Linguists Who Use Probabilistic Models Love Them: Quantification in Functional Distributional Semantics

Guy Emerson
Department of Computer Science and Technology
University of Cambridge
gete2@cam.ac.uk

Abstract

Functional Distributional Semantics provides a computationally tractable framework for learning truth-conditional semantics from a corpus. Previous work in this framework has provided a probabilistic version of first-order logic, recasting quantification as Bayesian inference. In this paper, I show how the previous formulation gives trivial truth values when a precise quantifier is used with vague predicates. I propose an improved account, avoiding this problem by treating a vague predicate as a distribution over precise predicates. I connect this account to recent work in the Rational Speech Acts framework on modelling generic quantification, and I extend this to modelling donkey sentences. Finally, I explain how the generic quantifier can be both pragmatically complex and yet computationally simpler than precise quantifiers.

1 Introduction

Model-theoretic semantics defines meaning in terms of *truth*, relative to *model structures*. In the simplest case, a model structure consists of a set of *individuals* (also called *entities*). The meaning of a content word is a *predicate*, formalised as a *truth-conditional function* which maps individuals to *truth values* (either *truth* or *falsehood*). Because of this precisely defined notion of truth, model theory naturally supports logic, and has become a prominent approach to formal semantics. For detailed expositions, see: Cann (1993); Allan (2001); Kamp and Reyle (2013).

Mainstream approaches to distributional semantics represent the meaning of a word as a vector (for example: Turney and Pantel, 2010; Mikolov et al., 2013; for an overview, see: Emerson, 2020b). In contrast, Functional Distributional Semantics represents the meaning of a word as a truth-conditional function (Emerson and Copestake, 2016; Emerson, 2018). It is therefore a promising framework for automatically learning truth-conditional semantics from large datasets.

In previous work (Emerson and Copestake, 2017b, §3.5, henceforth E&C), I sketched how this approach can be extended with a probabilistic version of first-order logic, where quantifiers are interpreted in terms of conditional probabilities. I summarise this approach in §2 and §3.

There are four main contributions of this paper. In §4.1, I first point out a problem with my previous approach. Quantifiers like *every* and *some* are treated as precise, but predicates are vague. This leads to trivial truth values, with *every* trivially false, and *some* trivially true.

Secondly, I show in §4.2–4.4 how this problem can fixed by treating a vague predicate as a distribution over precise predicates.

Thirdly, in §5 I look at vague quantifiers and generic sentences, which present a challenge for classical (non-probabilistic) theories. I build on Tessler and Goodman (2019)'s account of generics using Rational Speech Acts, a Bayesian approach to pragmatics (Frank and Goodman, 2012). I show how generic quantification is computationally simpler than classical quantification, consistent with evidence that generics are a "default" mode of processing (for example: Leslie, 2008; Gelman et al., 2015).

Finally, I show in §6 how this probabilistic approach can provide an account of donkey sentences, another challenge for classical theories. In particular, I consider generic donkey sentences, which are doubly challenging, and which provide counter-examples to the claim that donkey pronouns are associated with universal quantifiers.

Taking the above together, in this paper I show how a probabilistic first-order logic can be associated with a neural network model for distributional semantics, in a way that sheds light on long-standing problems in formal semantics.

Proceedings of the Probability and Meaning Conference (PaM 2020), pages 41–52

2 Generalised Quantifiers

Partee (2012) recounts how quantifiers have played an important role in the development of model-theoretic semantics, seeing a major breakthrough with Montague (1973)'s work, and culminating in the theory of *generalised quantifiers* (Barwise and Cooper, 1981; Van Benthem, 1984).

Ultimately, model theory requires quantifiers to give truth values to propositions. An example of a logical proposition is given in Fig. 1, with a quantifier for each logical variable. This also assumes a neo-Davidsonian approach to event semantics (Davidson, 1967; Parsons, 1990).

Equivalently, we can represent a logical proposition as a *scope tree*, as in Fig. 2. The truth of the scope tree can be calculated by working bottom-up through the tree. The leaves of the tree are logical expressions with free variables. They can be assigned truth values if each variable is fixed as an individual in the model structure. To assign a truth value to the whole proposition, we work up through the tree, quantifying the variables one at at time. Once we reach the root, all variables have been quantified, and we are left with a truth value.

Each quantifier is a non-terminal node with two children – its *restriction* (on the left) and its *body* (on the right). It quantifies exactly one variable, called its *bound variable*. Each node also has *free variables*. For each leaf, its free variables are exactly the variables appearing in the logical expression. For each quantifier, its free variables are the union of the free variables of its restriction and body, minus its own bound variable. For a well-formed scope tree, the root has no free variables. Each node in the tree defines a truth value, given a fixed value for each free variable.

The truth value for a quantifier node is defined based on its restriction and body. Given values for the quantifier's free variables, the restriction and body only depend on the quantifier's bound variable. The restriction and body therefore each define a set of individuals in the model structure – the individuals for which the restriction is true, and the individuals for which the body is true. We can write these as $\mathcal{R}(v)$ and $\mathcal{B}(v)$, respectively, where v denotes the values of all free variables.

Generalised quantifier theory says that a quantifier's truth value only depends on two quantities: the cardinality of the restriction $|\mathcal{R}(v)|$, and the cardinality of the intersection of the restriction and body $|\mathcal{R}(v) \cap \mathcal{B}(v)|$. Table 1 gives examples.

$$\forall x \; \textit{picture}(x) \rightarrow$$
$$\exists z \exists y \; \textit{tell}(y) \wedge \textit{story}(z) \wedge \text{ARG1}(y,x) \wedge \text{ARG2}(y,z)$$

Figure 1: A first-order logical proposition, representing the most likely reading of *Every picture tells a story*. Scope ambiguity is not discussed in this paper.

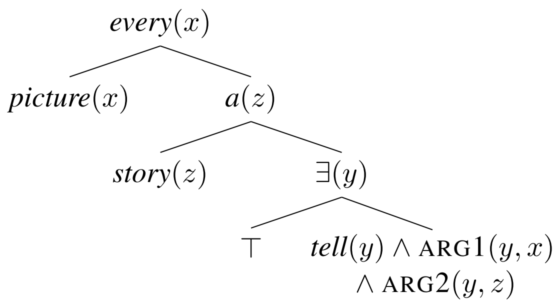


Figure 2: A scope tree, equivalent to Fig. 1 above. Each non-terminal node is a quantifier, with its bound variable in brackets. Its left child is its restriction, and its right child its body.

