Entities in Formal Distributional Semantics

Aurélie Herbelot

University of Trento
Centre for Mind/Brain Sciences

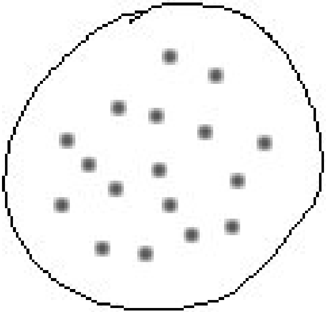
Geneva 2016

Introduction

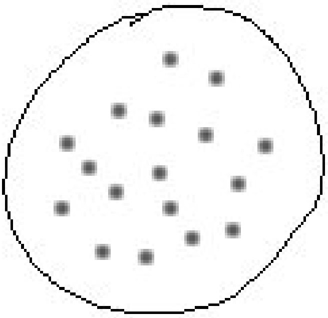
The meaning of life

$life'$

The meaning of life

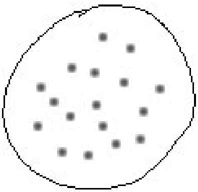


The meaning of cat



Distributional semantics

- A contentful representation of concepts.



Distributional semantics

- A contentful representation of concepts.

The (missing) dots in distributional semantics

Distributional semantics

The (missing) dots in distributional semantics

Distributional semantics

The (missing) dots in distributional semantics

Formal Distributional Semantics

A quick introduction to distributional semantics

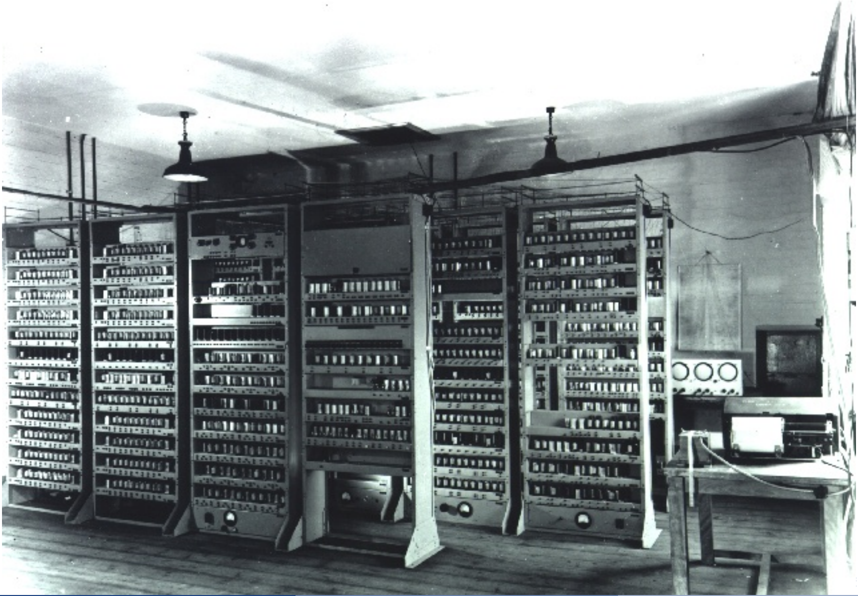
Distributional semantics: a short history

Ludwig Wittgenstein:
Meaning is defined by usage.

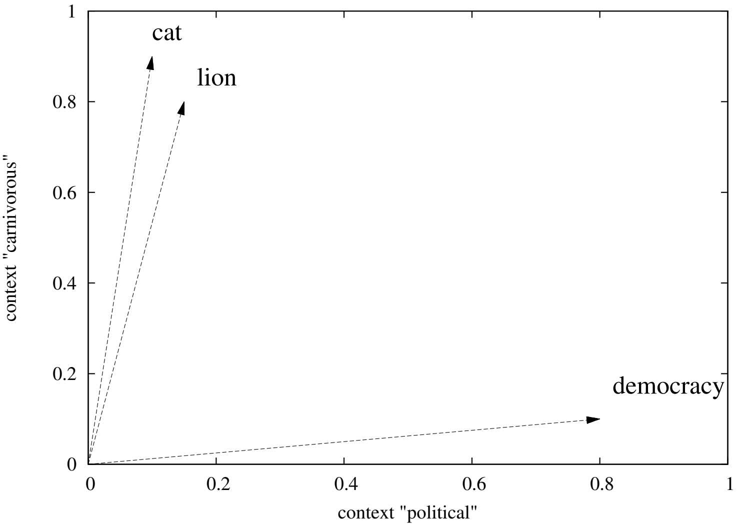
Margaret Masterman:
Cambridge Language Research Unit (CLRU: 1955–1986).

