Modelling language: words

Angus Roberts, Senior Lecturer in Health Informatics
Institute of Psychiatry, Psychology and Neuroscience
King's College London

Representing words and context

- BoW and TFIDF typically model a piece of text e.g. sentences or documents
- But how can we model words numerically?
  - Vector based representations
- And how can we take in to account
  - Their similarities
  - Their meaning, or semantics
- Distributional semantics and context as meaning
  - Contrast with other approaches to semantics

Distributional semantics

Wombling and snetches

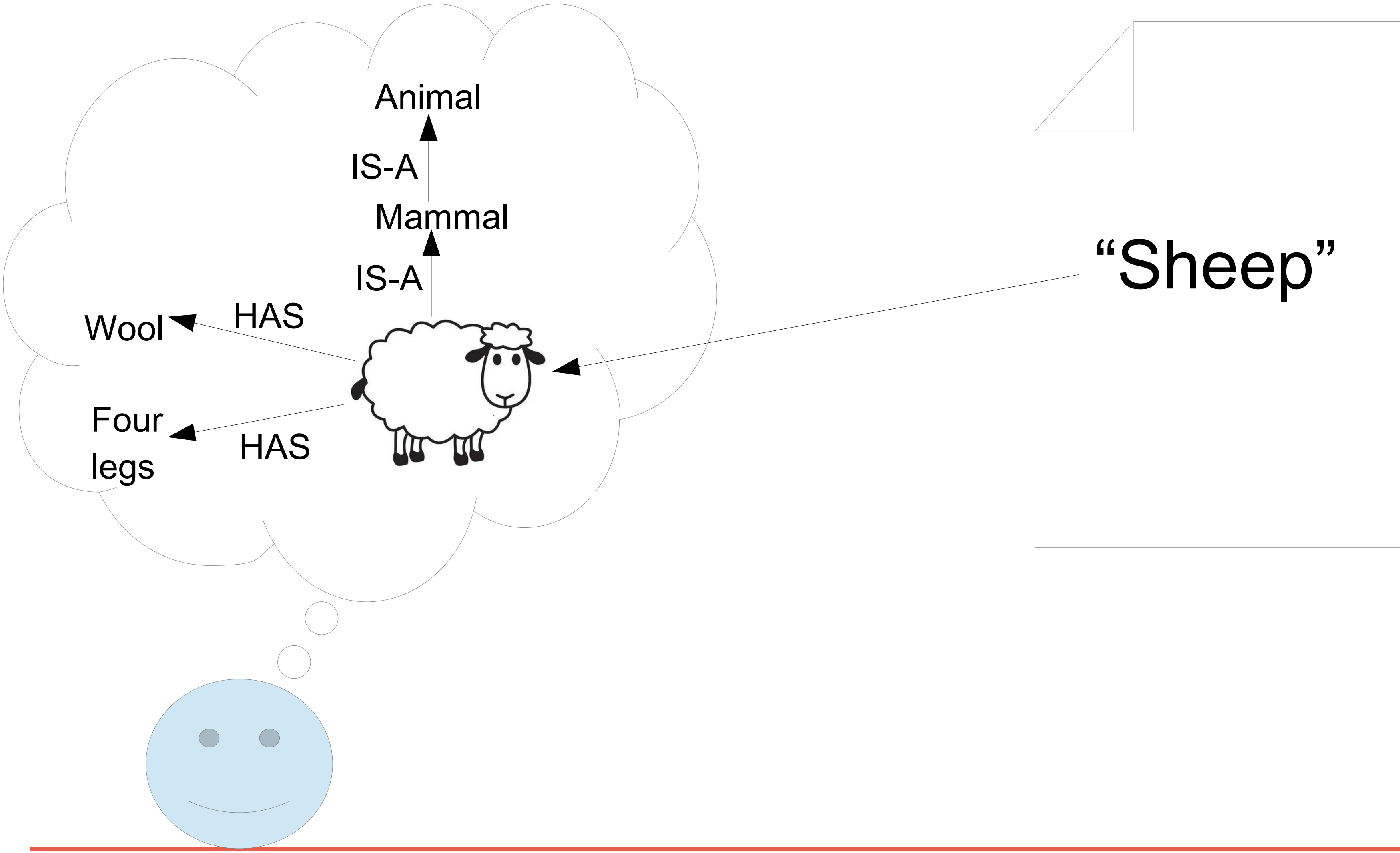
The Captain's side raked first. Tom staked. The hired sportsmen played so hard that they **wombled** too fast, and were shaky with the rakes. Tom fooled around the way he always did, and all his stakes dropped true. When it was his turn to rake he did not let Captain Najork and the hired sportsmen score a single rung, and at the end of the **snetch** he won by six ladders.

