Density Matrices with Metric for Derivational Ambiguity

Adriana D. Correia [*]

Institute for Theoretical Physics and Center for Complex Systems Studies, Utrecht University, P.O. Box 80.089, 3508 TB, Utrecht, The Netherlands.

Michael Moortgat

Utrecht Institute of Linguistics OTS and Center for Complex Systems Studies, Utrecht University, 3512 JK, Utrecht, The Netherlands

Henk T. C. Stoof

Institute for Theoretical Physics and Center for Complex Systems Studies, Utrecht University, P.O. Box 80.089, 3508 TB, Utrecht, The Netherlands.

Abstract

Recent work on vector-based compositional natural language semantics has proposed the use of density matrices to model lexical ambiguity and (graded) entailment. Ambiguous word meanings, in this work, are represented as mixed states, and the compositional interpretation of phrases out of their constituent parts takes the form of a strongly monoidal functor sending the derivational morphisms of a pregroup syntax to linear maps in FdHilb.

Our aims in this paper are threefold. Firstly, we replace the pregroup front end by a Lambek categorial grammar with directional implications expressing a word's selectional requirements. By the Curry-Howard correspondence, the derivations of the grammar's type logic are associated with terms of the (ordered) linear lambda calculus; these terms can be read as programs for compositional meaning assembly with density matrices as the target semantic spaces. Secondly, we extend on the existing literature and introduce a symmetric, non-degenerate bilinear form called a "metric" that defines a canonical isomorphism between a vector space and its dual, allowing us to keep a distinction between left and right implication. Thirdly, we use this metric to define density matrix spaces in a directional form, modeling the ubiquitous derivational ambiguity of natural language syntax, and show how this allows an integrated treatment of lexical and derivational forms of ambiguity controlled at the level of the interpretation.

[*]Corresponding author: a.duartecorreia@uu.nl

1 Introduction

Semantic representations of language using vector spaces are an increasingly popular approach to automate natural language processing, with early comprehensive accounts given in [4, 16]. This idea has found several implementations, both theoretically and computationally. On the theoretical side, the principle of compositionality [12] states that the meaning of a complex expression can be computed from the meaning of its simpler building blocks and the rules used to assemble them. On the computational side, the distributional hypothesis [11] asserts that a meaning of a word is adequately represented by looking at what words most often appear next to it. Joining these two approaches, a distributional compositional categorical (DisCoCat) model of meaning has been proposed [5], mapping the pregroup algebra of syntax to vectors spaces with tensor operations, by functorialy relating the properties of the categories that describe those structures, allowing one to interpret compositionality in a grammar-driven manner using data-extracted representations of words that are in principle agnostic to grammar. This method has been shown to give good results when used to compare meanings of complex expressions and with human judgements [10]. Developments in the computation of these vectors that use machine learning algorithms [15] provide representations of words that start deviating from the count-based models. However, each model still provides a singular vector embedding for each word, which allows the DisCoCat model to be applied with some positive results [30].

The principal limitation of these embeddings, designated *static* embeddings, is that it provides the same word representation independently of context. This hides polysemy, or even subtler gradations in meaning. Using the DisCoCat framework, this issue has been tackled using density matrices to describe lexical ambiguity [22, 23], and using the same framework also sentence entailment [24] and graded hyponymy [1], since the use of matrices allows the inclusion of correlations between context words. From the computational side, the most recent computational language models [7, 21] present contextual embeddings of words as an intrinsic feature. In this paper we aim at reconciling the compositional distributional model and these developments by presenting density matrices as the fundamental representations of words, thus leveraging previous results, and by introducing a refined notion of tensor contraction that can be applied even if we do not assume that we are working with static embeddings coming from the data, thus additionally presenting the possibility of eliminating the distinction between context and target words, because all words can be equally represented with respect to one another. To achieve this, we build the components of the density matrices as covariant or contravariant by introducing a metric that relates them, extending to the interpretation space the notion of direc-

tionality of word application, as a direct image of the directional Lambek calculus. After that, we attach permutation operations that act on either type of components to describe derivational ambiguity in a way that keeps multiple readings represented in formally independent vector spaces, thus opening up the possibility of integration between lexical and syntactic ambiguity.

Section 2 introduces our syntactic engine, the Lambek calculus $(\mathbf{N})\mathbf{L}_{/,\backslash}$, together with the Curry-Howard correspondence that associates syntactic derivations with programs of the ordered lambda calculus $\lambda_{/,\backslash}$. Section 3 motivates the use of a more refined notion of inner product and introduces the concept of a tensor and tensor contraction as a basis independent application of a dual vector to a vector, and introduces a metric as the mechanism to go from vectors extracted from the data to the dual vectors necessary to perform tensor contraction. Section 4 gives some background on density matrices, and on ways of capturing the directionality of our syntactic type logic in these semantic spaces using the previously described metric. Section 5 then turns to the compositional interpretation of the $\lambda_{/,\backslash}$ programs associated with $(\mathbf{N})\mathbf{L}_{/,\backslash}$ derivations. Section 6 shows how the directional density matrix framework can be used to capture simple forms of derivational ambiguity.

2   From proofs to programs

With his [13, 14] papers, Jim Lambek initiated the 'parsing as deduction' method in computational linguistics: words are assigned formulas of a type logic designed to reason about grammatical composition; the judgement whether a phrase is well-formed is the outcome of a process of deduction in that type logic. Lambek's original work was on a calculus of *syntactic* types, which he presented in two versions. With $\mathbf{L}_{/,\backslash}$ we refer to the simply typed (implicational) fragment of Lambek's [13] associative syntactic calculus, which assigns types to *strings*; $\mathbf{NL}_{/,\backslash}$ is the non-associative version of [14], where types are assigned to *phrases* (bracketed strings).[1]

Van Benthem [27] added semantics to the equation with his work on $\mathbf{LP}$, a commutative version of the Lambek calculus, which in retrospect turns out to be a precursor of (multiplicative intuitionistic) linear logic. $\mathbf{LP}$ is a calculus of *semantic* types. Under the Curry-Howard 'proofs-as-programs' approach, derivations in $\mathbf{LP}$ are in 1-to-1 correspondence with terms of the (linear) lambda calculus; these terms

---

[1]Neither of these calculi by itself is satisfactory for modelling natural language syntax. To handle the well-documented problems of over/undergeneration of $(\mathbf{N})\mathbf{L}_{/,\backslash}$ in a principled way, the logics can be extended with modalities that allow for controlled forms of reordering and/or restructuring. We address these extensions in [6].

Terms: $\quad t, u \quad ::= \quad x \quad | \quad \lambda^r x.t \quad | \quad \lambda^l x.t \quad | \quad t \triangleleft u \quad | \quad u \triangleright t$

Typing rules:

$$\frac{}{x : A \vdash x : A} \; Ax$$

$$\frac{\Gamma, x : A \vdash t : B}{\Gamma \vdash \lambda^r x.t : B/A} \; I/ \qquad \frac{x : A, \Gamma \vdash t : B}{\Gamma \vdash \lambda^l x.t : A\backslash B} \; I\backslash$$

$$\frac{\Gamma \vdash t : B/A \quad \Delta \vdash u : A}{\Gamma, \Delta \vdash t \triangleleft u : B} \; E/ \qquad \frac{\Gamma \vdash u : A \quad \Delta \vdash t : A\backslash B}{\Gamma, \Delta \vdash u \triangleright t : B} \; E\backslash$$

Figure 1: Proofs as programs for $(\mathbf{N})\mathbf{L}_{/,\backslash}$.

can be seen as *programs* for compositional meaning assembly. To establish the connection between syntax and semantics, the Lambek-Van Benthem framework relies on a homomorphism sending types and proofs of the syntactic calculus to their semantic counterparts.

In this paper, rather than defining semantic interpretation on a commutative type logic such as **LP**, we want to keep the distinction between the left and right implications $\backslash, /$ of the syntactic calculus in the vector-based semantics we aim for. To achieve this, our programs for meaning composition use the language of Wansing's [29] *directional* lambda calculus $\lambda_{/,\backslash}$. Wansing's overall aim is to study how the derivations of a family of substructural logics can be encoded by typed lambda terms. Formulas, in the substructural setting, are seen as information pieces, and the proofs manipulating these formulas as information processing mechanisms, subject to certain conditions that reflect the presence or absence of structural rules. The terms of $\lambda_{/,\backslash}$ faithfully encode proofs of $(\mathbf{N})\mathbf{L}_{/,\backslash}$; information pieces, in these logics, cannot be copied or deleted (absence of Contraction and Weakening), and information processing is sensitive to the sequential order in which the information pieces are presented (absence of Permutation).

We present the rules of $(\mathbf{N})\mathbf{L}_{/,\backslash}$ with the associated terms of $\lambda_{/,\backslash}$ in Fig 1. The presentation is in the sequent-style natural deduction format. The formula language has atomic types (say *s, np, n* for sentences, noun phrases, common nouns) for complete expressions and implicational types $A\backslash B$, B/A for incomplete expressions, selecting an A argument to the left (resp. right) to form a B.

Ignoring the term labeling for a moment, judgments are of the form $\Gamma \vdash A$, where the antecedent Γ is a non-empty list (for **L**) or bracketed list (**NL**) of formulas, and the succedent a single formula A. For each of the type-forming operations, there is an Introduction rule, and an Elimination rule.

