Distributional semantics for linguists

Lecture 5: Treating quantification in distributional semantics

Aurelie Herbelot

Universität Potsdam
Department Linguistik

ESSLLI 2012

Outline

Overview

- Introducing a relation between quantification and lexical semantics.
- A short recap on Lexicalised Compositionality.
- Doing model-theoretic quantification with distributions.
- Moving away from truth theory into a model of language comprehension.
- How to learn quantification? A real example.

Outline

Quantification

- Quantification is a phenomenon which cannot be directly represented in a distributional way.
- Quantifiers (*some*, *all*, *more than 32*) do not have a lexical representation. They are operators which 'count' over elements of a set.
- And still, there is a relation between lexical meaning and quantification.

The heffalump

Heffalumps eat grass. They are striped and have a long tail, as well as a trunk. They live in packs.
True or false: All heffalumps are animals. Most heffalumps live underwater. Some heffalumps are blind. All heffalumps are blind.

Conceptual representations

- A complete representation for a concept must involve some kind of quantificational information.
- Prototype theory (intension?) tells us what a representative instance of a concept should be like (for instance, a bird flies, has wings, build a nest, etc) but it is not able to account for the variety of utterances that people produce:
  - Most birds fly.
  - All birds have wings.
  - Some birds build nests.

Quantification in a distributional setting

- Quantification can be represented as a *relation* over distributions...
- ... but only in a setup where individual instances are available.
- The representation should account for the quantified sentences that humans produce with respect to particular concepts.
- It should also account for the fact that people can produce quantified sentences for concepts they don't master.

From production to understanding

- Another way to look at it: can we model what humans do when interpreting quantified statements?
  - I know 3 famous computational linguists.
  - I know 3000 famous computational linguists.
  - We baked 10 cakes yesterday! (20? 50? 90?)
  - 300 countries have signed the new global peace treaty.

Outline

An idealised representation

- **Ideal distributions** correspond to *complete distributional information* for a world *w*.
- They encapsulate information about *individual entities* and the *situations* in which those entities are found.
- They are hypothetical in the sense that they cannot be straightforwardly extracted from text. It is possible to regard them as the linguistic ‘competence’ of a speaker.

Example: a microworld w_0

- Two small elephants playing and, in another place and at another time, a zebra eating.
- Let's assume a speaker whose vocabulary consists of the terms *small, elephant, zebra, play, eat* and the quantifiers *a/an* and *two*.

A small elephant plays. (x2)
Two small elephants play.
An elephant plays. (x2)
Two elephants play.
A zebra eats.

Logical forms for w_0

- Logical forms in predicate logic (implicit conjunctions):

$$
\begin{aligned}
&\text{elephant}'(x_1), \text{small}'(x_1), \text{play}'(e_1, x_1)\\
&\text{elephant}'(x_2), \text{small}'(x_2), \text{play}'(e_2, x_2)\\
&\text{elephant}'(x_1), \text{small}'(x_1), \text{play}'(e_1, x_1), \text{elephant}'(x_2), \text{small}'(x_2), \text{play}'(e_2, x_2)\\
&\text{elephant}'(x_1), \text{play}'(e_1, x_1)\\
&\text{elephant}'(x_2), \text{play}'(e_2, x_2)\\
&\text{elephant}'(x_1), \text{play}'(e_1, x_1), \text{elephant}'(x_2), \text{play}'(e_2, x_2)\\
&\text{zebra}'(x_3), \text{eat}'(e_3, x_3)
\end{aligned}
$$

- **Note:** plural quantifiers are expressed by repeating the appropriate logical form for each entity in the plural set.

