Mathematical Structures in Computer Science (2019), **29**, pp. 783–809
doi:10.1017/S0960129518000324

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

A generalised quantifier theory of natural language in categorical compositional distributional semantics with bialgebras

Jules Hedges[1,*] and Mehrnoosh Sadrzadeh[2]

[1]Department of Computer Science, University of Oxford, Oxford, UK and [2]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

*Corresponding author. Email: julian.hedges@cs.ox.ac.uk

(Received 24 January 2016; revised 30 August 2017; accepted 1 June 2017; first published online 10 April 2019)

Abstract

Categorical compositional distributional semantics is a model of natural language; it combines the statistical vector space models of words with the compositional models of grammar. We formalise in this model the generalised quantifier theory of natural language, due to Barwise and Cooper. The underlying setting is a compact closed category with bialgebras. We start from a generative grammar formalisation and develop an abstract categorical compositional semantics for it, and then instantiate the abstract setting to sets and relations and to finite-dimensional vector spaces and linear maps. We prove the equivalence of the relational instantiation to the truth theoretic semantics of generalised quantifiers. The vector space instantiation formalises the statistical usages of words and enables us to, for the first time, reason about quantified phrases and sentences compositionally in distributional semantics.

Keywords: Compositional Distributional Semantics; Compact Closed Categories; Frobenius and Bialgebras; Vector Space Semantics; Generalised Quantifiers; Conservativity; Natural Language Semantics

1. Introduction

Distributional semantics is a statistical model of natural language; it is based on the hypothesis that words that have similar meanings often occur in the same contexts and meanings of words can be deduced from the contexts in which they often occur. Intuitively speaking and in a nutshell, words like ‘cat’ and ‘dog’ often occur in the contexts ‘pet’, ‘furry’ and ‘cute’, hence have a similar meaning, one which is different from ‘baby’, since the latter despite being ‘cute’ has not so often occurred in the context ‘furry’ or ‘pet’. This hypothesis has often been traced back to the philosophy of language discussed by Firth (1935) and the mathematical linguistic theory developed by Harris (1954). Distributional semantics has been used to reason about different aspects of word meaning, that is, similarity (Rubenstein and Goodenough 1965; Turney 2006), retrieval and clustering (Landauer and Dumais 1997; Lin 1998) and disambiguation (Schutze 1998). A criticism to these models has been that natural language is not only about words but also about sentences, but these models do not naturally extend to sentences, as sentences are not frequently occurring units of corpora of text.

Models of natural language are not restricted to distributional semantics. A tangential approach puts the focus on the compositional nature of meaning and its relationship with language constructions. This approach is inspired by a hypothesis often assigned to Frege that meaning of a sentence is a function of the meanings of its parts (Frege 1948). Informally speaking and very roughly put, meaning of a transitive sentence such as 'dogs chase cats' is a binary function of its subject and object. For instance, here the binary function is the verb 'chase' and the arguments are 'dogs' and 'cats'. This idea has been formalised in different ways, examples are the early works of Bar-Hillel (1953) and Ajdukiewicz (1935) on using classical logic, the context-free grammars (CFGs) of Chomsky (1956) and the first-order logical approach of Montague (1970). One criticism to all these settings, however, is that they do not say much about the meanings of the parts of the sentence. For instance, here we do not know anything more about the meaning of 'chase' and of 'dogs' and 'cats', apart from the fact that one is a function and others its arguments.

Compositional distributional semantics aims to combine the compositional models of grammar with the statistical models of distributional semantics in order to overcome the above-mentioned criticisms. Among the early grammar-based formalisms of the field is the work of Clark and Pulman (2007), and among the first corpus-based approaches is the work of Mitchell and Lapata (2010). The former model pairs the distributional meaning representation of a word with its grammatical role in a sentence and defines the meaning of a sentence to be a function of such pairs. The latter takes the distributional meaning of a sentence to be the addition or multiplication of the distributional meanings of its words. The model of Clark and Pulman has not been experimentally successful and its theory does not allow comparing meanings of different sentences. The model of Mitchell and Lapata has been experimentally successful but forgets the grammatical structure of sentences, since addition and multiplication are commutative.

Categorical compositional distributional semantics is an attempt to overcome these shortcomings and unify these models. This model was first described in Clark et al. (2008) and later published in Coecke et al. (2010). It is based on two major developments: first is the mathematical models of grammar introduced in the work of Lambek (1958, 1997), which either explicitly or implicitly use the theory of monoidal categories; second is the formulation of the distributional representations in terms of vectors, by many, for example, Lund and Salton (Lund and Burgess 1996; Salton et al. 1975). The categorical model uses the fact that the grammatical structures of language can be described within a compact closed category (Lambek 2010; Preller and Lambek 2007) and that finite-dimensional vector spaces and linear maps form such a category (Kelly and Laplaza 1980). The original formulation of this model consisted of the product of these two categories, which was later recast using a strongly monoidal functor (Coecke et al. 2013; Kartsaklis et al. 2013; Preller and Sadrzadeh 2010). The theoretical constructions of this model on an elementary fragment of language (adjective noun phrases and transitive sentences) were evaluated in Grefenstette and Sadrzadeh (2011, 2015) and in Kartsaklis et al. (2012) and Kartsaklis and Sadrzadeh (2013). Much of the recent work of the field is focused on using methods from machine learning (regression, tensor decomposition and neural embeddings) to implement them more efficiently (Grefenstette et al. 2013; Kartsaklis et al. 2014; Milajevs et al. 1935; Polajnar et al. 2014).

Despite all these, dealing with meanings of logical words such as pronouns, prepositions, quantifiers and conjunctives has posed challenges and open problems. In recent work Sadrzadeh et al. (2013, 2014) and also in Kartsaklis (2015), we showed how Frobenius algebras over compact closed categories can become useful in modelling relative pronouns and prepositions. In this paper, we take a step further and show how bialgebras over compact closed categories model generalised quantifiers (Barwise and Cooper 1981). We first present a preliminary account of compact closed categories and bialgebras over them and review how vector spaces and relations provide instances. The contributions of the paper start from Section 3, where we develop an abstract categorical semantics for the generalised quantifier theory in terms of diagrams and morphisms of compact closed categories with bialgebras. We present two concrete interpretations of this abstract setting: sets and relations, as well as finite-dimensional vector spaces and linear maps.

The former is the basis for truth theoretic semantics and the latter for corpus-based distributional semantics. We prove that the relational instantiation of the abstract model is equivalent to the truth theoretic model of generalised quantifier theory (as presented by Barwise and Cooper). We then prove how the relational model embeds into finite-dimensional vector spaces and more

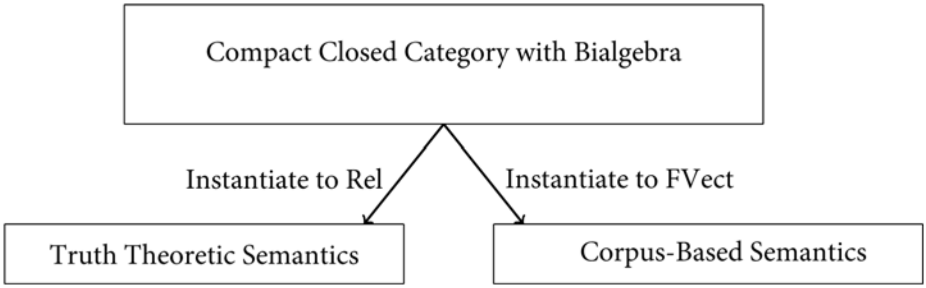


Figure 1. Abstract and concrete models for generalised quantifiers in compositional distributional semantics.

importantly show how it generalises to a compositional distributional semantic model of language. We provide vector interpretations for quantified sentences based on the grammatical structure of the sentences and the meaning vectors of their words. The meaning vectors of nouns, noun phrases and verbs are as previously developed. The meaning vectors of determiners and quantised phrases and sentences are novel.

There are two predecessors to this paper: Rypacek and Sadrzadeh (2014), where Frobenius algebras were used and the equivalence between relational instantiation and truth theoretic semantics could not be established, and Preller and Sadrzadeh (2011), where a two-sorted functional logic was used, but only a case for semantics of universal quantification was presented.

2. Preliminaries

2.1 Vector space models of natural language

Given a corpus of text, a set of contexts and a set of target words, a co-occurrence matrix has at each of its entries 'the degree of co-occurrence between the target word and the context'. This degree is determined using the notion of a *window*: a span of words or grammatical relations that slides across the corpus and records the co-occurrences that happen within it. A context can be a word, a lemma or a feature. A lemma is the canonical form of a word; it represents the set of different forms a word can take when used in a corpus. For example, the set {kills, killed, to kill, killing, killer, killers, . . . } is represented by the lemma 'kill'. A feature represents a set of words that together express a pertinent linguistic property of a word. These properties can be topical, lexical, grammatical or semantic. For example, the set {bark, miaow, neigh} represents a semantic feature of animal, namely the noise that it makes, whereas the set {fiction, poetry, science} represents the topical features of a book.

The lengths of the corpus and window are parameters of the model, as are the sizes of the feature and target sets. All of these depend on the task; for studies on these parameters, see for example Lapesa and Evert (2014) and Bullinaria and Levy (2007).

Given an $m \times n$ co-occurrence matrix, every target word t can be represented by a row vector of length n. For each feature f, the entries of this vector are a function of the raw co-occurrence counts, computed as follows:

$$\text{raw}_f(t) = \frac{\sum_c N(f,t)}{k}$$

for $N(f,t)$ the number of times the t and f have co-occurred in the window. Based on L, the total number of times that t has occurred in the corpus, the raw count is turned into various normalised degrees. Some common examples are probability, conditional probability, likelihood ratio and its logarithm:

$$\mathrm{P}_f(t) = \frac{\text{raw}_f(t)}{L}, \quad \mathrm{P}(f|t) = \frac{P(f,t)}{P(t)}, \quad \mathrm{LR}(f,t) = \frac{\mathrm{P}(f \mid t)}{\mathrm{P}(f)}, \quad \log \mathrm{LR}(f,t) = \log \frac{\mathrm{P}(f \mid t)}{\mathrm{P}(f)}$$

We denote a vector space model of natural language produced in this way with V_Σ, where Σ is the set of features, and V_Σ is the vector space spanned by it.

As an example, consider a corpus of 10^8 words, 10^6 target words and 10^5 features. Fix the window size to be 5 and suppose the co-occurrence matrix with raw counts to be as follows, where the column entries are the feature words and the row entries are the target words.

	fish	**horse**	**pet**	**blood**	...	total
dolphin	500	10	700	0	...	2000
shark	250	10	20	400	...	1000
plankton	250	10	1000	10	...	1700
pony	10	1000	10	10	...	1500

The vector representations of the target word 'dolphin' with the raw counts and its functions, as discussed above, are as follows:

$$\begin{aligned}
\text{raw} &= (500, 10, 700, 0)\\
\text{P} &:= \left(\frac{5}{20}, \frac{1}{200}, \frac{7}{20}, 0\right)\\
\text{LR} &:= (25000, 500, 17500, 0)\\
\log \text{LR} &:= (1.397, -0.301, 1.2430, 0)
\end{aligned}$$

Various notions of distance (length, angle) between the vectors have been used to measure the degree of similarity (semantic, lexical and information content) between the words. For instance, for the cosine of the angle between the vectors of dolphin and other target words we obtain:

$$\cos(\overrightarrow{\text{dolphin}}, \overrightarrow{\text{shark}}) = 0.87 \qquad \cos(\overrightarrow{\text{dolphin}}, \overrightarrow{\text{pony}}) = 0.009$$

This indicates that the degree of similarity between dolphin and shark is much higher than that of dolphin and pony. These degrees directly follow the co-occurrence degrees we have set above that dolphin and shark have co-occurred often with the same feature, but dolphin and pony have done so to a much lesser degree.

2.2 Generalised quantifier theory in natural language

We briefly review the theory of generalised quantifiers in natural language as presented in Barwise and Cooper (1981). Consider the fragment of English generated by the following CFG, referred to by G_Q:

$$\begin{array}{ll}
 & \text{NP} \to \text{John, Mary, something}, \ldots\\
\text{S} \to \text{NP VP} & \text{N} \to \text{cats, dogs, men}, \ldots\\
\text{VP} \to \text{V NP} & \text{VP} \to \text{sneeze, sleep}, \ldots\\
\text{NP} \to \text{Det N} & \text{V} \to \text{love, kiss}, \ldots\\
 & \text{Det} \to \text{a, the, some, every, each, all, no, most, few, one, two}, \ldots
\end{array}$$

A model for the language generated by this grammar is a pair $(U, [\![\ ]\!])$, where U is a universal reference set and $[\![\ ]\!]$ is an interpretation function defined by induction as follows.

