L114 Lexical Semantics

Session 3: Lexical Relations and Taxonomies

Simone Teufel

MPhil in Advanced Computer Science
Computer Laboratory Natural Language and Information Processing (NLIP) Group
UNIVERSITY OF CAMBRIDGE
Simone.Teufel@cl.cam.ac.uk

2013/2014

Last time: WSD

- "Simpler" algorithms for Word Sense Disambiguation (WSD)
  - Lesk
  - Supervised ML
  - Yarowsky

Today:

- Lexical relations in Wordnet
- Theory on lexical relations
  - Hyponymy
  - Meronymy
- Taxonomies
- A WN-based WSD algorithm

"interest/3" – a closer look

S: (n) **interest** (a fixed charge for borrowing money; usually a percentage of the amount borrowed) "how much interest do you pay on your mortgage?"
direct hyponym / **full hyponym**

- S: (n) compound interest (interest calculated on both the principal and the accrued interest)
- S: (n) simple interest (interest paid on the principal alone)

direct hyponym/ **inherited hypernym** / sister term:

- S: (n) fixed charge, fixed cost, fixed costs (a periodic charge that does not vary with business volume (as insurance or rent or mortgage payments etc.))
  - S: (n) charge (the price charged for some article or service) "the admission charge"
    - S: (n) cost (the total spent for goods or services including money and time and labor)
      - S: (n) outgo, spending, expenditure, outlay (money paid out; an amount spent)
        - S: (n) transferred property, transferred possession (a possession whose ownership changes or lapses)
          - S: (n) possession (anything owned or possessed)
            - S: (n) relation (an abstraction belonging to or characteristic of two entities or parts together)
              - S: (n) abstraction, abstract entity (a general concept formed by extracting common features from specific examples)
                - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

"interest/3" – co-hyponyms

direct hyponym/ inherited hypernym / **sister term**:

- S: (n) fixed charge, fixed cost, fixed costs (a periodic charge that does not vary with business volume (as insurance or rent or mortgage payments etc.))
  - S: (n) cover charge, cover (a fixed charge by a restaurant or nightclub over and above the charge for food and drink)
  - S: (n) **interest** (a fixed charge for borrowing money; usually a percentage of the amount borrowed) "how much interest do you pay on your mortgage?"
  - S: (n) fee (a fixed charge for a privilege or for professional services)
  - S: (n) due (a payment that is due (e.g., as the price of membership)) "the society dropped him for non-payment of dues"

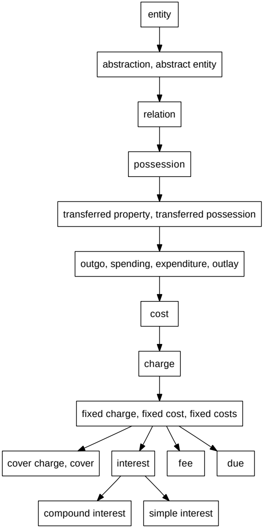
"interest/4" – a closer look

S: (n) **interest**, stake ((law) a right or legal share of something; a financial involvement with something) *"they have interests all over the world"; "a stake in the company's future"*

direct hyponym/ **inherited hypernym** / sister term:

- S: (n) share, portion, part, percentage (assets belonging to or due to or contributed by an individual person or group) "he wanted his share in cash"
  - S: (n) assets (anything of material value or usefulness that is owned by a person or company)
    - S: (n) possession (anything owned or possessed)
      - S: (n) relation (an abstraction belonging to or characteristic of two entities or parts together)
        - S: (n) abstraction, abstract entity (a general concept formed by extracting common features from specific examples)
          - S : (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

As a hierarchical graph



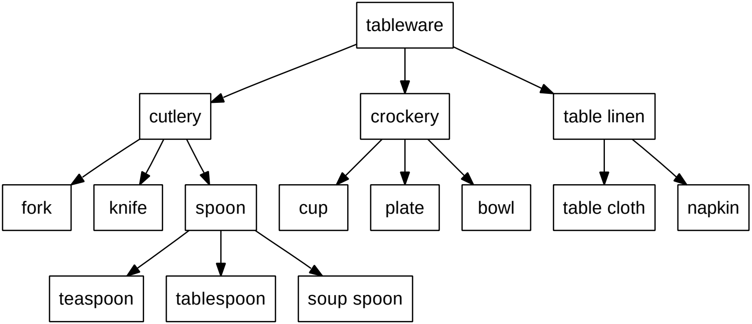
Lexical Relations

- Hyponymy
  - *Apple* is a hyponym of *fruit*.
  - *Fruit* is a superordinate/hypernym of *apple*.
- Meronomy
  - *Finger* is a meronym (rarely: partonym) of *hand*.
  - *Hand* is a holonym of *finger*.

