

Formal Concept Analysis for Linguistic Data Mining

Uta Priss

Ostfalia University
Wolfenbüttel, Germany
www.upriss.org.uk

November 2012

Outline

Formal Concept Analysis

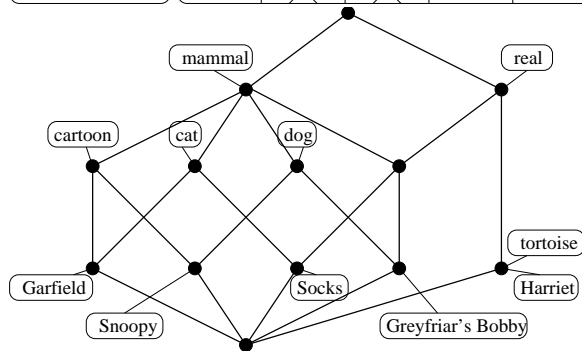
Paradigmatic

Syntagmatic

Associative concepts

Formal context and concept lattice

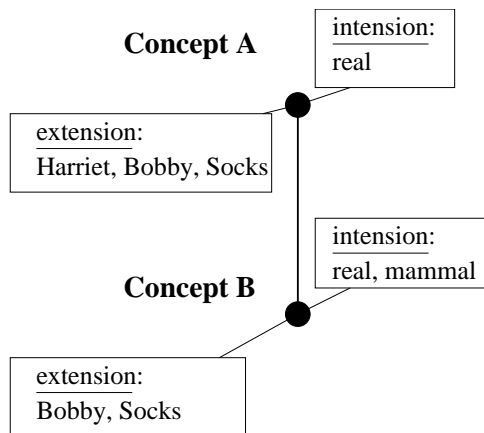
	cartoon	real	tortoise	dog	cat	mammal
Garfield	×				×	×
Snoopy	×			×		×
Socks		×			×	×
Greyfriar's Bobby		×		×		×
Harriet		×	×			



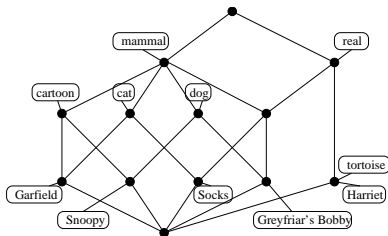
Concept lattices express duality

- ▶ objects/attributes
- ▶ extension/intension
- ▶ token/type
- ▶ value/data type
- ▶ data-driven/theory-driven
- ▶ bottom-up/top-down

The subconcept-superconcept relation



Other FCA topics: Implications

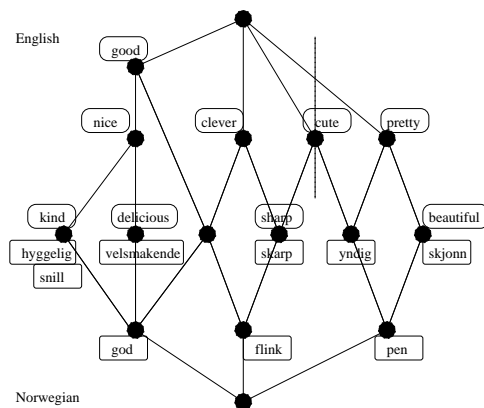


All cats are mammals.
The only real cat is Socks.
All tortoises are real.

Exploring paradigmatic relationships using ...