1. On terminals.

   (a) The interpretation of a determiner d generated by 'Det $\to d$' is a map with the following type:

   $$[\![d]\!] : \mathcal{P}(U) \to \mathcal{PP}(U)$$

   It assigns to each $A \subseteq U$, a family of subsets of U. These interpretations are referred to as *generalised quantifiers*. For logical quantifiers, they are defined as follows:

   $$\begin{aligned}
   [\![\text{some}]\!](A) &= \{X \subseteq U \mid X \cap A \neq \emptyset\}\\
   [\![\text{every}]\!](A) &= \{X \subseteq U \mid A \subseteq X\}\\
   [\![\text{no}]\!](A) &= \{X \subseteq U \mid A \cap X = \emptyset\}\\
   [\![n]\!](A) &= \{X \subseteq U \mid \mid X \cap A \mid = n\}
   \end{aligned}$$

   A similar method is used to define non-logical quantifiers. For example, for non-logical quantifiers most, few and many, they are defined as follows:

   $$\begin{aligned}
   [\![\text{most}]\!](A) &= \{X \subseteq U \mid X \text{ has most elements of } U\}\\
   [\![\text{few}]\!](A) &= \{X \subseteq U \mid X \text{ has few elements of } U\}\\
   [\![\text{many}]\!](A) &= \{X \subseteq U \mid X \text{ has many elements of } U\}
   \end{aligned}$$

   (b) The interpretation of a terminal $y \in \{np, n, vp\}$ generated by either of the rules 'NP $\to$ np, N $\to$ n, VP $\to$ vp' is $[\![y]\!] \subseteq U$. That is, noun phrases, nouns and verb phrases are interpreted as subsets of the reference set.

   (c) The interpretation of a terminal y generated by the rule V $\to$ y is $[\![y]\!] \subseteq U \times U$. That is, verbs are interpreted as binary relations over the reference set.

2. On non-terminals.

   (a) The interpretation of expressions generated by the rule 'NP $\to$ Det N' is as follows:

   $$[\![\text{Det N}]\!] = [\![d]\!]([\![n]\!]) \qquad \text{where} \qquad X \in [\![d]\!]([\![n]\!]) \textbf{ iff } X \cap [\![n]\!] \in [\![d]\!]([\![n]\!])$$
   $$\text{for all } d, n \text{ generated by Det} \to d \text{ and N} \to n$$

   (b) The interpretation of expressions generated by the rule 'VP $\to$ V NP' is as follows:

   $$[\![\text{V NP}]\!] = [\![v]\!]([\![np]\!]) \quad \text{for all } v, np \text{ generated by VP} \to v\ np$$

   where, R is a unary relation $R \subseteq U$ and for $A \subseteq U$, $R(A)$ is the forward image of R on A, that is $R(A) = \{y \mid y \in R, \text{for } x \in A, \text{s.t.} x = y\}$, indeed $R \cap A$.

   (c) The interpretation of expressions generated by the rule 'S $\to$ NP VP' is as follows

   $$[\![\text{NP VP}]\!] = [\![vp]\!]([\![np]\!]) \quad \text{for all } np, vp \text{ generated by S} \to np\ vp$$

   where, $R \subseteq U \times U$ is a binary relation and for $A \subseteq U$, $R(A)$ is the forward image of R on A, that is $R(A) = \{y \mid (x, y) \in R, \text{for some } x \in A\}$.

   In either of the 2.(b) and 2.(c), the cases where $[\![np]\!]$ is a family of sets, that is, obtained by application of a determiner to a noun, are defined using the induction hypothesis. More specifically, these definitions go through item 2.(a), which in turn is obtained by going through item 1.(a).

Generalised quantifiers in natural language satisfy a property referred to by *living on* or *conservativity*, defined below.

Definition 1. *For a generalised quantifier $Q : \mathcal{P}(U) \to \mathcal{PP}(U)$, we say that Q satisfies the* living on property *if, for all $X, A \subseteq U$, $X \in Q(A)$ iff $X \cap A \in Q(A)$.*

In the models of natural language, a generalised quantifier is the interpretation of a determiner. In G_Q, a determiner is generated by the syntactic rule 'Det $\to$ d'. Determiners are applied to nouns via the syntactic rule 'NP $\to$ Det N'. Thus, when employed in the language generated by G_Q, the above definition gets the following form:

For the expressions of G_Q, a determiner $[\![d]\!]$ satisfies the living on property if, for all $A, X \subseteq U$, $X \in [\![d]\!](A)$ iff $X \cap A \in [\![d]\!](A)$.

Originally discussed in Barwise and Cooper (1981) and subsequently in almost all the literature on generalised quantifier theory, it is easy to verify that the living on property makes the following true:

Lemma 1. *For d, n, np, vp, a determiner, a noun, a noun phrase and a verb phrase of G_Q, if $[\![d]\!]$ satisfies the* living on *property, equivalences of the following kind hold on the expressions of G_Q:*

$$d\ n\ vp \iff d\ n\ \textbf{are}\ n\ \textbf{who}\ vp$$

$$np\ v\ d\ n \iff np\ v\ d\ n\ \textbf{who}\ \ \textbf{are}\ n$$

Examples are as follows:

All men eat. $\iff$ All men are men who eat.

Many men run. $\iff$ Many men are men who run.

Men love some cats. $\iff$ Men love some cats who are cats.

Thus the quantifiers modifying the subjects and objects of sentences *live on* these subjects and objects. The equivalent sentences of the right-hand sides seem redundant. They seem to be a redundant way of expressing the same as their left-hand side sentences. However, they express the fact that only the part of the *vp* that is restricted to the quantified *np* matters. Barwise and Cooper note that this is a property of natural language, and that every natural language has determiners whose semantic role is to assign to nouns (more precisely to the common count nouns) quantifiers that *live on* them. For instance, the determiner 'all' in the first example above assigns to the noun phrase 'men' the quantifier *all* such that it lives on 'men'. This criteria is thus used to filter out determiners whose semantics is not definable by the generalised quantifier theory, an example is the determiner 'only'. There are also mathematical quantifiers for whom this property fails, an example is Härtig's *equinumerous* quantifier defined by $B \in I(A) \iff |A| = |B|$.

The 'meaning' of a sentence is its truth value, defined as follows:

Definition 2. *A sentence in generalised quantifier theory is true iff $[\![NP\ VP]\!] \neq \emptyset$.*

As an example, meaning of a sentence with a quantified phrase at its subject position becomes as follows:

$$[\![\text{Det N VP}]\!] = \begin{cases} \textit{true} & \text{if } [\![vp]\!] \cap [\![n]\!] \in [\![\text{Det N}]\!] \\ \textit{false} & \text{otherwise} \end{cases}$$

obtained by applying the inductive definition of $[\![\ ]\!]$ to $[\![\text{NP}\ \ \text{VP}]\!]$ to generate $[\![S]\!]$. Herein, $[\![np]\!]$ is generated by the inductive step $[\![\text{Det}\ \ \text{N}]\!]$. For instance, meaning of 'some men sneeze', which is of this form, is true iff $[\![\text{sneeze}]\!] \cap [\![\text{men}]\!] \in [\![\text{some men}]\!]$, that is, whenever the set of things that sneeze and are men is a non-empty set. As another example, consider the meaning of a sentence with a quantified phrase at its object position, whose meaning is as follows:

$$[\![\text{NP V Det N}]\!] = \begin{cases} \textit{true} & \text{if } [\![n]\!] \cap [\![v]\!]([\![np]\!]) \in [\![\text{Det N}]\!] \\ \textit{false} & \text{otherwise} \end{cases}$$

This is obtained by applying the inductive definition to $[\![\text{NP VP}]\!]$ to generate $[\![S]\!]$, wherein $[\![vp]\!]$ is obtained by the inductive step $[\![\text{V NP}]\!]$ and $[\![np]\!]$ is obtained by the inductive step $[\![\text{Det N}]\!]$. An example of this case is the meaning of 'John liked some trees', which is true iff $[\![\text{trees}]\!] \cap [\![\text{like}]\!]([\![\text{John}]\!]) \in [\![\text{some trees}]\!]$, that is, whenever, the set of things that are liked by John and are trees is a non-empty set. Similarly, the sentence 'John liked five trees' is true iff the set of things that are liked by John and are trees has five elements in it.

2.3 From CFG to pregroup grammars

A pregroup algebra $P = (P, \leq, \cdot, (-)^r, (-)^l)$ is a partially ordered monoid where every element has a left and a right adjoint (Lambek 1997). That is, for $p \in P$, there are $p^l, p^r \in P$ that satisfy the following four inequalities:

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

Let P be a pregroup algebra; a pregroup grammar based on P is a tuple $P = (P, \Sigma, \beta, s)$, where Σ is the vocabulary of the language, $s \in P$ is a designated sentence type and β is a relation $\beta \subseteq \Sigma \times P$ that assigns to words in Σ elements of the pregroup P. This relation is referred to as a 'type dictionary' and the elements of the pregroup as 'types'.

A pregroup grammar P assigns a type p to a string of words $w_1 \cdots w_n$, for $w_i \in \Sigma$, if there exist types $p_i \in \beta(w_i)$ for $1 \leq i \leq n$ such that $p_1 \cdot \cdots \cdot p_n \leq p$. We refer to this latter inequality as the *grammatical reduction* of the string. If $p_1 \cdot \cdots \cdot p_n \leq s$, then the string is a grammatical sentence.

A CFG is transformed into a pregroup grammar via the procedure described in Buszkowski (2001). In a nutshell, one first transforms the CFG into an Ajdukiewicz grammar (Ajdukiewicz 1935), using the procedure developed by Bar-Hillel et al. (1960). The procedure developed by Buszkowski is then applied to transform the result into a Lambek calculus (Buszkowski 1988). Via a translation between Lambek calculi and pregroup grammars (Lambek 2008), the result is finally turned into a pregroup grammar.

$$\text{CFG} \xrightarrow{\text{Ajdukiewicz (1935)}} \text{Ajdukiewicz Grammar} \xrightarrow{\text{Buszkowski (1988)}}$$

$$\text{Lambek Calculus} \xrightarrow{\text{Buszkowski (2001)}} \text{Pregroup Grammar}$$

In a CFG in Chomsky normal form, the rules are either of the form $A \to BC$ or $A \to x$, for A, B, C non-terminals and x a terminal. The rules of such a grammar are classified based on the model defined over their generated language. We recall that a rule $A \to BC$ is referred to by *right-to-left* if $[\![A]\!] \coloneqq [\![C]\!]([\![B]\!])$; it is referred to by *left-to-right* if $[\![A]\!] \coloneqq [\![B]\!]([\![C]\!])$; the rest of the rules are referred to by *atomic*. Based on this given classification and by collating the above, we define the concept of a pregroup grammar generated over a CFG as follows:

Definition 3. *A pregroup grammar P_G generated over a CFG $G = (T, N, S, \mathcal{R})$ and a set of atomic types $\mathcal{A}$ is the pregroup grammar $(P(\mathcal{A}), T, \beta, \sigma(S))$ defined as follows:*

- *$P(\mathcal{A})$ is the free pregroup algebra generated over the set of atomic types $\mathcal{A}$.*
- *β is $\{(x, \sigma(x)) \mid x \in T\}$.*
- *$\sigma : N \cup T \to P(\mathcal{A})$ is as given below:*
    - *To a non-terminal C in a left-to-right rule $A \to BC$ of G, it assigns $\sigma(C) \coloneqq \sigma(B)^r \cdot \sigma(A)$.*
    - *To a non-terminal B in a right-to-left rule $A \to BC$, it assigns $\sigma(B) \coloneqq \sigma(A) \cdot \sigma(C)^l$.*
    - *To all the other non-terminals A, it assigns an atomic type $\sigma(A)$.*
    - *To all terminals x, generated by an atomic rule $A \to x$, it assigns the type $\sigma(A)$.*

As an example, consider G_Q and the set of atomic types $\{s, n, p\}$. The pregroup grammar generated over these is $(P(\{s, n, p\}), T, \beta, s)$. This grammar is in Chomsky normal form. The rule 'S $\to$ NP VP' is left-to-right and the rules 'VP $\to$ V NP' and 'NP $\to$ Det N' are right-to-left; the rest of the rules are atomic. On the non-terminals VP, V and Det, σ is thus defined as follows:

$$\sigma(VP) \coloneqq \sigma(NP)^r \cdot \sigma(S) \qquad \sigma(V) \coloneqq \sigma(VP) \cdot \sigma(NP)^l \qquad \sigma(Det) \coloneqq \sigma(NP) \cdot \sigma(N)^l$$

On the rest of the non-terminals, σ is defined as follows:

$$\sigma(NP) = p \qquad \sigma(N) = n \qquad \sigma(S) = s$$

β is as follows:

$$\{(np, p), (n, n), (vp, p^r \cdot s), (v, p^r \cdot s \cdot p^l), (d, p \cdot n^l)\}$$

Noun phrases take type p, nouns type n, intransitive verbs type $p^r \cdot s$ and transitive verbs type $p^r \cdot s \cdot p^l$. Determiners take type $p \cdot n^l$. Sample elements of β are as follows:

$$\{(\text{John}, p), (\text{cats}, n), (\text{sneeze}, p^r \cdot s), (\text{stroked}, p^r \cdot s \cdot p^l), (\text{some}, p \cdot n^l), \ldots\}$$

In this pregroup grammar, a quantified noun phrase such as 'some cats', a sentence with a quantified phrase in its subject position such as 'some cats sneeze' and a sentence with a quantified phrase in its object position such as 'John stroked some cats'. The grammatical reductions of these phrases and sentences in the above pregroup grammar are as follows:

$$\begin{array}{ccccl}
 & & \text{some} & \text{cats} & \\
 & & (p \cdot n^l) & \cdot n & \leq p \cdot 1 = p \\
 & \text{some} & \text{cats} & \text{sneeze} & \\
 & (p \cdot n^l) & \cdot n & \cdot (p^r \cdot s) & \leq p \cdot 1 \cdot (p^r \cdot s) = p \cdot (p^r \cdot s) \leq 1 \cdot s = s \\
\text{John} & \text{stroked} & \text{some} & \text{cats} & \\
p & \cdot (p^r \cdot s \cdot p^l) & \cdot (p \cdot n^l) & \cdot n & \leq 1 \cdot (s \cdot p^l) \cdot p \cdot 1 = (s \cdot p^l) \cdot p \leq s \cdot 1 = s
\end{array}$$

In the first example, 'some' inputs 'cats' and outputs a noun phrase; in the second example, first 'some' inputs 'cats' and outputs a noun phrase, then 'sneeze' inputs this noun phrase and outputs a sentence; in the last example, again first 'some' inputs 'cats' and outputs a noun phrase, at the same time the verb inputs 'John' and outputs a verb phrase of type $s \cdot p^l$, which then inputs the p from the phrase 'some cats' and outputs a sentence.

