A Generalised Quantifier Theory of Natural Language in Categorical Compositional Distributional Semantics with Bialgebras

Jules Hedges and Mehrnoosh Sadrzadeh

School of Electronic Engineering and Computer Science,Queen Mary University of London
j.hedges, m.sadrzadeh@qmul.ac.uk

1 Introduction

Categorical compositional distributional semantics is a model of natural language that combines the statistical vector space models of words with the compositional models of grammar. Recently in the paper http://arxiv.org/pdf/1602.01635.pdf, submitted for publication elsewhere, we formalised in it the generalised quantifier theory of natural language, due to Barwise and Cooper [1]. The underlying setting is that of a compact closed category with bialgebras [2, 4]. We developed an abstract categorical compositional semantics, then instantiated it to sets and relations and to finite dimensional vector spaces and linear maps. We proved the equivalence of the relational instantiation to the truth theoretic semantics of generalized quantifiers and provided concrete corpus-based instantiations. The contributions of our work is three fold: first, it is the first time quantifiers are formalised in categorical compositional distributional semantics, second, it is the first time bialgebras are used, third, it is the first time equivalence of the setting to a truth-theoretic semantics is formally proved (and not just exemplified).

2 Preliminaries

Vector Models of Natural Language. Given a corpus of text, a set of contexts and a set of target words, these models work with a so called *co-occurrence matrix*. This has at each of its entries 'the degree of co-occurrence between the target word and the context' [6]. 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. A feature represents a set of words that together express a pertinent linguistic property of a word. Given an $m \times n$ co-occurrence matrix, every target word t can be represented by a row vector of length n. The lengths of the corpus and window are parameters of the model, as are the sizes of the feature and target sets. We denote a vector model of natural language produced in this way with V_Σ, where Σ is the set of contexts and V_Σ is the vector space spanned by it.

Generalised Quantifier Theory of Natural Language. Consider the fragment of English generated by the following context free grammar:

S $\to$ NP VP	NP $\to$ John, Mary	VP $\to$ sneeze, sleep
VP $\to$ V NP	N $\to$ cat, dog, man	V $\to$ love, kiss
NP $\to$ Det N		Det $\to$ a, some , all, no, most, few, one, two

According to [1], a generalised quantifier model for the language generated by this grammar is a pair $(U, [\![\ ]\!])$, where U is a universal reference set and $[\![\ ]\!]$ is an interpretation function. The interpretation of a determiner d generated by 'Det $\to$ d' is a map $[\![d]\!]\colon \mathcal{P}(U) \to \mathcal{P}\mathcal{P}(U)$, which assigns to each $A \subseteq U$, a family of subsets of U. The images of these interpretations are referred to as *generalised quantifiers*. Examples are $[\![\text{some}]\!](A) = \{X \subseteq U \mid X \cap A \neq \emptyset\}$ and $[\![\text{every}]\!](A) = \{X \subseteq U \mid A \subseteq X\}$.

Nouns, noun phrases and verb phrases are interpreted as subsets of U, verbs as subsets of $U \times U$. The interpretation of expressions of the form $[\![\text{Det N}]\!]$ are $[\![d]\!]([\![n]\!])$, where $X \in [\![d]\!]([\![n]\!])$ iff $X \cap [\![n]\!] \in [\![d]\!]([\![n]\!])$, for Det $\to d$ and N $\to n$. The meaning of a sentence is its truth value, defined as follows:

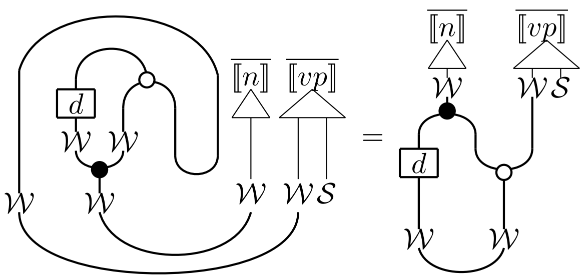
Definition 1. *The* meaning *of a sentence s in generalised quantifier theory is true iff* $[\![s]\!] \neq \emptyset$.

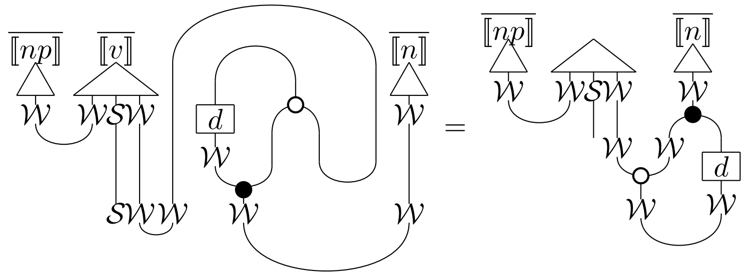
3 Abstract Compact Closed Semantics

An abstract compact closed categorical model for the language generated by a context free grammar is a tuple $(\mathcal{C}, \mathcal{W}, \mathcal{S}, \overline{[\![\ ]\!]})$, where $\mathcal{C}$ is a self adjoint compact closed category with two distinguished objects $\mathcal{W}$ and $\mathcal{S}$ with bialgebras on them, and $\overline{[\![\ ]\!]}$ is a strongly monoidal functor, defined as follows:

$$\overline{[\![x]\!]} := \begin{cases} \mathcal{W} & x \in P, x = p, x = n \\ \mathcal{S} & x \in P, x = s \\ I \to \overline{[\![\sigma(x)]\!]} & x \in P, A \to x \text{ is an atomic rule in } \mathcal{R} \text{ and } A \in \{\text{NP, N, VP, V}\} \\ \overline{[\![\sigma(x)]\!]} \to \overline{[\![\sigma(x)]\!]} & \text{same as above but } A = \text{Det} \\ I \to \overline{[\![\sigma(x)]\!]} & x \in T \end{cases}$$