Quantifier	Condition
some	$\lvert\mathcal{R}(v) \cap \mathcal{B}(v)\rvert > 0$
every	$\lvert\mathcal{R}(v) \cap \mathcal{B}(v)\rvert = \lvert\mathcal{R}(v)\rvert$
no	$\lvert\mathcal{R}(v) \cap \mathcal{B}(v)\rvert = 0$
most	$\lvert\mathcal{R}(v) \cap \mathcal{B}(v)\rvert > \frac{1}{2}\lvert\mathcal{R}(v)\rvert$

Table 1: Classical truth conditions for precise quantifiers, in generalised quantifier theory.

3 Generalised Quantifiers in Functional Distributional Semantics

Functional Distributional Semantics defines a probabilistic graphical model for distributional semantics. Importantly (from the point of view of formal semantics), this graphical model incorporates a probabilistic version of model theory.

This is illustrated in Fig. 3. The top row defines a distribution over situations, each situation being an event with two participants.[1] This generalises a model structure comprising a *set* of situations, as in classical situation semantics (Barwise and Perry, 1983). Each individual is represented by a *pixie*, a point in a high-dimensional space, which represents the features of the individual. Two individuals could be represented by the same pixie, and the space of pixies can be seen as a conceptual space in the sense of Gärdenfors (2000, 2014).

[1]For situations with different structures (multiple events or different numbers of participants), we can define a family of such graphical models. Structuring the graphical model in terms of semantic roles makes the simplifying assumption that situation structure is isomorphic to a semantic dependency graph such as DMRS (Copestake et al., 2005; Copestake, 2009). In the general case, the assumption fails. For example, the ARG3 of *sell* corresponds to the ARG1 of *buy*.

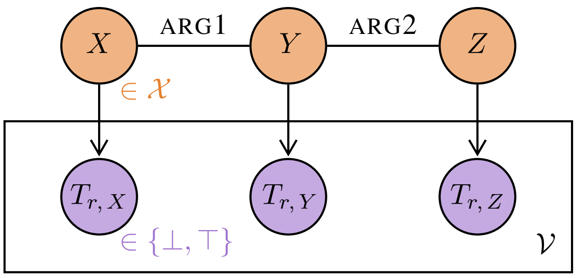


Figure 3: Probabilistic model theory, as formalised in Functional Distributional Semantics. Each node is a random variable. The plate (box in bottom row) denotes repetition of nodes.
Top row: pixie-valued random variables X, Y, Z together represent a situation composed of three individuals. They are jointly distributed according to the semantic roles ARG1 and ARG2. Their joint distribution can be seen as a probabilistic model structure.
Bottom row: each predicate r in the vocabulary $\mathcal{V}$ has a probabilistic truth-conditional function, which can be applied to each individual. This gives a truth-valued random variable for each individual for each predicate.

The bottom row of the graphical model defines a distribution over truth values, so that each predicate has some probability of being true of each individual. Each predicate can therefore be seen as a probabilistic truth-conditional function.

In this paper, I will not discuss learning such a model (for an up-to-date approach, see: Emerson, 2020a). Instead, the focus is on how we can manipulate a trained model, to move from single predicates to complex propositions.

In previous work (E&C), I sketched an account of quantification. The idea is to follow generalised quantifier theory, but with a truth-valued random variable for each node in the scope tree. Similarly to the classical case, the distributions for these nodes are defined bottom-up through the tree.

In the classical theory, we only need to know the cardinalities $|\mathcal{R}(v)|$ and $|\mathcal{R}(v) \cap \mathcal{B}(v)|$. In fact, all the conditions in Table 1 can be expressed in terms of the ratio $\frac{|\mathcal{R}(v) \cap \mathcal{B}(v)|}{|\mathcal{R}(v)|}$. It therefore makes sense to consider the conditional probability $\mathbb{P}(b \mid r, v)$, because this uses the same ratio, as shown in (1).[2]

$$\mathbb{P}(b \mid r, v) = \frac{\mathbb{P}(r, b \mid v)}{\mathbb{P}(r \mid v)} \qquad (1)$$

More precisely, B and R are truth-valued random variables for the body and restriction, and V is a tuple-of-pixies-valued random variable, with one pixie for each free variable. Intuitively, the truth of a quantified expression depends on how likely B is to be true, given that R is true.[3]

[2] I use uppercase for random variables, lowercase for values. I abbreviate $\mathbb{P}(X = x)$ as $\mathbb{P}(x)$, and $\mathbb{P}(T = \top)$ as $\mathbb{P}(t)$. For example, $\mathbb{P}(b \mid r, v)$ means $\mathbb{P}(B = \top \mid R = \top, V = v)$.

Quantifier	Condition
some	$\mathbb{P}(b \mid r, v) > 0$
every	$\mathbb{P}(b \mid r, v) = 1$
no	$\mathbb{P}(b \mid r, v) = 0$
most	$\mathbb{P}(b \mid r, v) > \frac{1}{2}$

Table 2: Truth conditions for precise quantifiers, in terms of the conditional probability of the body given the restriction (and given all free variables). These conditions mirror Table 1.

Truth conditions for quantifiers can be defined in terms of $\mathbb{P}(b \mid r, v)$, as shown in Table 2. For these precise quantifiers, the truth value is deterministic – if the condition in Table 2 holds, the quantifier's random variable Q has probability 1 of being true, otherwise it has probability 0. However, taking a probabilistic approach means that we can naturally model vague quantifiers like *few* and *many*. I did not give further details on this point in E&C, but I will expand on this in §5.

4 Quantification with Vague Predicates

Truth-conditional functions that give probabilities strictly between 0 and 1 are motivated for both practical and theoretical reasons. Practically, such a function can be implemented as a feedforward neural network with a final sigmoid unit (as used by E&C), whose output is never exactly 0 or 1. Theoretically, using intermediate probabilities of truth allows a natural account of vagueness (Goodman and Lassiter, 2015; Sutton, 2015, 2017).

However, as we will see in the following subsection, intermediate probabilities pose a problem for E&C's account of quantification.

4.1 Trivial Truth Values

Combining the conditions in Table 2 with vague predicates causes a problem, which can be illustrated with a simple example. Consider a model structure containing only a single individual, and consider only the single predicate *red*, which is true of this individual with probability p. Now consider the sentences (1) and (2).

[3] This would not seem to cover so-called *cardinal quantifiers* like *one* and *two*. Under Link (1983)'s lattice-theoretic approach, a model structure contains plural individuals, so numbers can be treated as normal predicates like adjectives.

(1) Everything is red.

(2) Something is red.

The body of each quantifier is simply the predicate *red*. For simplicity, we can assume that *everything* and *something* put no constraints on their restrictions. We need to calculate $\mathbb{P}(b \mid r, v)$. There are no free variables, and R is always true, so this is simply $\mathbb{P}(b)$. Because there is only one individual, this is simply the probability p.

