

A toy distributional model for fuzzy generalised quantifiers

Mehrnoosh Sadrzadeh

Department of Computer Science
University College London
m.sadrzadeh@ucl.ac.uk

Gijs Wijnholds

Utrecht Institute of Linguistics OTS
Utrecht University
g.j.wijnholds@uu.nl

Abstract

Recent work in compositional distributional semantics showed how bialgebras model generalised quantifiers of natural language. That technique requires working with vector space over power sets of bases, and therefore is computationally costly. It is possible to overcome the computational hurdles by working with fuzzy generalised quantifiers. In this paper, we show that the compositional notion of semantics of natural language, guided by a grammar, extends from a binary to a many valued setting and instantiate in it the fuzzy computations. We import vector representations of words and predicates, learnt from large scale compositional distributional semantics, interpret them as fuzzy sets, and analyse their performance on a toy inference dataset.

1 Introduction

The work of [10] showed how one can reason about generalised quantifiers using bialgebras over the category of sets and relations over a fixed powerset object (powerset of a universe of discourse). This provides us with an abstract categorical semantics, which when instantiated to category of sets and relations, one will obtain a truth-theoretic semantics. The abstract setting, however, can also be instantiated to category of vector spaces and linear maps, in which one obtains a compositional distributional semantics, in the style of [6, 9]. The downside is that the resulting vector spaces span over powersets of the usual set of bases and the complexity of reasoning in the setting explodes. It is also not very clear how can one learn the new basis vectors, consisting of sets of vectors, rather than just one vectors. One solution would be to move to a fuzzy setting, as done in [17]. The rationale behind this move is as follows: fuzzy sets have been encoded in the category of sets and many valued relations and the categorical set-

ting of [10] also instantiates to these categories. We demonstrate the details of this construction in the Springer Outstanding Contributions volume in honor of M. Ardeshir. In that paper, we show that the categorical version of fuzzy sets **V-Rel** of sets and many valued relations, is compact closed and define over it the necessary bialgebras to encode Zadeh’s fuzzy generalised quantifiers.

In this paper, we spare the categorical technicalities, and review the definitions of generalised quantifiers in a compositional relational setting (sets and relations) guided by an elementary generative grammar. Independently, we also review the definitions of fuzzy generalised quantifiers of Zadeh, using the notions of fuzzy sets and possibility distributions. We then explain how fuzzy sets can be modelled by many valued relations and define a compositional semantics for sentences with fuzzy generalised quantifiers in this setting. Finally, we interpret vectors as fuzzy sets and show how the many valued semantic computations can be done over vectorial data. We demonstrate the workings of our model on toy vectors extracted from real data and compute a degree of truth for quantified sentences containing them. In order to ground our semantics, i.e. conclude that these computations are sound, we use the results in a toy inference task and analyse the results.

Finally, although fuzzy concepts are often motivated by vague predicates such as short and tall, fuzzy generalised quantifiers have a large, if not full, overlap with natural language generalised quantifiers. Most of the latter are non-logical and consider words such as “almost, many, most, few”, and these are exactly the fuzzy generalised quantifiers that Zadeh deals with. Further, and as we will see in the examples that have been worked out in the paper, one can apply our methodology to deal with logical part of generalised quantifiers, i.e. words such as “all” and “some”.

2 Generalised Quantifiers as Relations

We briefly review the theory of generalised quantifiers in natural language as presented in [2]. Consider the fragment of English generated by the following context free grammar.

- S → NP VP
- VP → V NP
- NP → Det N
- NP → John, Mary, ...
- N → cat, dog, man, ...
- VP → sneeze, sleep, ...
- V → love, kiss, ...
- Det → some, all, no, most, almost all, several, ...

A model for the language generated by this grammar is a pair $(U, \llbracket \cdot \rrbracket)$, where U is a universal reference set and $\llbracket \cdot \rrbracket$ is an inductively defined interpretation function. In order to keep the semantic simple, we will not fully follow formal semantics guidelines and shall not treat noun phrases as general quantifiers. Noun phrases and nouns are treated similarly and by sets. The $\llbracket \cdot \rrbracket$ of terminals is thus defined via the following cases:

1. The interpretation of a determiner d generated by ‘Det → d ’ is the following map:

$$\llbracket d \rrbracket: \mathcal{P}(U) \rightarrow \mathcal{PP}(U)$$

It assigns to each $A \subseteq U$, a family of subsets of U . The images of these interpretations are referred to as *generalised quantifiers*. For logical quantifiers, these are:

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

A similar method is used to define non-logical quantifiers, for example “most A ” is defined to be the set of subsets of U that has ‘most’ elements of A , “few A ” is the set of subsets of U that contain ‘few’ elements of A , and similarly for ‘several’ and ‘many’.