(How Tom beat Captain Najork and his hired sportsmen
Russell Hoban and Quentin Blake)

Distributional semantics

- “You shall know a word by the company it keeps” (Firth)
- The contexts in which words appear correlate with their meaning
- We understand a word by its distribution: the set of contexts in which it is found
- “Don't think, but look!” (Wittgenstein)
  - i.e. the meaning of a word is the description of its use,

Lexical semantics



Formal semantics

- A contrast to distributional semantics
- Formal semantics
  - models the relationship between language and the world in formal logical terms
  - defines meaning in terms of this model
- The lexicon is defined as mappings from words to structured, conceptual knowledge

Complementary

- Distributional semantics is based on **statistics**, formal semantics on **formal mathematics**
- Distributional semantics is **differential**, lexical semantics is **referential**
- Distributional semantics is based on large **corpora**, lexical semantics (more often) on **structured lexicons**
- **Gathering a corpus** is easier than **building a lexicon and grammar**

Lack of grounding

- One criticism of distributional approaches is that they lack grounding in real world knowledge
- Consider the task of finding semantic features for sheep – which of these approaches are grounded?
  - Generated by psychology students (McRae, 2005):
    - have four legs, say “Baah”, have wool
  - Generated from texts using a rule based approach (Barbu, 2009):
    - live on farms, graze, get scrapie
  - Words found close to “sheep” in Bing, via WebCorp (collocates):
    - society, wool, association, breed… … electric

References:
McCrae et al 2005, Semantic feature production norms for a large set of living and nonliving things, https://doi.org/10.3758/bf03192726
Barbu, 2010, Extracting conceptual structures from multiple sources, PhD thesis, University of Trento
http://www.webcorp.org.uk/live/

NIHR | Maudsley Biomedical Research Centre

Representing context

Collocations

```
VE: The authors compared the efficacy of olanzapine and lithium in the prevention of mood
nd received open-label co-treatment with olanzapine and lithium for 6-12 weeks. Those meet
in Pharmacokinet. 1999 Sep;37(3):177-93. Olanzapine. Pharmacokinetic and pharmacodynamic p
patients with schizophrenia confirm that olanzapine is a novel antipsychotic agent with br
d with traditional antipsychotic agents, olanzapine causes a lower incidence of extrapyram
urbation of prolactin levels. Generally, olanzapine is well tolerated. The pharmacokinetic
okers and men have a higher clearance of olanzapine than women and nonsmokers. After admin
rred between olanzapine and alcohol, and olanzapine and imipramine, implying that patients
:485-92. doi: 10.1192/bjp.bp.107.037903. Olanzapine for the treatment of borderline person
o evaluate treatment with variably dosed olanzapine in individuals with borderline persona
double-blind trial, individuals received olanzapine (2.5-20 mg/day; n=155) or placebo (n=1
rried-forward methodology. RESULTS: Both olanzapine and placebo groups showed significant
p. CONCLUSIONS: Individuals treated with olanzapine and placebo showed significant but not
he types of adverse events observed with olanzapine treatment appeared similar to those ob
is study compared three dosage ranges of olanzapine (5 +/- 2.5 mg/day [Olz-L], 10 +/- 2.5
```

- Collocates from www.webcorp.org.uk
- Restricted to *.ncbi.nlm.nih.gov (i.e. mostly PubMed abstracts)