This means that (1) is true iff $p = 1$, and (2) is false iff $p = 0$. However, we have seen above how predicates will never be true with probability exactly 0 or exactly 1. This means (1) is always false, and (2) is always true, even though we have assumed nothing about the individual!

4.2 Distributions over Precise Predicates

To avoid the problem in §4.1, we must only combine precise quantifiers with precise predicates (i.e. classical truth-conditional functions). To do this, we can view a vague predicate not as defining a probability of truth for each individual, but as defining a distribution over precise predicates. This induces a distribution for the quantifier.

Consider the example in §4.1. With probability p, *red* is a precise predicate that is true of the individual. In this case, both (1) and (2) are true. With probability $1-p$, *red* is a precise predicate that is false of the individual. In this case, both (1) and (2) are false. Combining these cases, both (1) and (2) are true with probability p, which has avoided trivial truth values.

Formalising a vague predicate as a distribution over precise predicates was also argued for by Lassiter (2011). It can be seen as an improved version of supervaluationism (Fine, 1975; Kamp, 1975; Keefe, 2000, chapter 7), which avoids the problem of higher-order vagueness, as shown by Lassiter.

4.3 Probabilistic Scope Trees

To generalise the account in §4.2 to arbitrary scope trees (see §4.4) and vague quantifiers (see §5), it is helpful to introduce a graphical notation for *probabilistic scope trees*, illustrated in Fig. 4. This makes the E&C account easier to visualise. The improved proposal in this paper modifies how the distribution for each truth value node is defined.

For a classical scope tree, the truth of a quantifier node depends on its free variables, and is defined in terms of the extensions of its restriction and body, in a way that removes the bound variable. For a probabilistic scope tree, the distribution for a quantifier node is conditionally dependent on its free variables, and is defined in terms of the distributions for its restriction and body, marginalising out the bound variable. The distributions at the leaves of the tree are defined by predicates, inducing a distribution for each quantifier node as we work up through the tree.

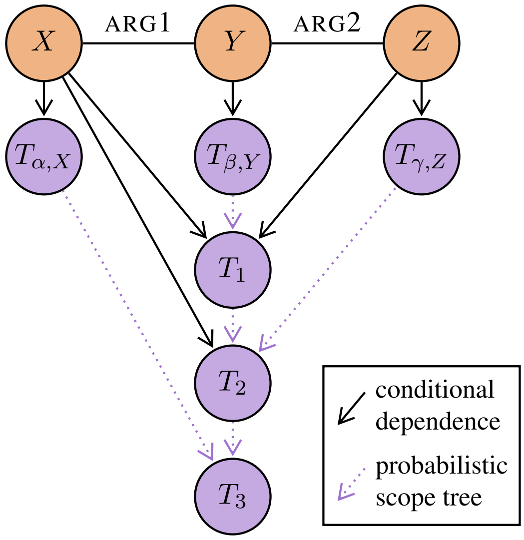


Figure 4: A probabilistic scope tree. T_1, T_2, T_3 correspond to non-terminal nodes in Fig. 2, going up through the tree. One random variable is marginalised out at a time, until T_3 is no longer dependent on any variables.

Fig. 4 corresponds to Fig. 2, if we set α, β, γ to be *picture*, *tell*, *story*. The distributions for $T_{\alpha,X}$, $T_{\beta,Y}$, $T_{\gamma,Z}$ are determined by the predicates. We have three quantifier nodes in the classical scope tree, and hence three additional truth value nodes in the probabilistic scope tree. We first define a distribution for T_1, which represents the $\exists(y)$ quantifier, and which depends on its free variables X and Z. It is true if, for situations involving the fixed pixies x and z, there is *nonzero* probability that they are the ARG1 and ARG2 of a telling-event pixie. Next, we define a distribution for T_2, which represents the $a(z)$ quantifier, and depends on the free variable X. It is true if, for situations involving the fixed pixie x and story pixie z, there is *nonzero* probability that T_1 is true. Finally, we define a distribution for T_3, which represents the *every*(x) quantifier, and has no free variables. It is true if, for situations involving a picture pixie X, we are *certain* that T_2 is true.

4.4 Probabilistic Scope Trees with Vague Predicates as Distributions

In this section I show how to define the quantifier nodes in §4.3 so that they are nontrivial.

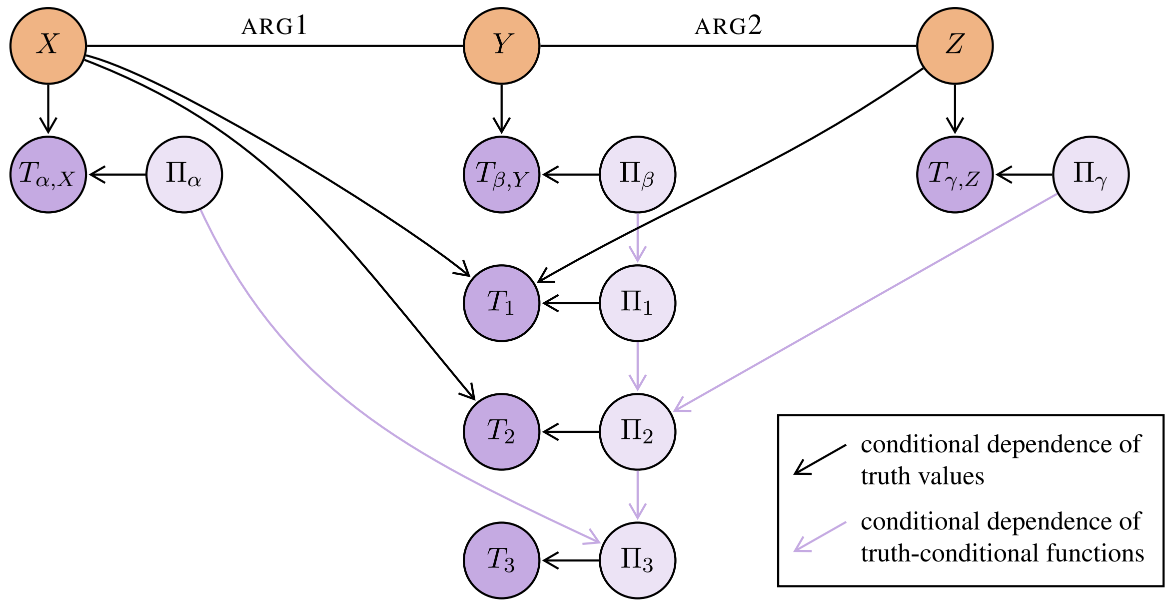


Figure 5: The probabilistic scope tree in Fig. 4, explicitly showing random variables over precise functions.

