Where do models come from and where do they go?
Underspecifying set-theoretic semantics with vector spaces

Aurélie Herbelot

University of Trento
Centre for Mind/Brain Sciences

Düsseldorf 2016
Introduction
Model-theoretic semantics in two slides

Photo: Adrian Kingsley-Hughes, Flickr, CC by-nc-dd.
Model-theoretic semantics in two slides

Tarski: "snow is white" is true iff snow is white

Photo: Adrian Kingsley-Hughes, Flickr, CC by-nc-dd.
Model-theoretic semantics in two slides

- Set-theoretic interpretation.
- Very successful at modelling logical phenomena, from quantification to modality.
Where do models come from? Where do they live?
This talk: a FAQ

- **Question:** Do we have models in our heads?  
  *Answer:* probably not. If we do, they are very bad ones.

- **Question:** Where do models come from?  
  *Answer:* from distributional data.

- **Question:** And you call those models?  
  *Answer:* Uh... yes, an underspecified kind of model.

- **Question:** Can I still refer?  
  *Answer:* we hope.
Do we have models in our heads?
The psychology of quantifiers

- Children acquire quantifiers after generics (Hollander et al 2002).
- Children acquire numerical abilities (counting) after the Approximate Number Sense (ANS) (Mazzocco et al 2011).

"Who has more crayons?"

- Adults make quantification ‘mistakes’: (All) ducks lay eggs. (Leslie et al 2011).
Non-grounded quantification

- All cats are mammals.
  We had profiteroles for dessert (at the restaurant last night).
- In non-grounded quantification, it is often unclear what exactly the restrictor’s set consists of. E.g. no one knows the exact composition of the set of cats.
How do native speakers of English model relations between non-grounded sets?

Given the generic *Bats are blind*, how do humans quantify the statement? (*some*, *most*, *all* bats?)

Problem: explicit quantification cannot directly be studied from corpora, being rare in naturally occurring text (7% of all NPs – see Herbelot & Copestake 2011).
Quantifying the McRae norms

- The McRae norms (2005): a set of feature norms elicited from 725 human participants for 541 concepts.
- The dataset contains 7257 concept-feature pairs such as:
  - *airplane used-for-passengers*
  - *bear is-brown*
- ... quantified.
Three native English speakers (one Southeast-Asian and two American speakers, all computer science students.

For each concept-feature pair \((C, f)\) in the norms, provide a label expressing the ratio of instances of \(C\) having the feature \(f\).

Allowable labels: NO, FEW, SOME, MOST, ALL.

An additional label, KIND, for usages of the concept as a kind (e.g. beaver symbol-of-Canada).
Minimising quantifier pragmatics

The quantification of *bats are blind* depends on:
- the speaker’s beliefs about the concepts *bat* and *blind* (lexical semantics, world knowledge);
- their personal interpretation of quantifiers in context (pragmatics of quantifier use).

The meaning of the labels *NO, FEW, SOME, MOST, ALL* must be fixed (as much as possible!)
See annotation guidelines in paper.
Example annotations

<table>
<thead>
<tr>
<th>Concept</th>
<th>Feature</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ape</strong></td>
<td>is_muscular</td>
<td>ALL</td>
</tr>
<tr>
<td></td>
<td>is_wooly</td>
<td>MOST</td>
</tr>
<tr>
<td></td>
<td>lives_on_coasts</td>
<td>SOME</td>
</tr>
<tr>
<td></td>
<td>is_blind</td>
<td>FEW</td>
</tr>
<tr>
<td><strong>tricycle</strong></td>
<td>has_3_wheels</td>
<td>ALL</td>
</tr>
<tr>
<td></td>
<td>used_by_children</td>
<td>MOST</td>
</tr>
<tr>
<td></td>
<td>is_small</td>
<td>SOME</td>
</tr>
<tr>
<td></td>
<td>used_for_transportation</td>
<td>FEW</td>
</tr>
<tr>
<td></td>
<td>a_bike</td>
<td>NO</td>
</tr>
</tbody>
</table>

Table: Example annotations for McRae feature norms.

