Theoretical Computer Science 752 (2018) 104–115

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Generalized relations in linguistics & cognition

Bob Coecke [1], Fabrizio Genovese [*,1], Martha Lewis [1], Dan Marsden [1], Alex Toumi [1]

Quantum Group, Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford, UK

A R T I C L E   I N F O

Article history:
Received 9 June 2017
Received in revised form 1 December 2017
Accepted 5 March 2018
Available online 7 March 2018

Keywords:
Categorical models of meaning
Generalized relations
Generalized spans

A B S T R A C T

Categorical compositional models of natural language exploit grammatical structure to calculate the meaning of phrases and sentences from the meanings of individual words. More recently, similar compositional techniques have been applied to conceptual space models of cognition.

Compact closed categories, particularly the category of finite dimensional vector spaces, have been the most common setting for categorical compositional models. When addressing a new problem domain, such as conceptual space models of meaning, a key problem is finding a compact closed category that captures the features of interest.

We propose categories of generalized relations as a source of new, practical models for cognition and NLP. We demonstrate using detailed examples that phenomena such as fuzziness, metrics, convexity, semantic ambiguity can all be described by relational models. Crucially, by exploiting a technical framework described in previous work of the authors, we also show how the above-mentioned phenomena can be combined into a single model, providing a flexible family of new categories for categorical compositional modelling.

1. Introduction

Distributional models of language describe the meaning of a word using co-occurrence statistics derived from corpus data. A central question with these models is how to combine meanings of individual words, in order to understand phrases and sentences. Categorical compositional models of natural language [1] address this problem, providing a principled approach to combining the meanings of words to form the meanings of sentences, by exploiting their grammatical structure. At the time, they went on to outperform conventional techniques for some standard NLP tasks [2,3]. Since then the focus of the research program has mostly turned its attention to:

- Identifying the appropriate compositional structure underpinning functional words such as relative pronouns and logical connectives,
- designing more interesting models of meaning, and
- employing these structures and models beyond the initial linguistic scope.

The latter two are what this paper is concerned with.

* Corresponding author.
E-mail address: fabrizio.genovese@cs.ox.ac.uk (F. Genovese).

[1] Authors in alphabetical order.

https://doi.org/10.1016/j.tcs.2018.03.008
0304-3975/

As explained in the expository paper [4], the main inspiration for these models was the category-theoretic description of quantum processes, initially, in particular, quantum-teleportation like protocols [5]. Subsequently, several other structures from categorical quantum mechanics have also found applications in natural language. For example, the representation of classical data [6] has been used to implement relative pronouns [7] and intonation [8], and the representation of open quantum systems [9] has been used to model ambiguity [10].

Distributional models of language can be thought of as "process theories" [11]. A process theory consists of a graphical language for reasoning about composite systems of abstract processes, and a categorical semantics modelling the application domain. A particularly important class of categorical models are the compact closed categories, which come equipped with an elegant graphical calculus. Besides quantum theory, process theoretic models built upon compact closed categories have been successfully exploited in many application areas including signal flow graphs [12], control theory [13], Markov processes [14], electrical circuits [15] and even basic linear algebra [16]. However, it should be stressed that for the compositional modelling of language, while convenient, compact closure is not vital and other categories could be taken as a starting point. In grammatical terms this means passing from Lambek's pregroups to his earlier proposals for grammatical structure [17].

Recently [18], the categorical compositional approach to meaning has been applied to the conceptual space models of human cognition introduced in [19,20]. With it comes another interesting connection with research in quantum theory, namely the use of general convex spaces, which have prominent in quantum theory research since Ludwig's work [21], and which have recently regained attention under the name of Generalised Probabilistic Theories [22].

More generally, when addressing a new application domain, it is necessary to identify a compact closed category with mathematical structure compatible with the application phenomena of interest. Amongst the compact closed categories the hypergraph categories [23] are a particularly well behaved class of practical interest. In [24] we presented a flexible parameterized mathematical framework for constructing hypergraph categories. We view this framework as a practical tool for building new models in a principled manner, by varying the parameter choices according to the needs of the application domain. These models are based upon generalizing the well understood notion of a binary relation, providing a concrete and intuitive setting for model development.

The present work is a significant extension of the ideas presented in the workshop paper [25]. We will demonstrate, via extensive examples, that categories of generalized relations present an attractive setting for constructing new models of language and cognition. We emphasize the intuitive interpretation of the models under construction, and their connections to concrete ideas in computation, NLP and further afield. These examples are structured as follows:

- In section 3 we introduce relations with generalized truth values, and exploit them to model features such as distances, forces, connectivity and fuzziness. Relations with generalized truth values are well known in the mathematical community, but seem to have received little attention from the perspective of compositional semantics, with the recent exception of [26]. In particular, we will use generalized relations to give a semantic interpretation of family trees, expressing degrees of kinship such as "parent" or "sibling" in a compositional way.
- In section 4 we generalize relations in another direction, considering relations that respect algebraic structure. These relations can capture features such as convexity, which is important in conceptual spaces models [19,20]. In this case, we recover a model first used in [18], originally constructed in an ad-hoc manner using techniques from monad theory and the theory of regular categories. Importantly, we then show that we can combine generalized truth values with relations respecting algebraic structure, providing conceptual space models with access to distance measures.
- In section 5 we view spans as generalized "proof aware" relations in which the apex of the span contains witnesses to relatedness between the domain and codomain. Spans can be extended to support generalized truth values, and to respect algebraic structure. Exploiting a combination of these features, we construct a new model of semantic ambiguity in conceptual space models of natural language, in which different proof witnesses allow us to vary how strongly different words are related, depending on how they are interpreted. We will moreover recast our family trees example in a proof-aware fashion.
- The previous examples were essentially built upon the category of sets. Our techniques can be applied with different choices of ambient topos. In section 6, we will briefly mention how the natural choice of truth values in toposes different from the category of sets is related with the subobject classifier of the topos, seen as an internal locale.

All of our models are preorder enriched, providing a natural candidate for modelling semantic entailment or hyponymy [27,28]. Preorder enrichment also means we can consider internal monads within our various categories of relations. We emphasize the importance of these internal monads throughout our discussions. They provide access to important structured objects such as preorders, generalized metric spaces and ultrametric spaces, and similar well behaved relationships when we combine various modelling features.