Generalising the two cases above, provides us with the following definition for semantics $\llbracket d \rrbracket(A)$ of any generalised quantifier d :

$$\{X \subseteq U \mid X \text{ has } d \text{ elements of } A\}$$

2. The interpretation of a terminal $y \in \{np, n, vp\}$ generated by either of the rules ‘NP → np , N → n , VP → vp ’ is $\llbracket y \rrbracket \subseteq U$. That is, noun phrases, nouns and verb phrases are interpreted as subsets of the reference set.
3. The interpretation of a terminal y generated by the rule $V \rightarrow y$ is $\llbracket y \rrbracket \subseteq U \times U$. That is, verbs are interpreted as binary relations over the reference set.

The semantics of $\llbracket \cdot \rrbracket$ on non-terminals is defined according to the following cases:

1. The interpretation of expressions generated by the rule ‘NP → Det N’ is:

$$\llbracket \text{Det N} \rrbracket = \llbracket d \rrbracket(\llbracket n \rrbracket)$$

where $X \in \llbracket d \rrbracket(\llbracket n \rrbracket)$ iff $X \cap \llbracket n \rrbracket \in \llbracket d \rrbracket(\llbracket n \rrbracket)$, for Det → d and N → n . This condition is often referred to as *conservativity* or the *living on* property of generalised quantifiers.

2. The interpretations of expressions generated by other rules are as usual:

$$\begin{aligned} \llbracket V \text{ NP} \rrbracket &= \llbracket v \rrbracket(\llbracket np \rrbracket) \\ \llbracket \text{NP VP} \rrbracket &= \llbracket vp \rrbracket(\llbracket np \rrbracket) \end{aligned}$$

Here, for $R \subseteq U \times U$ and $A \subseteq U$, by $R(A)$ we mean the forward image of R on A , that is $R(A) = \{y \mid (x, y) \in R, \text{ for } x \in A\}$. To keep the notation unified, for R a unary relation $R \subseteq U$, we use the same notation and define $R(A) = \{y \mid y \in R, \text{ for } x \in A\}$, i.e. $R \cap A$.

The ‘meaning’ of a sentence in this setting is its truth value. So we have that a sentence is true iff $\llbracket \text{NP VP} \rrbracket \neq \emptyset$ and false otherwise. For the cases of quantified sentences considered in this paper, i.e. sentences with quantified subject and object phrases, a truth value is defined as follows:

1. A sentence of the form ‘Det N VP’ is *true* iff $\llbracket \text{Det N VP} \rrbracket = \llbracket vp \rrbracket \cap \llbracket n \rrbracket \in \llbracket \text{Det N} \rrbracket$ and *false* otherwise.
2. A sentence of the form ‘NP V Det N’ is *true* iff $\llbracket \text{NP V Det N} \rrbracket = \llbracket n \rrbracket \cap \llbracket v \rrbracket(\llbracket np \rrbracket) \in \llbracket \text{Det N} \rrbracket$ and *false* otherwise.

For example, the sentence ‘some cats slept’ with a quantifier at the subject phrase is true iff $\llbracket \text{slept} \rrbracket \cap$

$\llbracket \text{cats} \rrbracket \in \llbracket \text{some cats} \rrbracket$, that is, whenever the set of things that sleep and are cats is a non-empty set. Similarly, a sentence with a quantified phrase at its object position, for instance, ‘Cats like some rats’ is true iff $\llbracket \text{rats} \rrbracket \cap \llbracket \text{likes} \rrbracket(\llbracket \text{cats} \rrbracket) \in \llbracket \text{some rats} \rrbracket$, that is, whenever, the set of things that are liked by cats and are rats is a non-empty set. Similarly, the sentence ‘Cats liked three rats’ is true iff the set of things that are liked by cats and are rats has three elements in it.

3 Zadeh’s Fuzzy Generalised Quantifiers

In this section we review definitions of fuzzy sets and quantifiers, as done by Zadeh [17]. A fuzzy set is a set whose elements have a corresponding weight associated to them. For a set A , the weight μ_i of element u_i is interpreted as the degree of membership of u_i in A . The fuzzy set A with n elements is represented symbolically by a sum:

$$A = \mu_1 u_1 + \mu_2 u_2 + \dots + \mu_n u_n$$

The cardinality of a fuzzy set is defined via the notion of *sigma-count*, defined below:

$$\Sigma \text{Count}(A) = \sum_{i=1}^n \mu_i$$

Terms whose degrees of membership fall below a certain threshold, may be omitted from the sum. This is to avoid a situation where a large number of terms with low degrees become equivalent to a small number of terms with high degrees.

The quantified sentences Zadeh considers are built from two basic forms: ‘‘There are Q A ’s’’ and ‘‘ Q A ’s are B ’s’’. Each of these propositions induces a possibility distribution. Zadeh provides the following insights for the analysis of these quantified propositions. ‘‘There are Q A ’s’’ implies that the probability of event A is a fuzzy probability equal to Q . ‘‘ Q A ’s are B ’s’’ implies that the conditional probability of event B given event A is a fuzzy probability which is equal to Q . Most statements involving fuzzy probabilities may be replaced by semantically equivalent propositions involving fuzzy quantifiers and this is the statement we work with in this paper.

