From Functional to Distributional Models

Anne Preller

From Functional to Distributional Models

Anne Preller * †

Abstract

This paper defines logical functional models in the category of finite dimensional vector spaces over the field of real numbers. The functional models are given by functors from the free compact closed lexical category associated to a pregroup grammar. Any functional model is completely compositional and includes first order predicate logic. Each functional model is mapped to a vector space model in the sense of Kartsaklis, Sadrzadeh, et al. via a 'canonical' probability of the functional model. The logical connectives of the functional model are transferred to the vector space model where they become algebraic operators. The algebraic logical operators subsume the quantum logical operators of van Rijsbergen. The transfer provides an insight into how logical operators and other function words interpreted in abstract semantics change when implemented in vector space semantics.

Keywords:compositional semantics for natural language logical models vector space models compact closed category quantum logic

1 Content, Definitions and Context

Semantic models for natural language vary from logical functional models, such as first-order models or Montague models, to conceptual models. Conceptual models comprise variants of higher order type theory, (Asher 2011) and vector space models based on the geometry of quantum logic, (Widdows 2004), (Rijsbergen 2004). They all involve reasoning, an essential ingredient of any compositional semantics according to (Kracht 2007).

This paper traces the switch from the logical functional models based on pregroup grammars to the distributional vector space models considered in (Kartsaklis et al. 2013) and shows how the classical logic of functional models changes to the quantum logic of vector space models. It also proposes a guideline helping to extend distributional vector semantics to function words.

A *compact closed* category is a monoidal category in which every object has a right and left adjoint. It is not required to be symmetric. A *lexical category* is the free compact closed category $\mathcal{L}(\mathcal{B})$ generated by a partially ordered set $\mathcal{B}$

*LIRMM, 161 rue Ada, 34194 Montpellier, France. E-mail: preller@lirmm.fr

†to appear in the Proceedings of the 10th conference of Quantum Physics and Logic (QPL 2013)

and a finite number of new, 'distinguished' morphisms occurring in the lexicon of a pregroup grammar. A *functor* from a lexical category to the category of finite dimensional vector spaces $\mathbf{Vect}_{\mathbb{R}}$ is always a functor that commutes with the tensor product and maps the right and left adjoints to the dagger. Moreover, a *concept word* is a noun, an adjective or a verb. Any word that is not a concept word is called a *noise word*. The notation V_A designates a vector space with a fixed set A of orthogonal basis vectors.

The pregroup grammar defines the meaning of a grammatical string of words as that morphism of the lexical category that is the composition of the reduction of the string with the tensor product of the words. An example of this definition is given at the end of (Preller 2005). Obviously, any functor that is defined for all words in the lexicon provides a compositional interpretation for all grammatical strings of words.

The important question concerning compositional semantics is how it handles the logical content of words and reasoning. To compare the logical models with vector space models I show that any logical model 'lives' in $\mathbf{FVect}_{\mathbb{R}}$ under the form of a functor. A functional vector space model interprets the sentence type by a two-dimensional space S, the noun phrase type by an arbitrary space, say V_A, and every noise word in the lexicon by a linear map. In particular, propositional connectives and quantifiers are rendered by linear maps. For any functional vector model $\mathcal{F}$ and any set of concept words in the lexicon there is a vector space model V_C and a map $\mathcal{J}_C$ that assigns to every $\mathcal{F}$-value of a concept word a vector in V_C. The coefficients of this vector are conditional probabilities with respect to the elements of C. The propositional connectives on V_C are given as algebraic operators, both for vectors and for self-adjoint endomorphisms. They subsume the quantum logical operations of (Rijsbergen 2004) and generalise the logical operators of predicate logic, because the choice $C = A$ makes $\mathcal{J}_A$ an isomorphism from the Boolean algebra of predicates on V_A onto the Boolean algebra of projectors of V_A for which A forms a set of eigenvectors. For any C, $\mathcal{J}_C$ reflects the consequence relation and commutes with negation.

Under sufficient conditions, it commutes with the binary operators and replaces composition of endomorphisms by the pointwise product $\odot$ of vectors $\mathcal{J}_C(\mathcal{F}(\mathtt{w}_1 \circ \mathtt{w}_2)) = \mathcal{J}_C(\mathcal{F}(\mathtt{w}_1)) \odot \mathcal{J}_C(\mathcal{F}(\mathtt{w}_2))$.

For any space V_C there is a unique functor defined on the free pregroup generated by a partially ordered set $\mathcal{B}$ that maps all elements in $\mathcal{B}$ to V_C and all inequalities of basic types to the identity of V_C. The free pregroup is a strict subcategory of every lexical pregroup category. Hence, (Kartsaklis et al. 2013)'s claim that they use a functor from the free pregroup can be strengthened: For any set W of concept words and any map F from W into the set of vectors ofV_C, they use a functor G_F from the lexical category defined by the pregroup lexicon listing the words in W. The functor is determined by the requirement that it extends the unique functor and satisfies $F = \mathcal{I}(\mathcal{G}(w))$ for all $w \in W$. The map $\mathcal{I}$ is essentially the one-to-one correspondence of diagonal matrices and vectors. The existence of G_F is based on two facts. A grammatical string of words yanks in lexical category to a composition on basic morphisms and, second, $\mathcal{I}(\mathcal{G}(\mathtt{w}_1 \circ \mathtt{w}_2)) = \mathcal{I}(\mathcal{G}(\mathtt{w}_1)) \odot \mathcal{I}(\mathcal{G}(\mathtt{w}_2))$, where the $\mathtt{w}_i$ is the morphism given to the concept word w_i by the lexicon. The functions F considered by the authors are determined by a probability distribution based on context. Any functional vector model $\mathcal{F}$ defines a functor G_F via the restriction F of $\mathcal{J}_C \circ \mathcal{F}$ to W. If $\mathcal{F}$ renders the statements of the text true the G_F-values for words in W have truth-theoretical probabilistic content. This extends distributional vector semantics to vector semantics based on a contextual probability and vector semantics based on a truth-theoretical probability.

Lifting the 'yanking' to the level of the lexical category also gives an insight into how noise words work in different kinds of semantics. For example, the distinguished morphism $\mathtt{and} : \boldsymbol{s} \otimes \boldsymbol{s} \to \boldsymbol{s}$ occurring in the lexicon for the function word *and* becomes the linear map μ_{V_C} given by the Frobenius algebra, because

$$\mathcal{J}_C(\mathcal{F}(\mathtt{and} \circ (\mathtt{w}_1 \otimes \mathtt{w}_2))) = \mathcal{J}_C(\mathcal{F}(\mathtt{w}_1)) \wedge \mathcal{J}_C(\mathcal{F}(\mathtt{w}_2))) = \mathcal{J}_C(\mathcal{F}(\mathtt{w}_1)) \odot \mathcal{J}_C(\mathcal{F}(\mathtt{w}_2)).$$

On the other hand, such a functor cannot be extended to include the basic morphism $\mathtt{not} : \boldsymbol{s} \to \boldsymbol{s}$ intervening in the lexical entries of *not*, *no*. Any functor maps it to an endomorphism of V_C. This would result in a logic where the negation of false is false, because the zero-map when composed and endomorphism is again the zero-map.

