

# Modelling semantics

---

developing a  
cognitively plausible,  
data-driven approach

# Objective

---

- Develop a model of semantics that is wide-coverage, cognitively plausible and computationally useful
- Data-driven approach:
  - technically feasible, empirically grounded, scale, potential for practical utility
  - but linguistic and cognitive motivation?

# Semantics in computational linguistics

---

- Compositional semantics
  - `deep' grammars
  - shallow/intermediate grammars
- Lexical semantics
  - manually constructed ontologies: e.g., WordNet
  - data-driven: e.g., clustering
- Combined, data-driven approaches
  - Lin et al, Curran, Lapata
  - but surprisingly little work

# Integrated approaches

---

- Compositional semantics

the dog doesn't like peppermint

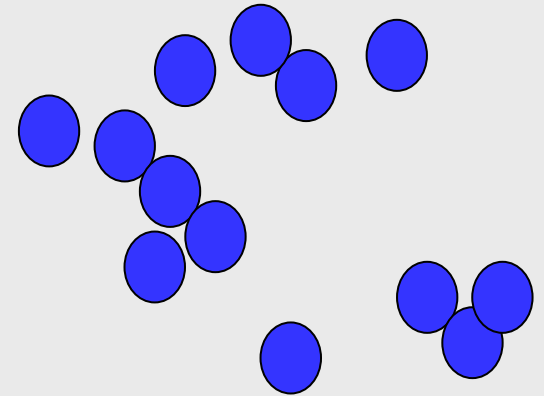
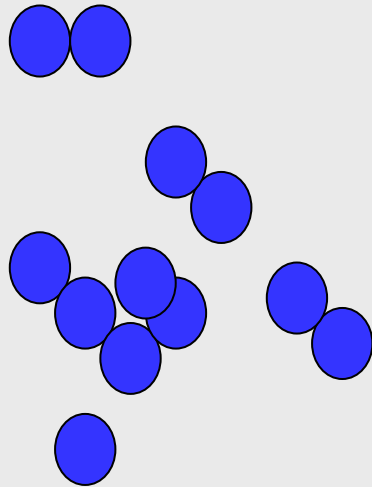
$the'(x, dog'(x), h1), not'(like'(e,x,y)),$

$bnpq(y, peppermint'(y), h2)$

- Open-class predicates correspond to region(s) in semantic `space'
  - peppermint' – unary predicate
  - like' – three regions – event, experiencer, stimulus

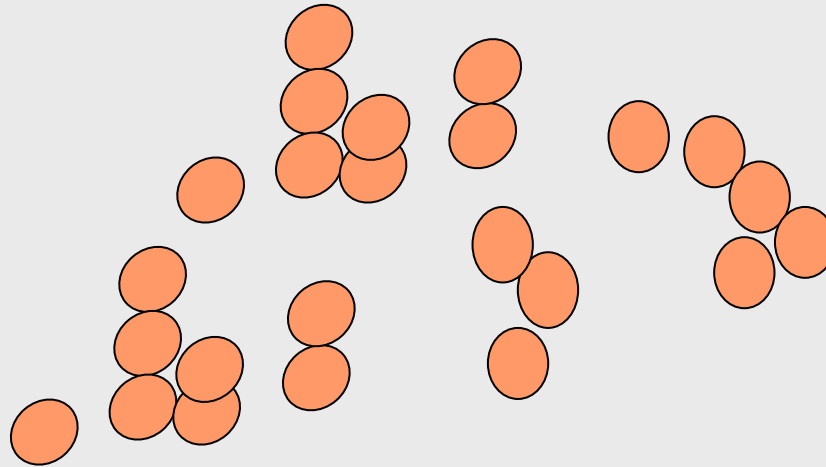
# Polysemy: *bank*

---



# Polysemy: *twist*

---



# Vector-space models from corpora

---

- Hypothesis: semantic space can be derived from textual context in corpora
  - Relationship to classical lexical semantics? polysemy, synonymy, antonymy, metonymy etc
  - Relationship to psycholinguistic experiments? Quantifiable predictions?
  - Task-based evaluation: word/phrase prediction?

# From distribution to semantics

---

- Robust morphological, syntactic and compositional semantic processing
- Iterated sense disambiguation with respect to derived soft clusters
- Document structure, anaphora resolution etc



# Some text corpora issues

---

- Spoken language vs written language
  - speech transcription, quantity of data, disfluencies etc
- Personal vs non-personal settings
  - shared context, background knowledge
- Individual experience: compare balanced and longitudinal corpora

# Summary

---

- Develop a model of semantics that is cognitively and linguistically plausible while practically tractable and useful
  - Exploit text corpora to provide scale
  - Exploit and further develop tools for large-scale text processing
  - Investigate how balanced corpora relate to individual experience
  - Evaluate against human experiments

# Potential participants include

---

- Cambridge: Copestake, Briscoe, Marslen-Wilson
- Sheffield: Lapata
- Edinburgh: Keller, Pickering