Turning to the Curry-Howard encoding of $\mathbf{NL}_{/,\backslash}$ proofs, we introduce a language of directional lambda terms, with variables as atomic expressions, left and right λ abstraction, and left and right application. The inference rules now become *typing* rules for these terms, with judgments of the form

$$x_1 : A_1, \ldots, x_n : A_n \vdash t : B. \tag{1}$$

The antecedent is a typing environment providing type declarations for the variables x_i; a proof constructs a program t of type B out of these variables. In the absence of Contraction, Weakening and Permutation structural rules, the program t contains $x_1, \ldots, x_n$ as free variables exactly once, and in that order. Intuitively, one can see a term-labelled proof as an algorithm to compute a meaning t of type B with parameters x_i of type A_i. In parsing a particular phrase, one substitutes the meaning of the constants (i.e. words) that make it up for the parameters of this algorithm.

3 Directionality in interpretation

In order to introduce the directionality of the syntactic calculus in the semantic calculus, we expand on the existing literature that uses **FdVect** as the interpretation category by calling attention to the implied inner product. We introduce a more abstract notion of tensor, tensor contraction and the need to introduce explicitly the existence of a metric, coming from the literature of general relativity, following the treatment in [28].[2] Formally, a metric is a function that assigns a distance between two elements of a set, but if applied to the elements of a set that is closed under addition and scalar multiplication, that is, the elements of a vector space, it becomes an inner product. Since we will be looking at vector spaces, we use the terms metric and inner product interchangeably.

To motivate the need for a more careful treatment regarding the inner product, lets look at a very simple yet illustrative example. Suppose that a certain language model provides word embeddings that correspond to two-dimensional, real valued vectors. In this model, the words "vase" and "wall" have the vector representations $\vec{v}$ and $\vec{w}$, respectively

---

[2]An alternative introductory treatment of tensor calculus can be found in [8].

$$\vec{v} = (0,1), \quad \text{and} \quad \vec{w} = (1,0). \tag{2}$$

This representation could mean that they are context words in a count-based model, since they form the standard (orthogonal) basis of $\mathbb{R}^2$, or that they have this particular representation in a particular context-dependent language model. To compute cosine similarity, the notion of Euclidean inner product is used, where the components corresponding to a certain index are multiplied:

$$\vec{v} \cdot \vec{w} = 0 \cdot 1 + 1 \cdot 0 = 0, \tag{3}$$

which we can use to calculate the cosine of the angle θ between these vectors,

$$\cos(\theta) = \frac{\vec{v} \cdot \vec{w}}{\|\vec{v}\| \cdot \|\vec{w}\|} = \frac{0 \cdot 1 + 1 \cdot 0 = 0}{1 \cdot 1} = 0. \tag{4}$$

Thus, if the representations of these words are orthogonal, then using this measure to evaluate similarity we conclude that these words are not related. However, there is a degree of variation in the vectors that are assigned to the distributional semantics of each word. *Static embeddings* are unique vector representations given by a global analysis of a word over a corpus. The unique vector assigned to the semantics of a word depends on the model used to analyze the data, so different models do not necessarily put out the same vector representations. Alternative to this are *dynamic embeddings*, which assign different vector representations to the same word depending on context, within the same model.

Therefore, there are at least three ways in which the result of eq.4 and subsequent interpretation can be challenged:

1. **Static Embeddings.** If the representations come from a count-based model, choosing other words as context words changes the vector representation and therefore these words are not orthogonal to one another anymore; in fact this can happen with any static embedding representation when the basis of the representation changes. Examples of models that give static embeddings are Word2Vec [15] and GloVe [20].

2. **Dynamic Embeddings.** When the vector representations come from a context-dependent embedding, changing the context in which the words are evaluated influences their representation, which might not be orthogonal anymore. Dynamic embeddings can be obtained with i.e. ELMo [21] and BERT [7].

3. **Expectation of meaning.** Human judgements, which are the outcomes of experiments where subjects are explicitly asked to rate the similarity of words, predict that some words should have a degree of relationship. Therefore, the conclusion we derive from orthogonal representations of certain words might not be valid if there is a disagreement with their human assessment. These judgements are condensed in datasets such as the MEN dataset [3].

While points 1 and 2 can be related, caution is necessary in establishing that link. On a preliminary inspection, comparing the cosine similarity of context-free embeddings of nouns extracted from pre-trained BERT [7] with the normalized human judgements from the MEN dataset [3], we find that the similarity between two words given by the language model is systematically overrated when compared to its human counterpart. One possible explanation is that the language model is comparing all words against one another, so it is an important part of similarity that the two words belong to the the same part of speech, namely nouns, while humans assume that as a condition for similarity evaluation. Further, though we can ask the language model to rate the similarity of words in specific contexts, that has not explicitly been done with human subjects. A more detailed comparison between context-depend representations and human judgement constitutes further research.

One way to reconcile the variability of representations and the notion of similarity is to expand the notion of inner product to be invariant under the change of representations. Suppose now that by points 1 or 2 the representations of "vase" and "wall" change, respectively, to

$$\vec{v}' = (1, 1), \ \vec{w}' = (-1, 2). \tag{5}$$

These vectors also form a basis of $\mathbb{R}^2$, but not an orthogonal one. If we use the same measure to compute similarity, taking normalization into account, the Euclidean inner product gives $\vec{v}' \cdot \vec{w}' = (-1) \cdot 1 + 1 \cdot 2 = 1$ and cosine similarity gives

$$cos(\theta') = \frac{\vec{v}' \cdot \vec{w}'}{\|\vec{v}'\| \cdot \|\vec{w}'\|} = \frac{1}{\sqrt{2} \cdot \sqrt{5}} = \frac{1}{\sqrt{10}}. \tag{6}$$

If now we have a conflict between which representations are the correct ones, we can look at the human evaluations of similarity. Suppose that it corresponds too to $\frac{1}{\sqrt{10}}$.

We argue in this paper that, by introducing a different notion of inner product, we can fine-tune a relationship between the components of the vectors with the goal to preserve a particular value, for example a human similarity judgement. In this framework, the different representations of words in dynamic embeddings are

brought about by a change of basis, similarly to what happens when the context words change in static embeddings, in which case the value of the inner product should be preserved. This can be achieved by describing the inner product as a tensor contraction between a vector and a dual vector, with the latter computed using a metric.

Let V be a finite dimensional vector space and let V^* denote its dual vector space, constituted by the linear maps from V to the field $\mathbb{R}$. A tensor T of type (k, l) over V is a multilinear map

$$T : \underbrace{V^* \times \cdots \times V^*}_{k} \times \underbrace{V \times \cdots \times V}_{l} \to \mathbb{R}. \tag{7}$$

Once applied on k dual vectors and l vectors, a tensor outputs an element of the field, in this case a real number. By this token, a tensor of type $(0, 1)$ is a dual vector, which is the map from the vector space to the field, and a tensor of type $(1, 0)$, being technically the dual of a dual vector, is naturally isomorphic to a vector. Given a basis $E = \{\hat{e}_i\}$ in V and its dual basis $_dE = \{\hat{e}^j\}$ in V^*, with $\hat{e}^j(\hat{e}_i) = \delta_i^j$, the tensor product between the basis vectors and dual basis vectors forms a basis $B = \{\hat{e}_{i_1} \otimes \cdots \otimes \hat{e}_{i_k} \otimes \hat{e}^{j_1} \otimes \cdots \otimes \hat{e}^{j_l}\}$ of a tensor of type (k, l), allowing the tensor to be expressed with respect to this basis as

$$T = \sum_{i_1,\ldots,i_k,j_1,\ldots,j_l} T^{i_1\ldots i_k}{}_{j_1\ldots j_l} \hat{e}_{i_1} \otimes \cdots \otimes \hat{e}_{i_k} \otimes \hat{e}^{j_1} \otimes \cdots \otimes \hat{e}^{j_l}. \tag{8}$$

The basis expansion coefficients $T^{i_1\ldots i_k}{}_{j_1\ldots j_l}$ are called the *components* of the tensor.

We can perform two important operations on tensors: apply the tensor product between them, $T' \otimes T$, and contract components of the tensor, CT. The first operation happens in the obvious way, while the second corresponds to applying one of the basis dual vectors to a basis vector, resulting in an identification and summing of the corresponding components:

$$(CT)^{i_1\ldots i_{k-1}}{}_{j_1\ldots j_{l-1}} = \sum_{\sigma} T^{i_1\ldots \sigma\ldots i_{k-1}}{}_{j_1\ldots \sigma\ldots j_{l-1}}. \tag{9}$$

The outcome is a tensor of type $(k-1, l-1)$. Note that this procedure is basis independent, because of the relationship between the basis and dual basis. For a tensor of type $(1, 1)$, which represents a linear operator from V to V, tensor contraction corresponds precisely to taking the trace of that operator. To simplify the notation, we will use primed indices instead of numbered ones when the tensors have a low rank. We define a special $(0, 2)$ tensor called a *metric d*:

$$d = \sum_{j,j'} d_{jj'} \hat{e}^j \otimes \hat{e}^{j'}. \tag{10}$$

This tensor is symmetric and non-degenerate. The contraction of this tensor with two vectors v and w gives the value of the inner product:

$$d(v,w) = \sum_{j,j'} d_{jj'} v^j w^{j'}. \tag{11}$$

Because of symmetry, $d(v,w) = d(w,v)$, and because of non-degeneracy, the metric is invertible, with its inverse d^{-1} expressed as

$$d^{-1} = \sum_{i,i'} d^{ii'} \hat{e}_i \otimes \hat{e}_{i'}. \tag{12}$$

Given that the elements extracted from the data are elements of V, the contractions that need to be performed, for example for the application of the compositionality principle in vector spaces, must involve a passage from vectors to dual vectors as seen in the DisCoCat model, before contraction takes place. The metric can be used to define a canonical map between V and V^* via the partial map that is obtained when only one vector is used as an argument of the metric, giving rise to the dual vector $_dv : v \mapsto d(-,v)$, with the slash indicating the empty argument slot:

$$d(v,w) \equiv d(v,-)(w) \equiv {}_dv(w). \tag{13}$$

This formulation is basis independent, since it results from tensor contraction. Once a basis is defined, the resulting dual vector can be expressed as

$$v^d = \sum_{i,j,j'} d_{jj'} v^i \hat{e}^j \otimes \hat{e}^{j'}(\hat{e}_i) = \sum_{j,j'} d_{jj'} v^{j'} \hat{e}^j = \sum_{j'} v_{j'} \hat{e}^{j'}, \tag{14}$$

where we rewrite $v_{j'} = \sum_j d_{jj'} v^{j'}$.