2 Syntactic and semantic graphs for pregroup grammars

A pregroup grammar is determined by a partially ordered set $\mathcal{B}$ and a lexicon. The *lexicon* of a pregroup grammar is a finite list of pairs *word* :: $\overline{\mathtt{word}} : I \to T$, where $\overline{\mathtt{word}} : I \to T$ designates a morphism in the language of compact closed categories and T is an object of the free pregroup introduced in (Lambek 1999). The free pregroups and the pregroup dictionaries in (Lambek 2008) consisting of *word* : T do not provide semantics for pregroup grammars. The semantics of (Clark et al. 2008) interprets every entry *conceptword* : T by a vector in $\mathbf{Vect}_{\mathbb{R}}$. This is a special case of the semantics provided by the lexical category $\mathcal{L}(\mathcal{B})$

The objects of $\mathcal{L}(\mathcal{B})$ are called *types*, among them are the elements of $\mathcal{B}$, called *basic types*. The *simple* types are the basic types and their iterated right and left adjoints. Any object of $\mathcal{L}(\mathcal{B})$ can be written as a finite tensor product of simple types.

The *basic morphisms* are the (in)equalities $in_{\boldsymbol{ab}} : \boldsymbol{a} \to \boldsymbol{b}$ between elements of $\mathcal{B}$ and the distinguished morphisms given by the lexicon. Every morphism of the free category can be designated by a graph where all links are labelled by basic morphisms. The overlinks represent names of basic morphisms and the underlinks conames of basic morphisms. All paths of the graph have length 1. By convention, the graph displays the domain at the top, the codomain at the bottom. If the label is an (in)equality of $\mathcal{B}$ it is omitted. For example, a basic morphism $f : A \to B$ and its right and left adjoint are represented by the graphs

$$f : A \to B = \begin{tikzcd} A \arrow[d, "f"] \\ B \end{tikzcd} \qquad f^r : B^r \to A^r = \begin{tikzcd} B^r \\ A^r \arrow[u, "f"'] \end{tikzcd} \qquad f^\ell : B^\ell \to A^\ell = \begin{tikzcd} B^\ell \\ A^\ell \arrow[u, "f"] \end{tikzcd}$$

The tensor product is not required to be symmetric so that any object A has a right unit $\eta_A : I \to A^r \otimes A$ and a left unit, $\eta_{A^\ell} : I \to A \otimes A^\ell$ and any morphism $f : A \to B$ has two names

right name

$$\begin{tikzcd}[column sep=small] & I & \\ A^r \arrow[rr, bend left, "f"] & \otimes & B \end{tikzcd} = \eta_f \quad = (f^r \otimes 1_B) \circ \eta_B = \begin{tikzcd}[column sep=small] & I & \\ B^r \arrow[rr, bend left] & \otimes & B \arrow[d] \\ A^r \arrow[u, "f"] & \otimes & B \end{tikzcd}$$

left name

$$\begin{tikzcd}[column sep=small] & I & \\ B & \otimes & A^\ell \arrow[ll, bend right, "f"'] \end{tikzcd} = \eta_{f^\ell} = (f \otimes 1_{A^\ell}) \circ \eta_{A^\ell} = \begin{tikzcd}[column sep=small] & I & \\ A \arrow[d, "f"'] & \otimes & A^\ell \arrow[ll, bend right] \\ B & \otimes & A^\ell \arrow[u] \end{tikzcd} \quad .$$

The lefthand graphs above are the connection-free abbreviations for the right-hand graphs, which connects the graph of a unit with the graph of a tensor product. Similarly, the algebraic name is a composition-free abbreviation for a compound algebraic expression with one composition symbol. Analogous definitions and notations apply to counits and conames.

Composition of morphisms is computed graphically by connecting the corresponding graphs at their joint interface and replacing every maximal path by a simple link, labelled by the composition of the labels. The paths of the resulting graph have length 1 and every link is a name, a coname or an eventually iterated adjoint of basic morphisms.

The *lexicon* of a pregroup grammar is a list of pairs *word* :: $\overline{\texttt{word}}$, where $\overline{\texttt{word}} : I \to T$ designates a morphism in the language of compact closed categories. For example,

$$\begin{array}{lll} \textit{new} & :: \overline{\texttt{new}} & : I \to \boldsymbol{n}_2 \otimes {\boldsymbol{c}_2}^\ell \\ \textit{no} & :: \overline{\texttt{no}} & : I \to \boldsymbol{s} \otimes \boldsymbol{s}^\ell \otimes \boldsymbol{n}_2 \otimes {\boldsymbol{c}_2}^\ell \\ \textit{triangles} & :: \overline{\texttt{triangles}} & : I \to \boldsymbol{c}_2 \\ \textit{are} & :: \overline{\texttt{are}} & : I \to {\boldsymbol{n}_2}^r \otimes \boldsymbol{s} \otimes \boldsymbol{p}^\ell \otimes \boldsymbol{n}_2 \\ \textit{blue} & :: \overline{\texttt{blue}} & : I \to \boldsymbol{n}^r \otimes \boldsymbol{p} \end{array}$$

where $\boldsymbol{c}_2 \leq \boldsymbol{n}_2 \leq \boldsymbol{n}$ are the basic types for plural common nouns, plural noun phrases and noun phrases where the number does not matter, in that order.

The basic types $\boldsymbol{p} \leq \boldsymbol{s}$ correspond to predicative adjectives and to sentences.

$$\overline{\texttt{no}} = \eta_{\texttt{not}^\ell} \otimes \eta_{in^\ell} = \begin{array}{c} I \\ \overset{\texttt{not}}{\boldsymbol{s} \otimes \boldsymbol{s}^\ell} \otimes \boldsymbol{n}_2 \otimes {\boldsymbol{c}_2}^\ell \end{array} \qquad \overline{\texttt{triangles}} = \begin{array}{c} I \\ \downarrow \\ \overset{\texttt{triangles}}{\boldsymbol{c}_2} \end{array}$$

$$\overline{\texttt{are}} = (1_{{\boldsymbol{n}_2}^r} \otimes \eta_{in^\ell_{\boldsymbol{ps}}} \otimes 1_{\boldsymbol{n}}) \circ \eta_{\boldsymbol{n}_2} = \begin{array}{c} I \\ {\boldsymbol{n}_2}^r \otimes \boldsymbol{n} \\ \nearrow \qquad \searrow \\ {\boldsymbol{n}_2}^r \otimes \boldsymbol{s} \otimes \boldsymbol{p}^\ell \otimes \boldsymbol{n} \end{array} = \begin{array}{c} I \\ {\boldsymbol{n}_2}^r \otimes \boldsymbol{s} \otimes \boldsymbol{p}^\ell \otimes \boldsymbol{n} \end{array}$$

$$\overline{\texttt{blue}} = \begin{array}{c} I \\ \overset{\texttt{blue}}{\boldsymbol{n}^r \otimes \boldsymbol{p}} \end{array} \qquad \overline{\texttt{new}} = \begin{array}{c} I \\ \overset{\texttt{new}}{\boldsymbol{n}_2 \otimes {\boldsymbol{c}_2}^\ell} \end{array} .$$

The distinguished morphisms introduced by this lexicon are $\texttt{triangles} : I \to \boldsymbol{c}_2$, $\texttt{not} : \boldsymbol{s} \to \boldsymbol{s}$, $\texttt{blue} : \boldsymbol{n} \to \boldsymbol{p}$ and $\texttt{new} : \boldsymbol{c}_2 \to \boldsymbol{n}_2$.