In the pregroup grammar of English presented in Lambek (2008), Lambek proposes to type the quantifiers as follows:

$$\text{when modifying the subject} : ss^l \pi \pi^l \qquad \text{when modifying the object} : os^r so^l$$

For the subject case, we have the identity $ss^l \pi \pi^l = s(\pi^r s)^l \pi^l$, which means that the quantifier inputs the subject (of type π) and the whole verb phrase and produces a sentence. Similarly, in the object case we have $os^r so^l = (so^l)^r so^l$. These types are translations of the original Lambek calculus types for quantifiers, where they were designed such that they would get a first-order logic semantics through a correspondence with lambda calculus (van Benthem 1987). However, as explained in Lambek (2008), due to the ambiguities in Lambek calculus-pregroup translations such a correspondence fails for pregroups. Consequently, the above types fail to provide a logical semantics for quantifiers. In this paper, we have taken a different approach and go by the types coming from the CFG of generalised quantifier theory. It will become apparent in the proceeding sections how this together with the use of compact closed categories offers a solution.

2.4 Category theoretic and diagrammatic definitions

This subsection briefly reviews compact closed categories and bialgebras. For a formal presentation, see Kelly and Laplaza (1980), Kock (1972) and McCurdy (2012). A compact closed category, $\mathcal{C}$, has objects A, B; morphisms $f: A \to B$ and a monoidal tensor $A \otimes B$ that has a unit I, that is we have $A \otimes I \cong I \otimes A \cong A$. Furthermore, for each object A there are two objects A^r and A^l and the following morphisms:

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

These morphisms satisfy the following equalities, where 1_A is the identity morphism on object A:

$$(1_A \otimes \epsilon_A^l) \circ (\eta_A^l \otimes 1_A) = 1_A \qquad (\epsilon_A^r \otimes 1_A) \circ (1_A \otimes \eta_A^r) = 1_A$$
$$(\epsilon_A^l \otimes 1_A) \circ (1_{A^l} \otimes \eta_A^l) = 1_{A^l} \qquad (1_{A^r} \otimes \epsilon_A^r) \circ (\eta_A^r \otimes 1_{A^r}) = 1_{A^r}$$

These express the fact the A^l and A^r are the left and right adjoints, respectively, of A in the 1-object bicategory whose 1-cells are objects of $\mathcal{C}$. A self-adjoint compact closed category is one in which for even object A we have $A^l \equiv A^r \equiv A$.

Given a morphism $f: X \to Y$ in a self-adjoint compact closed category, its *transpose* is the morphism $f^\top : Y \to X$ defined by

$$(\epsilon_Y \otimes X) \circ (Y \otimes f \otimes X) \otimes (Y \otimes \eta_X)$$

Given two compact closed categories $\mathcal{C}$ and $\mathcal{D}$ a strongly monoidal functor $F: \mathcal{C} \to \mathcal{D}$ is defined as follows:

$$F(A \otimes B) = F(A) \otimes F(B) \qquad F(I) = I$$

One can show that this functor preserves the compact closed structure, that is we have:

$$F(A^l) = F(A)^l \qquad F(A^r) = F(A)^r$$

A bialgebra in a symmetric monoidal category $(\mathcal{C}, \otimes, I, \sigma)$ is a tuple $(X, \delta, \iota, \mu, \zeta)$ where, for X an object of $\mathcal{C}$, the triple (X, δ, ι) is an internal comonoid; that is, the following are coassociative and counital morphisms of $\mathcal{C}$:

$$\delta: X \to X \otimes X \qquad \iota: X \to I$$

Moreover, (X, μ, ζ) is an internal monoid; that is, the following are associative and unital morphisms:

$$\mu: X \otimes X \to X \qquad \zeta: I \to X$$

And finally δ and μ satisfy the four equations (McCurdy 2012)

$$\iota \circ \mu = \iota \otimes \iota \tag{Q1}$$
$$\delta \circ \zeta = \zeta \otimes \zeta \tag{Q2}$$
$$\delta \circ \mu = (\mu \otimes \mu) \circ (\mathrm{id}_X \otimes \sigma_{X,X} \otimes \mathrm{id}_X) \circ (\delta \otimes \delta) \tag{Q3}$$
$$\iota \circ \zeta = \mathrm{id}_I \tag{Q4}$$

Informally, the comultiplication δ dispatches to copies the information contained in one object into two objects, and the multiplication μ unifies or merges the information of two objects into one. In what follows, we present three examples of compact closed categories, two of which with bialgebras.

2.5 Three examples of compact closed categories

Example 1. Pregroup Algebras A pregroup algebra $P = (P, \leq, \cdot, (-)^l, (-)^r)$ is a compact closed category whose objects are the elements of the set $p \in P$ and the partial ordering between the elements are the morphisms. That is, for $p, q \in P$, we have that $p \to q$ is a morphism of the category iff $p \leq q$ in the partial order. The tensor product of the category is the monoid multiplication, whose unit is 1, and the adjoints of objects are the adjoints of the elements of the algebra. The epsilon and eta morphism are thus as follows:

$$p \cdot p^r \xrightarrow{\epsilon_p^r} 1 \xrightarrow{\eta_p^r} p^r \cdot p \qquad\qquad p^l \cdot p \xrightarrow{\epsilon_p^l} 1 \xrightarrow{\eta_p^l} p \cdot p^l$$

The above directly follow from the pregroup inequalities on the adjoints. A pregroup with a bialgebra structure on it becomes degenerate. To see this, suppose we have such an algebra on the object p of such a pregroup. Then the unit morphism of the internal comonoid of this algebra becomes the partial ordering $\iota\colon p \leq 1$; taking the right adjoints of both sides of this inequality will yield $1 = 1^r \leq p^r$, and by multiplying both sides of this with p we will obtain $p \leq p \cdot p^r$, which by adjunction results in $p \leq p \cdot p^r \leq 1$, hence we have $p \leq 1$ and also $1 \leq p$, thus p must be equal to 1. That is, assuming that we have a bialgebra on an object will mean that that object is 1.

Example 2. Finite-Dimensional Vector Spaces over $\mathbb{R}$. These structures together with linear maps form a compact closed category, which we refer to as FdVect. Finite-dimensional vector spaces V, W are objects of this category; linear maps $f\colon V \to W$ are its morphisms with composition being the composition of linear maps. The tensor product $V \otimes W$ is the linear algebraic tensor product, whose unit is the scalar field of vector spaces; in our case this is the field of reals $\mathbb{R}$. Here, there is a natural isomorphism $V \otimes W \cong W \otimes V$. As a result of the symmetry of the tensor, the two adjoints reduce to one and we obtain the isomorphism $V^l \cong V^r \cong V^*$, where V^* is the dual space of V. When the basis vectors of the vector spaces are fixed, it is further the case that we have $V^* \cong V$. Thus, the compact closed category of finite-dimensional vector spaces with fixed basis is self-adjoint.

Given a basis $\{r_i\}_i$ for a vector space V, the epsilon maps are given by the inner product extended by linearity; that is, we have:

$$\epsilon^l = \epsilon^r\colon V \otimes V \to \mathbb{R} \quad \text{given by} \quad \sum_{ij} c_{ij}\, (\psi_i \otimes \phi_j) \quad \mapsto \quad \sum_{ij} c_{ij} \langle \psi_i \mid \phi_j \rangle$$

Similarly, eta maps are defined as follows:

$$\eta^l = \eta^r\colon \mathbb{R} \to V \otimes V \quad \text{given by} \quad 1 \mapsto \sum_i (|r_i\rangle \otimes |r_i\rangle)$$

Transposes in the category of finite-dimensional vector spaces are given by linear algebraic transposes of linear maps.

Let V be a vector space with basis $\mathcal{P}(U)$, where U is an arbitrary set. We give V a bialgebra structure as follows:

$$\iota|A\rangle = 1$$
$$\delta|A\rangle = |A\rangle \otimes |A\rangle$$
$$\zeta = |U\rangle$$
$$\mu(|A\rangle \otimes |B\rangle) = |A \cap B\rangle$$

Note that an arbitrary basis element of $V \otimes V$ is of the form $|A\rangle \otimes |B\rangle$ for $A, B \subseteq U$. For example, the verification of the bialgebra axiom (Q3) is as follows:

$$\begin{aligned}((\mu \otimes \mu) \circ (\text{ id} \otimes \sigma \otimes \text{id}) \circ (\delta \otimes \delta))(|A\rangle \otimes |B\rangle) &= ((\mu \otimes \mu) \circ (\text{ id} \otimes \sigma \otimes \text{id}))(|A\rangle \otimes |A\rangle \otimes |B\rangle \otimes |B\rangle) \\ &= (\mu \otimes \mu)(|A\rangle \otimes |B\rangle \otimes |A\rangle \otimes |B\rangle) \\ &= |A \cap B\rangle \otimes |A \cap B\rangle \\ &= \delta|A \cap B\rangle \\ &= (\delta \circ \mu)(|A\rangle \otimes |B\rangle)\end{aligned}$$

Example 3. Sets and Relations. Another important example of a compact closed category is Rel, the category of sets and relations. Here, $\otimes$ is Cartesian product with the singleton set as its unit $I = \{\star\}$, and * is identity on objects. Hence Rel is also self-adjoint. Closure reduces to the fact that a relation between sets $A \times B$ and C is equivalently a relation between A and $B \times C$. Given a set S with elements $s_i, s_j \in S$, the epsilon and eta maps are given as follows:

$$\begin{aligned}\epsilon^l = \epsilon^r &: S \times S \nrightarrow I \quad \text{given by} \quad (s_i, s_j)\epsilon\star \iff s_i = s_j \\ \eta^l = \eta^r &: I \nrightarrow S \times S \quad \text{given by} \quad \star\,\eta(s_i, s_j) \iff s_i = s_j\end{aligned}$$

Transposes in the category of relations are given by inverse relations.

For an object in Rel of the form $S = \mathcal{P}(U)$, we give S a bialgebra structure by taking

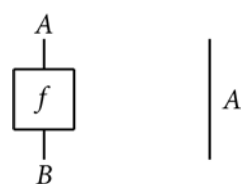
$$\begin{aligned}&\delta : S \nrightarrow S \times S && \text{given by} \quad A\delta(B, C) \iff A = B = C \\ &\iota : S \nrightarrow I && \text{given by} \quad A\iota\star \iff \text{(always true)} \\ &\mu : S \times S \nrightarrow S && \text{given by} \quad (A, B)\mu C \iff A \cap B = C \\ &\zeta : I \nrightarrow S && \text{given by} \quad \star\,\zeta A \iff A = U\end{aligned}$$

The axioms (Q1)–(Q4) can be easily verified by the reader.

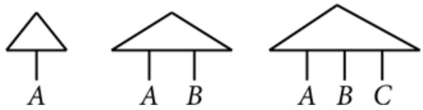
It should be noted that since both FdVect and Rel are †-categories, these constructions dualise to give two pairs of bialgebras. However, these bialgebras are not interacting in the sense of Bonchi et al. (2014), and the Frobenius axiom does not hold for either.

