Collaborations with
Abhijeet Gupta\textsuperscript{1}
Marco Baroni\textsuperscript{2}
Gemma Boleda\textsuperscript{2}
Gabriella Lapesa\textsuperscript{1}
V Thejas\textsuperscript{3}
Matthijs Westera\textsuperscript{2}

\textsuperscript{1} University of Stuttgart
\textsuperscript{2} UPF Barcelona
\textsuperscript{3} BITS Pilani

Entities as a Window into (Distributional) Semantics

Sebastian Padó
Macron tells PM revised Brexit deal 'not an option'.

The leaders will meet hours after the French president insisted the Brexit deal could not be reopened.

UK POLITICS
**deal**  noun (2)

**Definition of deal (Entry 3 of 4)**

1  business
   a  : an act of dealing (see DEAL entry 2 sense 3)
      // big corporate deals
      // a real estate deal
   b  : BARGAIN
      // got a great deal on a new TV
      // accepted a plea deal
   c  : CONTRACT sense 1a
      // signed a 2-year deal

2 a  : the power or right to choose : freedom of choice
      // He has the option to cancel the deal.
   b  : a privilege of demanding fulfillment of a contract on any day within a specified time
   c  : a contract conveying a right to buy or sell designated securities, commodities, or property interest at a specified price during a stipulated period
      also : the right conveyed by an option
      // The ad is for a condo to rent with an option to buy.
   d  : a right of an insured person to choose the form in which payments d policy shall be made or applied

3  : something that may be chosen: such as
   a  : an alternative course of action
      // didn't have many options open
• **deal, option** are categories (concepts)
• Listed in dictionary

• **Macron, Brexit** are individual entities/events
• Listed in encyclopedia
Model-theoretic semantics

• Meaning of language units defined relative to world model (Gamut 1991: Universe U = set of individuals)

• Proper nouns and other entities:
  • Mapped onto elements of the universe

• Common nouns, adjectives, and other categories:
  • Mapped onto sets of elements of the universe
Model-theoretic semantics

- Meaning of language units defined relative to world model (Universe $U$ of individuals)
- Proper nouns and other entities:
  - Mapped onto elements of the universe $U$
- Common nouns, adjectives, and other categories:
  - Mapped onto sets of elements of the universe $U$

Entities and categories are fundamentally different

What about current NLP?
Distributional Semantics (DS)

• Dominant paradigm to acquire lexical information:
  • Learn linear algebra representations of linguistic units from context
  • A.k.a. Vector spaces, embeddings, distributed representations
  • Still DS because all use the “distributional hypothesis”: “You shall know a word by the company it keeps” (Firth, Harris, Miller & Charles 1991, etc.)
Distributional Semantics (DS)

• Dominant paradigm to acquire lexical information:
  • Learn linear algebra representations of linguistic units from context
  • A.k.a. Vector spaces, embeddings
  • Still DS because all use the "distributional hypothesis": “You shall know a word by the company it keeps” (Firth, Harris, Miller & Charles 1991, etc.)

How is this applied to categories / entities in NLP?

Split by subcommunity

deal
option
Brexit
Computational Lexical Semantics

- Strong focus on modelling linguistic aspects of meaning: categories and relations among categories
  - Hyponymy/hypernymy (entailment), synonymy, meronymy
  - Also diachronic change

“Interested in generalizations”
Semantic Web / Information Extraction

• Complementary focus on modelling world knowledge aspects of meaning: entities and relations among entities

• Knowledge bases / knowledge graphs

“Interested in particularities”
The Current Situation

• So Distributional Semantics is applied
  • to both entities and categories
  • to learn fairly different things
  • How is this possible?

• “It just works”
  • DS is a practice without a theory
Agenda for this presentation

• Q: Are there relevant differences in the way we can apply DS to modelling entities and categories?
• Research strand 1: Knowledge Bases
  • How far can we push DS in learning world knowledge?
• Research strand 2: The Instantiation Relation
  • How do categories and entities behave distributionally?