We call the components of vectors, with indices "up", the *contravariant* components, and those of dual vectors, with indices "down", the *covariant* components. Thus, consistent with our notation, the metric can be used to "lower" or "raise" indices, applying contraction between the metric and the tensor and relabeling the components:

$$d(T) = \sum_{i_1,\ldots,i_k,j_1,\ldots,j_{l+2}} d_{j_{l+1},j_{l+2}} T^{i_1,\ldots,i_k}{}_{j_1,\ldots,j_l} \hat{e}^{j_{l+1}} \otimes \hat{e}^{j_{l+2}}(\hat{e}_{i_1}) \otimes \ldots \otimes \hat{e}^{j_l}$$

$$= \sum_{i_1,\ldots,i_k,j_1,\ldots,j_{l+1}} d_{j_{l+1},i_1} T^{i_1,\ldots,i_k}{}_{j_1,\ldots,j_l} \hat{e}^{j_{l+1}} \otimes \hat{e}_{i_2} \otimes \ldots \otimes \hat{e}^{j_l}$$

$$= \sum_{i_2,\ldots,i_k,j_1,\ldots,j_{l+1}} T_{j_{l+1}}{}^{i_2,\ldots,i_k}{}_{j_1,\ldots,j_l} \hat{e}^{j_{l+1}} \otimes \hat{e}_{i_2} \otimes \ldots \otimes \hat{e}^{j_l}. \tag{15}$$

The effect of the metric on a tensor can be captured by seeing how we rewrite the components of some example tensors:

- $\sum_{j'} d_{jj'} T^{j'}{}_{j''} = T_{jj''}$;

- $\sum_{i'} T^i{}_{i'} d^{i'i''} = T^{ii''}$;

- $\sum_{j',j'''} d_{jj'} d_{j''j'''} T^{j'j'''} = T_{jj''}$.

Most importantly, a proper tensor is only defined in the form of eq.8, so whenever we have a tensor that has components "up" and "down" in different orders, for example in $T_j{}^i$, this is in fact a tensor of type $(1,1)$ of which the actual value of the components is

$$\sum_{i',j'} d^{ii'} d_{jj'} T^{j'}{}_{i'}. \tag{16}$$

Returning to our toy example with the words "vase" and "wall", we can look at the change in vector representations as a change of basis $\hat{e}_i = \sum_{i'} \Lambda_i{}^{i'} \hat{e}'_{i'}$:

$$\vec{v} = \sum_i v^i \hat{e}_i = \sum_{ii'} v^i \Lambda_i{}^{i'} \hat{e}'_{i'} = \sum_{i'} v'^{i'} \hat{e}'_{i'}, \tag{17}$$

corresponding to a change in the vector components $v'^{i'} = v^i \Lambda_i{}^{i'}$. The components of the metric also change with the basis:

$$d'_{j''j'''} = \Lambda_{j''}{}^{j'} \Lambda_{j'''}{}^{j} d_{jj'}. \tag{18}$$

With this change, we can show that inner product remains invariant under a basis change:

$$w'^{i'} v'_{i'} = w'^{i'} v'^{j'} d'_{j'i'} = w'^{i'} v'^{j'} \Lambda_{i'}{}^{i} \Lambda_{j'}{}^{j} d_{ji} = w^i v^j d_{ji} = w^i v_i. \tag{19}$$

In this way, finding the right metric allows us to preserve a value that is constant in the face of context dependent representations. Assuming a metric that has the following matrix representation in the standard basis,

$$d = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}, \tag{20}$$

its application to the vector elements in eqs.2 gives a value of the inner product calculated in the new representation:

$$v'_{i'} w'^{i'} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1. \tag{21}$$

Since norms of the vectors have to be calculated using the same notion of inner product,

$$\|\vec{v}\| = \sqrt{v^i g_{ij} v^j}, \tag{22}$$

we find exactly the cosine similarity calculated in eq.6. Note that this formalism allows us to deal with non-orthogonal basis, but does not require it: in fact, there is an implicit metric already when we compute the Euclidean inner product in eq.2, given by $d_{orth} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ in the standard basis, which is the one assumed when talking about an orthonormal basis.

Since these new tools allow us to preserve a quantity in the face of a change of representation, we can start reversing the question on similarity: given a certain human judgement on similarity, or another constant of interest, what is the metric that preserves it across different representations?[3] Once the vector spaces are endowed with specific metrics, the new inner product definitions permeate all higher-rank tensor contractions that are performed between higher and lower rank tensors, namely the ones that will be used in the interpretation of the Lambek rules,[4]

---

[3]In case the quantity we wish to preserve is other than that of the Euclidean inner product in either representation, there is an option to expand the vector representation of our words by adding vector components that act as parameters, to ensure that the quantity is indeed conserved. This would be similar to the role played by the time dimension in Einstein's relativity theory.

[4]Using this formalism, we can replace the unit and counit maps ϵ and η maps of the compact closed category **FdVect** by

$$\eta^l : \mathbb{R} \to V \otimes V^* :: 1 \mapsto \mathbb{1} \otimes d(\mathbb{1}, -)$$
$$\eta^r : \mathbb{R} \to V^* \otimes V :: 1 \mapsto d(-, \mathbb{1}) \otimes \mathbb{1}$$
$$\epsilon^l : V^* \otimes V \to \mathbb{R} :: d(-, v) \otimes u \mapsto d(u, v)$$
$$\epsilon^r : V \otimes V^* \to \mathbb{R} :: v \otimes d(u, -) \mapsto d(u, v).$$

and can further be extended to density matrices.

3.1   Metric in Dirac Notation

We want to lift our description to the realm of density matrices. We now show how the concept of a metric can also be introduced in that description, such that the previously described advantages carry over.

Dirac notation is the usual notation for vectors in the quantum mechanics literature. To make the bridge with the previous concepts from tensor calculus, we introduce it simply as a different way to represent the basis and dual basis of a vector space. Let us rename their elements as *kets* $|i\rangle \equiv \hat{e}_i$ and as *bras* $\langle j| \equiv \hat{e}^j$. The fact that the bases are dual to one another is expressed by the orthogonality condition $\langle j|i\rangle = \delta_{ij}$, which, if the vector basis elements are orthogonal to each other, is equivalent to applying the Euclidean metric to $|i\rangle$ and $|j\rangle$. Using Dirac notation, a vector and dual vector are represented as $v \equiv |v\rangle = \sum_i v^i |i\rangle$ and $v^d \equiv \langle u| = \sum_j v_j \langle j|$.[5] If the basis elements are not orthogonal, this mapping has to be done through a more involved metric. To express this, in this paper we introduce a modified Dirac notation over the field of real numbers, inspired by the one used in [9] for the treatment of quantum states related by a specific group structure.[6] The previous basis elements of V are written now as $|_i\rangle \equiv \hat{e}_i$ and the corresponding dual basis as $\langle^j| \equiv \hat{e}^j$, such that $\langle^j|_i\rangle = \delta_i^j$. In this basis, the metric is expanded as $d = \sum_{j,j'} d_{j'j} \langle^j| \otimes \langle^{j'}|$ while the inverse metric is expressed as $d^{-1} = \sum_{ii'} d^{i'i} |_i\rangle \otimes |_{i'}\rangle$. The elements of the metric and inverse metric are related by $\sum_i d_{j'i'} d^{i'i} = \delta_{j'}^{i'}$. Applying the metric to a basis element of V, we get

$$\langle_i| \equiv d(-,|_i\rangle) = \sum_{jj'} d_{j'j} \langle^j| \otimes \langle^{j'}|_i\rangle = \sum_j d_{ij} \langle^j|. \tag{23}$$

Acting with this on $|_{i'}\rangle$ to extract the value of the inner product, the following formulations are equivalent:

$$d(|_{i'}\rangle, |_i\rangle) = d(-,|_i\rangle) |_{i'}\rangle = \sum_j d_{ij} \langle^j|_{i'}\rangle = \langle_i|_{i'}\rangle = d_{ii'}. \tag{24}$$

When the inverse metric is applied to $\langle^j|$ it gives

---

[5]For orthonormal basis over the field of complex numbers, the covariant components are simply given by the complex conjugate of the contravariant ones, $v_i = \bar{v}^i$.