Note that the distinguished morphism corresponding to a concept word is entirely determined by the entry *word* : T in the pregroup lexicon. If $T = \boldsymbol{a}^r \otimes \boldsymbol{b}$ the expression $\overline{\texttt{word}}$ is the right name of a unique morphism $\texttt{word} : \boldsymbol{a} \to \boldsymbol{b}$, if $T = \boldsymbol{a} \otimes \boldsymbol{b}^\ell$ it is the left name of a unique morphism $\texttt{word} : \boldsymbol{b} \to \boldsymbol{a}$ and so on. In opposition, function words may introduce more that one distinguished morphism. For example, the pregroup lexicon of (Preller & Sadrzadeh 2011) defines the relative pronoun *who* as the name of $(1_{\boldsymbol{n}} \otimes \texttt{who}) \circ (d_{\boldsymbol{n}} \otimes 1_{\boldsymbol{s}})$, hence introduces two distinguished morphisms, $\texttt{who} : \boldsymbol{n} \otimes \boldsymbol{s} \to \boldsymbol{n}$ and $d_{\boldsymbol{n}} : \boldsymbol{n} \to \boldsymbol{n} \otimes \boldsymbol{n}$. The latter is mapped by a functor to the 'diagonal' $d_{\mathcal{F}(\boldsymbol{n})} : \mathcal{F}(\boldsymbol{n}) \to \mathcal{F}(\boldsymbol{n}) \otimes \mathcal{F}(\boldsymbol{n})$, which according to the definition (3) coincides with $\sigma_{\mathcal{F}(gn)}$.

A string of words $word_1 \ldots word_n$ is *grammatical* if there are entries $word_i$ :: $\overline{\texttt{word}}_i : I \to T_i$ in the lexicon, a basic type $\boldsymbol{b}$ and a reduction $r : T_1 \ldots T_n \to \boldsymbol{b}$. Any successful syntactical analysis of a pregroup grammar results in a reduction of compact bilinear logic or, equivalently, a morphism of the free compact closed category involving counits and basic morphisms only. For example, the sentence *No triangles are grammatical* is recognised by the morphism

$$\begin{array}{cccc} \textit{No} & \textit{triangles} & \textit{are} & \textit{blue} \end{array}$$

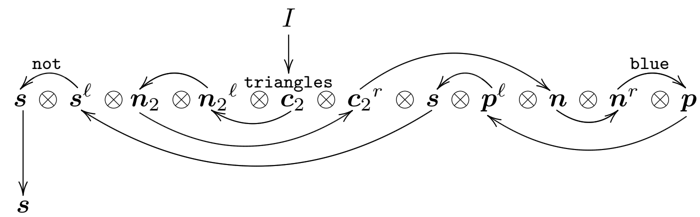
$$r = \begin{array}{l} \boldsymbol{s} \otimes \boldsymbol{s}^\ell \otimes \boldsymbol{n}_2 \otimes {\boldsymbol{n}_2}^\ell \otimes \boldsymbol{c}_2 \otimes {\boldsymbol{c}_2}^r \otimes \boldsymbol{s} \otimes \boldsymbol{p}^\ell \otimes \boldsymbol{n} \otimes \boldsymbol{n}^r \otimes \boldsymbol{p} \\ \downarrow \\ \boldsymbol{s} \end{array}$$

The *meaning* of the string $word_1 \ldots word_n$ recognised by the reduction r is

$$r \circ (\overline{\texttt{word}}_1 \otimes \ldots \otimes \overline{\texttt{word}}_n) .$$

For example,

$$r \circ (\overline{\texttt{no}} \otimes \overline{\texttt{triangles}} \otimes \overline{\texttt{are}} \otimes \overline{\texttt{blue}}) =$$



$$= \texttt{not} \circ \texttt{blue} \circ \texttt{triangles} : I \to \boldsymbol{s} \quad .$$

Clearly, a compact closed functor $\mathcal{F}$ from the lexical category into $\mathbf{FVect}_{\mathbb{R}}$ preserves the graphical representation of the morphisms of the lexical category. For example, if

$$N = \mathcal{F}(\boldsymbol{n}) = \mathcal{F}(\boldsymbol{n}_2) = \mathcal{F}(\boldsymbol{c}_2) \qquad S = \mathcal{F}(\boldsymbol{s}) = \mathcal{F}(\boldsymbol{p})$$

and the inequalities are mapped to identities then $\mathcal{F}(\texttt{not}) : S \to S$, $\mathcal{F}(\texttt{blue}) : N \to S$, $\mathcal{F}(\texttt{triangles}) : I \to N$ and

$$\mathcal{F}(r \circ (\overline{\texttt{no}} \otimes \overline{\texttt{triangles}} \otimes \overline{\texttt{are}} \otimes \overline{\texttt{blue}})) = \mathcal{F}(\texttt{not}) \circ \mathcal{F}(\texttt{blue}) \circ \mathcal{F}(\texttt{triangles}) \,. \quad (1)$$

If all basic types are mapped to a single space W and the inequalities to the identity of W then the functor maps the nouns to vectors of W, the adjectives and verbs to endomorphisms of W, but also the noise-word `not` . Assume now that the endomorphisms corresponding to adjectives and verbs are self-adjoint and have a common orthonormal basis of eigenvectors. If $\mathcal{I}$ is the one-to-one correspondence that maps vectors to diagonal matrices then the restriction of $F = \mathcal{I} \circ \mathcal{F}$ is a map that assigns to every concept word w a vector in W. The converse also holds due to the remark saying that basic morphisms are determined by the entry $w : T$ in the lexicon. Any map F from words to vectors in W defines a functor $\mathcal{F}$ from the lexical category (defined by the pregroup lexicon listing these words and nothing else) which maps every basic type to W and every inequality of basic types to the identity and satisfies $F = \mathcal{I} \circ \mathcal{F}$. The functor $\mathcal{F}$ guarantees compositionality for grammatical strings consisting of concept words only.

3 Functional Vector Semantics

In the following, all entries in the lexicon are interpreted in the category of finite-dimensional vector spaces over the field of real numbers.

A compact closed functor is a *functional vector space model* if it maps $\boldsymbol{s}$ and $\boldsymbol{p}$ to the two-dimensional space $S = V_{\{\top,\bot\}}$ with orthogonal basis vectors $\top$ and $\bot$, nouns to sums of basic vectors, verbs and predicative adjectives to ‘predicates’, logical words to ‘logical connectives’, determiners and attributive adjectives to 'projectors', where the terms 'predicates' and 'logical connectives' are defined below. Functional vector space models extend first order predicate logic as we will see.

Note that first order predicates extend to linear maps of vector spaces thus

PREDICATES

Let $A = \{a_1, \ldots, a_n\}$ be an orthogonal basis of $\mathbb{R}^n = V_A$. A linear map $p : V_A \to S$ is a *predicate on* V_A if $p(a) \in \{\top, \bot\}$ for any $a \in A$.

