A proof-theoretic approach
to scope ambiguity
in compositional vector space models

Gijs Jasper Wijnholds
School of Electronic Engineering and Computer Science,
Queen Mary University of London

ABSTRACT

We investigate the extent to which compositional vector space models can be used to account for scope ambiguity in quantified sentences (of the form *Every man loves some woman*). Such sentences containing two quantifiers introduce two readings, a direct scope reading and an inverse scope reading. This ambiguity has been treated in a vector space model using bialgebras by Hedges and Sadrzadeh (2016) and Sadrzadeh (2016), though without an explanation of the mechanism by which the ambiguity arises. We combine a polarised focussed sequent calculus for the non-associative Lambek calculus **NL**, as described in Moortgat and Moot (2011), with the vector-based approach to quantifier scope ambiguity. In particular, we establish a procedure for obtaining a vector space model for quantifier scope ambiguity in a derivational way.

1   INTRODUCTION

There is a long standing tradition in formal semantics on compositionality: in order to separate the meaning of basic elements (lexical semantics) from the construction of higher-level meaning (derivational semantics) one assigns a homomorphism from a *syntactic algebra* to a *semantic algebra.* Having been rigorously formalised by Montague in his seminal papers (Montague 1970, 1973), these ideas have been made concrete in the field of grammar, where syntactic types are

mapped onto semantic types so that any derivation gives rise to a *meaning recipe*. Traditionally, meaning is taken to be a linear lambda term that evaluates to a truth value.

Ongoing research on distributional semantics, based on the idea that word meaning is defined relative to a word's context, has revealed an appealing way to incorporate type-logical grammar into distributional models (Coecke *et al.* 2010). This approach, also known as the DisCoCat approach (Distributional Compositional Categorical models), treats compositionality in the Montagovian style as a functorial passage from syntactic types and proofs to vectors and linear maps. Given that this line of research is still in its early phase, there is much to be done to formalise details of the model, give accounts for semantic phenomena, and evaluate the effectiveness of the chosen approach.

Though traditional categorial syntax and semantics go hand in hand, some aspects of the set-theoretic formal semantics get lost in the switch to a vector space model of meaning. First, the interpretation of constants that one can appeal to in formal semantics are not directly available in a vector-based setting; a logical word like "not" can be computed in the formal setting by taking set complement, but negating a vector or matrix is not trivial.[1] Similarly, for coordinators like "and" and "or" the standard set intersection and union are not available in a vectorial setting. One could replace intersection by vector multiplication and union by vector summation, but in the presence of concrete distributional vectors it is not clear that such operations indeed perform well in an experimental setting. Second, the DisCoCat approach assumes a tight categorical correspondence between a syntactic formalism and the concrete vector semantics: when we want to stay in the realm of finite dimensional vector spaces, we are dealing with a compact closed category; to model a categorial grammar as a category, one needs to fully explicate its proof-theoretic logical and structural rules, an exposition that is not trivially available for any categorial system.[2] Another issue with this categorical treatment is

---

[1] Although there is work on simulating negation in a tensor-based setting (Grefenstette 2013), it is not clear what negation really means in a distributional setting. For instance, an alternative view is to treat distributional negation as conversational (Kruszewski *et al.* 2016).

[2] For instance, the composition and type-raising combinators one finds in Combinatorial Categorial Grammar (Steedman 2000) don't easily translate into

that a simple vector-based model does not have the non-linearity that some models would assume. As an example, allowing a non-linearity in lexical lambda terms or as a syntactic mechanism means the copying of material which is not possible with all vectors. We discuss this issue in more detail in the rest of the paper.

Some of the above issues have been addressed in recent work by Sadrzadeh *et al.* (2013), Hedges and Sadrzadeh (2016), Sadrzadeh (2016), giving accounts of subject/object relativisation, generalised quantifiers, and quantifier scope. In Sadrzadeh *et al.* (2013), the meaning of pronoun relative clauses is explained by using Frobenius algebras in the lexicon, and assigning different pregroup grammar types to the subject relative pronoun "who" and the object relative pronoun "whom". Two different derivations then naturally arise, giving an intersectional meaning to subject relative clauses of the type "Men who like Mary", and object relative clauses such as "Men whom Mary likes". Such an approach does not lend itself to certain Germanic languages where the ambiguity has to be derivational: in Dutch, the subject relative and object relative interpretations above share the surface form "Mannen die Marie mogen".[3] To deal with this issue without specifying lexical alternatives, i.e. different possible typings of the relative pronoun "die", Moortgat and Wijnholds (2017) provide a derivational account that results in the same intersective vector space meaning as the ones of Sadrzadeh *et al.* (2013).

An element that lacks in the results obtained so far on quantifier scope ambiguity is a detailed discussion of the derivational process, giving rise to ambiguities. Quantifier scope ambiguity as opposed to pronoun relativisation is more pressing as the former exists in English and does not come from the lexicon, but rather from different ways of reading the same surface form. The account of Hedges and Sadrzadeh (2016) explores the use of bialgebras to represent quantifiers, using context free grammars as the syntactic engine; its follow up (Sadrzadeh 2016) discusses scope ambiguity but assumes the ambiguity to be given before detailing the direct scope and inverse scope readings of phrases of the shape "Every man loves some woman". In

---

a standard category, and the Displacement Calculus of Morrill *et al.* (2011) subsumes its structural rules in the rules of the system (Valentín 2014).

[3] "Die" can mean "who", "whom", "that", "mogen" is an inflection of "like".

order to explain how the ambiguity comes about, we need to detail the syntactic process, and integrate it with a vector-based semantics.

Our goal in this paper, then, is to pave the way to fully explain compositionality in vector space models of meaning while also taking into account the desirable mechanisms of e.g. Frobenius algebras and bialgebras. The contribution of this paper is to show how we can represent quantifier scope ambiguity in a derivational manner, fully determined by the syntactic process combined with a suitable lexical semantics.

We will make use of a polarised non-associative Lambek calculus, and use focussing as a technique to gain control over the space of sequent derivations. A *continuation-passing-style* translation from syntactic types into semantic objects then gives rise to the expected reading for quantifier scope ambiguity. This technique has been worked out by Moortgat and Moot (2011) (following Bernardi and Moortgat 2010 and Bastenhof 2012), but has not, until now, been put in the context of vector space models.

This paper is structured as follows: in Section 2, we briefly discuss quantifier scope ambiguity and its apparent non-linearity. Next, in Section 3 we define the basic, compositional DisCoCat model. We proceed to review quantifier scope ambiguity in vector space models in Section 4, and show in Section 5 how we can derive quantifier scope ambiguity in a compositional way using a polarised focussed sequent calculus that is interpreted in a vector space model. We conclude in Section 6 by explaining how our results can be further expanded and we introduce some potential new areas of investigation.