[6]This treatment can be extended to the field of complex numbers by considering that the metric has conjugate symmetry, $d_{ij} = \bar{d}_{ji}$ [25].

$$\left|^j\right\rangle \equiv d\left(-,\left\langle^j\right|\right) = \left\langle^j\right| \sum_{ii'} d^{i'i} \left|_i\right\rangle \otimes \left|_{i'}\right\rangle = \sum_{i'} d^{i'j} \left|_{i'}\right\rangle, \qquad (25)$$

with a subsequent application on $\left\langle^{j'}\right|$ giving

$$d^{-1}\left(\left\langle^{j'}\right|,\left\langle^j\right|\right) = \left\langle^{j'}\right| d\left(-,\left\langle^j\right|\right) = \left\langle^{j'}\right| \sum_{i'} d^{i'j} \left|_{i'}\right\rangle = \left\langle^{j'}\middle|^j\right\rangle = d^{j'j}. \qquad (26)$$

Consistently, we can calculate the value of the new bras and kets defined in eqs.23 and 25 applied to one other, showing that they too form a basis/dual basis pair:

$$\left\langle_i\middle|^j\right\rangle = \sum_{j'} d_{ij'} \left\langle^{j'}\right| \sum_{i'} d^{i'j} \left|_{i'}\right\rangle = \sum_{i'j'} d_{ij'} d^{i'j} \left\langle^{j'}\middle|_{i'}\right\rangle = \sum_{j'} d_{ij'} d^{j'j} = \delta_i^j. \qquad (27)$$

If the basis elements are orthogonal, the components of the metric and inverse metric coincide with the orthogonality condition.

4   Density Matrices: Capturing Directionality

The semantic spaces we envisage for the interpretation of the syntactic calculus are density matrices. A *density matrix* or density operator is used in quantum mechanics to describe systems for which the state is not completely known. For lexical semantics, it can be used to describe the meaning of a word by placing distributional information on its components. As standardly presented,[7] density matrices that are defined on a tensor product space indicate no preference with respect to contraction from the left or from the right. Because we want to keep the distinction between left and right implications in the semantics, we set up the interpretation of composite spaces in such a way that they indicate which parts will and will not contract with other density matrices.

The *basic* building blocks of the interpretation are density matrix spaces $\tilde{V} \equiv V \otimes V^*$. For this composite space, we choose the basis formed by $\left|_i\right\rangle$ tensored with $\left\langle_{i'}\right|$, $\tilde{E} = \{\left|_i\right\rangle \left\langle_{i'}\right|\} = \left\{\tilde{E}_J\right\}$. Carrying over the notion of duality to the density matrix space, we define the dual density matrix space $\tilde{V}^* \equiv V \otimes V^*$. The dual basis in this space is the map that takes each basis element of $\tilde{V}$ and returns the appropriate orthogonality conditions. It is formed by $\left\langle^j\right|$ tensored with $\left|^{j'}\right\rangle$, $_d\tilde{E} = \left\{\left|^{j'}\right\rangle \left\langle^j\right|\right\} = \left\{\tilde{E}^J\right\}$, and is applied on the basis vectors of $\tilde{V}$ via the trace operation

---

[7]A background for the non-physics reader can be found in [19].

$$\tilde{E}^J\left(\tilde{E}_I\right) = \mathrm{Tr}\left(\left|_i\right\rangle\left\langle_{i'}\right|^{j'}\right\rangle\left\langle^j\right|\right) = \sum_l \left\langle^l\right|_i\right\rangle\left\langle_{i'}\right|^{j'}\right\rangle\left\langle^j\right|_l\right\rangle$$

$$= \sum_{jj'} \left\langle^j\right|_i\right\rangle\left\langle_{j'}\right|^{i'}\right\rangle \delta_i^j \delta_{i'}^{j'} \equiv \delta_I^J. \tag{28}$$

Because density operators are hermitian, their matrices do not change under conjugate transposition, which extends to elements of the basis of the density matrix space. In this way, we can extend our notion of metric to the space of density matrices, where a new metric D emerges from d, expanded in the basis of V^* as

$$D = \sum_{J,J'} D_{JJ'} \tilde{E}^J \otimes \tilde{E}^{J'} \tag{29}$$

$$= \sum_{jj',j''j'''} d_{j''j'} d_{j'''j} \left|^{j'}\right\rangle\left\langle^j\right| \otimes \left|^{j'''}\right\rangle\left\langle^{j''}\right|. \tag{30}$$

We can see how both definitions are equivalent by their action on a density matrix tensor $T \equiv \sum_I T^I \tilde{E}_I \equiv \sum_{ii'} T^{ii'} \left|_i\right\rangle\left\langle_{i'}\right|$. Staying at the level of $\tilde{V}$ and $\tilde{V}^*$, we use eq.29 to obtain

$$D(-,T) = \sum_{I,J,J'} D_{JJ'} T^I \tilde{E}^J \otimes \tilde{E}^{J'}\left(\tilde{E}_I\right) = \sum_{I,J,J'} D_{JJ'} T^I \tilde{E}^J \delta_I^{J'}$$

$$= \sum_{J,J'} D_{JJ'} T^{J'} \tilde{E}^J \equiv \sum_J T_J \tilde{E}^J = \sum_{jj'} T_{j'j} \left|^{j'}\right\rangle\left\langle^j\right|, \tag{31}$$

where we redefine $T_J \equiv D_{JJ'} T^{J'}$, thus establishing covariance and contravariance of the tensor components defined over the density matrix space. Looking in its turn at the level of V and V^*, using eq.30, we see that both definitions are equivalent:

$$D(-,T) = \sum_{ii',jj',j''j'''} T^{ii'} d_{j''j'} d_{j'''j} \left|^{j'}\right\rangle\left\langle^j\right| \otimes \mathrm{Tr}\left(\left|^{j'''}\right\rangle\left\langle^{j''}\right|_i\right\rangle\left\langle_{i'}\right|\right)$$

$$= \sum_{ii',jj',j''j'''} T^{ii'} d_{j''j'} d_{j'''j} \delta_i^{j''} \delta_{i'}^{j'''} \left|^{j'}\right\rangle\left\langle^j\right|$$

$$= \sum_{ii'jj'} T^{ii'} d_{ij'} d_{i'j} \left|^{j'}\right\rangle\left\langle^j\right| \equiv \sum_{jj'} T_{jj'} \left|^{j'}\right\rangle\left\langle^j\right|, \tag{32}$$

where we rewrite $T_{jj'} \equiv T^{ii'} d_{ij'} d_{i'j}$.[8]

From these basic building blocks, *composite* spaces are formed via the binary operation $\otimes$ (tensor product) and a unary operation $()^*$ (dual functor) that sends the elements of a density matrix basis to its dual basis, using the metric defined above. In the notation, we use $\tilde{A}$ for density matrix spaces (basic or compound), and ρ, or subscripted $\rho_x, \rho_y, \rho_z, \ldots \in \tilde{A}$ for elements of such spaces. The $()^*$ operation is involutive; it interacts with the tensor product as $(\tilde{A} \otimes \tilde{B})^* = \tilde{B}^* \otimes \tilde{A}^*$ and acts as identity on matrix multiplication.

Below in (†) is the general form of a density matrix defined on a single space in the standard basis, and (‡) in the dual basis:

$$(\dagger) \quad \rho_x^{\tilde{A}} = \sum_{ii'} X^{ii'} \left|i\right\rangle {}_{\tilde{A}}\!\left\langle i'\right|, \qquad (\ddagger) \quad \rho_x^{\tilde{A}^*} = \sum_{jj'} X_{j'j} \left|{}^{j'}\right\rangle {}_{\tilde{A}^*}\!\left\langle {}^{j}\right|.$$

Over the density matrix spaces, we can see these matrices as *tensors* as we defined them previously, with $X^I \equiv X^{ii'}$ the *contravariant* components and with $X_{J'} \equiv X_{j'j}$ the *covariant* components.

A density matrix of a composite space can be an element of the tensor product space between the standard space and the dual space either from the left or from the right:

$$\rho_y^{\tilde{A} \otimes \tilde{B}^*} = \sum_{ii', jj'} Y^{ii'}_{j'j} \left|{}_i^{\,j'}\right\rangle {}_{\tilde{A} \otimes \tilde{B}^*}\!\left\langle {}_{i'}^{\,j}\right|; \tag{33}$$

$$\rho_w^{\tilde{B}^* \otimes \tilde{A}} = \sum_{ii', jj'} W_{j'j}^{\;ii'} \left|{}^{j'}_{\,i}\right\rangle {}_{\tilde{B}^* \otimes \tilde{A}}\!\left\langle {}^{j}_{\,i'}\right|. \tag{34}$$

Although both tensors are of the form $(1, 1)$, the last one is a tensor with components $Y^I_{J'}$, which relate with a true tensor form by $D^{II'} Y_{I'}{}^J D_{JJ'}$. Recursively, density matrices that live in higher-rank tensor product spaces can be constructed, taking a tensor product with the dual basis either from the left or from the right. Multiplication between two density matrices of a standard and a dual space follows the rules of tensor contraction:

$$\rho_y^{\tilde{A}^*} \cdot \rho_x^{\tilde{A}} = \sum_{jj'} Y_{j'j} \left|{}^{j'}\right\rangle {}_{\tilde{A}^*}\!\left\langle {}^{j}\right| \cdot \sum_{ii'} X^{ii'} \left|i\right\rangle {}_{\tilde{A}}\!\left\langle i'\right| = \sum_{i', jj'} Y_{j'j} X^{ji'} \left|{}^{j'}\right\rangle {}_{\tilde{A}}\!\left\langle i'\right|. \tag{35}$$

---

[8]Here we can compare our formalism to that of the compact closed category of completely positive maps **CPM(FdVect)** developed in [26]. The categorical treatment applies here at a higher level, however, introducing the metric defines explicitly the canonical isomorphisms $V \cong V^*$ and $\tilde{V} \cong \tilde{V}^*$, which trickles down to knowing exactly how the symmetry of the tensor product acts on the compenents of a tensor: $\sigma_{V,V^*} : V^* \otimes V \to V \otimes V^* :: \sum_{ij} T_i{}^j \hat{e}^i \otimes \hat{e}_j \mapsto \sum_{ii',jj'} d^{ii'} d_{jj'} T^{j'}{}_{i'} \hat{e}_i \otimes \hat{e}^j$.