The fuzzy semantics of a proposition p is interpreted as ‘‘the degree of truth of p ’’, or the possibility of p , where possibility is treated in an elementary way, i.e. a function from a set to the unit interval. In order to compute this, one translates p into a *possibility assignment equation*, denoted

by $\Pi_{(X_1, \dots, X_n)} = F$, where F is a fuzzy subset of the universe of discourse U and $\Pi_{(X_1, \dots, X_n)}$ is the joint possibility distribution over (explicit or implicit) variables X_1, \dots, X_n of p . For instance, the proposition ‘‘Vickie is tall’’ is translated as:

$$\Pi_{\text{Height}(\text{Vickie})} = \text{TALL}$$

Here, TALL is a fuzzy subset, $\text{Height}(\text{Vickie})$ is a variable implicit in ‘‘Vickie is tall’’, and $\Pi_{\text{Height}(\text{Vickie})}$ is the possibility distribution of this variable. Following Zadeh, we use the $=$ sign, but are aware that this makes the reading awkward. The reader is encouraged to treat (as we did) $=$ as an informal assignment. The above possibility assignment equation implies that

$$\text{Poss}\{\text{Height}(\text{Vickie}) = u\} = \mu_{\text{TALL}}(u)$$

where $\text{Poss}\{X = u\}$ the possibility that X is u , for u a specified value. The above thus states that ‘‘the possibility that height of Vickie is u is equal to $\mu_{\text{TALL}}(u)$, that is, is the grade of membership of u in the fuzzy set TALL . Quantified sentences are translated in a similar way. For instance, ‘‘Vickie has several credit cards’’, is translated to the following:

$$\Pi_{\text{Count}(\text{CreditCards}(\text{Vickie}))} = \text{SEVERAL}$$

Suppose that 4 is compatible with the meaning of ‘‘several’’ with degree 0.8, then the above implies that, for instance, the possibility that Vickie has 4 credit cards is

$$\text{Poss}\{\text{Count}(\text{CreditCards}(\text{Vickie})) = 4\} = 0.8$$

In order to analyse sentences of the general forms ‘‘There are Q A ’s’’ and ‘‘ Q A ’s are B ’s’’, Zadeh assumes that they are semantically equivalent to the following:

$$\begin{aligned} \text{There are } Q \text{ } A\text{'s} &\rightsquigarrow \Sigma \text{Count}(A) \text{ is } Q \\ Q \text{ } A\text{'s are } B\text{'s} &\rightsquigarrow \text{Proportion}(B|A) \text{ is } Q \end{aligned}$$

Here, $\text{Proportion}(B|A)$ is the proportion of elements of B that are in A , aka the relative cardinality of B in A , formally defined as follows:

$$\Pi_{\text{Proportion}(B|A)} := \frac{\Sigma \text{Count}(A \cap B)}{\Sigma \text{Count}(A)}$$

Both $\text{Proportion}(B|A)$ and $\Sigma \text{Count}(A)$ may be fuzzy or non-fuzzy counts. Zadeh then formalises

the above counts as possibility assignment equations as follows

$$\begin{aligned} \Sigma \text{Count}(A) \text{ is } Q &\rightsquigarrow \Pi_{\Sigma \text{Count}(A)} = Q \\ \text{Proportion}(B|A) \text{ is } Q &\rightsquigarrow \Pi_{\text{Proportion}(B|A)} = Q \end{aligned}$$

In the spirit of truth-conditional semantics, the weight of each of the elements of the set can be interpreted as the degree of truth of the proposition denoted by the element. This weight is $Q(\Sigma \text{Count}(A))$ for sentences of the form ‘‘There are Q A ’s’’ and $Q(\text{Proportion}(B|A))$ for sentences of the form ‘‘ Q A ’s are B ’s’’.

Writing $\mu_A(u)$ for the degree of membership of u in the fuzzy set A , we define the intersection of two fuzzy sets A and B as

$$A \cap B = \sum_i \min(\mu_A(u_i), \mu_B(u_i)) u_i$$

where i is understood to range over all the elements in A and B (when an element is in A but not in B it will still be represented in A with a degree of 0). A similar version without the Σ is used to define it for the non-fuzzy case.

Example. Let’s say we have a universe

$$U = \{u_1, u_2, u_3, u_4, u_5\}$$

and fuzzy sets KP for ‘‘kind people’’ and BP for ‘‘big men’’, defined as follows:

$$\begin{aligned} KP &= 0.5u_1 + 0.8u_2 + 0.2u_3 + 0.6u_4 \\ BM &= 0.8u_1 + 0.3u_2 + 0.1u_3 + 0.9u_4 + 1u_5 \end{aligned}$$

The quantified sentence ‘‘Most big men are kind’’, is translated to the following possibility assignment equation $\Pi_{\text{Proportion}(KP|BM)} = \text{MOST}$. The intersection of KP and BM is computed as follows:

$$KP \cap BM = 0.5u_1 + 0.3u_2 + 0.1u_3 + 0.6u_4$$

The proportion of big men that are kind, i.e. $\text{Proportion}(KP|BM)$, is computed as follows:

$$\frac{\Sigma \text{Count}(KP \cap BM)}{\Sigma \text{Count}(BM)} = \frac{1.5}{3.1}$$

Suppose that proportions between 0.6 and 0.7 are compatible with the meaning of *MOST* with degree 0.75. Then, since $\frac{1.5}{3.1} = 0.48$, the degree of truth of our sentence is below 0.75. For the crisp quantifier *ALL*, the sentence ‘‘All big men

are kind’’ is, since only the proportion 1 is compatible with the meaning of *ALL* with degree 1, which is not the case here.