Benefit: insights into capabilities and limits of distributional approaches to meaning
Agenda for this presentation

• Q: Are there relevant differences in the way we can apply DS to modelling entities and categories?

• Research strand 1: Knowledge Bases
  • How far can we push DS in learning world knowledge?

• Research strand 2: The Instantiation Relation
  • How do categories and entities behave distributionally?

Benefit: insights into capabilities and limits of distributional approaches to meaning
Strand 1: Knowledge Base Completion

- Challenge: KBs are incomplete [Min et al. 2013, West et al. 2014]
- **Knowledge Base Completion (KBC)**: Add missing edges to knowledge graph
- Very active area of research
- Representation learning
- Learn embeddings for entities and relations
Entity Embeddings and KBC

- KBC embeddings can be learned from text, KB, or both
  - Our Interest: limits of distributional semantics
    - Focus on text-based embeddings of entities
- Entities have fine-grained attributes with specific values
- Research Question: Can all attributes be predicted from vanilla word embeddings? (And if not, why not?)

Italy
- Sunny 30
- Wine 15
- Beach 12
- Rome 10
- Naples 6

Italy
- Population: 61 million
- Area: 301,000 sq.km
- Language: Italian
- Contained by: Europe
- Currency used: Euro

RANLP, September 3, 2019
Simple Supervised KBC [Gupta et al. 15,17]

• Task: Use entity embeddings to predict entity attributes with Multi-Layer Perceptron (MLP)
  • Numeric: predict value(s)
  • Categorical: predict embedding for relatum (Italy, currency, Euro)

Our hypothesis is that intermediate values of $\sigma$ will improve prediction quality for the two types of attributes. We expect this to be the case since joint training can be seen as an instance of multi-task learning, which is known to often positively impact the quality of the learned intermediate representations (Zhang and Yang, 2017). Note that this effect is not guaranteed, since we introduce competition among the two output layers, which may deteriorate the output of the 'losing' layer.

2.4 Discussion
Note that all three model assume that all entities share a common set of attributes: in the numeric model, these determine the shape of the output layer, and in the categorical mode, they determine the shape of the attribute input layer. While it is still possible to train global models, CCKBs are typically organized into top-level domains that share little to no attributes. For example, people (which have, e.g., birth and death dates) or organizations (which have e.g., personnel, turnover, profit numbers) have no attributes in common. Consequently, in the remainder of the paper, we adopt a domain-specific approach, learning and evaluating separate models for each domain.

3 Experimental Setup
3.1 Dataset and Embeddings
To our knowledge, there are no existing datasets that include both numeric and categorical attributes. For example, the widely used FB15K and WN18 datasets (Bordes et al., 2013) focus exclusively on categorical attributes. For this reason, we construct our own dataset which we make freely available on DANS at URL https://doi.org/10.17026/dans-zxp-t7tf.

We construct the dataset on the basis of the FreeBase CCKB (Bollacker et al., 2008). As sketched above, we proceed by domains and extract entities and attributes for six of the most populous top-level FreeBase domains (animal, book, citytown, country, employer, organization, people). Since we build on pretrained embeddings for the entities in question, we only include entities if they are covered by the largest existing pretrained embedding space for proper names. This is the "Google News" embedding space that used a 100G token news corpus to compute embeddings specifically for FreeBase entities (Mikolov et al., 2013). The embeddings are computed with the Word2Vec 2

https://code.google.com/p/word2vec/
Evaluation of Attributes

• **Categorical** attributes: Mean Reciprocal Rank (MRR)
  • Mean rank of predicted relatum embedding among nearest neighbors of true relatum embedding

• **Numeric** attributes: Correlation
  • Spearman correlation between predicted and true rankings of entities w.r.t. attribute