Ideal context sets for w_0

$$
\begin{array}{lll}
\text{elephant}^\circ \equiv & \{ & <[x1][\text{small}^\circ(x1), \text{play}^\circ(e1, x1)], S_1>, \\
 & & <[x1][\text{play}^\circ(e1, x1)], S_1>, \\
 & & <[x2][\text{small}^\circ(x2), \text{play}^\circ(e2, x2)], S_1>, \\
 & & <[x2][\text{play}^\circ(e2, x2)], S_1>\} \\
\text{zebra}^\circ \equiv & \{ & <[x3][\text{eat}^\circ(e3, x3)], S_2>\} \\
\text{small}^\circ \equiv & \{ & <[x1][\text{elephant}^\circ(x1), \text{play}^\circ(e1, x1)], S_1>, \\
 & & <[x2][\text{elephant}^\circ(x2), \text{play}^\circ(e2, x2)], S_1>\} \\
\text{play}^\circ \equiv & \{ & <[e1, x1][\text{elephant}^\circ(x1), \text{small}^\circ(x1)], S_1>, \\
 & & <[e1, x1][\text{elephant}^\circ(x1)], S_1>, \\
 & & <[e2, x2][\text{elephant}^\circ(x2), \text{small}^\circ(x2)], S_1>, \\
 & & <[e2, x2][\text{elephant}^\circ(x2)], S_1>\} \\
\text{eat}^\circ \equiv & \{ & <[e3, x3][\text{zebra}^\circ(x3)], S_2>\}
\end{array}
$$

Figure: Full context sets for w_0

Correspondence between LC and models

- There is a very straightforward correspondence between LC and the standard notion of extension.
- We only need to know the real world equalities between the constants corresponding to distributional arguments.
- So... we can just do what model-theoretic semantics does?

Outline

Counting with distributions

- Truth of sentences such as *An elephant eats*, *some elephants eat*, *more than 36 elephants eat*, given situations that have been observed by both speaker and hearer (the hearer is an omniscient being).

World w_1

- w_1 comprises one situation with two playing elephants, one eating elephant and one elephant that eats and plays.
  We omit the situation variable in what follows.

$$
\begin{array}{lll}
\text{elephant}^\circ(x) = & \{ & <[x1][\text{play}^\circ(e1, x1)]>, \\
 & & <[x2][\text{play}^\circ(e2, x2)]>, \\
 & & <[x3][\text{eat}^\circ(e3, x3)]>, \\
 & & <[x4][\text{play}^\circ(e4, x4)]>, \\
 & & <[x4][\text{eat}^\circ(e4, x4)]>\} \\
\text{play}^\circ(e, x) = & \{ & <[e1, x1][\text{elephant}^\circ(x1)]>, \\
 & & <[e2, x2][\text{elephant}^\circ(x2)]>, \\
 & & <[e5, x4][\text{elephant}^\circ(x4)]>\} \\
\text{eat}^\circ(e, x) = & \{ & <[e3, x3][\text{elephant}^\circ(x3)]>, \\
 & & <[e4, x4][\text{elephant}^\circ(x4)]>\}
\end{array}
$$

Counting in LC

- We need to get to the cardinality of the sets involved in the full context sets.

The original full context set

$$
\begin{aligned}
\text{elephant}^\circ(x) \equiv & \quad \{ \quad < [x1][\text{play}^\circ(e1, x1)] >, \\
& \qquad < [x2][\text{play}^\circ(e2, x2)] >, \\
& \qquad < [x3][\text{eat}^\circ(e3, x3)] >, \\
& \qquad < [x4][\text{play}^\circ(e4, x4)] >, \\
& \qquad < [x4][\text{eat}^\circ(e4, x4)] > \} \\
\text{play}^\circ(e, x) \equiv & \quad \{ \quad < [e1, x1][\text{elephant}^\circ(x1)] >, \\
& \qquad < [e2, x2][\text{elephant}^\circ(x2)] >, \\
& \qquad < [e5, x4][\text{elephant}^\circ(x4)] > \} \\
\text{eat}^\circ(e, x) \equiv & \quad \{ \quad < [e3, x3][\text{elephant}^\circ(x3)] >, \\
& \qquad < [e4, x4][\text{elephant}^\circ(x4)] > \}
\end{aligned}
$$

