Distributional semantics for linguists
Lecture 5: Treating quantification in distributional semantics

Aurelie Herbelot

Universität Potsdam
Department Linguistik

ESSLLI 2012
Outline

1 Overview

2 Relating quantification and lexical semantics

3 Lexicalised Compositionality (again)

4 Quantification and truth

5 Quantification in ungrounded situations

6 Learning quantification
   - Learning over individuals
   - Pragmatic matters

7 Conclusion
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

1. Overview

2. Relating quantification and lexical semantics

3. Lexicalised Compositionality (again)

4. Quantification and truth

5. Quantification in ungrounded situations

6. Learning quantification
   - Learning over individuals
   - Pragmatic matters

7. Conclusion
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.
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

1. Overview
2. Relating quantification and lexical semantics
3. Lexicalised Compositionality (again)
4. Quantification and truth
5. Quantification in ungrounded situations
6. Learning quantification
   - Learning over individuals
   - Pragmatic matters
7. Conclusion
**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.
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 in predicate logic (implicit conjunctions):

\[
\begin{align*}
elephant'(x_1), & \quad \text{small}'(x_1), \quad \text{play}'(e_1, x_1) \\
elephant'(x_2), & \quad \text{small}'(x_2), \quad \text{play}'(e_2, x_2) \\
elephant'(x_1), & \quad \text{small}'(x_1), \quad \text{play}'(e_1, x_1), \quad \text{elephant}'(x_2), \quad \text{small}'(x_2), \quad \text{play}'(e_2, x_2) \\
elephant'(x_1), & \quad \text{play}'(e_1, x_1) \\
elephant'(x_2), & \quad \text{play}'(e_2, x_2) \\
elephant'(x_1), & \quad \text{play}'(e_1, x_1), \quad \text{elephant}'(x_2), \quad \text{play}'(e_2, x_2) \\
\text{zebra}'(x_3), & \quad \text{eat}'(e_3, x_3)
\end{align*}
\]

**Note:** plural quantifiers are expressed by repeating the appropriate logical form for each entity in the plural set.
Ideal context sets for $w_0$

\[
elephant^\circ \equiv \{< [x1][\text{small}^\circ(x1), \text{play}^\circ(e1, x1)], S_1 >,
\quad < [x1][\text{play}^\circ(e1, x1)], S_1 >,
\quad < [x2][\text{small}^\circ(x2), \text{play}^\circ(e2, x2)], S_1 >,
\quad < [x2][\text{play}^\circ(e2, x2)], S_1 >\}\n\]
\[
zebra^\circ \equiv \{< [x3][\text{eat}^\circ(e3, x3)], S_2 >\}\n\]
\[
\text{small}^\circ \equiv \{< [x1][\text{elephant}^\circ(x1), \text{play}^\circ(e1, x1)], S_1 >,
\quad < [x2][\text{elephant}^\circ(x2), \text{play}^\circ(e2, x2)], S_1 >\}\n\]
\[
\text{play}^\circ \equiv \{< [e1, x1][\text{elephant}^\circ(x1), \text{small}^\circ(x1)], S_1 >,
\quad < [e1, x1][\text{elephant}^\circ(x1)], S_1 >,
\quad < [e2, x2][\text{elephant}^\circ(x2), \text{small}^\circ(x2)], S_1 >,
\quad < [e2, x2][\text{elephant}^\circ(x2)], S_1 >\}\n\]
\[
\text{eat}^\circ \equiv \{< [e3, x3][\text{zebra}^\circ(x3)], S_2 >\}\n\]

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

1. Overview
2. Relating quantification and lexical semantics
3. Lexicalised Compositionality (again)
4. Quantification and truth
5. Quantification in ungrounded situations
6. Learning quantification
   - Learning over individuals
   - Pragmatic matters
7. Conclusion
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.