(Leaving out details here; see papers)
Experimental Setup

- **Embeddings**: Google News vectors (Mikolov et al. 2013)
- **Word2Vec skipgram**, 300 dimensions
- **Experimental setup**: Train/Test on 7 FreeBase domains

| Domain      | # Entities (train/val/test) | |C| | |N| |
|-------------|-----------------------------|---|---|---|
| Animal      | 279/93/93                   | 22 | 118 |
| Book        | 16/5/6                      | 8  | 2  |
| Citytown    | 1783/594/595                | 57 | 62 |
| Country     | 155/53/51                   | 79 | 698|
| Employer    | 720/140/141                 | 50 | 53 |
| Organization| 187/63/62                   | 36 | 32 |
| People      | 85/28/29                    | 25 | 76 |
| **Sum**     | 3225/976/977                | 277| 1043|
**Experimental Setup**

- **Embeddings**: Google News vectors (Mikolov et al. 2013)
- **Word2Vec skipgram**, 300 dimensions

**Experimental Setup**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Entities (train/val/test)</th>
<th>C</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal</td>
<td>279/93/93</td>
<td>22</td>
<td>118</td>
</tr>
<tr>
<td>Book</td>
<td>16/5/6</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Citytown</td>
<td>1783/594/595</td>
<td>57</td>
<td>62</td>
</tr>
<tr>
<td>Country</td>
<td>155/53/51</td>
<td>79</td>
<td>698</td>
</tr>
<tr>
<td>Employer</td>
<td>720/140/141</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Organization</td>
<td>187/63/62</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>People</td>
<td>85/28/29</td>
<td>25</td>
<td>76</td>
</tr>
<tr>
<td>Sum</td>
<td>3225/976/977</td>
<td>277</td>
<td>1043</td>
</tr>
</tbody>
</table>

**Table 2: Data set statistics.**

- **C**: number of categorical attribute types.
- **N**: number of numeric attribute types.

**Figure 3: Hyperparameter exploration.** Impact of different values of $\alpha$ on the animal domain for categorical (solid) and numeric (dashed) attributes (validation set).

The results for the joint model are shown in Figure 3. As expected, there is a trade-off between the two objectives: Results for categorical prediction improve for high values of $\alpha$, where the model focuses on these attributes. Conversely, results for numeric prediction improve when the model pays more attention to these attributes, for low values of $\alpha$ (recall that lower NRS values are better). We chose $\alpha = 0.6$ as a value that gives both models a chance to profit from the joint setup.

**3.4 Inference**

Regarding inference in the models, the two individual prediction models are trivial, and so is the categorical part of the joint model: To predict the value of a categorical attribute for an entity, the numeric output can simply be ignored. To predict the value of a numeric attribute of entity, however, different inference procedures are possible. We used the simplest one, namely activating a random categorical attributes to query a numeric attribute (cf. Figure 2). We did not observe meaningful variance across the choice of different categorical attributes.

**3.5 Baselines**

We use two baseline models from previous studies. For categorical attributes, our baseline model ignores the entity. For each attribute, it predicts the frequency-ordered list of all values seen in the training set (Frequency Baseline). We also report on a baseline that simply models each attribute as a linear operation in embedding space (Mikolov et al., 2013; Bordes et al., 2013) defined as the centroid of all difference vectors for a given attribute between entities and their values for this attribute (Linear Baseline).

For numeric attributes, our baseline model predicts the mean value of the attribute seen in the training set (Mean Baseline).

**4 Results and Discussion**

**Numeric Attributes.** Table 3 shows the results as averaged normalized rank scores (NRS) for each domain as well as macro-averaged (Avg) scores for the complete test set. Recall that for NRS lower values are better.