Examples are the linear maps $\mathtt{true} : V_A \to S$ and $\mathtt{false} : V_A \to S$ satisfying

$$\mathtt{true}(a_i) = \top \text{ respectively } \mathtt{false}(a_i) = \bot \quad \text{for } i = 1, \ldots, n\,.$$

A predicate p on A 'counts' the number of basis vectors for which it takes the value $\top$. Indeed, the subset of $B = \{a_{i_1}, \ldots, a_{i_m}\} \subseteq A$ identifies with the vector $\overrightarrow{B} = \sum_{l=1}^{m} a_{i_l}$. Let n_{pB} be the number of elements of B for which p returns the value $\top$. Then

COUNTING PROPERTY

$$p(\overrightarrow{B}) = n_{pB}\top + (m - n_{pB})\bot \text{ and } n_{\mathtt{true}B} = |B|\,. \tag{2}$$

The counting property gives us a clue how to generalise truth-values to real vector spaces

TRUTH-VALUES

Let p be a linear predicate on V_A and X any vector of V_A. We say that
$p(X)$ is *true* if $p(X)$ is co-linear to $\top$ and $p(X) \neq 0$
$p(X)$ is *false* if $p(X)$ is co-linear to $\bot$ and $p(X) \neq 0$
$p(X)$ is *mixed* if $p(X) = \alpha\top + \beta\bot$ for some $\alpha \neq 0, \beta \neq 0$
$p(X)$ is *mute* if $p(X) = 0$.

The corresponding logic has four truth values, namely 'true', 'false', 'mixed' and 'mute'. Their intuitive meaning will be explained below. Truth-values are invariant under scaling. The vectors X and λX have identical truth-values for $\lambda \neq 0$. Saying 'p is not true on X' only means that $p(X)$ is not co-linear to the basis vector $\top$. This does not imply that 'p is false on X'.

LOGICAL CONNECTIVES

The *logical connectives* are the linear maps $\mathtt{not} : S \to S$, $\mathtt{and} : S \otimes S \to S$, $\mathtt{or} : S \otimes S \to S$ and $\mathtt{ifthen} : S \otimes S \to S$ determined by their values on the basis vectors $z \in \{\top\otimes\top, \top\otimes\bot, \bot\otimes\top, \bot\otimes\bot\}$ thus

$$\begin{aligned}
\mathtt{and}(z) &= \begin{cases} \top & \text{if } z = \top\otimes\top \\ \bot & \text{else} \end{cases} & \mathtt{or}(z) &= \begin{cases} \bot & \text{if } z = \bot\otimes\bot \\ \top & \text{else} \end{cases} \\
\mathtt{ifthen}(z) &= \begin{cases} \bot & \text{if } z = \top\otimes\bot \\ \top & \text{else} \end{cases} & \mathtt{not}(\top) &= \bot \quad \mathtt{not}(\bot) = \top\,.
\end{aligned}$$

The logical connectives induce a Boolean algebra structure on the set of predicates on V_A with the largest element $\mathtt{true}$. Let $d_A : V_A \to V_A \otimes V_A$ be the unique linear map satisfying $d_A(a_i) = a_i \otimes a_i$ for $i = 1, \ldots, n$ and

$$\langle p, q\rangle = (p \otimes q) \circ d_A : V_A \to S \otimes S\,. \tag{3}$$

Then the linear maps

$$\texttt{not} \circ p : V_A \to S,\ \texttt{and} \circ \langle p, q\rangle,\ \texttt{or} \circ \langle p, q\rangle,\ \texttt{ifthen} \circ \langle p, q\rangle : V_A \to S$$

are predicates on V_A.

LOGICAL CONSEQUENCE RELATION

A predicate q is said to be a *logical consequence* of a predicate p if

$$\texttt{ifthen} \circ \langle p, q\rangle = \texttt{true}\,.$$

The logic introduced above extends first order predicate logic. Indeed, assume that a vector $X = \sum_{i=1}^{n} \alpha_i a_i \neq 0$ satisfies $\alpha_i \geq 0$, for $i = 1, \ldots, n$, and let $\overline{X} = \{a_{i_1}, \ldots a_{i_m}\}$ be the subset of basis vectors a_{i_k} for which $\alpha_{i_k} \neq 0$. Then the following holds

FUNDAMENTAL PROPERTY

$$\begin{array}{ll} p \text{ is true on } X & \Leftrightarrow \forall x (x \in \overline{X} \Rightarrow p(x) = \top) \\ p \text{ is false on } X & \Leftrightarrow \forall x (x \in \overline{X} \Rightarrow p(x) = \bot) \\ p \text{ is mixed on } X & \Leftrightarrow \exists x \exists y (x, y \in \overline{X} \ \&\ p(x) = \top \ \&\ p(y) = \bot)\,. \end{array} \tag{4}$$

Words are interpreted by vectors with non-negative coordinates in the functional vector models. Hence, the Fundamental Property applies to all of them.

Example 1.

Consider a game involving chips that come in different shapes and colours. Each shape has a combination of any number of the colours *red*, *yellow* and *blue*. The machine that distributes the chips can recognise colours, but not shapes. Players who want a certain shape therefore must describe the shape, say *triangle*, *square*, *circle*, in terms of colour combinations.

A player who believes in functional models represents thirty chips $A = \{a_1, \ldots, a_{30}\}$ extracted from the machine by a functional model $\mathcal{F}$. His observation is summed up thus

$$\mathcal{F}(\boldsymbol{n}) = N = V_A$$
$$\mathcal{F}(\texttt{triangle})(1) = a_1 + \cdots + a_{10} \quad \mathcal{F}(\texttt{square})(1) = a_{11} + \cdots + a_{20}$$
$$\mathcal{F}(\texttt{circle})(1) = a_{21} + \cdots + a_{30}$$
$$\begin{array}{llll} \mathcal{F}(\texttt{new})(a_i) & = a_i \text{ if } i = 5, 7-15, 20, 25, 30 & \mathcal{F}(\texttt{new})(a_i) & = 0 \text{ else} \\ \mathcal{F}(\texttt{blue})(a_i) & = \top \text{ if } i = 16-20 & \mathcal{F}(\texttt{blue})(a_i) & = \bot \text{ else} \\ \mathcal{F}(\texttt{red})(a_i) & = \top \text{ if } i = 1-9, 11-20 & \mathcal{F}(\texttt{red})(a_i) & = \bot \text{ else} \\ \mathcal{F}(\texttt{yellow})(a_i) & = \top \text{ if } i = 7-15, 21 & \mathcal{F}(\texttt{yellow})(a_i) & = \bot \text{ else}\,. \end{array}$$

Because of his preference for new chips, he computes the noun phrases *new triangles*, *new squares* etc.