$$\rho_x^{\tilde{A}} \cdot \rho_y^{\tilde{A}^*} = \sum_{ii'} X^{ii'} \left|i\right\rangle_{\tilde{A}} \left\langle i'\right| \cdot \sum_{jj'} Y_{j'j} \left|j'\right\rangle_{\tilde{A}^*} \left\langle j\right| = \sum_{i,jj'} X^{ij'} Y_{j'j} \left|i\right\rangle_{\tilde{A}} \left\langle j\right|, \qquad (36)$$

respecting the directionality of composition. To achieve full contraction, the trace in the appropriate space is applied, corresponding to a partial trace if the tensors involve more spaces:

$$\text{Tr}_{\tilde{A}} \left(\sum_{i',jj'} Y_{j'j} X^{ji'} \left|j'\right\rangle_{\tilde{A}} \left\langle i'\right| \right) = \sum_{l,i',jj'} Y_{j'j} X^{ji'} {}_{\tilde{A}} \left\langle l \middle| j'\right\rangle_{\tilde{A}^*} {}_{\tilde{A}} \left\langle i' \middle| l\right\rangle_{\tilde{A}^*} = \sum_{jj'} Y_{j'j} X^{jj'},$$
$$(37)$$

$$\text{Tr}_{\tilde{A}} \left(\sum_{i,jj'} X^{ij'} Y_{j'j} \left|i\right\rangle_{\tilde{A}} \left\langle j\right| \right) = \sum_{l,j',ij} X^{ij'} Y_{j'j} {}_{\tilde{A}^*} \left\langle l \middle| i\right\rangle_{\tilde{A}} {}_{\tilde{A}^*} \left\langle j \middle| l\right\rangle_{\tilde{A}} = \sum_{jj'} X^{jj'} Y_{j'j}. \quad (38)$$

We see that the cyclic property of the trace is preserved.

In §6 we will be dealing with derivational ambiguity, and for that the concepts of *subsystem* and *permutation operation* introduced here will be useful. A subsystem can be thought of as a copy of a space, described using the same basis, but formally treated as a different space. In practice, this means that different subsystems do not interact with one another. In the quantum setting, they represent independent identical quantum systems. For example, when we want to describe the spin states of two electrons, despite the fact that each spin state is defined on the same basis, it is necessary to distinguish which electron is in which state and so each is attributed to their own subsystem. Starting from a space $\tilde{A}$, two different subsystems are referred to as $\tilde{A}_1$ and $\tilde{A}_2$. If different words are described in the same space, subsystems can be used to formally assign them to different spaces. The permutation operation extends naturally from the one in standard quantum mechanics. We define two permutation operators: $P^{\tilde{A}_1 \tilde{A}_2}$ permutes the elements of the basis of the respective spaces, while $P_{\tilde{A}_1 \tilde{A}_2}$ permutes the elements of the dual basis. If only one set of basis elements is inside the scope of the permutation operators, then either the subsystem assignment changes,

$$P^{\tilde{A}_1 \tilde{A}_2} \left|i\right\rangle_{\tilde{A}_1} \left\langle i'\right| P^{\tilde{A}_1 \tilde{A}_2} = \left|i\right\rangle_{\tilde{A}_2} \left\langle i'\right|; \qquad P_{\tilde{A}_1 \tilde{A}_2} \left|i'\right\rangle_{\tilde{A}_1^*} \left\langle i\right| P_{\tilde{A}_1 \tilde{A}_2} = \left|i'\right\rangle_{\tilde{A}_2^*} \left\langle i\right|; \qquad (39)$$

or the respective space of tracing changes,

$$\mathrm{Tr}_{\tilde{A}_1}\left(P_{\tilde{A}_1\tilde{A}_2}\left|i'\right\rangle_{\tilde{A}_2}{}^*\langle i|P_{\tilde{A}_1\tilde{A}_2}\right) = \mathrm{Tr}_{\tilde{A}_2}\left(\left|i'\right\rangle_{\tilde{A}_2}{}^*\langle i|\right). \tag{40}$$

Note that permutations take precedence over traces. If two words are assigned to different subsystems, the permutations act to swap their space assignment:[9]

$$P^{\tilde{A}_1\tilde{A}_2}\left|i\right\rangle_{\tilde{A}_1}\langle i'|\otimes|j\rangle_{\tilde{A}_2}\langle j'|P^{\tilde{A}_1\tilde{A}_2} = \left|i\right\rangle_{\tilde{A}_2}\langle i'|\otimes|j\rangle_{\tilde{A}_1}\langle j'|, \tag{41}$$

$$P_{\tilde{A}_1\tilde{A}_2}\left|i'\right\rangle_{\tilde{A}_1}{}^*\left\langle i\right|\otimes\left|j'\right\rangle_{\tilde{A}_2}{}^*\left\langle j\right|P_{\tilde{A}_1\tilde{A}_2} = \left|i'\right\rangle_{\tilde{A}_2}{}^*\left\langle i\right|\otimes\left|j'\right\rangle_{\tilde{A}_1}{}^*\left\langle j\right|. \tag{42}$$

If no word has that subsystem assignment then the permutation has no effect.

5 Interpreting Lambek Calculus derivations

Let us turn now to the syntax-semantics interface, which takes the form of a homomorphism sending the types and derivations of the syntactic front end $\mathbf{(N)L}_{/,\backslash}$ to their semantic counterparts. Consider first the action of the interpretation homomorphism on *types*. We write $\lceil.\rceil$ for the map that sends syntactic types to the interpreting semantic spaces. For primitive types we set

$$\lceil s \rceil = \tilde{S}, \ \lceil np \rceil = \lceil n \rceil = \tilde{N}, \tag{43}$$

with S the vector space for sentence meanings and N the space for nominal expressions (common nouns, full noun phrases). For compound types we have

$$\lceil A/B \rceil = \lceil A \rceil \otimes \lceil B \rceil^*, \ \text{and} \ \lceil A\backslash B \rceil = \lceil A \rceil^* \otimes \lceil B \rceil. \tag{44}$$

Given semantic spaces for the syntactic types, we can turn to the interpretation of the syntactic *derivations*, as coded by their $\lambda_{/,\backslash}$ proof terms. We write $[\![\cdot]\!]_g$ for the map that associates each term t of type A with a semantic value, i.e. an element of $\lceil A \rceil$, the semantic space where meanings of type A live. The map $[\![.]\!]$ is defined relative to a assignment function g that provides a semantic value for the basic building blocks, viz. the variables that label the axiom leaves of a proof. As we saw

---

[9]We define this as a shorthand application of the permutation operations as defined in eq.39, such that eq.41 can be calculated w.r.t. that definition as

$$P^{\tilde{A}_1\tilde{A}_2}\left|i\right\rangle_{\tilde{A}_1}\left(_{\tilde{A}_1}\langle i'|P^{\tilde{A}_1\tilde{A}_2}\right)\otimes\left(P^{\tilde{A}_1\tilde{A}_2}\left|j\right\rangle_{\tilde{A}_2}\right)_{\tilde{A}_2}\langle j'|P^{\tilde{A}_1\tilde{A}_2}$$

$$= P^{\tilde{A}_1\tilde{A}_2}\left|i\right\rangle_{\tilde{A}_1}{}_{\tilde{A}_2}\langle i'|\otimes|j\rangle_{\tilde{A}_1}{}_{\tilde{A}_2}\langle j'|P^{\tilde{A}_1\tilde{A}_2} = \left|i\right\rangle_{\tilde{A}_2}\langle i'|\otimes|j\rangle_{\tilde{A}_1}\langle j'|,$$

and similarly for eq.42.

above, a proof term is a generic meaning recipe that abstracts from particular lexical meanings. Specific lexical items, as we will see in §6, have the status of *constants*. These constants are mapped to their distributional meaning by an interpretation function I. The distributional meaning corresponds to the embeddings assigned by a particular model to the lexicon. Below we show that this calculus is sound with respect to the semantics of section 4.

Axiom

$$\left[\!\left[x^A\right]\!\right]_g = g(x^A) = \rho_x^{\lceil A\rceil} = \sum_{ii'} X^{ii'} \, |i\rangle_{\lceil A\rceil}\langle i'|. \tag{45}$$

Elimination   Recall the inference rules of fig.1
$E_/$: Premises $t^{B/A}$, u^A; conclusion $(t \triangleleft u)^B$:

$$\left[\!\left[(t \triangleleft u)^B\right]\!\right]_g \equiv \mathrm{Tr}_{\lceil A\rceil}\left(\left[\!\left[t^{B/A}\right]\!\right]_g \cdot \left[\!\left[u^A\right]\!\right]_g\right) \tag{46}$$

$$= \mathrm{Tr}_{\lceil A\rceil}\left(\sum_{ii',jj'} T^{ii'}_{j'j}\left|^{j'}_i\right\rangle_{\lceil B\rceil\otimes\lceil A\rceil^*}\left\langle^j_{i'}\right| \cdot \sum_{kk'} U^{kk'}|k\rangle_{\lceil A\rceil}\langle k'|\right) \tag{47}$$

$$= \sum_{ii',jj'}\sum_{kk'} T^{ii'}_{j'j} \cdot U^{kk'}\,\delta^j_k\delta^{j'}_{k'}|i\rangle_{\lceil B\rceil}\langle i'| = \sum_{ii',jj'} T^{ii'}_{j'j} \cdot U^{jj'}|i\rangle_{\lceil B\rceil}\langle i'|. \tag{48}$$

