Distributional semantics for linguists: 4

Ann Copestake and Aurélie Herbelot

Computer Laboratory, University of Cambridge
and
Department Linguistik, Universität Potsdam

August 2012

Generative lexicon and distributional semantics

Introduction to the Generative Lexicon

Contextual coercion

- GL account of contextual coercion
- Corpus-based approach to logical metonymy

English compound noun relations

Polysemy

Outline.

The Generative Lexicon (Pustejovsky, 1991, 1995)

- Polysemy is pervasive.
- Sense enumeration is not an adequate treatment.
- Several types of polysemy including: regular polysemy, constructional polysemy and sense modulation.
- Lexical semantic information in qualia structure (also argument structure, event structure and inheritance structure).
- Later work emphasizes dot object: not discussed here.

Qualia structure in GL

Certain aspects of meaning are highly salient:

- FORMAL ROLE
- CONSTITUTIVE ROLE
- TELIC ROLE (purpose)
- AGENTIVE ROLE

From Pustejovsky (1991):

```
novel(*x*)
  Const: narrative(*x*)
  Form: book(*x*), disk(*x*)
  Telic: read(T,y,*x*)
  Agentive: artifact(*x*), write(T,z,*x*)
```

Qualia structure in GL

- GL: qualia structure provides metonymic interpretations in a range of contexts.
- Also, perhaps, controls application of certain processes.
- Computational approaches to qualia: encode on lexical entries (via feature structures), default inheritance over a semantic hierarchy.
- Semi-automatic acquisition from Machine Readable Dictionaries.

GL and distributional semantics

Today: GL account and distributional experiments:

- Contextual coercion
- Compound nouns
- Regular polysemy

Encoding semantics: feature structures vs distributions?

Outline.

Contextual coercion

- After 6pm, most of the bars are full.
- After running for an hour, Kim was very thirsty.
- After the talk, we could go for a drink.
- After three martinis, Kim felt much happier.

Contextual coercion:

- After drinking three martinis, Kim felt much happier.

Adjectives

- fast runner: someone who runs fast
- fast typist: someone who (can) type fast
- fast car: car which can go fast

Not plausible that there are different senses of fast for each different context.

enjoy

- Mary enjoyed the book.
- Mary enjoyed reading the book.
- * Mary enjoyed that she read the book.
- ? Mary enjoyed the table.

‘Sylvie and Bruno Concluded’ (Lewis Carroll)

“You seem to enjoy that cake?” the Professor remarked.
“Doos that mean ‘munch?” Bruno whispered to Sylvie.
Sylvie nodded. “It means ‘to munch and ‘to like to munch.”
Bruno smiled at the Professor. “I doos enjoy it,” he said.
The Other Professor caught the word. “And I hope you’re enjoying yourself, little Man?” he enquired.
Bruno’s look of horror quite startled him. “No, indeed I aren’t!” he said.

Logical metonymy

Additional meaning systematically arises for some verb/noun or adjective/noun combinations:

- Kim enjoyed the cake
- Kim enjoyed eating the cake
- semantically, *enjoy*, *finish* etc always take an eventuality: different syntactic variants are thus closely related.
- contextual coercion of object-denoting NP to an appropriate eventuality

enjoy the cake would mean something like:

$$enjoy(e, x, e') \wedge cake(y) \wedge R(e', x, y)$$

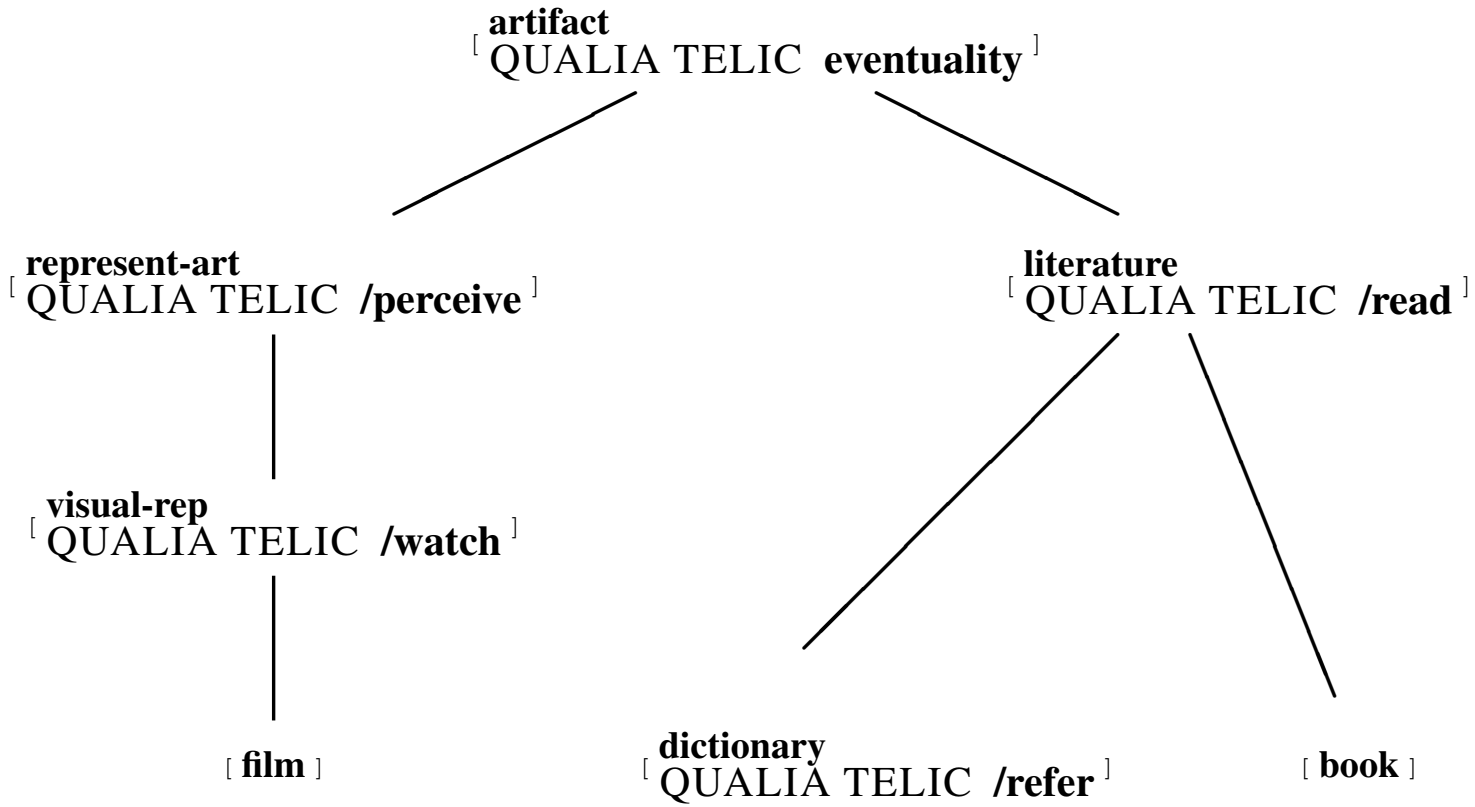
The metonymic event

- The default interpretation is supplied lexically
- A noun has qualia structure, e.g., telic role of *cake* is 'eat', agentive role is 'bake'

$$enjoy(e, x, e') \wedge cake(y) \wedge purpose(cake)(e', x, y)$$

where $purpose(cake) = eat$.

Qualia structure in feature structures



Interaction with pragmatics 1

Assume lexical defaults because of interpretation in unmarked contexts (Briscoe et al, 1990).

> *Willie enjoyed the hot sweet tea, standing on the deck in the cool of the night.*

Less common to find explicit verb when default is meant (e.g., don't tend to find *enjoy drinking the tea*).

But unusual verbs are specified.

Interaction with pragmatics 2

Defaults can be contextually overridden:

- John Grisham enjoyed that book.
- The goat enjoyed the book.

Override if Grisham is an author and goats don't read.
Similarly:

- After that book, Kim felt her knowledge of semantics had greatly improved.
- After that book, John Grisham became a household name.
- After that book, the goat had indigestion.

Formalisation in terms of typed default feature structures and non-monotonic logic (e.g., Lascarides and Copestake, 1998).

Context and adjectives

All the office personnel took part in the company sports day last week. One of the programmers was a good athlete, but the other was struggling to finish the courses. The fast programmer came first in the 100m.
cf Pollard and Sag discussion of *good linguist*

Logical metonymy restrictions

Logical metonymy has restrictions:

? Kim enjoyed the pebble.