$$\begin{array}{ll} \mathcal{F}(\texttt{new} \circ \texttt{square}) & = \mathcal{F}(\texttt{new}) \circ \mathcal{F}(\texttt{square}) = \mathcal{F}(\texttt{new})(a_{11} + \cdots + a_{20}) \\ & = a_{11} + \cdots + a_{15} + a_{20} \\ \mathcal{F}(\texttt{new} \circ \texttt{triangle}) & = \mathcal{F}(\texttt{new}) \circ \mathcal{F}(\texttt{triangle}) = \mathcal{F}(\texttt{new})(a_1 + \cdots + a_{10}) \\ & = a_5 + a_7 + \cdots + a_{10} \\ \mathcal{F}(\texttt{new} \circ \texttt{circle}) & = \mathcal{F}(\texttt{new}) \circ \mathcal{F}(\texttt{circle}) = \mathcal{F}(\texttt{new})(a_{21} + \cdots + a_{30}) \\ & = a_{25} + a_{30} \end{array}$$

Concentrating on *triangles*, he finds that the meaning of the sentence '*No triangles are blue*' computes to $\mathtt{not} \circ \mathtt{blue} \circ \mathtt{triangle}$ in the lexical category, by (1). Hence, the interpretation of the sentence in the functional vector space model $\mathcal{F}$ is

$$\mathcal{F}(\mathtt{not}) \circ \mathcal{F}(\mathtt{blue}) \circ \mathcal{F}(\mathtt{triangle}) = \mathcal{F}(\mathtt{not})(10 \cdot \bot) = 10 \cdot \top = \begin{pmatrix} 10 \\ 0 \end{pmatrix}.$$

The resulting vector is colinear to $\top$, hence the sentence *No triangles are blue* is true in the model. The Fundamental Property implies that $\mathcal{F}(\mathtt{blue})(x) = \bot$ for every basis vector $x \in \mathcal{F}(\mathtt{triangle})$. The predicates $\mathcal{F}(\mathtt{red})$ and $\mathcal{F}(\mathtt{yellow})$ are mixed on $\mathcal{F}(\mathtt{triangle})$

$$\begin{aligned} \mathcal{F}(\mathtt{red} \circ \mathtt{triangle}) &= \mathcal{F}(\mathtt{red})(a_1 + \cdots + a_9) + \mathcal{F}(\mathtt{red})(a_{10}) = 9 \cdot \top + 1 \cdot \bot \\ \mathcal{F}(\mathtt{yellow} \circ \mathtt{triangle}) &= 4 \cdot \top + 6 \cdot \bot \ . \end{aligned}$$

Example 2.
This looks rather confusing, so the player decides to describe the concepts *triangle*, *square*, *circle* by their colours and compute the probability that a chip with colour combination c_j has shape p. Therefore he needs the number k_{pj} of chips of shape p appearing with a given colour combination c_j and the number m_j of chips having colour combination c_j. Using $\mathtt{r}$ if the red colour is present and $\neg\mathtt{r}$ if the red colour is absent and similarly for the other colours, he first arranges the chips according to their colour combinations

combination	chips	number
$c_1 = \ \mathtt{r} \ \neg\mathtt{y} \ \ \mathtt{b}$	$C_1 = \{a_{16}, \ldots, a_{20}\}$	$m_1 = 5$
$c_2 = \ \mathtt{r} \ \ \mathtt{y} \neg\mathtt{b}$	$C_2 = \{a_7, a_8, a_9\} \cup \{a_{11}, \ldots, a_{15}\}$	$m_2 = 8$
$c_3 = \ \mathtt{r} \ \neg\mathtt{y} \neg\mathtt{b}$	$C_3 = \{a_1, \ldots, a_6\}$	$m_3 = 6$
$c_4 = \neg\mathtt{r} \ \ \mathtt{y} \neg\mathtt{b}$	$C_4 = \{a_{10}, a_{21}\}$	$m_4 = 2$
$c_5 = \neg\mathtt{r} \neg\mathtt{y} \neg\mathtt{b}$	$C_5 = \{a_{22}, \ldots, a_{30}\}$	$m_5 = 9$.

	shape	combination
$k_{s1} = 5$	squares	$c_1 = \ \mathtt{r} \neg\mathtt{y} \ \ \mathtt{b}$
$k_{s2} = 5$	squares	$c_2 = \ \mathtt{r} \ \ \mathtt{y} \neg\mathtt{b}$
$k_{c4} = 1$	circles	$c_4 = \neg\mathtt{r} \ \ \mathtt{y} \neg\mathtt{b}$
$k_{c5} = 9$	circles	$c_5 = \neg\mathtt{r} \neg\mathtt{y} \neg\mathtt{b}$

	shape	combination
$k_{t3} = 6$	triangles	$c_3 = \ \mathtt{r} \neg\mathtt{y} \neg\mathtt{b}$
$k_{t2} = 3$	triangles	$c_2 = \ \mathtt{r} \ \ \mathtt{y} \neg\mathtt{b}$
$k_{t4} = 1$	triangles	$c_4 = \neg\mathtt{r} \ \ \mathtt{y} \neg\mathtt{b}$.

Representing each shape p by the vector $p = \sum_j k_{pj}/m_j \cdot c_j$ he obtains the following overview of his findings

$$\begin{aligned} \mathcal{F}(\mathtt{square}) &\mapsto \textit{square} &&= 5/5 \cdot c_1 + 5/8 \cdot c_2 \\ \mathcal{F}(\mathtt{triangle}) &\mapsto \textit{triangle} &&= 3/8 \cdot c_2 + 6/6 \cdot c_3 + 1/2 \cdot c_4 \\ \mathcal{F}(\mathtt{circle}) &\mapsto \textit{circle} &&= 1/2 \cdot c_4 + 9/9 \cdot c_5 \, . \end{aligned}$$

It suffices to ask for a red chip that is not yellow to obtain a triangle.

4 Conceptual vector semantics

Compositional natural language semantics requires reasoning with 'properties'. For the reset of the paper, 'property' stands for the meaning of a noun, an adjective, a verb, a noun phrase or a sentence. Function words contribute essentially to the meanings of grammatical strings, but they are absent in a semantic vector model that only represent concept words. Any compositional method ignoring noise words returns the same result for *No apples are juicy* and *All apples are juicy*. Such a retrieval method only tells us what the sentences are about but not what they mean.

Compositional vector semantics does not require a functor, at least not all the way. One can use a lexical functor which maps the concept words to vectors or endomorphisms of some space V_C and each time the functor asks for the value of a noise word use some other method, for example an operator defined on V_C. The method must capture the logical content, so we need a notion of truth and of logical consequence in V_C. Otherwise, we could choose any operation on the vector space, like vector sum, and say that the sum of the vectors in the string is the 'meaning' of the whole string. The algebraic connectives and the associated consequence relation given below generalise the corresponding logical connectives for sets and predicates and the quantum connectives on projectors. They also allow reasoning with probability.