2. Compositional models of meaning

The grammatical structure of natural language can be modelled using Lambek's pregroup grammars [29]. Multiple other choices, as context-free grammars, are available to accomplish this task, but the categorical properties of pregroups such as compact closure, see Definition 2, make them very appealing from our perspective. This choice is quite a common one in the categorical approach to natural language.

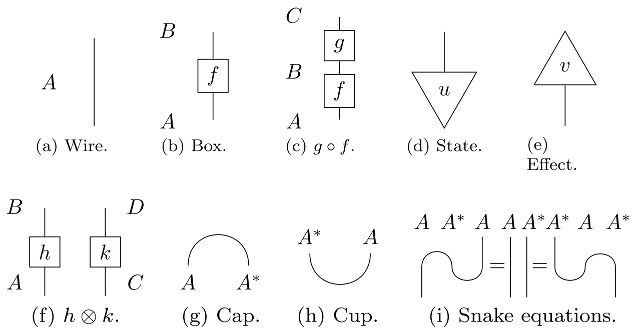


Fig. 1. Graphical calculus for compact closed categories.

Definition 1. A **pregroup** is a tuple $(X, \cdot, 1, (-)^l, (-)^r, \leq)$ where $(X, \cdot, 1, \leq)$ is a partially ordered monoid, or pomonoid, and $(-)^r, (-)^l$ are unary functions of type $X \to X$ such that for all $x \in X$ the following conditions hold,

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

We say that x **reduces** to x' if $x \leq x'$.

A grammar is typically described using the free pregroup over some set of basic types. For example, we may consider the free pregroup of the set $\{n, s\}$ where n and s are basic types for nouns and sentences respectively. More complex terms are then built up using the algebraic operations, for example the type of a transitive verb is $n^r s n^l$. We can calculate the type of a phrase by composing the types of the individual terms using the monoid multiplication. For example, the phrase "mice eat cheese" has type $n(n^r s n^l)n$, where "mice" and "cheese" have type n (noun), and "eat" has type $(n^r s n^l)$ (verb). A composite term is a well typed sentence if its type reduces to the sentence type. For example:

$$n(n^r s n^l)n = (nn^r)s(n^l n) \leq s(n^l n) \leq s$$

and so "mice eat cheese" is a well typed sentence. In this way, pregroups give us access to the *compositional* features of language.

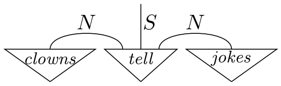
On the other hand, distributional models [30] of the *meaning* of words in natural language are built using vector space models automatically derived from co-occurrence statistics in a large corpus of text. The key observation of the categorical compositional approach to natural language is that both pregroups and the category of finite dimensional real vector spaces carry the same categorical structure, that of an autonomous category.

Definition 2. A monoidal category $\mathcal{V}$ has left/right duals if every object has an internal left/right adjoint when $\mathcal{V}$ is regarded as a one object bicategory. An **autonomous category** is a monoidal category in which every object has both left and right duals. A **compact closed category** is a symmetric monoidal category in which every object has right duals.

This observation can be exploited to derive the meanings of sentences from the meanings of words. We fix a strong monoidal functor from a pregroup describing grammatical structure to the category of finite dimensional vector spaces. This functor maps type reductions to linear maps, allowing us to automatically derive the meaning of a sentence from its constituent parts. Clearly, this approach can be seen as an instance of functorial semantics. By varying the domain and preserved structure we can consider different categorial grammars [17]. By varying the codomain we can consider different models, as has been important in recent work broadening the scope to mathematical models of cognition [18,31]. When varying the category of meanings, it is desirable to remain within the domain of compact closed categories, in order to exploit connections with previous linguistic developments, and to retain access to the powerful graphical calculus that we are now going to recall. A straightforward application oriented introduction to monoidal categories and compact closed categories can be found in [32].

In the graphical calculus for compact closed categories, an object A is denoted by a wire, as shown in Fig. 1a, while a morphism $f : A \to B$ is represented by a box as depicted in Fig. 1b. If $g : B \to C$, the composite $g \circ f : A \to C$ is formed by vertical composition, as in Fig. 1c. By convention, the monoidal unit I is drawn as the empty diagram. Morphisms of type $u : I \to A$ and $v : A \to I$ are referred to as **states** and **effects** of A, and are drawn using the special notation of Figs. 1d and 1e. If $h : A \to B$ and $k : C \to D$, their tensor product $h \otimes k : A \otimes C \to B \otimes D$ is formed by horizontal juxtaposition as in Fig. 1f. The existence of a right dual A^* for every object A means that for every object there exist morphisms $\epsilon_A : A \otimes A^* \to I$ and $\eta_A : I \to A^* \otimes A$, referred to as a **cap** and **cup** and displayed graphically as in Figs. 1g and 1h. They satisfy the equations $(\epsilon \otimes 1_A) \circ (1_A \otimes \eta) = 1_A$ and $1_{A^*} = (1_{A^*} \otimes \epsilon) \circ (\eta \otimes 1_{A^*})$. These conditions are suggestively referred to as the **snake equations** given their graphical formulation in Fig. 1i.

The functor from grammar to semantics gives us our "wiring", that allows us to calculate meanings graphically as follows: our meaning category supplies the qualitative meanings of words, like *clowns*, *tell*, and *jokes*. Our grammar category tells us how to stitch these together. This corresponds to "telling us where to put cups and caps." The essence of the method should be thought of as the diagram



where we think of the words as states in our semantics and of the wires as the image through our functor of the pregroup reduction witnessing that they form a well-typed sentence.

The question then becomes: *How can we find or construct compact closed categories with desirable mathematical properties?* This is what we explore in this paper: our constructions produce a subclass of compact closed categories, referred to as hypergraph categories [33,23], and so this is where we shall focus our attention.

Definition 3. A **hypergraph category** is a symmetric monoidal category in which every object is equipped with a choice of special commutative Frobenius algebra, coherently with the monoidal structure.