Taxonomies

- Taxonomies are a subtype of hyponymy.
  - *horse:animal* forms part of a taxonomy.
  - *stallion:horse* does not, although it is a hyponymy.
- a taxonym further specifies the supertype's core characteristic:
  - *A strawberry blonde is a type of blonde.*
  - *?A blonde is a type of woman.*
- In a taxonomy, there is a unique mother constraint
- Example: tableware

Example taxonomy



Basic Level categories

Examples:

- *vehicle*–**car**–*hatchback*
- *object–implement–cutlery*–**spoon**–*teaspoon*

Properties:

- Pattern of behavioural interaction (you could mime how you'd interact with it)
- Visual image (you could visualise it)
- Part-whole relationships make sense (handle–implement?)
- Membership can be most rapidly decided (Alsatian–dog–mammal)
- Neutral, everyday reference
- Morphologically simple, original
- Level at which best categories are formed: maximize distintness from neighbours, internal homogeneity, informativeness

Super- and subordinate level categories

Superordinate categories:

- less good categories because not internally homogenous
- but distinct from sister categories
- often change mass/count properties with basic categories (*metals–silver* but *footwear–shoe*)

Subordinate level categories:

- show low distinctiveness from sister categories
- but are internally homogenous
- names frequently morphologically complex (e.g., *herring gull, coffee cup*

Taxonomies in everyday language

- Taxonomic hierarchies in everyday language rarely have more than 5 or 6 levels, typically fewer
- Taxonomic hierarchies appear mostly as fragments, not as fully developed structures.
- Expert, technical vocabularies (zoological ones included) do not show such limiations

Example for an everyday taxonomy

- Clothing
- Taxonyms at basic level: *trousers, jacket, dress, skirt, shoe. . .*
- Restricted perspective terms
  - where worn on body (*footwear, headwear;* all distinct)
  - when worn (*eveningwear*)
  - worn while doing what (*sportswear, outdoor wear, leisurewear*)
- Default category: everyday, publicly observable, not-for-special purpose clothing; unnamed.
- Virtually impossible to create a well-formed hierarchy of clothing terms

Clothing, problems

- Each perspective potentially yields a separate hierarchy: *shoe* is hyponymic to *evening wear* and *footwear*.
- Cross-classification if perspective is changed. *tennis shoes* hyponym of *sportswear* and *high heels* hyponym of *evening wear*, but both hyponyms of *shoe*
- Different hierarchies can intersect in various ways
- Senses are not fully lexically distinguished *dress/1* is a taxonym of neutralwear, and also a hyponym of *dress/2* (a mass term) with a more general sense

Exercise

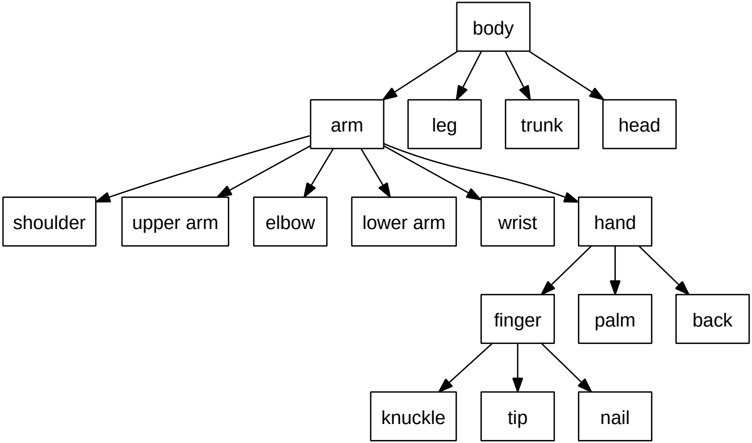
Create the best taxonomy you can from the following word forms:

tablecloth	*wine glass*	*table mat*	*salt*	*napkin*
teaspoon	*bread knife*	*coaster*	*tumbler*	*vinegar*
water jug	*fork*	*cake dish*	*saucer*	*napkin ring*
knife	*butter knife*	*corkscrew*	*cake slice*	*pepper*
breadboard	*butter dish*	*soupspoon*	*teaspoon*	*serving spoon*
soup bowl	*dessert spoon*	*mug*		

