathrm{op}} \to \mathcal{C}(!\mathbf{L}^*)$ such that:*
  - *i. For objects $A, B \in \mathcal{C}(!\mathbf{L}^*)$, we have objects $(A \Rightarrow B), (A \Leftarrow B) \in \mathcal{C}(!\mathbf{L}^*)$ and a pair of morphisms, called* right and left evaluation*, given below:*

$$\mathrm{ev}^r_{A,(A\Rightarrow B)}\colon A \otimes (A \Rightarrow B) \longrightarrow A, \qquad \mathrm{ev}^l_{(A\Leftarrow B),B}\colon (A \Leftarrow B) \otimes B \longrightarrow A$$

  - *ii. For morphisms $f\colon A \otimes C \longrightarrow B, g\colon C \otimes B \longrightarrow A$, we have unique* right and left curried *morphisms, given below:*

$$\Lambda^l(f)\colon C \longrightarrow (A \Rightarrow B), \qquad \Lambda^r(g)\colon C \longrightarrow (A \Leftarrow B)$$

  - *iii. The following hold*

$$\mathrm{ev}^l_{A,B} \circ (\mathrm{id}_A \otimes \Lambda^l(f)) = f, \qquad \mathrm{ev}^r_{A,B} \circ (\Lambda^r(g) \otimes \mathrm{id}_B) = g$$

[2] We follow the convention that products are not symmetric unless stated, hence a monoidal product is not symmetric unless referred to by 'symmetric monoidal'.

- *A* **coalgebra modality** ! *on* $\mathcal{C}(!\mathbf{L}^*)$*. That is, a lax monoidal comonad* $(!, \delta, \varepsilon)$ *such that:*

  For every object $A \in \mathcal{C}(!\mathbf{L}^*)$*, the object* $!A$ *has a comonoid structure* $(!A, \Delta_A, e_A)$ *in* $\mathcal{C}(!\mathbf{L}^*)$*. Where the comultiplication* $\Delta_A : !A \to !A \otimes !A$*, and the counit* $e_A : !A \to I$ *satisfy the usual comonoid equations. Further, we require* $\delta_A : !A \to !!A$ *to be a morphism of comonoids [4].* [3]

- ***Restricted symmetry*** *over the coalgebra modality, that is, natural isomorphisms* $\sigma^r : 1_{\mathcal{C}(!\mathbf{L}^*)} \otimes ! \to ! \otimes 1_{\mathcal{C}(!\mathbf{L}^*)}$ *and* $\sigma^l : ! \otimes 1_{\mathcal{C}(!\mathbf{L}^*)} \to 1_{\mathcal{C}(!\mathbf{L}^*)} \otimes !$.

$$\sigma^r_{A,B} : A \otimes !B \longmapsto !B \otimes A, \qquad \sigma^l_{A,B} : !A \otimes B \longmapsto B \otimes !A.$$

We now define a categorical semantics for $!\mathbf{L}^*$ as the map $[\![\ ]\!] : !\mathbf{L}^* \to \mathcal{C}(!\mathbf{L}^*)$ and prove that it is sound.

Definition 2. *The* semantics *of formulae and sequents of* $!\mathbf{L}^*$ *is the image of the interpretation map* $[\![\ ]\!] : !\mathbf{L}^* \to \mathcal{C}(!\mathbf{L}^*)$*. To elements* φ *in* $\mathrm{Typ}_{\mathbf{L}}$*, this map assigns objects* C_φ *of* $\mathcal{C}(!\mathbf{L}^*)$*, as defined below:*

$$\begin{array}{lcl@{\qquad\qquad}lcl}
[\![\emptyset]\!] & := & C_\emptyset = I & [\![\varphi]\!] & := & C_\varphi \\
[\![(\varphi,\varphi)]\!] & := & C_\varphi \otimes C_\otimes & [\![!\varphi]\!] & := & !C_\varphi \\
[\![(\varphi/\varphi)]\!] & := & (C_\varphi \Leftarrow C_\varphi) & [\![(\varphi\backslash\varphi)]\!] & := & (C_\varphi \Rightarrow C_\varphi)
\end{array}$$

To the sequents $\Gamma \vdash A$ *of* $!\mathbf{L}^*$*, for* $\Gamma = \{A_1, A_2, \cdots A_n\}$ *where* $A_i, A \in \mathrm{Typ}_{\mathbf{L}}$*, it assigns morphism of* $\mathcal{C}(!\mathbf{L}^*)$ *as follows* $[\![\Gamma \vdash A]\!] := C_\Gamma \longrightarrow C_A$*, for* $C_\Gamma = [\![A_1]\!] \otimes [\![A_2]\!] \otimes \cdots \otimes [\![A_n]\!]$.

Since sequents are not labelled, we have no obvious name for the linear map $[\![\Gamma \vdash A]\!]$, so we will label such morphisms by lower case roman letters as needed.

Definition 3. *A* ***categorical model*** *for* $!\mathbf{L}^*$*, or a* $!\mathbf{L}^*$***-model****, is a pair* $(\mathcal{C}, [\![\ ]\!]_{\mathcal{C}})$*, where* $\mathcal{C}$ *is a monoidal biclosed category with a coalgebra modality and restricted symmetry, and* $[\![\ ]\!]_{\mathcal{C}}$ *is a mapping* $\mathrm{Typ}_{!\mathbf{L}^*} \to \mathcal{C}$ *factoring through* $[\![\ ]\!] : \mathrm{Typ}_{\mathbf{L}} \to \mathcal{C}(!\mathbf{L}^*)$.