Details of the notion of a Frobenius algebra, and linguistic applications including modelling relative pronouns can be found in [7,34]. Morphisms of type $I \to I$ are referred to as **numbers**.

Example 1. The category **Rel** of sets and binary relations between them can be given the structure of a hypergraph category. The monoidal structure is given by forming Cartesian products of sets. A state of a set X is a subset of X and the numbers are the Boolean truth values. The Frobenius algebra is given by the copying relation $x \sim (x, x) : X \to X \times X$, the deletion relation $x \sim * : X \to I$, and their converses.

All the compact closed categories discussed in this paper will be hypergraph categories, generalizing Example 1 along different axes of variation.

3. Generalized truth values

A binary relation $R : A \to B$ between sets can be identified with a characteristic function of type $A \times B \to \{\top, \bot\}$ mapping the related pairs of elements to $\top$. It is fruitful to consider generalizing the codomain of such characteristic functions to a set Q, thought of as a collection of truth values. We can then consider functions of the form $A \times B \to Q$ as generalized relations, with truth values in Q. In order for the corresponding binary relations to have satisfactory notions of identities and composition, the set Q must carry the structure of a quantale.

Definition 4 *(Quantale).* A **quantale** is a join complete partial order Q with a monoid structure $(\otimes, k)$ satisfying the following distributivity axioms, for all $a, b \in Q$ and $A, B \subseteq Q$:

$$a \otimes \left[\bigvee B\right] = \bigvee \{a \otimes b \mid b \in B\} \qquad \left[\bigvee A\right] \otimes b = \bigvee \{a \otimes b \mid a \in A\}$$

A quantale is said to be **commutative** if its monoid structure is commutative.

All the quantales encountered in this paper will be commutative. We introduce some examples of importance in later developments.

Example 2. The **Boolean quantale** is given by the two element complete Boolean algebra $\mathbf{B} = \{\top, \bot\}$, with the join and multiplication given by the join and meet in the Boolean algebra.

Example 3. The **Lawvere quantale L** is given by the chain $[0, \infty]$ of extended positive reals with the *reverse* ordering, hence minima in $[0, \infty]$ provide the joins of the quantale, and the monoid structure is given by addition.

Example 4. The quantale **F** has again the extended positive reals with reverse order as its partial order, but now with max as the monoid multiplication.

Example 5. The **interval quantale I** is given by the ordered interval $[0, 1]$ with minima as the monoid structure.

For a quantale Q, the Q-relations form a category $\mathbf{Rel}(Q)$ with composition and identities[2]

$$(S \circ R)(a,c) = \bigvee_b R(a,b) \otimes S(b,c) \qquad 1_A(a,b) = \bigvee\{k|a=b\}$$

If Q is a commutative quantale, $\mathbf{Rel}(Q)$ carries a symmetric monoidal structure, with the tensor product of objects given by the Cartesian product of sets, and the action on relations given for $R : A \to C$ and $S : B \to D$ by

$$(R \otimes S)(a,b,c,d) = R(a,c) \otimes S(b,d)$$

The singleton set is the monoidal unit. A key observation from the perspective of this paper is:

Theorem 1. $\mathbf{Rel}(Q)$ *is compact closed with respect to this monoidal structure.*

Now that we have described how Q-relations compose, we can consider computational interpretations for our example choices of quantale.

Example 6. The relations over the Lawvere quantale $\mathbf{L}$ can be thought of as describing costs. The value $R(a,b)$ describes the cost of converting a into b. A cost of 0 means they are maximally related and can be freely inter-converted. A cost of ∞ indicates completely unrelated values, that cannot be converted between each other for finite cost. The value $(S \circ R)(a,c)$ describes the cheapest way of converting a into some b, and then converting that b into c, and adds the associated costs. If we perform two conversions in parallel $(R \otimes R')(a,a',b,b')$ describes the sum of the two individual conversion costs.

In this setting, we can think of a state $I \to A$ as giving a table of costs for acquiring the resources in A, and similarly an effect $A \to I$ is a table of costs for disposing of resources in A.

Example 7. The quantale $\mathbf{F}$ has the same underlying set as the Lawvere quantale, but its different algebraic structure leads to a very different interpretation. We think of $R(a,b)$ as the peak force required to move a to b. The value given by the composite $(S \circ R)(a,c)$ then describes optimum peak force we will require to move a to c. For example if we can convert a to b with one unit of force, and then move b to c for two units of force, then the peak force required is two units. An alternative procedure converting a to b' for zero units of cost, and then converting b' to c for 2.5 units of cost has a peak cost of 2.5 units, so we would prefer the first procedure to minimize our peak effort. Similarly, the truth value $(R \otimes R')(a,a',b,b')$ gives the peak force required to complete both conversions, assuming these costs are independently incurred. As with Example 6, we can think of states and effects as tables of acquisition and elimination forces.

Example 8. We can interpret ordinary relations over the Boolean quantale as modelling connectivity. $R(a,b)$ tells us that a is connected to b, composition tells us that we can chain connections together, and the tensor product tells us that we can connect pairs of elements together using a pair of connections between their components. Generalizing to the interval quantale, we now think of $R(a,b)$ as a "connection strength" between a and b. The composite $(S \circ R)(a,c)$ gives the best connection quality that we can achieve in two steps via B. Similarly, the parallel composite $(R \otimes R')(a,a',b,b')$ gives a conservative judgment of the connection quality we can achieve simultaneously between both a and b and a' and b' as the lower of the two individual connection strengths. States describe the "transmission strength" with which signals enter the system from the environment, and effects describe the "reception quality" on output signals.

Alternatively, we could view relations over $\mathbf{I}$ as fuzzy relations, with states and effects sets with fuzzy membership, and fuzzy predicates. Graded membership is widely used in cognitive science, for example in [35–39]. Concepts such as 'tall' have no crisp boundary and are better modelled using grades of membership. Although human concept use does not obey fuzzy logic [40], fuzzy relations may prove useful.