$E_\backslash$: Premises u^A, $t^{A\backslash B}$; conclusion $(u \triangleright t)^B$:

$$\left[\!\left[(u \triangleright t)^B\right]\!\right]_g \equiv \mathrm{Tr}_{\lceil A\rceil}\left(\left[\!\left[u^A\right]\!\right]_g \cdot \left[\!\left[t^{A\backslash B}\right]\!\right]_g\right) \tag{49}$$

$$= \mathrm{Tr}_{\lceil A\rceil}\left(\sum_{kk'} U^{kk'}|k\rangle_{\lceil A\rceil}\langle k'| \cdot \sum_{ii',jj'} T_{jj'}^{\ ii'}\left|^{j'}_i\right\rangle_{\lceil A\rceil^*\otimes\lceil B\rceil}\left\langle^j_{i'}\right|\right) = \tag{50}$$

$$= \sum_{kk'}\sum_{ii',jj'} U^{kk'} \cdot T_{j'j}^{\ ii'}\,\delta^j_k\delta^{j'}_{k'}|i\rangle_{\lceil B\rceil}\langle i'| = \sum_{ii',jj'} U^{jj'} \cdot T_{j'j}^{\ ii'}|i\rangle_{\lceil B\rceil}\langle i'|. \tag{51}$$

Introduction $I_/$: Premise t^B, with x^A as its rightmost parameter; conclusion $(\lambda^r x.t)^{B/A}$:

$$\left[\!\left[(\lambda^r x.t)^{B/A}\right]\!\right]_g \equiv \sum_{kk'}\left(\left[\!\left[t^B\right]\!\right]_{g^x_{kk'}} \otimes |k'\rangle_{\lceil A\rceil^*}\langle k|\right) \tag{52}$$

$I_\backslash$: Premise t^B, with x^A as its leftmost parameter; conclusion $(\lambda^l x.t)^{A\backslash B}$:

$$\left[\!\!\left[\left(\lambda^l x.t\right)^{A\backslash B}\right]\!\!\right]_g \equiv \sum_{kk'} \left(\left|k'\right\rangle_{\lceil A\rceil^*} \left\langle k\right| \otimes [\![t^B]\!]_{g^x_{kk'}}\right) \tag{53}$$

Here $g^x_{kk'}$ is the assignment exactly like g except possibly for the parametric variable x which takes the value of the basis element $|k\rangle_{\lceil A\rceil}\langle k'|$. More generally, the interpretation of the introduction rules lives in a compound density matrix space representing a linear map from $\tilde{A}$ to $\tilde{B}$. The semantic value of that map, applied to any object $m \in \tilde{A}$, is given by $[\![t^B]\!]_{g'}$, where g' is the assignment exactly like g except possibly for the bound variable x^A, which is assigned the value m. Note that now, given the introduction of the metric, the interpretations of A/B and $B\backslash A$ are related by it: if the components of the first are $T_J{}^I$, then those of the second are given by those in eq.16 adapted for density matrices. This is what introduces directionality in our interpretation: using the metric, we can extract a certain representation for a function word and distinguish by the values of the components whether it will contract from the left or from the right.

6 Derivational Ambiguity

The density matrix construction has been successfully used to address lexical ambiguity [22], as well as lexical and sentence entailment [1,24], where different measures of entropy are used to perform the disambiguation. Here we arrive at disambiguation in a different way, by storing in the diagonal elements of a higher order density matrix the different interpretations that result from the different contractions that the proof-as-programs prescribes. This is possible due to the the set-up that is formed by a multi-partite density matrices space, so that, by making use of permutation operations, it happens automatically that the two meanings are expressed independently. This is useful because it can be integrated with a lexical interpretation in density matrices optimized to other tasks, such as lexical ambiguity or entailment. It is also appropriate to treat the existence of these ambiguities in the context of incrementality, since it keeps the meanings separated in their interaction with posterior fragments.

We give a simple example of how the trace machinery can be used on an ambiguous fragment, providing a passage from one reading to the other at the interpretation level, and how the descriptions are kept separated. For this application, the coefficients in the interpretation of the words contain distributional information harvested from data, either from a count-base model or a more sophisticated language model.

The final coefficient of each outcomes is the vector-based representation of that reading.

We illustrate the construction with the phrase "tall person from Spain". The lexicon below has the syntactic type assignments and the corresponding semantic spaces.

	syn type A	$\lceil A \rceil$
tall	n/n	$N^* \otimes N \otimes (N^* \otimes N)^*$
person	n	$N^* \otimes N$
from	$(n \backslash n)/np$	$(N^* \otimes N)^* \otimes N^* \otimes N \otimes (N^* \otimes N)^*$
Spain	np	$N^* \otimes N$

Given this lexicon, "tall person from Spain" has two derivations, corresponding to the bracketings "(tall person) from Spain" ($x/tall, y/person, w/from, z/Spain$):

$$\cfrac{\cfrac{\overline{x : n/n \vdash x : n/n}^{\ ax} \quad \overline{y : n \vdash y : n}^{\ ax}}{(x : n/n, y : n) \vdash (x \triangleleft y) : n}^{/E_2} \quad \cfrac{\overline{w : (n\backslash n)/np \vdash w : (n\backslash n)/np}^{\ ax} \quad \overline{z : np \vdash z : np}^{\ ax}}{(w : (n\backslash n)/np, z : n) \vdash (w \triangleleft z) : n\backslash n}^{/E_1}}{[(x : n/n, y : n), (w : (n\backslash n)/np, z : n)] \vdash ((x \triangleleft y) \triangleright (w \triangleleft z)) : n}^{\backslash E_3}$$

versus "tall (person from Spain)":

$$\cfrac{\overline{x : n/n \vdash x : n/n}^{\ ax} \quad \cfrac{\overline{y : n \vdash y : n}^{\ ax} \quad \cfrac{\overline{w : (n\backslash n)/np \vdash w : (n\backslash n)/np}^{\ ax} \quad \overline{z : np \vdash z : np}^{\ ax}}{(w : (n\backslash n)/np, z : n) \vdash (w \triangleleft z) : n\backslash n}^{/E_1}}{[y : n, (w : (n\backslash n)/np, z : n)] \vdash (y \triangleright (w \triangleleft z)) : n}^{\backslash E_2}}{(x : n/n, [y : n, (w : (n\backslash n)/np, z : n)]) \vdash (x \triangleleft (y \triangleright (w \triangleleft z))) : n}^{/E_3}$$

In the first reading, the adjective "tall" is evaluated with respect to all people, before it is specified that this person happens to be from Spain, whereas in the second reading the adjective "tall" is evaluated only in the restricted universe of people from Spain.

Taking "from Spain" as a unit for simplicity, let us start with the following primitive interpretations:

- $[\![tall^{n/n}]\!]_I = \sum_{ii',jj'} \mathbf{T}_{ii'}^{j'j} \left| {}_{j'}^{i} \right\rangle_{N \otimes N^*} \left\langle {}_{j}^{i'} \right|,$

- $[\![person^n]\!]_I = \sum_{kk'} \mathbf{P}_{kk'} \left| {}^{k} \right\rangle_{N} \left\langle {}^{k'} \right|,$

- $[\![from\_Spain^{n\backslash n}]\!]_I = \sum_{ll',mm'} \mathbf{F}_{mm'}^{l'l} \left| {}_{l''}^{m} \right\rangle_{N^* \otimes N} \left\langle {}_{l}^{m'} \right|.$

Interpreting each step of the derivation in the way described in the previous section will give two different outcomes. The first one is

$$[\![tall\_person\_from\_Spain^n]\!]^1_I =$$

$$= \mathrm{Tr}_N \left(\mathrm{Tr}_N \left(\sum_{ii',jj'} \mathbf{T}^{ii'}_{j'j} \left| ^{j'}_i \right\rangle_{N \otimes N^*} \left\langle ^j_{i'} \right| \cdot \sum_{kk'} \mathbf{P}^{kk'} \left| k \right\rangle_N \left\langle k' \right| \right) \right.$$

$$\left. \cdot \sum_{ll',mm'} \mathbf{F}_{l'l}^{\ mm'} \left| ^{l'}_m \right\rangle_{N^* \otimes N} \left\langle ^l_{m'} \right| \right)$$

$$= \sum_{ii',jj',mm'} \mathbf{T}^{ii'}_{j'j} \, \mathbf{P}^{jj'} \, \mathbf{F}_{i'i}^{\ mm'} \left| m \right\rangle_N \left\langle m' \right|, \tag{54}$$

while the second one is

$$[\![tall\_person\_from\_Spain^n]\!]^2_I =$$

$$= \mathrm{Tr}_N \left(\sum_{ii',jj'} \mathbf{T}^{ii'}_{j'j} \left| ^{j'}_i \right\rangle_{N \otimes N^*} \left\langle ^j_{i'} \right| \cdot \mathrm{Tr}_N \left(\sum_{kk'} \mathbf{P}_{kk'} \left| k \right\rangle_N \left\langle k' \right| \right. \right.$$

$$\left. \left. \cdot \sum_{ll',mm'} \mathbf{F}_{l'l}^{\ mm'} \left| ^{l'}_m \right\rangle_{N^* \otimes N} \left\langle ^l_{m'} \right| \right) \right)$$

$$= \sum_{ii',jj',ll'} \mathbf{T}^{ii'}_{j'j} \, \mathbf{P}^{ll'} \, \mathbf{F}_{l'l}^{\ jj'} \left| i \right\rangle_N \left\langle i' \right|. \tag{55}$$