$$\text{elephant}^\circ(x) = \{ \langle [x1][\text{play}^\circ(e1, x1)] \rangle, \langle [x2][\text{play}^\circ(e2, x2)] \rangle, \langle [x3][\text{eat}^\circ(e3, x3)] \rangle, \langle [x4][\text{play}^\circ(e4, x4)] \rangle, \langle [x4][\text{eat}^\circ(e4, x4)] \rangle \}$$

$$\text{play}^\circ(e, x) = \{ \langle [e1, x1][\text{elephant}^\circ(x1)] \rangle, \langle [e2, x2][\text{elephant}^\circ(x2)] \rangle, \langle [e5, x4][\text{elephant}^\circ(x4)] \rangle \}$$

$$\text{eat}^\circ(e, x) = \{ \langle [e3, x3][\text{elephant}^\circ(x3)] \rangle, \langle [e4, x4][\text{elephant}^\circ(x4)] \rangle \}$$
We need to get to the cardinality of the sets involved in the full context sets.
The original full context set

\[
\text{elephant}^{\circ}(x) \equiv \{ \langle x_1][\text{play}^{\circ}(e_1, x_1)] \rangle, \\
\langle x_2][\text{play}^{\circ}(e_2, x_2)] \rangle, \\
\langle x_3][\text{eat}^{\circ}(e_3, x_3)] \rangle, \\
\langle x_4][\text{play}^{\circ}(e_4, x_4)] \rangle, \\
\langle x_4][\text{eat}^{\circ}(e_4, x_4)] \rangle \}
\]

\[
\text{play}^{\circ}(e, x) \equiv \{ \langle e_1, x_1][\text{elephant}^{\circ}(x_1)] \rangle, \\
\langle e_2, x_2][\text{elephant}^{\circ}(x_2)] \rangle, \\
\langle e_5, x_4][\text{elephant}^{\circ}(x_4)] \rangle \}
\]

\[
\text{eat}^{\circ}(e, x) \equiv \{ \langle e_3, x_3][\text{elephant}^{\circ}(x_3)] \rangle, \\
\langle e_4, x_4][\text{elephant}^{\circ}(x_4)] \rangle \}
\]
Assume each lexeme co-occurs with itself

\[
elephant \circ (x) \equiv \{  \langle [x1]\elephant \circ (x1) \rangle,  \\
\langle [x2]\elephant \circ (x2) \rangle,  \\
\langle [x3]\elephant \circ (x3) \rangle,  \\
\langle [x4]\elephant \circ (x4) \rangle,  \\
\langle [x1]\play \circ (e1, x1) \rangle,  \\
\langle [x2]\play \circ (e2, x2) \rangle,  \\
\langle [x3]\eat \circ (e3, x3) \rangle,  \\
\langle [x4]\play \circ (e4, x4) \rangle,  \\
\langle [x4]\eat \circ (e4, x4) \rangle \}  \\
\]

\[
\play \circ (e, x) \equiv \{  \langle [e1, x1]\play \circ (e1) \rangle,  \\
\langle [e2, x2]\play \circ (e2) \rangle,  \\
\langle [e5, x4]\play \circ (e5) \rangle,  \\
\langle [e1, x1]\elephant \circ (x1) \rangle,  \\
\langle [e2, x2]\elephant \circ (x2) \rangle,  \\
\langle [e5, x4]\elephant \circ (x4) \rangle \}  \\
\]

\[
\eat \circ (e, x) \equiv \{  \langle [e3, x3]\eat \circ (e3) \rangle,  \\
\langle [e4, x4]\eat \circ (e4) \rangle,  \\
\langle [e3, x3]\elephant \circ (x3) \rangle,  \\
\langle [e4, x4]\elephant \circ (x4) \rangle \}  \\
\]
Underspecify entities

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

<table>
<thead>
<tr>
<th></th>
<th>elephant (\circ(x))</th>
<th>play (\circ(e, x))</th>
<th>eat (\circ(e, x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>elephant (\circ(x))</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>play (\circ(e, x))</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>eat (\circ(e, x))</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
The semantic space has dimensions derived from elementary predications (EPs) of the type $P^\circ(x)$, e.g. play $^\circ(e, x)$, eat $^\circ(e, x)$ and 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:

  $|\{x_1, x_2, x_3, x_4\}| = \text{elephant}_{UG} : \text{elephant}_{UG} = 4$.