? Kim enjoyed the dictionary.

Differences between verbs:

Kim began the salad. (eating, making)

Kim enjoyed the salad. (eating)

Kim began the book. (reading, writing)

Kim enjoyed the book. (reading)

Kim began the tunnel / bridge / path / road (constructing only)

(examples first pointed out by Godard and Jayez, 1993)

tunnels

Why is the tunnel example strange with a telic interpretation, when the explicit version is impeccable?

? Kim began the tunnel.

Kim began driving through the tunnel.

Context does not make this better:

The drive to the Alps had been long and tiring, and Kim was prone to claustrophobia.

*Therefore it was with considerable trepidation that Kim began the first tunnel.

and again . . .

tunnel can have a telic interpretation with other verbs:
But after the first tunnel, Kim felt much happier.
But much to his surprise, Kim enjoyed the first tunnel.
'telic' interpretation for *begin* in corpora seems to be almost completely restricted to foodstuffs, drinks and books (Verspoor, 1997)

Lapata and Lascarides (2003)

- corpus-based model that can retrieve plausible verbs for logical metonymy
- rough idea: given a metonymic verb (*begin, enjoy* etc) and a candidate object noun, find most frequent verbs that occur with the metonymic verb, and most frequent verbs that occur with the noun
- reasonable agreement with human judgements
- model is better than noun-only baseline
- alternative to qualia (arguably, even more lexical!)

Lapata and Lascarides (2003)

(9) a. Siegfried bustled in, muttered a greeting and began to pour his coffee.
b. She began to pour coffee.
c. Jenna began to serve the coffee.
d. Victor began dispensing coffee.

(10) a. I was given a good speaking part and enjoyed making the film.
b. He's enjoying making the film.
c. Courtenay enjoyed making the film.
d. I enjoy most music and enjoy watching good films.
e. Did you enjoy acting alongside Marlon Brando in the recent film The Freshman?

Lapata and Lascarides (2003): estimating probabilities

$P(e, o, v) = P(e).P(v|e).P(o|e, v)$
v = verb (enjoy), o = object, e = event
Maximum likelihood estimates via frequencies for $P(e)$ and $P(v|e)$, but not $P(o|e, v)$ because 'usual' verbs are not made explicit.
Assume $P(o|e, v) \approx P(o|e)$
i.e., count frequencies of verbs with objects regardless of enjoyment etc

Lapata and Lascarides (2003): some results

begin book - read (15.49) /write (15.52)
enjoy book - read (16.48) / write (16.48)
But: begin sandwich - bite into (18.12) / eat (18.23)
Adding in subject:
author book - write - 14.87
author book - read - 17.98
student book - read - 16.12
student book - write - 16.48
Correlation with human judgements: verbs r=.64, adjectives r=.40
Lower probabilities for weird examples. (but compare with infrequent good examples)?

Lapata and Lascarides (2003)

Problems:

- disambiguation: e.g., *fast plane*
- the model recovers individual verbs, but this is too specific:

  Kim enjoyed the soup

- sparse data (trained on BNC)
- titles etc:

  Sandy did not enjoy 'Sylvie and Bruno'

  back-off to semantic classes required in such cases
- not clear that it fully accounts for the semi-productivity facts model gives interpretations for 'enjoy the ice cream' but also 'begin the rock'

General distributional approach to logical metonymy?

- Still require a syntax-semantics interface component: distributions as a replacement for qualia, rather than whole of GL account.
- Ideally, want the metonymic interpretation to 'fall out' of a general distributional model of meaning.
  Distributional models with more complex feature spaces might allow this.
- Metonymic interpretation should be a distribution, approximately realizable as word(s)?
- How to allow for restrictions on *enjoy* etc, given that 'usual' verbs aren't explicit?
- Pragmatic overriding based on distribution associated with individual entity?

Outline.

Compound noun relations

- *cheese knife*: knife for cutting cheese
- *steel knife*: knife made of steel
- *kitchen knife*: knife characteristically used in the kitchen

Very limited syntactic/phonological cues in English, so assume parser gives: N1(x), N2(y), compound(x,y).