Contexts as matrices

- Build matrices of event frequencies, where events are words in documents
  - **Row:** words (or terms)
  - **Column:** colocated words (or documents or sentences, or….)
- This is very similar to bag of words, with the bag represented as a row (vector)
- The row gives a “signature” of the word / term
- Sequential information is lost (at least in the simplest models)
- The matrix will be sparse

Word-Word matrices

	treatment	mg	anti-psychotic	placebo	patients
olanzapine	110	86	76	75	73

- Top 5 collocates for olanzapine
- Collocates four to the left and right, from www.webcorp.org.uk
- Restricted to *.ncbi.nlm.nih.gov (i.e. mostly PubMed abstracts)
- Normalised to collocates per 1000 hits

Word-Word matrices

	treatment	mg	anti-psychotic	placebo	patients
olanzapine	110	86	76	75	73
clozapine	70	30	78	0	89

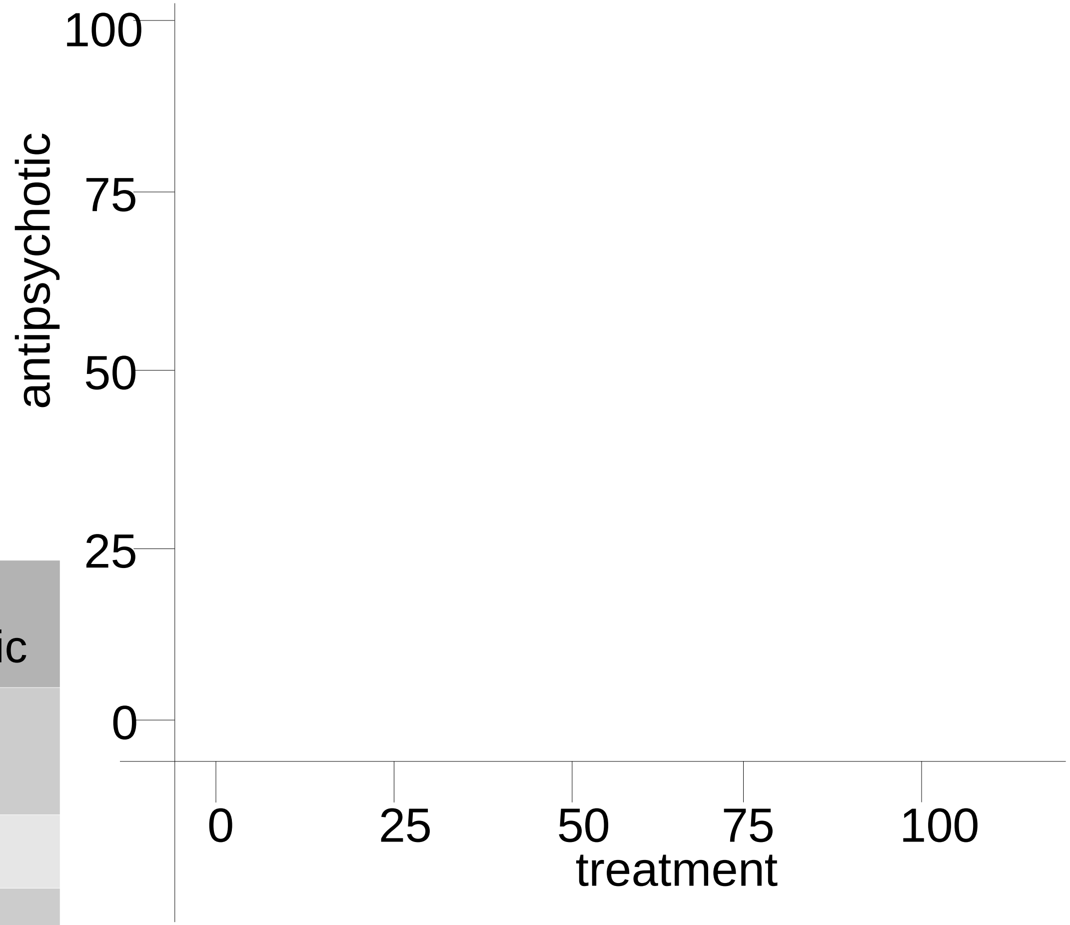
Word-Word matrices

	treatment	mg	anti-psychotic	placebo	patients
olanzapine	110	86	76	75	73
clozapine	70	30	78	0	89
vinegar	15	0	0	0	0