- Participants took 20 or less hours to complete the task, which they did at their own pace, in as many sessions as they wished.
Inter-annotator agreement

- We need an inter-annotator agreement measure that assumes separate distributions for all three coders.
- We would also like to account for the seriousness of the disagreements: a disagreement between NO and ALL should be penalised more than one between MOST and ALL.
- Weighted Kappa ($\kappa_w$, Cohen 1968) satisfies both requirements:

$$\kappa_w = 1 - \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} o_{ij}}{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} e_{ij}}$$ (1)
The weight matrix

- Weighted kappa requires a weight matrix to be set, to quantify disagreements.
- Setup 1: we use prevalence estimates from the work of Khemlani et al (2009) (after some mapping of their classification to ours).
- Setup 2: we exhaustively search the space of possible weights and report the highest agreement – under the assumption that more accurate prevalence estimates will result in higher agreement.
Prevalence estimates (Khemlani et al 2009)

<table>
<thead>
<tr>
<th>Predication type</th>
<th>Example</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principled</td>
<td>Dogs have tails</td>
<td>92%</td>
</tr>
<tr>
<td>Quasi-definitional</td>
<td>Triangles have three sides</td>
<td>92%</td>
</tr>
<tr>
<td>Majority</td>
<td>Cars have radios</td>
<td>70%</td>
</tr>
<tr>
<td>Minority characteristic</td>
<td>Lions have manes</td>
<td>64%</td>
</tr>
<tr>
<td>High-prevalence</td>
<td>Canadians are right-handed</td>
<td>60%</td>
</tr>
<tr>
<td>Striking</td>
<td>Pit bulls maul children</td>
<td>33%</td>
</tr>
<tr>
<td>Low-prevalence</td>
<td>Rooms are round</td>
<td>17%</td>
</tr>
<tr>
<td>False-as-existentials</td>
<td>Sharks have wings</td>
<td>5%</td>
</tr>
</tbody>
</table>

**Table**: Classes of generic statements with associated prevalence, as per Khemlani (2009).
### Results

<table>
<thead>
<tr>
<th></th>
<th>$\kappa_{12}^w$</th>
<th>$\kappa_{13}^w$</th>
<th>$\kappa_{23}^w$</th>
<th>$\kappa_{A}^w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>full</td>
<td>.37</td>
<td>.34</td>
<td>.50</td>
<td>.40</td>
</tr>
<tr>
<td>KH09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEST</td>
<td>.44</td>
<td>.40</td>
<td>.50</td>
<td>.45</td>
</tr>
<tr>
<td>maj</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KH09</td>
<td>.49</td>
<td>.48</td>
<td>.60</td>
<td>.52</td>
</tr>
<tr>
<td>BEST</td>
<td>.57</td>
<td>.53</td>
<td>.67</td>
<td>.59</td>
</tr>
</tbody>
</table>

**Table:** $\kappa_{w}$ for MCRAE\textsubscript{full} and MCRAE\textsubscript{maj}. Best estimates for exhaustive search are NO (0%), FEW (5%), SOME (35%), MOST (95%), ALL (100%)
**Per-feature agreement**

<table>
<thead>
<tr>
<th>BR Label</th>
<th>Example</th>
<th>Freq.</th>
<th>$\kappa_{12}^W$</th>
<th>$\kappa_{13}^W$</th>
<th>$\kappa_{23}^W$</th>
<th>$\kappa_A^W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>taxonomic</td>
<td>axe a_tool</td>
<td>713</td>
<td>.66</td>
<td>.48</td>
<td>.56</td>
<td>.57</td>
</tr>
<tr>
<td>visual-form</td>
<td>ball is_round</td>
<td>2330</td>
<td>.48</td>
<td>.44</td>
<td>.54</td>
<td>.49</td>
</tr>
<tr>
<td>function</td>
<td>hoe used_for_farming</td>
<td>1489</td>
<td>.36</td>
<td>.35</td>
<td>.50</td>
<td>.40</td>
</tr>
<tr>
<td>encyclopaedic</td>
<td>wasp builds_nests</td>
<td>1361</td>
<td>.39</td>
<td>.34</td>
<td>.37</td>
<td>.37</td>
</tr>
<tr>
<td>visual-colour</td>
<td>pen is_red</td>
<td>421</td>
<td>.44</td>
<td>.27</td>
<td>.30</td>
<td>.34</td>
</tr>
<tr>
<td>visual-motion</td>
<td>canoe floats</td>
<td>332</td>
<td>.28</td>
<td>.20</td>
<td>.46</td>
<td>.31</td>
</tr>
<tr>
<td>smell</td>
<td>skunk smells_bad</td>
<td>24</td>
<td>.34</td>
<td>.48</td>
<td>.12</td>
<td>.31</td>
</tr>
<tr>
<td>taste</td>
<td>pear tastes_sweet</td>
<td>84</td>
<td>.22</td>
<td>.29</td>
<td>.36</td>
<td>.29</td>
</tr>
<tr>
<td>tactile</td>
<td>toaster is _hot</td>
<td>242</td>
<td>.19</td>
<td>.31</td>
<td>.30</td>
<td>.27</td>
</tr>
<tr>
<td>sound</td>
<td>tuba is_loud</td>
<td>143</td>
<td>.11</td>
<td>.10</td>
<td>.36</td>
<td>.19</td>
</tr>
</tbody>
</table>