- Similarly, we can derive the number of playing elephants by counting the number of entities $x$ in $\text{elephant}(x)$ which fill the argument $x$ in $\text{play}(e, x)$. In UG form:

  $|\{x_1, x_2, x_4\}| = \text{play}_{UG}(e, x) : \text{elephant}_{UG}(x) = 3$. 
Cardinalities in LC

• Assuming generalised quantifiers of the type $Q(x)[rstr(x) \land scp(x)]$ 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, $rstr \circ UG(x) : scp \circ UG(e, x)$. 
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:** $S = Three$ $elephants$ $eat$.

\[
\text{elephant}^\circ = \{ < [x6][\text{eat}^\circ(e6, x6)] >, \\
\quad < [x7][\text{eat}^\circ(e7, x7)] >, \\
\quad < [x8][\text{eat}^\circ(e8, x8)] > \}
\]

\[
\text{eat}^\circ = \{ < [e6, x6][\text{elephant}^\circ(x6)] >, \\
\quad < [e7, x7][\text{elephant}^\circ(x7)] >, \\
\quad < [e8, x8][\text{elephant}^\circ(x8)] > \}
\]

In $S$, $\text{elephant}^\circ_U(x) : \text{eat}^\circ_U(e, x) = 3$.
In $w_1$, $\text{elephant}^\circ_U(x) : \text{eat}^\circ_U(e, x) = 2$.

$(\text{elephant}^\circ_U(x) : \text{eat}^\circ_U(e, x))_S \neq (\text{elephant}^\circ_U(x) : \text{eat}^\circ_U(e, x))_{w_1}$ so $S$ is false.
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

<table>
<thead>
<tr>
<th></th>
<th>elephant °=</th>
<th>play °=</th>
<th>eat °=</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>{</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>⟨[x₆]play °(e₆,x₆)⟩,</td>
<td>⟨[e₆,x₆][elephant °(x₆)]⟩,</td>
<td>⟨[e₆,x₆][elephant °(x₆)]⟩,</td>
</tr>
<tr>
<td></td>
<td>⟨[x₇]play °(e₇,x₇)⟩,</td>
<td>⟨[x₇][play °(e₇,x₇)]⟩,</td>
<td>⟨[x₇][play °(e₇,x₇)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⟨[x₈][play °(e₈,x₈)]⟩,</td>
<td>⟨[x₈][play °(e₈,x₈)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⟨[x₉][eat °(e₉,x₉)]⟩,</td>
<td>⟨[x₉][eat °(e₉,x₉)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⟨[x₉][play °(e₉,x₉)]⟩,</td>
<td>⟨[x₉][play °(e₉,x₉)]⟩,</td>
</tr>
<tr>
<td></td>
<td>⟨[x₉][elephant °(x₉)]⟩,</td>
<td>⟨[e₇,x₇][elephant °(x₇)]⟩,</td>
<td>⟨[e₇,x₇][elephant °(x₇)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⟨[e₈,x₈][elephant °(x₈)]⟩,</td>
<td>⟨[e₈,x₈][elephant °(x₈)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⟨[e₉,x₉][elephant °(x₉)]⟩,</td>
<td>⟨[e₉,x₉][elephant °(x₉)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>⟨[x₆]play °(e₆,x₆)⟩,</td>
<td>⟨[x₆][play °(e₆,x₆)]⟩,</td>
<td>⟨[x₆][play °(e₆,x₆)]⟩,</td>
</tr>
<tr>
<td></td>
<td>⟨[x₇][play °(e₇,x₇)]⟩,</td>
<td>⟨[x₇][play °(e₇,x₇)]⟩,</td>
<td>⟨[x₇][play °(e₇,x₇)]⟩,</td>
</tr>
<tr>
<td></td>
<td>⟨[x₈][play °(e₈,x₈)]⟩,</td>
<td>⟨[x₈][play °(e₈,x₈)]⟩,</td>
<td>⟨[x₈][play °(e₈,x₈)]⟩,</td>
</tr>
<tr>
<td></td>
<td>⟨[x₉][eat °(e₉,x₉)]⟩,</td>
<td>⟨[x₉][eat °(e₉,x₉)]⟩,</td>
<td>⟨[x₉][eat °(e₉,x₉)]⟩,</td>
</tr>
<tr>
<td></td>
<td>⟨[x₉][play °(e₉,x₉)]⟩,</td>
<td>⟨[x₉][play °(e₉,x₉)]⟩,</td>
<td>⟨[x₉][play °(e₉,x₉)]⟩,</td>
</tr>
<tr>
<td></td>
<td>⟨[x₉][elephant °(x₉)]⟩,</td>
<td>⟨[e₇,x₇][elephant °(x₇)]⟩,</td>
<td>⟨[e₇,x₇][elephant °(x₇)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⟨[e₈,x₈][elephant °(x₈)]⟩,</td>
<td>⟨[e₈,x₈][elephant °(x₈)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>⟨[e₉,x₉][elephant °(x₉)]⟩,</td>
<td>⟨[e₉,x₉][elephant °(x₉)]⟩,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Quantification and truth