2   QUANTIFIER SCOPE AMBIGUITY

There seems to be an intrinsic non-linearity associated with quantifiers. Consider the word "all" in a phrase "all men sleep". One way of modelling the universal quantification in the phrase is to let "all" refer to an operation that decides whether the set of "men" is a subset of those entities that are sleeping, i.e. if "men" refers to some set A, and "sleep" to some set B, then "all men sleep" computes whether $A \subseteq B$. This can be given an alternative definition:

$$[all](A)(B) = \begin{cases} 1 & \text{if } A = A \cap B \\ 0 & \text{otherwise} \end{cases}$$

When one tries to give this interpretation in terms of a λ-term, the usual approach is to model both "men" and "sleep" as a *characteristic function* of a set of entities, where "all" is given a non-linear λ-term:

$$[\![all]\!] = \lambda P.\lambda Q.(\forall \, (\lambda x.(P \; x) \rightarrow (Q \; x)))$$

This λ-term will effectively decide whether $A \subseteq B$, or alternatively whether $A = A \cap B$. Both the modellings sacrifice linearity in a sense: where the first, relational interpretation needs to use A as an operand to the intersection operation and as an argument to decide equality, the second interpretation has to *copy* the variable x to decide whether everything in the universe satisfying the property P also satisfies Q. We argue that this required non-linearity that is introduced by allowing non-linear λ-terms to be inserted through the lexicon, is exactly the same kind of non-linearity that is introduced to vector space models by means of bialgebra operations. It has been argued before that modelling quantification in vector space models forces one to use non-linear maps (Grefenstette 2013). However, this issue has been partially resolved by Hedges and Sadrzadeh (2016) by admitting a powerset structure to the basis vectors of the model. The then obtained bialgebra operations are linear in the algebraic sense, but non-linear in terms of typing information. That is, they allow for copying a resource X into a resource $X \otimes X$ and deleting a resource in the opposite direction. That this kind of operation would jeopardize a Lambek-style grammar formalism is immediate as the bialgebra operations would correspond to contraction and expansion, respectively. Our argument will proceed by claiming that a continuation-passing-style translation that allows for lexical insertion of non-linear λ-terms can instead be interpreted by means of the bialgebra operations of Hedges and Sadrzadeh (2016).

3   COMPOSITIONAL DISTRIBUTIONAL SEMANTICS

Compositional distributional semantics in a categorical setting takes a mathematically rigorous approach to compositionality. Much like traditional Montagovian semantics, there is a syntactic algebra involved that provides grammaticality by means of a proof system, in this case it can be either a *pregroup grammar* or the *Lambek calculus* (Lambek 1958, 1997). The *semantic algebra* is, in the basic setup, the category of

finite dimensional vector spaces, denoted **FVect**: content words are assigned a vector that represents its position in the space of word meanings, obtained through some method of co-occurrence extraction on a corpus. Whenever a sequence of words, annotated with their *syntactic types*, leads to a derivation that proves grammaticality, the proof term associated with that derivation provides a linear map on the vectors associated with basic words which, after evaluation, gives us the *phrase meaning* of that sequence of words.

3.1   *Lambek grammars*

We make the model sketched above concrete by giving the relevant definitions. These are based on work by Wijnholds (2014) in combination with the work of Coecke *et al.* (2013).

Definition 1 (Lambek types). *Given a set T of basic types, the set of Lambek types $F(T)$ is the smallest set such that:*

1. *If $p \in T$ then $p \in F(T)$,*
2. *If A, $B \in F(T)$ then $A \otimes B$, $A \backslash B$, $B/A \in F(T)$.*

We proceed to define a Lambek calculus in terms of a labelled deductive system, i.e. we use the notation of an inference system to show how proofs are derived:

Definition 2 (Non-associative Lambek calculus). *The (non-associative, non-unitary) Lambek calculus **NL** over T is given by the types in $F(T)$ and the proofs generated by the following (labelled) inference system:*

$$\frac{}{1_A : A \to A} \; Ax \qquad \frac{f : A \to B \quad g : B \to C}{g \circ f : A \to C} \; T$$

$$\frac{f : A \otimes B \to C}{\rhd f : A \to C/B} \; R1 \qquad \frac{f : A \otimes B \to C}{\lhd f : B \to A\backslash C} \; R2$$

$$\frac{g : A \to C/B}{\rhd^{-1} g : A \otimes B \to C} \; R1^{-1} \qquad \frac{g : B \to A\backslash C}{\lhd^{-1} g : A \otimes B \to C} \; R2^{-1}$$

One can show that monotonicity laws for each of the connectives are derived rules of inference:

$$\frac{f : A \to C \quad g : B \to D}{f \otimes g : A \otimes B \to C \otimes D} \; M_\otimes$$

$$\frac{f : A \to C \quad g : B \to D}{g/f : B/C \to D/A} \; M_{/}$$

$$\frac{f : A \to C \quad g : B \to D}{f \backslash g : C \backslash B \to A \backslash D} \; M_{\backslash}$$

where we have:

$$f \otimes g := \rhd^{-1} ((\rhd \lhd^{-1} ((\lhd 1_{C \otimes D}) \circ g)) \circ f)$$
$$g/f := \rhd (g \circ (\lhd^{-1} ((\lhd \rhd^{-1} 1_{B \backslash C}) \circ f)))$$
$$f \backslash g := \lhd (g \circ (\rhd^{-1} ((\rhd \lhd^{-1} 1_{C \backslash B}) \circ f)))$$

Leaving aside the issue of global associativity and its desirability from a linguistic perspective, we note how it can be added using two additional axioms:

$$\frac{}{a_{A,B,C} : (A \otimes B) \otimes C \to A \otimes (B \otimes C)} \; Ass$$

$$\frac{}{a^{-1}_{A,B,C} : A \otimes (B \otimes C) \to (A \otimes B) \otimes C} \; Ass^{-1}$$

The *categorical* version of the Lambek calculus can be obtained by imposing the relevant standard equivalences on proofs, amongst others stipulating that composing with the identity proof is a vacuous operation, and that all two-way inference rules are isomorphims. For more detail we refer the reader to Wijnholds (2014).

In order to make grammaticality judgments to sequences of words, we need a *lexicon* assigning types to words over an alphabet. For the sake of completeness we define the lexicon as a relation, but in the remainder of this paper we will freely abuse notation and treat the lexicon as if it were a function.

Definition 3 (Lexicon). *Let Σ be a finite, non-empty set of words (an alphabet). A lexicon over Σ is a relation $\delta \subseteq \Sigma \times F(T)$.*

Definition 4 (Lambek grammar). *Given a set of basic types T, a Lambek grammar over T is a triple (Σ, δ, S) where Σ is an alphabet, δ is a lexicon over T, and $S \in F(T)$ is a distinguished goal type.*

Definition 5 (Grammaticality). *Given a Lambek grammar (Σ, δ, S) over T, we say that a sequence of words $w_1...w_n$ over Σ is grammatical iff there is a merged sequence $W_1 \otimes W_2... \otimes W_n$ (where for each i we have $w_i \delta W_i$), and there exists a proof of $W_1 \otimes W_2... \otimes W_n \to S$ in the Lambek calculus.*

The presented definitions so far give a procedure to obtain a proof of sentencehood for a sequence of words. Moreover, there might be several proofs of the same sequence of words. This may be desirable (in cases of derivational ambiguity) or not (in the case proof-theoretic redundancy, e.g. the successive to and fro use of two-way rules). In the categorical variant of the Lambek calculus, we can simply take the proofs of sentencehood of a sequence to be the hom-set of morphisms $Hom(W_1 \otimes W_2...\otimes W_n, S)$. This produces fewer proofs as unnecessary ambiguity of the proof system is brought down by categorical equations. The structure of the (non-associative) Lambek calculus **NL** is that of a *biclosed magmatic category*.[4]

3.2 *Finite dimensional vector space models*

Lambek grammars are easily interpretable in vector space semantics, as vector spaces enjoy compact closure – a weaker variant of the biclosure of the Lambek calculus. We define the category **FVect** and show that it enjoys compact closure:

Definition 6 (Compact Closure). *A compact closed category is a monoidal category $\mathscr{C}$ with dual objects A^l, A^r for every object A in $\mathscr{C}$ and additional morphisms:*

$$A^l \otimes A \xrightarrow{\epsilon_A^l} I \xrightarrow{\eta_A^l} A \otimes A^l$$
$$A \otimes A^r \xrightarrow{\epsilon_A^r} I \xrightarrow{\eta_A^r} A^r \otimes A$$

that satisfy the yanking properties:[5]

$$(id_A \otimes \epsilon_A^l) \circ (\eta_A^l \otimes id_A) = id_A \qquad (\epsilon_A^r \otimes id_A) \circ (id_A \otimes \eta_A^r) = id_A$$

$$(\eta_A^l \otimes id_{A^l}) \circ (id_{A^l} \otimes \eta_A^l) = id_{A^l} \quad (id_{A^r} \otimes \eta_A^r) \circ (\eta_A^r \otimes id_{A^r}) = id_{A^r}$$

In the category of finite dimensional vector spaces **FVect** we have that the dual space A^* is isomorphic to A when we fix a basis (which is the case for concrete models). The ϵ and η maps, now reduced to just two maps, are given by:

$$\epsilon := \sum_{ij} c_{ij}(v_i \otimes v_j) \mapsto \sum_{ij} c_{ij}\langle v_i \mid v_j \rangle$$

$$\eta := 1 \mapsto \sum_i (v_i \otimes v_i)$$

---

[4] A magmatic category is a weaker version of a monoidal category: the tensor has no unit and is not necessarily associative. See Wijnholds (2017).

[5] Note that we left out the hidden associativity morphism.

In concrete vector models, we will have vectors learnt for content words. For instance, the noun phrases "John" and "Mary" can be interpreted as vectors $\overrightarrow{n_1}, \overrightarrow{n_3} \in \mathbf{N}$, respectively. This means that they are essentially single points in a vector space. Setting the sentence space to be the real numbers, a transitive verb such as "loves" would live in the vector space $\mathbf{N} \otimes \mathbb{R} \otimes \mathbf{N}$, and would carry information about the degree with which individuals love one another. In vector terms:

$$\sum_{ij} c_{ij}(\overrightarrow{n_i} \otimes 1 \otimes \overrightarrow{n_j})$$

The c_{ij} is the respective degree for any pair of individuals i, j. The meaning of the phrase "John loves Mary" should then reduce by taking the inner product of the noun phrases with the verbs to give:

$$\sum_{ij} c_{ij}\langle \overrightarrow{n_1} \mid \overrightarrow{n_i} \rangle \langle \overrightarrow{n_j} \mid \overrightarrow{n_3} \rangle$$

In the next section, we show how to relate derivations in a Lambek grammar to concrete computations in a vector space model.

3.3 *Interpretation*

Given that the compact closedness of **FVect** instantiates the closure of the Lambek calculus, we can easily interpret proofs in a Lambek grammar in a vector space model by passing from words and their lexical types to vectors in a homomorphically obtained vector space. Any proof of grammaticality will be interpreted through the η and ϵ maps:

Definition 7 (Interpretation). *Let (Σ, δ, S) be a Lambek grammar over T. An interpretation is a pair of maps $I_0 : F(T) \to$ **FVect**, $I_1 : \Sigma \to \delta(\Sigma)$, where $\delta(\Sigma)$ is the relational image of δ, such that I_0 respects typing and I_1 respects lexical type assignment. That is,*

$$I_0(A \otimes B) = I_0(A \backslash B) = I_0(A/B) = I_0(A) \otimes I_0(B)$$

and

$$I_1(w) = \overrightarrow{v} \quad iff \quad w\delta W \text{ and } \overrightarrow{v} \in I_0(W)$$

An interpretation map sends words to vectors that respect the syntactic types associated with those words. We need to give a vectorial interpretation of proofs as well, in order to know how to compute meanings

of a tuple of vectors. The identity proof and transitivity of proofs carries over to the identity map on vector spaces and the composition of linear maps. The remaining rules of residuation are interpreted as shown below:

$$\frac{f' : I_0(A) \otimes I_0(B) \to I_0(C)}{\left(f' \otimes id_{I_0(B)}\right) \circ \left(id_{I_0(A)} \otimes \eta_{I_0(B)}\right) : I_0(A) \to I_0(C) \otimes I_0(B)} \ R1$$

$$\frac{f' : I_0(A) \otimes I_0(B) \to I_0(C)}{\left(id_{I_0(A)} \otimes f'\right) \circ \left(\eta_{I_0(A)} \otimes id_{I_0(B)}\right) : I_0(B) \to I_0(A) \otimes I_0(C)} \ R2$$

$$\frac{g' : I_0(A) \to I_0(C) \otimes I_0(B)}{\left(id_{I_0(C)} \otimes \epsilon_{I_0(B)}\right) \circ \left(g' \otimes id_{I_0(B)}\right) : I_0(A) \otimes I_0(B) \to I_0(C)} \ R1^{-1}$$

$$\frac{g' : I_0(B) \to I_0(A) \otimes I_0(C)}{\left(\epsilon_{I_0(A)} \otimes id_{I_0(C)}\right) \circ \left(id_{I_0(A)} \otimes g'\right) : I_0(A) \otimes I_0(B) \to I_0(C)} \ R2^{-1}$$

It is a nice puzzle for the reader to verify that by the yanking equations, we preserve isomorphicity of residuation, for example one can show that the interpretation of $\rhd^{-1} \rhd f$ is equal to the interpretation of f.

3.4 *Illustration*

Recall that we have vectors for "John", "Mary" and "loves" and we have an intended meaning of the phrase "John loves Mary". We take a Lambek grammar over the set of basic types $\{np, s\}$, where np will be interpreted as **N** and s will be mapped to $\mathbb{R}$. We define a lexicon as follows:

w	$\delta(w)$	$I_1(w)$	$I_0(\delta(w))$
"John"	np	n_1	**N**
"Mary"	np	n_3	**N**
"loves"	$(np\backslash s)/np$	$\sum_{ij} c_{ij}(\vec{n_i} \otimes 1 \otimes \vec{n_j})$	$\mathbf{N} \otimes \mathbb{R} \otimes \mathbf{N}$

Given that $\lhd^{-1}\rhd^{-1}(1_{(np\backslash s)/np})$ proves the grammaticality of "John loves Mary", the associated meaning computation will be:

$$(\epsilon_N \otimes id_{\mathbb{R}}) \circ (id_N \otimes ((id_{N \otimes \mathbb{R}} \otimes \epsilon_N) \circ (id_{N \otimes \mathbb{R} \otimes N} \otimes id_N)))$$

$$(\overrightarrow{n_1} \otimes \textstyle\sum_{ij} c_{ij}(\overrightarrow{n_i} \otimes 1 \otimes \overrightarrow{n_j}) \otimes \overrightarrow{n_3})$$

$$= (\epsilon_N \otimes id_{\mathbb{R}}) \circ (id_{N \otimes N \otimes \mathbb{R}} \otimes \epsilon_N)(\overrightarrow{n_1} \otimes \textstyle\sum_{ij} c_{ij}(\overrightarrow{n_i} \otimes 1 \otimes \overrightarrow{n_j}) \otimes \overrightarrow{n_3})$$

$$= (\epsilon_N \otimes id_{\mathbb{R}})(\overrightarrow{n_1} \otimes \textstyle\sum_{ij} c_{ij}(\overrightarrow{n_i} \otimes \langle \overrightarrow{n_j} \mid \overrightarrow{n_3} \rangle))$$

$$= \textstyle\sum_{ij} c_{ij} \langle \overrightarrow{n_1} \mid \overrightarrow{n_i} \rangle \langle \overrightarrow{n_j} \mid \overrightarrow{n_3} \rangle$$

$$= c_{13}$$

This result is exactly the intended meaning we wanted to obtain. Note that the result of the computation relies on the fact that the content words in the vector space model are taken to be the basis vectors, hence they are orthogonal. The result c_{13} indicates the distributional strength of John loving Mary in a corpus that the vectors have been learnt from. Until now, we have neglected discussion about function words: logical words, relative pronouns, and quantifiers are not intuitively represented well by co-occurrence data. The logical word "and" may occur with many different words, but that statistic does not tell us much about the meaning of the word. So, although all the basic operations from a Lambek grammar are directly interpretable in vector space models, more advanced semantic phenomena lack an explanation in the simple models.