The following definitions refer to an orthogonal basis $A = \{a_1, \ldots, a_n\}$ of $\mathbb{R}^n$. A vector is *Boolean* if its coordinates with respect to A are equal to 0 or 1. Let $\mathcal{B}_A$ denote the set of Boolean vectors and $\ulcorner f \urcorner_{AA}$ the matrix that an endomorphism f of $\mathbb{R}^n$ defines with respect to A. The endomorphisms of $\mathbb{R}^n$ that are diagonalisable with respect to A and the projectors among them form the sets

$$\begin{aligned} \mathcal{D}_A &= \{f : \mathbb{R}^n \longrightarrow \mathbb{R}^n : \ulcorner f \urcorner_{AA} \text{ is a diagonal matrix}\} \\ \mathcal{P}_A &= \{f \in \mathcal{D}_A : f \circ f = f\}\,. \end{aligned}$$

There is an obvious one-to-one correspondence between vectors of $\mathbb{R}^n$ and $\mathcal{D}_A$ via the correspondence

$$X = \sum_{i=1}^{n} \alpha_i a_i \leftrightarrows D_X = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix}.$$

Clearly, the map $\mathcal{I}$ that assigns to $f \in \mathcal{D}_A$ the vector X for which $\ulcorner f \urcorner_{AA} = D_X$ is a bijection from $\mathcal{P}_A$ onto $\mathcal{B}_A$.

ALGEBRAIC CONNECTIVES

The *algebraic connectives* are defined for scalars and for arbitrary endomorphisms thus

$$\begin{array}{llll} \text{negation} & \neg\alpha = 1 - \alpha & \neg D = 1 - D \\ \text{conjunction} & \alpha \wedge \beta = \alpha\beta & D \wedge E = DE \\ \text{disjunction} & \alpha \vee \beta = \alpha + \beta - \alpha\beta & D \vee E = D + E - DE \\ \text{implication} & \alpha \to \beta = 1 - \alpha + \alpha\beta & D \to E = 1 - D + DE & . \end{array}$$

The algebraic connectives are lifted to vectors by

$$\begin{array}{lll} \neg X & = \neg(\sum_{i=1}^{n} \alpha_i a_i) & = \sum_{i=1}^{n} (\neg \alpha_i) a_i \\ X \triangledown Y & = (\sum_{i=1}^{n} \alpha_i a_i) \triangledown (\sum_{i=1}^{n} \beta_i a_i) & = \sum_{i=1}^{n} (\alpha_i \triangledown \beta_i) a_i \,, \end{array}$$

where $\triangledown$ stands for any of the binary algebraic connectives. It follows that

$$\neg D_X = D_{\neg X} \quad \text{and} \quad D_X \triangledown D_Y = D_{X \triangledown Y} \tag{5}$$

The two equalities above say that the one-to-one correspondence $\mathcal{I}$ is an isomorphism of algebraic connectives. Note that the algebraic conjunction coincides on vectors with the pointwise product $\odot$ introduced by (Kartsaklis et al. 2013) and composition is related to the algebraic conjunction thus

$$D_X \circ D_Y = D_X \wedge D_Y = D_{X \wedge Y} \qquad D_X \circ \langle Y| = \langle X \wedge Y| \,, \tag{6}$$

where $\langle Z|$ denotes the matrix of the linear map $v : I \to V_A$ that assigns the vector $Z \in V_A$ to the unique basis vector of I. Reformulating the last equalities with the help of $\mathcal{I}$, we get for any $f, g \in \mathcal{D}_A,\ v : I \to V_A$

$$\mathcal{I}(f \circ g) = \mathcal{I}(f) \wedge \mathcal{I}(g) = \mathcal{I}(f) \odot \mathcal{I}(g) \qquad \mathcal{I}(f \circ v) = \mathcal{I}(f) \wedge v = \mathcal{I}(f) \odot v \,.$$

The vectors considered in semantic vector models have positive components. In general, they are replaced by vectors with components in the real interval $[0, 1]$. Think of the latter vectors as *concept vectors*. Concept vectors are closed under the algebraic connectives, because the interval $[0, 1]$ is closed under the algebraic connectives. With a lexicon that assigns a meaning expression to every noise word as well, the meaning of a grammatical string can be computed in the lexical category, resulting in a composition of distinguished morphisms $\mathtt{w}_1 \circ \cdots \circ \mathtt{w}_n$. If the lexical functor $\mathcal{F}$ of Section 2 maps the distinguished morphisms $\mathtt{w}_i$ to self-adjoint endomorphisms then $\mathcal{I}(\mathcal{F}(\mathtt{w}_1) \circ \cdots \circ \mathcal{F}(\mathtt{c}_2)) = \mathcal{I}(\mathcal{F}(\mathtt{c}_1)) \odot \cdots \odot \mathcal{I}(\mathcal{F}(\mathtt{c}_2))$.

The algebraic connectives do not define a lattice structure, for example the algebraic conjunction is not idempotent unless the involved scalars are 0 or 1. They have however several properties with a logical flavour, among them the laws of a weak conditional logic in the sense of (Rijsbergen 2004).

Any real numbers $\alpha, \beta \in [0, 1]$ and diagonal matrices $D, E \in \mathcal{D}_A$ with entries in $[0, 1]$ satisfy

$$\alpha \to \beta = 1 \text{ if and only if } \alpha = 0 \text{ or } \beta = 1 \tag{7}$$

$$\begin{gathered} \alpha \to \beta = 1 \text{ if and only if } \alpha\beta = \alpha \\ \text{If } \alpha \to \beta = 1 \text{ then } \alpha \leq \beta \\ D \to E = 1 \text{ if and only if } E \circ D = D \\ \text{If } D \to E = 1 \text{ then } D \leq E \,. \end{gathered}$$

ALGEBRAIC CONSEQUENCE RELATION

The endomorphism defined by E is an *algebraic consequence* of that defined by D if and only if

$$D \to E = 1 \,.$$

The last statement in (7) means that the algebraic consequence relation implies the probabilistic consequence relation, because by definition E is a *probabilistic consequence* of D if and only if $D \leq E$.

(Rijsbergen 2004) uses projectors to represent properties and the geometrical connectives of quantum logic. They are introduced as geometrical operators via the range of the involved linear maps based on the fact that for every subspace there is a unique projector which maps the whole space onto the subspace.

GEOMETRICAL CONNECTIVES

Let $p, q : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be projectors. Then

the *geometrical negation* $\neg p$ is the unique projector that has range

$$(\neg p)(\mathbb{R}^n) = p(\mathbb{R}^n)^\perp$$

the *geometrical conjunction* $p \wedge q$ is the unique projector that has range

$$(p \wedge q)(\mathbb{R}^n) = p(\mathbb{R}^n) \cap q(\mathbb{R}^n)$$

the *geometrical disjunction* $p \vee q$ is the unique projector that has range

$$(p \vee q)(\mathbb{R}^n) = p(\mathbb{R}^n) + q(\mathbb{R}^n)$$

the *geometrical implication* $p \Rightarrow q$ is the unique projector that has range

$$(p \Rightarrow q)(\mathbb{R}^n) = \{x \in \mathbb{R}^n \,:\, (q \circ p)(x) = p(x)\}\,.$$

QUANTUM CONSEQUENCE RELATION

Projector q is said to be a *geometrical consequence* of projector p if and only if

$$p \Rightarrow q = 1\,.$$

The definition makes the detour via the subspaces, because there is no obvious algebraic operator defining the projector. For example, $p \circ q$ maps $\mathbb{R}^n$ onto the intersection of the image of p and the image of q, but $p \circ q$ is not a projector unless p and q commute. If p and q do not commute, they are not both diagonalisable in any single basis. Are we not loosing representatives of properties in probability when replacing the projectors by $\mathcal{D}_A$? The answers is that to the contrary, we are gaining representatives at least as long as we accept the geometric consequence relation.