Distributional models for *some*

- We can refer to those models as $M_{2.2\ldots 4.3\ldots m.n}^{\text{some}}$. The superscript indicates that the model satisfies a certain quantifier and the subscript indicates the cardinalities $rstr^U_G : rstr^U_G$ and $rstr^U_G : scp^U_G$.

- **Example (continued):** the model $M_{4.3}^{\text{some}}$ satisfies the equalities
  - $(\text{elephant}^U_G(x) : \text{elephant}^U_G(x))_{M_{4.3}^{\text{some}}} = (\text{elephant}^U_G(x) : \text{elephant}^U_G(x))_{w_1}$
  - $(\text{elephant}^U_G(x) : \text{play}^U_G(e, x))_{M_{4.3}^{\text{some}}} = (\text{elephant}^U_G(x) : \text{play}^U_G(e, x))_{w_1}$.

- So the sentence *Some elephants play* is true with regard to $w_1$. 
Computing truth (summary)

- Given a quantified sentence $S : Qx[rstr(x) \land scp(e, x)]$ 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:

\[
t = \begin{cases} 
1 & \text{if there is one model } M^{Q}_{m,k} \text{ such that } m = (rstr_{UG}(x) : rstr_{UG}(x)) \\
0 & \text{if not}
\end{cases}
\]

(1)

- 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

1. Overview

2. Relating quantification and lexical semantics

3. Lexicalised Compositionality (again)

4. Quantification and truth

5. Quantification in ungrounded situations

6. Learning quantification
   - Learning over individuals
   - Pragmatic matters

7. Conclusion
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.
We can generalise the truth-theoretic model to unobserved situations by assuming **probabilistic underspecified generalised** (PUG) distributions.

In PUG distributions, the value of \( r^{\circ}(x) \) along \( scp^{\circ}(e, x) \) is the probability for an individual in \( r^{\circ} \) to fill the empty argument of \( scp^{\circ} \). We will initially assume that this probability is computed over the observed individuals in the full context set.

**Example:** for the dimension \( scp^{\circ}(e, x) \):

\[
r^{\circ}_{PUG}(x) : scp^{\circ}_{PUG}(e, x) = \frac{r^{\circ}_{UG}(x) : scp^{\circ}_{UG}(e, x)}{r^{\circ}_{UG}(x) : r^{\circ}_{UG}(x)}
\]  

(2)
Example

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

<table>
<thead>
<tr>
<th></th>
<th>elephant°($x$)</th>
<th>play°($e, x$)</th>
<th>eat°($e, x$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>elephant°($x$)</td>
<td>1</td>
<td>0.75</td>
<td>0.5</td>
</tr>
<tr>
<td>play°($e, x$)</td>
<td>0.75</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>eat°($e, x$)</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**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 resembling $w_1$ 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 \( r_{\text{str}_{\text{UG}}} \) is often not meaningful:

\[
I \text{ went to see Joe Bloggs yesterday. He has 50 cats, all of them black.}
\]
\[
I \text{ went to see Granny Weatherwax yesterday. She has 50 cats, all of them black.}
\]

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

1. Overview
2. Relating quantification and lexical semantics
3. Lexicalised Compositionality (again)
4. Quantification and truth
5. Quantification in ungrounded situations
6. Learning quantification
   - Learning over individuals
   - Pragmatic matters
7. Conclusion
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 $scp^{°}_{1...n}$ 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 elephant $^{°}(x)$ along the dimension eat $^{°}(e, x)$. 
Outline

1 Overview
2 Relating quantification and lexical semantics
3 Lexicalised Compositionality (again)
4 Quantification and truth
5 Quantification in ungrounded situations
6 Learning quantification
   - Learning over individuals
   - Pragmatic matters
7 Conclusion
Heffalumps again...