Karen Spärck-Jones: Early experiments on distributional semantics: 1963, 1967.

'The' computer: the EDSAC



The semantic space



A distributional cat

0.124 pet-N	0.074 tiger-N	0.063 hate-V
0.123 mouse-N	0.073 jump-V	0.063 asleep-A
0.099 rat-N	0.073 tom-N	0.063 stance-N
0.097 owner-N	0.073 fat-A	0.062 unfortunate-A
0.096 dog-N	0.071 spell-V	0.061 naked-A
0.092 domestic-A	0.071 companion-N	0.061 switch-V
0.090 wild-A	0.070 lion-N	0.061 encounter-V
0.090 duck-N	0.068 breed-V	0.061 creature-N
0.087 tail-N	0.068 signal-N	0.061 dominant-A
0.084 leap-V	0.067 bite-V	0.060 black-A
0.084 prey-N	0.067 spring-V	0.059 chocolate-N
0.083 breed-N	0.067 detect-V	0.058 giant-N
0.080 rabbit-N	0.067 bird-N	0.058 sensitive-A
0.078 female-A	0.066 friendly-A	0.058 canadian-A
0.075 fox-N	0.066 odour-N	0.058 toy-N
0.075 basket-N	0.066 hunting-N	0.058 milk-N
0.075 animal-N	0.066 ghost-N	0.057 human-N
0.074 ear-N	0.065 rub-V	0.057 devil-N
0.074 chase-V	0.064 predator-N	0.056 smell-N
0.074 smell-V	0.063 pig-N	

Distributional semantics in 2016

Linguistic representation: selectional preference, adjective semantics, quantifiers, phrasal composition, *meaning* of words.

Cognitive representation: simulates language acquisition, priming, fMRI measurements.

Useful hack: representation of the lexicon for NLP applications.

Meaning in DS

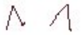
- A representation of *lexical* meaning?
- A representation of concepts? Of a certain type of genericity?
- *Not* a representation of the entities and events involved in a sentence.

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...



Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

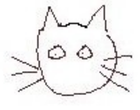
- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

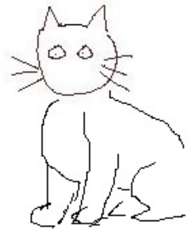


Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...



Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

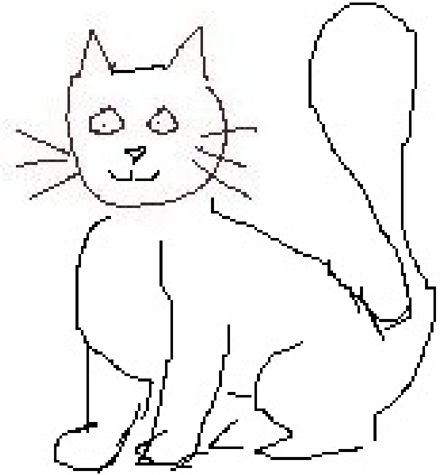
Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

A distributional cat (the theory)



A distributional cat (the reality)

Why does that cat look so bad?

Only 10% of NPs are references to kind

... an entirely black cat, like ...

... she owned a big ginger cat ...

... the cat was striped ...

... two long-haired white cats ...

... was a small grey cat ...

... cats are mammals ...

Just build vectors for individuals...

- Distributions are *statistical* representations.
  They need (enough) data.
- Yes, vectors can be built for individuals but...
- ... distributional semantics is a (lexical) semantics of
  - mid-frequency words
  - famous people with unambiguous names
  - capital cities...

A semantics without reference

- No notion of extension:
  - Barack Obama is on television.
  - The president of the United States is on television.
- Inference/entailment can only be done at a general level:
  - The cat sleeps –> The animal sleeps.
  - Sylvester sleeps –> Sylvester must be ill (because he would be chasing Tweety otherwise).
  - This is not what Wittgenstein or Masterman would have wanted...