2.6 String diagrams

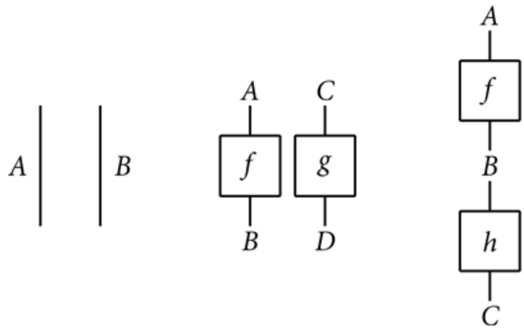
The framework of compact closed categories and bialgebras comes with a diagrammatic calculus that visualises derivations, and which also simplifies the categorical and vector space computations. Morphisms are depicted by boxes and objects by lines, representing their identity morphisms. For instance, a morphism $f : A \to B$, and an object A with the identity arrow $1_A : A \to A$, are depicted as follows:



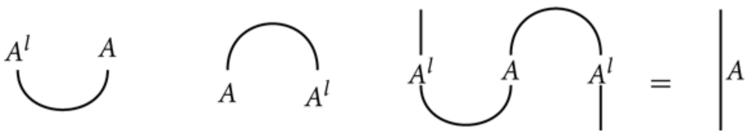
Morphisms from I to objects are depicted by triangles with strings emanating from them. In concrete categories, these morphisms represent elements within the objects. For instance, an element a in A is represented by the morphism $a : I \to A$ and depicted by a triangle with one string emanating from it. The number of strings of such triangles depict the tensor rank of the element; for instance, the diagrams for $a \in A$, $a' \in A \otimes B$ and $a'' \in A \otimes B \otimes C$ are as follows:



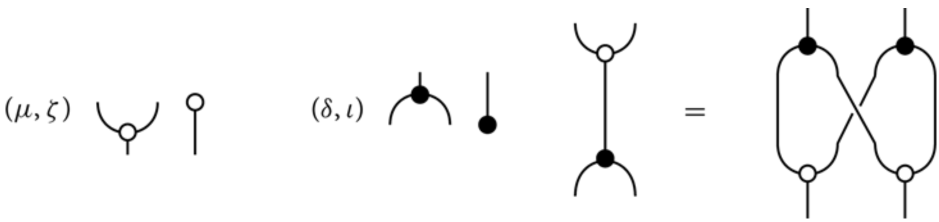
The tensor products of the objects and morphisms are depicted by juxtaposing their diagrams side by side, whereas compositions of morphisms are depicted by putting one on top of the other; for instance, the object $A \otimes B$, and the morphisms $f \otimes g$ and $h \circ f$, for $f \colon A \to B, g \colon C \to D$ and $h \colon B \to C$, are depicted as follows:



The ϵ maps are depicted by cups, η maps by caps and yanking by their composition and straightening of the strings. For instance, the diagrams for $\epsilon^l \colon A^l \otimes A \to I$, $\eta \colon I \to A \otimes A^l$ and $(\epsilon^l \otimes 1_A) \circ (1_A \otimes \eta^l) = 1_A$ are as follows:



As for the bialgebra, the diagrams for the monoid and comonoid morphisms and their interaction (the bialgebra law Q3) are as follows:



3. Abstract compact closed semantics

Definition 4. *An abstract compact closed categorical model for the language generated by the grammar $G = (T, N, S, \mathcal{R})$ is a tuple $(\mathcal{C}, W, S, \overline{[\![\ ]\!]})_{\mathcal{B}}$ where:*

- *$\mathcal{C}$ is a self-adjoint compact closed category with two distinguished objects W and S, where W has a bialgebra on it,*
- *$\mathcal{B}$ is the compact closed category freely generated over the generators of the pregroup grammar*

P_G as in Definition 3 and the atomic morphisms $I \to \sigma(A)$ introduced by the atomic rules $A \to x$ of G.

– *$\overline{[\![\ ]\!]} : \mathcal{B} \to \mathcal{C}$ is a strongly monoidal functor defined as follows:*

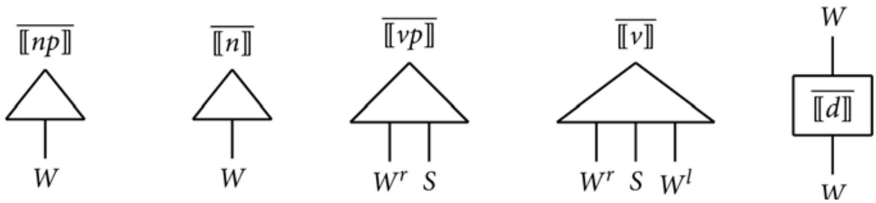
$$\overline{[\![x]\!]} \coloneqq \begin{cases} S & x = \sigma(S) \text{ for } S \text{ the designated starting symbol of } G \\ W & x \in \mathcal{A} \setminus \{\sigma(S)\} \\ I \to \overline{[\![\sigma(x)]\!]} & x \in \{NP, N, VP, V\} \\ \overline{[\![\sigma(x)]\!]} \to \overline{[\![\sigma(x)]\!]} & x \in \{Det\} \\ I \to \overline{[\![\sigma(x)]\!]} & x \in T, (x, \sigma(A)) \in \beta \end{cases}$$

Using a free compact closed category as opposed to a free pregroup is a solution suggested in Preller (2013) to the fact that there is no corresponding strongly monoidal functor on a free pregroup that assigns to atomic types, spaces of more than one dimension in the relational and distributional instantiations. In what follows, we drop the source category $\mathcal{B}$ and denote the semantics by $(\mathcal{C}, W, S, \overline{[\![\ ]\!]})$ in cases where the source category is fixed. In particular, the abstract compact closed categorical model for the language generated by G_Q is the one used in the rest of this paper, and thus we will drop the $\mathcal{B}$ from the tuple notation in the relational and vector space instantiations that follow.

As an example of the application of the above definition, consider the language of G_Q, wherein the map $\overline{[\![\ ]\!]}$ on the terminals is defined as follows:

$$\begin{aligned} \mathrm{NP} \to np \quad &\Longrightarrow \quad \overline{[\![np]\!]} \coloneqq I \to \overline{[\![\sigma(np)]\!]} : I \to W \\ \mathrm{N} \to n \quad &\Longrightarrow \quad \overline{[\![n]\!]} \coloneqq I \to \overline{[\![\sigma(n)]\!]} : I \to W \\ \mathrm{VP} \to vp \quad &\Longrightarrow \quad \overline{[\![vp]\!]} \coloneqq I \to \overline{[\![\sigma(vp)]\!]} : I \to W^r \otimes S \\ \mathrm{V} \to v \quad &\Longrightarrow \quad \overline{[\![v]\!]} \coloneqq I \to \overline{[\![\sigma(v)]\!]} : I \to W^r \otimes S \otimes W^l \\ \mathrm{Det} \to d \quad &\Longrightarrow \quad \overline{[\![d]\!]} \coloneqq \overline{[\![\sigma(d)]\!]} \to \overline{[\![\sigma(d)]\!]} : W \to W \end{aligned}$$

In diagrammatic form we have :



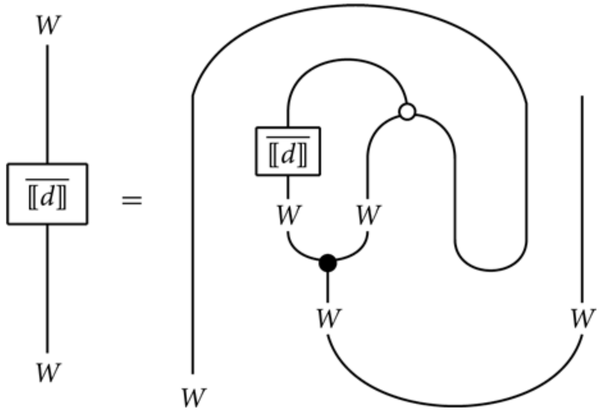
Intuitively, noun phrases and nouns are elements within the object W. Verb phrases are elements within the object $W^r \otimes S$; the intuition behind this representation is that in a compact closed category we have that $W^r \otimes S \cong W \to S$, where $W^r \to S = \hom(W, S)$ is an internal hom object of the category, coming from its monoidal closedness. Hence, we are modelling verb phrases as morphisms with input W and output S. Similarly, verbs are elements within the object $W^r \otimes S \otimes W^r$, equivalent to morphisms $W \otimes W \to S$ with pairs of input from W and output S. Determiners are morphisms $W \to W$ that further satisfy the categorical version of the *living on* property, defined below.

Definition 5. *A determiner d satisfies the* categorical living-on *property in the abstract compact closed categorical model $(\mathcal{C}, W, S, \overline{[\![\ ]\!]})$ generated by G_Q, if*

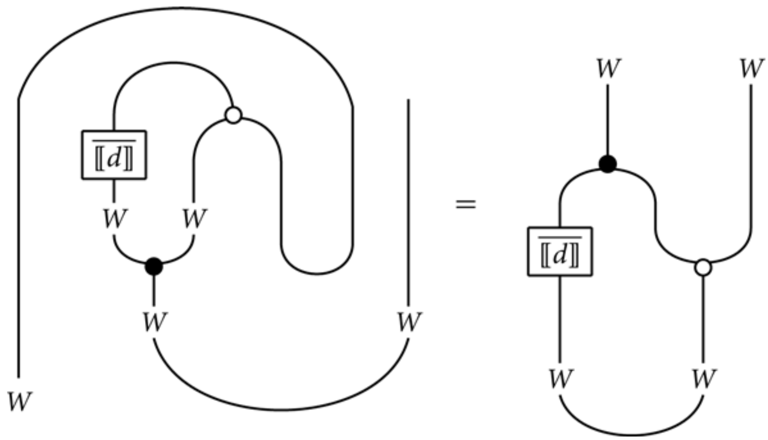
$$\overline{[\![d]\!]} = (1_W \otimes \epsilon_W) \circ (1_W \otimes \mu_W \otimes \epsilon_W \otimes 1_W) \circ (1_W \otimes \overline{[\![d]\!]}^{\top} \otimes \delta_W \otimes 1_{W \otimes W})$$
$$\circ (1_W \otimes \eta_W \otimes 1_{W \otimes W}) \circ (\eta_W \otimes 1_W)$$

where $-^{\top}$ *denotes transposition in* $\mathcal{C}$.

In the concrete relational and Boolean vector interpretations that we will define, this condition will be equivalent to Barwise and Cooper's living on property. Diagrammatically, this stipulation means that we have the following equality of diagrams:



Note that we also have



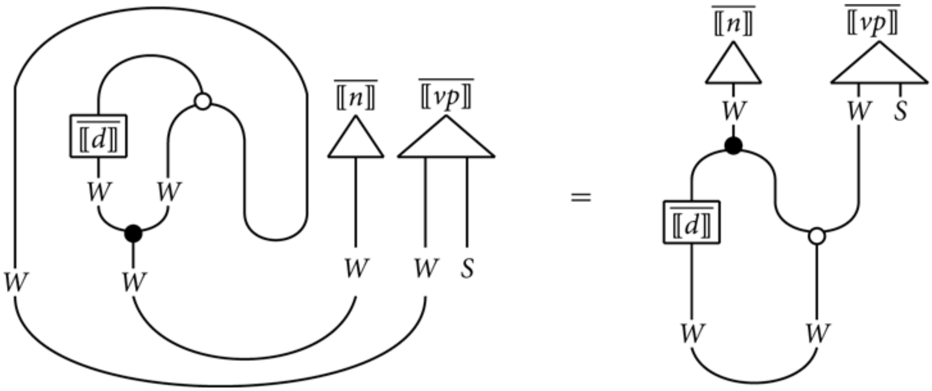
Intuitively, semantics of $\overline{[\![d]\!]}$ ends up being in $W \otimes W$, obtained by making a copy (via the bialgebra map δ) of one of the inputs in W, applying the determiner to one copy and taking the intersection of the other copy (via the bialgebra map μ) with the other input in W.

Meanings of expressions of language are obtained according to the following definition:

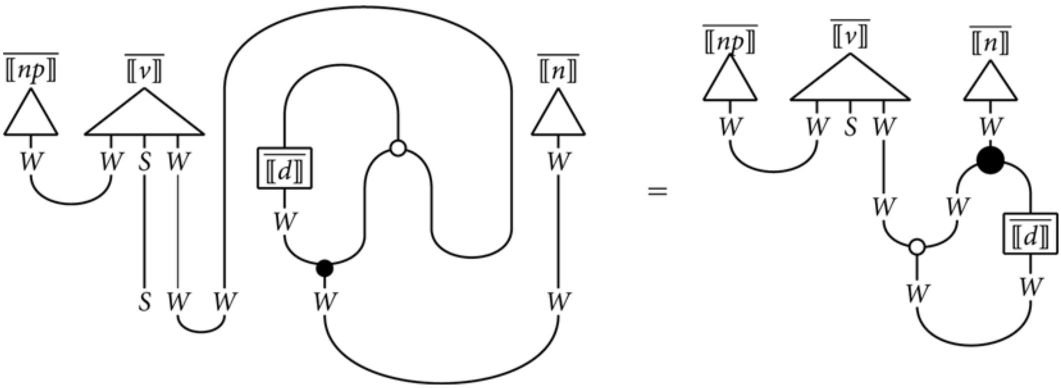
Definition 6. *The interpretation of a string* $w_1 \cdots w_n$, *for* $w_i \in T$ *with a grammatical reduction* α *is*

$$\overline{[\![w_1 \cdots w_n]\!]} \coloneqq \overline{[\![\alpha]\!]} \circ (\overline{[\![w_1]\!]} \otimes \cdots \otimes \overline{[\![w_n]\!]})$$

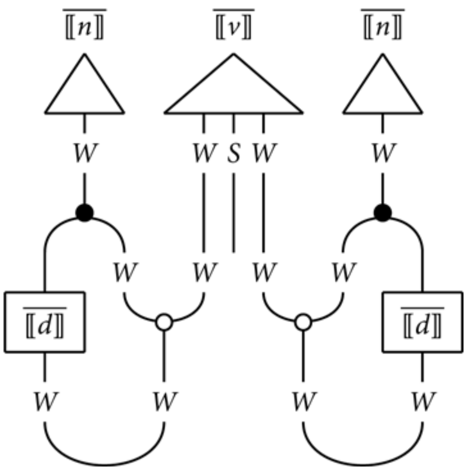
For example, the interpretation of an intransitive sentence with a quantified phrase in subject position and its simplified forms are as follows:



The interpretation of a transitive sentence with a quantified phrase in object position is as follows:



Putting the two cases together, the interpretation of a sentence with quantified phrases both at subject and at an object position is as follows:



4. Truth theoretic interpretation in Rel

A model $(U, [\![\ ]\!])$ of the language of generalised quantifier theory is made categorical via the instantiation to Rel of the abstract compact closed categorical model.

 https://doi.org/10.1017/S0960129518000324

Definition 7. *The instantiation of the abstract model of definition 4 to* Rel *is a tuple* $(\mathrm{Rel}, \mathcal{P}(U), \{\star\}, \overline{[\![\ ]\!]})$*, for* $\mathcal{U}$ *the universe of reference. The interpretations of words in this model are defined by the following relations:*

- *The interpretation of a terminal x generated by any of the non-terminals N, NP and VP is*

$$\star\overline{[\![x]\!]}A \iff A = [\![x]\!]$$

- *The interpretation of a terminal x generated by the non-terminal V is*

$$\star\overline{[\![x]\!]}(A, \star, B) \iff [\![x]\!](A) = B$$

  where $[\![x]\!](A)$ *is the forward image of A in the binary relation* $[\![x]\!]$*.*
- *The interpretation of a terminal d generated by the non-terminal Det is*

$$A\overline{[\![d]\!]}B \iff B \in [\![d]\!](A)$$

For the types, note that the interpretation of a terminal x generated by any of the non-terminals N, NP and VP has type $\overline{[\![x]\!]} : \{\star\} \nrightarrow \mathcal{P}(U)$. The interpretation of a VP is the initial morphism to $\mathcal{P}(U) \otimes \{\star\}$, which is isomorphic to $\mathcal{P}(U)$, hence it gets the same concrete instantiation as N and NP. The interpretation of a terminal x generated by the non-terminal V has type $\overline{[\![x]\!]} : \{\star\} \nrightarrow \mathcal{P}(U) \otimes \{\star\} \otimes \mathcal{P}(U) \cong \mathcal{P}(U) \otimes \mathcal{P}(U)$. Finally, the interpretation of a terminal d generated by the non-terminal Det has type $\overline{[\![d]\!]} : \mathcal{P}(U) \nrightarrow \mathcal{P}(U)$.

Informally, the bialgebra map μ is the analogue of set-theoretic intersection and the compact closed epsilon map is the analogue of set-theoretic application. It is not hard to show that the truth theoretic interpretation of the compact closed semantics of quantified sentences provides us with the same meaning as the generalised quantifier semantics. We make this formal as follows.

Definition 8. *The interpretation of a quantified sentence s is true in* $(\mathrm{Rel}, \mathcal{P}(U), \{\star\}, \overline{[\![\ ]\!]})$ *iff* $\star\overline{[\![s]\!]}\star$.

Theorem 1. $\star\overline{[\![s]\!]}\star$ in $(\mathrm{Rel}, \mathcal{P}(U), \{\star\}, \overline{[\![\ ]\!]})$ iff $[\![S]\!]$ is true in generalised quantifier theory, as defined in Definition 2.

Proof. If a sentence is quantified, it is either of the form 'Det N VP' or of the form 'NP V Det N'. For either case, since $\{\star\}$ is the unit of tensor in Rel, the S objects and morphisms can be dropped from the meaning morphism.

- For the first case, we have to calculate the $\overline{[\![s]\!]}$ relation:

$$\epsilon_{\mathcal{P}}(U) \circ (\overline{[\![d]\!]} \otimes \mu_{\mathcal{P}}(U)) \circ (\delta_{\mathcal{P}}(U) \otimes \mathrm{id}_{\mathcal{P}}(U)) \circ (\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]}) : \{\star\} \nrightarrow \{\star\}$$

  We will calculate this relation in stages. First:

$$\begin{aligned}\star(\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]})(A, B) &\iff \star\overline{[\![n]\!]}A \text{ and } \star\overline{[\![vp]\!]}B\\ &\iff A = [\![n]\!] \text{ and } B = [\![vp]\!]\end{aligned}$$

  since $(\star, \star) \cong \star$. Second:

$$\begin{aligned}\star((\delta_{\mathcal{P}}(U) \otimes \mathrm{id}_{\mathcal{P}}(U)) \circ (\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]}))(A, B, C) &\iff \star(\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]})(A, C) \text{ and } A = B\\ &\iff A = B = [\![n]\!] \text{ and } C = [\![vp]\!]\end{aligned}$$

Third:

$$
\begin{aligned}
& \star\,((\overline{[\![d]\!]} \otimes \mu_{\mathcal{P}}(U)) \circ (\delta_{\mathcal{P}}(U) \otimes \mathrm{id}_{\mathcal{P}}(U)) \circ (\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]}))(A, B) \\
\iff\; & A'\overline{[\![d]\!]}A \text{ and } B = B' \cap C' \text{ for some } \star\,((\delta_{\mathcal{P}}(U) \otimes \mathrm{id}_{\mathcal{P}}(U)) \circ (\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]}))(A', B', C') \\
\iff\; & A \in [\![d]\!]([\![n]\!]) \text{ and } B = [\![n]\!] \cap [\![vp]\!]
\end{aligned}
$$

Finally:

$$
\begin{aligned}
& \star\,(\epsilon_{\mathcal{P}}(U) \circ (\overline{[\![d]\!]} \otimes \mu_{\mathcal{P}}(U)) \circ (\delta_{\mathcal{P}}(U) \otimes \mathrm{id}_{\mathcal{P}}(U)) \circ (\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]}))\star \\
\iff\; & \star\,((\overline{[\![d]\!]} \otimes \mu_{\mathcal{P}}(U)) \circ (\delta_{\mathcal{P}}(U) \otimes \mathrm{id}_{\mathcal{P}}(U)) \circ (\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]}))(A, A) \text{ for some } A \\
\iff\; & [\![n]\!] \cap [\![vp]\!] \in [\![d]\!]([\![n]\!])
\end{aligned}
$$

This is the same as the set-theoretic meaning of the sentence in generalised quantifier theory.

- For the second case, we have:

$$
\overline{[\![s]\!]} \quad = \quad \epsilon_{\mathcal{P}(U)} \circ (\mu_{\mathcal{P}}(U) \otimes \overline{[\![d]\!]}) \circ (\epsilon_{\mathcal{P}(U)} \otimes \mathrm{id}_{\mathcal{P}(U)} \otimes \delta_{\mathcal{P}}(U)) \circ (\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]})
$$

Again we calculate in stages. First:

$$
\begin{aligned}
\star(\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]})(A, B, C, D) \iff\; & \star\overline{[\![np]\!]}A \text{ and } \star\,\overline{[\![v]\!]}(B, C) \text{ and } \star\,\overline{[\![n]\!]}D \\
\iff\; & A = [\![np]\!] \text{ and } C = [\![v]\!](B) \text{ and } D = [\![n]\!]
\end{aligned}
$$

Second:

$$
\begin{aligned}
& \star\,((\epsilon_{\mathcal{P}(U)} \otimes \mathrm{id}_{\mathcal{P}(U)} \otimes \delta_{\mathcal{P}}(U)) \circ (\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]}))(C, D, E) \\
\iff\; & D = E, \text{ and } \star\,(\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]})(A, A, C, D) \text{ for some } A \\
\iff\; & C = [\![v]\!]([\![np]\!]) \text{ and } D = E = [\![n]\!]
\end{aligned}
$$

Third:

$$
\begin{aligned}
& \star\,((\mu_{\mathcal{P}}(U) \otimes \overline{[\![d]\!]}) \circ (\epsilon_{\mathcal{P}(U)} \otimes \mathrm{id}_{\mathcal{P}(U)} \otimes \delta_{\mathcal{P}}(U)) \circ (\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]}))(F, G) \\
\iff\; & F = C \cap D \text{ and } D\overline{[\![d]\!]}G \text{ for some } \star\,((\epsilon_{\mathcal{P}(U)} \otimes \mathrm{id}_{\mathcal{P}(U)} \otimes \delta_{\mathcal{P}}(U)) \\
& \qquad \circ (\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]}))(C, D, E) \\
\iff\; & F = [\![v]\!]([\![np]\!]) \cap [\![n]\!] \text{ and } G \in [\![d]\!]([\![n]\!])
\end{aligned}
$$

Fourth:

$$
\begin{aligned}
& \star\,(\epsilon_{\mathcal{P}(U)} \circ (\mu_{\mathcal{P}}(U) \otimes \overline{[\![d]\!]}) \circ (\epsilon_{\mathcal{P}(U)} \otimes \mathrm{id}_{\mathcal{P}(U)} \otimes \delta_{\mathcal{P}}(U)) \circ (\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]}))\star \\
\iff\; & \star\,((\mu_{\mathcal{P}}(U) \otimes \overline{[\![d]\!]}) \circ (\epsilon_{\mathcal{P}(U)} \otimes \mathrm{id}_{\mathcal{P}(U)} \otimes \delta_{\mathcal{P}}(U)) \circ (\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]})) \text{ for some } F \\
\iff\; & [\![v]\!]([\![np]\!]) \cap [\![n]\!] \in [\![d]\!]([\![n]\!])
\end{aligned}
$$

Again, this is exactly the truth theoretic definition of the meaning of the sentence in generalised quantifier theory. This completes the proof.

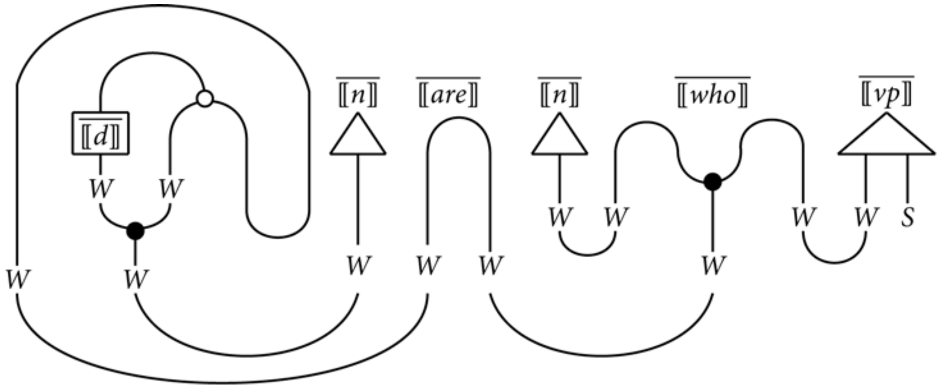
In the previous work (Clark et al. 2013; Sadrzadeh et al. 2013, 2014), we modelled relative pronouns in compact closed categories with Frobenius algebras. Using those results and Theorem 1, we show that the living on equivalences of Lemma 1 holds in Rel.

Corollary 1. *If d satisfies the categorical living on property, then the following equivalences hold in* $(\mathrm{Rel}, \mathcal{P}(U), \{\star\}, \overline{[\![\ ]\!]})$*:*

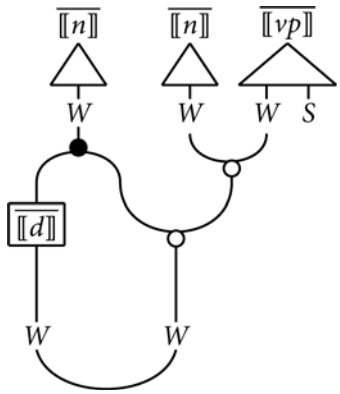
$$\star[\![\overline{d\ n\ vp}]\!]\star \iff \star[\![\overline{d\ n\ \textbf{are}\ n\ \textbf{who}\ vp}]\!]\star$$
$$\star[\![\overline{np\ v\ d\ n}]\!]\star \iff \star[\![\overline{np\ v\ d\ n\ \textbf{who}\ \ \textbf{are}\ n}]\!]\star$$

where $\overline{[\![who]\!]} = (1_{\mathcal{P}(U)} \otimes \mu_{\mathcal{P}(U)} \otimes 1_{\mathcal{P}(U)}) \circ (\eta_{\mathcal{P}(U)} \otimes \eta_{\mathcal{P}(U)})$ *and* $\overline{[\![are]\!]} = \eta_{\mathcal{P}(U)}$.