Definition 4. *A sequent* $\Gamma \vdash A$ *of* $!\mathbf{L}^*$ *is sound in* $(\mathcal{C}(!\mathbf{L}^*), [\![\ ]\!])$*, iff* $C_\Gamma \longrightarrow C_A$ *is a morphism of* $\mathcal{C}(!\mathbf{L}^*)$*. A rule* $\frac{\Gamma \vdash A}{\Delta \vdash B}$ *of* $!\mathbf{L}^*$ *is* sound *in* $(\mathcal{C}(!\mathbf{L}^*), [\![\ ]\!])$ *iff whenever* $C_\Gamma \longrightarrow C_A$ *is sound then so is* $C_\Delta \longrightarrow C_B$*. We say* $!\mathbf{L}^*$ *is sound with regards to* $(\mathcal{C}(!\mathbf{L}^*), [\![\ ]\!])$ *iff all of its rule are.*

Theorem 1. $!\mathbf{L}^*$ *is sound with regards to* $(\mathcal{C}(!\mathbf{L}^*), [\![\ ]\!])$.

Proof. See full paper [25]. □

4 Vector Space Semantics for $\mathcal{C}(!\mathbf{L}^*)$

Following [5], we develop vector space semantics for $!\mathbf{L}^*$, via a *quantisation* functor to the category of finite dimensional vector spaces and linear maps $F : \mathcal{C}(!\mathbf{L}^*) \to \mathbf{FdVect}_{\mathbb{R}}$. This functor interprets objects as finite dimensional vector spaces, and derivations as linear maps. Quantisation is the term first introduced by Atiyah in Topological Quantum Field Theory, as a functor from the category of manifolds and cobordisms to the category of vector spaces and linear maps. Since the cobordism category is monoidal, quantisation was later generalised to refer to a functor that 'quantises' any category in $\mathbf{FdVect}_{\mathbb{R}}$. Since $\mathcal{C}(!\mathbf{L}^*)$ is free, there is a unique functor $\mathcal{C}(!\mathbf{L}^*) \to (\mathbf{FdVect}_{\mathbb{R}}, !)$ for any choice of ! such that $(\mathbf{FdVect}_{\mathbb{R}}, !)$ is a $!\mathbf{L}^*$-model. In definition 5 we introduce the necessary nomenclature to define quantisations in full.

[3] Strictly speaking, this definition applies to symmetric monoidal categories, however we may abuse notation without worrying, as we have symmetry in the image of ! coming from the restricted symmetries σ^l, σ^r.

Definition 5. *A* ***quantisation*** *is a functor* $F : \mathcal{C}(!\mathbf{L}^*) \to (\mathbf{FdVect}_{\mathbb{R}}, !)$*, defined on the objects of* $\mathcal{C}(!\mathbf{L}^*)$ *using the structure of the formulae of* $!\mathbf{L}^*$*, as follows:*

$$\begin{aligned} F(C_\emptyset) &:= \mathbb{R} & F(C_\varphi) &:= V_\varphi \\ F(C_{\varphi\otimes\varphi}) &:= V_\varphi \otimes V_\varphi & F(C_{!\varphi}) &:= !V_\varphi \\ F(C_{\varphi/\varphi}) &:= (V_\varphi \Leftarrow V_\varphi) & F(C_{\varphi\backslash\varphi}) &:= (V_\varphi \Rightarrow V_\varphi) \end{aligned}$$

Here, V_φ *is the vector space in which vectors of words with an atomic type live and the other vector spaces are obtained from it by induction on the structure of the formulae they correspond to. Morphisms of* $\mathcal{C}(!\mathbf{L}^*)$ *are of the form* $C_\Gamma \longrightarrow C_A$*, associated with sequents* $\Gamma \vdash A$ *of* $!\mathbf{L}^*$*, for* $\Gamma = \{A_1, A_2, \cdots, A_n\}$*. The quantisation functor is defined on these morphisms as follows:*

$$F(C_\Gamma \longrightarrow C_A) := F(C_\Gamma) \longrightarrow F(C_A) = V_{A_1} \otimes V_{A_2} \otimes \cdots \otimes V_{A_n} \longrightarrow V_A$$

Note that the monoidal product in $\mathbf{FdVect}_{\mathbb{R}}$ is symmetric, so there is formally no need to distinguish between $([\![A]\!] \Rightarrow [\![B]\!])$ and $([\![B]\!] \Leftarrow [\![A]\!])$. However it may be practical to do so when calculating things by hand, for example when retracing derivations in the semantics. We should also make clear that the freeness of $\mathcal{C}(!\mathbf{L}^*)$ makes F a *strict* monoidal closed functor, meaning that $F(C_A \otimes C_B) = FC_A \otimes FC_B$, or rather, $V_{(A\otimes B)} = (V_A \otimes V_B)$, and similarly, $V_{(A\Rightarrow B)} = (V_A \Rightarrow V_B)$ etc. Further, since we are working with finite dimensional vector spaces we know that $V_\varphi^\perp \cong V_\varphi$, thus our internal homs have an even simpler structure, which we exploit when computing, which is $V_\varphi \Rightarrow V_\varphi \cong V_\varphi \otimes V_\varphi$.

5 Concrete Constructions

In this section we present three different coalgebra modalities on $\mathbf{FdVect}_{\mathbb{R}}$ defined over two different underlying comonads, treated in individual subsections. Defining these modalities lets us reason about sound vector space semantics of $\mathcal{C}(!\mathbf{L}^*)$ in terms of !-preserving monoidal biclosed functors $\mathcal{C}(!\mathbf{L}^*) \to \mathbf{FdVect}_{\mathbb{R}}$.

We point out here that we do not aim for *complete* model in that we do not require the tensor of our vector space semantics to be non-symmetric. This is common practice in the DisCoCat line of research and also in the standard set theoretic semantics of Lambek calculus [41]. Consider the English sentence ”John likes Mary” and the Farsi sentence ”John Mary-ra Doost-darad(likes)”. These two sentences have the same semantics, but different word orders, thus exemplifying the lack of syntax within semantics.

5.1 ! as the Dual of a Free Algebra Functor

Following [34] we interpret ! using the Fermionic Fock space functor $\mathcal{F} : \mathbf{FdVect}_{\mathbb{R}} \to \mathbf{Alg}_{\mathbb{R}}$. In order to define $\mathcal{F}$ we first introduce the simpler free algebra construction, typically studied in the theory of representations of Lie algebras [19]. It turns out that $\mathcal{F}$ is itself a free functor, giving us a comonad structure on $U\mathcal{F}$ upon dualising [34]. The choice of the symbol $\mathcal{F}$ comes from “Fermionic Fock space” (as opposed to “Bosonic”), and is also known as the exterior algebra functor, or the Grassmannian algebra functor [19].