**Table**: Per-feature agreement for $\text{MCRAE}_{\text{full}}$, sorted by $\kappa_A^W$
General observations

- Substantial agreement on the majority test set: humans do have similar ‘models’ of the world (phew!)
- Even when features are reliably produced for a given concept, their quantification may vary significantly between annotators.
- Agreement is highly dependent on the corresponding functional or sensory type.
- No wonder children acquire generics before quantifiers...
- No wonder explicit quantification is infrequent (a cause for disagreements)...)
Where do models come from?

Herbelot & Vecchi (2015)
‘Meaning is use’.

DS is a general representation of the usages of a word. Akin to concept representation.
Where do models come from?

A state-of-the-art distributional cat (Baroni et al, 2014)

<table>
<thead>
<tr>
<th>Term</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>seussentennial</td>
<td>0.042</td>
</tr>
<tr>
<td>scaredy</td>
<td>0.041</td>
</tr>
<tr>
<td>saber-toothed</td>
<td>0.035</td>
</tr>
<tr>
<td>un-neutered</td>
<td>0.034</td>
</tr>
<tr>
<td>meow</td>
<td>0.034</td>
</tr>
<tr>
<td>unneutered</td>
<td>0.033</td>
</tr>
<tr>
<td>fanciers</td>
<td>0.033</td>
</tr>
<tr>
<td>pussy</td>
<td>0.033</td>
</tr>
<tr>
<td>pedigree</td>
<td>0.033</td>
</tr>
<tr>
<td>sabre-toothed</td>
<td>0.032</td>
</tr>
<tr>
<td>tabby</td>
<td>0.032</td>
</tr>
<tr>
<td>civet</td>
<td>0.032</td>
</tr>
<tr>
<td>redtail</td>
<td>0.032</td>
</tr>
<tr>
<td>meowing</td>
<td>0.032</td>
</tr>
<tr>
<td>felis</td>
<td>0.032</td>
</tr>
<tr>
<td>whiskers</td>
<td>0.032</td>
</tr>
<tr>
<td>silvestris</td>
<td>0.028</td>
</tr>
<tr>
<td>strangling</td>
<td>0.028</td>
</tr>
<tr>
<td>non-pedigree</td>
<td>0.029</td>
</tr>
<tr>
<td>sabertooth</td>
<td>0.029</td>
</tr>
<tr>
<td>woodpile</td>
<td>0.029</td>
</tr>
<tr>
<td>mewing</td>
<td>0.029</td>
</tr>
<tr>
<td>ragdoll</td>
<td>0.029</td>
</tr>
<tr>
<td>purring</td>
<td>0.029</td>
</tr>
<tr>
<td>whiskas</td>
<td>0.029</td>
</tr>
<tr>
<td>shorthair</td>
<td>0.029</td>
</tr>
<tr>
<td>scalded</td>
<td>0.029</td>
</tr>
<tr>
<td>retranslation</td>
<td>0.029</td>
</tr>
<tr>
<td>feral</td>
<td>0.029</td>
</tr>
<tr>
<td>whisker</td>
<td>0.028</td>
</tr>
<tr>
<td>silvestris</td>
<td>0.028</td>
</tr>
<tr>
<td>laziest</td>
<td>0.028</td>
</tr>
<tr>
<td>flap</td>
<td>0.028</td>
</tr>
<tr>
<td>purred</td>
<td>0.028</td>
</tr>
<tr>
<td>mummified</td>
<td>0.028</td>
</tr>
<tr>
<td>cryptozoological</td>
<td>0.028</td>
</tr>
<tr>
<td>english/french</td>
<td>0.030</td>
</tr>
<tr>
<td>straining</td>
<td>0.030</td>
</tr>
</tbody>
</table>

Herbelot, Aurélie (University of Trento)
Do cats have heads?

- grep "head" state-of-the-art-cat-distribution.txt
- 0.031179 headbutts
- 0.030823 flat-headed
- 0.016109 two-headed
- 0.009172 headless
- 0.002176 pilgrim
- 0.002176 out
- 0.002173 head
- 0.002169 merge
- 0.002165 idiot
Do cats have heads?