$\mathbf{Rel}(Q)$ is partial order-enriched if we order relations pointwise with respect to the underlying quantale order. It therefore makes sense to consider internal monads in $\mathbf{Rel}(Q)$ as interesting "structured objects". An internal monad on an object in a partially ordered category is an endomorphism R satisfying:

$$(R \circ R) \subseteq R, \qquad 1_A \subseteq R \tag{1}$$

Example 9. If we specialize condition (1) to $\mathbf{Rel}(\mathbf{L})$, it is equivalent to:

$$R(a,b) + R(b,c) \geq R(a,c), \qquad 0 = R(a,a)$$

[2] The slightly unusual formulation of identities is to avoid definition by cases. This means they can be interpreted in the internal language of an arbitrary topos.

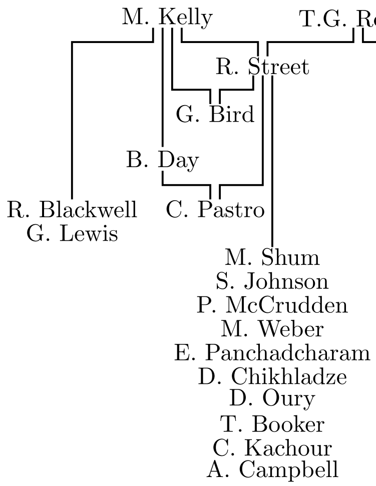


Fig. 2. The mathematical family tree.

We therefore consider these internal monads as describing *generalized metric spaces*. This observation is important in the field of monoidal topology [41].

As before, we can also interpret our internal monad as giving a well behaved collection of conversion costs between resources. Converting a resource to itself is free, and converting a resource via an intermediate state is at least as expensive as taking the direct route. Similarly, if we consider $\mathbf{Rel}(\mathbf{F})$ the conditions of (1) become:

$$\max(R(a,b), R(b,c)) \geq R(a,c), \qquad 0 = R(a,a)$$

and we can therefore see such internal monads as *generalized ultrametric spaces*. Again, the interpretation in terms of maximum force requirements extends to a sensible interpretation of these axioms.

Example 10. Internal monads in the category of ordinary relations are preorders on their underlying set. The generalization to the interval quantale then gives a fuzzy generalization of the notion of preorder. We can also apply our intuition in terms of connection strengths. Reflexivity tells us that every element can be perfectly connected to itself. Transitivity tell us that the optimal connection strength available is always at least as good as connecting via an intermediate node.

We now provide an extended example of the application of relations over generalized truth values.

Example 11 *(Family trees).* We will assume our universe of discourse to be the "mathematical family tree" in Fig. 2, built using information about supervisor relationships freely available from the mathematics genealogy project [42]. A vertical line represents a Supervisor–PhD student relationship, with the supervisor in diagrammatically higher position. For example, T.G. Room is the supervisor of A. Horadon, Ross Street was supervised by both Room and Kelly, and Kelly supervised five different students. We will define two individuals to be "academic siblings" if they share one or more supervisors. For example M. Shum and S. Johnson are academic siblings.

What makes this family tree interesting is that there are relationships that rarely occur in ordinary genealogy trees. For instance, Bird is both an academic sibling and a student of Street. In a real family graph this would imply an unconventional relationship in which Street is both a parent and sibling of Bird. Such possibilities make the academic family tree an interesting relationship with non-trivial structure.

We will freely borrow terms from genealogy, saying for instance that Shum is the cousin of Shannon, or that Kelly is an ancestor of Weber. We set the following goals:

- We want to use a relational model to give meaning to sentences such as "Bird is a student of Kelly".
- If we define other genealogical relationships such as "grandparent", "cousin" or "ancestor"in the natural way, we expect these definitions to coincide with the ones obtained compositionally in our model. Ideally, "Kelly is a academic grandparent of Shum" and "Kelly is a supervisor of a supervisor of Shum" should have the same meaning.
- We would like to express more complicated degrees of kinship, such as "Blackwell is the second-degree academic cousin of Horadon", again in a purely compositional way.
- We want this process of defining complicated relations from simpler ones to be scalable, such that it can be used on family trees of arbitrary size.

We model the compositional structure of these relationships using a very simple pregroup grammar, with only one basic type N denoting nouns. In particular, our sentence type will simply be the pregroup unit, meanings sentences will be

Fig. 3. Simple verbs for family trees.

interpreted as numbers[3] in our monoidal category. This is a rather heterodox choice: usually, the sentence type is assumed to have non-trivial structure because we are interested in comparing the meaning of a rich space of potential sentences. In our setting however, it is not particularly interesting to compare the sentences "Ralph is the brother of Mary" and "John is the son of Mark". Instead, what we would really like is to measure *how true* the individual sentences are, ideally quantifying the degree of kinship between the people involved. We can achieve this by creatively varying our choice of truth values.

As sentence meanings are interpreted as numbers, they correspond to a single truth value. If we choose **B** as quantale for truth values, in the spirit of Montague there are only two possible choices, a sentence is either true or false. Things will get more exciting once we move to less conventional truth values, but we begin with some simple examples.

Taking **B** as our quantale, we define the following relation pointwise in the obvious way:

$$C(x, y) = x \textit{ is the academic child of } y$$

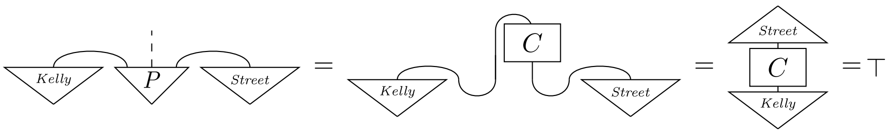
$C(x, y)$ is $\top$ if x is a child of y, and $\bot$ otherwise. Letting F° denote the converse of a relation F, as displayed diagrammatically by bending wires (Fig. 3a), we can build many interesting academic relationships out of the child relation C, for example:

$$\begin{aligned} S &= \left(C^\circ \circ C\right) \setminus 1_N && \textit{the sibling relationship} \\ P &= C^\circ && \textit{the parent relationship} \\ G &= P \circ P && \textit{the grandparent relationship} \\ K &= P \circ S \circ C && \textit{the cousin relationship} \end{aligned}$$

where 1_N is the identity relation on our nouns, given by

$$1_N(x, y) = \begin{cases} \top & \text{if } x = y \\ \bot & \text{otherwise} \end{cases}$$

We can interpret our various family tree relations as simple verbs, as illustrated in Fig. 3b. We draw the sentence space as a dashed wire, as it is actually the monoidal unit and would not normally be explicitly drawn. A simple graphical calculation establishes that "Kelly is a parent of Street", as follows:



Similar calculations show that "Shum is the cousin of Shannon", whereas "Shum is the cousin of Street" is false. More surprisingly, Pastro is his own cousin!