Assume each lexeme co-occurs with itself

$$
\begin{aligned}
\text{elephant}^\circ(x) \equiv \quad & \{ \; <[x1][\text{elephant}^\circ(x1)]>, \\
& <[x2][\text{elephant}^\circ(x2)]>, \\
& <[x3][\text{elephant}^\circ(x3)]>, \\
& <[x4][\text{elephant}^\circ(x4)]>, \\
& <[x1][\text{play}^\circ(e1, x1)]>, \\
& <[x2][\text{play}^\circ(e2, x2)]>, \\
& <[x3][\text{eat}^\circ(e3, x3)]>, \\
& <[x4][\text{play}^\circ(e4, x4)]>, \\
& <[x4][\text{eat}^\circ(e4, x4)]> \} \\
\text{play}^\circ(e, x) \equiv \quad & \{ \; <[e1, x1][\text{play}^\circ(e1)]>, \\
& <[e2, x2][\text{play}^\circ(e2)]>, \\
& <[e5, x4][\text{play}^\circ(e5)]>, \\
& <[e1, x1][\text{elephant}^\circ(x1)]>, \\
& <[e2, x2][\text{elephant}^\circ(x2)]>, \\
& <[e5, x4][\text{elephant}^\circ(x4)]> \} \\
\text{eat}^\circ(e, x) \equiv \quad & \{ \; <[e3, x3][\text{eat}^\circ(e3)]>, \\
& <[e4, x4][\text{eat}^\circ(e4)]>, \\
& <[e3, x3][\text{elephant}^\circ(x3)]>, \\
& <[e4, x4][\text{elephant}^\circ(x4)]> \}
\end{aligned}
$$

Underspecify entities

$$
\begin{aligned}
\text{elephant}^\circ(x) \equiv \quad \{ \quad & < [x][\text{elephant}^\circ(x)] >, \\
& < [x][\text{elephant}^\circ(x)] >, \\
& < [x][\text{elephant}^\circ(x)] >, \\
& < [x][\text{elephant}^\circ(x)] >, \\
& < [x][\text{play}^\circ(e, x)] >, \\
& < [x][\text{play}^\circ(e, x)] >, \\
& < [x][\text{eat}^\circ(e, x)] >, \\
& < [x][\text{play}^\circ(e, x)] >, \\
& < [x][\text{eat}^\circ(e, x)] > \} \\
\text{play}^\circ(e, x) \equiv \quad \{ \quad & < [e, x][\text{play}^\circ(e)] >, \\
& < [e, x][\text{play}^\circ(e)] >, \\
& < [e, x][\text{play}^\circ(e)] >, \\
& < [e, x][\text{elephant}^\circ(x)] >, \\
& < [e, x][\text{elephant}^\circ(x)] >, \\
& < [e, x][\text{elephant}^\circ(x)] > \} \\
\text{eat}^\circ(e, x) \equiv \quad \{ \quad & < [e, x][\text{eat}^\circ(e)] >, \\
& < [e, x][\text{eat}^\circ(e)] >, \\
& < [e, x][\text{elephant}^\circ(x)] >, \\
& < [e, x][\text{elephant}^\circ(x)] > \}
\end{aligned}
$$

Underspecified Generalised form

- The LC context sets have been converted into an underspecified generalised (UG) form.
- The UG form can be expressed as a (frequency-based) vector space:

	elephant $^\circ(x)$	play $^\circ(e,x)$	eat $^\circ(e,x)$
elephant $^\circ(x)$	4	3	2
play $^\circ(e,x)$	3	3	0
eat $^\circ(e,x)$	2	0	2

Form of the semantic space

- The semantic space has dimensions derived from elementary predications (EPs) of the type $\text{P}^\circ(x)$, e.g. $\text{play}^\circ(e,x)$, $\text{eat}^\circ(e,x)$ and $\text{elephant}^\circ(x)$
- Each dimension is an EP with one (and only one) uninstantiated argument variable ('curried EP'): e.g. $\lambda x\text{hear}^\circ(e,x,y)$ or $\lambda y\text{hear}^\circ(e,x,y)$.
- The points in that multidimensional space are also curried EPs: each curried EP is represented both as a dimension and as a point, which in turns implies that *any curried EP can be defined in terms of all the others*.