- grep "head" state-of-the-art-cat-distribution.txt
- 0.031179 head
- 0.030823 flat-headed
- 0.016109 two-headed
- 0.009172 headless
- 0.002176 pilgrim
- 0.002176 out
- 0.002173 head
- 0.002169 merge
- 0.002165 idiot
Where do models come from?

Do cats have heads?

- grep "head" state-of-the-art-cat-distribution.txt
- 0.031179 head butts
  0.030823 flat-headed
  0.016109 two-headed
  0.009172 headless
- 0.002176 pilgrim
  0.002176 out
  0.002173 head
  0.002169 merge
  0.002165 idiot
I picked some pears today. They’re really nice.

The reporters asked questions at the press conference.

The addax is a mammal.

[Pictures: CC by beautifulcataya, NASA and Zachi Evenor.]
A set-theoretic vector space
The context *meow* is very related to *cat*.
The context *sleep* is moderately related to *cat*.

Weight: how lexically characteristic a context is for a target.
Set-theoretic vector spaces

The attribute *has head* applies to **ALL** cats.
The attribute *is ginger* applies to **SOME** cats.
**Weight:** the set overlap between target and attribute.
### QMR: The McRae norms, quantified

<table>
<thead>
<tr>
<th>Concept</th>
<th>Feature</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ape</td>
<td>is_muscular</td>
<td>ALL</td>
</tr>
<tr>
<td></td>
<td>is_wooly</td>
<td>MOST</td>
</tr>
<tr>
<td></td>
<td>lives_on_coasts</td>
<td>SOME</td>
</tr>
<tr>
<td></td>
<td>is_blind</td>
<td>FEW</td>
</tr>
<tr>
<td>tricycle</td>
<td>has_3_wheels</td>
<td>ALL</td>
</tr>
<tr>
<td></td>
<td>used_by_children</td>
<td>MOST</td>
</tr>
<tr>
<td></td>
<td>is_small</td>
<td>SOME</td>
</tr>
<tr>
<td></td>
<td>used_for_transportation</td>
<td>FEW</td>
</tr>
</tbody>
</table>
## Axes and hatchets

<table>
<thead>
<tr>
<th><strong>axe</strong></th>
<th><strong>hatchet</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>a tool</td>
<td>a tool</td>
</tr>
<tr>
<td>is sharp</td>
<td>is sharp</td>
</tr>
<tr>
<td>has a handle</td>
<td>has a handle</td>
</tr>
<tr>
<td>used for cutting</td>
<td>used for cutting</td>
</tr>
<tr>
<td>has a metal blade</td>
<td>made of metal</td>
</tr>
<tr>
<td>a weapon</td>
<td>an axe</td>
</tr>
<tr>
<td>has a head</td>
<td>is small</td>
</tr>
<tr>
<td>used for chopping</td>
<td>–</td>
</tr>
<tr>
<td>has a blade</td>
<td>–</td>
</tr>
<tr>
<td>is dangerous</td>
<td>–</td>
</tr>
<tr>
<td>is heavy</td>
<td>–</td>
</tr>
<tr>
<td>used by lumberjacks</td>
<td>–</td>
</tr>
<tr>
<td>used for killing</td>
<td>–</td>
</tr>
</tbody>
</table>

- Inconsistencies in McRae.
- Ideally, each concept would be annotated against all features. That is $541 \times 2172 = 1,175,052$ annotations!
AD: The animal-only dataset

- Additional animal data from Herbelot (2013): a set of 72 animal concepts with quantification annotations along 54 features.
- Comprehensiveness of annotation: the 72 concepts were annotated along all 54 features. This ensures the availability of a large number of negatively quantified pairs (e.g. cat is-fish).
Both McRae and AD datasets are annotated with natural language quantifiers rather than set cardinality ratios, so we convert the annotation into a numerical format:

- ALL $\rightarrow$ 1
- MOST $\rightarrow$ 0.95
- SOME $\rightarrow$ 0.35
- FEW $\rightarrow$ 0.05
- NO $\rightarrow$ 0

These weights correspond to the best weighted kappa obtained for the McRae dataset (see H&V).
Converting annotated data into vectors

<table>
<thead>
<tr>
<th>Concept</th>
<th>Features</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>hatchet</td>
<td>an_axe, a_tool, has_a_handle, is_sharp, is_made_of_metal, is_used_for_cutting, is_small</td>
<td>ALL, ALL, ALL, MOST, MOST, MOST, SOME</td>
</tr>
</tbody>
</table>
## Converting annotated data into vectors