- ▶ bilingual corpora
- ▶ bilingual dictionaries
- ▶ monolingual semantic dictionaries/thesauri

Bilingual corpora - Semantic Mirrors (Dyvik, 2004)



Words, senses and entries in Roget's Thesaurus

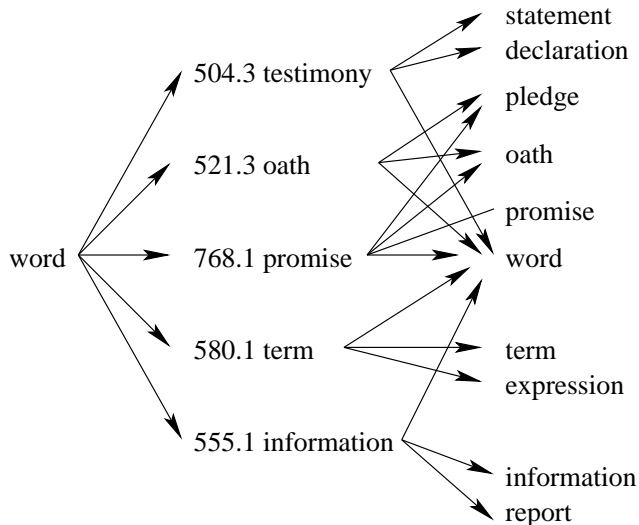
WORDS → SYNONYMY

SENSES
↓

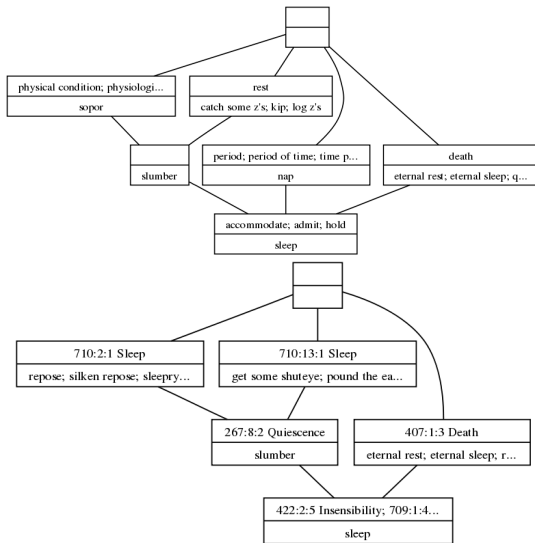
	over	above	beyond	across	past	in excess of	...
9:12:1	X						
206:24:2	X	X					
206:27:4	X	X	X			X	
227:40:1	X	X					
...	X	X	X	X	X		