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

  - elephant $°(x)$
  - zebra $°(x)$
  - lion $°(x)$
  - fish $°(x)$
  - hasTrunk $°(e, x)$
  - eatGrass $°(e, x)$
  - hasStripes $°(e, x)$
  - jump $°(e, x)$
  - underwater $°(e, x)$
Heffalumps again...

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

<table>
<thead>
<tr>
<th></th>
<th>elephant</th>
<th>zebra</th>
<th>lion</th>
<th>fish</th>
<th>hasTrunk</th>
<th>striped</th>
<th>eatGrass</th>
<th>jump</th>
<th>underwater</th>
</tr>
</thead>
<tbody>
<tr>
<td>elephant</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>zebra</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>lion</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>fish</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.75</td>
<td>0</td>
<td>0.25</td>
<td>1</td>
</tr>
<tr>
<td>hasTrunk</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>striped</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.75</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>eatGrass</td>
<td>1</td>
<td>1</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>jump</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>underwater</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**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:

<table>
<thead>
<tr>
<th></th>
<th>elephant</th>
<th>zebra</th>
<th>lion</th>
<th>fish</th>
<th>hasTrunk</th>
<th>hasStripes</th>
<th>eatGrass</th>
<th>jump</th>
<th>underwater</th>
</tr>
</thead>
<tbody>
<tr>
<td>heffalump</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

- 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°(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°(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°(x).

- E.g. we learn the feature underwater°(e, x) and subsequently use our learned model to predict the value of underwater°(e, x) for heffalump°(x) (which, we hope, should be low).
Outline

1. Overview
2. Relating quantification and lexical semantics
3. Lexicalised Compositionality (again)
4. Quantification and truth
5. Quantification in ungrounded situations
6. Learning quantification
   - Learning over individuals
   - Pragmatic matters
7. Conclusion
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:

<table>
<thead>
<tr>
<th></th>
<th>bird</th>
<th>insect</th>
<th>mammal</th>
<th>domestic</th>
<th>graze</th>
<th>hibernate</th>
<th>layeggs</th>
<th>poisonous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>0.684</td>
<td>0.491</td>
<td>0.774</td>
<td>0.743</td>
<td>0.76</td>
<td>1</td>
</tr>
</tbody>
</table>
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 corresponding to elephant °.
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 corresponding to elephant.
### Results

- **Baseline results (repeated):**

<table>
<thead>
<tr>
<th>bird</th>
<th>insect</th>
<th>mammal</th>
<th>domestic</th>
<th>graze</th>
<th>hibernate</th>
<th>layeggs</th>
<th>poisonous</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.684</td>
<td>0.491</td>
<td>0.774</td>
<td>0.743</td>
<td>0.76</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Bootstrapped results (on test data):**

<table>
<thead>
<tr>
<th>bird</th>
<th>insect</th>
<th>mammal</th>
<th>domestic</th>
<th>graze</th>
<th>hibernate</th>
<th>layeggs</th>
<th>poisonous</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.951</td>
<td>0.491</td>
<td>0.774</td>
<td>0.743</td>
<td>0.951</td>
<td>1</td>
</tr>
</tbody>
</table>
Baseline classifier for \textit{be\_v+mammal\_n}

\begin{verbatim}
J48 unpruned tree
-------------

\texttt{genus\_n+of\_p()} \leq 0.159023
| \texttt{in\_p()}+\texttt{appearance\_n} \leq 0.059356
| | \texttt{be\_v+inhabitant\_n} \leq 0
| | | \texttt{kiss\_n+of\_p()} \leq 0: all (20.0/1.0)
| | | \texttt{kiss\_n+of\_p()} > 0: no (2.0)
| | \texttt{be\_v+inhabitant\_n} > 0: no (2.0)
| \texttt{in\_p()}+\texttt{appearance\_n} > 0.059356: no (4.0)
\texttt{genus\_n+of\_p()} > 0.159023: no (12.0)
\end{verbatim}
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

1. Overview
2. Relating quantification and lexical semantics
3. Lexicalised Compositionality (again)
4. Quantification and truth
5. Quantification in ungrounded situations
6. Learning quantification
   - Learning over individuals
   - Pragmatic matters
7. 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.