To explicitly represent each vague truth-conditional function as a random variable over precise functions, we need to add a function node for each truth value node in the graphical model. For example, this transforms Fig. 4 into Fig. 5.

For a truth value node T that is a leaf of the scope tree (the second row of Fig. 5), the distribution $\mathbb{P}(t)$ over truth values follows the description in §4.2. A precise predicate $\pi : \mathcal{X} \to \{\top, \bot\}$ maps pixies to truth values. Given π and a pixie x, the distribution for T is deterministic: $T = \pi(x)$ with probability 1. A distribution Π over precise predicates π defines a vague predicate ψ, by marginalising out this distribution:[4]

$$\mathbb{P}(t \,|\, x) = \psi(x) = \mathbb{E}_\pi\left[\pi(x)\right] \tag{2}$$

More generally, a truth value node Q is dependent on its free variables V. We can represent this in terms of a precise function $\pi : \mathcal{X}^n \to \{\top, \bot\}$, where n is the number of free variables. Given values v for the free variables, a distribution Π over precise predicates π defines a vague predicate ψ, by marginalising out this distribution:

$$\mathbb{P}(q \,|\, v) = \psi(v) = \mathbb{E}_\pi\left[\pi(v)\right] \tag{3}$$

What remains to be shown is that the E&C account of quantification (in §3) can be adapted so that a quantifier's distribution Π_Q over precise functions π_Q can be defined in terms of its restriction function π_R and body function π_B. This can be seen as probabilistic semantic composition: the aim is to combine two truth-conditional functions to produce a distribution over truth-conditional functions. This is illustrated by the nodes Π_1, Π_2, Π_3 in Fig. 5, which are conditionally dependent on other function nodes (indicated by the purple edges), forming a probabilistic scope tree.

Expanding (3) so it is dependent on the restriction and body functions, we have (4). The aim is now to re-write the distribution for Q, using an adapted version of E&C, in order to derive π_Q in terms of π_R and π_B. As explained in §3, the E&C account defines Q using the conditional probability $\mathbb{P}(b \,|\, r, v)$. More precisely, $\mathbb{P}(q \,|\, v) = f_Q(\mathbb{P}(b \,|\, r, v))$ for some f_Q, such as those defined by Table 2. With vague functions now considered as distributions over precise functions, the conditional probability must be amended to $\mathbb{P}(b \,|\, r, v, \pi_R, \pi_B)$, as in (5), given precise functions π_R and π_B for the restriction and body. This can be re-written as a ratio of probabilities (corresponding to the classical sets), summing over possible values for the bound variable(s) U, as in (6). We can factorise out the distribution for U, according to the conditional dependence structure (illustrated in Fig. 5), as in (7). Finally, we can express R and B in terms of the functions π_R and π_B, and write the sum as an expectation, as in (8). Note that π_R and π_B take $u \cup v$ as an argument – by definition of a scope tree, if we combine a quantifier's bound and free variables, we get the free variables of its restriction and body. I have written $u \cup v$ rather than $\{u\} \cup v$, to leave open the possi-

[4] I write expectations with a subscript to indicate the random variable being marginalised out. To write the expectation in (2) explicitly as a sum: $\mathbb{E}_\pi\left[\pi(x)\right] = \sum_\pi [\pi(x)\mathbb{P}(\pi)]$.

bility that the quantifier has more than one bound variable, which will be relevant in §5.

$$\begin{aligned}
\mathbb{P}\left(q \mid v, \pi_R, \pi_B\right) &= \mathbb{E}_{\pi_Q \mid \pi_R, \pi_B}\left[\pi_Q(v)\right] && (4)\\
&= f_Q\left(\mathbb{P}\left(b \mid r, v, \pi_R, \pi_B\right)\right) && (5)\\
&= f_Q\left(\frac{\sum_u \mathbb{P}\left(b, r, u \mid v, \pi_R, \pi_B\right)}{\sum_u \mathbb{P}\left(r, u \mid v, \pi_R, \pi_B\right)}\right) && (6)\\
&= f_Q\left(\frac{\sum_u \mathbb{P}\left(u \mid v\right)\mathbb{P}\left(r, b \mid u, v, \pi_R, \pi_B\right)}{\sum_u \mathbb{P}\left(u \mid v\right)\mathbb{P}\left(r \mid u, v, \pi_R\right)}\right) && (7)\\
&= f_Q\left(\frac{\mathbb{E}_{u|v}\left[\pi_R(u \cup v)\, \pi_B(u \cup v)\right]}{\mathbb{E}_{u|v}\left[\pi_R(u \cup v)\right]}\right) && (8)
\end{aligned}$$

(8) gives a probability of truth, hence a vague function. Viewing it as a distribution over precise functions (as in §4.2), we finally have a definition of π_Q in terms of π_R and π_B. Concretely, π_Q returns truth iff (8) is above a threshold. A uniform distribution over thresholds in $[0, 1]$ gives a distribution over such functions.

Abbreviating the notation, we can write (9). A quantifier's truth-conditional function depends on the restriction and body functions, marginalising out the bound variable. The ratio of expectations mirrors the classical ratio of cardinalities.

$$\pi_Q \sim f_Q\left(\frac{\mathbb{E}_u\left[\pi_R \pi_B\right]}{\mathbb{E}_u\left[\pi_R\right]}\right) \qquad (9)$$

We can now recursively define functions for quantifier nodes, given functions in the leaves. We can therefore see Fig. 4 as an abbreviated notation for Fig. 5. The dotted edges do not indicate conditional dependence of *truth values*, but conditional dependence of *truth-conditional functions*.

5 Vague Quantifiers and Generics

While *some*, *every*, *no*, and *most* can be given precise truth conditions, other natural language quantifiers are vague. In particular, we can consider the terms *few* and *many*.[5]

Under a classical account (for example: Barwise and Cooper, 1981), *many* means that $\mathcal{R}(v) \cap \mathcal{B}(v)$ is large compared to $\mathcal{R}(v)$, but how large is underspecified; similarly, *few* means this ratio is small. The underspecification of a proportion can naturally be represented as a distribution. So, we can define the meaning of a vague generalised quantifier to be a function from $\mathbb{P}\left(b \mid r, v\right)$ to a probability of truth, as illustrated in Fig. 6.

[5] Partee (1988) surveys work suggesting that *few* and *many* are ambiguous between a vague cardinal reading and a vague proportional reading. As mentioned in Footnote 3, we can treat cardinals as predicates rather than quantifiers.