Possibility distributions can be learnt, e.g. Zadeh develops a test-score procedure by sampling from a database of related data.

4 Fuzzy Generalised Quantifiers as Many Valued Relations

A *many-valued relation* between two sets A and B is denoted by $R : A \dashv\vdash B$ and is a function $R : A \times B \rightarrow \mathbf{V}$, where \mathbf{V} is a commutative quantale of values, usually the unit interval $[0, 1]$. This function is viewed as a \mathbf{V} -valued matrix. We compose two relations $R : A \dashv\vdash B$ and $S : B \dashv\vdash C$ to get a relation $S \circ R : A \dashv\vdash C$ such that

$$(S \circ R)(a, c) = \bigvee_{b \in B} (R(a, b) \bullet S(b, c))$$

holds in \mathbf{V} . Here, \bullet and \bigvee are operations on the numbers in the quantale \mathbf{V} . When \mathbf{V} is the real interval $[0, 1]$ with operations \min and \max , the composition of two \mathbf{V} -relations becomes as follows. Given two \mathbf{V} -relations $R : A \dashv\vdash B$ and $S : B \dashv\vdash C$ (so two functions $R : A \times B \rightarrow [0, 1]$ and $S : B \times C \rightarrow [0, 1]$), the composite $S \circ R : A \dashv\vdash C$ is given by

$$(S \circ R)(a, c) = \max_{b \in B} \min(R(a, b), S(b, c)).$$

We refer to sets and many valued relations on them as **V-Rel**.

A non-fuzzy generalised quantifier d is interpreted as a relation $\llbracket d \rrbracket$ over the power set of the universe of discourse $P(U)$, where it relates a subset $A \subseteq U$ to subsets $B \subseteq U$, based on the cardinalities of A and B . The fuzzy version of this quantifier is interpreted as a many valued relation over $P(U)$, where, in fuzzy set notation, it relates A to subsets $u_i \subseteq U$ and assigns to each such subset a degree of membership μ_i . The result is a fuzzy set whose weights come from a possibility distribution over the relative cardinalities of A and u_i ’s. In Zadeh’s notation:

$$\llbracket d \rrbracket(\text{Proportion}(u_i|A)) = \mu_i \quad (1)$$

For $\mathbf{V} = [0, 1]$ and given a fuzzy generalised quantifier for which we have $\Pi_{\text{Proportion}(B|A)} = \llbracket d \rrbracket$, we define its **V-Rel** encoding to be the many valued relation $\llbracket \llbracket d \rrbracket \rrbracket : P(U) \dashv\vdash P(U)$, with values

Table 1: Sentences of our Toy Dataset and their Annotated Degrees of Entailment

Entry 1	Entry 2	Deg.
people strike	group attacks	4.31
notice advertises	sign announces	5.37
clarify rule	explain process	5.00
recommend development	suggest improvement	5.37
people clarify rule	group explain process	5.00
corporation recommend development	firm suggest improvement	5.375
office arrange task	staff organize work	5.50
editor threatens	application predicts	1.12
progress reduces	development replaces	1.22
confirm number	approve performance	1.81
editor threatens man	application predicts number	1.12
man recall time	firm cancel term	1.62

coming from the possibility distribution of $\llbracket d \rrbracket$, defined as follows:

$$\llbracket d \rrbracket(A, B) = \mu_i, \text{ for } \mu_i = \llbracket d \rrbracket(\text{Proportion}(B|A))$$

In order to obtain a many valued relation in **V-Rel**, we need a numerical value assigned to subsets A and B of universe. This number is nothing but the weight of $\llbracket d \rrbracket(\text{Proportion}(B|A))$.

The semantics of a sentence of our grammar extends from sets and relations to sets and many valued relations. We define a **V-Rel** model to be the tuple $(\mathbf{V-Rel}, P(U), \llbracket \rrbracket)$ over a universe of discourse U . In this model, the language constructions are interpreted as follows:

1. A terminal x of either category N, NP, or VP is interpreted as a many valued relation whose value is the degree to which a subset A of the universe is $\llbracket x \rrbracket$. This is the relative sigma count of the subset A in $\llbracket x \rrbracket$, that is:

$$\star \llbracket x \rrbracket A := \text{Proportion}(A|\llbracket x \rrbracket)$$

2. A terminal x of category V is interpreted as a many valued relation whose value is the degree to which its image on a subset A of universe is a subset B of the universe, that is the relative sigma count of B in $\llbracket x \rrbracket(A)$:

$$\star \llbracket x \rrbracket(A, B) = \text{Proportion}(B|\llbracket x \rrbracket(A))$$

where $\llbracket x \rrbracket(A)$ is the application of $\llbracket x \rrbracket$ to A , resulting in a set $\sum_{i=1}^n \mu_i b$ where the subscripts of the μ 's vary over elements of fuzzy sets A and $\llbracket v \rrbracket$, so we have

$$\max_{a_i} \min(\mu_A(a_i), \mu_{\llbracket v \rrbracket}(a_i, b_i))$$

Here, μ_A and $\mu_{\llbracket v \rrbracket}$ are degrees of memberships of elements of fuzzy sets A and $\llbracket v \rrbracket$, respectively.