Definition 6. *The* **free algebra functor** $T : \mathbf{Vect}_{\mathbb{R}} \to \mathbf{Alg}_{\mathbb{R}}$ *is defined on objects as:*

$$V \longmapsto \bigoplus_{l\geq 0} V^{\otimes n} = \mathbb{R} \oplus V \oplus (V \otimes V) \oplus (V \otimes V \otimes V) \ \otimes \cdots$$

and for morphisms $f : V \to W$, we get the algebra homomorphism $T(f) : T(V) \to T(W)$ defined layer-wise as

$$T(f)(v_1 \otimes v_2 \otimes \cdots \otimes v_n) := f(v_1) \otimes f(v_2) \otimes \cdots \otimes f(v_n).$$

T is free in the sense that it is left adjoint to the forgetful functor $U : \mathbf{Alg}_{\mathbb{R}} \to \mathbf{Vect}_{\mathbb{R}}$, thus giving us a monad UT on $\mathbf{Vect}_{\mathbb{R}}$ with a monoidal algebra modality structure, i.e. the dual of what we are looking for. However note that even when restricting T to finite dimensional vector spaces $V \in \mathbf{FdVect}_{\mathbb{R}}$ the resulting $UT(V)$ and $UT(V^{\perp})^{\perp}$ are infinite-dimensional. The necessity of working in $\mathbf{FdVect}_{\mathbb{R}}$ motivates us to use $\mathcal{F}$, defined below, rather than T.

Definition 7. *The* **Fermionic Fock space functor** $\mathcal{F} : \mathbf{Vect}_{\mathbb{R}} \to \mathbf{Alg}_{\mathbb{R}}$[4] *is defined on objects as*

$$V \mapsto \bigoplus_{n \geq 0} V^{\wedge n} = \mathbb{R} \oplus V \oplus (V \wedge V) \oplus (V \wedge V \wedge V) \otimes \cdots$$

where $V^{\wedge n}$ is the coequaliser of the family of maps $(-\tau_{\sigma})_{\sigma \in S_n}$, defined as $-\tau_{\sigma} : V^{\otimes n} \to V^{\otimes n}$ and given as follows:

$$(-\tau_{\sigma})(v_1 \otimes \cdots \otimes v_n) := \mathrm{sgn}(\sigma)(v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)}).$$

$\mathcal{F}$ applied to linear maps gives an analogous algebra homomorphism as in 6.

Concretely, one may define $V^{\wedge n}$ to be the n-fold tensor product of V where we quotient by the layer-wise equivalence relations $v_1 \otimes v_2 \otimes \cdots \otimes v_n \sim \mathrm{sgn}(\sigma)(v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)})$ for $n = 0, 1, 2 \ldots$ and denoting the equivalence class of a vector $v_1 \otimes v_2 \otimes \cdots \otimes v_n$ by $v_1 \wedge v_2 \wedge \cdots \wedge v_n$.

Note that simple tensors in $V^{\wedge n}$ with repeated vectors are zero. That is, if $v_i = v_j$ for some $1 \leq i, j \leq n$ and $i \neq j$ in the above, the permutation $(ij) \in S_n$ has odd sign, and so $v_1 \wedge v_2 \wedge \cdots \wedge v_n = 0$, since $v_1 \wedge v_2 \wedge \cdots \wedge v_n = \mathrm{sgn}(ij)(v_1 \wedge v_2 \wedge \cdots \wedge v_n) = -(v_1 \wedge v_2 \wedge \cdots \wedge v_n)$.

Remark 1. *Given a finite dimensional vector space V, the antisymmetric algebra $\mathcal{F}(V)$ is also finite dimensional. This follows immediately from the note in definition 7, as basis vectors in layers of $\mathcal{F}(V)$ above the* $\dim(V)$*-th are forced to repeat entries.*

Remark 1 shows that restricting $\mathcal{F}$ to finite dimensional vector spaces turns $U\mathcal{F}$ into an endofunctor on $\mathbf{FdVect}_{\mathbb{R}}$. We note that $\mathcal{F}$ is the free antisymmetric algebra functor [34] and conclude that $U\mathcal{F}$ is a monad $(U\mathcal{F}, \mu, \eta)$ on $\mathbf{FdVect}_{\mathbb{R}}$.

Given $\mathcal{F}$, there are two ways to obtain a comonad structure $(U\mathcal{F}(V), \Delta_V, e_V)$, thus define a coalgebra modality $(U\mathcal{F}, \delta, \varepsilon)$ on $\mathbf{FdVect}_{\mathbb{R}}$, as desired. One is referred to by *Cogebra* construction and is given below, for a basis $\{e_i\}_i$ of V, and thus a basis $\{1, e_{i_1}, e_{i_2} \wedge e_{i_3}, e_{i_4} \wedge e_{i_5} \wedge e_{i_6}, \cdots\}_{i_j}$ of $U\mathcal{F}(V)$ as:

$$\Delta(1, e_{i_1}, e_{i_2} \wedge e_{i_3}, e_{i_4} \wedge e_{i_5} \wedge e_{i_6}, \cdots) = (1, e_{i_1}, e_{i_2} \wedge e_{i_3}, e_{i_4} \wedge e_{i_5} \wedge e_{i_6}, \cdots) \otimes (1, e_{i_1}, e_{i_2} \wedge e_{i_3}, e_{i_4} \wedge e_{i_5} \wedge e_{i_6}, \cdots)$$

The map $e_V : U\mathcal{F}(V) \to V$ is given by projection onto the first layer, that is

$$1, e_{i_1}, e_{i_2} \wedge e_{i_3}, e_{i_4} \wedge e_{i_5} \wedge e_{i_6}, \cdots \longmapsto e_{i_1}.$$