Here $\overline{[\![\ ]\!]} \colon P \to \mathcal{C}$, for P the pregroup obtained via the cfg-pregroup transformation σ, for details please see the main paper. Meaning of sentences with a quantified subject and object are each depicted below, where the box containing d represents the morphism corresponding to $\overline{[\![d]\!]}$:





4 Truth Theoretic Interpretation in Rel

Rel, the cateogry of sets and relations, is a self adjoint compact closed category, with $\otimes$ the cartesian product and $I = \{\star\}$ its unit and with adjoints as identities on objects. For an object in Rel of the form $W = \mathcal{P}(U)$, we give W a bialgebra structure by taking

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

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. This instantiation is the tuple $(\text{Rel}, \mathcal{P}(\mathcal{U}), \{\star\}, \overline{[\![\ ]\!]})$, where the interpretations of terminals are defined by the following relations:

- The interpretation of a terminal x generated by any of N,NP, and VP is $\star\overline{[\![x]\!]}A \iff A = [\![x]\!]$.
- The interpretation of a terminal x generated by 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 Det is $A\overline{[\![d]\!]}B \iff B \in [\![d]\!](A)$.

It is not hard to show that $(\text{Rel}, \mathcal{P}(\mathcal{U}), \{\star\}, \overline{[\![\ ]\!]})$ provides us with the same meaning as defined in the generalised quantifier theory. This is made formal as follows:

Definition 2. *The interpretation of a quantified sentence s is true in* $(\text{Rel}, \mathcal{P}(\mathcal{U}), \{\star\}, \overline{[\![\ ]\!]})$ *iff* $\star\overline{[\![s]\!]}\star$.

Theorem 1. $\star\overline{[\![s]\!]}\star$ *iff* $[\![s]\!] \neq \emptyset$.

The relational model embeds into a vector spaces model using the usual embedding of sets and relations into vector spaces and linear maps, resulting in the tuple $(\text{FdVect}, V_{\mathcal{P}(\mathcal{U})}, V_{\{\star\}}, \overline{[\![\ ]\!]})$, for $V_{\mathcal{P}(\mathcal{U})}$ the free vector space generated over the set of subsets of $\mathcal{U}$ and $V_{\{\star\}}$ the one dimensional space. In this tuples, the terminals generated by N, NP, VP, and V rules are interpreted as: $\overline{[\![x]\!]}(\star) = |[\![x]\!]\rangle$. The interpretation of a terminal d generated by the Det rule is $\overline{[\![d]\!]}(|A\rangle) = \sum_{B\in[\![d]\!](A)} |B\rangle$ on subsets A of $\mathcal{U}$. This embedding provides us with an imitation of the truth theoretic model of Rel in FdVect.

Definition 3. *The interpretation of a quantified sentence s is true in* $(\text{FdVect}, V_{\mathcal{P}(\mathcal{U})}, V_{\{\star\}}, \overline{[\![\ ]\!]})$ *iff* $\overline{[\![s]\!]}(\star) \neq 0$.

Corollary 1. $\overline{[\![s]\!]}(\star) \neq 0$ *in* $(\text{FdVect}, V_{\mathcal{P}(\mathcal{U})}, V_{\{\star\}}, \overline{[\![\ ]\!]})$ *iff* $\star\overline{[\![s]\!]}\star$ *in* $(\text{Rel}, \mathcal{P}(\mathcal{U}), \{\star\}, \overline{[\![\ ]\!]})$.

5 Corpus-Based Instantiation in FdVect

A corpus-based vector space instantiation of the model is obtained via a construction similar to the above, but with real number (rather than boolean) weights, retrievable from corpora of text using distributional methods. The non-quantified part of this instantiation closely follows that of the general model of [3] and the concrete constructions of [5].

Definition 4. *The distributional instantiation of the model to* FdVect *is the tuple* $(\text{FdVect}, V_{\mathcal{P}(\Sigma)}, Z, \overline{[\![\ ]\!]})$*, 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:*

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

Examples are obtained by taking $Z = \mathbb{R}$ for degrees of truth, or supposing $\mathcal{S}$ contains sentence dimensions according to concrete construction of [5] and taking $Z = V_{\mathcal{S}}$. In the tuple obtained by this latter case, one can embed the distributional vectors via singleton constructions, e.g. 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 $\sum_i c_i^x |v_i\rangle$. One can also work with sets of subsets consisting of lemmas or features as basis. Implementing this setting on real data and experimenting with it constitutes work in progress.

References

1. Barwise, J., Cooper, R.: Generalized quantifiers and natural language. Linguistics and Philosophy 4, 159–219 (1981)
2. Bonchi, F., Sobocinski, P., Zanasi, F.: Interacting bialgebras are Frobenius. In: Proceedings of FoSSaCS 2014. vol. 8412, pp. 351–365. Springer (2014)
3. Coecke, B., Sadrzadeh, M., Clark, S.: Mathematical Foundations for Distributed Compositional Model of Meaning. Lambek Festschrift. Linguistic Analysis 36, 345–384 (2010)
4. Coecke, B., Duncan, R.: Interacting Quantum Observables. Automata, Languages and Programming pp. 298–310 (2008)
5. Grefenstette, E., Sadrzadeh, M.: Experimental support for a categorical compositional distributional model of meaning. In: Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1394–1404 (2011)
6. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM 18, 613–620 (1975)