So far, so good. We now move to more expressive truth values that will allow us to quantify "how related" two individuals in the hierarchy are.

Definition 5. The **step quantale N** is given by the extended natural numbers $\mathbb{N} \cup \{\infty\}$ with the *reverse* ordering. Joins are minima and we take addition as the monoid multiplication. This can be seen as a discrete version of the Lawvere quantale **L**.

As expected, we now use **Rel**(**N**) as our semantics. In this case, we re-define C as follows:

$$C(x, y) = \begin{cases} 1 & \text{iff } x \text{ is directly below } y \\ \infty & \text{otherwise} \end{cases}$$

We then define the parent, grandparent and cousin relations as we did before. The sibling relation S is defined as $P \circ C$. It is easy to see how our truth values represent the degree of kinship between our individuals: a parent–child relation between

[3] Recall numbers are morphisms of type $I \to I$.

x and y can assume value one or ∞, depending if it is satisfied or not according to our tree. The sibling relationship S can have value two or ∞: we are considering being a sibling as a more distant relationship than parenthood. Although slightly surprising at first sight, this observation makes sense from an heraldry perspective, where the parent–child relationship is considered to be stronger than that of siblings. If two individuals are cousins, the degree of kinship will be 4, and so on. The strongest degree of all, zero, can only be attained by the identity relation, corresponding to "being oneself". Note how in this framework an individual can be considered "their own sibling" but, in doing so, this relation will be satisfied only with value two, while considering an individual as "oneself" attains value zero.

The impact of using the truth values in the quantale **N** is most pronounced when we consider relations such as "ancestor" and "relative". In order to do so, we extend the notion of transitive closure to relations over a quantale. Firstly, we define for a relation $F : X \to X$:

$$F^1 = F \quad \text{and} \quad F^{n+1} = F \circ F^n$$

The transitive closure can then be defined as the relation:

$$\overline{F}(x, y) = \bigvee\{F^n(x, y) \mid n \geq 1\}$$

The ancestor relation A is the transitive closure of the child relation C. The value $A(x, y)$ is lowest number of child relation "steps" from x to y, returning ∞ when x is not an ancestor of y.

The relative relation R is slightly more complex, we define it using the transitive closure as follows:

$$R = \overline{P \cup C \cup 1_N}$$

$R(x, y)$ is the shortest number of steps between x and y assuming that we can travel in either direction, and that we can always reach ourselves in zero steps.

4. Incorporating convexity

Up to this point, the domain and codomain of our relations have been sets. If we fix an algebraic structure (Σ, E) with set of operations Σ and equations between terms E, we can define a notion of binary relation between these algebras.

Definition 6. An **algebraic Q-relation** of type $A \to B$ is an ordinary Q-relation R between the underlying sets, such that for each operation $\sigma \in \Sigma$ of arity n the following inequation holds in the quantale order:

$$R(a_1, b_1) \otimes \ldots \otimes R(a_n, b_n) \leq R(\sigma(a_1, \ldots, a_n), \sigma(b_1, \ldots, b_n))$$

As shown in [24], algebraic Q-relations form a hypergraph category:

Theorem 2. *For commutative quantale Q and algebraic signature (Σ, E) there is a hypergraph category $\mathbf{Rel}_{(\Sigma,E)}(Q)$ with objects (Σ, E)-algebras and morphisms algebraic Q-relations.*

In the conceptual spaces literature, convexity is conceptually important. In [18] this convexity was captured using relations between convex algebras. We refer to [18] and the extended paper [31] for explicit modelling of toy computations of composed concepts in this category.

These convex algebras can be described as the Eilenberg–Moore algebras of the finite distribution monad. They can in fact be presented by a family Σ_c of binary operations

$$+^p, \quad p \in (0, 1)$$

satisfying suitable axioms. We can read $x +^p y$ as "choose x with probability p and y with probability $(1 - p)$". By considering algebraic **B**-relations over this signature, we can construct a category isomorphic to the category **ConvexRel** of convex relations from [18]. By changing our quantale of truth values, we can go further than this.

Proposition 1. *In the category of convex* **L***-relations, the internal monads are generalized metric spaces satisfying the additional axioms for $p \in (0, 1)$:*

$$R(a_1, b_1) + R(a_2, b_2) \geq R(a_1 +^p b_1, a_2 +^p b_2)$$

So internal monads in the category of convex relations over the Lawvere quantale are generalized metric spaces that interact well with formation of convex mixtures. The usual distance on $\mathbb{R}^n$ is an example of such a metric.

As shown in [24], every quantale homomorphism $h : Q_1 \to Q_2$ induces a strict monoidal functor of type $\mathbf{Rel}_{(\Sigma,E)}(Q_1) \to \mathbf{Rel}_{(\Sigma,E)}(Q_2)$. If the quantale morphism is injective, this functor is faithful. In particular, the mapping

$$\bot \mapsto \infty \qquad \top \mapsto 0$$

is an injective quantale homomorphism from the Boolean to the Lawvere quantale. This means we can find the ordinary Boolean binary relations as a monoidal subcategory of the category $\mathbf{Rel}(\mathbf{L})$. This presents some flexible modelling possibilities. If U and V are two subsets of a set X, they induce two states $U, V : I \to X$ in $\mathbf{Rel}(\mathbf{B})$. If we consider the number $V^\circ \circ U$, where R° denotes relational converse, it evaluates to true if and only if $U \cap V \neq \emptyset$.

Proposition 2. *If $U, V \subseteq X$ and d is an internal monad in $\mathbf{Rel}(\mathbf{L})$, the composite $V^\circ \circ d \circ U$ is the infimum of the distances between elements in U and V.*

This gives us the greatest lower bound on the distances between elements in U and V, providing a finer grain measure of similarity than can conventionally be achieved in relational models. We note that as distances are in general asymmetric, the number $U^\circ \circ d \circ V$ may give a different measure of similarity. Similarly, we can find the ordinary Boolean convex relations within the category of $\mathbf{L}$-valued convex relations, presenting analogous opportunities for performing calculations with discrete convex relations, and then measuring their separation on a continuum of values.