Another coalgebra modality arises from dualising the monad $U\mathcal{F}$, and the monoid structure on $\mathcal{F}(V)$, or strictly speaking on $U\mathcal{F}(V)$. Following [7], we dualise $U\mathcal{F}$ to define a comonad structure on $U\mathcal{F}$ as follows. We take the comonad comultiplication to be $\delta_V := \mu_V^{\perp} : U\mathcal{F}U\mathcal{F}(V)^{\perp} \to U\mathcal{F}(V)^{\perp}$, and the comonad counit to be $\varepsilon_V := \eta_V^{\perp} : U\mathcal{F}(V)^{\perp} \to V^{\perp}$. To avoid working with dual spaces one may chose to

[4]One may wish to think of $\mathcal{F}$ as having codomain $\mathbf{Aalg}_{\mathbb{R}}$, the category of antisymmetric $\mathbb{R}$-algebras, which is itself a subcategory $\mathbf{Aalg}_{\mathbb{R}} \hookrightarrow \mathbf{Alg}_{\mathbb{R}}$.

formally consider $!(V) := U\mathcal{F}(V^\perp)^\perp$, as in [34], since $U\mathcal{F}(V^\perp)^\perp \cong U\mathcal{F}(V)$ (although this is not strictly necessary, we choose this notation to stay close to its original usage [34, 7]). Note that a dualising in this manner only makes sense for finite dimensional vector spaces, as in general, for an arbitrary family of vector spaces $(V_i)_{i\in I}$, we have $(\bigoplus_{i\in I} V_i)^\perp \cong \prod_{i\in I}(V_i^\perp)$. Finite dimensionality of a vector space V makes the direct sum in $U\mathcal{F}(V)$ finite, making the right-hand product a direct sum, i.e. for a finite index I we have $(\bigoplus_{i\in I} V_i)^\perp \cong \bigoplus_{i\in I}(V_{i\in I}^\perp)$, meaning that we indeed have $U\mathcal{F}(V)^\perp \cong U\mathcal{F}(V^\perp)$. This lets us dualise the monoid structure of $U\mathcal{F}(V)$, giving a comonoid structure on $U\mathcal{F}(V)$ hence making $U\mathcal{F}$ into a coalgebra modality. To compute the comultiplication it suffices to transpose the matrix for the multiplication on $U\mathcal{F}(V)$. However, this is in general intractable, as for V an n dimensional space, $U\mathcal{F}(V)$ will have 2^n dimensions, and its multiplication will be a $(2^n)^2 \times 2^n$-matrix. We leave working with a dualised comultiplication to another paper, but in the next subsection use this construction to obtain a richer copying than the Cogebra one mentioned above.

5.2 ! as the Identity Functor

The above Cogebra construction can be simplified when one works with free vector spaces, for details of which we refer to the full version of the paper [25]. The simplified version resembles half of a bialgebra over $\mathbf{FdVect}_\mathbb{R}$, known as *Special Frobenius bialgebras*, which were used in [36, 28, 26] to model relative pronouns in English and Dutch. As argued in [42], however, the copying map resulting from this comonoid structure only copies the basis vectors and does not seem adequate for a full copying operation. In fact, a quick computation shows that this Δ in a sense only half-copies of the input vector. In order to see this, consider a vector $\overrightarrow{v} = \sum_i C_i s_i$, for $s_i \in S$. Extending the comultiplication Δ linearly provides us with

$$\Delta(\overrightarrow{v}) = \sum_i C_i \Delta(s_i) = \sum_i C_i(s_i \otimes s_i) = (\sum_i C_i s_i) \otimes (\sum_i s_i) = \overrightarrow{v} \otimes \vec{1},$$

In the second term, we have lost the C_i weights, in other words we have replaced the second copy with a vector of 1's, denoted by $\vec{1}$.

The above problem can be partially overcome by noting that this Δ map is just one of a family of copying maps, parametrised by reals, where for any $k \in \mathbb{R}$ we may define the a *Cofree-inspired* comonoid (V_φ, Δ_k, e) over a vector spafce V_φ with a basis $(v_i)_i$, as:

$$\Delta_k : V_\varphi \to V_\varphi \otimes V_\varphi :: v \mapsto (v \otimes \vec{k}) + (\vec{k} \otimes v), \quad e : V_\varphi \to \mathbb{R} :: \sum_i C_i v_i \mapsto \sum_i C_i$$

Here, $\overrightarrow{v}$ is as before and $\vec{k}$ stands for an element of V padded with number k. In the simplest case, when $k = 1$, we obtain two copies of the weights $\overrightarrow{v}$ and also of its basis vectors, as the following calculation demonstrates. Consider a two dimensional vector space and the vector $ae_1 + be_2$ in it. The 1 vector $\overrightarrow{1}$ is the 2-dimensional vector $e_1 + e_2$ in V. Suppose $\overrightarrow{v}$ and $\vec{1}$ are column vectors, then applying Δ results in the matrix $2a\, e_1 \otimes e_1 + ab\, e_1 \otimes e_2 + ab\, e_2 \otimes e_1 + 2b\, e_2 \otimes e_2$, where we have two copies of the weights in the diagonal and also the basis vectors have obviously multiplied.

This construction is inspired by the graded algebra construction on vector spaces, whose dual construction is referred to as a *Cofree Coalgebra*. The Cofree-inspired coalgebra over a vector space defines a coalgebra modality structure on the identity comonad on $\mathbf{FdVect}_\mathbb{R}$, which provides another $!\mathbf{L}^*$-model, or rather, another quantization $\mathcal{C}(!\mathbf{L}^*) \to \mathbf{FdVect}_\mathbb{R}$.