Using the above interpretation, a quantified sentence s gets a degree of truth $r \in [0, 1]$ as its semantics iff $\llbracket s \rrbracket = r$ in $(\mathbf{V-Rel}, P(U), \llbracket \rrbracket)$. Using this definition, we compute the semantics of the sentence ‘‘several cats sleep’’ with a fuzzy quantifier at subject position becomes as follows:

$$\max_{(A,B)} \min \left(\star \llbracket \text{cats} \rrbracket A, \star \llbracket \text{sleep} \rrbracket B, A \llbracket \text{several} \rrbracket A \cap B \right)$$

This formula will get a maximal value for $A = \llbracket \text{cats} \rrbracket$, $B = \llbracket \text{sleep} \rrbracket$ and when assuming that $\prod_{\text{Proportion}(A \cap B|A)} = \text{several}$, in which case the value of semantics becomes as follows:

$$\llbracket \text{several} \rrbracket \left[\frac{\sum \text{Count}(\llbracket \text{cats} \rrbracket \cap \llbracket \text{sleep} \rrbracket)}{\sum \text{Count}(\llbracket \text{cats} \rrbracket)} \right]$$

To compute this concretely, suppose that the fuzzy sets $\llbracket \text{cats} \rrbracket$ and $\llbracket \text{sleep} \rrbracket$ are defined as follows:

$$\llbracket \text{cats} \rrbracket = 0.2c_1 + 0.3c_2 + 0.8c_3$$

$$\llbracket \text{sleep} \rrbracket = 0.5c_1 + 0.4c_2 + 0.4c_3$$

Then the value for ‘‘several cats sleep’’ will be

$$\begin{aligned} & \llbracket \text{several} \rrbracket \left[\frac{\sum \text{Count}(0.2c_1 + 0.3c_2 + 0.4c_3)}{0.2c_1 + 0.3c_2 + 0.8c_3} \right] \\ &= \llbracket \text{several} \rrbracket \left[\frac{0.9}{1.3} \right] \end{aligned}$$

Suppose that the possibility distribution $\llbracket \text{several} \rrbracket$ will map low values to low values and very high values to low values, but intermediate values would be mapped to a high number as they still represent ‘‘several’’. Thus the proportion $\frac{0.9}{1.3}$, which is a high number, will evaluate to a high number. Thus the many valued relation of this statement will be high (a number close to 1). For examples of possibility distributions of some other fuzzy quantifiers, see [17].

Table 2: Degrees of Truth of the Non Quantified Sentences of the Toy Data Set

Entry 1	Entry 2	Deg.
people strike	group attacks	[0.63, 0.42]
notice advertises	sign announces	[0.57, 0.6]
clarify rule	explain process	[0.54, 0.47]
recommend development	suggest improvement	[0.69, 0.59]
people clarify rule	group explain process	[0.50, 0.34]
corporation recommend development	firm suggest improvement	[0.60, 0.49]
office arrange task	staff organize work	[0.61, 0.62]
editor threatens	application predicts	[0.36, 0.49]
progress reduces	development replaces	[0.69, 0.61]
confirm number	approve performance	[0.47, 0.67]
editor threatens man	application predicts number	[0.42, 0.36]
man recall time	firm cancel term	[0.65, 0.11]

The semantics of a sentence “Mice eat several plants” with a fuzzy quantifier at object place is computed as follows. Suppose we have fuzzy sets

$$\begin{aligned} \llbracket mice \rrbracket &= 0.7c_1 + 0.6c_2 + 0.2c_3 \\ \llbracket eat \rrbracket &= 0.5(c_1, c_1) + 0.8(c_1, c_3) + 0.2(c_2, c_1) \\ &\quad + 0.3(c_2, c_3) + 0.9(c_3, c_3) \\ \llbracket plants \rrbracket &= 0.2c_1 + 0.3c_2 + 0.6c_3 \end{aligned}$$

Then the semantics we get is

$$\llbracket several \rrbracket \left[\frac{\Sigma Count(\llbracket eat \rrbracket(\llbracket mice \rrbracket) \cap \llbracket plants \rrbracket)}{\Sigma Count(\llbracket plants \rrbracket)} \right]$$

The application of the verb to its subject gives

$$\llbracket eat \rrbracket(\llbracket mice \rrbracket) = 0.5c_1 + 0.7c_3$$

As a result, the whole expression now evaluates to

$$\begin{aligned} \llbracket several \rrbracket \left[\frac{\Sigma Count(0.2c_1 + 0.6c_3)}{\Sigma Count(0.2c_1 + 0.3c_2 + 0.6c_3)} \right] \\ = \llbracket several \rrbracket \left[\frac{0.8}{1.1} \right] \end{aligned}$$

This gives another relatively high value for the many valued semantics of this sentence, as $Proportion(\llbracket eat \rrbracket(\llbracket mice \rrbracket) | \llbracket plants \rrbracket)$ certainly indicates a case of “several” mice eating plants.