Taxonomies in other languages

- English sense of *animal* = { *mammals, amphibians, reptiles* } does not exist in French or German
- { *walnuts, hazelnuts, almonds* . . . } do not form a natural class *nuts* in French or Italian.
- In English, *marmelade* is not a hyponym of *jam*
- Boots and *sandals* are not (necessarily) hyponyms of *shoe* in English

Example meronymy



Meronymy

- Much less sharply defined relationship than hyponymy.
- Is a lid a part of the pot?
- What makes a good part:
  - necessary/functionality to the proper functioning of holonym
  - integrality/attachedness
    - The hand is attached to the arm
    - ? The fingers are attached to the hand
  - Moves indepently from holonym

Meronymy

- Has the same range as holonym
  - There are no hands without fingers and no fingers without hands.
  - There are doors without handles and handles without doors
- Exist at the same time – unlike ingredients.
  - ?Milk is part of the cake
- Parts and wholes are of same ontological type.
  - ?Wood is part of the table (material vs. object)
  - ?A nerve is part of a leg (systemic vs. segmental parts)

More on meronymy

- Part-of vs. Piece-of relationship
  - Pieces are always concrete
  - Replicas of pieces are not pieces (a piece must have once been part of an undamaged whole); but: spare parts
  - Motivated vs arbitrary boundaries
  - A part has a function
- Meronymy is often not transitive:
  - Handle is part of a door
  - Door is part of a house
  - ? Handle is part of a door
- Co-meronymy is a relation of exclusion; sister parts do not overlap. If X and Z are co-meronyms of Y, then no meronym of X is a meronym of Z.

Meronomies

- Example: segmental version of human body
- Lexical gaps: meaning not salient enough to merit lexical distinction (*palm+back of hand=?*; *watch+clock=?*)
- Automeronymy (if a lexical gaps filled by extended sense of item directly below or above; *body (trunk)+head=body*)
- Examples from other languages:
  - Greek *xeri/podi* is hand/foot up to elbow/knee
  - Turkish: thumb is simply "big finger"
  - Conceptual gap: distinction not perceived to exist (dark and light blue; Russian)

Graph-Based WSD (Navigli and Lapata; 2010)

- The internal structure of sense inventories can be exploited even further.
- Represent Wordnet as a graph whose nodes are synsets and whose edges are relations between synsets.
- The edges are not labeled, i.e., the type of relation between the nodes is ignored.

Figures and tables in this section from Navigli and Lapata (2010).

Example

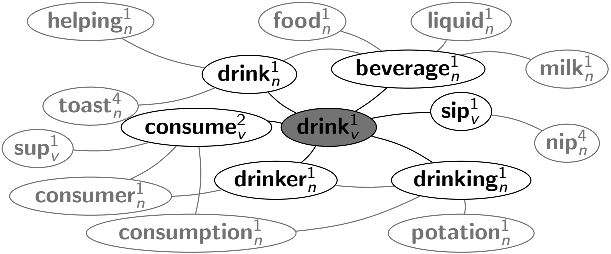
Wordnet Synsets (senses) of drink/v:

- {**drink**^1_v, *imbibe*^3_v} (take in liquids)
- {**drink**^2_v, *booze*^1_v, *fuddle*^2_v} (consume alcohol)
- {*toast*^2_v, **drink**^3_v, *pledge*^2_v, *salute*^1_v, *wassail*^2_v} (propose a toast)
- {*drink in*^1_v, **drink**^4_v} (be fascinated, pay close attention)
- {**drink**^5_v, *tope*^1_v} (be an alcoholic)

Wordnet Synsets (senses) of milk/n:

- {**milk**^1_n} (a white nutritious liquid secreted by mammals and used as food by human beings)
- {**milk**^2_n} (produced by mammary glands of female mammals for feeding their young)
- {**Milk**^3_n, *Milk River*^1_n} (a river that rises in the Rockies in northwestern Montana and flows eastward to become a tributary of the Missouri River)
- {**milk**^4_n} (any of several nutritive milklike liquids)

Graph for first sense of *drink*



Graph Construction

Disambiguation algorithm:

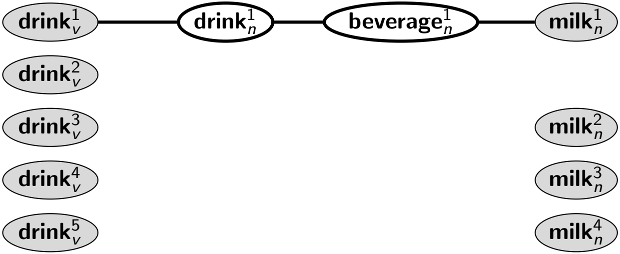
1. Use the Wordnet graph to construct a graph that incorporates each content word in the sentence to be disambiguated;
2. Rank each node in the sentence graph according to its importance using **graph connectivity measures;**
   - **Local measures:** give a connectivity score to an individual node in the graph; use this directly to select a sense;
   - **Global measures:** assign a connectivity score the to the graph as a whole; apply the measure to each interpretation and select the highest scoring one.