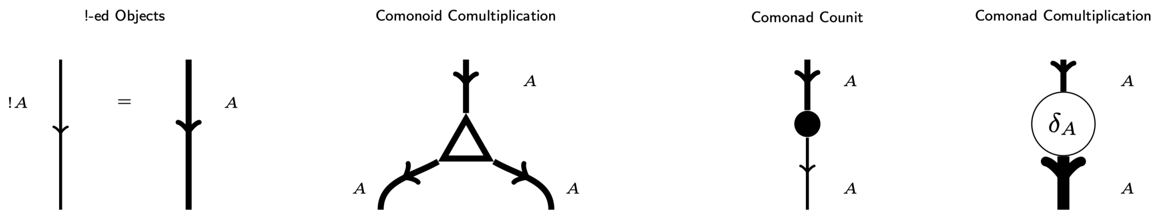

Figure 1: Diagrammatic Structure of ! in $\mathcal{C}(!\mathbf{L}^*)$

6 Clasp Diagrams

In order to show the semantic computations for the parasitic gap, we introduce a diagrammatic semantics. The derivation of the parasitic gap phrase is involved and its categorical or vector space interpretations require close inspection to read. The diagrammatic notation makes it easier to visualise the steps of the derivation and the final semantic form. In what follows we first introduce notation for the Clasp diagrams, then extend them with extra prospective notation necessary to model the ! coalgebra modality. The basic structure of the $\mathcal{C}(!\mathbf{L}^*)$ category, i.e. its objects, morphisms, monoidal product and its internal homs are as in [2]. To these, we add the necessary diagrams for the coalgebra modality, that is the coalgebra comultiplication (copying) Δ, the counit of the comonad ε, and the comonad comultiplication δ, found in figure 1.

7 Linguistic Examples

The motivating example of [20] was the parasitic gap example "the paper that John signed without reading", with the following lexicon:

$$\{(\text{The}, NP\backslash N), (\text{paper}, N), (\text{that}, (N\backslash N)/(S/!NP)), (\text{John}, NP), (\text{signed}, (NP\backslash S)/NP),$$
$$(\text{without}, ((NP\backslash S)\backslash(NP\backslash S))/NP), (\text{reading}, NP/NP)\}\,.$$

The $!\mathbf{L}^*$ derivation of "the paper that John signed without reading" is in the full version of the paper [25]. The categorical semantics of this derivation is the following linear map.

$$(\llbracket NP\rrbracket \Leftarrow \llbracket N\rrbracket) \otimes \llbracket N\rrbracket \otimes ((\llbracket N\rrbracket \Rightarrow \llbracket N\rrbracket) \Leftarrow (\llbracket S\rrbracket \Leftarrow \llbracket !NP\rrbracket)) \otimes \llbracket NP\rrbracket \otimes ((\llbracket NP\rrbracket \Rightarrow \llbracket S\rrbracket) \Leftarrow \llbracket NP\rrbracket) \otimes$$
$$(((\llbracket NP\rrbracket \Rightarrow \llbracket S\rrbracket) \Rightarrow (\llbracket NP\rrbracket \Rightarrow \llbracket S\rrbracket)) \Leftarrow \llbracket NP\rrbracket) \otimes (\llbracket NP\rrbracket \Leftarrow \llbracket NP\rrbracket) \longrightarrow \llbracket NP\rrbracket$$

defined on elements as follows, for the bracketed subscripts in Sweedler notation:

$$\begin{aligned} &the(-) \otimes \overrightarrow{paper} \otimes that(-,-) \otimes \overrightarrow{John} \otimes signed(-,-) \otimes without(-,-,-) \otimes reading(-) \\ &\quad\mapsto \quad the(that(\overrightarrow{paper}, without(\overrightarrow{John}, signed(-,-_{(1)}), reading(-_{(2)})))) \end{aligned}$$

The diagrammatic interpretation of the $!\mathbf{L}^*$-derivation is depicted in figure 2

This is obtained via steps mirroring the steps of the derivation tree of the example, please see the full version of the paper [25].

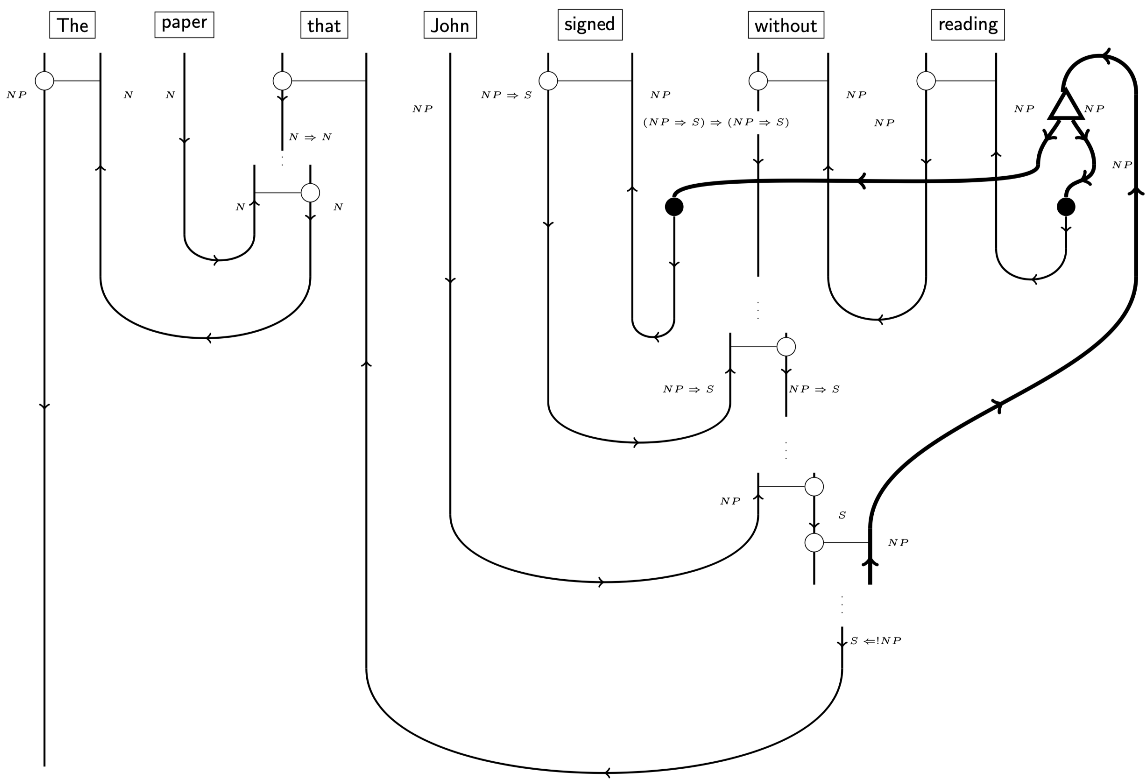

Figure 2: Diagrammatic interpretation of “The paper that John signed without reading”