4   QUANTIFIER SCOPE AMBIGUITY IN VECTOR SPACE MODELS

In this section we review the use of bialgebras in vector space models as exhibited by Hedges and Sadrzadeh (2016) and Sadrzadeh (2016) and show how the two scope readings can be obtained. The treatment of quantifiers in vector space models relies on the use of powersets: as long as we can know of our vector space that its basis vectors are given by the powerset of some set A, we can perform additional operations on the vector space.

Definition 8 (Bialgebra). *Given a symmetric monoidal category* $(\mathscr{C}, \otimes, I, \sigma)$, *a bialgebra on an object* X *in* $\mathscr{C}$ *is a tuple of maps:*

$$X \xrightarrow{\delta_X} X \otimes X \xrightarrow{\mu_X} X$$

$$X \xrightarrow{\iota_X} I \xrightarrow{\zeta_X} X$$

that satisfy the conditions of a monoid for (X, μ, ζ) and a comonoid for (X, δ, ι) and furthermore satisfy the bialgebra axioms:

$$\iota \circ \mu = \iota \otimes \iota$$
$$\delta \circ \zeta = \zeta \otimes \zeta$$
$$\iota \circ \zeta = id_I$$
$$\delta \circ \mu = (\mu \otimes \mu) \circ (id_X \otimes \sigma \otimes id_X) \circ (\delta \otimes \delta)$$

The last of the four equations tells us that in a bialgebra, the order of copying and merging is irrelevant given that we can switch copies by means of the symmetry of the category. What is interesting to note is that any powerset $P(U)$ bears a bialgebra structure if we consider the Cartesian product to be the tensor and the singleton set $\{\star\}$ as the identity object. What follows is that any vector space over a powerset, denoted $V_{P(U)}$, carries a bialgebra structure. Both bialgebras are given below:

$$
\begin{array}{llll}
A & \stackrel{\delta}{\Longleftrightarrow} & A \times A & \qquad |A\rangle \stackrel{\delta}{\mapsto} |A\rangle \otimes |A\rangle \\
A \times B & \stackrel{\mu}{\Longleftrightarrow} & A \cap B & \qquad |A\rangle \otimes |B\rangle \stackrel{\mu}{\mapsto} |A \cap B\rangle \\
A & \stackrel{\iota}{\Longleftrightarrow} & \{\star\} & \qquad |A\rangle \stackrel{\iota}{\mapsto} 1 \\
\{\star\} & \stackrel{\zeta}{\Longleftrightarrow} & U & \qquad 1 \stackrel{\zeta}{\mapsto} |U\rangle
\end{array}
$$

The existence of a bialgebra on powerset vector spaces allows for a neat treatment of quantification. Given that nouns and noun phrases are represented as vectors on a powerset, universal quantification and existential quantification are treated as:

$$|A\rangle \stackrel{[\![all]\!]}{\mapsto} \sum_{A \subseteq B \subseteq U} |B\rangle$$

$$|A\rangle \stackrel{[\![some]\!]}{\mapsto} \sum_{\substack{B \text{ s.t.} \\ A \cap B \neq \emptyset}} |B\rangle$$

To get a feel for how the meaning of a quantified sentence should be computed according to Hedges and Sadrzadeh (2016), we show the example of "all men sleep", which gets assigned the meaning:

$$\epsilon_{V_{P(U)}} \circ ([\![all]\!] \otimes \mu_{V_{P(U)}}) \circ (\delta_{V_{P(U)}} \otimes id_{V_{P(U)}})(|[\![men]\!]\rangle \otimes |[\![sleep]\!]\rangle)$$

$$= \epsilon_{V_{P(U)}} \circ ([\![all]\!] \otimes \mu_{V_{P(U)}})(|[\![men]\!]\rangle \otimes |[\![men]\!]\rangle \otimes |[\![sleep]\!]\rangle)$$

$$= \epsilon_{V_{P(U)}} \Big(\sum_{[\![men]\!] \subseteq B \subseteq U} |B\rangle \otimes |[\![men]\!] \cap [\![sleep]\!]\rangle \Big)$$

$$= \sum_{[\![men]\!] \subseteq B \subseteq U} \langle B | [\![men]\!] \cap [\![sleep]\!]\rangle$$

$$= \langle [\![men]\!] | [\![men]\!] \cap [\![sleep]\!]\rangle$$

Although this approach works for statements with a single quantifier, it fails to deliver both readings for a doubly quantified statement such as "every student likes some teacher" as the computations for the subject and object quantifiers will be independent of each other. Hence, both readings will collapse to the same meaning. This lack of explanatory power of the model is amended in a subsequent paper (Sadrzadeh 2016), where the implicit quantified variable is passed on to the computation of the second quantifier. A transitive verb such as "likes" is modelled as an element:

$$[\![\text{likes}]\!] = \sum_{ij} c_{ij}(|A_i\rangle \otimes |A_j\rangle)$$

in $V_{P(U)} \otimes V_{P(U)}$, and we can model the forward image of an element in U as:

$$[\![\text{likes}_a]\!] = \sum_{ij} w_{ij}\langle\{a\} \mid A_i\rangle|A_j\rangle$$

The backward image is computed similarly by taking the inner product of $\overrightarrow{v_a}$ with $\overrightarrow{v_j}$. This construction now allows for both readings of "every student likes some teacher", though there is no procedure given to obtain these readings through a syntactic process.

5   QUANTIFIER SCOPE AMBIGUITY USING FOCUSSING AND POLARISATION

Focussing is a proof-theoretic technique stemming from the work of (Andreoli 2001) that aims to eliminate redundancy from regular sequent systems. Focussed proof search proceeds by distinguishing those formulas that enjoy invertible introduction rules (*asynchronous formulas*), and those that do not (*synchronous formulas*). Asynchronous formulas are decomposed in a backward chaining proof search until there is no more decomposition possible. Then, one of the synchronous formulas is selected to be put in focus, after which the process of decomposition continues. This implies that now only the number of synchronous formulas determines the number of distinct proofs. This approach has been applied to the Lambek-Grishin calculus, a symmetric extension of the Lambek calculus, by Bernardi and Moortgat (2010), and is worked out in more detail by Moortgat and Moot (2011).

In order to obtain a compositional Montagovian semantics from a display style presentation of focussed proofs for the Lambek-Grishin calculus, Bastenhof (2012) applies a polarisation technique, whereby formulas are assigned either positive or negative polarity. Atomic formulas are assigned an arbitrary polarity; the choice of this *bias* affects the set of proofs obtained. The polarity also influences semantics: under the continuation semantics of Bernardi and Moortgat (2010), a negative formula will be *negated* in its interpretation. Though the focussing and polarisation approaches are described by Bernardi and Moortgat (2010) and Bastenhof (2012), respectively, here we follow the focussed sequent presentation of Moortgat and Moot (2011).