Graph Construction

- Given a word sequence $\sigma = (w_1, w_2, \ldots, w_n)$, find all possible word senses of all words; call this set V_σ.
- Perform a depth-first search of the Wordnet graph: every time we encounter a node $v' \in V_\sigma$ ($v' \neq v$) along a path $v \rightarrow v_1 \rightarrow \cdots \rightarrow v_k \rightarrow v'$ of length L, we add all intermediate nodes and edges on the path from v to v' to the graph G.
- For tractability, we set the maximum path length to 6.

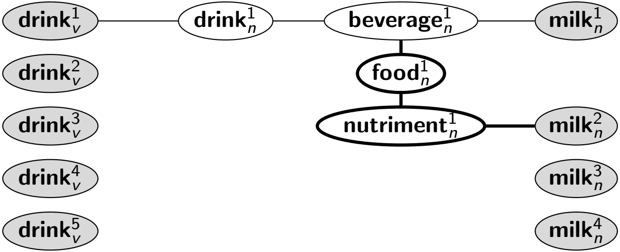
Graph Construction

Example: graph for *drink milk*.



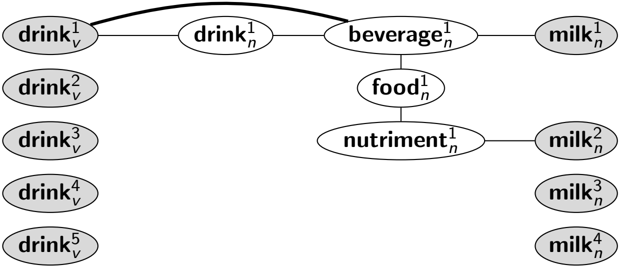
Graph Construction

Example: graph for *drink milk*.



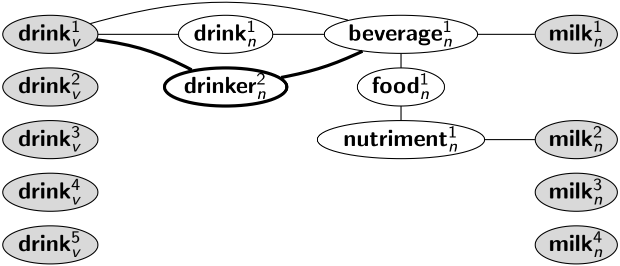
Graph Construction

Example: graph for *drink milk*.



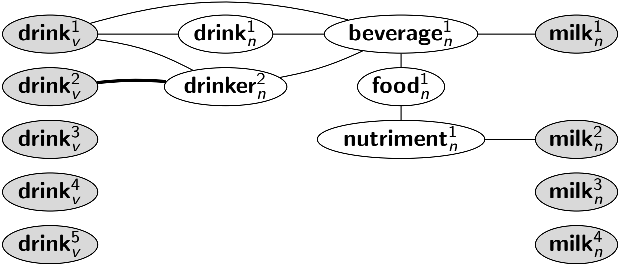
Graph Construction

Example: graph for *drink milk*.



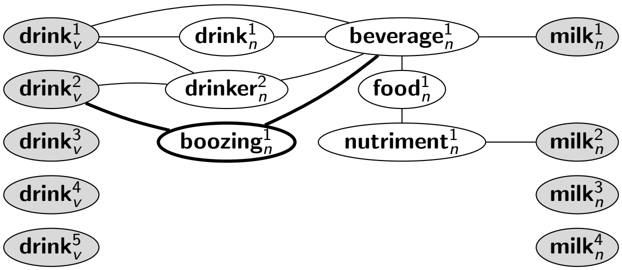
Graph Construction

Example: graph for *drink milk*.