8 Experimental Comparison

The reader might have rightly been wondering which one of these interpretations, the Cogebra or the Cofree-inspired coalgebra model, produces the correct semantic representation. We implement the resulting vector representations on large corpora of data and experiment with a disambiguation task to provide insights. The disambiguation task was that originally proposed in [15], but we work with the data set of [22], which contains verbs deemed as *genuinely ambiguous* by [33], as those verbs whose meanings are not related to each other. We extended this latter with a second verb and a preposition that provided enough data to turn the dataset from a set of pairs of transitive sentences to a set of pairs of parasitic gap phrases. As an example, consider the verb **file**, with meanings **register** and **smooth**. Example entries of the original dataset and its extension are below; the full dataset is available from `https://msadrzadeh.com/datasets/`.

S: accounts that the local government filed
S1: accounts that the local government registered
S2: accounts that the local government smoothed
P: accounts that the local government filed after inspecting
P1: accounts that the local government registered after inspecting
P2: accounts that the local government smoothed after inspecting

P': nails that the young woman filed after cutting
P'1: nails that the young woman registered after cutting
P'2: nails that the young woman smoothed after cutting
S': nails that the young woman filed
S'1: nails that the young woman registered
S'2: nails that the young woman smoothed

We follow the same procedure as in [22] to disambiguate the phrases with the ambiguous verb: (1) build vectors for phrases P, P1, P2, and also P', P'1, P'2, (2) check whether vector of P is closer to vector of P1 or vector of P2 and whether P' is close to P'2 or P'1. If yes, then we have two correct outputs, (3) compute a mean average precision (MAP), by counting in how many of the pairs, the vector of the phrase with the ambiguous verb is closer to that of the phrase with its appropriate meaning.

In order to instantiate our categorical model on this task and experiment with the different copying maps, we proceed as follows. We work with the parasitic gap phrases that have the general form: "A's the B C'ed Prep D'ing". Here, C and D are verbs and their vector representations are multilinear maps. C is a bilinear map that takes A and B as input and D is a linear map that takes A as input. For now, we represent the preposition Prep by the trilinear map *Prep*. The vector representation of the parasitic gap phrase with a proper copying operator is $Prep(C(\overrightarrow{B}, \overrightarrow{A}), D(\overrightarrow{A}))$, for C and D multilinear maps and $\overrightarrow{A}$ and $\overrightarrow{B}$, vectors, and where $\overrightarrow{A} = \sum_i C_i^A n_i$. The different types of copying applied to this, provide us with the following options.

$$\textbf{Cogebra copying} \qquad (a)\, Prep\left(C(\overrightarrow{B}, \overrightarrow{A}), D(\sum_i n_i)\right), \quad (b)\, Prep\left(C(\overrightarrow{B}, \sum_i n_i), D(\overrightarrow{A})\right)$$

$$\textbf{Cofree-inspired copying} \qquad Prep\left(C(\overrightarrow{B}, \overrightarrow{A}) + D(\vec{1}), C(\overrightarrow{B}, \vec{1}) + D(\overrightarrow{A})\right)$$

Table 2: Parasitic Gap Phrase Disambiguation Results

Model	MAP	Model	MAP	Model	MAP
BERT	0. 65	FT(+)	0.55	W2V (+)	0.46
Full	0.48	Full	0.57	Full	0.54
Cofree-inspired	0.47	Cofree-inspired	0.56	Cofree-inspired	0.54
Cogebra (a)	0.46	Cogebra (a)	0.56	Cogebra (a)	0.46
Cogebra (b)	0.42	Cogebra (b)	0.37	Cogebra (b)	0.39

In the *copy object* model of [22], these choices become as follows:

Cogebra copying

$$(a)\, Prep\left(\overrightarrow{A} \odot (C \times \overrightarrow{B}), D \times \sum_i n_i\right)$$

$$(b)\, Prep\left((\sum_i n_i) \odot (C \times \overrightarrow{B}), (D \times \overrightarrow{A})\right)$$

Cofree-inspired copying

$$Prep\left((\overrightarrow{A} \odot (C \times \overrightarrow{B})) + (D \times \vec{1}), (\vec{1} \odot (C \times \vec{B})) + (D \times \vec{A})\right)$$

For comparison, we also implemented a model where a *Full* copying operation $\Delta(\overrightarrow{v}) = \overrightarrow{v} \otimes \overrightarrow{v}$ was used, resulting in a third option $Prep\left(C(\overrightarrow{B}, \overrightarrow{A}), D(A)\right)$, with the *copy-object* model

$$Prep\left(\overrightarrow{A} \odot (C \times \overrightarrow{B}), D \times \overrightarrow{A}\right)$$

Note that this copying is non-linear and thus cannot be an instance of our $\mathbf{FdVect}_{\mathbb{R}}$ categorical semantics; we are only including it to study how the other copying models will do in relation to it.

The results of experimenting with these models are presented in table 2. We experimented with three neural embedding architectures: BERT [10], FastText (FT) [6], and Word2Vec CBOW (W2V) [40]. For details of the training, please see the full version of the paper [25].