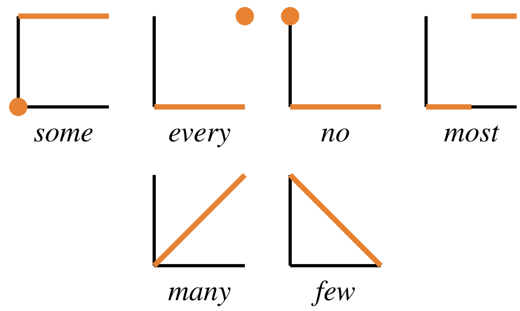


Figure 6: Probabilities of truth for various quantifiers. Each x-axis is $\mathbb{P}\left(b \mid r, v, \pi_R, \pi_B\right)$, and each y-axis is $\mathbb{P}\left(q \mid v, \pi_R, \pi_B\right)$, plotting the function f_Q in orange. All axes range from 0 to 1. Quantifiers in the bottom row are vague, requiring intermediate probabilities.

A particularly challenging case of natural language quantification involves *generic* sentences, such as: *dogs bark*, *ducks lay eggs*, and *mosquitoes carry malaria*. Generics are ubiquitous in natural language, but they are challenging for classical models, because the truth conditions seem to depend heavily on lexical semantics and on the context of use (for discussion, see: Carlson, 1977; Carlson and Pelletier, 1995; Leslie, 2008).

While it is tempting to treat generic quantification as underspecification of a precise quantifier (for example: Herbelot, 2010; Herbelot and Copestake, 2011), this is at odds with evidence that generics are easier for children to acquire than precise quantifiers (Hollander et al., 2002; Leslie, 2008; Gelman et al., 2015), and also easier for adults to process (Khemlani et al., 2007).

In contrast, Tessler and Goodman (2019) analyse generic sentences as being semantically simple, with the complexity coming down to pragmatic inference. They use Rational Speech Acts (RSA), a Bayesian approach to pragmatics (Frank and Goodman, 2012; Goodman and Frank, 2016). In this framework, literal truth is separated from pragmatic meaning. Communication is viewed as a game where a listener has a prior belief about a situation, and a speaker wants to update the listener's belief. Given a truth-conditional function, a *literal listener* updates their belief by conditioning on truth, ruling out situations for which the function returns false. A *pragmatic speaker* who observes a situation can choose an utterance which is informative for a literal listener – in particular, the utterance which maximises a literal listener's posterior probability for the observed situation. A *pragmatic listener* can update their belief by conditioning on a pragmatic speaker's utterance.

Tessler and Goodman's insight is that this inference of *pragmatic* meanings can account for the behaviour of generic sentences. The literal meaning of a generic can be simple (it is more likely to be true as the proportion increases), but the pragmatic meaning can have a rich dependence on the world knowledge encoded in the prior over situations. For example, *Mosquitoes carry malaria* does not mean that all mosquitoes do (in fact, many do not) but it can be informative for the listener: as most animals never carry malaria, even a small proportion is pragmatically relevant.

Building on this, we could model the generic quantifier by setting f_Q as the identity function (the same as *many* in Fig. 6). From (8), the probability of truth is then as shown in (10). However, marginalising out Π_R and Π_B is computationally expensive, as it requires summing over all possible functions. We can approximate this by reversing the order of the expectations, and so marginalising out Π_R and Π_B before U, as shown in (11), where ψ_R and ψ_B are vague functions. Evaluating a vague function is computationally simple.

$$\mathbb{E}_{\pi_R,\pi_B}\left[\frac{\mathbb{E}_{u|v}\left[\pi_R(u\cup v)\,\pi_B(u\cup v)\right]}{\mathbb{E}_{u|v}\left[\pi_R(u\cup v)\right]}\right] \tag{10}$$

$$\approx \frac{\mathbb{E}_{u|v}\left[\psi_R(u\cup v)\,\psi_B(u\cup v)\right]}{\mathbb{E}_{u|v}\left[\psi_R(u\cup v)\right]} \tag{11}$$

Abbreviating this, similarly to (9), we can write:

$$\psi_Q = \frac{\mathbb{E}_u\left[\psi_R\psi_B\right]}{\mathbb{E}_u\left[\psi_R\right]} \tag{12}$$

For precise quantifiers, using vague functions gives trivial truth values (discussed in §4.1), but for generics, (10) and (11) give similar probabilities of truth. To put it another way, a vague quantifier doesn't need precise functions. Modelling generics with (10) was driven by the intuition that generics are vague but semantically simple. The alternative in (11) is even simpler, because we only need to calculate $\mathbb{E}_{u|v}$ once in total, rather than once for each possible π_R and π_B. This would make generics computationally simpler than other quantifiers, consistent with the evidence that they are easier to acquire and to process.

In fact, (11) takes us back to E&C's conditional probability, as shown in (13).

$$\psi_Q(v) = \frac{\sum_u \mathbb{P}(u\,|\,v)\,\mathbb{P}(r\,|\,u,v)\,\mathbb{P}(b\,|\,u,v)}{\sum_u \mathbb{P}(u\,|\,v)\,\mathbb{P}(r\,|\,u,v)} = \mathbb{P}(b\,|\,r,v) \tag{13}$$

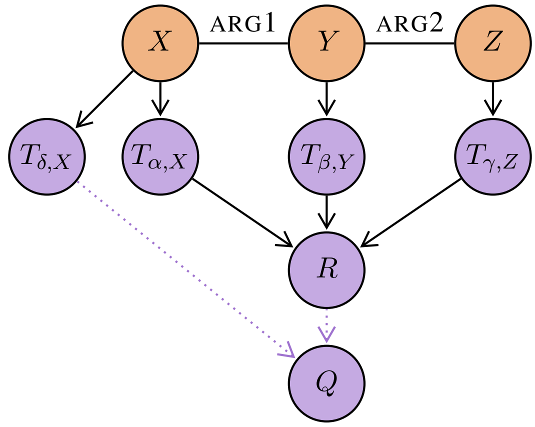


Figure 7: Emerson and Copestake (2017a)'s logical inference, re-analysed as generic quantification. R is the restriction, the logical conjunction of $T_{\alpha,X}$, $T_{\beta,Y}$, and $T_{\gamma,Z}$, while $T_{\delta,X}$ is the body. Generic quantification gives $\mathbb{P}(q) = \mathbb{P}(t_{\delta,X}\,|\,t_{\alpha,X}, t_{\beta,Y}, t_{\gamma,Z})$, marginalising out all three bound variables (X, Y, and Z).

This means the logical inference proposed by Emerson and Copestake (2017a) can in fact be seen as generic quantification. This is illustrated in Fig. 7, which corresponds to a sentence like *Rooms that have stoves are kitchens*, if α, β, γ, δ are set to *room*, *have*, *stove*, *kitchen*.[6]

Not only does this approach to quantification deal with both precise and vague quantifiers in a uniform way, it can also explain why generics are easier to process than precise quantifiers.