The respective graphical representations of these contractions can be found in fig.2. Though the coefficients might be different for each derivation, it is not clear how both interpretations are carried separately if they are part of a larger fragment, since their description takes place on the same space. Also, this recipe gives a fixed ordering and range for each trace. To be able to describe each final meaning separately, we use here the concept of *subsystem*. Because different subsystems act formally as different syntactic types and in each derivation the words that interact are different, it follows that each word should be assigned to a different subsystem:

- $[\![tall^{n/n}]\!]_{I_1} = [\![tall^{n/n}]\!]_{I_2} = \sum_{ii',jj'} \mathbf{T}^{ii'}_{j'j} \left| ^{j'}_i \right\rangle_{N^1 \otimes N^{2*}} \left\langle ^j_{i'} \right|,$
- $[\![person^n]\!]_{I_1} = \sum_{kk'} \mathbf{P}^{kk'} \left| k \right\rangle_{N^2} \left\langle k' \right|,$
  $[\![person^n]\!]_{I_2} = \sum_{kk'} \mathbf{P}^{kk'} \left| k \right\rangle_{N^3} \left\langle k' \right|,$
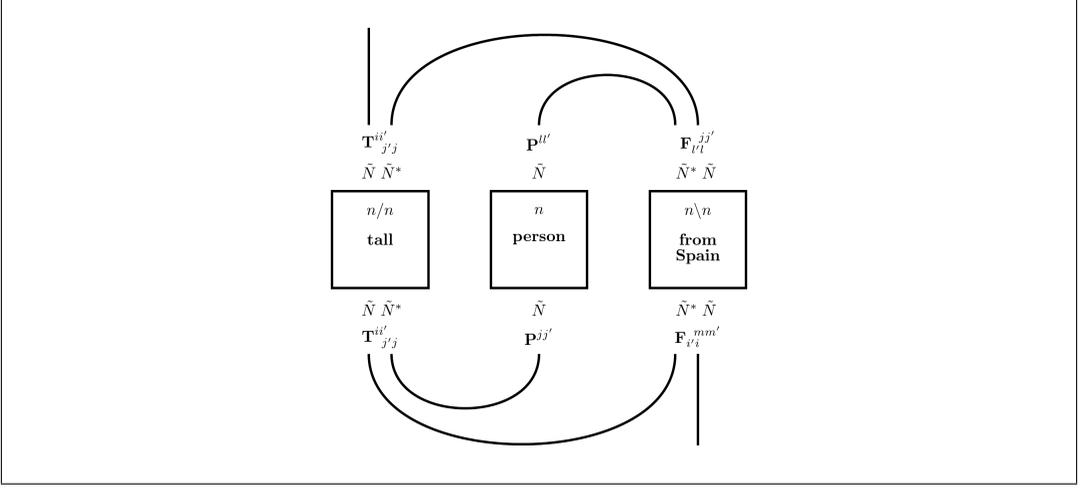
Figure 2: Representation of contractions corresponding to the first reading (lower links) and to the second reading (upper links), without subsystems. The final value is a coefficient in the $\tilde{N}$ space as in eq.54 and in eq.55, respectively.

- $[\![from\_Spain^{n\backslash n}]\!]_{I_1} = \sum_{ll',mm'} \mathbf{F}_{l'l}^{mm'} \left|\begin{smallmatrix} l' \\ m \end{smallmatrix}\right\rangle_{N^{1*}\otimes N^3} \left\langle\begin{smallmatrix} l \\ m' \end{smallmatrix}\right|,$

  $[\![from\_Spain^{n\backslash n}]\!]_{I_2} = \sum_{ll',mm'} \mathbf{F}_{l'l}^{mm'} \left|\begin{smallmatrix} l' \\ m \end{smallmatrix}\right\rangle_{N^{3*}\otimes N^2} \left\langle\begin{smallmatrix} l \\ m' \end{smallmatrix}\right|.$

Notice that the value of the coefficients given by the interpretation functions I_1 and I_2 that describe the words does not change from the ones given in I, only possibly the subsystem assignment does. Rewriting the derivation of the interpretations in terms of subsystems, the ordering of the traces does not matter anymore since the contraction is restricted to its own subsystem. For the first reading we obtain

$$[\![tall\_person\_from\_Spain^n]\!]^1_{I_1} =$$

$$= \mathrm{Tr}_{N^1}\left(\mathrm{Tr}_{N^2}\left(\sum_{ii',jj'} \mathbf{T}^{ii'}_{j'j}\left|\begin{smallmatrix} j' \\ i \end{smallmatrix}\right\rangle_{N^1\otimes N^{2*}}\left\langle\begin{smallmatrix} j \\ i' \end{smallmatrix}\right| \cdot \sum_{kk'} \mathbf{P}^{kk'}\left|k\right\rangle_{N^2}\left\langle k'\right|\right.\right.$$

$$\left.\left.\cdot \sum_{ll',mm'} \mathbf{F}_{l'l}^{mm'}\left|\begin{smallmatrix} l' \\ m \end{smallmatrix}\right\rangle_{N^{1*}\otimes N^3}\left\langle\begin{smallmatrix} l \\ m' \end{smallmatrix}\right|\right)\right)$$

$$= \sum_{ii',jj',mm'} \mathbf{T}^{ii'}_{j'j}\,\mathbf{P}^{jj'}\,\mathbf{F}_{i'i}^{mm'}\left|m\right\rangle_{N^3}\left\langle m'\right| \tag{56}$$

and for the second

$$[\![tall\_person\_from\_Spain^n]\!]_{I_2}^2 =$$

$$= \mathrm{Tr}_{N^2}\left(\sum_{ii',jj'}\mathbf{T}^{ii'}_{j'j}\left|{}^{j'}_i\right\rangle_{N^1\otimes N^{2*}}\left\langle{}^{j}_{i'}\right| \cdot \mathrm{Tr}_{N^3}\left(\sum_{kk'}\mathbf{P}^{kk'}\left|k\right\rangle_{N^3}\left\langle k'\right|\right.\right.$$

$$\left.\left.\cdot \sum_{mm',ll'}\mathbf{F}^{mm'}_{l'l}\left|{}^{l'}_m\right\rangle_{N^{3*}\otimes N^2}\left\langle{}^{l}_{m'}\right|\right)\right)$$

$$= \mathrm{Tr}_{N^3}\left(\mathrm{Tr}_{N^2}\left(\sum_{ii',jj'}\mathbf{T}^{ii'}_{j'j}\left|{}^{j'}_i\right\rangle_{N^1\otimes N^{2*}}\left\langle{}^{j}_{i'}\right| \cdot \sum_{kk'}\mathbf{P}^{kk'}\left|k\right\rangle_{N^3}\left\langle k'\right|\right.\right.$$

$$\left.\left.\cdot \sum_{ll',mm'}\mathbf{F}^{mm'}_{l'l}\left|{}^{l'}_m\right\rangle_{N^{3*}\otimes N^2}\left\langle{}^{l}_{m'}\right|\right)\right)$$

$$= \sum_{ii',jj',ll'}\mathbf{T}^{ii'}_{j'j}\,\mathbf{P}^{ll'}\,\mathbf{F}^{jj'}_{l'l}\left|i\right\rangle_{N^1}\left\langle i'\right|. \tag{57}$$

The interpretation of each derivation belongs now to different subsystems, which keeps the information about the original word to which the free "noun" space is attached. We can see this by comparing the upper and lower links in fig.3.

However, it is not very convenient to attribute each word to a different subsystem depending on the interpretation it will be part of, since that is information that comes from the derivation itself and not from the representations of words. To tackle this problem, one uses permutation operations over the subsystems. Since these have precedence over the trace, when the traces are taken the contractions change accordingly. This changes the subsystem assignment at specific points so it is possible to go from one interpretation to the other, without giving different interpretations to each word initially. Thus, there is a way to go directly from the first interpretation to the second:
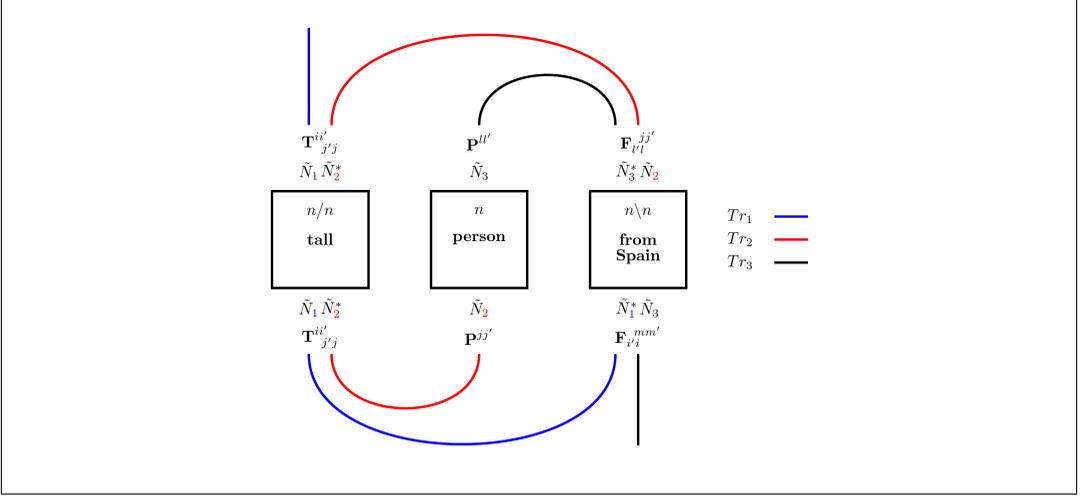
Figure 3: Representation of contractions corresponding to the first reading (lower links) and to the second reading (upper links), with subsystems. The final value is a coefficient in the $\tilde{N}$ space as in eq.56 and in eq.57, respectively.