Distributional individuals

Properties of individuals

1. **Uniqueness:** individuals should be separate from each other within the distributional space (two Smiths should occupy two different points in the semantic space).
2. **Instantiation:** individuals should stand in a learnable relationship to the concept they are instantiating (Mr Darcy should clearly be an instance of *man, person*, etc.)
3. **Individuality:** individuals should be distinguishable from concepts.

Uniqueness

- **Uniqueness** is satisfied by building the distribution over contexts which uniquely refer to an individual (use co-reference chain, or simply a unique person's name).

Instantiation

- **Instantiation** can be tested using hyponymy detection measures.
- Using invCL (Lenci & Benotto, 2012): a measure relying on feature inclusion. For a given individual, calculate invCL to its 50 nearest neighbours. Take highest scores as potentially instantiated concepts.

Individuality

- Tests from the literature on generics (difficult to implement):
  - Some predicates are only applicable to kind: **Mr Toad is widespread*.
  - Some predicates are infelicitous when applied to a kind: **Badgers are male* (‘positive alternatives’).
- An alternative test:
  - A concept may include extensionally exclusive contexts in at the top of its distribution (e.g. *rich* and *poor*).
  - The top of an individual’s distribution should be more coherent (i.e. a given man is either rich or poor).
  - Calculate coherence of top 50 characteristic contexts in a distribution:

$$Coherence(w_{1...n}) = mean\{Sim(w_i, w_j), ij \in 1...n, i < j\}$$

Implementation

Corpora

- Three corpora and their corresponding semantic spaces (BOW model):
  - the British National Corpus (BNC). 100M words. ‘General’ semantic space for English.
  - *Pride and prejudice* (*P&P*) by Jane Austen (1813), a novel of around 13,000 tokens.
  - *The wind in the willows* (*WitW*) by Kenneth Grahame (1908), a children’s novel of around 6000 tokens.

Standard individuals

Some name distributions

- Characteristic contexts of Mr Darcy and Mr Toad:

Darcy	(Mr, acquaint, Miss, interest, eye, Pemberley, degree, stand, wholly, walk, nephew, sake, civility, surprise)
Toad	(Hall, sternly, ho, paddock, smash, whack, popular, nonsense, trot, seize, disguise, cushion, terror, necessaries)

- Data sparsity problem (*Mr Denny* only appears 11 times in *P&P*).

Results: Instantiation - *P&P*

Darcy	Elizabeth	Bingley	Jane	...
0.47 **gentleman**	0.47 moment	0.48 **gentleman**	0.48 feeling	
0.47 word	0.46 subject	0.48 lady	0.47 **sister**	
0.46 manner	0.46 feeling	0.46 sister	0.46 pleasure	
0.46 feeling	0.46 pleasure	0.46 party	0.46 aunt	
0.46 conversation	0.45 house	0.46 answer	0.46 letter	

...	Collins	Denny
	0.50 daughter	0.47 news
	0.48 house	0.47 intention
	0.47 family	0.47 aunt
	0.46 **cousin**	0.45 journey
	0.46 lady	0.44 home

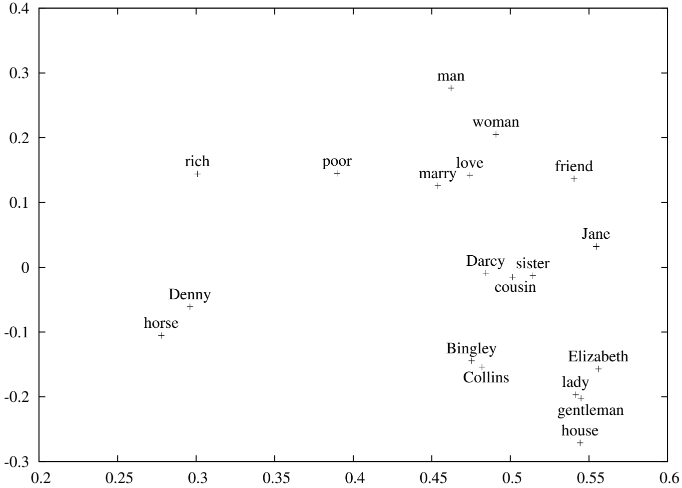
Results: Instantiation *WitW*

Toad	Rat	Mole	Badger
0.41 **animal**	0.40 water	0.40 **animal**	0.43 time
0.38 **toad**	0.39 **animal**	0.38 time	0.43 **animal**
0.38 time	0.39 time	0.37 thing	0.40 thing
0.37 way	0.37 thing	0.36 way	0.39 friend
0.36 thing	0.36 way	0.35 water	0.38 toad