Counting elephants

- The number of elephants in w_1 is given by counting the number of *distinct* real-world entities in the set of elephants, as given by the full context sets. This is equivalent to:
  $|\{x1, x2, x3, x4\}| = \text{elephant}^{\circ}_{UG} : \text{elephant}^{\circ}_{UG} = 4$.
- Similarly, we can derive the number of playing elephants by counting the number of entities x in elephant$^{\circ}(x)$ which fill the argument x in play$^{\circ}(e, x)$. In UG form:
  $|\{x1, x2, x4\}| = \text{play}^{\circ}_{UG}(e, x) : \text{elephant}^{\circ}_{UG}(x) = 3$.

Cardinalities in LC

- Assuming generalised quantifiers of the type $Q(x)[rstr(x) \wedge scp(x)]$ where Q can be *any* quantifier (*some, all, three, ten out of thirty, the majority of*...)
- ... the cardinality of a particular set of entities x of type *rstr* filling the argument of *scp* is given by the position of *rstr* along the axis representing *scp*, that is,
  $\mathrm{rstr}^{\circ}_{UG}(x) : \mathrm{scp}^{\circ}_{UG}(e, x)$.

Computing truth

- Computing the truth value of a sentence S means comparing the ideal distributions of the 'real world' (w_1) (as seen by an omniscient being) with the ideal distributions of the world assumed by the speaker.
- **Example:** $S =$ *Three elephants eat*.

$$\begin{aligned}
\text{elephant}^\circ = \quad & \{ \quad < [x6][\text{eat}^\circ(e6, x6)] >, \\
& \quad\;\; < [x7][\text{eat}^\circ(e7, x7)] >, \\
& \quad\;\; < [x8][\text{eat}^\circ(e8, x8)] > \} \\
\text{eat}^\circ = \quad & \{ \quad < [e6, x6][\text{elephant}^\circ(x6)] >, \\
& \quad\;\; < [e7, x7][\text{elephant}^\circ(x7)] >, \\
& \quad\;\; < [e8, x8][\text{elephant}^\circ(x8)] > \}
\end{aligned}$$

- In S, $\text{elephant}^\circ_{UG}(x) : \text{eat}^\circ_{UG}(e, x) = 3$.
  In w_1, $\text{elephant}^\circ_{UG}(x) : \text{eat}^\circ_{UG}(e, x) = 2$.
  $(\text{elephant}^\circ_{UG}(x) : \text{eat}^\circ_{UG}(e, x))_S \neq (\text{elephant}^\circ_{UG}(x) : \text{eat}^\circ_{UG}(e, x))_{w_1}$ so S is false.

Consistency of models with the quantifier

- When evaluating statements involving quantifiers such as *some* or *most*, it is necessary to consider all possible distributional models which satisfy the constraint imposed by the quantifier.
- **Example:** *Some elephants play* is true for any number of elephants greater than 1 and may be expressed by any of the following distributional models...

Distributional models for *some*

$$
\begin{aligned}
\text{elephant}^\circ = \{ \quad & <[x6][\text{play}^\circ(e6,x6)]>, \\
& <[x7][\text{play}^\circ(e7,x7)]>\} \\
\text{play}^\circ = \{ \quad & <[e6,x6][\text{elephant}^\circ(x6)]>, \\
& <[e7,x7][\text{elephant}^\circ(x7)]>\}
\end{aligned}
$$