Overgeneration: German compounds with non-compound translations

Arzttermin	*doctor appointment
Terminvorschlag	* date proposal
Terminvereinbarung	* date agreement
Januarhälfte	* January half
Frühlingsanfang	* spring beginning

1. **doctor appointment*/*doctor's appointment* — possessive compounds
2. **date agreement*/*agreement on a date* — head derived from a PP-taking verb (*water seeker* vs **water looker*)
3. **spring beginning*/*beginning of spring* — relational nouns as heads?

GL approach

Johnston and Busa (1996)

- *bread knife*: telic interpretation. Italian *da* (*coltello da pane*)
- *lemon juice*: origin interpretation. Italian *di* (*succo di limone*)
- *glass door*: constitutive interpretation. Italian *a* (*porta a vetri*)

Problems:

- Only applicable to a limited number of English compounds.
- How are readings selected?

Data-driven approaches to compound relation learning

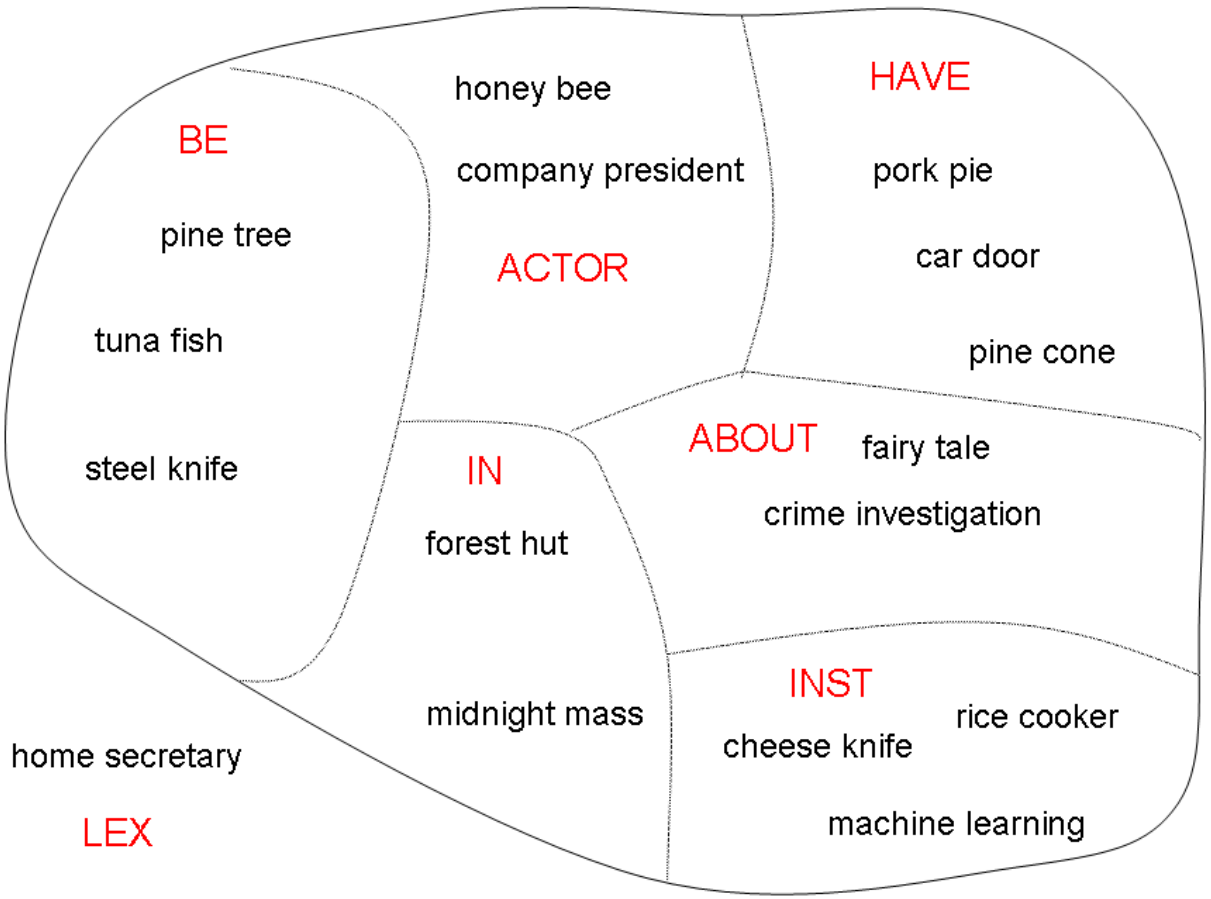
- Find paraphrases by looking for explicit relationships (Lauer: prepositions, Lapata: verbal compounds)
- OR human annotation of compounds, use distributional techniques to compare unseen to seen examples. Girju et al, Turner, Ó Séaghdha among others.