We start by defining polarity of types:

Definition 9 (Polarity). *Given a set of basic types T, a polarity assignment on types is a map $pol : F(T) \to \{-, +\}$ that assigns to the types in T an arbitrary polarity but fixes the polarity for complex types:*

$$
\begin{aligned}
pol(A \otimes B) &= + \\
pol(A \backslash B) &= - \\
pol(B / A) &= -
\end{aligned}
$$

Given a Lambek grammar G over a set T, grammaticality is defined similarly to Definition 5, where the set of proofs is given by the underlying proof system. The only difference is that the final sequent should have the consequent formula in focus. Any proof is encoded by its abstract label, according to the abstract sequent system defined in Figure 1.
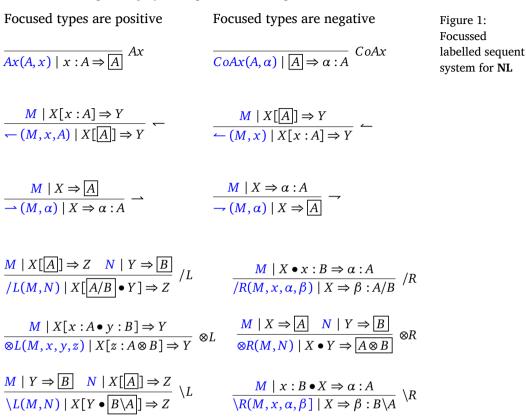
5.1 *CPS translation*

The translation of types and proofs given by Moortgat and Moot (2011) into a target semantic algebra is a two-step process:

$$
\text{source } \mathbf{NL}_{\otimes, \backslash, /} \xrightarrow{I} \text{continuation semantics } \mathbf{LP}_{\otimes, \perp} \xrightarrow{J} \text{target } \mathbf{FVect}
$$

Instead of considering a proof to be a simple transformation of values (the assumptions) to a value (the conclusion), we consider a proof to be a *continuation*, a function that awaits an evaluation context to compute a final value. The intermediate semantics is the Lambek calculus

Focused types are positive	Focused types are negative	Figure 1: Focussed labelled sequent system for **NL**

$$\frac{}{Ax(A,x) \mid x : A \Rightarrow \boxed{A}} \; Ax \qquad\qquad \frac{}{CoAx(A,\alpha) \mid \boxed{A} \Rightarrow \alpha : A} \; CoAx$$

$$\frac{M \mid X[x:A] \Rightarrow Y}{\leftharpoonup (M,x,A) \mid X[\boxed{A}] \Rightarrow Y} \; \leftharpoonup \qquad\qquad \frac{M \mid X[\boxed{A}] \Rightarrow Y}{\leftharpoonup (M,x) \mid X[x:A] \Rightarrow Y} \; \leftharpoonup$$

$$\frac{M \mid X \Rightarrow \boxed{A}}{\rightharpoonup (M,\alpha) \mid X \Rightarrow \alpha : A} \; \rightharpoonup \qquad\qquad \frac{M \mid X \Rightarrow \alpha : A}{\rightharpoonup (M,\alpha) \mid X \Rightarrow \boxed{A}} \; \rightharpoonup$$

$$\frac{M \mid X[\boxed{A}] \Rightarrow Z \quad N \mid Y \Rightarrow \boxed{B}}{/L(M,N) \mid X[\boxed{A/B} \bullet Y] \Rightarrow Z} \; /L \qquad\qquad \frac{M \mid X \bullet x : B \Rightarrow \alpha : A}{/R(M,x,\alpha,\beta) \mid X \Rightarrow \beta : A/B} \; /R$$

$$\frac{M \mid X[x:A \bullet y:B] \Rightarrow Y}{\otimes L(M,x,y,z) \mid X[z : A \otimes B] \Rightarrow Y} \; \otimes L \qquad\qquad \frac{M \mid X \Rightarrow \boxed{A} \quad N \mid Y \Rightarrow \boxed{B}}{\otimes R(M,N) \mid X \bullet Y \Rightarrow \boxed{A \otimes B}} \; \otimes R$$

$$\frac{M \mid Y \Rightarrow \boxed{B} \quad N \mid X[\boxed{A}] \Rightarrow Z}{\backslash L(M,N) \mid X[Y \bullet \boxed{B\backslash A}] \Rightarrow Z} \; \backslash L \qquad\qquad \frac{M \mid x : B \bullet X \Rightarrow \alpha : A}{\backslash R(M,x,\alpha,\beta] \mid X \Rightarrow \beta : B\backslash A} \; \backslash R$$

with permutation and negation, $\mathbf{LP}_{\otimes,\perp}$, a system that only uses a product operation but introduces a negation. Furthermore, permutation of resources is allowed to compensate for the lack of directionality without the $/, \backslash$ connectives. We will define a direct mapping from source to target, to skip the administrative details of the intermediate semantics.

In order to replicate the effect of the negation in $\mathbf{LP}_{\otimes,\perp}$, we use vector spaces over sets; given some type A, we define its interpretation to be a vector space over a set. In this way, we enjoy the bialgebras defined over those vector spaces. First, a type W is mapped to some set A, using the Cartesian product and powerset operations. Then, the final interpretation of a type will be the vector space over the given set, V_A. We get the intended tensor products on spaces due to the fact that $V_{A \times B} \cong V_A \otimes V_B$.

Definition 10 (Type interpretation). *Given a set of basic types T and a basic interpretation map $I_0 : T \to$ **Set**, the type interpretation is a map $I_1 : F(T) \to$ **Set** defined as follows:*

1. *For basic types $p \in T$ we have:*

$$I_1(p) = \begin{cases} I_0(p) & \text{if } pol(p) = + \\ P(I_0(p)) & \text{if } pol(p) = - \end{cases}$$

2. *For complex types, the interpretation depends both on the polarity of subtypes and the connective involved:*

A	B	$I_1(A \otimes B)$	$I_1(A \backslash B)$	$I_1(B/A)$
$-$	$-$	$P(I_1(A)) \times P(I_1(B))$	$P(I_1(A)) \times I_1(B)$	$I_1(B) \times P(I_1(A))$
$-$	$+$	$P(I_1(A)) \times I_1(B)$	$P(I_1(A)) \times P(I_1(B))$	$I_1(B) \times I_1(A)$
$+$	$-$	$I_1(A) \times P(I_1(B))$	$I_1(A) \times I_1(B)$	$P(I_1(B)) \times P(I_1(A))$
$+$	$+$	$I_1(A) \times I_1(B)$	$I_1(A) \times P(I_1(B))$	$P(I_1(B)) \times I_1(A)$

3. *We stipulate that for any type A, its interpretation $I_1(A)$ is lifted to the vector space spanned by its elements, that is we define the final interpretation $I_2 : F(T) \to$ **FVect** as $I_2(W) = V_{I_1(W)}$.*

Definition 11 (Word interpretation). *Given a Lambek grammar (Σ, δ, S) over a set of basic types T and an interpretation map $I_2 : F(T) \to I_2(\delta(\Sigma))$, where $\delta(\Sigma)$ is the relational image of Σ under the lexicon, and $I_2(\delta(\Sigma))$ is the image under interpretation (i.e. vector spaces), the word interpretation is a map I_3 that respects the following:*

$$I_3(w) \in I_2(W) \qquad \text{iff} \quad w\,\delta\,W \text{ and } pol(W) = +$$
$$I_3(w) \in I_2(W) \to \mathbb{R} \quad \text{iff} \quad w\,\delta\,W \text{ and } pol(W) = -$$

That is, words with a positive type are translated as vectors, while words with a negative type are translated as linear maps.