Word-Word matrices

	treatment	mg	anti-psychotic	placebo	patients	balsamic
olanzapine	110	86	76	75	73	0
clozapine	70	30	78	0	89	0
vinegar	15	0	0	0	0	109

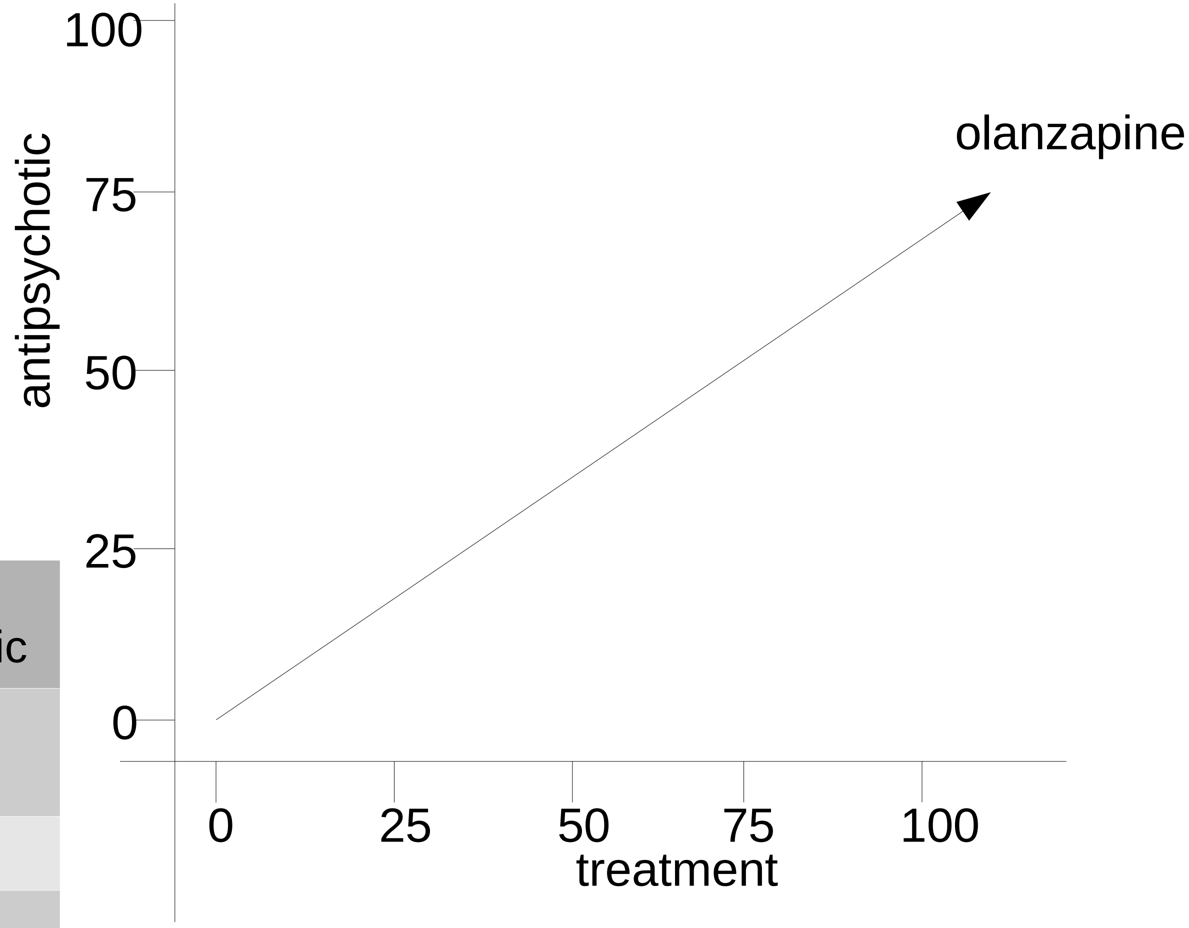
Semantic spaces



	treatment	anti-psychotic
olanzapine	110	76
clozapine	70	78
vinegar	15	0

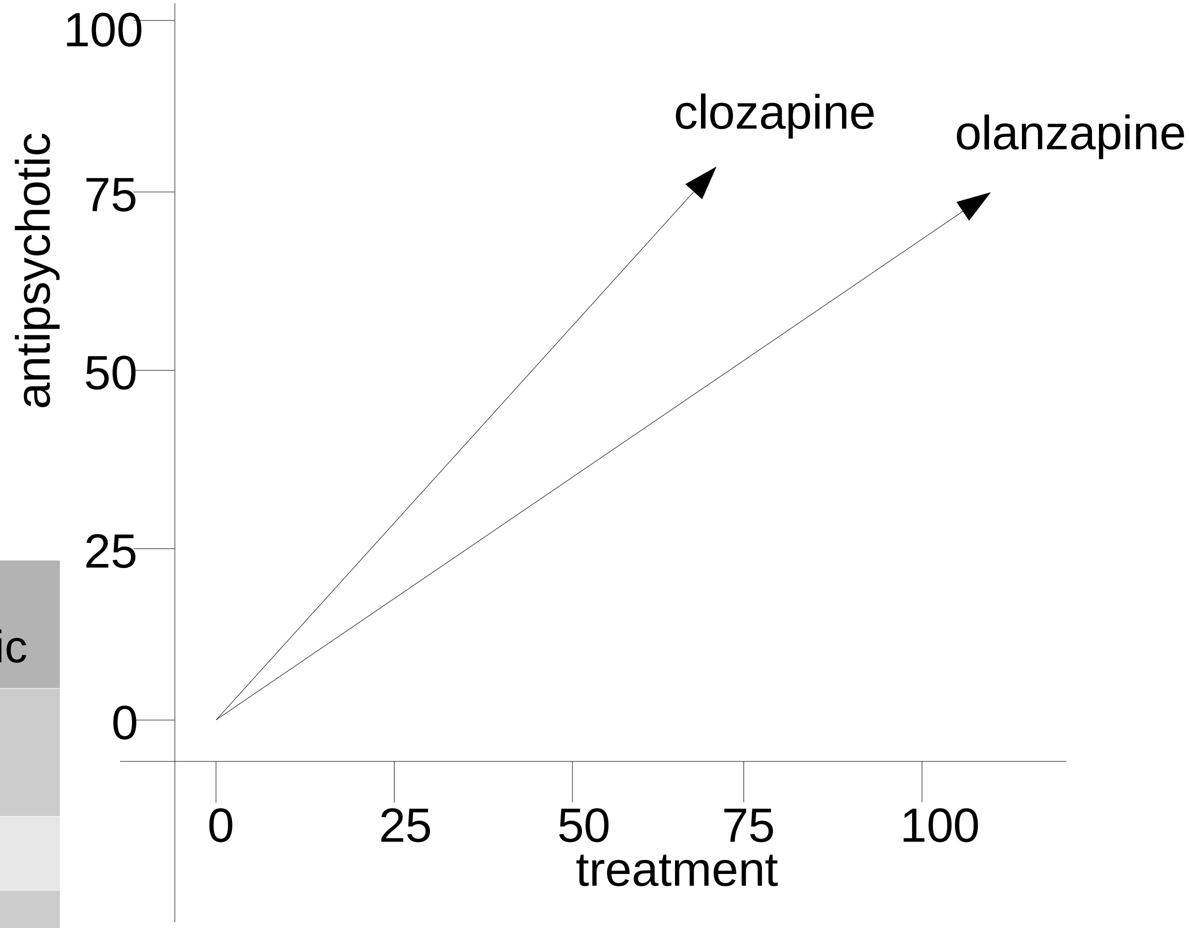
Semantic spaces