Results: Individuality

	Ind. coh.	Kind coherence: woman	lady	man	gent.	toad	rat	mole	badger	animal
Darcy	0.22	-	-	0.24	0.25	-	-	-	-	-
Elizabeth	0.24	0.24	0.28	-	-	-	-	-	-	-
Bingley	0.23	-	-	0.24	0.25	-	-	-	-	-
Jane	0.23	0.24	0.28	-	-	-	-	-	-	-
Collins	0.22	-	-	0.24	0.25	-	-	-	-	-
Denny	0.23	-	-	0.24	0.25	-	-	-	-	-
Toad	0.21	-	-	-	-	0.24	-	-	-	0.22
Rat	0.23	-	-	-	-	-	0.24	-	-	0.22
Mole	0.22	-	-	-	-		-	0.22	-	0.22
Badger	0.21	-	-	-	-		-	-	0.23	0.22

The Pride & Prejudice semantic space



Individuals as contextualised kinds

Some remarks

- A reader does not come to *P&P* or *WitW* without some extensive prior knowledge of the lexicon (i.e. when *Mr* is used, what *men* and *toads* are, etc)
- Knowledge about new entities is built dynamically (as information about them is provided).
- Knowledge about the kind is preserved in the individual (Mr Darcy most probably has two feet).

Kind contextualisation

- Assumption: we have run NER on the corpus:
  - The characters in *P&P* have been classified as either *man* or *woman*;
  - The characters in *WitW* have trivially been recognised as members of the kind indicated by their names: Mr Toad is a *toad*, Mr Badger is a *badger*, etc.
- Take a concept and re-weight its distribution to produce an individual.

Balanced individuals

- The individuality property is at odds with instantiation. To make Darcy a distinguishable individual, more weight must be given to contexts that distinguish him from the concept, but doing so might 'break' his relation to the concept.
- Produce balanced individuals.

From concepts to individuals (1)

- N is a proper name, instance of kind K.
- N has a 'standard' distribution $\boldsymbol{v}(N)$, with m characteristic contexts $c_1 ... c_m \in C$.
- K also has a distribution $\boldsymbol{v}(K)$ which lives in a space S with dimensions $d_{1...n} \in D$, as obtained from the BNC.
- We define $\boldsymbol{v}(K)$ in terms of S's basis vectors $\{\boldsymbol{e}_{d'} | d' \in D\}$ and a weighting function w (in our case, PPMI):

$$\boldsymbol{v}(K) = \sum_{d' \in D} w(K, d').\boldsymbol{e}_{d'} \tag{1}$$

From concepts to individuals (2)

- We contextualise $\boldsymbol{v}(K)$ with respect to each of the characteristic contexts $c' \in C$ in $\boldsymbol{v}(N)$, using the following function (see Thater et al, 2011):

$$C(K, c') = \sum_{d' \in D} cos(c', d')^p w(K, d').\boldsymbol{e}_{d'} \tag{2}$$

- We introduce a weight p acting on the cosine function to increase or decrease the effect of the contextualisation: a higher p makes the individual more ‘unique’ but less like its kind.

From concepts to individuals (3)

- The name vector for N is the sum of the contextualisations with respect to all characteristic contexts in C:

$$\sum_{c' \in C} \sum_{d' \in D} cos(c', d')^p w(K, d').\boldsymbol{e}_{d'} \tag{3}$$

Results: instantiation

Darcy	Elizabeth	Bingley	Jane	Toad	Badger
0.97 **man**	0.97 **woman**	0.98 **man**	0.98 **woman**	0.97 **toad**	0.97 **badger**
0.91 girl	0.90 girl	0.91 boy	0.82 girl	0.75 sea	0.72 sight
0.91 face	0.89 eye	0.90 girl	0.82 man	0.74 desert	0.72 dog
0.91 boy	0.88 man	0.88 eye	0.81 other	0.73 rock	0.71 boy
0.90 smile	0.88 face	0.88 face	0.79 eye	0.73 mountain	0.71 fox