Finally, for the case where we have fuzzy quantifiers at both subject and object places, e.g. in the sentence “Several mice eat most plants”, a semantics is computed as follows. Given that the fuzzy sets representing mice and plants are as before and taking the same fuzzy relation for $\llbracket eat \rrbracket$, we compute the meaning of this sentence. Suppose further that $\llbracket most \rrbracket$ is a possibility distribution that assigns the value 0 to numbers below 0.5, and gradually increasing the value for numbers from 0.5 to 1. In

this case, first, we compute the application of the quantifiers to their respective noun phrases:

$$\begin{aligned} \llbracket several \rrbracket \llbracket \llbracket mice \rrbracket \rrbracket &= \\ \arg \max_B \left(\llbracket several \rrbracket \left[\frac{\Sigma Count(\llbracket mice \rrbracket \cap B)}{\Sigma Count(\llbracket mice \rrbracket)} \right] \right) \end{aligned}$$

If we assume that “several” has the highest value for 0.4, then it would for instance assign to the set $0.4 \llbracket mice \rrbracket$ the value $\sum_i 0.4 \mu_i u_i$ for $\mu_i u_i$ in $\llbracket mice \rrbracket$. The second application gives

$$\begin{aligned} \llbracket most \rrbracket \llbracket \llbracket plants \rrbracket \rrbracket &= \\ \arg \max_A \left(\llbracket most \rrbracket \left[\frac{A \cap \llbracket plants \rrbracket}{\Sigma Count(\llbracket plants \rrbracket)} \right] \right) \end{aligned}$$

This will set $A = \llbracket plants \rrbracket$, given that 1 has the highest probability of being “most”.

The value of the whole sentence will be the verb applied to the quantified subject and object, hence we obtain

$$\begin{aligned} \llbracket eat \rrbracket \left[\llbracket several \rrbracket \llbracket \llbracket mice \rrbracket \rrbracket, \llbracket most \rrbracket \llbracket \llbracket plants \rrbracket \rrbracket \right] \\ = \llbracket eat \rrbracket \left[0.4 \llbracket mice \rrbracket, \llbracket plants \rrbracket \right] \\ = \max_{a,b} \min(\mu_{0.4 \llbracket mice \rrbracket}(a), \mu_{\llbracket eat \rrbracket}(a, b), \mu_{\llbracket plants \rrbracket}(b)) \\ = \max \left(\min(0.28, 0.5, 0.2), \min(0.28, 0.8, 0.6), \right. \\ \quad \min(0.24, 0.2, 0.2), \min(0.24, 0.3, 0.6), \\ \quad \left. \min(0.08, 0.9, 0.6) \right) \\ = \max(0.2, 0.28, 0.2, 0.24, 0.08) = 0.28 \end{aligned}$$

That is, the extent to which several mice eat most plants is 28%.

5 From Many Valued Relations to Vectors Spaces and Linear Maps

By transferring our natural language semantics of quantified sentences from sets and relations to

Table 3: Degrees of Truth of the Quantified Sentences of the Toy Data Set

Entry 1	Entry 2	Deg.
all people strike	several groups attacks	[0.8, 0.8]
all noticesadvertise	many signs announce	[0.8, 0.8]
clarify several rules	explain some processes	[0.8, 0.8]
recommend many developments	suggest several improvements	[0.8, 0.8]
all people clarify rule	several groups explain process	[0.2, 0.8]
all corporations recommend development	many firms suggest improvement	[0.8, 0.2]
several offices arrange task	some staff organize work	[0.8, 0.8]
few editors threaten	all applications predict	[0.8, 0.2]
many progresses reduce	all developments replace	[0.8, 0.2]
confirm several numbers	approve few performances	[0.8, 0.2]
few editors threaten man	all applications predict number	[0.8, 0.2]
few men recall time	many firms cancel term	[0.2, 0.2]

sets and many valued relations, we generalised our Boolean-valued true-false semantics to a many valued semantics with degrees of truth from the unit interval $[0, 1]$. Transferring sets and many valued relations to vector spaces and linear maps enables us to compute the meaning of our sentences via quantitative reasoning on the statistical data provided in distributional semantics. A distributional vector for a target word w is seen as a fuzzy set whose degrees of membership are the degrees of co-occurrences of w with a set of context words c , or the degrees of contextual relevance of w to c , or other similar readings. In this section we use this interpretation and implement the formulae for obtaining the degrees of truth of fuzzy generalised quantifiers on vectors obtained from the combined UKWac/Wackypedia corpus [16], extracted using normalised co-occurrence counts, to ensure that the vectors indeed represent fuzzy sets. As Zadeh only provides semantics for a set of two atomic quantifiers without invoking grammatical compositionality, we use the many valued semantics of the previous section to compute the semantics of our quantified sentences compositionally.