As an example, if we define the associated vector space of the type np to be U and n to be $P(U)$, then the interpretation of a noun like "student" will be a constant $I_3(\text{"student"}) \in V_{P(U)}$, whereas a word like "all" that is typed np/n will be a linear map:

$$I_3(\text{"all"}) \in V_{P(U)} \otimes V_{P(U)} \to \mathbb{R}$$

We proceed to define how we interpret proof terms. The intuitive idea is that a proof term is translated into a linear map which will subsequently be applied to the word interpretations of its antecedents.

Though the proof system builds up terms with potentially unbound variables, we require for grammaticality (see above) that the conclusion formula be in focus; this means that the only unbound variables in the proof term are those of the antecedent formula, which will be substituted by word interpretations.

Definition 12 (Proof term interpretation). *Given a proof in the focussed sequent calculus for* **NL**, *there is a proof term that encodes the proof. We define the interpretation of a proof by giving the translation of proof terms into linear maps:*

$$Ax(A, x) \quad \overset{I_4}{\Longrightarrow} \quad x \in I_3(A)$$

$$CoAx(A, \alpha) \quad \overset{I_4}{\Longrightarrow} \quad \alpha \in I_3(A)$$

$$\leftharpoonup (M, x, A) \quad \overset{I_4}{\Longrightarrow} \quad |\{x \in I_3(A) | I_4(M) \neq 0\}\rangle$$

$$\leftharpoonup (M, x) \quad \overset{I_4}{\Longrightarrow} \quad x(I_4(M))$$

$$\rightharpoonup (M, \alpha) \quad \overset{I_4}{\Longrightarrow} \quad \alpha(I_4(M))$$

$$\rightharpoonup (M, \alpha) \quad \overset{I_4}{\Longrightarrow} \quad \alpha \mapsto I_4(M)$$

$$/L(M, N) \quad \overset{I_4}{\Longrightarrow} \quad I_4(M) \otimes I_4(N)$$

$$/R(M, x, \alpha, \beta) \quad \overset{I_4}{\Longrightarrow} \quad I_4(M)[\beta \rightarrow \alpha \otimes x]$$

$$\otimes L(M, x, y, z) \quad \overset{I_4}{\Longrightarrow} \quad I_4(M)[z \rightarrow x \otimes y]$$

$$\otimes R(M, N) \quad \overset{I_4}{\Longrightarrow} \quad I_4(M) \otimes I_4(N)$$

$$\backslash L(M, N) \quad \overset{I_4}{\Longrightarrow} \quad I_4(M) \otimes I_4(N)$$

$$\backslash R(M, x, \alpha, \beta) \quad \overset{I_4}{\Longrightarrow} \quad I_4(M)[\beta \rightarrow x \otimes \alpha]$$

Finally, as the interpretation is a continuation-passing-style translation, we will end up with a map that needs an evaluation context before finishing computation. So, given that a proof gives a linear map, we apply it to the identity map, and we instantiate the unbound variables with the relevant *word interpretations*.

5.2   *Deriving quantifier scope ambiguity*

Quantifier scope ambiguity as exemplified by the phrase "Every student likes some teacher", is already shown to be obtainable using the two-step translation process of Bernardi and Moortgat (2010) in a Lambek-Grishin grammar, and in a Lambek grammar (Moortgat and Moot 2011). Here, we alter the latter example given to translate into

the vector space model as employed by Hedges and Sadrzadeh (2016) and Sadrzadeh (2016) to show that both readings (narrow/wide and wide/narrow) can be obtained and give exactly the kind of meaning we would expect from a vector space model. This means that we can obtain the intended meaning in a *derivational* way. What is more, given that we have both a grammar available and we have learned concrete vectors, the process can potentially be fully automated. Each word has to be associated with a syntactic type, and we have to give a word interpretation mapping the words to a vector or linear map. We assume a set of basic types $\{np, n, s\}$ where s is the distinguished goal type. Polarity assignment is handled by stipulating that np and n are positive, and s is negative. Basic types np and n are interpreted as U and $P(U)$, respectively, and s gets translated to $\mathbb{R}$. The syntactic types and the word interpretation are given by the following table:

w	$\delta(w)$	$\lceil w \rceil$
every	np/n	$\epsilon_{V_{P(U)}} \circ (\llbracket \text{all} \rrbracket \otimes \mu_{V_{P(U)}}) \circ (\delta_{V_{P(U)}} \otimes id_{V_{P(U)}}) \circ \sigma$
student	n	$\llbracket \text{student} \rrbracket$
likes	$(np \backslash s)/np$	$\vec{a} \otimes f \otimes \vec{b} \mapsto f(\llbracket (\text{likes}_b)_a \rrbracket)$
some	np/n	$\epsilon_{V_{P(U)}} \circ (\llbracket \text{some} \rrbracket \otimes \mu_{V_{P(U)}}) \circ (\delta_{V_{P(U)}} \otimes id_{V_{P(U)}}) \circ \sigma$
teacher	n	$\llbracket \text{teacher} \rrbracket$

As a reminder, we also note the vectorial interpretation of lexical constants in the word interpretation:

$$\llbracket \text{all} \rrbracket(|A\rangle) \quad = \quad \sum_{A \subseteq B \subseteq U} |B\rangle$$

$$\llbracket \text{some} \rrbracket(|A\rangle) \quad = \quad \sum_{\substack{B \subseteq U \text{ s.t.} \\ A \cap B \neq \emptyset}} |B\rangle$$

$$\llbracket \text{student} \rrbracket \quad = \quad |A\rangle \text{ for some } A \subseteq U$$

$$\llbracket \text{teacher} \rrbracket \quad = \quad |B\rangle \text{ for some } B \subseteq U$$

$$\llbracket \text{likes} \rrbracket \quad = \quad \sum_{ij} c_{ij}(|A_i\rangle \otimes 1 \otimes |A_j\rangle) \text{ for each } A_x \subseteq U$$

The two proofs that we get from the focussed sequent calculus are displayed in Figures 2 and 3 (without labelling).

$$\cfrac{\cfrac{\overline{a : np \Rightarrow \boxed{np}}\ Ax \quad \overline{\boxed{s} \Rightarrow \alpha : s}\ CoAx}{a : np \bullet \boxed{np\backslash s} \Rightarrow \alpha : s}\ \backslash L \quad \overline{b : np \Rightarrow \boxed{np}}\ Ax}{\cfrac{a : np \bullet (\boxed{(np\backslash s)/np} \bullet b : np) \Rightarrow \alpha : s}{\cfrac{a : np \bullet (z : (np\backslash s)/np \bullet b : np) \Rightarrow \alpha : s}{\cfrac{a : np \bullet (z : (np\backslash s)/np \bullet \boxed{np}) \Rightarrow \alpha : s}{\cdots}}}}\ /L$$

(etc.)