$$
\begin{aligned}
\text{elephant}^\circ = \{ \quad & <[x6][\text{play}^\circ(e6,x6)]>, \\
& <[x7][\text{play}^\circ(e7,x7)]>, \\
& <[x8][\text{play}^\circ(e8,x8)]>, \\
& <[x9][\text{eat}^\circ(e9,x9)]>\} \\
\text{play}^\circ = \{ \quad & <[e6,x6][\text{elephant}^\circ(x6)]>, \\
& <[e7,x7][\text{elephant}^\circ(x7)]>, \\
& <[e8,x8][\text{elephant}^\circ(x8)]>\} \\
\text{eat}^\circ = \{ \quad & <[e9,x9][\text{elephant}^\circ(x9)]>\}
\end{aligned}
$$

$$
\begin{aligned}
\text{elephant}^\circ = \{ \quad & <[x6][\text{play}^\circ(e6,x6)]>, \\
& \vdots \\
& <[x_k][\text{play}^\circ(e_k,x_k)]>\} \\
\text{play}^\circ = \{ \quad & <[e6,x6][\text{elephant}^\circ(x6)]>, \\
& \vdots \\
& <[e_k,x_k][\text{elephant}^\circ(x_k)]>\}
\end{aligned}
$$

$$
\begin{aligned}
\text{elephant}^\circ = \{ \quad & <[x6][\text{play}^\circ(e6,x6)]>, \\
& \vdots \\
& <[x_n][\text{play}^\circ(e_n,x_n)]>\} \\
\text{play}^\circ = \{ \quad & <[e6,x6][\text{elephant}^\circ(x6)]>, \\
& \vdots \\
& <[e_n,x_n][\text{elephant}^\circ(x_n)]>\}
\end{aligned}
$$

Distributional models for *some*

- We can refer to those models as $M^{some}_{2.2...4.3...m.n}$. The superscript indicates that the model satisfies a certain quantifier and the subscript indicates the cardinalities $\text{rstr}^{\circ}_{UG} : \text{rstr}^{\circ}_{UG}$ and $\text{rstr}^{\circ}_{UG} : \text{scp}^{\circ}_{UG}$.
- **Example (continued):** the model $M^{some}_{4.3}$ satisfies the equalities
  - $(\text{elephant}^{\circ}_{UG}(x) : \text{elephant}^{\circ}_{UG}(x))_{M^{some}_{4.3}} = (\text{elephant}^{\circ}_{UG}(x) : \text{elephant}^{\circ}_{UG}(x))_{w_1}$
  - $(\text{elephant}^{\circ}_{UG}(x) : \text{play}^{\circ}_{UG}(e, x))_{M^{some}_{4.3}} = (\text{elephant}^{\circ}_{UG}(x) : \text{play}^{\circ}_{UG}(e, x))_{w_1}$.
- So the sentence *Some elephants play* is true with regard to w_1.

Computing truth (summary)

- Given a quantified sentence $S : Qx[rstr(x) \wedge scp(e,x)]$ where Q is a quantifier, $rstr(x)$ the restriction of Q and $scp(e,x)$ the scope of Q, we will define the truth t of S in world w as:

$$t = \begin{cases} 1 & \text{if there is one model } M^{Q}_{m.k} \text{ such that } m = (\text{rstr}^{\circ}_{UG}(x) : \text{rstr}^{\circ}_{UG}(x)) \\ & \text{and } k = (\text{rstr}^{\circ}_{UG}(x) : \text{scp}^{\circ}_{UG}(e,x))_w \\ 0 & \text{if not} \end{cases} \tag{1}$$

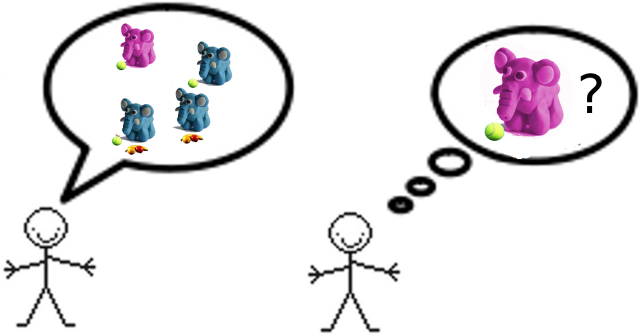
where $M^{Q}_{i.i...n.n}$ are all the models satisfying the constraint Q and $M^{Q}_{m.k}$ is *any* of those models.