<table>
<thead>
<tr>
<th>Vector</th>
<th>Dimensions</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>hatchet</td>
<td>an_axe</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>a_tool</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>has_a_handle</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>is_sharp</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>is_made_of_metal</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>is_used_for_cutting</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>is_small</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>has_a_beak</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>taste_good</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Experiments
### Three configurations

<table>
<thead>
<tr>
<th>Space</th>
<th># train vec.</th>
<th># test vec.</th>
<th># dims</th>
<th># test inst.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{MT}_{QMR}$</td>
<td>400</td>
<td>141</td>
<td>2172</td>
<td>1570</td>
</tr>
<tr>
<td>$\text{MT}_{AD}$</td>
<td>60</td>
<td>12</td>
<td>54</td>
<td>648</td>
</tr>
<tr>
<td>$\text{MT}_{QMR+AD}$</td>
<td>410</td>
<td>145</td>
<td>2193</td>
<td>1595</td>
</tr>
</tbody>
</table>
Two distributional spaces:
- a co-occurrence based space ($\text{DS}_\text{cooc}$ – see paper for details);
- context-predicting vectors ($\text{DS}_\text{Mikolov}$) available as part of the
  word2vec project (Mikolov et al, 2013).

We learn a function $f : \text{DS} \rightarrow \text{MT}$ that transforms a distributional
semantic vector for a concept to its model-theoretic equivalent.

$f$: linear function. We estimate the coefficients of the function
using (multivariate) partial least squares regression (PLSR).
## Results

<table>
<thead>
<tr>
<th></th>
<th>Model-Theoretic</th>
<th>Distributional</th>
<th>human</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>train</strong></td>
<td><strong>test</strong></td>
<td><strong>DS_{cooc}</strong></td>
<td><strong>DS_{Mikolov}</strong></td>
</tr>
<tr>
<td>MT_{QMR}</td>
<td>MT_{QMR}</td>
<td>0.350</td>
<td>0.346</td>
</tr>
<tr>
<td>MT_{AD}</td>
<td>MT_{AD}</td>
<td><strong>0.641</strong></td>
<td>0.634</td>
</tr>
<tr>
<td>MT_{QMR+AD}</td>
<td>MT_{QMR+AD}</td>
<td>0.569</td>
<td>0.523</td>
</tr>
</tbody>
</table>

- Results for the QMR and AD dataset taken separately, as well as their concatenation.
- Performance on the domain-specific AD is very promising, at 0.641 correlation.
- Performance increases substantially when we train and test over the two datasets (MT_{QMR+AD}).
Results

<table>
<thead>
<tr>
<th>Model-Theoretic</th>
<th>Distributional</th>
<th>human</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>train</strong></td>
<td><strong>test</strong></td>
<td></td>
</tr>
<tr>
<td>$\text{MT}_{QMR+AD}$</td>
<td>$\text{MT}_{\text{animals}}$</td>
<td>0.663</td>
</tr>
<tr>
<td>$\text{MT}_{QMR+AD}$</td>
<td>$\text{MT}_{\text{no-animals}}$</td>
<td>0.353</td>
</tr>
<tr>
<td></td>
<td>$\text{DS}_{\text{cooc}}$</td>
<td>0.612</td>
</tr>
<tr>
<td></td>
<td>$\text{DS}_{\text{Mikolov}}$</td>
<td>0.341</td>
</tr>
<tr>
<td></td>
<td></td>
<td>–</td>
</tr>
</tbody>
</table>

- We investigate whether merging the datasets generally benefits all McRae concepts or just the animals.
- The result on the $\text{MT}_{\text{animals}}$ test set, which includes animals from the AD and the McRae datasets, shows that this category fares very well, at $\rho = 0.663$.
- No improvements for concepts of other classes.
We quantify the specific improvement to the McRae animal concepts by comparing the correlation obtained on the McRae animal features ($MT_{QMR^{animals}}$) after training on a) the McRae data alone and b) the merged dataset.

Performance increases from 0.419 to 0.666 on that specific set. This is in line with the inter-annotator agreement (0.663).
Underspecified Formal Semantics
A model-theoretic space? Really?