Such asymmetric distance measures are of practical use in cognitive science applications. A fundamental concept in psychology is that of similarity, which can be used as the basis of concept formation. Similarity between objects or concepts can be explained by locating objects in some sort of conceptual or feature space, and modelling similarity as a function of distance, for example in [43]. However, judgements of similarity are not necessarily symmetric [44]. In one study examining the similarity between pairs of countries, participants are asked to choose between statements ‘Country A is similar to country B’ or ‘Country B is similar to country A’. In all cases, a majority of participants preferred the statement where the latter country was considered more prominent.

5. Proof relevance

A **span** S of sets, between sets A and B, is a set X and a pair of functions $X \xrightarrow{p_1} A$ and $X \xrightarrow{p_2} B$. Parallelling the notation for relations, we will write

$$S_x(a, b) := x \in X \ \wedge \ p_1(x) = a \ \wedge \ p_2(x) = b$$

We can think of such a span as a *proof relevant relation* in which $S_x(a, b)$ tells us that x witnesses that a and b are related. In a computational linguistics or cognition application where relations may have been derived automatically from data in some way, we can exploit these proof witnesses to track evidence for our beliefs that certain relationships hold.

Sets and spans between them form a hypergraph category **Span** with composition given by pullback, and tensor product induced by a choice of products.[4] In fact, as we did for relations, we can extend these spans with algebraic structure and a choice of truth values in a partially ordered monoid. We no longer require full quantale structure on our truth values, as multiple proof witnesses mean we don’t need to choose a single representative truth value when composing relations.

Definition 7. For an algebraic signature (Σ, E) and pomonoid Q an **algebraic Q-span** of type $A \to B$ between (Σ, E)-algebras is a span $A \xleftarrow{p_1} X \xrightarrow{p_2} B$ between the underlying objects, with a **characteristic morphism** $\chi : X \to Q$. We require that the algebraic structure is respected in that for all $\sigma \in \Sigma$, with arity n:

$$\bigwedge_{1 \le i \le n} (p_1(x_i) = a_i \wedge p_2(x_i) = b_i) \Rightarrow \bigotimes_{1 \le i \le n} \chi(x_i) \le \chi(\sigma(x_1, ..., x_n))$$

Intuitively, these are intensional relations in which proof witnesses are weighted by a truth value, and the relations respect the algebraic structure. As shown in [24], algebraic Q-spans also form a hypergraph category:

Theorem 3. *For commutative pomonoid Q and algebraic signature (Σ, E) there is a hypergraph category $\mathbf{Span}_{(\Sigma, E)}(Q)$ with objects (Σ, E)-algebras and morphisms algebraic Q-spans.*

For algebraic Q span S we define

$$S^q_x(a, b) := x \in X \ \wedge \ p_1(x) = a \ \wedge \ p_2(x) = b \ \wedge \ \chi(x) = q$$

We then read $S^q_x(a, b)$ as telling us that x witnesses that a and b are related with strength q. In fact, we can order algebraic Q-spans in a manner similar to that for relations, but accounting for proof witnesses.

[4] In fact, in order for composition to be associative, it is necessary to work with equivalence classes of spans. It is sufficient to consider representatives, and we do so to avoid distracting technicalities.

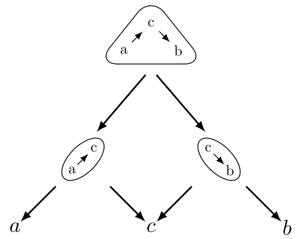


Fig. 4. Interpretation of pullbacks as composition of paths.

Definition 8. For pomonoid Q, we define a preorder on algebraic Q-spans by setting $(X_1, f_1, g_1, \chi_1) \subseteq (X_2, f_2, g_2, \chi_2)$ if there is a **Set**-monomorphism $\varphi : X_1 \to X_2$ such that $f_1 = f_2 \circ \varphi$, $g_1 = g_2 \circ \varphi$ and $\forall x. \chi_1(x) \leq \chi_2(\varphi(x))$.

The ordering accounts pointwise for strengths of relatedness in a natural way. The requirement that the function φ in Definition 8 is a monomorphism ensures that even if our truth values are trivial, we take account of the "number" of proof witnesses available.

As internal monads provided interesting objects in the setting of relations, we should consider them in the span setting as well.

Proposition 3. *An internal monad on A in* $\mathbf{Span}(\mathbf{L})$ *is an* $\mathbf{L}$*-span $S : A \to A$ such that if $S_x^p(a_1, a_2)$ and $S_y^q(a_2, a_3)$ we can choose an element $\varphi(x, y)$ of the apex such that $S^r_{\varphi(x,y)}(a_1, a_3)$ and $p + q$ is greater than r in the usual ordering on the real numbers. Furthermore, we can do this in a way such that the assignment φ is injective.*

So internal **L**-span monads further generalize metric spaces to incorporate multiple possible distances, which we can think of as describing different paths between points. We discuss two concrete examples.

Example 12 *(Semantic ambiguity via spans).* In natural language, we often encounter ambiguous situations. For example the word "bank" can refer to either a "river bank" or a "financial bank". A compositional account of semantic ambiguity was presented in [10], using mathematical models of incomplete information from quantum theory. The techniques applied implicitly assume meanings are built upon a vector space model, to which we apply Selinger's CPM construction [9] to yield a new category of ambiguous meanings. The CPM construction can also be applied to categories of relations, but in this case it does not provide a satisfactory model of ambiguity [45].

An alternative approach to ambiguity in relational models is to use spans. We consider how the ambiguous word "bank" is related to the word "water"

- In the "river bank" context, we would expect a strong relationship;
- In the "financial bank" context, we would expect a weaker relationship.

By using spans rather than relations, we can introduce two different proof witnesses for the different contexts under consideration. By choosing our quantale of truth values to be the Lawvere quantale **L**, we can attach a different choice of distance to each of these choices. As we compose spans to describe the meanings of phrases and sentences, the proof witnesses will keep track of the different possible relationships in play.