- In other words, a quantified statement is true in w if its quantifier allows the existence of a model equal to w (as far as the restriction and scope of the quantifier are concerned).

Outline

Ungrounded situations

- The speaker utters a sentence about some world. This sentence is plausible given his or her model of the world (in our example, the distributions elephant°, eat° and play°).



Assumptions

- The speaker is talking about a situation (or world) which is not directly observable.
- This situation is *comparable* to situations he or she knows about.
- Retrieving comparable situations from ideal distributions is possible, although we won't talk about it today.

PUG distributions

- We can generalise the truth-theoretic model to unobserved situations by assuming **probabilistic underspecified generalised** (PUG) distributions.
- In PUG distributions, the value of $\mathrm{rstr}^\circ(x)$ along $\mathrm{scp}^\circ(e, x)$ is the probability for an individual in rstr° to fill the empty argument of scp°. We will initially assume that this probability is computed over the observed individuals in the full context set.
- **Example:** for the dimension $\mathrm{scp}^\circ(e, x)$:

$$\mathrm{rstr}^\circ_{PUG}(x) : \mathrm{scp}^\circ_{PUG}(e, x) = \frac{\mathrm{rstr}^\circ_{UG}(x) : \mathrm{scp}^\circ_{UG}(e, x)}{\mathrm{rstr}^\circ_{UG}(x) : \mathrm{rstr}^\circ_{UG}(x)} \quad (2)$$

Example

- If w_1 corresponds to an observed world (and ignoring the data sparsity issue), we have the following PUG distribution.

	elephant$^\circ(x)$	play$^\circ(e,x)$	eat$^\circ(e,x)$
elephant$^\circ(x)$	1	0.75	0.5
play$^\circ(e,x)$	0.75	1	0
eat$^\circ(e,x)$	0.5	0	1

Figure: Vectors corresponding to probabilistic underspecified generalised context sets for w_1

Accounting for utterances

- The PUG distribution directly reflects the assumptions of a speaker with regard to quantification in a world that he/she hasn't directly observed.
- For instance, a speaker acquainted with sufficiently many situations ressembling $w1$ might utter the following with respect to an imaginary world with 10 elephants:
    - 7 out of 10 elephants were playing.
    - Half of the elephants were eating.

Problem!

- Calculating a probability for rstr $^{\circ}_{UG}$ is often not meaningful:

  I went to see Joe Bloggs yesterday. He has 50 cats, all of them black.
  I went to see Granny Weatherwax yesterday. She has 50 cats, all of them black.

- Again, doing this relies on identifying comparable situations in the ideal distribution. (Not for today!)

Outline

The case of sparse (or null) data

The heffalump

Heffalumps eat grass. They are striped and have a long tail, as well as a trunk.
True or false: All heffalumps are animals. Most heffalumps live underwater. Some heffalumps are blind. All heffalumps are blind.

- Impossible to calculate probabilities... this cannot be treated in a pure model-theoretic setting.
- But we have lexical information...

The distributional dependence hypothesis

- Let us assume a distributional space with n dimensions.
- The **distributional dependence hypothesis**: we hypothesise that the value of a distribution rstr° along a dimension scp°_k is dependent on the value of rstr° along all other dimensions $\text{scp}^\circ_{1...n}$ in that space.

Intuitively...

... the probability that an elephant (habitually) eats is dependent on the probability of that elephant to (habitually) sleep, run, communicate, to be made of stone or to be sold in department stores. The distribution of a typical elephant x reflects its status as a living being, which in turn implies a high probability of $\text{elephant}^\circ(x)$ along the dimension $\text{eat}^\circ(e, x)$.