Proposition 1. *If projector q is a geometrical consequence of projector p then there is an orthogonal basis of $\mathbb{R}^n$ consisting of eigenvectors of both p and q.*

The geometrical consequence relation and the algebraic consequence relation coincide on projectors. If one of p or q is a geometrical consequence of the other then the geometrical connectives coincide with the algebraic connectives for p and q.

Proof. (Outline) Clearly, the second statement follows from the first. To see the first statement, assume that q is a geometrical consequence of p and let A be an orthonormal basis of $\mathbb{R}^n$ formed by eigenvectors of p. The eigenvectors in A left invariant by p are also left invariant by q. Let A_1 denote this set and V be the subspace generated by A_1. Then q maps the orthogonal complement of V onto itself whereas p maps it to 0. Hence any set C of orthonormal eigenvectors of q belonging to V are also eigenvectors of p. Thus $A_1 \cup C$ is a basis of orthonormal eigenvectors for both p and q. □

The proposition above also implies that the algebraic connectives can be captured by geometrical properties, at least on Boolean vectors. In particular, two distinct basis vectors contradict each other. This raises the question how to choose the basis vectors so that they represent contradictory properties.

5 Distributional interpretations

Everyday language switches commonly from asserting facts about sone real or imagined world and updating the concepts intervening in the statements about the facts. I model this switch using the canonical distribution based on the counting property of predicates. The modelling interprets the coefficient of a concept, say *apple*, for a basis vector, say *juicy*, as the conditional probability of the event *apple* given the event *juicy*. I give sufficient conditions for concept logic to be reflected and predicate logic to be preserved.

Predicates $q_1, \ldots, q_k$ on V_A are said to be *mutually contradictory* if from $j \neq l$ and $a \in A$ follows that $q_j(a) = \bot$ or $q_l(a) = \bot$. They are *pertinent* if for any $a \in A$ there is a predicate q_j for which $q_j(a) = \top$.

Choose some orthogonal basis $A = \{a_1, \ldots, a_n\}$ of $\mathbb{R}^n$ and $C = \{c_1, \ldots, c_k\}$ of $\mathbb{R}^k$. Think of the basis vectors $a_1, \ldots, a_n$ as 'individuals' and the basis vectors $c_1, \ldots, c_k$ as 'basic events on A' or as 'basic properties' of the individuals. That is to say, identify $\{c_1, \ldots, c_k\}$ with a partition $\{C_1, \ldots, C_k\}$ of pairwise disjoint subsets covering A or, equivalently, with a family of mutually contradictory pertinent predicates $q_1, \ldots, q_k$ on V_A. The predicates $q_1, \ldots, q_k$ are related to the sets $C_1, \ldots, C_k$ by the equalities

$$C_j = \{a \in A \,:\, q_j(a) = \top\},\ j = 1, \ldots, k\,.$$

Let $m_j = |C_j|$ so that $\sum_j m_j = n = |A|$. Assuming that every individual in A has probability $1/n$, the real number

$$\mu_j = m_j/n$$

can be understood as the probability that an arbitrary individual $a \in A$ has property q_j, for $j = 1, \ldots, k$. The density operator defined by the diagonal matrix

$$D_\mu = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_k \end{pmatrix}$$

summarises the first order model consisting of A and the predicates q_j, for $j = 1, \ldots, k$. Composing the density operator D_μ with the projector p_j onto the line spanned by basis vector c_j we obtain the vector $\mu_j c_j$. The component μ_j is the probability that an arbitrary individual of the model has property q_j, with other words that the model is in 'state' q_j.

Recall that for any predicate p on V_A, the integer n_{pC_j} denotes the number of elements in C_j for which p has value $\top$. Define

$$\mathcal{J}_C(p) = \sum_{j=1}^{k} \alpha_{pj} c_j, \text{ where } \alpha_{pj} = \begin{cases} n_{pC_j}/m_j & \text{if } m_j \neq 0 \\ 0 & \text{else} \end{cases}, \text{ for } j = 1, \ldots, k\,.$$

The number α_{pj} is the conditional probability that an individual has property p provided it has property q_j. It follows from the linearity of p that $n_{pA} = \sum_j n_{pC_j}$. Therefore, the probability that an arbitrary element of A has property p is equal to

$$n_{pA}/n = \sum_j n_{pC_j}/n = \sum_j (n_{pC_j}/m_j)(m_j/n) = \mathit{trace}(D_\mu \circ D_{\mathcal{J}(p)}) = \sum_j \alpha_{pj}\mu_j\,.$$

The interpretation $\mathcal{J}_C$ is an isomorphism if C consists of the singleton sets $\{a_1\}, \ldots, \{a_n\}$, because the sufficient conditions of Lemma 1 and Theorem 1 are satisfied. In general, however, it is not even one-to-one. Indeed, $\mathcal{J}_C(p)$ only tells us for how many individuals p is true in each set C_j and not for which individuals. Moreover, $\mathcal{J}_C$ does not commute with the connectives in general. The following sufficient conditions are of practical relevance.

Lemma 1. *Let $p, q : V_A \to S$ be any predicates on A and assume that the sets C_j are not empty, for $j = 1, \ldots, k$. Then the following holds*

$$\mathcal{J}_C(\mathtt{not} \circ p) = \neg \mathcal{J}_C(p)$$

If in addition for every j at least one of p or q is constant on C_j then

$$\begin{aligned} \mathcal{J}_C(\mathtt{and} \circ \langle p, q \rangle) &= \mathcal{J}_C(p) \wedge \mathcal{J}_C(q) \\ \mathcal{J}_C(\mathtt{or} \circ \langle p, q \rangle) &= \mathcal{J}_C(p) \vee \mathcal{J}_C(q) \\ \mathcal{J}_C(\mathtt{ifthen} \circ \langle p, q \rangle) &= \mathcal{J}_C(p) \to \mathcal{J}_C(q)\,. \end{aligned}$$

Negation is preserved exactly when D_μ is positive definite. This condition alone is not sufficient for $\mathcal{J}_C$ to preserve the binary connectives. The supplementary condition needed, however, corresponds to the way we explain new concepts in terms of more basic concepts.