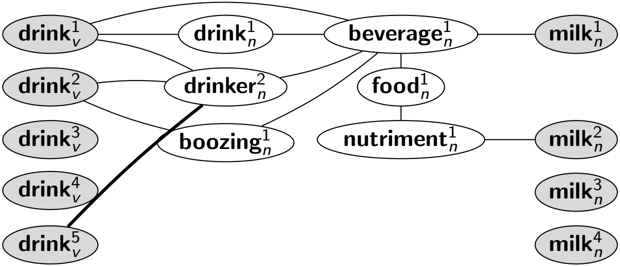
Graph Construction

Example: graph for *drink milk*.



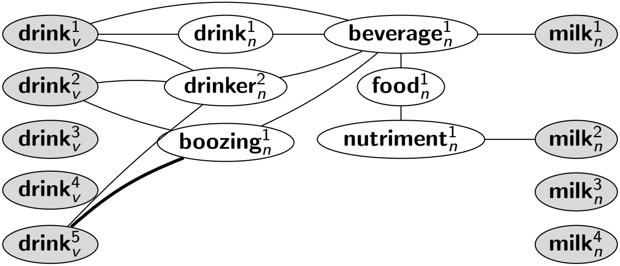
Graph Construction

Example: graph for *drink milk*.



Graph Construction

Example: graph for *drink milk*.



We get $3 \cdot 2 = 6$ interpretations, i.e., subgraphs obtained when only considering one connected sense of *drink* and *milk*.

A Local Measure: Degree Centrality

Assume a graph with nodes V and edges E. Then the **degree** of $v \in V$ is the number of edges terminating in it:

$$deg(v) = |\{\{u, v\} \in E : u \in V\}| \quad (1)$$

Degree centrality is the degree of a node normalized by the maximum degree:

$$C_D(v) = \frac{deg(v)}{|V| - 1} \quad (2)$$

For the previous example, $C_D(drink_v^1) = \frac{3}{14}$, $C_D(drink_v^2) = C_D(drink_v^5) = \frac{2}{14}$, and $C_D(milk_n^1) = C_D(milk_n^2) = \frac{1}{14}$. So we pick $drink_v^1$, while $milk_n$ is tied.

A Global Measure: Edge Density

Edge density of a graph is the number of edges compared to a complete graph with $|V|$ nodes (given by $\binom{|V|}{2}$):

$$ED(G) = \frac{|E(G)|}{\binom{|V|}{2}} \tag{3}$$

The first interpretation of **drink milk** has $ED(G) = \frac{6}{\binom{5}{2}} = \frac{6}{10} = 0.60$, the second one $ED(G) = \frac{5}{\binom{5}{2}} = \frac{5}{10} = 0.50$.

Evaluation on SemCor

	Measure	WordNet All	WordNet Poly	EnWordNet All	EnWordNet Poly
	Random	39.13	23.42	39.13	23.42
	ExtLesk	47.85	34.05	48.75	35.25
Local	**Degree**	**50.01**	**37.80**	**56.62**	**46.03**
Local	PageRank	49.76	37.49	56.46	45.83
Local	HITS	44.29	30.69	52.40	40.78
Local	KPP	47.89	35.16	55.65	44.82
Local	Betweenness	48.72	36.20	56.48	45.85
Global	Compactness	43.53	29.74	48.31	35.68
Global	Graph Entropy	42.98	29.06	43.06	29.16
Global	Edge Density	43.54	29.76	52.16	40.48
	First Sense	74.17	68.80	74.17	68.80

Evaluation on Semeval All-words Data

System	F
Best Unsupervised (Sussex)	45.8
ExtLesk	43.1
Degree Unsupervised	52.9
Best Semi-supervised (IRST-DDD)	56.7
First Sense	62.4
Best Supervised (GAMBL)	65.2

Discussion

Strengths:

- exploits the structure of the sense inventory/dictionary;
- conceptually simple, doesn't require any training data, not even a seed set;
- achieves good performance for unsupervised system.

Weaknesses:

- performance not good enough for real applications (F-score of 0.53 on Semeval);
- sense inventories take a lot of effort to create (Wordnet has been under development for more than 15 years).

Summary

- Ontologies such as WN are based on lexical relations such as hyponomy (subtype taxonomy) and meronymy, which are non-trivial phenomena in the real world
- **Unsupervised graph-based WSD** finds the most connected nodes (senses) in a graph of lexical relations that represents all possible interpretations of a sentence.

Reading for today

Cruse chapters 3.2.3.6 (p. 61/62); 6 and 8

Navigli and Lapata (2010): An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 32(4), IEEE Press, 2010, pp. 678-692.