Table: Top invCL scores for various characters – contextualised individuals

Results: individuality

	Ind.	Kind coherence								
	coh.	woman	lady	man	gent.	toad	rat	mole	badger	animal
Darcy	**0.42**	-	-	0.24	0.25	-	-	-	-	-
Elizabeth	**0.40**	0.24	0.28	-	-	-	-	-	-	-
Bingley	**0.42**	-	-	0.24	0.25	-	-	-	-	-
Jane	**0.34**	0.24	0.28	-	-	-	-	-	-	-
Collins	**0.34**	-	-	0.24	0.25	-	-	-	-	-
Denny	**0.40**	-	-	0.24	0.25	-	-	-	-	-
Toad	**0.28**	-	-	-	-	0.24	-	-	-	0.22
Rat	**0.32**	-	-	-	-	-	0.24	-	-	0.22
Mole	**0.24**	-	-	-	-	-	-	0.22	-	0.22
Badger	**0.28**	-	-	-	-	-	-	-	0.23	0.22

From man to gentleman

man

0.120 gay_A
0.113 isle_N
0.112 woman_N
0.102 young_A
0.099 tall_A
0.097 wise_A
0.091 brave_A
0.089 handsome_A
0.088 suit_N
0.087 arrest_V
0.085 uniform_N
0.084 decent_A
0.082 married_A
0.078 trent_N
0.077 blind_A
0.076 armed_A
0.075 thin_A
0.074 murder_V
0.072 man_N
0.072 marry_V

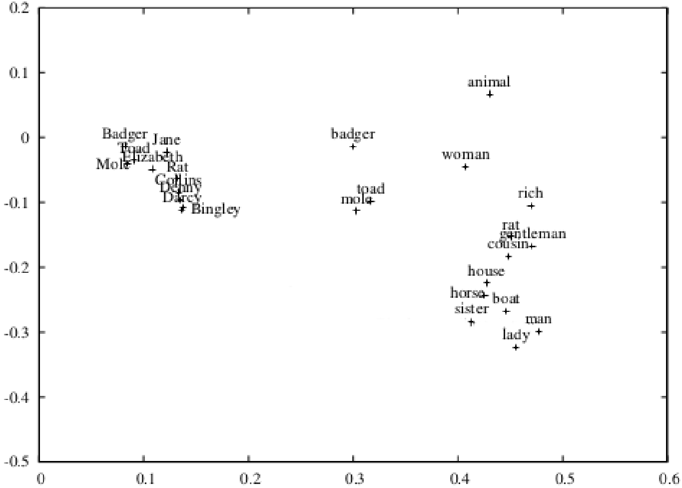
contexts of *Darcy*

$\longrightarrow$

Darcy

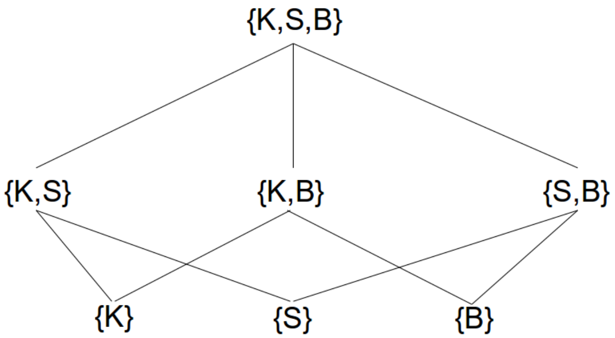
0.046 smile_V
0.045 stand_V
0.030 eye_N
0.030 sake_N
0.028 brother_N
0.025 smile_N
0.024 look_V
0.015 wholly_A
0.014 stare_V
0.012 face_N
0.011 look_N
0.010 appear_V
0.005 sit_V
0.004 walk_V
0.004 glance_V
0.004 surprise_N
0.004 laugh_V
0.003 lean_V
0.003 nod_V
0.003 kiss_V