Proof. For the first case, consider the diagram corresponding to the relation $\star[\![\overline{d\ n\ \textbf{are}\ n\ \textbf{who}\ vp}]\!]\star$:



It simplifies to the following diagram:



The morphism corresponding to this diagram is as follows:

$$(\epsilon_{\mathcal{P}(U)} \otimes 1_{\{\star\}}) \circ (\overline{[\![d]\!]} \otimes \mu_{\mathcal{P}(U)} \otimes 1_{\{\star\}}) \circ (\delta_{\mathcal{P}(U)} \otimes \mu_{\mathcal{P}(U)} \otimes 1_{\{\star\}})(\overline{[\![n]\!]} \otimes \overline{[\![n]\!]} \otimes \overline{[\![vp]\!]}) : \{\star\} \nrightarrow \{\star\}$$

Following almost identical calculation steps as in the proof of Theorem 1 for sentences with a quantified subject, the above calculates to:

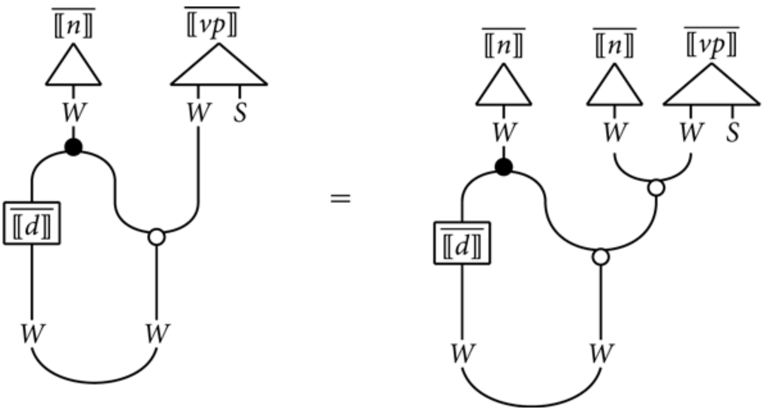
$$([\![vp]\!] \cap [\![n]\!]) \cap [\![n]\!] \in [\![d]\!]([\![n]\!])$$

evidently equivalent to

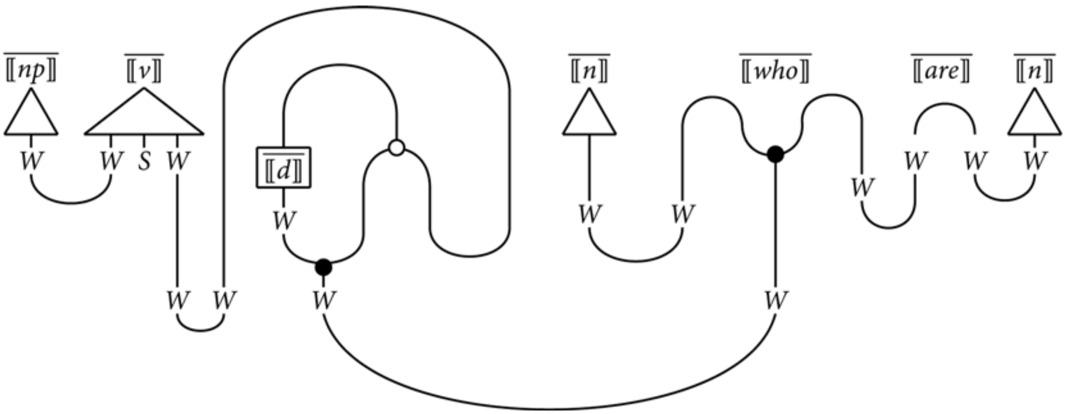
$$[\![vp]\!] \cap [\![n]\!] \in [\![d]\!]([\![n]\!])$$

which as proved in Theorem 1 is the result of calculating the relation corresponding to $\star[\![\overline{d\ n\ vp}]\!]\star$.

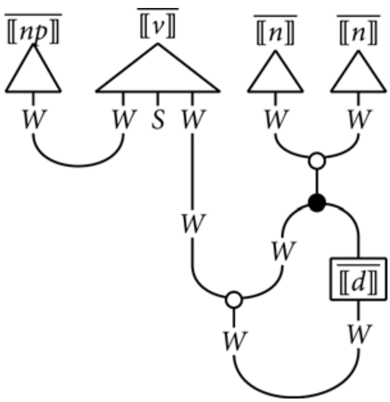
Diagrammatically, we have the following equality of the simplified forms of the diagrams of these sentences:



For the second case, consider the diagram corresponding to the relation $\star\overline{[\![np\ v\ d\ n\ \textbf{who}\ \ \textbf{are}\ n]\!]}\star$:



It simplifies to the following diagram:



The morphism corresponding to this diagram is as follows:

$$(1_{\{\star\}} \otimes \epsilon_{\mathcal{P}(U)}) \circ (1_{\{\star\}} \otimes \mu_{\mathcal{P}(U)} \otimes \overline{[\![d]\!]}) \circ (1_{\{\star\}} \otimes 1_{\mathcal{P}(U)} \otimes \delta_{\mathcal{P}(U)}) \circ (\epsilon_{\mathcal{P}(U)} \otimes 1_{\{\star\}} \otimes 1_{\mathcal{P}(U)} \otimes \mu_{\mathcal{P}(U)})$$
$$(\overline{[\![np]\!]} \otimes \overline{[\![v]\!]} \otimes \overline{[\![n]\!]} \otimes \overline{[\![n]\!]}) \colon \{\star\} \mathrel{\rlap{\hspace{0.45em}+}\longrightarrow} \{\star\}$$

Similar to the previous case, following almost identical calculation steps as in the proof of Theorem 1 for sentences with a quantified object, the above calculates to:

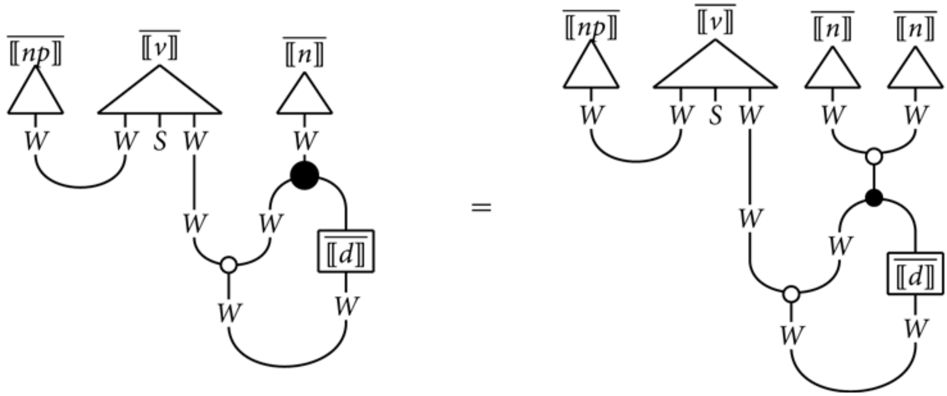
$$[\![v]\!]([\![np]\!]) \cap ([\![n]\!] \cap [\![n]\!]) \in [\![d]\!]([\![n]\!])$$

which is equivalent to

$$[\![v]\!]([\![np]\!]) \cap [\![n]\!] \in [\![d]\!]([\![n]\!])$$

which as proved in Theorem 1 is the result of calculating the relation corresponding to $\star\overline{[\![np\ v\ d\ n]\!]}\star$.

Diagrammatically, we have the following equality of the simplified forms of the diagrams of these sentences:



We end this section with three notes. First is that following Barwise and Cooper (1981), our G_Q does not generate relative clauses, copulous sentences and adjectival phrases. These are, however, needed for stating the equivalences related to the living on property. In the relational instantiations of Corollary 1, we are implicitly working with an extended form of G_Q that generates these expressions. In this extended form, interpreting the to-be verb 'are' in its copulous form as an eta map amounts to stipulating an equality such as $[\![\text{a is b}]\!] = [\![\text{b}]\!]([\![\text{a}]\!])$. Secondly, the μ and ζ maps of the previous work (Clark et al. 2013; Sadrzadeh et al. 2013, 2014) were the coalgebra maps of the Frobenius algebra over Rel. These Frobenius algebras were defined over the universe of reference U. The above results, however, are over $\mathcal{P}(U)$ and thus use the μ and ζ of the coalgebra maps of our bialgebra over Rel. Finally, since the sentence space of our model is the monoidal unit of Rel, the ζ_S map becomes identity; for simplicity we have dropped it from the type of 'who'.

5. Corpus-based instantiation in FdVect

The relational model embeds into a vector spaces model using the usual embedding of sets and relations into vector spaces and linear maps. This embedding sends a set T to a vector space V_T spanned by elements of T and a relation $R \subseteq T \times T$ to a linear map $V_T \to V_T$. By taking T to be $\mathcal{P}(U)$ for the distinguished space W and by taking it to be $\{\star\}$ for the distinguished space S, this embedding provides us with a vector space instantiation of the categorical model. This instantiation imitates the truth theoretic model presented in Rel. We refer to it by the *Boolean* FdVect instantiation.

Definition 9. *The Boolean instantiation of the abstract model of definition 4 to* FdVect *is the tuple* (FdVect, $V_{\mathcal{P}(U)}$, $V_{\{\star\}}$, $\overline{[\![\ ]\!]}$)*, for* $V_{\mathcal{P}(U)}$ *the free vector space generated over the set of subsets of* $\mathcal{U}$ *and* $V_{\{\star\}}$ *the one-dimensional space. Words are interpreted by the following linear maps:*

– *The terminals generated by N, NP, VP and V rules are given by:*

$$\overline{[\![x]\!]}(\star) = |[\![x]\!]\rangle$$

– *The interpretation of a terminal d generated by the Det rule is defined as follows on subsets A of $\mathcal{U}$:*

$$\overline{[\![d]\!]}(|A\rangle) = \sum_{B \in [\![d]\!](A)} |B\rangle$$

The types of these linear maps are as in definition 7, since $V_{\{\star\}} \cong \mathbb{R}$ is the unit of tensor in FdVect. Thus, the terminals generated by N, NP and VP rules have type $V_{\{\star\}} \to V_{\mathcal{P}(U)}$; the type of terminals generated by the V rule is $V_{\{\star\}} \to V_{\mathcal{P}(U)} \otimes V_{\{\star\}} \otimes V_{\mathcal{P}(U)} \cong V_{\mathcal{P}(U)} \otimes V_{\mathcal{P}(U)}$. A terminal generated by the Det rule has type $V_{\mathcal{P}(U)} \to V_{\mathcal{P}(U)}$.

Theorem 1 is carried over from Rel to FdVect by defining vector representations of sentences to be true iff they are non-zero elements of $V_{\{\star\}}$.

Definition 10. *The interpretation of a quantified sentence s is true in* $(\text{FdVect}, V_{\mathcal{P}(U)}, V_{\{\star\}}, \overline{[\![\ ]\!]})$ *iff* $\overline{[\![s]\!]}(\star) \neq 0$.

Corollary 2. $\overline{[\![s]\!]}(\star) \neq 0$ *in* $(\text{FdVect}, V_{\mathcal{P}(U)}, V_{\{\star\}}, \overline{[\![\ ]\!]})$ *iff* $\star[\![s]\!]\star$ *in* $(\text{Rel}, \mathcal{P}(U), \{\star\}, \overline{[\![\ ]\!]})$.

Proof. The proof goes through the same cases and steps as in Theorem 1. Consider a quantified sentence of the form 'Det N VP'. Its interpretation is obtained by calculating $\overline{[\![s]\!]}(\star)$, defined to be:

$$\epsilon_{V_{\mathcal{P}(U)}} \circ (\overline{[\![d]\!]} \otimes \mu_{V_{\mathcal{P}(U)}}) \circ (\delta_{V_{\mathcal{P}(U)}} \otimes \text{id}_{V_{\mathcal{P}(U)}}) \circ (\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]})(\star)$$

The four stages of this computation are as follows:

$$(\overline{[\![n]\!]} \otimes \overline{[\![vp]\!]})(\star) = \overline{[\![n]\!]}(\star) \otimes \overline{[\![vp]\!]}(\star) = |[\![n]\!]\rangle \otimes |[\![vp]\!]\rangle \tag{1}$$

$$(\delta_{V_{\mathcal{P}(U)}} \otimes \text{id}_{V_{\mathcal{P}(U)}})(|[\![n]\!]\rangle \otimes |[\![vp]\!]\rangle) = |[\![n]\!]\rangle \otimes |[\![n]\!]\rangle \otimes |[\![vp]\!]\rangle \tag{2}$$

$$(\overline{[\![d]\!]} \otimes \mu_{V_{\mathcal{P}(U)}})(|[\![n]\!]\rangle \otimes |[\![n]\!]\rangle \otimes |[\![vp]\!]\rangle) = \sum_{B \in [\![d]\!]([\![n]\!])} |B\rangle \otimes |[\![n]\!] \cap [\![vp]\!]\rangle \tag{3}$$

$$\epsilon_{V_{\mathcal{P}(U)}}\left(\sum_{B \in [\![d]\!]([\![n]\!])} |B\rangle \otimes |[\![n]\!] \cap [\![vp]\!]\rangle\right) = \sum_{B \in [\![d]\!]([\![n]\!])} \langle B \mid [\![n]\!] \cap [\![vp]\!]\rangle \tag{4}$$

The interpretation of a sentence with a quantified object 'NP V Det N' is computed similarly, resulting in the following expression:

$$\sum_{B \in [\![d]\!]([\![n]\!])} \langle [\![v]\!]([\![np]\!]) \cap [\![n]\!] \mid B\rangle$$

The result of the first case is non-zero iff there is a subset $B \in [\![d]\!]([\![n]\!])$ that is equal to $[\![n]\!] \cap [\![vp]\!]$. The result of the second case is non-zero iff there is a subset $B \in [\![d]\!]([\![n]\!])$ that is equal to $[\![v]\!]([\![np]\!]) \cap [\![n]\!]$. These are respectively equivalent to their corresponding cases in $\star[\![s]\!]\star$, as computed in the proof of theorem 1.

A corpus-based distributional vector space instantiation of the model is obtained via a construction similar to the above, but this time with real number weights (rather than Boolean ones). These weights are retrievable from corpora of text using distributional methods. The non-quantified part of this instantiation closely follows that of the previous work Coecke et al. (2010): nouns and noun phrases live in distributional spaces similar to the one described in Section 2.1;

verb phrases and transitive verbs live in tensor spaces, built using the methods described in the concrete instantiations of the theoretical model of previous work, for example, see Grefenstette and Sadrzadeh (2011) and Kartsaklis et al. (2012).