POLYSEMY

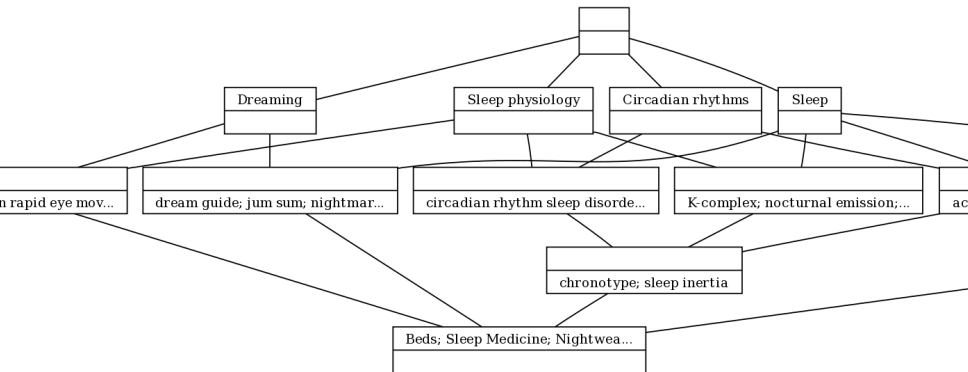
The plus operator



A Neighbourhood in WordNet (top) and ROGET



Wikipedia: Sleep

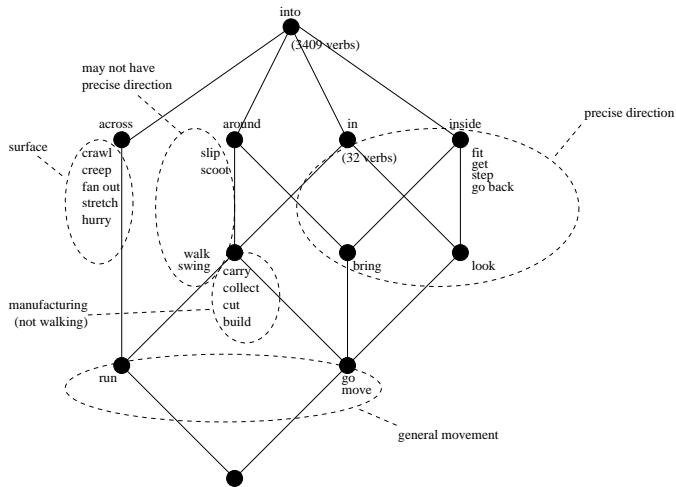


Syntagmatic relationships

- ▶ She eats a banana. → eat food
- ▶ We eat at 5 pm. → eat a meal
- ▶ What is eating you? → being distressed

→ semantically similar verbs have syntactically similar argument structures (Beth Levin)

Verb clusters in a corpus



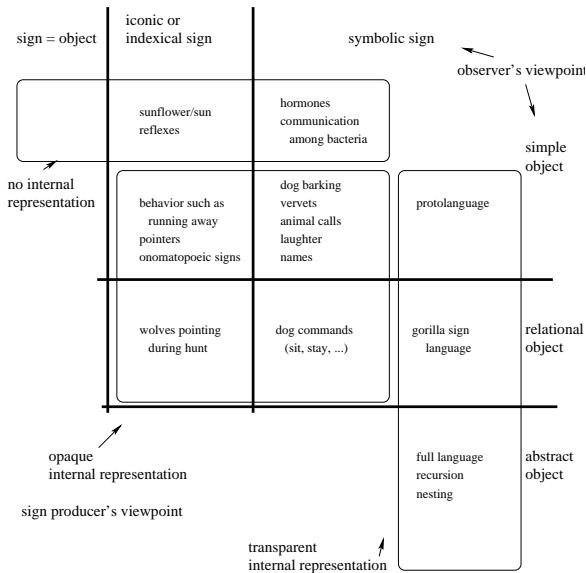
Exploiting syntagmatic relationships

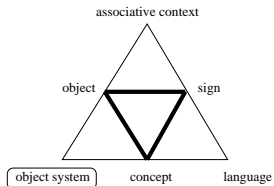
- ▶ Verb paradigms - Grosskopf & Harras (1999)
- ▶ Lexical Tuning - Basili et al. (1997)
- ▶ Ontology engineering - Cimiano et al. (2003)

Dichotomies: contradictory and complementary!

- ▶ Physics: photons as waves and particles
- ▶ Psychology: Sloman (1996): “two systems of reasoning”
- ▶ AI: symbolic and connectionist models
- ▶ Classification theory: “Aristotelian” versus prototype models
- ▶ History of science: externalists and internalists
- ▶ Positivism versus Post-modernism
- ▶ Bertin’s semiology of graphics versus Gibson’s ecological theory of perception
- ▶ Pinker: rule-based and associative rules of language (e.g., regular and irregular verbs)

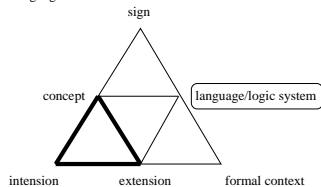
	formal	associative
design:	top-down	bottom-up
objects:	symbolic, abstract	complex, grounded
concepts:	precise, defined	fuzzy, emergent
systems:	simple, organised correct, complete discrete, static	complex, decentralised probabilistic, dynamic continuous, "chaotic"
extensions:	set of "objects"	prototypes, exemplars
intensions:	formal definitions	stereotypes, image schemata
reasoning:	logical, formal causal explanation hierarchical inference, deduction classification correctness	similarity, fuzzy co-occurrence observation associative analogy, abduction clustering resonance
implementations	algorithmic, modular	learning-based, neural networks





Peirce's sign triad

formal concept analysis



Thank you!