	treatment	anti-psychotic
olanzapine	110	76
clozapine	70	78
vinegar	15	0



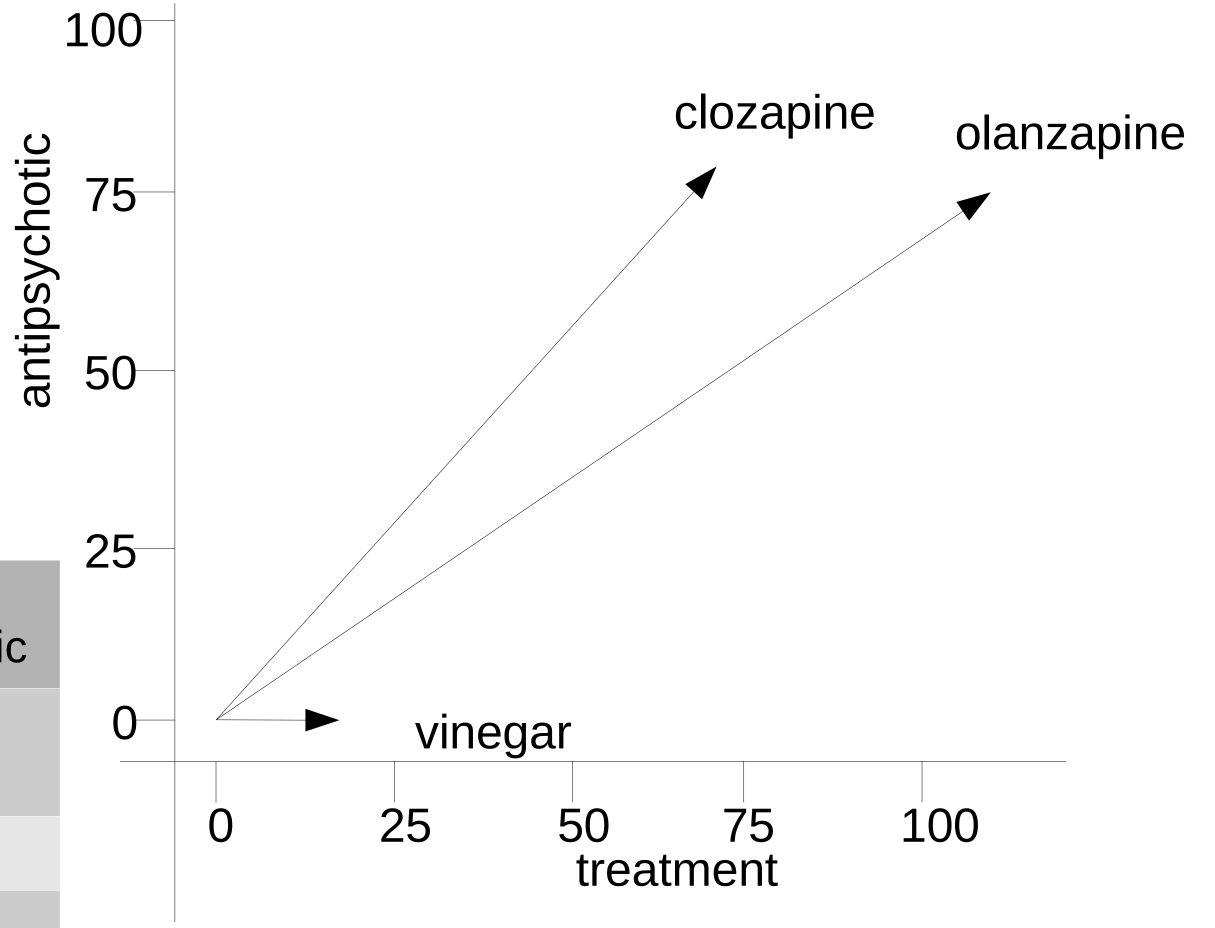
Semantic spaces

	treatment	anti-psychotic
olanzapine	110	76
clozapine	70	78
vinegar	15	0



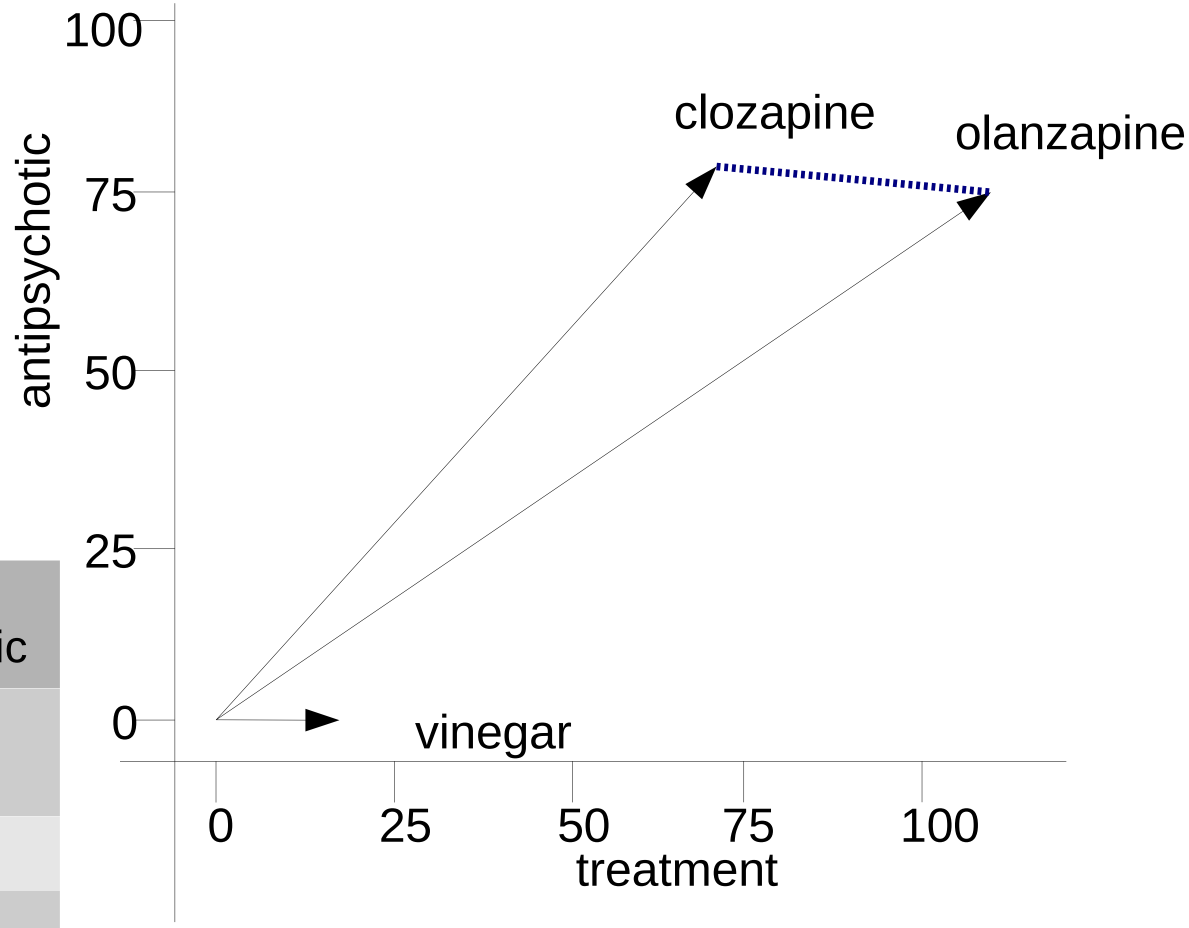
Semantic spaces

	treatment	anti-psychotic
olanzapine	110	76
clozapine	70	78
vinegar	15	0