Definition 11. *The distributional instantiation of the abstract model of definition 4 to* FdVect *is the tuple* (FdVect, $V_{\mathcal{P}(\Sigma)}, Z, \overline{[\![\;]\!]}$), *for* $V_{\mathcal{P}(\Sigma)}$ *the vector space freely generated over the set* Σ *and* Z *a vector space wherein interpretations of sentences live. The interpretations of terminals are defined as follows:*

- *A terminal* x *generated by N or NP rules is given by* $\overline{[\![x]\!]}(1) \coloneqq \sum_i c_i^x |A_i\rangle$ *for* $A_i \subseteq \Sigma$.
- *A terminal* x *generated by the VP rule is given by* $\overline{[\![x]\!]}(1) \coloneqq \sum_{jk} c_{jk}^x |A_j \otimes A_k\rangle$, *for* $A_j \subseteq \Sigma$ *and* $|A_k\rangle$ *a basis vector of* Z.
- *A terminal* x *generated by the V rule is given by* $\overline{[\![x]\!]}(1) \coloneqq \sum_{lmn} c_{lmn}^x |A_l \otimes A_m \otimes A_n\rangle$, *for* $A_l, A_n \subseteq \Sigma$ *and* $|A_m\rangle$ *a basis vector of* Z.
- *A terminal* d *generated by the Det rule is concretely given on subsets* A *of* Σ *by* $\overline{[\![d]\!]}(|A\rangle) = \sum_{B \in [\![d]\!](A)} c_B^d |B\rangle$.

As for the types, a terminal generated by either of the N and NP rules has type $\mathbb{R} \to V_{\mathcal{P}(\Sigma)}$, a VP terminal has type $\mathbb{R} \to V_{\mathcal{P}(\Sigma)} \otimes Z$; the type of a V terminal is $\mathbb{R} \to V_{\mathcal{P}(\Sigma)} \otimes Z \otimes V_{\mathcal{P}(\Sigma)}$. A terminal d generated by the Det rule has type $V_{\mathcal{P}(\Sigma)} \to V_{\mathcal{P}(\Sigma)}$.

Examples of this model are obtained by setting three sets of parameters: (1) instantiating Z to different sentence spaces, (2) different ways of embedding the distributional vectors of V_Σ in the space $V_{\mathcal{P}(\Sigma)}$ and (3) different ways in which word vectors and tensors are built. The concrete constructions for the weighted interpretations of quantifiers depend on these choices, but can be implemented according to the same general guidelines. The weight c_B^d of a quantifier d over the basis A can stand for a *degree of set membership*. In this case $\sum_{B \in [\![d]\!](A)} c_B^d |B\rangle$ can be implemented as 'c_B^d is the degree to which d elements of A are in B'. This weight can also stand for a *degree of co-occurrence* and be retrieved from a corpus. In this case, $\sum_{B \in [\![d]\!](A)} c_B^d |B\rangle$ is read as 'c_B^d is the degree to which d elements of A *have co-occurred with* B'. We provide three example instantiations below.

Scalar Sentence Dimensions. Suppose $Z = \mathbb{R}$. The interpretation of a sentence with a quantified subject becomes as follows:

$$\sum_{ij} \sum_{B \in [\![d]\!]([\![n]\!])} c_i^n c_j^{vp} c_B^d \langle B \mid A_i \cap A_j \rangle$$

Similarly, the interpretation of a sentence with a quantified object becomes as follows:

$$\sum_{ijlm} \sum_{B \in [\![d]\!]([\![n]\!])} c_i^{np} c_{jl}^v c_m^n c_B^d \langle A_i \mid A_j \rangle \langle A_l \cap A_m \mid B \rangle$$

Here, take $\Sigma = \mathcal{U}$ and one can use the Rel-to-FdVect embedding and obtain a weighted version of the Boolean model of definition 9.

Distributional Sentence Dimensions. Suppose $\mathcal{S}$ contains the sentence dimensions of a compositional distributional model of meaning and take $Z = V_{\mathcal{S}}$. The sentence dimensions can be constructed in different ways. In Grefenstette and Sadrzadeh (2011), they were taken to be $\mathbb{R}$, whereas in Kartsaklis et al. (2012), we took them to be the same as the dimensions of V_Σ. In either case, there are different options on how to interpret the dimensions of $V_{\mathcal{P}(\Sigma)}$ in a distributional model. We present three different constructions below.

1. **The singleton construction.** Take the interpretation of a terminal x generated by either of the N or NP rules to be $\sum_i c_i^x |\{v_i\}\rangle$ whenever $\sum_i c_i^x |v_i\rangle$ is the vector interpretation of x in

the distributional space V_Σ. Similarly, a terminal x generated by the VP rule is embedded as $\sum_{ij} c^x_{ij} |\{v_i\} \otimes s_j\rangle$ whenever $\sum_{ij} c^x_{ij} |v_i \otimes s_j\rangle$ is the matrix interpretation of x in $V_\Sigma \otimes V_\mathcal{S}$. In the same fashion, a terminal x generated by the V rule embeds as $\sum_{ijk} c^x_{ijk} |\{v_i\} \otimes s_j \otimes \{v_k\}\rangle$, for $\sum_{ijk} c^x_{ijk} |v_i \otimes s_j \otimes v_k\rangle$ the cube interpretation of x in $V_\Sigma \otimes V_\mathcal{S} \otimes V_\Sigma$.

The interpretation of a sentence with a quantified subject becomes as follows:

$$\sum_{ijk} \sum_{B \in [\![d]\!]([\![n]\!])} c^n_i c^{vp}_{jk} c^d_B \langle B \mid \{v_i\} \cap \{v_j\} \rangle | s_k \rangle$$

Similarly, for the interpretation of a sentence with a quantified object we obtain:

$$\sum_{ijklm} \sum_{B \in [\![d]\!]([\![n]\!])} c^{np}_i c^v_{jkl} c^n_m c^d_B \langle \{v_i\} \mid \{v_j\} \rangle | s_k \rangle \langle \{v_l\} \cap \{v_m\} \mid B \rangle$$

The weights in the above formulae come from the underlying compositional distributional model. The vector constructions for nouns and noun phrases are obtained by following a distributional model; the matrix and cube constructions for verbs are constructed as detailed in Grefenstette and Sadrzadeh (2011) or in Kartsaklis et al. (2012), depending on the choice of $\mathcal{S}$.

2. **Sets of dimensions as lemmas.** A lemma is a set of different forms of a word. In this instantiation, each dimension of $V_{\mathcal{P}(\Sigma)}$ stands for a lemma.

   The interpretation of a sentence with a quantified sentence becomes:

$$\sum_{ijk} \sum_{B \in [\![d]\!]([\![n]\!])} c^n_i c^{vp}_{jk} c^d_B \langle B \mid A_i \cap A_j \rangle | s_k \rangle$$

   Similarly, the interpretation of a sentence with a quantified object becomes:

$$\sum_{ijklm} \sum_{B \in [\![d]\!]([\![n]\!])} c^{np}_i c^v_{jkl} c^n_m c^d_B \langle A_i \mid A_j \rangle | s_k \rangle \langle A_l \cap A_m \mid B \rangle$$

   The weights are retrieved from a corpus by, for example, adding, normalising and clustering (e.g. average or k-means) of the co-occurrence weights of the elements of the lemma set.

3. **Sets of dimensions as features.** A feature is the set of words that together represent a pertinent property. In this instantiating, each such dimension of $V_{\mathcal{P}(\Sigma)}$ represents a set of such words. For instance, {miaow, purr} is the sound feature for the 'animals', {run, sleep} is its action feature and {cat, kitten} is its species feature. Each dimension of $V_{\mathcal{P}(\Sigma)}$ stands for a feature. The interpretations of quantified sentences are obtained by computing the same formulae as in the lemma instantiation, but the concrete values of the weights are obtained differently.

Lemmas and features are sets of words. While lemmas are syntactic objects (different syntactic forms of a word), features represent semantic properties of words. Any set of words can in principle be a feature, namely, a common property shared by the words of the corresponding set. For instance, the words in the set {sky, sea, blueberry, winter} represent a feature corresponding to them being 'blue'; the words in the set {milk, beer, Lucozade, soya sauce} represent a feature since they are all drinkable. Thus every subset of Σ is feature and every dimension in $V_{\mathcal{P}(\Sigma)}$ becomes interpretable in the third embedding above. This is not the case in the second embedding. Not every subset of Σ corresponds to a lemma, thus not every dimension of $V_{\mathcal{P}(\Sigma)}$ is interpretable. In effect, for the theory and also the practical parts of this project, we do not need the whole of the $\mathcal{P}(\Sigma)$, as we have only used the intersection property of Rel with bialgebras. Working in a subset thereof, such as down or up sets is work under progress.

As an example of the third embedding, consider the feature set instantiation and suppose the following are among the features of $V_{\mathcal{P}(\Sigma)}$:

$$\{\text{cats, kittens}\}, \{\text{miaow, purr}\}, \{\text{sleep, snore}\} \quad \in \quad \mathcal{P}(\Sigma)$$

Take the instantiation of the universal quantifier over these to be:

$\overline{[\![\text{all}]\!]}$	$\lvert\{\text{cats, kittens}\}\rangle$	$\lvert\{\text{miaow, purr}\}\rangle$	$\lvert\{\text{sleep, snore}\}\rangle$
$\lvert\{\text{cats, kittens}\}\rangle$	`small`	0.7	0.5
$\lvert\{\text{miaow, purr}\}\rangle$	0.9	`small`	0.3
$\lvert\{\text{sleep, snore}\}\rangle$	0.2	0.3	`small`

In the first row, 0.7 is the degree to which *all* elements of {cats, kittens} have feature {miaow, purr}, witnessed by the fact that, for instance, all occurrences of cats and kittens in the corpus have occurred in sentences which have a verb such as miaow or purr. Similarly, 0.5 is the degree to which *all* elements of {cats, kittens} have feature {sleep, snore}. The intersection of a term with itself has no information content and is thus taken to be a very small fraction, so as not to play a role in deductions.

For the existential quantifier, a similar instantiation results in higher degrees as the quantifier is more relaxed, witnessed by the fact that, for instance, 'kittens' have more of the miaow feature than 'cats' since they miaow more. Suppose this provides us with the following:

$\overline{[\![\text{some}]\!]}$	$\lvert\{\text{cats, kittens}\}\rangle$	$\lvert\{\text{miaow, purr}\}\rangle$	$\lvert\{\text{sleep, snore}\}\rangle$
$\lvert\{\text{cats, kittens}\}\rangle$	`small`	0.9	0.6
$\lvert\{\text{miaow, purr}\}\rangle$	0.9	`small`	0.5
$\lvert\{\text{sleep, snore}\}\rangle$	0.5	0.5	`small`

Suppose the vectors of 'animal' and 'run' in this space are as follows:

	$\lvert\{\text{cats, kittens}\}\rangle$	$\lvert\{\text{miaow, purr}\}\rangle$	$\lvert\{\text{sleep, snore}\}\rangle$
$\overline{[\![\text{animal}]\!]}$	0.5	0.4	0.3
$\overline{[\![\text{run}]\!]}$	$0.6\,\lvert s_1\rangle$	$0.4\lvert s_2\rangle$	$0.2\lvert s_3\rangle$

In the first row, 0.5 is the degree to which the word 'animal' has had the feature {cats, kittens} in the corpus, for example, due to the fact that it has occurred in sentences such as 'a cat is an animal' and 'kittens are small and cute animals'. Similarly, 0.4 is the degree to which 'animal' has feature $\lvert\{\text{miaow, purr}\}\rangle$ and 0.3 the degree to which it has feature $\lvert\{\text{sleep, snore}\}\rangle$. The values of the $\lvert s_i\rangle$ in $\overline{[\![\text{run}]\!]}$ depend on the concrete instantiation of the sentence dimensions; to keep things simple we do not instantiate them.