We find that the joint model (which predicts numeric and categorical attributes at the same time) yields substantially better results than the individual model on all domains, ranging between 0.03 (for animal and people) and 0.1 (books). Average performance on all domains improves from 0.3 by 0.07 to 0.23. In turn, the individual model outperforms the baseline on all domains except people, corresponding to a similar improvement by 0.07. We see the best results of the joint model for city-town and the worst results for people. These numbers correlate with the numbers of entities present.
## Domain Country: Numeric Attributes

<table>
<thead>
<tr>
<th>Feature</th>
<th>Correlation of MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation (Lat. / Long.)</td>
<td>0.93</td>
</tr>
<tr>
<td>GDP_per_capita</td>
<td>0.89</td>
</tr>
<tr>
<td>CO2_emissions_per_capita</td>
<td>0.88</td>
</tr>
<tr>
<td>GDP_nominal</td>
<td>0.78</td>
</tr>
<tr>
<td>DateFounded</td>
<td>0.54</td>
</tr>
<tr>
<td>Religion_percentage</td>
<td>0.42</td>
</tr>
</tbody>
</table>

- Attributes differ greatly in difficulty
- Geographical attributes easy (Louwerse et al. 2009)
Geolocation: The Good

- A: Hong Kong
- B: Bangladesh
- C: Cocos Islands
- D: Eritrea
- E: Latvia
- F: Belarus
- G: Iran

Map showing actual and predicted locations.
Geolocation: The Bad

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>New Caledonia</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Cocos Islands</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Cook Islands</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Mauritius</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Niue</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Tuvalu</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Vanuatu</td>
<td></td>
</tr>
</tbody>
</table>

RANLP, September 3, 2019
### Domain Country: GDP

<table>
<thead>
<tr>
<th>Feature</th>
<th>Correlation of MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation (Lat. / Long.)</td>
<td>0.93</td>
</tr>
<tr>
<td>GDP_per_capita</td>
<td>0.89</td>
</tr>
<tr>
<td>CO2_emissions_per_capita</td>
<td>0.88</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>GDP_nominal</td>
<td>0.78</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Date_founded</td>
<td>0.54</td>
</tr>
<tr>
<td>Religion_percentage</td>
<td>0.42</td>
</tr>
</tbody>
</table>

- Even very similar attributes differ substantially (?)
### Domain Country: Difficult Attributes

<table>
<thead>
<tr>
<th>Feature</th>
<th>Correlation of MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geolocation (Lat. / Long.)</td>
<td>0.93</td>
</tr>
<tr>
<td>GDP_per_capita</td>
<td>0.89</td>
</tr>
<tr>
<td>CO2_emissions_per_capita</td>
<td>0.88</td>
</tr>
<tr>
<td>GDP_nominal</td>
<td>0.78</td>
</tr>
<tr>
<td>DateFounded</td>
<td>0.54</td>
</tr>
<tr>
<td>Religion_percentage</td>
<td>0.42</td>
</tr>
</tbody>
</table>

- The most difficult attributes appear to be very **specific**
Contextual Support

• Our KBC task = learn mappings from context-derived embedding space to attribute space

1. Attribute must correlate with prominent context cues
2. Entities with similar values of attribute must co-occur with similar context cues
Contextual Support

- Our KBC task = learn mappings from (BOW) embedding space to attribute space

1. Attribute must correlate with **prominent context cues**
2. Entities with **similar values** of attribute must co-occur with **similar context cues**

The extent to which this holds: **degree of contextual support**
Contextual Support Accounts for…

• The island displacement: “Hubness effect”
  • Predictions for sparse entities dominated by similar, more frequent entities

<table>
<thead>
<tr>
<th>A</th>
<th>New Caledonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Cocos Islands</td>
</tr>
<tr>
<td>C</td>
<td>Cook Islands</td>
</tr>
<tr>
<td>D</td>
<td>Mauritius</td>
</tr>
<tr>
<td>E</td>
<td>Niue</td>
</tr>
<tr>
<td>F</td>
<td>Tuvalu</td>
</tr>
<tr>
<td>G</td>
<td>Vanuatu</td>
</tr>
</tbody>
</table>
Contextual Support Accounts for...

- The island displacement: “Hubness effect”
- Predictions for sparse entities dominated by similar, more frequent entities

Context cues: Ocean, islands, palms, ...