Contextualised individuals in the BNC semantic space



Towards a Formal Distributional Semantics

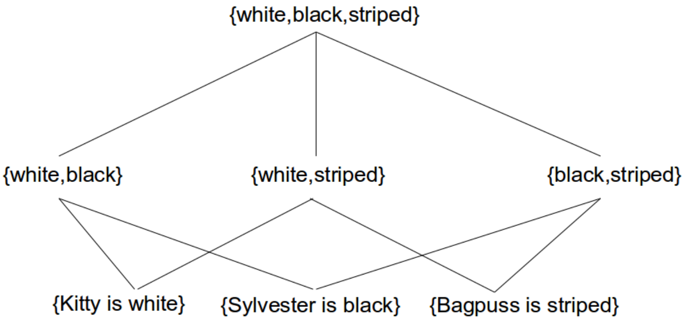
The Linkian semi-lattice



Distributions as suprema

- The distribution of a word is the aggregate of the contexts for many referents.
- Intuitively, a distributional 'supremum' is the top of a lattice where each node corresponds to the contexts for a (sub)set. (Let's ignore problems with collectives for the moment!)
- There is a straightforward relation between sums (nodes) in the semi-lattice: the sum S of two sums s_1 and s_2 is simply their addition $s_1 + s_2$. Or: the sum of two distributions is the distribution of their sums (in terms of frequency of occurrence!)

The distributional semi-lattice



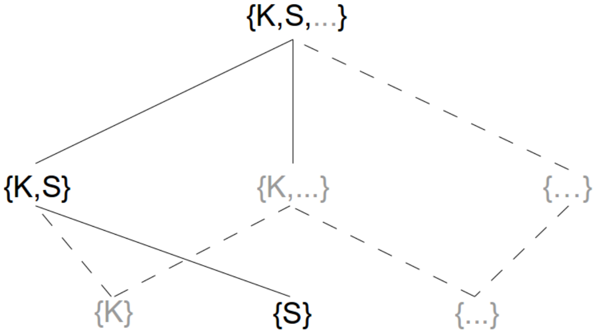
Reusing some formal semantics

- Linkian semantics can be ported to distributions.
- E.g. underspecified quantification (Herbelot & Copestake 2011):

  $\{K, S, B\} = \sigma^* x\ cat'(x) \wedge \exists Y[Y \prod X \wedge black(Y)]$
  (There is the superset of cats, and a subset of that superset, and the cats in that subset are black.)
- Equivalent in the distributional semi-lattice:

  $\overrightarrow{cat} = \sigma^* c\ cat(c) \wedge \exists Y[Y \prod X \wedge black(Y)]$
  (There is the distribution of 'cat' which is a superset of contexts, and there is a subset of that superset, and the contexts in that subset are 'black'.)

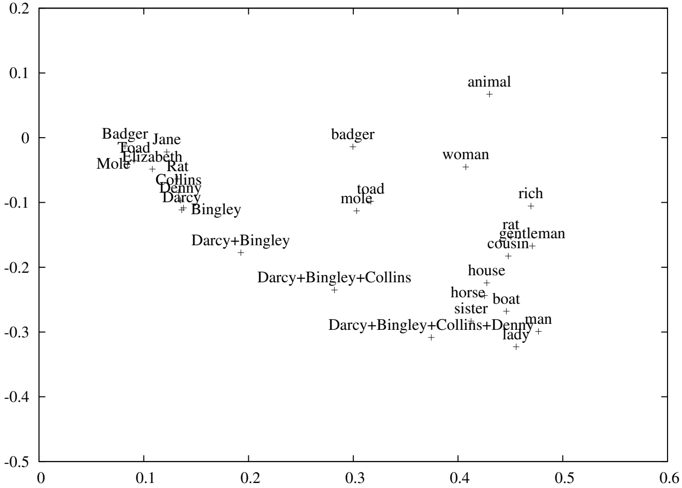
The underspecified semi-lattice



Issues with the distributional semi-lattice

- We know that the individuals at the bottom of the lattice will be poor representations of their referents (data sparsity).
- If using the contextualisation method, addition doesn't work cleanly anymore.
- (At least) two ways to go about this:
  - Don't be clean (it's distributional semantics after all...)
  - Rethink the way we conceptualise distributions

The dirty way: adding gentlemen



The clean way: rethinking distributions

- ‘Standard distributions’ give us information about the usages of a *word*. It is fine to be sparse.
- Distributions of referents: a word does not just activate the related concept but also its hypernyms. (See neurolinguistic results of Pulvermueller et al.)
- ‘Hallucinate’ contexts in the process of building distributions.

Conclusion

Conclusion

- Why distributional individuals?
  - Building blocks of a full semantics.
  - Psycholinguistic validity: fair to assume that some individuals (e.g. *my dog*, *Kim's boss*) might have distributions attached to them.
  - Make progress on issues related to antonymy, genericity, etc.
- Method applicable to common nouns in co-reference chain.
- Semantic space as dynamic system: effects of individuals over concepts.

Thank you!

Thank you!