One computes the vector interpretations of $\overline{[\![\text{all}]\!]}(\overline{[\![\text{animals}]\!]})$ and $\overline{[\![\text{some}]\!]}(\overline{[\![\text{animals}]\!]})$ by linearly expanding $\overline{[\![\text{all}]\!]}$ and $\overline{[\![\text{some}]\!]}$ over the vector of 'animal':

$$\begin{aligned}
\overline{[\![\text{all}]\!]}(\overline{[\![\text{animals}]\!]}) &= 0.5\overline{[\![\text{all}]\!]}(\lvert\{\text{cats, kittens}\}\rangle) + 0.4\overline{[\![\text{all}]\!]}(\lvert\{\text{miaow, purr}\}\rangle) \\
&\quad +0.3\overline{[\![\text{all}]\!]}(\lvert\{\text{sleep, snore}\}\rangle) \\
\overline{[\![\text{some}]\!]}(\overline{[\![\text{animals}]\!]}) &= 0.5\overline{[\![\text{some}]\!]}(\lvert\{\text{cats, kittens}\}\rangle) + 0.4\overline{[\![\text{some}]\!]}(\lvert\{\text{miaow, purr}\}\rangle) \\
&\quad +0.3\overline{[\![\text{some}]\!]}(\lvert\{\text{sleep, snore}\}\rangle)
\end{aligned}$$

The interpretations of the quantified sentences 'all animals run' and 'some animals run' are be computed by substituting these numbers in the formula $\sum_{ijk} \sum_{B \in [\![d]\!]([\![n]\!])} c_i^n c_{jk}^{vp} c_B^d \langle B \mid A_i \cap A_j \rangle |s_k\rangle$. It results in the following summands of their corresponding linear expansions:

	$\|s_1\rangle$	$\|s_2\rangle$	$\|s_3\rangle$
$\overline{[\![\text{all animals run}]\!]}$	$0.5 \times (0.4 \times 0.9 + 0.3 \times 0.2)$	$0.4 \times (0.5 \times 0.7 + 0.3 \times 0.3)$	$0.3(0.5 \times 0.5 + 0.4 \times 0.3)$
$\overline{[\![\text{some animals run}]\!]}$	$0.5 \times (0.4 \times 0.3 + 0.5 \times 0.2)$	$0.4 \times (0.5 \times 0.9 + 0.3 \times 0.5)$	$0.3(0.5 \times 0.6 + 0.4 \times 0.5)$

In the literature on *distributional inclusion hypothesis* (Geffet and Dagan 2005; Weeds et al. 1935) different types of orderings on feature vectors are used to model and experiment with word-level entailment. Wherein, a word 'v' entails a word 'w', written as '$v \vdash w$', if features of 'v' are also features of 'w'. The simplest such ordering is the point-wise ordering on vector dimensions. In our model, the point-wise ordering on the feature sets provide us with the following entailments:

$$\overline{[\![\text{all animals}]\!]} \vdash \overline{[\![\text{some animals}]\!]} \qquad \overline{[\![\text{all animals run}]\!]} \vdash \overline{[\![\text{some animals run}]\!]}$$

This opens the way to reason about entailment on quantified phrases and sentences compositionally and using statistical data from corpora of text. Implementing some of the above instantiations and experimenting with their applications to entailments on datasets constitute work in progress.

6. Conclusion and future work

After a review of the setting of distributional semantics and a context-free and pregroup grammatical formalisation of the fragment of language concerning quantified phrases and sentences (and the necessary preliminaries on compact closed categories and bialgebras), we developed an abstract compact closed categorical semantics for quantifiers with the help of bialgebras. We instantiated the abstract setting to the category of sets and relations and proved its equivalence to the truth theoretic semantics of generalised quantifier theory of Barwise and Cooper. We extended the existing instantiation of the categorical compositional distributional semantics to finite-dimensional vector spaces and linear maps to develop a corpus-based instantiation for our model. Implementing this setting on real data and experimenting with it constitute work in progress. Extending the theory to other sets of subsets, for example, down or up sets, rather than the full powerset is a future direction, as is developing a logic based on any of these collections (powerset vs down or up sets). Extending the grammar to more expressive fragments of English and addressing advanced language phenomena such as co-reference resolution, following the work of Preller (2014), are other future directions.

Acknowledgements. We thank the anonymous reviewers for their comments. Hedges thanks ESPRC for postdoctoral fellowship EP/N021282/1. Sadrzadeh thanks EPSRC for Career Acceleration Fellowship EP/J002607/1 and AFOSR for International Scientific Collaboration Grant FA9550-14-1-0079.

References

Ajdukiewicz, K. (1935). Die syntaktische konnexitat. *Studia Philosophica* **1** 1–27.

Bar-Hillel, Y. (1953). A quasi-arithmetical notation for syntactic description. *Language* **29** 47–58.

Bar-Hillel, Y., Gaifman, C. and Shamir, E. (1960). On categorial and phrase-structure grammars. *Bulletin of the Research Council of Israel* **9F**, 1–16.

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy* **4** 159–219.

Bonchi, F., Sobocinski, P. and Zanasi, F. (2014). Interacting bialgebras are Frobenius. In: Muscholl, A. (ed.) *Proceedings of FoSSaCS* 2014, vol. **8412,** Grenoble, France Springer, 351–365.

Bullinaria, J. A. and Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods* **39** 510–526.

Buszkowski, W. (1988). Generative power of categorial grammars. In: Oehrle, R., Bach, E. and Wheeler, D. (eds.) *Categorial Grammars and Natural Language Structures*, Studies in Linguistics and Philosophy, vol. **32**, Springer Netherlands, 69–94.

Buszkowski, W. (2001). Lambek grammars based on pregroups. In: *Logical Aspects of Computational Linguistics*, Lecture Notes in Computer Science, vol. **2099,** Springer Berlin Heidelberg, 95–109.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory* **2** 113–124.

Clark, S., Coecke, B. and Sadrzadeh, M. (2008). A compositional distributional model of meaning. In: Bruza, P., Lawless, W., Coecke, B. (eds.) *Proceedings of the Second Symposium on Quantum Interaction (QI)*, Oxford University, College Publications, 133–140.

Clark, S., Coecke, B. and Sadrzadeh, M. (2013). The Frobenius anatomy of relative pronouns. In: Kornai, A., Kuhlmann, M. (eds.), *13th Meeting on Mathematics of Language (MoL).*, Sofia, Bulgaria, ACL, 41–51.

Clark, S. and Pulman, S. (2007). Combining symbolic and distributional models of meaning. In: Bruza, P., Lawless, W., van Rijsbergen, C. J. (eds.) *Proceedings of the AAAI Spring Symposium on Quantum Interaction*, Technical Report SS-07-08, Stanford University, AAAI Press, 52–55.

Coecke, B., Grefenstette, E. and Sadrzadeh, M. (2013). Lambek vs. Lambek: Functorial vector space semantics and string diagrams for Lambek calculus. *Annals of Pure and Applied Logic* **164**(11) 1079–1100, special issue on Seventh Workshop on Games for Logic and Programming Languages (GaLoP VII).

Coecke, B., Sadrzadeh, M. and Clark, S. (2010). Mathematical foundations for distributed compositional model of meaning. Lambek Festschrift. *Linguistic Analysis* **36** 345–384.

Firth, J. (1957). A synopsis of linguistic theory 1930–1955. In: Palmer, F. R. (ed.) *Studies in Linguistic Analysis* Longmans, 168–205.

Frege, G. (1948). On sense and reference. *The Philosophical Review* **57** 209–230.

Geffet, M. and Dagan, I. (2005). The distributional inclusion hypotheses and lexical entailment. In: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, Association for Computational Linguistics, 107–114.

Grefenstette, E., Dinu, G., Zhang, Y., Sadrzadeh, M. and Baroni, M. (2013). Multi-step regression learning for compositional distributional semantics. In: 10th International Conference on Computational Semantics (IWCS). Postdam.

Grefenstette, E. and Sadrzadeh, M. (2011). Experimental support for a categorical compositional distributional model of meaning. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, *Computational Linguistics* **41**. MIT Press 1394–1404.

Grefenstette, E. and Sadrzadeh, M. (2015). Concrete models and empirical evaluations for the categorical compositional distributional model of meaning. *Computational Linguistics* **41** 71–118.

Harris, Z. (1954). Distributional structure. *Word* **10**, 146–162, Routledge.

Kartsaklis, D. (2015). *Compositional Distributional Semantics with Compact Closed Categories and Frobenius Algebras*. PhD thesis, Department of Computer Science, University of Oxford.

Kartsaklis, D., Kalchbrenner, N. and Sadrzadeh, M. (2014). Resolving lexical ambiguity in tensor regression models of meaning. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Volume **2**: Short Papers, June 22–27, 2014, ACL 2014, Baltimore, MD, USA, 212–217.

Kartsaklis, D. and Sadrzadeh, M. (2013). Prior disambiguation of word tensors for constructing sentence vectors. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)* Association for Computational Linguistics, 1590–1601.

Kartsaklis, D., Sadrzadeh, M. and Pulman, S. (2012). A unified sentence space for categorical distributional-compositional semantics: Theory and experiments. In: *Proceedings of 24th International Conference on Computational Linguistics (COLING 2012): Posters*, Mumbai, India, 549–558.

Kartsaklis, D., Sadrzadeh, M., Pulman, S. and Coecke, B. (2013). Reasoning about meaning in natural language with compact closed categories and Frobenius algebras. In: Chubb, A., Eskandarian, J. and Harizanov, V. (eds.) *Logic and Algebraic Structures in Quantum Computing and Information*, Cambridge University Press. 199–222.

Kelly, G. and Laplaza, M. (1980). Coherence for compact closed categories. *Journal of Pure and Applied Algebra* **19**, 193–213. http://www.sciencedirect.com/science/article/pii/0022404980901012

Kock, A. (1972). Strong functors and monoidal monads. *Archiv der Mathematik* **23** 113–120.

Lambek, J. (1958). The mathematics of sentence structure. *American Mathematics Monthly* **65** 154–170.

Lambek, J. (1997). Type grammars revisited. In: *Proceedings of LACL 97*, Lecture Notes in Artificial Intelligence, vol. **1582**, Springer Verlag. 1–27.

Lambek, J. (2008). *From Word to Sentence: A Computational Algebraic Approach to Grammar*. Polimetrica.

Lambek, J. (2010). Compact monoidal categories from linguistics to physics. In: Coecke, B. (ed.) *New Structures for Physics*, Lecture Notes in Physics, Springer, 451–469.

Landauer, T. and Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* **104** 211–240.

Lapesa, G. and Evert, S. (2014). A large scale evaluation of distributional semantic models: Parameters, interactions and model selection. *Transactions of the Association for Computational Linguistics* **2** 531–545.

Lin, D. (1998). Automatic retrieval and clustering of similar words. In: *Proceedings of the 17th international conference on Computational linguistics*, vol. **2**, Association for Computational Linguistics, 768–774.

Lund, K. and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods Instruments and Computers* **28** (2) 203–208.

McCurdy, M. (2012). Graphical methods for Tannaka duality of weak bialgebras and weak Hopf algebras. *Theory and Applications of Categories* **26** (9) 233–280.

Milajevs, D., Kartsaklis, D., Sadrzadeh, M. and Purver, M. (2014). Evaluating neural word representations in tensor-based compositional settings. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, 708–719.

Mitchell, J. and Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science* **34** 1388–1439.

Montague, R. (1970). English as a formal language. In: Visentini, B. (ed.) *Linguaggi nella Società e nella Tecnica*, Edizioni di Comunità, 189–224.

Polajnar, T., Fagarasan, L. and Clark, S. (2014). Reducing dimensions of tensors in type-driven distributional semantics. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1036–1046. Association for Computational Linguistics.

Preller, A. (2013). From logical to distributional models. In: *Proceedings of the 10th International Workshop on Quantum Physics and Logic*, QPL 2013, Castelldefels (Barcelona), Spain, July 17–19, 113–131.

Preller, A. (2014). Natural language semantics in biproduct dagger categories. *Journal of Applied Logic* **12** (1) 88–108. https://doi.org/10.1016/j.jal.2013.08.001

Preller, A. and Lambek, J. (2007). Free compact 2-categories. *Mathematical Structures in Computer Science* **17** 309–340.

Preller, A. and Sadrzadeh, M. (2010). Bell states and negative sentences in the distributed model of meaning. In: Coecke, B., Panangaden, P., Selinger, P. (eds.) *Proceedings of the 6th QPL Workshop on Quantum Physics and Logic*, Electronic Notes in Theoretical Computer Science, University of Oxford. 141–153.

Preller, A. and Sadrzadeh, M. (2011). Semantic vector models and functional models for pregroup grammars. *Journal of Logic Language and Information* **20** 419–443.

Rubenstein, H. and Goodenough, J. (1965). Contextual correlates of synonymy. *Communications of the ACM* **8** (10) 627–633.

Rypacek, O. and Sadrzadeh, M. (2014). A low-level treatment of generalised quantifiers in categorical compositional distributional semantics. In: *Joint Proceedings of the Second International Workshop on Natural Language and Computer Science (NLCS14) and First International Workshop on Natural Language Services for Reasoners (NLSR 2014)*, TR 2014/02, Center for Informatics and Systems of the University of Coimbra, 165–177.

Sadrzadeh, M., Clark, S. and Coecke, B. (2013). Frobenius anatomy of word meanings i: Subject and object relative pronouns. *Journal of Logic and Computation* **23** 1293–1317.

Sadrzadeh, M., Clark, S. and Coecke, B. (2014). Frobenius anatomy of word meanings 2: Possessive relative pronouns. *Journal of Logic and Computation* **26** 785–815.

Salton, G., Wong, A. and Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM* **18** 613–620.

Schuetze, H. (1998). Automatic word sense discrimination. *Computational Linguistics* **24** (1) 97–123.

Turney, P. D. (2006). Similarity of semantic relations. *Computational Linguistics* **32** (3) 379–416.

van Benthem, J. (1987). Categorial grammar and lambda calculus. In: Skordev, Dimiter G. (ed.) *Mathematical Logic and Its Applications*, Springer, 39–60.

Weeds, J., Weir, D. and McCarthy, D. (2004). Characterising measures of lexical distributional similarity. In: *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Association for Computational Linguistics. 1015.

Cite this article: Hedges J and Sadrzadeh M (2019). A generalised quantifier theory of natural language in categorical compositional distributional semantics with bialgebras. *Mathematical Structures in Computer Science* **29**, 783–809. https://doi.org/10.1017/S0960129518000324