Uniformly, in all the neural architectures, the Full model provided a better disambiguation than other linear copying models. This better performance was closely followed by the Cofree-inspired model: in BERT, the Full model obtained an MAP of 0.48, and the Cofree-inspired model an MAP of 0.47; in FT, we have 0.57 for Full and 0.56 for Cofree-inspired; and in W2V we have 0.54 for both models. Also uniformly, in all of the neural architectures, the Cogebra (a) did better than the Cogebra (b). It is not surprising that the Full copying did better than other two copyings, since this is the model that provides two identical copies of the head noun A. This kind of copying can only be obtained via the application of a non-linear Δ. The fact that our linear Cofree-inspired copying closely followed the Full model, shows that in the absence of Full copying, we can always use the Cofree-inspired as a reliable approximation. It was also not surprising that the Cofree-inspired model did better than either of the Cogebra models, as this model uses the sum of the two possibilities, each encoded in one of the Cogebra (a) or (b). That Cogebra (a) performed better than Cogebra (b), shows that it is more important to have a full copy of the object for the main verb rather than the secondary verb of a parasitic gap phrase. Using this, we can say that verb C that got a full copy of its object A, played a more important role in disambiguation, than verb D, which only got a vector of 1's as a copy of A. Again, this is natural, as the secondary verb only provides subsidiary information.

The most effective disambiguation of the new dataset was obtained via the BERT phrase vectors, followed by the Full model. BERT is a contextual neural network architecture that provides different meanings for words in different contexts, using a large set of tuned parameters on large corpora of data. There is evidence that BERT's phrase vectors do encode some grammatical information in them. So it is not surprising that these embeddings provided the best disambiguation result. In the other neural embeddings: W2V and FT, however, the Full and its Cofree-inspired approximation provided better results. Recall that in these models, phrase embeddings are obtained by adding the word embeddings, and addition forgets the grammatical structure. That the type-driven categorical model outperformed these models is a very promising result.

9 Future Directions

There are plenty of questions that arise from the theory in this paper, concerning alternative syntaxes, coherence, and optimisation.

One avenue we are pursuing is to bound the !-modality of $!\mathbf{L}^*$. This is desirable from a natural language point of view, as the ! of linear logic symbolises *infinite* reuse, however at no point in natural language is this necessary. Thus bounding ! by indexing with natural numbers, similar to Bounded Linear Logic [13] may allow for a more intuitive notion of resource insensitivity closer to that of natural language.

Showing the coherence of the diagrammatic semantics by using the proof nets of Modal Lambek Calculus [29], developed for clasp-string diagrams in [44] constitutes work in progress. Proving coherence would allow us to do all our derivations diagrammatically, making the sequent calculus labour superfluous. However, we suspect there are better notations for the diagrammatic semantics perhaps more closely related to the proof nets of linear logic.

Applications of type-logics with limited contraction and permutation to movement phenomena is a line of research initiated in [14, 3] with a recent boost in [1, 30, 31], and also in [12]. Finding commonalities with these approaches is future work.

We would also like to see how much we can improve the implementation of the cofree-inspired model in this paper. This involves training better tensors, hopefully by using neural networks methods.

10 Acknowledgement

Part of the motivation behind this work came from the Dialogue and Discourse Challenge project of the Applied Category Theory adjoint school during the week 22–26 July 2019. We would like to thank the organisers of the school. We would also like to thank Adriana Correia, Alexis Toumi, and Dan Shiebler, for discussions. McPheat acknowledges support from the UKRI EPSRC Doctoral Training Programme scholarship, Sadrzadeh from the Royal Academy of Engineering Industrial Fellowship IF-192058.

References

[1] Pogadalla Amblard, de Groote and Retoré, editors. *On the Logic of Expansion in Natural Language*, volume 10054, Nancy, France. In Amblard, de Groote, Pogadalla and Retoré (eds.), 2016. Springer. doi:10.1007/978-3-662-53826-5_14.

[2] J. Baez and M. Stay. Physics, topology, logic, and computation: A rosetta stone. In B. Coecke, editor, *New Structures in Physics*, volume 813 of *Lecture Notes in Physics. Springer*. Springer, 2011. doi:DOI:10.1007/978-3-642-12821-9_2.

[3] Guy Barry, Mark Hepple, Neil Leslie, and Glyn Morrill. Proof figures and structural operators for categorial grammar. 1991. doi:10.3115/977180.977215.

[4] R. F. Blute, J. R. B. Cockett, and R. A. G. Seely. Differential categories. *Mathematical Structures in Computer Science*, 16(6):1049–1083, 2006. doi:10.1017/S0960129506005676.

[5] Mehrnoosh Sadrzadeh Bob Coecke, Edward Grefenstette. Lambek vs. lambek: Functorial vector space semantics and string diagrams for lambek calculus. *Ann. Pure and Applied Logic*, 164:1079 – 1100, 2013. doi:10.1016/j.apal.2013.05.009.

[6] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 2017. arXiv:1607.04606, doi:10.1162/tacl_a_00051.

[7] Alain Bruguières and Alexis Virelizier. Hopf monads. *Advances in Mathematics*, 2007. doi:10.1016/j.aim.2007.04.011.

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

[9] JR Curran. From distributional to semantic similarity. *University of Edimburgh*, 2003. doi:10.1.1.10.6068.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018. URL: http://arxiv.org/abs/1810.04805, arXiv:1810.04805.

[11] John R Firth. A synopsis of linguistic theory, 1930-1955, 1957.

[12] M. Sadrzadeh G. Wijnholds. A type-driven vector semantics for ellipsis with anaphora using lambek calculus with limited contraction. *J of Log Lang and Inf*, 28:331–358, 2019. doi:DOI:10.1007/s10849-019-09293-4.

[13] Jean Yves Girard, Andre Scedrov, and Philip J. Scott. Bounded linear logic: a modular approach to polynomial-time computability. *Theoretical Computer Science*, 1992. doi:10.1016/0304-3975(92)90386-T.

[14] Mark Hepple Glyn Morrill, Neil Leslie and Guy Barry. Categorial deductions and structural operations. In *Studies in Categorial Grammar, Edinburgh Working Papers in CognitiveScience*, volume 5, pages 1–21. Centre for Cognitive Science, 1990.