Outline

Heffalumps again...

- World w_2 has 3 elephants, 2 zebras, 5 lions and 4 fish.
- We assume a space with 9 dimensions:

$$
\begin{array}{l}
\text{elephant}^{\circ}(x) \\
\text{zebra}^{\circ}(x) \\
\text{lion}^{\circ}(x) \\
\text{fish}^{\circ}(x) \\
\text{hasTrunk}^{\circ}(e, x) \\
\text{eatGrass}^{\circ}(e, x) \\
\text{hasStripes}^{\circ}(e, x) \\
\text{jump}^{\circ}(e, x) \\
\text{underwater}^{\circ}(e, x)
\end{array}
$$

Heffalumps again...

- The (imaginary) PUG distributions for w_2 are represented below:

	elephant °	zebra °	lion °	fish °	hasTrunk °	striped °	eatGrass °	jump °	underwater °
elephant °	1	0	0	0	1	0	1	0	0
zebra °	0	1	0	0	0	1	1	1	0
lion °	0	0	1	0	0	0	0.4	1	0
fish °	0	0	0	1	0	0.75	0	0.25	1
hasTrunk °	1	0	0	0	1	0	0	0	0
striped °	0	1	0	0.75	0	1	0	0	0
eatGrass °	1	1	0.4	0	0	0	1	0	0
jump °	0	1	1	0.25	0	0	0	1	0
underwater °	0	0	0	1	0	0	0	0	1

Figure: Vectors corresponding to probabilistic underspecified generalised context sets for w_2

Heffalumps again...

- Let us now assume a speaker who knows (and believes) that all heffalumps have a trunk, are striped and eat grass.
- We can write the PUG distribution of *heffalump* in our 9-dimensional space as follows:

	elephant °	zebra °	lion °	fish °	hasTrunk °	hasStripes °	eatGrass °	jump °	underwater °
heffalump °	?	?	?	?	1	1	1	?	?

- How likely is it for that speaker to utter ‘All heffalumps live underwater’?

Learning distributions

- We assume that, having heard a few things about heffalumps, our speaker has built a conceptual representation of heffalumps which 'fills in' some of the missing information. (cf *rancid/off*)
- This process can be modelled using a classifier: something that takes training data (the distributions, or conceptual representations, that the speaker already has) and returns a model of how certain features/contexts are likely to associate with a new concept.

How?

- We can now transform each PUG distribution in the reference world into a training instance for a classifier. E.g., there is a training instance for fish$^\circ(x)$ which is the vector $[0, 0, 0, 1, 0, 0.75, 0, 0.25, 1]$.
- By feeding all available training vectors to a classifier, and withholding the component corresponding to the logical form scp$^\circ(x)$ for which we need a probability estimate, we can produce a model which tells us how likely an instance of type *rstr* is to fill the argument of scp$^\circ(x)$.
- E.g. we learn the feature underwater$^\circ(e, x)$ and subsequently use our learned model to predict the value of underwater$^\circ(e, x)$ for heffalump$^\circ(x)$ (which, we hope, should be low).

Outline

How to create an ideal distribution?

- Need co-reference resolution for large amounts of text:
  - including co-reference of definite NPs and underquantified NPs
  - across texts, not only within a single text.
- Need a way to mark situations.
- Must identify and appropriately process ‘encyclopedic knowledge’ (*The elephant is a mammal*).
- etc...

Corpus-based and ideal distributions

- At first glance, there is no relation between frequencies obtained from a corpus and real-world frequencies.

  My cat Kitty, who is a mammal, is 2 years old.
  My cat Kitty (a mammal) likes playing in the garden.
  Kitty, my cat – and a mammal –, is hungry.

- In some cases, though, there is one...
- Seeing the predicate *mammal* applied to *cat* **only once** in a corpus is sufficient to know that **all** cats are mammals.