Example 13 *(Proof relevant family trees).* We return to the family tree Example 11, this time formalizing our semantics in $\mathbf{Span}(\mathbf{N})$. The intuition for such a span $(N \xleftarrow{f} X \xrightarrow{g} N, \chi)$ is that an element $x \in X$ witnesses a path from $f(x)$ to $g(x)$ of length $\chi(x)$. For example, we can introduce a span C describing the child relationship, admitting a path from a to b of length 1 if and only if a is a child of b. The parent span P is the converse of the child span, given by reversing its legs. A composite of two spans encodes composites of compatible paths and the sum of their corresponding lengths. The sibling span is the composite $P \circ C$, illustrated in Fig. 4.

If a and b are siblings, they must have some common parent c, resulting in a length two path $a \to c \to b$, as illustrated in Fig. 4. If the pair a and b have two different common parents, in contrast to the case of relations where this information is lost, the composite span will record two distinct paths between them. Similarly, if we generalize the ancestor relation to a span, it would witness every possible way of relating two members of the family tree, and record the corresponding path length. In this way, we would explicitly record that Bird is related to Kelly in two distinct ways, directly in one step, and via Street in two steps.

6. Varying the underlying topos

Our definitions of algebraic Q-relations and algebraic Q-spans are constructive. This means that Theorems 2 and 3 continue to hold for any elementary topos, as proved in [24]. Background on topos theory can be found in [46]. We will write $\mathbf{Rel}^{\mathcal{E}}_{(\Sigma,E)}(Q)$ and $\mathbf{Span}^{\mathcal{E}}_{(\Sigma,E)}(Q)$ for the categories of spans and relations, to make the choice of topos $\mathcal{E}$ explicit. We have already seen that **Rel** and $\mathbf{Rel}(\mathbf{B})$ are isomorphic as categories.

We conclude by establishing a similar connection between our framework of generalized relations and the standard notion of the category of relations over a regular category. This will involve the internal locale given by the subobject classifier.

Definition 9. A category $\mathcal{C}$ is **regular** if it is finitely complete, every kernel pair has a coequalizer and regular epimorphisms are stable under pullback.

There is standard construction of a category of relations $\mathbf{Rel}(\mathcal{C})$ of a regular category $\mathcal{C}$, see for example [47]. For the category **Set**, this construction recovers exactly the usual category of binary relations. Every topos is regular, and in fact for any algebraic theory (Σ, E), the category of internal (Σ, E)-algebras in a regular category [48], meaning we can consider the impact of algebraic structure. In fact, the resulting category of relations is equivalent to the one produced by our construction with the subobject classifier as the object of truth values. In this way, we see that relations over suitable regular categories are a special case of our construction.

Theorem 4. *Let $\mathcal{E}$ be a topos, Ω its subobject classifier and (Σ, E) an algebraic signature. The category $\mathbf{Rel}^{\mathcal{E}}_{(\Sigma,E)}(\Omega)$ resulting from the algebraic Q-relations construction is equivalent to the category of internal relations over the regular category of internal (Σ, E)-algebras in $\mathcal{E}$.*

7. Conclusion

We have demonstrated that categories of generalized relations present a flexible modelling tool for categorical compositional models of natural language and cognition, presenting case studies to motivate our claims. We also outlined various potential models worthy of further investigation, capturing features such as fuzziness, distances, convexity and ambiguity, and showed how these features can be used in combination within a generic framework. One natural direction for further work would be empirical investigation of the compatibility of these theoretical models with concrete applications. Another one would be to investigate whether the techniques in [49] can be used to build models with either non-commutative or typed quantales, known as quantaloids.

Acknowledgements

This work was funded by AFOSR grant “Algorithmic and Logical Aspects when Composing Meanings” (grant number FA9550-14-1-0079) and FQXi grant “Categorical Compositional Physics” (grant number 2015-145578).

References

[1] B. Coecke, M. Sadrzadeh, S. Clark, Mathematical foundations for distributed compositional model of meaning, in: Lambek Festschrift, Linguist. Anal. 36 (2010) 345–384.
[2] E. Grefenstette, M. Sadrzadeh, Experimental support for a categorical compositional distributional model of meaning, in: The 2014 Conference on Empirical Methods on Natural Language Processing, 2011, pp. 1394–1404, arXiv:1106.4058.
[3] D. Kartsaklis, M. Sadrzadeh, Prior disambiguation of word tensors for constructing sentence vectors, in: The 2013 Conference on Empirical Methods on Natural Language Processing, ACL, 2013, pp. 1590–1601.
[4] S. Clark, B. Coecke, E. Grefenstette, S. Pulman, M. Sadrzadeh, A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure, Malays. J. Math. Sci. 8 (2014) 15–25, arXiv:1305.0556.
[5] S. Abramsky, B. Coecke, A categorical semantics of quantum protocols, in: Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 2004, IEEE, 2004, pp. 415–425.
[6] B. Coecke, É.O. Paquette, D. Pavlović, Classical and quantum structuralism, in: S. Gay, I. Mackie (Eds.), Semantic Techniques in Quantum Computation, Cambridge University Press, 2010, pp. 29–69, arXiv:0904.1997.
[7] M. Sadrzadeh, S. Clark, B. Coecke, The Frobenius anatomy of word meanings I: subject and object relative pronouns, J. Logic Comput. (2013), ext044.
[8] D. Kartsaklis, M. Sadrzadeh, A Frobenius model of information structure in categorical compositional distributional semantics, in: The 14th Meeting on the Mathematics of Language, 2015, p. 62.
[9] P. Selinger, Dagger compact closed categories and completely positive maps, Electron. Notes Theor. Comput. Sci. 170 (2007) 139–163.
[10] R. Piedeleu, D. Kartsaklis, B. Coecke, M. Sadrzadeh, Open system categorical quantum semantics in natural language processing, in: 6th Conference on Algebra and Coalgebra in Computer Science, CALCO 2015, 2015, pp. 270–289.
[11] B. Coecke, A. Kissinger, Picturing Quantum Processes. A First Course in Quantum Theory and Diagrammatic Reasoning, Cambridge University Press, 2017.
[12] F. Bonchi, P. Sobocinski, F. Zanasi, Full abstraction for signal flow graphs, ACM SIGPLAN Not. 50 (1) (2015) 515–526.
[13] J.C. Baez, J. Erbele, Categories in control, Theory Appl. Categ. 30 (24) (2015) 836–881.
[14] J.C. Baez, B. Fong, B.S. Pollard, A compositional framework for Markov processes, J. Math. Phys. 57 (3) (2016) 033301.
[15] J.C. Baez, B. Fong, A compositional framework for passive linear networks, arXiv:1504.05625.