Relation schemes for learning experiments

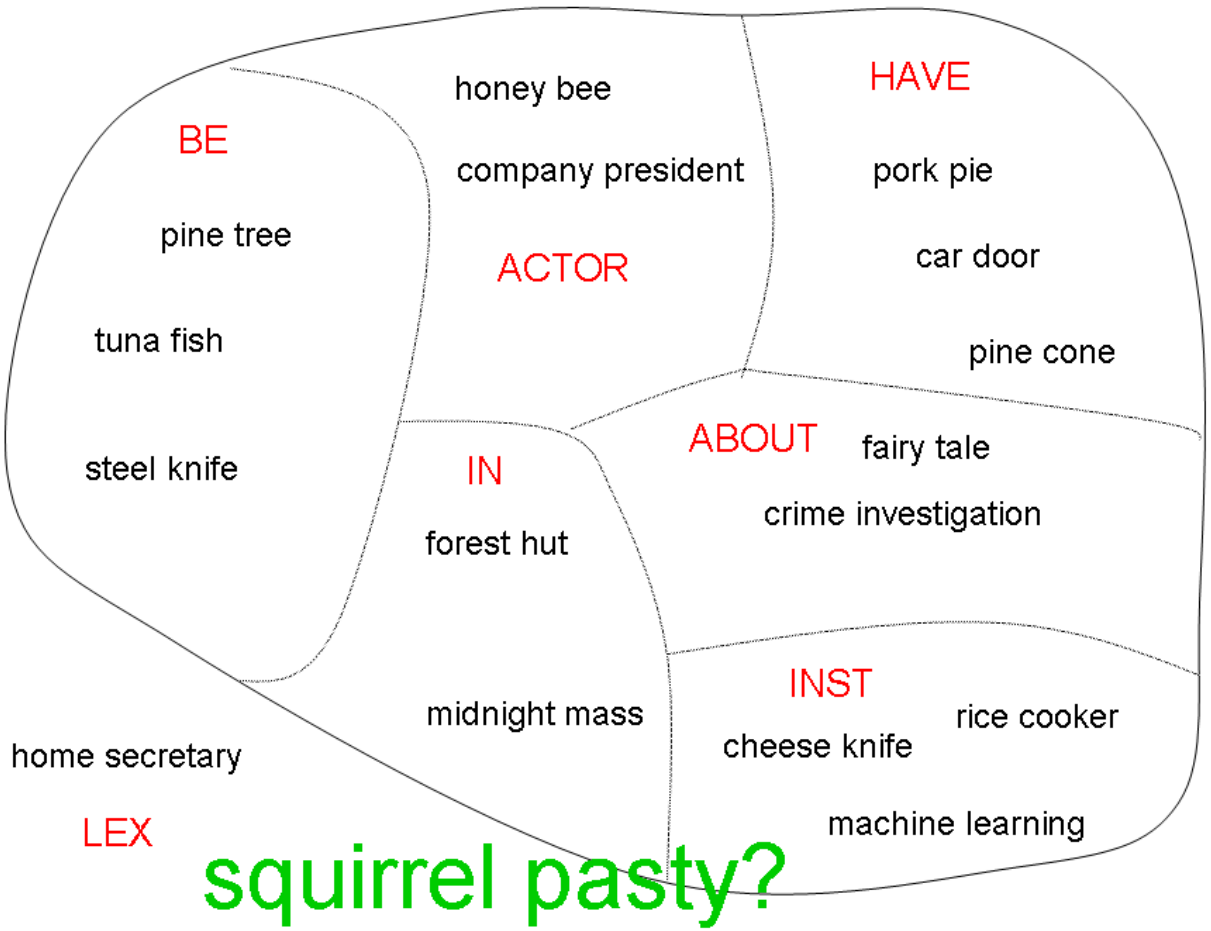
Ó Séaghdha (2007):

- BE, HAVE, INST, ACTOR, IN, ABOUT: (with subclasses) LEX: lexicalised, REL: weird, MISTAG: not a noun compound.
  - Based on Levi (1978)
  - Considerable experimentation to define a usable scheme: some classes very rare (therefore not annotated reliably)
  - Annotation of 1400 examples from BNC by two annotators.