Experimental setup

- A small data set of 59 animal names, with their distributions *ant, bat, beaver, bee, cat, chicken*...
- 8 features (vector components): *be_v+bird_n, be_v+insect_n, be_v+mammal_n, domestic_a, graze_v, hibernate_v, lay_v+egg_n, poisonous_a*
- The task: classifying every {animal, feature} pair into quantificational classes *no, a few, some, most all*.
- A manual annotation is performed and the data separated into training and test data.

Running the baseline

- In this experiment, we test a system based on the corpus-based distributions alone.
  i.e. whether *no, a few, some, most* or *all* elephants are domestic animals is decided on the basis of the elephant distribution alone.
- Results by features:

bird	*insect*	*mammal*	*domestic*	*graze*	*hibernate*	*layeggs*	*poisonous*
1	1	0.684	0.491	0.774	0.743	0.76	1

Incremental learning

- We don't have ideal distributions, but we would like to learn them.
- Incremental learning: incrementally 'correct' the actual distribution to obtain distributions which are closer and closer to the ideal distribution.

Incremental learning

- **Bootstrapping algorithm.**
- Iteration 1:
  - Learn classifiers for each feature (i.e. *bird, mammal, poisonous*, etc) using distributions only.
  - Calculate precision on training data.
    e.g. *mammal* classified with 0.491 precision.
  - Record best classifier and decisions made on training data.
    e.g. *bird* is the best classifier with precision of 1.
  - Add classified instances to training vectors.
    e.g. add feature *bird-learned* with value *no* to the training vector correponding to elephant$^\circ$.

Incremental learning

- **Bootstrapping algorithm.**
- Iterations 2-n:
  - Learn classifiers for each feature (i.e. *bird, mammal, poisonous*, etc) using new training data (including best learned feature from previous iteration).
  - Calculate precision on training data.
    e.g. *mammal* classified with 0.976 precision.
  - Record best classifier and decisions made on training data.
    e.g. *hibernate* is the best classifier with precision of 1.
  - Add classified instances to training vectors.
    e.g. add feature *hibernate-learned* with value *no* to the training vector correponding to elephant $^{\circ}$.

Results

- Baseline results (repeated):

bird	*insect*	*mammal*	*domestic*	*graze*	*hibernate*	*layeggs*	*poisonous*
1	1	0.684	0.491	0.774	0.743	0.76	1

- Bootstrapped results (on test data):

bird	*insect*	*mammal*	*domestic*	*graze*	*hibernate*	*layeggs*	*poisonous*
1	1	0.951	0.491	0.774	0.743	0.951	1

Baseline classifier for *be_v+mammal_n*

```
J48 unpruned tree
------------------

genus_n+of_p() <= 0.159023
|   in_p()+appearance_n <= 0.059356
|   |   be_v+inhabitant_n <= 0
|   |   |   kiss_n+of_p() <= 0: all (20.0/1.0)
|   |   |   kiss_n+of_p() > 0: no (2.0)
|   |   be_v+inhabitant_n > 0: no (2.0)
|   in_p()+appearance_n > 0.059356: no (4.0)
genus_n+of_p() > 0.159023: no (12.0)
```

Improved classifier for *be_v+mammal_n*

```
J48 unpruned tree
------------------

lay_v+egg_n:learned = no: all (20.0/1.0)
lay_v+egg_n:learned = afew: no (2.0)
lay_v+egg_n:learned = some: no (18.0)
lay_v+egg_n:learned = most: no (0.0)
lay_v+egg_n:learned = all: no (0.0)
```

Outline

Conclusion

- Standard accounts of quantification can be retained in a distributional setting.
- To do this, individual, real-world entities must be represented in lexemes' distributions.
- Using the distributional setup of Lexicalised Compositionality, it is possible to account for
  - The truth of quantified sentences, as in model-theoretic semantics.
  - The likelihood, for a particular speaker, that they will utter a certain quantified sentence about an ungrounded situation.