[16] P. Sobocinski, Graphical linear algebra, mathematical blog, https://graphicallinearalgebra.net/.
[17] B. Coecke, E. Grefenstette, M. Sadrzadeh, Lambek vs. Lambek: functorial vector space semantics and string diagrams for Lambek calculus, Ann. Pure Appl. Logic 164 (11) (2013) 1079–1100.
[18] J. Bolt, B. Coecke, F. Genovese, M. Lewis, D. Marsden, R. Piedeleu, Interacting conceptual spaces, in: Proceedings of the 2016 Workshop on Semantic Spaces at the Intersection of NLP, Physics and Cognitive Science, 2016, pp. 11–19.
[19] P. Gärdenfors, Conceptual Spaces: The Geometry of Thought, MIT Press, 2004.
[20] P. Gärdenfors, The Geometry of Meaning: Semantics Based on Conceptual Spaces, MIT Press, 2014.
[21] G. Ludwig, An Axiomatic Basis of Quantum Mechanics. 1. Derivation of Hilbert Space, Springer-Verlag, 1985.
[22] J. Barrett, Information processing in generalized probabilistic theories, Phys. Rev. A 75 (2007) 032304.
[23] B. Fong, The Algebra of Open and Interconnected Systems, Ph.D. thesis, University of Oxford, 2016.
[24] D. Marsden, F. Genovese, Custom hypergraph categories via generalized relations, in: 7th Conference on Algebra and Coalgebra in Computer Science (CALCO 2017), 2017.
[25] B. Coecke, F. Genovese, M. Lewis, D. Marsden, Generalized relations in linguistics & cognition, in: International Workshop on Logic, Language, Information, and Computation (WOLLIC), 2017.
[26] M. Dostal, M. Sadrzadeh, Many Valued Generalised Quantifiers for Natural Language in the DisCoCat Model, Tech. rep., Queen Mary University of London, 2016.
[27] D. Bankova, Comparing Meaning in Language and Cognition – p-Hyponoymy, Concept Combination, Asymmetric Similarity, Master's thesis, University of Oxford, 2015.
[28] D. Bankova, B. Coecke, M. Lewis, D. Marsden, Graded entailment for compositional distributional semantics, arXiv:1601.04908.
[29] J. Lambek, Type grammar revisited, in: Logical Aspects of Computational Linguistics, Springer, 1999, pp. 1–27.
[30] H. Schütze, Automatic word sense discrimination, Comput. Linguist. 24 (1) (1998) 97–123.
[31] J. Bolt, B. Coecke, F. Genovese, M. Lewis, D. Marsden, R. Piedeleu, Interacting conceptual spaces I: grammatical composition of concepts, arXiv:1703.08314.
[32] B. Coecke, E.O. Paquette, Categories for the practising physicist, in: New Structures for Physics, Springer, 2010, pp. 173–286.
[33] A. Kissinger, Finite matrices are complete for (dagger-)hypergraph categories, arXiv:1406.5942.
[34] M. Sadrzadeh, S. Clark, B. Coecke, The Frobenius anatomy of word meanings II: possessive relative pronouns, J. Logic Comput. (2014), exu027.
[35] E. Rosch, C.B. Mervis, Family resemblances: studies in the internal structure of categories, Cogn. Psychol. 7 (4) (1975) 573–605.
[36] L.W. Barsalou, Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories, J. Exper. Psychol., Learn., Mem., Cogn. 11 (4) (1985) 629.
[37] J.A. Hampton, Overextension of conjunctive concepts: evidence for a unitary model of concept typicality and class inclusion, J. Exper. Psychol., Learn., Mem., Cogn. 14 (1) (1988) 12.
[38] J.A. Hampton, Disjunction of natural concepts, Mem. Cogn. 16 (6) (1988) 579–591.
[39] R. Dale, C. Kehoe, M.J. Spivey, Graded motor responses in the time course of categorizing atypical exemplars, Mem. Cogn. 35 (1) (2007) 15–28.
[40] D.N. Osherson, E.E. Smith, Gradedness and conceptual combination, Cognition 12 (3) (1982) 299–318, https://doi.org/10.1016/0010-0277(82)90037-3.
[41] D. Hofmann, G.J. Seal, W. Tholen, Monoidal Topology: A Categorical Approach to Order, Metric, and Topology, vol. 153, Cambridge University Press, 2014.
[42] H.B. Coonce, The mathematics genealogy project, website provided by North Dakota State University, https://www.genealogy.math.ndsu.nodak.edu/index.php.
[43] R.N. Shepard, et al., Toward a universal law of generalization for psychological science, Science 237 (4820) (1987) 1317–1323.
[44] A. Tversky, Features of similarity, Psychol. Rev. 84 (4) (1977) 327.
[45] D. Marsden, A graph theoretic perspective on CPM(Rel), in: Proceedings 12th International Workshop on Quantum Physics and Logic, QPL 2015, Oxford, UK, July 15–17, 2015, 2015, pp. 273–284.
[46] S. MacLane, I. Moerdijk, Sheaves in Geometry and Logic: A First Introduction to Topos Theory, Springer Science & Business Media, 2012.
[47] F. Borceux, Handbook of Categorical Algebra, vol. 2, Categories and Structures, Cambridge University Press, 1994.
[48] M. Barr, Exact categories, in: Exact Categories and Categories of Sheaves, Springer, 1971, pp. 1–120.
[49] I. Stubbe, Categorical Structures Enriched in A Quantaloid: Categories and Semicategories, Ph.D. thesis, Université Catholique de Louvain, 2003.