Figure 2: Proof for wide over narrow scope

Figure 3: A proof for narrow over wide scope

Figure 2: Proof for wide over narrow scope

Figure 3: A proof for narrow over wide scope

If we take the proof term for the first proof and translate this into a vectorial map we get:

$$(1a)\ \alpha \mapsto x\ (\ |\{a \in U \mid u\ (\ |\{b \mid z\ (a \otimes \alpha \otimes b) \neq 0\}\rangle \otimes w) \neq 0\}\rangle \otimes y)$$

For the second proof term, we get a slightly different map:

$$(2a)\ \alpha \mapsto u\ (\ |\{a \in U \mid x\ (\ |\{b \mid z\ (b \otimes \alpha \otimes a) \neq 0\}\rangle \otimes y) \neq 0\}\rangle \otimes w)$$

The unfolded maps are quite intimidating so the complete computation is taken up in the Appendix. Here we just note that the two maps reduce to the readings shown below:

$$(1b)\ \langle [\![\text{student}]\!] \mid [\![\text{student}]\!] \cap \{a \in U \mid \sum_{\substack{B \subseteq U \text{ s.t.} \\ [\![\text{teacher}]\!] \cap B \neq \emptyset}} \langle B \mid [\![\text{teacher}]\!] \cap C \rangle\}\rangle\rangle$$

$$\text{where } C = \{b \in U \mid [\![(\text{likes}_b)_a]\!] \neq 0\}$$

$$(2b)\ \sum_{\substack{B \subseteq U \text{ s.t.} \\ [\![\text{teacher}]\!] \cap B \neq \emptyset}} \langle B \mid [\![\text{teacher}]\!] \cap \{a \in U \mid \langle [\![\text{student}]\!] \mid [\![\text{student}]\!] \cap D \rangle \neq 0\}\rangle$$

$$\text{where } D = \{b \in U \mid [\![(\text{likes}_a)_b]\!] \neq 0\}$$

We can see that these interpretations will give different results depending on the instantiation of the vectors. In fact, these interpretations correspond to the result of Sadrzadeh (2016). This effectively shows that quantifier scope ambiguity can be achieved in vector space models by the use of appropriate proof-theoretic notions.

6   CONCLUDING REMARKS

In this paper, we elaborated on quantifier scope ambiguity in compositional distributional models of meaning. In particular, the approach of Moortgat and Moot (2011) using a continuation-passing-style translation for a polarised and focussed proof system for the Lambek calculus was combined with the approach to generalised quantifiers of Hedges and Sadrzadeh (2016). The result is fully derivational and provides a fully worked out compositional way to obtain two readings for phrases of the type "Every student likes some teacher", thereby resolving the issue of manually assigning appropriate meaning vectors to such phrases.

Although we illustrate this with examples of two generalised quantifiers in a sentence, the approach works for a single quantifier, and since the applied strategy exploits the combinatorial choices of the proof system (focus on the first quantifier and then on the second one, or vice versa) we expect the approach to generalise to more quantifiers, though the possibility of overgeneration needs to be investigated.

As for experimental validation, since the writing of this paper, it has been recognised that using a powerset construction in vector spaces, to be able to make use of bialgebras, may not be very feasible in practical models: having a powerset as a basis may lead to an exponential blowup in vector space size, and could potentially give sparsity issues. One approach to deal with this could be to use fuzzy quantification (Zadeh 1983), which has already been explored by Dostal and Sadrzadeh (2016).

Another interesting avenue is to work out how several phenomena involving the copying of linguistic material can be analysed in a compositional distributional model. Coordination and pronoun relativisation have been given an account using Frobenius algebras over vector spaces (Kartsaklis 2016; Sadrzadeh *et al.* 2013), where the Frobenius operations allow one to express elementwise multiplication on arbitrary tensors. In future work we hope to analyse ellipsis, a phenomenon for which it can be argued that copying has to be part of the syntactic process. Rules of controlled copying, then, can be interpreted using the Frobenius or bialgebra operations. A first step has already been taken by Kartsaklis *et al.* (2016), and we wish to approach the problem from the type-logical perspective.

7   ACKNOWLEDGMENTS

APPENDIX

(**1a**) $\alpha \to x \left(\, |\{a \in U \mid u \left(|\{b \in U \mid z \, (a \otimes \alpha \otimes b) \neq 0\}| \otimes w \right) \neq 0\}| \otimes y \right)$

Which, after lexical insertion gives

$\alpha \to \llbracket\text{every}\rrbracket \left(\, |\{a \in U \mid \llbracket\text{some}\rrbracket \left(|\{b \in U \mid \llbracket\text{likes}\rrbracket \, (a \otimes \alpha \otimes b) \neq 0\}| \otimes \llbracket\text{teacher}\rrbracket \right) \neq 0\}| \otimes \llbracket\text{student}\rrbracket \right)$

Unfolding the definition and inserting the identity map gives

$\epsilon_{V_{P(U)}} \, \circ \, \left(\llbracket\text{all}\rrbracket \otimes \mu_{V_{P(U)}} \right) \, \circ \, \left(\delta_{V_{P(U)}} \otimes id_{V_{P(U)}} \right)$

$\circ \, \sigma \Big(\, |\{a \in U \mid \epsilon_{V_{P(U)}} \circ \big(\llbracket\text{some}\rrbracket \otimes \mu_{V_{P(U)}}\big) \circ \big(\delta_{V_{P(U)}} \otimes id_{V_{P(U)}}\big) \circ \sigma \big(|\{b \in U \mid \llbracket(\text{likes}_b)_a\rrbracket \neq 0\}| \otimes \llbracket\text{teacher}\rrbracket \big) \neq 0\}| \otimes \llbracket\text{student}\rrbracket \Big)$

$= \epsilon_{V_{P(U)}} \, \circ \, \left(\llbracket\text{all}\rrbracket \otimes \mu_{V_{P(U)}} \right) \, \circ \, \left(\delta_{V_{P(U)}} \otimes id_{V_{P(U)}} \right) \circ \sigma \Big(\, |\{a \in U \mid \sum\limits_{\substack{B \subseteq U \text{ s.t.} \\ \llbracket\text{teacher}\rrbracket \cap B \neq \emptyset}} \langle B \mid \llbracket\text{teacher}\rrbracket \cap \{b \in U \mid \llbracket(\text{likes}_b)_a\rrbracket \neq 0\}\rangle \} | \rangle \otimes \llbracket\text{student}\rrbracket \Big)$

$= \sum\limits_{\substack{\llbracket\text{student}\rrbracket \subseteq C \subseteq U}} \langle C \mid \llbracket\text{student}\rrbracket \cap \{a \in U \mid \sum\limits_{\substack{B \subseteq U \text{ s.t.} \\ \llbracket\text{teacher}\rrbracket \cap B \neq \emptyset}} \langle B \mid \llbracket\text{teacher}\rrbracket \cap \{b \in U | \llbracket(\text{likes}_b)_a\rrbracket \neq 0\}\}\rangle\rangle$

(2a) $\alpha \mapsto u\left(\mid\{a \in U \mid x\left(\mid\{b \in U \mid z\,(b \otimes a \otimes a) \neq 0\}\right) \otimes y) \neq 0\}\right) \otimes w)$

Which, after lexical insertion gives

$$\alpha \mapsto \lceil\text{some}\rceil\left(\mid\{a \in U \mid \lceil\text{every}\rceil\left(\mid\{b \in U \mid \lceil\text{likes}\rceil\,(b \otimes a \otimes a) \neq 0\}\right) \otimes \lceil\text{student}\rceil) \neq 0\}\right) \otimes \lceil\text{teacher}\rceil)$$