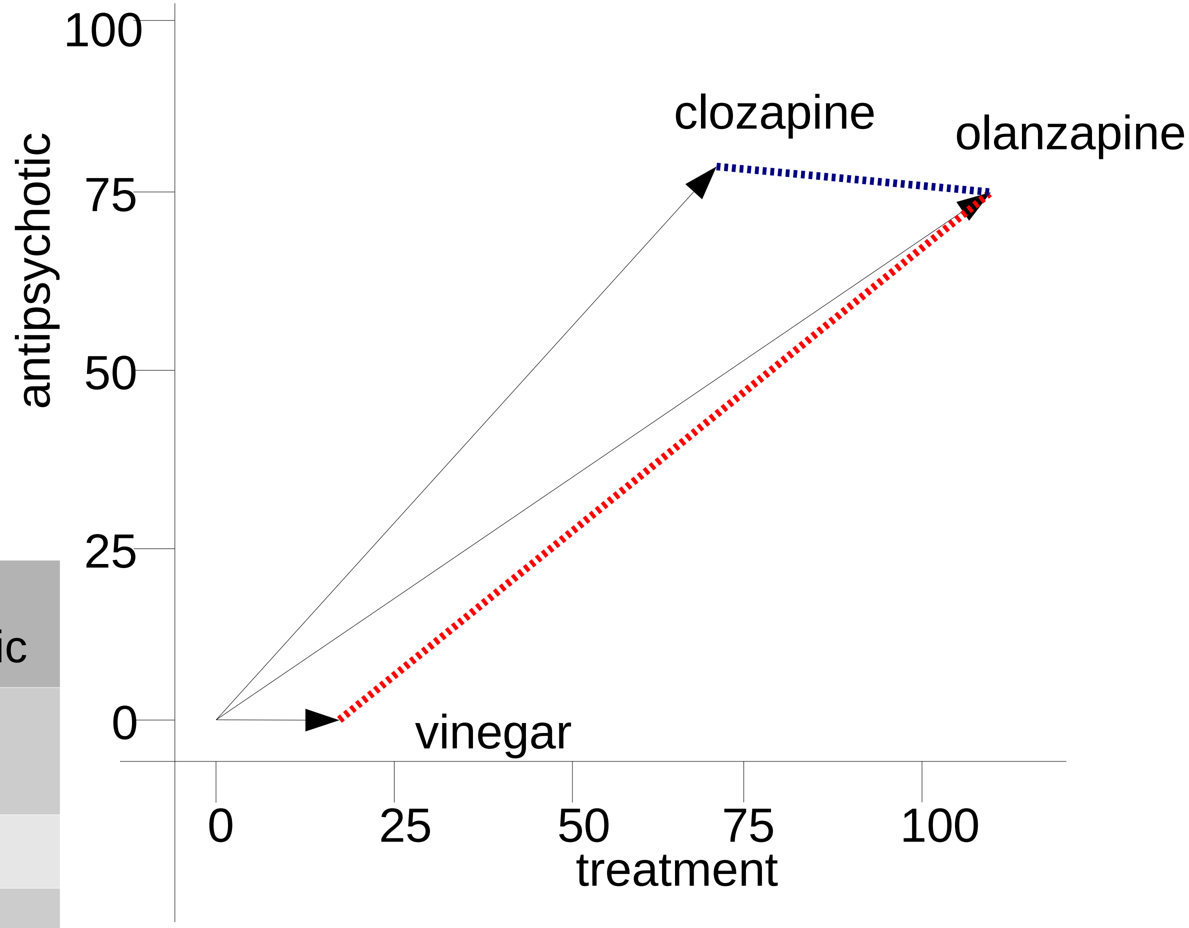
Semantic spaces

	treatment	anti-psychotic
olanzapine	110	76
clozapine	70	78
vinegar	15	0



Semantic spaces

	treatment	anti-psychotic
olanzapine	110	76
clozapine	70	78
vinegar	15	0



Semantic vectors

- Our vectors capture something about the context, and therefore the meaning, of words
- They could potentially be used to replace words in any NLP, so that instead of the word, we would manipulate or classify the vector representation of the word
- The simple result we have shown would create very sparse, crude vectors
- In reality, we use more sophisticated techniques

Tools for the job

- Rationalist NLP – facetiously referred to as armchair linguistics
  - Needs an armchair, within which to dream up examples and rules that match those examples
- Empirical NLP
  - Needs a pile of documents (corpus)
  - And a representation
- What about the algorithm to classify our words in representation space? Are they important?

NIHR | Maudsley Biomedical Research Centre

Thank you.
Any questions?

angus.roberts@kcl.ac.uk