Hubs with these contexts: Maldives, Seychelles, Mauritius
Contextual Support Accounts for

- **GDP_per_capita** being easier than **GDP_nominal**
- GDP per capita comes with **more consistent context cues**

<table>
<thead>
<tr>
<th>List of countries 2</th>
<th>GDP per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxembourg</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td></td>
</tr>
<tr>
<td>Qatar</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>List of countries 1</th>
<th>GDP nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td></td>
</tr>
</tbody>
</table>
Contextual Support Accounts for

- **GDP_per_capita** being easier than **GDP_nominal**
- GDP per capita comes with more consistent context cues

**List of countries 2**
- Luxembourg
- Switzerland
- Norway
- Ireland
- Iceland
- Qatar

**Context cues:** Luxury, finance, tax evasion, ...

**List of countries 1**
- United States
- China
- Japan
- Germany
- UK
- India

**Context cues:** ...

RANLP, September 3, 2019
Contextual Support Accounts for

• Difficulty of learning very specific attributes (date of foundation, countries exported to..)
  • Indicated by highly specific, low frequency context cues
  • “Drowned out” by other information in pretrained BOW vectors
  • Compare to pattern-based approach (Hearst 1992):

The modern state of Italy was created in the year 1861. In 1861, Italy was largely unified. The Kingdom of Italy was founded on this day in 1861.

Italy
Date founded : 1861
Area : 301,000 sq.km
Language : Italian
Contained by : Europe
Currency used: Euro
Take-home from Strand 1

• Knowledge can only be learned distributionally if has a substantial degree of contextual support

• Future directions:
  • Measuring / quantifying contextual support
  • Increasing contextual support
    • Fine-tuning on labeled data – not a panacea (?)
    • Present specific patterns to learner (Roller & Erk 2016)
  • Use meta-information about attribute-attribute relations

GDP_{per\_capita} = \frac{GDP_{nominal}}{\text{population}}
Agenda for this presentation

• Q: Are there relevant differences in the way we can apply DS to modelling entities and categories?

• Research strand 1: Knowledge Bases
  • How far can we push DS in learning world knowledge?

• Research strand 2: The Instantiation Relation
  • How do categories and entities behave distributionally?

Benefit: insights into capabilities and limits of distributional approaches to meaning
Strand 2: Instantiation

[Gupta et al. EACL 2017, ArXiv]

• We introduce a new semantic relation: instantiation

• Hypernymy -- relation between two categories
  [Baroni et al. 12, Roller et al. 14, Santus et al. 14, Levy et al. 15, etc.]

• Instantiation -- relation between entity and category
  • Many-to-many, not reflexive, not symmetrical, not transitive
An Instantiation Dataset

- 22k pairs: 5.5k positive pairs + 3* 5.5k negative pairs
  - Positive: Group entity with category
    - “Instance hypernym” relation from WordNet
  - Negative 1: INVERSE (switch entity and category)
  - Negative 2: INST2INST(entity + random other entity)
  - Negative 3: NOTINST (entity + wrong related category)

<table>
<thead>
<tr>
<th>Positive</th>
<th>Abraham Lincoln – POTUS</th>
<th>Mumbai – city</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverse</td>
<td>POTUS – Abraham Lincoln</td>
<td>city – Mumbai</td>
</tr>
<tr>
<td>Inst2Inst</td>
<td>Abraham Lincoln – Duncan Grant</td>
<td>Mumbai – Vicksburg</td>
</tr>
<tr>
<td>NotInst-inClass</td>
<td>Abraham Lincoln – doctor</td>
<td>Mumbai – residential area</td>
</tr>
</tbody>
</table>
Modeling Instantiation

- Architecture: let’s use an MLP again (1 hidden layer)
- Inspiration: hypernymy classifier (Roller et al. 14)
- Input: Embeddings for two words, e.g. $v = w_1 || w_2$
- Output: Binary decision (instantiation or not)