6 Donkey Sentences

An example of a donkey sentence is shown in (3). They are challenging for classical semantic theories, because naive composition, shown in (4), leaves a variable (y) outside the scope of its quantifier (Geach, 1962). The tempting solution in (5) requires a universal quantifier for an indefinite (*a donkey*), which would be non-compositional.[7]

(3) Every farmer who owns a donkey feeds it.

(4) $\forall x\big[\big(\textit{farmer}(x) \wedge \exists y[\textit{donkey}(y) \wedge \textit{own}(x,y)]\big) \rightarrow \textit{feed}(x,y)\big]$

(5) $\forall x \forall y\big[\big(\textit{farmer}(x) \wedge \textit{donkey}(y) \wedge \textit{own}(x,y)\big) \rightarrow \textit{feed}(x,y)\big]$

Kanazawa (1994), Brasoveanu (2008), and King and Lewis (2016) discuss how donkey sentences seem to admit multiple readings, which vary in the strength of their truth conditions, and which depend on both lexical semantics and the

[6] An example from RELPRON (Rimell et al., 2016).

[7] For simplicity, (4) and (5) suppress event variables.

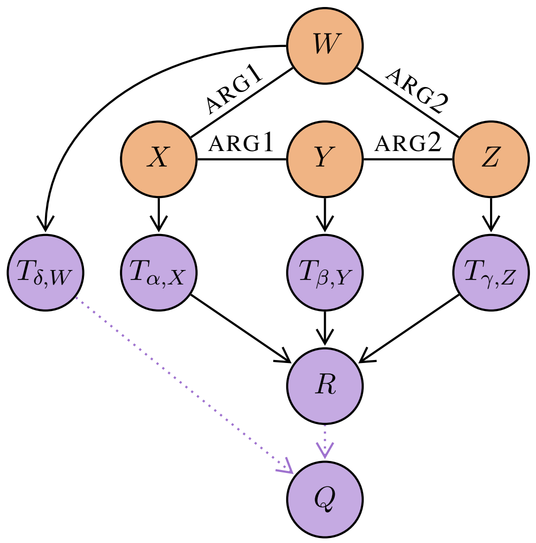


Figure 8: Analysis of a generic donkey sentence, using generic quantification. The quantifier node Q has restriction R (a logical conjunction) and body $T_{\delta,W}$.

context of use. This kind of dependence is exactly what Tessler and Goodman (2019) explained using RSA, so I will apply the same tools here.

As discussed in §5, generics are more basic than classical quantifiers, so I first consider generic donkey sentences, as illustrated in (6)–(8). An analysis of (3) is given in Appendix A.

(6) Farmers who own donkeys feed them.

(7) Linguists who use probabilistic models love them.

(8) Mosquitoes which bite birds infect them with malaria.

Example (8) shows it is inappropriate to use a universal quantifier: not all mosquitoes carry malaria, and not all bitten birds are infected (even if bitten by a malaria-carrying mosquito). However, this sentence still communicates that malaria is spread between birds by mosquitoes. This relies on pragmatic inference, from prior knowledge that most animals cannot spread malaria.

Despite the challenge for classical theories, generic donkey sentences can be straightforwardly handled by my proposed probabilistic approach. An example is shown in Fig. 8, which corresponds to (6), if α, β, γ, δ are set to *farmer*, *own*, *donkey*, *feed*. Intuitively, the more likely it is that a farmer owning a donkey implies the farmer feeding the donkey, the more likely it is for the sentence to be true. Given world knowledge and a discourse context, this can lead to a sharp threshold for being uttered, using RSA's pragmatic inference.

7 Related Work

Functional Distributional Semantics is related to other probabilistic semantic approaches. Goodman and Lassiter (2015) and Bernardy et al. (2018, 2019) represent meaning as a probabilistic program. This paper brings Functional Distributional Semantics closer to their work, because a probabilistic scope tree can be seen as a probabilistic program. An important practical difference is that Functional Distributional Semantics represents all predicates in the same way (as functions of pixies), allowing a model to be trained on corpus data.

Probabilistic TTR (Cooper, 2005; Cooper et al., 2015) also represents meaning as a probabilistic truth-conditional function. However, in this paper I have provided an alternative compositional semantics, in order to deal with vague quantifiers and generics. In principle, my proposal could be incorporated into a probabilistic TTR appproach. Furthermore, although Cooper et al. (2015) discuss learning, they assume a richer input than available in distributional semantics.

Some hybrid distributional-logical systems exist (for example: Lewis and Steedman, 2013; Grefenstette, 2013; Herbelot and Vecchi, 2015; Beltagy et al., 2016), but these do not discuss challenging cases like generics and donkey sentences.

Explaining the multiple readings of donkey sentences using pragmatic inference has been proposed using non-probabilistic tools (for example: Champollion, 2016; Champollion et al., 2019). I have provided a concrete computational method to calculate such inferences, in the same way that Tessler and Goodman (2019) have provided a concrete account of generics.

8 Conclusion

In this paper, I have presented a compositional semantics for both precise and vague quantifiers, in the probabilistic framework of Functional Distributional Semantics. I have re-interpreted previous work in this framework as performing generic quantification, building on the approach of Tessler and Goodman (2019). I have shown how generic quantification is computationally simpler than classical quantification, consistent with evidence that generics are a "default" mode of processing. Finally, I have presented examples of generic donkey sentences, which are doubly challenging for classical theories, but straightforward under my proposed approach.

Acknowledgements

This paper builds on chapter 7 of my PhD thesis (Emerson, 2018), and I would like to thank my PhD supervisor Ann Copestake, for her support, advice, and suggestions. I would also like to thank the anonymous reviewers, for pointing out areas that were unclear, and suggesting additional areas for discussion.

I am supported by a Research Fellowship at Gonville & Caius College, Cambridge.

References

Keith Allan. 2001. *Natural language semantics*. Blackwell.

Jon Barwise and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2):159–219.

Jon Barwise and John Perry. 1983. *Situations and Attitudes*. Massachusetts Institute of Technology (MIT) Press.

Islam Beltagy, Stephen Roller, Pengxiang Cheng, Katrin Erk, and Raymond J. Mooney. 2016. Representing meaning with a combination of logical and distributional models. *Computational Linguistics*, 42(4):763–808.

Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, and Shalom Lappin. 2018. A compositional Bayesian semantics for natural language. In *Proceedings of the 1st International Workshop on Language, Cognition and Computational Models*, pages 1–10.

Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, Shalom Lappin, and Aleksandre Maskharashvili. 2019. Bayesian Inference Semantics: A modelling system and a test suite. In *Proceedings of the 8th Joint Conference on Lexical and Computational Semantics (*SEM)*, pages 263–272.

Adrian Brasoveanu. 2008. Donkey pluralities: plural information states versus non-atomic individuals. *Linguistics and Philosophy*, 31(2):129–209.

Ronnie Cann. 1993. *Formal semantics: an introduction*. Cambridge University Press.

Gregory N. Carlson. 1977. *Reference to kinds in English*. Ph.D. thesis, University of Massachusetts at Amherst.

Gregory N. Carlson and Francis Jeffry Pelletier, editors. 1995. *The generic book*. University of Chicago Press.

Lucas Champollion. 2016. Homogeneity in donkey sentences. In *Proceedings of the 26th Semantics and Linguistic Theory Conference (SALT 26)*, pages 684–704.

Lucas Champollion, Dylan Bumford, and Robert Henderson. 2019. Donkeys under discussion. *Semantics and Pragmatics*, 12(1):1–45.

Robin Cooper. 2005. Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3(2-3):333–362.

Robin Cooper, Simon Dobnik, Staffan Larsson, and Shalom Lappin. 2015. Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology (LiLT)*, 10.

Ann Copestake. 2009. Slacker semantics: Why superficiality, dependency and avoidance of commitment can be the right way to go. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 1–9.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal Recursion Semantics: An introduction. *Research on Language and Computation*, 3(2-3):281–332.

Donald Davidson. 1967. The logical form of action sentences. In Nicholas Rescher, editor, *The Logic of Decision and Action*, chapter 3, pages 81–95. University of Pittsburgh Press. Reprinted in: Davidson (1980/2001), *Essays on Actions and Events*, Oxford University Press.

Guy Emerson. 2018. *Functional Distributional Semantics: Learning Linguistically Informed Representations from a Precisely Annotated Corpus*. Ph.D. thesis, University of Cambridge.

Guy Emerson. 2020a. Autoencoding pixies: Amortised variational inference with graph convolutions for Functional Distributional Semantics. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Guy Emerson. 2020b. What are the goals of distributional semantics? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Guy Emerson and Ann Copestake. 2016. Functional Distributional Semantics. In *Proceedings of the 1st Workshop on Representation Learning for NLP (RepL4NLP)*, pages 40–52. Association for Computational Linguistics.

Guy Emerson and Ann Copestake. 2017a. Variational inference for logical inference. In *Proceedings of the Conference on Logic and Machine Learning in Natural Language (LaML)*, pages 53–62. Centre for Linguistic Theory and Studies in Probability (CLASP).

Guy Emerson and Ann Copestake. 2017b. Semantic composition via probabilistic model theory. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS)*, pages 62–77. Association for Computational Linguistics.

Kit Fine. 1975. Vagueness, truth and logic. *Synthese*, 30(3-4):265–300.

Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Peter Gärdenfors. 2000. *Conceptual spaces: The geometry of thought*. Massachusetts Institute of Technology (MIT) Press.

Peter Gärdenfors. 2014. *Geometry of meaning: Semantics based on conceptual spaces*. Massachusetts Institute of Technology (MIT) Press.

Peter Thomas Geach. 1962. *Reference and generality: An examination of some medieval and modern theories*. Cornell University Press.

Susan A. Gelman, Sarah-Jane Leslie, Alexandra M. Was, and Christina M. Koch. 2015. Children's interpretations of general quantifiers, specific quantifiers and generics. *Language, Cognition and Neuroscience*, 30(4):448–461.

Noah D. Goodman and Michael C. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, 20(11):818–829.

Noah D. Goodman and Daniel Lassiter. 2015. Probabilistic semantics and pragmatics: Uncertainty in language and thought. In Shalom Lappin and Chris Fox, editors, *The Handbook of Contemporary Semantic Theory*, 2nd edition, chapter 21, pages 655–686. Wiley.

Edward Grefenstette. 2013. Towards a formal distributional semantics: Simulating logical calculi with tensors. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics (*SEM)*, pages 1–10. Association for Computational Linguistics.

Aurélie Herbelot. 2010. *Underspecified quantification*. Ph.D. thesis, University of Cambridge.

Aurélie Herbelot and Ann Copestake. 2011. Formalising and specifying underquantification. In *Proceedings of the 9th International Conference on Computational Semantics (IWCS)*, pages 165–174. Association for Computational Linguistics.

Aurélie Herbelot and Eva Maria Vecchi. 2015. Building a shared world: mapping distributional to model-theoretic semantic spaces. In *Proceedings of the 20th Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 22–32. Association for Computational Linguistics.

Michelle A. Hollander, Susan A. Gelman, and Jon Star. 2002. Children's interpretation of generic noun phrases. *Developmental Psychology*, 38(6):883–394.

Hans A. W. Kamp. 1975. Two theories about adjectives. In Edward L. Keenan, editor, *Formal Semantics of Natural Language*, chapter 9, pages 123–155. Reprinted in: Stephen Davis and Brendan S. Gillon, editors (2004), *Semantics: A Reader*, chapter 26, pages 541–562; Kamp (2013), *Meaning and the Dynamics of Interpretation: Selected Papers of Hans Kamp*, pages 225–261.

Hans A. W. Kamp and Uwe Reyle. 2013. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, volume 42 of *Studies in Linguistics and Philosophy*. Springer.

Makoto Kanazawa. 1994. Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting. *Linguistics and Philosophy*, 17(2):109–158.

Rosanna Keefe. 2000. *Theories of Vagueness*. Cambridge Studies in Philosophy. Cambridge University Press.

Sangeet Khemlani, Sarah-Jane Leslie, Sam Glucksberg, and Paula Rubio Fernandez. 2007. Do ducks lay eggs? how people interpret generic assertions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 29, pages 395–400.

Jeffrey C. King and Karen S. Lewis. 2016. Anaphora. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, fall 2016 edition. Metaphysics Research Lab, Stanford University.

Daniel Lassiter. 2011. Vagueness as probabilistic linguistic knowledge. In Rick Nouwen, Robert van Rooij, Uli Sauerland, and Hans-Christian Schmitz, editors, *Vagueness in Communication: Revised Selected Papers from the 2009 International Workshop on Vagueness in Communication*, chapter 8, pages 127–150. Springer.

Sarah-Jane Leslie. 2008. Generics: Cognition and acquisition. *The Philosophical Review*, 117(1):1–47.

Mike Lewis and Mark Steedman. 2013. Combined distributional and logical semantics. *Transactions of the Association for Computational Linguistics (TACL)*, 1:179–192.