- seussentennial 0.042
- scaredy 0.041
- saber-toothed 0.035
- un-neutered 0.034
- meow 0.034
- unneutered 0.033
- pussy 0.033
- pedigreed 0.033
- sabre-toothed 0.032
- tabby 0.032
- civet 0.032
- redtail 0.032
- meowing 0.032
- felis 0.032
- whiskers 0.032
- morphosys 0.032
- meows 0.032
- scratches 0.032
- 1 walks
- 1 purrs
- 1 meows
- 1 has-eyes
- 1 has-a_heart
- 1 has-a_head
- 1 has-whiskers
- 1 has-paws
- 1 has-fur
- 1 has-claws
- 1 has-a_tail
- 1 has-4_legs
- 1 an-animal
- 1 a-mammal
- 1 a-feline
- 0.7 is-independent
- 0.7 eats-mice
- 0.7 is-carnivorous
- 0.3 is-domestic
- Herbelot, Aurélie (University of Trento)
- Where do models come from?
- Düsseldorf 2016
The Linkian semi-lattice

The distributional semi-lattice

\{F, K, S, B\}
\[4212\]

\{F, K, S\}
\[3201\]
\{F, K, B\}
\[3112\]
\{F, S, B\}
\[3212\]
\{K, S, B\}
\[3111\]

\{F, K\}
\[2101\]
\{F, S\}
\[2201\]
\{F, B\}
\[2112\]
\{K, S\}
\[2100\]
\{K, B\}
\[2011\]
\{S, B\}
\[2111\]

\{F\}
\[1101\]
\{K\}
\[1000\]
\{S\}
\[1100\]
\{B\}
\[1011\]

A semi-lattice of cats. Dimensions=[cat, black, striped, lazy]
In set theory, the union $A \cup B = A + B - A \cap B$.

Equivalent in distributional terms: $\overrightarrow{A} \cup \overrightarrow{B} = \overrightarrow{A} + \overrightarrow{B} - \overrightarrow{A \cap B}$.

Example:

$\overrightarrow{F} = [1101]$ \quad $\overrightarrow{F}, \overrightarrow{K} = [2101]$ \quad $\overrightarrow{F}, \overrightarrow{S} = [2201]$

$\overrightarrow{F, K, S}$
$= \overrightarrow{F, K} + \overrightarrow{F, S} - \overrightarrow{F}$
$= [2101] + [2201] - [1101]$
$= [3201]$

3 elements of the set have the property *cat*, 2 have the property *black*, etc.
What is $\rightarrow cat$ in the ‘model-theoretic’ space?

Underspecified version of the supremum (i.e. cardinality of the set is not known).

Semi-lattices can be generated from the distribution (see Erk 2016 for a distributional account of probabilistic world generation).

Example:

$\rightarrow cat = [1 \ 0.35 \ 0.35 \ 0.95]$

Generate a semi-lattice with, say, 100 cats.

Supremum of the generated lattice is $[100 \ 35 \ 35 \ 95]$. 
The underspecified semi-lattice

An underspecified semi-lattice of cats. Dimensions=[cat, black, striped, lazy]
Equivalent squiggles

\[ X = \sigma^* x \text{ cat}'(x) \land \exists Y [Y \prod X \land \text{black}(Y)] \]

There is the superset of cats, and a subset of that superset, and the cats in that subset are black. (Herbelot & Copestake 2011)
Models as generative processes

- The type of model we might acquire from distributional data is underspecified.
- Very often, specification is not necessary (see use of generics, see inter-speaker differences in views of the world.)
- *But* if needed, a standard model can be generated from the underspecified blueprint. This model will have all the properties normally attributable to a formal semantics model.
- Advantages of specification: do cardinals, talk about grounded/known situations, etc.
Conclusion: can I still refer???
Reference in underspecified formal semantics

- Correspondence theory in an underspecified model: muddy.
- Reference acts (Searle 1979): the generation of a referring expression. The reference act is successful if the hearer does not need to ask ‘which...?’
Referring expressions

- Referring expression generation (Krahmer & van Deemter, 2010).
- Minimal system: identify and order discriminatory properties for the referent.
Referring expressions

Instructions:
Try to guess what we're thinking about based on the hints we show you
After each hint, you can guess
if you aren't sure just hit Enter
Ready? To begin press Enter

New set of questions...
Next hint: clothing
dress
Next hint: is_warm
coat
Next hint: associated_with_Eskimos
parka
Congrats! You guessed correctly after only 3 (out of 9) hints!

Ready for another one? (Press Enter)
Models: where do they go?

Set-theoretic model

Run REG algorithm over set-theoretic model to produce new references

Referring expressions
  - observed
  - generated

Map distributional vectors to set-theoretic vectors

Distributional model

Use references in large corpora to create distributions for sets of entities