$$
[\![tall\_person\_from\_Spain^n]\!]^2_{I_1} =
$$

$$
= \text{Tr}_{N^1} \left(P_{13} \, \text{Tr}_{N^2} \left(\sum_{ii',jj'} \mathbf{T}^{ii'}_{j'j} \left| {j' \atop i} \right\rangle_{N^1 \otimes N^{2*}} \left\langle {j \atop i'} \right| \cdot P_{13} P^{23} \sum_{kk'} \mathbf{P}^{kk'} \left| k \right\rangle_{N^2} \left\langle k' \right| \right.\right.
$$

$$
\left.\left. \cdot \sum_{ll',mm'} \mathbf{F}_{l'l}^{\,mm'} \left| {l' \atop m} \right\rangle_{N^{1*} \otimes N^3} \left\langle {l \atop m'} \right| P^{23} P_{13} \right) P_{13} \right)
$$

$$
= \text{Tr}_{N^3} \left(\text{Tr}_{N^2} \left(\sum_{ii',jj'} \mathbf{T}^{ii'}_{j'j} \left| {j' \atop i} \right\rangle_{N^1 \otimes N^{2*}} \left\langle {j \atop i'} \right| \cdot \sum_{kk'} \mathbf{P}^{kk'} \left| k \right\rangle_{N^3} \left\langle k' \right| \right.\right.
$$

$$
\left.\left. \cdot \sum_{ll',mm'} \mathbf{F}_{l'l}^{\,mm'} \left| {l' \atop m} \right\rangle_{N^{3*} \otimes N^2} \left\langle {l \atop m'} \right| \right) \right). \tag{58}
$$

The reasoning behind this is as follows: the permutation P^{23} swaps the space assignment between that of "person" and the free space in "from_Spain", according to eq.42; after that a permutation P_{13} is used as in eq.39 to change the argument space of "from_Spain" from N^{1*} to N^{3*}, and then the same permutation is applied again to change the space of tracing, following eq.40. In this way, all the coefficients
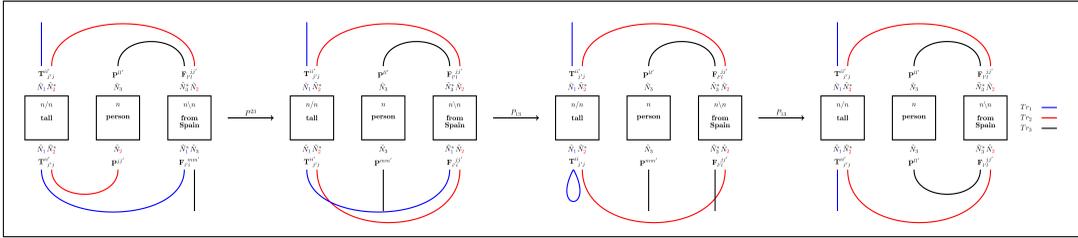
Figure 4

will have the correct contractions and in a different space from the first reading. In fig.4 we can see the action of the permutations by visualizing how both the spaces and the traces change as we go from the lower to the upper links.

Although the metric is not used explicitly in the application of the permutation operators, it is necessary to generate the correct tensors where the permutation operator is applied in the first place, by going from the vector representation that comes directly from the data to one that allows contraction. As an example, the adjective "tall" would have a vector representation from the data as an element of $\tilde{V} \otimes \tilde{V}$, of the form $\mathbf{T}^{ii', kk'}$. We need the metric $d_{kj'} d_{k'j}$ to change its form to $T^{ii'}_{j'j}$. By defining the interpretation space of adjectives as $\tilde{N} \otimes \tilde{N}^*$, we assume this passage has already been made when we assign an interpretation to a word in this space. As an alternative to this derivation, we mention that it is possible to apply a P^{23} permutation followed by a P^{13} permutation that results in the correct contraction of the indices, but fails to deliver the results of the two derivations in different subspaces; it is however noteworthy that, in order to start with a unique assignment for each word, this alternative derivation can, in any case, only be achieved by distinguishing between subsystems, as well as the covariant and contravariant indices.

7 Conclusion and Future Work

In this paper we provided a density matrix model for a simple fragment of the Lambek Calculus, differently from what is done in [2] who uses density matrices to interpret dependency parse trees. The syntax-semantics interface takes the form of a compositional map assigning semantic values to the $\lambda_{/,\backslash}$ terms coding syntactic derivations. We proposed the use of a metric as a way to reconcile the various vector representations of the same word that come from different treatments, assuming that there is a quantity that is being preserved, such as human judgements. If we know the metric, we can confidently assign only one embedding to each word as its semantic value. A metric can relate these various representations so that we can assign

only one vector as its semantic value. The density matrix model enables the integration of lexical and derivational forms of ambiguity. Additionally, it allows for the transfer of methods and techniques from quantum mechanics and general relativity to computational semantics. One example of such transfer is the permutation operator. In quantum mechanics, this operator permits a description of indistinguishable particles. In the linguistic application, it allows one to go from an interpretation that comes from one derivation to another, without the need to to go through the latter, but keeping this second meaning in a different subsystem. Another example is the introduction of covariant and contravariant components, associated with a metric, that allow the permutation operations to be properly applied. In future work, we want to explore the preservation of human judgements found in the literature via a metric that represents the variability of vector representations of words, either static or dynamic. We also want to extend our simple fragment with modalities for structural control (cf [17]), in order to deal with cases of derivational ambiguity that are licensed by these control modalities. Finally, we want to consider derivational ambiguity in the light of an *incremental* left-to-right interpretation process, so as to account for the evolution of interpretations over time. In enriching the treatment with a metric, we want to explore the consequences of having this new parameter in treating context dependent embeddings.

Acknowledgements

References

[1] Dea Bankova, Bob Coecke, Martha Lewis, and Dan Marsden. Graded hyponymy for compositional distributional semantics. *Journal of Language Modelling*, 6(2):225–260, 2019.

[2] William Blacoe. Semantic composition inspired by quantum measurement. In *International Symposium on Quantum Interaction*, pages 41–53. Springer, 2014.

[3] Elia Bruni, Nam-Khanh Tran, and Marco Baroni. Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, 49:1–47, 2014.

[4] Stephen Clark. Vector space models of lexical meaning. *Handbook of Contemporary Semantics*, 10:9781118882139, 2015.

[5] Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. Mathematical foundations for a compositional distributional model of meaning. *Lambek Festschrift, Linguistic Analysis 36(1–4)*, pages 345–384, 2010.

[6] AD Correia, HTC Stoof, and M Moortgat. Putting a spin on language: A quantum interpretation of unary connectives for linguistic applications. *arXiv preprint arXiv:2004.04128*, 2020.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pretraining of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[8] Kees Dullemond and Kasper Peeters. Introduction to tensor calculus. *Kees Dullemond and Kasper Peeters*, 1991.

[9] Howard Georgi. Lie algebras in particle physics. from isospin to unified theories. *Front. Phys.*, 54:1–255, 1982.

[10] Edward Grefenstette and Mehrnoosh Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404. Association for Computational Linguistics, 2011.

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

[12] Theo M.V. Janssen and Barbara H. Partee. Chapter 7 - Compositionality. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 417 – 473. North-Holland, Amsterdam, 1997.

[13] Joachim Lambek. The mathematics of sentence structure. *The American Mathematical Monthly*, 65(3):154–170, 1958.

[14] Joachim Lambek. On the calculus of syntactic types. In Roman Jakobson, editor, *Structure of Language and its Mathematical Aspects*, volume XII of *Proceedings of Symposia in Applied Mathematics*, pages 166–178. American Mathematical Society, 1961.

[15] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[16] Jeff Mitchell and Mirella Lapata. Composition in distributional models of semantics. *Cognitive science*, 34(8):1388–1429, 2010.

[17] Michael Moortgat. Chapter 2 - Categorial type logics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 93–177. Elsevier, Amsterdam, 1997.

[18] Richard Moot and Christian Retoré. *The logic of categorial grammars: a deductive account of natural language syntax and semantics*, volume 6850. Springer, 2012.

[19] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[20] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[21] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[22] Robin Piedeleu. *Ambiguity in categorical models of meaning*. PhD thesis, University of Oxford Master's thesis, 2014.

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

[24] Mehrnoosh Sadrzadeh, Dimitri Kartsaklis, and Esma Balkir. Sentence entailment in compositional distributional semantics. *Ann. Math. Artif. Intell.*, 82(4):189–218, 2018.

[25] Lorenzo Adlai Sadun. *Applied linear algebra: The decoupling principle*. American Mathematical Soc., 2007.

[26] Peter Selinger. Dagger compact closed categories and completely positive maps. *Electronic Notes in Theoretical computer science*, 170:139–163, 2007.

[27] Johan van Benthem. The semantics of variety in categorial grammar. Technical Report 83-29, Simon Fraser University, Burnaby (B.C.), 1983. Revised version in W. Buszkowski, W. Marciszewski and J. van Benthem (eds) Categorial grammar, Benjamin, Amsterdam.

[28] Robert M Wald. General relativity. *University of Chicago Press*, 1984.

[29] Heinrich Wansing. Formulas-as-types for a hierarchy of sublogics of intuitionistic propositional logic. In David Pearce and Heinrich Wansing, editors, *Nonclassical Logics and Information Processing*, pages 125–145, Berlin, Heidelberg, 1992. Springer Berlin Heidelberg.

[30] 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, 2019.