Godehard Link. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow, editors, *Meaning, Use and the Interpretation of Language*, chapter 18, pages 303–323. Walter de Gruyter. Reprinted in: Paul Portner and Barbara H. Partee, editors (2002), *Formal semantics: The essential readings*, chapter 4, pages 127–146.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *Proceedings of the 1st International Conference on Learning Representations (ICLR), Workshop Track*.

Richard Montague. 1973. The proper treatment of quantification in ordinary English. In K. Jaakko J. Hintikka, Julius M. E. Moravcsik, and Patrick Suppes, editors, *Approaches to Natural Language*, number 49 in Synthese Library, chapter 10, pages 221–242. Kluwer Academic Publishers. Reprinted in: Paul Portner and Barbara H. Partee, editors (2002), *Formal semantics: The essential readings*, chapter 1, pages 17–34.

Terence Parsons. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. Current Studies in Linguistics. Massachusetts Institute of Technology (MIT) Press.

Barbara H. Partee. 1988. Many quantifiers. In *Proceedings of the Eastern States Conference on Linguistics (ESCOL)*, pages 383–402. Ohio State University. Reprinted in: Partee (2004), *Compositionality in Formal Semantics*, pages 241–258, Blackwell.

Barbara H. Partee. 2012. The starring role of quantifiers in the history of formal semantics. In *The Logica Yearbook 2012*, pages 113–136. College Publications.

Laura Rimell, Jean Maillard, Tamara Polajnar, and Stephen Clark. 2016. RELPRON: A relative clause evaluation dataset for compositional distributional semantics. *Computational Linguistics*, 42(4):661–701.

Peter R. Sutton. 2015. Towards a probabilistic semantics for vague adjectives. In Henk Zeevat and Hans-Christian Schmitz, editors, *Bayesian Natural Language Semantics and Pragmatics*, chapter 10, pages 221–246. Springer.

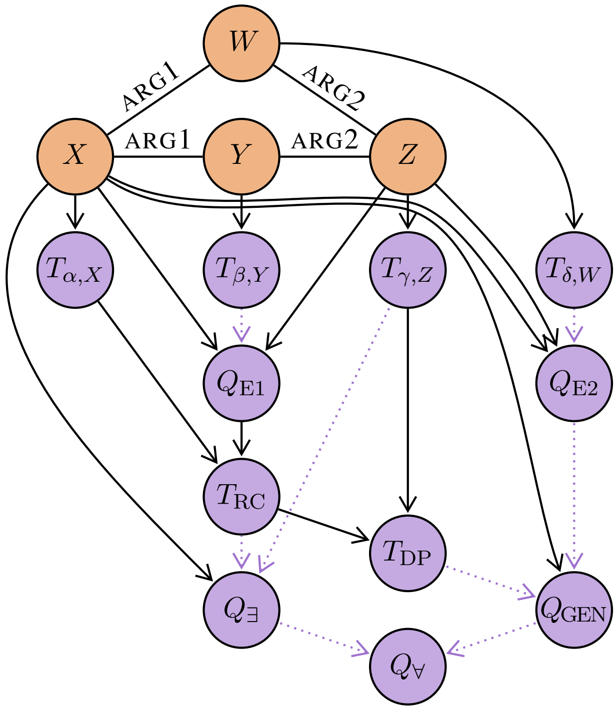
Peter R. Sutton. 2017. Probabilistic approaches to vagueness and semantic competency. *Erkenntnis*.

Michael Henry Tessler. 2018. *Communicating generalizations: Probability, vagueness, and context*. Ph.D. thesis, Stanford University.

Michael Henry Tessler and Noah D. Goodman. 2019. The language of generalization. *Psychological Review*, 126(3):395–436.

Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188.

Johan Van Benthem. 1984. Questions about quantifiers. *The Journal of Symbolic Logic*, 49(2):443–466.



A Classical Donkey Sentences

In this analysis of a classical donkey sentence, the donkey pronoun is associated with a generic quantifier, while all other quantifiers are precise. The generic quantifier allows the range of readings associated with donkey sentences.

The above figure corresponds to example (3), if α, β, γ, δ are set to *farmer*, *own*, *donkey*, *feed*. Intuitively, this analysis says that, if all farmers who own at least one donkey feed at least a proportion p of their donkeys, then this sentence is true with probability p.

The probability of truth gradually increases with the proportion p. Given world knowledge and a discourse context, this can lead to a sharp threshold proportion above which it is uttered, using pragmatic inference in the RSA framework. If distinguishing small proportions is pragmatically relevant, the *weak* reading becomes preferred. If distinguishing large proportions is pragmatically relevant, the *strong* reading becomes preferred.

I will now go over all nodes in the graph. Firstly, the distributions for $T_{\alpha,X}$, $T_{\beta,Y}$, $T_{\gamma,Z}$, $T_{\delta,W}$ are determined by the predicates.

The remaining truth value nodes are labelled for convenience. T_{RC} and T_{DP} are logical conjunctions (for the relative clause and donkey pronoun, respectively). The remaining five nodes are quantifier nodes, each quantifying one variable.

Note that Z is quantified twice (by $Q_{\exists}$ and Q_{GEN}). This would be surprising in a classical logic, but is not a problem here – marginalising out a random variable means that the quantifier node is not dependent on that variable, but the random variable is still part of the joint distribution, so it can be referred to by other nodes. Because of this double quantification, the scope "tree" is actually a scope DAG (directed acyclic graph).

Q_{E1} and Q_{E2} marginalise out the event variables, respectively Y and W, with trivially true restrictions and bodies $T_{\beta,Y}$ and $T_{\delta,W}$, leaving free variables X and Z. They can be treated like *some* in Fig. 6. For given pixies x and z, Q_{E1} is true if x owns z; Q_{E2} is true if x feeds z.

$Q_{\exists}$ marginalises out Z, with T_{RC} as restriction and Q_{E1} as body, leaving the free variable X. It can be treated like *some* in Fig. 6. For a given x, it is true if x is a farmer and there is a donkey z such that Q_{E1} is true.

Q_{GEN} also marginalises out Z, with T_{DP} as restriction and Q_{E2} as body, leaving the free variable X. It uses the generic quantifier, as in (11). For a given x, it considers donkeys z for which Q_{RC} is true; the probability of truth is the proportion of such z for which Q_{E2} is true (out of donkeys owned by farmer x, the proportion fed by x).

Finally, $Q_{\forall}$ marginalises out X, with $T_{\exists}$ as restriction and Q_{GEN} as body, leaving no free variables. It is treated as in Fig. 6. It is true if, whenever $T_{\exists}$ is true of x, Q_{GEN} is true of x, considering Q_{GEN} as a distribution over precise functions.