Unfolding the definition and inserting the identity map gives

$$\epsilon_{V_{P(U)}} \circ \left(\llbracket\text{some}\rrbracket \otimes \mu_{V_{P(U)}}\right) \circ \left(\delta_{V_{P(U)}} \otimes id_{V_{P(U)}}\right)$$

$$\circ\, \sigma\left(\middle|\left\{a \in U \mid \epsilon_{V_{P(U)}} \circ \left(\llbracket\text{all}\rrbracket \otimes \mu_{V_{P(U)}}\right) \circ \left(\delta_{V_{P(U)}} \otimes id_{V_{P(U)}}\right) \circ \sigma\left(\middle|\{b \in U \mid \llbracket(\text{likes}_a)_b\rrbracket \neq 0\}\right) \otimes \llbracket\text{student}\rrbracket\right) \neq 0\right\}\right) \otimes \llbracket\text{teacher}\rrbracket\right)$$

$$= \epsilon_{V_{P(U)}} \circ \left(\llbracket\text{some}\rrbracket \otimes \mu_{V_{P(U)}}\right) \circ \left(\delta_{V_{P(U)}} \otimes id_{V_{P(U)}}\right) \circ \sigma\left(\middle|\left\{a \in U \mid \sum_{\llbracket\text{student}\rrbracket \subseteq B \subseteq U} \langle B \mid \llbracket\text{student}\rrbracket \cap \{b \in U \mid \llbracket(\text{likes}_a)_b\rrbracket \neq 0\}\rangle \neq 0\right\}\right) \otimes \llbracket\text{teacher}\rrbracket\right)$$

$$= \sum_{\substack{C \subseteq U \text{ s.t.} \\ \llbracket\text{teacher}\rrbracket \cap C \neq \emptyset}} \langle C \mid \llbracket\text{teacher}\rrbracket \cap \left\{a \in U \mid \sum_{\llbracket\text{student}\rrbracket \subseteq B \subseteq U} \langle B \mid \llbracket\text{student}\rrbracket \cap \{b \in U \mid \llbracket(\text{likes}_a)_b\rrbracket \neq 0\}\rangle \neq 0\right\}\rangle$$

REFERENCES

Jean-Marc ANDREOLI (2001), Focussing and proof construction, *Annals of Pure and Applied Logic*, 107(1):131–163, doi:https://doi.org/10.1016/S0168-0072(00)00032-4.

Arno BASTENHOF (2012), Polarized Montagovian semantics for the Lambek-Grishin calculus, in Philippe DE GROOTE and Mark-Jan NEDERHOF, editors, *15th and 16th International Conference on Formal Grammar*, volume 7395, pp. 1–16, Springer, Springer-Verlag Berlin Heidelberg, doi:http://dx.doi.org/10.1007/978-3-642-32024-8.

Raffaella BERNARDI and Michael MOORTGAT (2010), Continuation semantics for the Lambek–Grishin calculus, *Information and Computation*, 208(5):397–416, doi:https://doi.org/10.1016/j.ic.2009.11.005.

Bob COECKE, Edward GREFENSTETTE, and Mehrnoosh SADRZADEH (2013), Lambek vs. Lambek: Functorial vector space semantics and string diagrams for Lambek calculus, *Annals of Pure and Applied Logic*, 164(11):1079–1100, doi:https://doi.org/10.1016/j.apal.2013.05.009.

Bob COECKE, Mehrnoosh SADRZADEH, and Stephen CLARK (2010), Mathematical foundations for a compositional distributional model of meaning, *arXiv preprint arXiv:1003.4394*, https://arxiv.org/pdf/1003.4394.

Matej DOSTAL and Mehrnoosh SADRZADEH (2016), Many valued generalised quantifiers for natural language in the DisCoCat model, Technical report, Czech Technical University Prague and Queen Mary University of London, https://qmro.qmul.ac.uk/xmlui/bitstream/handle/123456789/17382/DisCoCat%20Maodel%20Paper%20M.Sadrzadeh.pdf.

Edward GREFENSTETTE (2013), Towards a formal distributional semantics: simulating logical calculi with tensors, in *Proceedings of the Second Joint Conference on Lexical and Computational Semantics*, pp. 1–10, Association for Computational Linguistics, http://aclweb.org/anthology/S13-1001.

Jules HEDGES and Mehrnoosh SADRZADEH (2016), A generalised quantifier theory of natural language in categorical compositional distributional semantics with bialgebras, *arXiv preprint arXiv:1602.01635*, https://arxiv.org/pdf/1602.01635.

Dimitri KARTSAKLIS (2016), Coordination in categorical compositional distributional semantics, *arXiv preprint arXiv:1606.01515*, https://arxiv.org/pdf/1606.01515.

Dimitri KARTSAKLIS, Matthew PURVER, and Mehrnoosh SADRZADEH (2016), Verb phrase ellipsis using Frobenius algebras in categorical compositional distributional semantics, *DSALT Workshop, European Summer School on Logic, Language and Information*, https://pdfs.semanticscholar.org/6c56/137ffb008ee5f94a482e0c74e494d7f7bc04.pdf.

Germán KRUSZEWSKI, Denis PAPERNO, Raffaella BERNARDI, and Marco BARONI (2016), There is no logical negation here, but there are alternatives: Modeling conversational negation with distributional semantics, *Computational Linguistics*, 42(4):637–660, doi:https://doi.org/10.1162/COLI_a_00262.

Joachim LAMBEK (1958), The mathematics of sentence structure, *The American Mathematical Monthly*, 65(3):154–170, doi:https://doi.org/10.1080/00029890.1958.11989160.

Joachim LAMBEK (1997), Type grammar revisited, in *International Conference on Logical Aspects of Computational Linguistics*, pp. 1–27, Springer, doi:https://doi.org/10.1007/3-540-48975-4_1.

Richard MONTAGUE (1970), English as a formal language, *Linguaggi nella Società e nella Tecnica*.

Richard MONTAGUE (1973), The proper treatment of quantification in ordinary English, in *Approaches to Natural Language*, pp. 221–242, Springer, doi:https://doi.org/10.1007/978-94-010-2506-5_10.

Michael MOORTGAT and Richard MOOT (2011), Proof nets for the Lambek-Grishin calculus, *arXiv preprint arXiv:1112.6384*, `https://arxiv.org/pdf/1112.6384`.

Michael MOORTGAT and Gijs WIJNHOLDS (2017), Lexical and derivational meaning in vector-based models of relativisation, *Proceedings of the 21st Amsterdam Colloquium*, pp. 55–64, `https://semanticsarchive.net/Archive/jZiM2FhZ/AC2017-Proceedings.pdf`.

Glyn MORRILL, Oriol VALENTÍN, and Mario FADDA (2011), The displacement calculus, *Journal of Logic, Language and Information*, 20(1):1–48, doi:https://doi.org/10.1007/s10849-010-9129-2.

Mehrnoosh SADRZADEH (2016), Quantifier scope in categorical compositional distributional semantics, *arXiv preprint arXiv:1608.01404*, `https://arxiv.org/pdf/1608.01404`.

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

Mark STEEDMAN (2000), *The Syntactic Process*, MIT Press.

Oriol VALENTÍN (2014), The hidden structural rules of the discontinuous Lambek calculus, in *Categories and Types in Logic, Language, and Physics*, pp. 402–420, Springer, doi:https://doi.org/10.1007/978-3-642-54789-8_23.

Gijs WIJNHOLDS (2014), *Categorical foundations for extended compositional distributional models of meaning*, Master's thesis, Universiteit van Amsterdam, `https://www.illc.uva.nl/Research/Publications/Reports/reportlist/MoL-2014-22.text.pdf`.

Gijs Jasper Wıjnholds (2017), Coherent diagrammatic reasoning in compositional distributional semantics, in *International Workshop on Logic, Language, Information, and Computation*, pp. 371–386, Springer, doi:https://doi.org/10.1007/978-3-662-55386-2_27.

Lotfi A. Zadeh (1983), A computational approach to fuzzy quantifiers in natural languages, *Computers & Mathematics with Applications*, 9(1):149–184, ISSN 0898-1221, doi:http://dx.doi.org/10.1016/0898-1221(83)90013-5.