Theorem 1. *Suppose that the sets $C_1, \ldots, C_k$ partition A and are not empty. Then $\mathcal{J}_C$ reflects the consequence relation, i.e. for any predicates p and q on A*

$$\mathcal{J}_C(p) \to \mathcal{J}_C(q) = 1 \textit{ implies } \mathtt{ifthen} \circ \langle p, q \rangle = \mathtt{true} \tag{8}$$

Moreover, if one of $\mathcal{J}_C(q)$ and $\mathcal{J}_C(p)$ is an algebraic/geometrical consequence of the other then $\mathcal{J}$ preserves the logical connectives. The algebraic/geometrical *connectives preserve the probabilistic interpretation of concept vectors.*

$$\begin{aligned}
\neg\mathcal{J}_C(p) &= \mathcal{J}_C(\mathtt{not}\circ p)\\
\neg\mathcal{J}_C(q) &= \mathcal{J}_C(\mathtt{not}\circ q)\\
\mathcal{J}_C(p)\wedge\mathcal{J}_C(q) &= \mathcal{J}_C(\mathtt{and}\circ\langle p,q\rangle)\\
\mathcal{J}_C(p)\vee\mathcal{J}_C(q) &= \mathcal{J}_C(\mathtt{or}\circ\langle p,q\rangle)\\
\mathcal{J}_C(p)\to\mathcal{J}_C(q) &= \mathcal{J}_C(\mathtt{ifthen}\circ\langle p,q\rangle)\\
\mathcal{J}_C(q)\to\mathcal{J}_C(p) &= \mathcal{J}_C(\mathtt{ifthen}\circ\langle q,p\rangle)\,.
\end{aligned} \tag{9}$$

Proof. Assume that $\mathcal{J}_C(p) \to \mathcal{J}_C(q) = 1\,.$ Then $1 - \alpha_{pj} + \alpha_{pi}\alpha_{qj} = 1$ and therefore $\alpha_{pj} = 0$ or $\alpha_{qj} = 1$, for $j = 1, \dots, k$. With other words, p maps every element of C_j to $\perp$ or q maps every element of C_j to $\top$. As every element of A belongs to some C_j, we have

$$\mathtt{ifthen}\circ\langle p,q\rangle(x) = \top,\ \text{for all } x \in A\,.$$

The equality $\mathtt{ifthen}\circ\langle p,q\rangle = \mathtt{true}$ follows. This completes the proof of (8).

The equalities (9) hold, because the assumptions of the preceding lemma are satisfied. □

If the covering partition $\{\{a_1\}, \dots, \{a_n\}\}$ consists of the singleton sets formed by the basis vectors $a_1, \dots, a_n$ of V_A the interpretation $\mathcal{J}_A$ is an isomorphism of the Boolean algebra of predicates defined on V_A onto the Boolean algebra of Boolean vectors $\mathcal{B}_A$ of V_A. This is the degenerate case where the 'individuals', i.e. the basis vectors of V_A are identified with the 'basic concepts', again the basis vectors of V_A. The functional model interprets noun phrases by vectors in $\mathcal{B}_A$ adjectives in attributive position by projectors in $\mathcal{P}_A$ and verbs and adjectives in predicative position by predicates. Boolean vectors in $\mathcal{B}_A$ identify with projectors in $\mathcal{P}_A$ by (5) and with predicates by $\mathcal{J}_A$. Therefore $\mathcal{J}_C$ can be extended to all non-noise verbs such that Theorem 1 remains valid.

6 Practice and Conclusion

The method given in Example 2 generalises to an arbitrary number of words. Indeed, let $P = \{w_1, \dots, w_d\}$ be a set of concept words in the lexicon. Invent a two-dimensional space $S_i = V_{\{\mathtt{w}_i, \neg\mathtt{w}_i\}}$ with basis vectors $\mathtt{w}_i$ and $\neg\mathtt{w}_i$, for $i = 1, \dots, d$ and define the *concept space generated by* P as

$$C(P) = S_1 \otimes \ldots \otimes S_d\,.$$

The basis vectors of $C(P)$ are of the form

$$c_j = c_j(1) \otimes \ldots \otimes c_j(d)$$

where $c_j(i) \in \{\mathtt{w}_i, \neg\mathtt{w}_i\}$, for $j = 1, \dots, k$.

Without loss of generality, we may assume that the functional model interprets the words $w_1, \ldots, w_d$ as predicates $p_1, \ldots, p_d$ on a space V_A. The basis vectors correspond to the following partition $C_1, \ldots, C_{2^d}$ of subsets of A

$$a \in C_j \text{ if and only if } p_i(a) = \begin{cases} \top & \text{if } c_j(i) = \mathtt{w}_i \\ \bot & \text{if } \neg\mathtt{w}_i \end{cases}, \text{ for } i = 1, \ldots, d\,.$$

If we work within the subspace C of $C(P)$ generated by the basis vectors c_j for which $C_j \neq \emptyset$, the interpretation $\mathcal{J}_C$ reflects concept logic. It preserves predicate logic under the conditions of Theorem 1. The concept words w_i 'live' in $C(P)$ under the form

$$\overrightarrow{w_i} = \sum_{\{j\,:\,c_j(i)=\mathtt{w}_i\}} c_j\,,$$

and the algebraic operators induce on the set of Boolean vectors of $C(P)$ the structure of a Boolean algebra which is isomorphic to the free Boolean algebra generated by the set $\{w_1, \ldots, w_d\}$, (Preller 2012).

Example 3.
A player who has not seen any chips, but has access to text mentioning the colours of the various shapes still can use the the conceptual representation.

$$\mathit{triangle} = \alpha_1 c_1 + \cdots + \alpha_8 c_8,\ 0 \leq \alpha_i \leq 1$$

The truth of the statement *No triangle is blue* implies that $\alpha_i = 0$ for all colour combinations c_i involving `b` (blue) without the negation symbol. Hence

$$\mathit{triangle} = \alpha_2 c_2 + \alpha_3 c_3 + \alpha_4 c_4 + \alpha_5 c_5\,.$$

Note that the vector *triangle* is orthogonal to every basis vector that lists the colour blue as present, namely c_1, c_6, c_7 and c_8. Therefore *triangle* is orthogonal to the subspace generated by the four basis vectors with 'blue' present.

References

Asher, N. (2011), *Lexical Meaning in Context*, Cambridge University Press.

Clark, S., Coecke, B. & Sadrzadeh, M. (2008), A compositional distributional model of meaning, *in* W. L. P. Bruza & J. van Rijsbergen, eds, 'Proceedings of Conference on Quantum Interactions', University of Oxford, College Publications.

Kartsaklis, D., Sadrzadeh, M., Pulman, S. & Coecke, B. (2013), *Reasoning about Meaning in Natural Language with Compact Closed Categories and Frobenius Algebras*, Cambridge University Press.

Kracht, M. (2007), 'Compositionality: The very idea', *Research in Language and Computation* **5**, 287–308.

Lambek, J. (1999), Type grammar revisited, *in* A. Lecomte, ed., 'Logical Aspects of Computational Linguistics', Vol. 1582 of *LNAI*, Springer, Heidelberg, pp. 1–27.

Lambek, J. (2008), *From word to sentence*, Polimetrica, Milano, Italia.

Preller, A. (2005), Category theoretical semantics for pregroup grammars, *in* P. Blache & E. Stabler, eds, 'Logical Aspects of Computational Linguistics', Vol. 3492 of *Lecture Notes in Artificial Intelligence*, pp. 254–270.

Preller, A. (2012), *From Sentence to Concept: Predicate Logic and Quantum Logic in Compact Closed Categories*, Oxford University Press, pp. 00–29.

Preller, A. & Sadrzadeh, M. (2011), 'Semantic vector models and functional models for pregroup grammars', *Journal of Logic, Language and Information* **20**(4), 419–423.

Rijsbergen, C. v. (2004), *The Geometry of Information Retrieval*, Cambridge University Press.

Widdows, D. (2004), *Geometry and Meaning*, number 172 *in* 'CSLI lecture notes', CSLI Publications.