We start with the sentence entailment dataset of [13] and later combine it with the entailed generalised quantifiers of [1]. The former dataset is a small dataset of 12 pairs of sentences, classified in three bands: high entailment, medium entailment, and low entailment. Each band is annotated with human judgements with a number in the range 1-7, representing the degree of entailments between each pair of sentences. The bands of entailment are decided upon based on the averages of the annotations. For the purpose of this paper, we only work with clear (non-)entailments and thus only present data for the high and low bands. In the high band, both subjects/objects and

verbs/verb phrases/intransitive verbs entail each other. These got an average annotation of 4 and above. In the low band are the non-entailing entries, i.e. neither of the subjects/objects or verbs/verb phrases/intransitive verbs entail each other. These got an average annotation of 2 or under. The entries of the dataset and their annotated degrees of entailment are given in table 1.

We implement our fuzzy semantics on the sentences of each entry and obtain degrees of truth. These degrees are given in table 2, with a horizontal line separating the high and low bands of entailment. A list of entailing and non-entailing quantifiers were provided in [1]. The list is a mixture of logical and generalised quantifiers. Entailment for logical quantifiers, e.g. ‘all’ and ‘some’, is clear and so we drop these as well as numerical quantifiers such as ‘both’ which are not fuzzy. The case of generalised quantifiers is the interesting one. Here, the degrees of truth of the non quantified sentences will change after applying the quantifiers to them, as we saw in the examples of the previous section for the quantifier ‘several’. Thus we work with the generalised quantifier subset of the data set of [1]. These are as follows:

Entailing	Non-Entailing
all, several	several, all
all, many	many, all
several, some	several, few
many, several	few, all
	few, many

We use the entailing ones to strengthen the entailments of the high band and the non-entailing ones to weaken the entailments of the low band. This provides us with the following dataset, the positive entailments are separated from the negative ones by a bar in the table. Hardcoding the meaning of generalised quantifiers, in the same way as we did in the previous section, provides

Table 4: Degrees of Entailment for Sentences Before and After Quantification

Entry 1	Entry 2	Deg.1	Deg.2
people strike	group attacks	-	+
notice advertises	sign announces	+	+
clarify rule	explain process	-	+
recommend development	suggest improvement	-	+
people clarify rule	group explain process	-	+
corporation recommend development	firm suggest improvement	-	-
office arrange task	staff organize work	+	+
editor threatens	application predicts	+	-
progress reduces	development replaces	-	-
confirm number	approve performance	+	-
editor threatens man	application predicts number	-	-
man recall time	firm cancel term	-	+

us with new degrees of truth for each pair of sentences of our dataset, which we give in table 3.

The intuitions for the hard coding of quantifiers are obtained following [17] and are stipulated in the table below:

quantifier	hard coding
all	high \rightarrow 0.8, low \rightarrow 0.2
some	very low \rightarrow 0.2, the rest \rightarrow 0.8
several	low \rightarrow 0.2, high \rightarrow 0.2, interm. \rightarrow 0.8
many	low \rightarrow 0.2, high \rightarrow 0.8, interm. \rightarrow 0.2
few	low \rightarrow 0.8, high \rightarrow 0.2, interm. \rightarrow 0.2

In this table, the quantifier ‘all’ expectedly sends high numbers to high numbers and low numbers to low numbers, whereas e.g. ‘several’ maps low values to low values, very high values to low values, but intermediate values to high values. In order to be able to compare the resulting numbers in a uniform way we apply the convention that ‘low’ is 0.2, ‘high’ is 0.8 and intermediate is 0.5.

Given a pair of degrees of truth, we now compute an entailment and a degree for it using the definition of fuzzy entailment [17, 8]. A fuzzy proposition p entails another fuzzy proposition q iff q is less specific than p . For p and q two fuzzy sets, this is defined to be the point wise ordering between their possibilistic distributions. After a proportion is computed on the fuzzy sets, the ordering becomes the ordering between their computed degrees of truth, i.e. on $\Pi_{Proportion(B|A)}$ for the proportion of elements of B in A . A degree of entailment is computed from a pair of degrees of truth, by subtracting them. Whenever this number is positive, we mark it with a + sign and whenever it is negative we mark it with a - sign. If the result is 0, it must have come from a case where the degrees of truth of each entry of the pair is the same. Since the ordering is \leq and not strict, these cases stand for as a full entailment and given a + sign. The degrees of entailment of the entries of

our dataset thus becomes as in the Deg. 1 column of table 4, the two high and low bands are combined in one table with a bar separating them.