Compound noun relation learning



Compound noun relation learning



Squirrels and pasties

Compound noun relation learning

- Ó Séaghdha, 2008 (also Ó Séaghdha and Copestake, forthcoming)
- Treat compounds as single words: doesn't work!
- Constituent similarity: compounds x1 x2 and y1 y2, compare x1 vs y1 and x2 vs y2.
  squirrel vs *pork*, *pasty* vs *pie*
- Relational similarity: **sentences** with x1 and x2 vs sentences with y1 and y2.
  squirrel is very tasty, especially in a pasty vs
  pies are filled with tasty pork
- Comparison using kernel methods: allows combination of kernels.
- Best accuracy: about 65% (slightly lower than agreement between annotators) using combined kernels.

Analogical reasoning using distributions

- Using distributional similarity to match known cases: a type of analogical reasoning.
- Known examples explicitly annotated (this approach to compounds) or based on observation (adjectives and binomials).
- Relatively sophisticated techniques allow combination of evidence types (Ó Séaghdha's use of kernel methods).
- Explicit relations could be thought of as a label for distributions?

Outline.

Polysemy and regular polysemy

- homonymy — unrelated meanings — *bank* (river bank) vs *bank* (financial institution)
- general polysemy — related meanings but no systematic connection — *bank* (financial institution) vs *bank* (in a casino)
- regular polysemy — regularly related meaning — *bank* (N) (financial institution) vs *bank* (V) (put money in a bank), compare *store*, *cache* etc
- vague/general terms — e.g. *teacher* may be male or female (in English, other languages may distinguish)

Some types of regular polysemy/sense extension

- Count/mass: animal/meat, tree/wood . . . (generically thing/derived substance grinding)
  After several lorries had run over the body, there was rabbit splattered all over the road.
- Mass/count: portions , kinds.
  two beers: 'two servings of beer' or 'two types of beer'
- Verb alternations: causative/inchoative . . .
- Noun-verb conversions: *sugar*, *hammer*, *tango* (cf derivational endings *-ize*)

Established senses tend to have additional conventional meaning.

More regular polysemy/sense extension

- Container-contents: *bottle*
  He drank a bottle of whisky
  paralleled by suffixation with -ful
  He drank a bottleful of whisky
- Plant/fruit: *olive, grapefuit* (cf *aceituna/aceituno, pomela/pomelo*)
- Broadening of senses (maybe):
  cloud: normal sense is weather, but also e.g., *cloud of dust*
  forest, *bank* . . .
- Figure/ground (maybe):
  Kim painted the door. vs *Kim walked through the door.*
  ? Kim painted the door but got paint on herself when she walked through it.

Systematic polysemy and translation

- If similar polysemy patterns: no need to disambiguate.
- Marked: disambiguate but straightforward translation.
  Plant/fruit examples: grapefuit (pomela/pomelo).
- Translation mismatch due to differences in systematic polysemy patterns.
  - hammer a nail into a frame
  - enfoncer un clou dans un cadre avec un marteau
    Literally: drive a nail into a frame with a hammer
  - mettre un clou dans un cadre avec un marteau
    Literally: put a nail into a frame with a hammer

Metonymy: country names

- Location: About 300 Australians will remain inside Iraq on logistical and air surveillance duties.
- Government: The US and Libya have agreed to work together to resolve compensation claims . . .
- Teams: England are comfortable 3-0 winners in their end-of-season friendly against Trinidad and Tobago. (football)
  Stand-in England boss Rob Andrew said Sunday's laboured win over the Barbarians was a "useful" exercise. (rugby union)

Metonymy: object to person

- *The cello is playing badly.*
  (the person playing the cello)
- *His Dad was a Red Beret.*
  (i.e., someone who wore a red beret: here a British paratrooper)
- *Chester serves not just country folk, but farming, suburban and city folk too. You'll see Armani drifting into the Grosvenor Hotel's exclusive (but exquisite) Arkle Restaurant and C&A giggling out of its streetfront brasserie next door.*
  (i.e., people who wear Armani/C&A clothes)

Nunberg (1978)

Restaurant context:

- The ham sandwich wants his check
  (meaning ‘person who ordered a ham sandwich’)
- The french fries wants his check
- * The ham sandwich wants a coke and has gone stale
- * The brown suit is in the microwave

Compare:

- I’m parked out back.