(We experimented with different variations)
Experimental Setup

• Own dataset
  • Train-dev-test split with memorization filtering
  • No entity or concept appears in more than one section

• Embeddings: Google News

• Baseline: Always predict instantiation

• Evaluation: F1 for class instantiation
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pos:Neg</th>
<th>Neg ex.</th>
<th>$\text{BL}_{\text{pos}}$</th>
<th>MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos + <em>Inverse</em></td>
<td>1:1</td>
<td>POTUS – Lincoln</td>
<td>0.67</td>
<td>0.96</td>
</tr>
<tr>
<td>Pos + <em>Inst2Inst</em></td>
<td>1:1</td>
<td>Lincoln – Grant</td>
<td>0.67</td>
<td>0.91</td>
</tr>
<tr>
<td>Pos + <em>NotInst</em></td>
<td>1:1</td>
<td>Lincoln – doctor</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>Pos + <em>Union</em></td>
<td>1:3</td>
<td><em>all</em></td>
<td>0.40</td>
<td>0.63</td>
</tr>
</tbody>
</table>

- Inverse and Inst2Inst are very simple
- Entities and categories are simple to distinguish
- NotInst is very difficult: hardly beats baseline!
- Corresponds well to findings about hypernymy
What makes NotInst hard?
Category representation

- Our assumption: Embedding of **noun x** is a good representation of **category x**
  - (Universally assumed in lexical semantic modeling)
- That is actually questionable:
  - Informativity
  - Ambiguity, including metaphors
  - Lexical choice and speaker intent (Lapesa et al. 2017, Westera and Boleda 2019)

Grass is green

Elephant in the room

Fotograf vs. Fotografin
  (generic/female photographer)
Re-representing Categories

• Are there alternatives for concept representation?

U

E. Macron

politician

B. Johnson

• Formal semantics: (extension of) concept = set of entities instantiating it

• In our context: Represent categories by the **centroid of their entity embeddings** („centroid embedding“)

• Vs. traditional approach: „concept embedding“
Does It Work?

Entity – Entity \hspace{1cm} \text{mean cos} = 0.22
Entity – Concept \hspace{1cm} \text{mean cos} = 0.16
Entity – Centroid \hspace{1cm} \text{mean cos} = 0.55
### Experimental Validation

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pos:Neg</th>
<th>BL&lt;sub&gt;Pos&lt;/sub&gt;</th>
<th>Concept emb.</th>
<th>Centroid emb.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos + <em>Inverse</em></td>
<td>1:1</td>
<td>0.67</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>Pos + <em>Inst2Inst</em></td>
<td>1:1</td>
<td>0.67</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>Pos + <em>NotInst</em></td>
<td>1:1</td>
<td>0.67</td>
<td>0.67</td>
<td>0.79</td>
</tr>
<tr>
<td>Pos + <em>Union</em></td>
<td>1:3</td>
<td>0.40</td>
<td>0.63</td>
<td>0.76</td>
</tr>
</tbody>
</table>

- Extension of previous experiment
- Centroids built on training set
- Improvement for centroid-based category representation
Take-home from Strand 2

- Categories and entities **differ in distributional behavior**
- Can be distinguished easily
- But capturing entity-category relations is tricky
  - Analogy to difficult attributes in Strand 1
- How to improve comparability?
  - Here: Centroid-based representation
    - Conceptually appealing
    - Requires more information about categories than just their names, namely instances

How many instances are needed? Sneak preview!
Wrap-up

• The Distributional Hypothesis – usage determines meaning – is at the heart of many NLP applications
  • But is it really true?
• My proposal today: Let’s relate the properties of information we want to learn to the properties of the linguistic material we want to learn it from
  • This presentation: Entities vs. categories
  • Other direction: Speaker intention vs. linguistic usage
Thank you!

Sebastian Padó

e-mail  pado@ims.uni-stuttgart.de
phone  +49 (0) 711 685-81394
www.ims.uni-stuttgart.de/~pado

University of Stuttgart