[15] Edward Grefenstette and Mehrnoosh Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics. URL: https://www.aclweb.org/anthology/D11-1129.

[16] Edward Grefenstette and Mehrnoosh Sadrzadeh. Concrete models and empirical evaluations for the categorical compositional distributional model of meaning. *Computational Linguistics*, 2015. doi:10.1162/COLI_a_00209.

[17] Gregory Grefenstette. *Explorations in Automatic Thesaurus Discovery*. 1994. doi:10.1007/978-1-4615-2710-7.

[18] Zellig S. Harris. Distributional Structure. *WORD*, 1954. doi:10.1080/00437956.1954.11659520.

[19] James Edward Humphreys. *Introduction to lie algebras and representation theory*. Springer-Verlag, 1972. doi:10.1007/978-1-4612-6398-2.

[20] Max Kanovich, Stepan Kuznetsov, and Andre Scedrov. Undecidability of the Lambek calculus with a relevant modality. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9804 LNCS:240–256, 2016. arXiv:1601.06303, doi:10.1007/978-3-662-53042-9_14.

[21] Dimitri Kartsaklis and Mehrnoosh Sadrzadeh. Prior disambiguation of word tensors for constructing sentence vectors. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*,

pages 1590–1601, Seattle, Washington, USA, October 2013. Association for Computational Linguistics. URL: https://www.aclweb.org/anthology/D13-1166.

[22] Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Stephen Pulman. Separating disambiguation from composition in distributional semantics. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 114–123, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL: https://www.aclweb.org/anthology/W13-3513.

[23] Thomas K. Landauer and Susan T. Dumais. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 1997. doi:10.1037/0033-295X.104.2.211.

[24] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze. *Introduction to Information Retrieval*. 2008. doi:10.1017/cbo9780511809071.

[25] Lachlan McPheat, Mehrnoosh Sadrzadeh, Hadi Wazni, and Gijs Wijnholds. Categorical vector space semantics for lambek calculus with a relevant modality,https://arxiv.org/abs/2005.03074, 2020. arXiv:2005.03074.

[26] Michael Moortgat Mehrnoosh Sadrzadeh and Gijs Wijnholds. A frobenius algebraic analysis for parasitic gaps. In *Workshop on Semantic Spaces at the Intersection of NLP, Physics, and Cognitive Science*, Riga, Latvia, 2019.

[27] Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Matthew Purver. Evaluating neural word representations in tensor-based compositional settings. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 708–719, Doha, Qatar, October 2014. Association for Computational Linguistics. doi:10.3115/v1/D14-1079.

[28] M. Moortgat and G. Wijnholds. Lexical and derivational meaning in vector-based models of relativisation. In *Proceedings of the 21st Amsterdam Colloquium*, 2017.

[29] Michael Moortgat. Multimodal Linguistic Inference. *Logic Journal of IGPL*, 1995. doi:10.1093/jigpal/3.2-3.371.

[30] Glyn Morrill. Grammar logicised: relativisation. *Linguistics and Philosophy*, 40:119–163, 2017. doi:10.1007/s10988-016-9197-0.

[31] Glyn Morrill. A note on movement in logical grammar. *Journal of Language Modelling*, pages 353–363, 2018. doi:10.15398/jlm.v6i2.233.

[32] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global vectors for word representation. In *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014. doi:10.3115/v1/d14-1162.

[33] Martin Pickering and Steven Frisson. Processing ambiguous verbs: Evidence from eye movements. *Journal of experimental psychology. Learning, memory, and cognition*, 27:556–73, 03 2001. doi:10.1037/0278-7393.27.2.556.

[34] Prakash Panangaden Richard Blute and Robert Seely. Fock space: a model of linear exponential types. *Manuscript, revised version of the MFPS paper Holomorphic models of exponential types in linear logic*, pages 474–512, 1994.

[35] Herbert Rubenstein and John B. Goodenough. Contextual correlates of synonymy. *Commun. ACM*, 8(10):627–633, October 1965. doi:10.1145/365628.365657.

[36] Mehrnoosh Sadrzadeh, Stephen Clark, and Bob Coecke. The Frobenius anatomy of word meanings I: Subject and object relative pronouns. *Journal of Logic and Computation*, 2013. arXiv:1404.5278, doi:10.1093/logcom/ext044.

[37] Mehrnoosh Sadrzadeh, Dimitri Kartsaklis, and Esma Balkır. Sentence entailment in compositional distributional semantics. *Annals of Mathematics and Artificial Intelligence*, 2018. arXiv:1512.04419, doi:10.1007/s10472-017-9570-x.

[38] G. Salton. A document retrieval system for man-machine interaction. In *Proceedings of the 1964 19th ACM national conference*, pages 122.301–122.3020, New York, New York, USA, 1964. ACM Press. doi:10.1145/800257.808923.

[39] P. Selinger. A survey of graphical languages for monoidal categories, 2011. arXiv:0908.3347, doi:10.1007/978-3-642-12821-9_4.

[40] Kai Chen Tomas Mikolov, Ilya Sutskever, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, page 3111–3119, 2013. doi:10.5555/2999792.2999959.

[41] Johan Van Benthem. The Lambek Calculus. 1988. doi:10.1007/978-94-015-6878-4_3.

[42] Gijs Wijnholds and Mehrnoosh Sadrzadeh. Classical copying versus quantum entanglement in natural language: The case of vp-ellipsis. *EPTCS Proceedings of the second workshop on Compositional Approaches for Physics, NLP, and Social Sciences (CAPNS)*, 2018. doi:10.4204/EPTCS.283.8.

[43] Gijs Wijnholds and Mehrnoosh Sadrzadeh. Evaluating composition models for verb phrase elliptical sentence embeddings. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 261–271, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi:10.18653/v1/N19-1023.

[44] Gijs Jasper Wijnholds. Coherent diagrammatic reasoning in compositional distributional semantics. In *International Workshop on Logic, Language, Information, and Computation*, pages 371–386. Springer, 2017. doi:DOI:10.1007/978-3-662-55386-2_27.