GL accounts of regular polysemy

- Regular polysemy often involves syntactic effects.
- Lexical rules in a feature structure framework can capture syntax and semantics.
- This requires a generative account of the derived meaning.
- Application of rules has to be controlled: probabilities and productivity metrics (but practical problems with deriving these from corpora).

Metonymy disambiguation

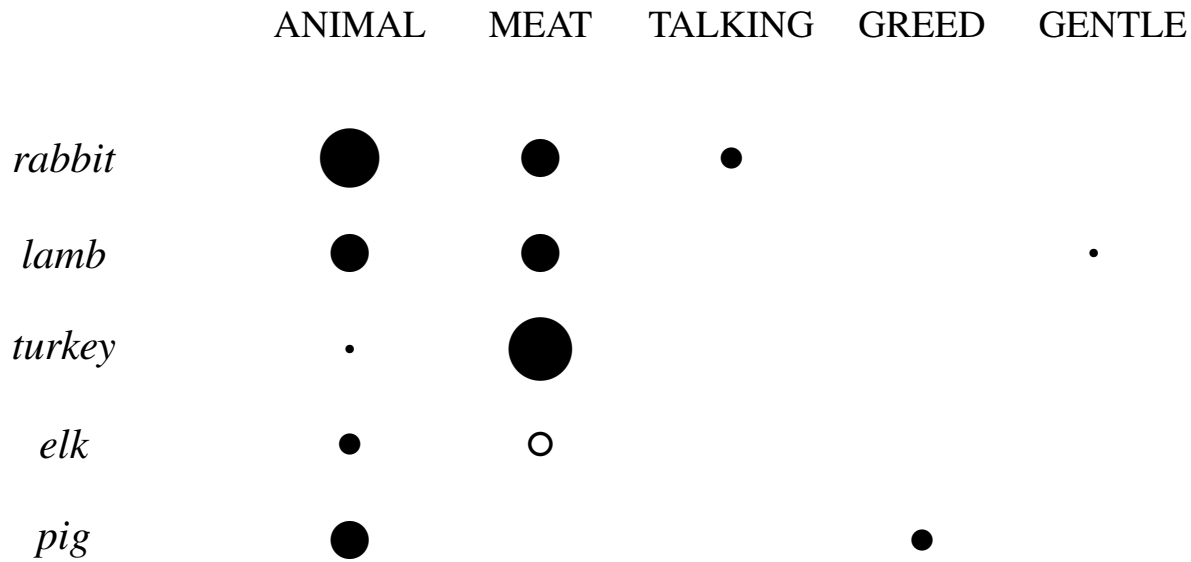
Regular sense distinctions/metonymy (e.g., place/government for countries):

> *China_org admits to climate failings.*
> *The company already owns nearly 50 stores in China_place, . . .*

Generalisation is possible, human agreement is often better that for WSD: better disambiguation performance.

Distributional approaches to regular polysemy

LC account: subspaces correspond to distributions for individuals and for groups of individuals (senses/usages) (also Rapp, 2004 on clustering; Boleda and Padó (2012) on regular polysemy).



Summary: qualia structure versus distributions

Issues with qualia structure:

- What types of thing are the fillers for qualia roles? (disjunctions of) lexemes, concepts? *smoke*: cigars, fish?
- Should they have associated probabilities?
- Can symbolic qualia values account for the observed data?
- How are qualia learned by humans?

Distributional accounts look more promising (but only if we can build single models, and don't use implausible amounts of data).

Feature structures versus distributions in general

- Feature structures are appropriate when:
  - Small number of relevant roles.
  - Role fillers can be isolated.
  - Defined processes (e.g., in grammars) which access those roles.
- Distributions are appropriate when:
  - No fixed set of roles or no role/filler distinction. Abstraction over any concept is possible (can't abstract over features).
  - Data source for distributions exists.
  - Sometimes: as a way of learning appropriate role fillers.
- Interfaces are necessary and not yet well understood.