This means that out of the seven cases of positive entailment, only two (the ones marked with a +) are predicted correctly and out of the five cases of negative entailment, only three (the ones marked with -) re predicted correctly. However, after applying the generalised quantifiers to them, the results improve, as shown in the Deg. 2 column of table 4. Here, the + signs increase from two to six, predicting all but one case correctly. The number of the - signs also increase by 1, so the model is predicting that 4 out of the 5 entries do not entail each other correctly.

6 Conclusion and Future Work

We showed how the compositional semantics of the generalised quantifiers of natural language extends from a binary setting to a many valued setting and how this latter can be used to model fuzzy generalised quantifier. In a quest to relate these developments to large scale data learnt from distributional semantics, we interpreted vectors as fuzzy sets and computed degrees of truth for the sentences of a toy dataset. We then extended the dataset with quantifiers and showed how these computations can be used for an inference task.

One essential piece of work is experimenting with the model on main stream inference datasets such as SNLI and Fracas. We mainly chose to work with this small dataset since the details of its high and low bands and the annotations for them where published and we could use them to provide a detailed case by case analysis. A theoretical direction is to use the logic of fuzzy sets, e.g. that of [12], to develop a logic for distributional

data. Quantifiers are known to impose contextual restrictions on their domains, e.g. in the donkey sentences. In previous work [15, 14] authors have shown how compositional distributional semantics deals with these issues. Finding out the added value of working in a fuzzy setting for such examples remains to be worked out.

Finally, we are aware of the tension that exists between the membership values of fuzzy sets and the probabilities that come from the normalised distributional vector representations. In their simplest forms, the latter probabilities are log likelihood estimates resulting from co-occurrence counts of a corpus of text. Fuzzy values, however, are obtained from distribution of the individuals in the domain of vague predicates in a model. Relating these two should be possible by machine learning and alongside recent work on Bayesian inference semantics [5], unified functional distributional models [7], and distributional model theoretic approach [11, 3, 4].

References

- [1] Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do, and Chung-chieh Shan. Entailment above the word level in distributional semantics. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 23–32, Avignon, France, 2012. Association for Computational Linguistics.
- [2] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219, 1981.
- [3] I. Beltagy, Stephen Roller, Pengxiang Cheng, Katrin Erk, and Raymond J. Mooney. Representing meaning with a combination of logical and distributional models. *Computational Linguistics*, 42:763–808, 2016.
- [4] Islam Beltagy, Cuong Chau, Gemma Boleda, Dan Garrette, Katrin Erk, and Raymond Mooney. Montague meets Markov: Deep semantics with probabilistic logical form. In *Second Joint Conference on Lexical and Computational Semantics*, pages 11–21, Atlanta, Georgia, USA, 2013. Association for Computational Linguistics.
- [5] Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, Shalom Lappin, and Aleksandre Maskharashvili. Bayesian inference semantics: A modelling system and a test suite. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019)*, pages 263–272, Minneapolis, Minnesota, 2019. Association for Computational Linguistics.
- [6] B. Coecke, M. Sadrzadeh, and S. Clark. Mathematical Foundations for Distributed Compositional Model of Meaning. *Lambek Festschrift. Linguistic Analysis*, 36:345–384, 2010.
- [7] Guy Emerson and Ann Copestake. Functional distributional semantics. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 40–52, Berlin, Germany, 2016. Association for Computational Linguistics.
- [8] L. Godo and L. Valverde. Entailment and inference in fuzzy logic using fuzzy preorders. In *1992 Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 587–594, San Diego, CA, USA, 1992. IEEE.
- [9] Edward Grefenstette and Mehrnoosh Sadrzadeh. Concrete models and empirical evaluations for the categorical compositional distributional model of meaning. *Computational Linguistics*, 41:71–118, 2015.
- [10] Jules Hedges and Mehrnoosh Sadrzadeh. A generalised quantifier theory of natural language in categorical compositional distributional semantics with bialgebras, 2016.
- [11] Aurélie Herbelot and Eva Maria Vecchi. Building a shared world: mapping distributional to model-theoretic semantic spaces. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 22–32, Lisbon, Portugal, 2015. Association for Computational Linguistics.
- [12] Vilém Novák. *Fuzzy Sets in Natural Language Processing*, volume 165 of *The Springer International Series in Engineering and Computer Science*, pages 185–200. Springer, 1992.
- [13] M. Sadrzadeh, D. Kartsaklis, and E. Balkır. Sentence entailment in compositional distributional semantics. *Ann Math Artif Intell*, 82:189–218, 2018.
- [14] Mehrnoosh Sadrzadeh. Quantifier scope in categorical compositional distributional semantics. In *Proceedings of the Workshop on Semantic Spaces at the Intersection of NLP, Physics and Cognitive Science*, volume 221 of *EPTCS*, pages 49–57, 2016.
- [15] G. Wijnholds. A proof-theoretic approach to scope ambiguity in compositional vector space models. *Journal of Language Modelling*, 6:261–286, 2018.
- [16] Gijs Wijnholds and Mehrnoosh Sadrzadeh. Evaluating composition models for verb phrase elliptical sentence embeddings. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 2019.
- [17] Lotfi A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computers & Mathematics with Applications*, 9(1):149 – 184, 1983.