Linguistic Applications of Formal Concept Analysis

OntoQuery - Lecture 2

Uta Priss
School of Computing,
Napier University,
Edinburgh, UK
u.priss@napier.ac.uk

May, 2004

1) Analysing Linguistic Structures

2) Lexical Semantics

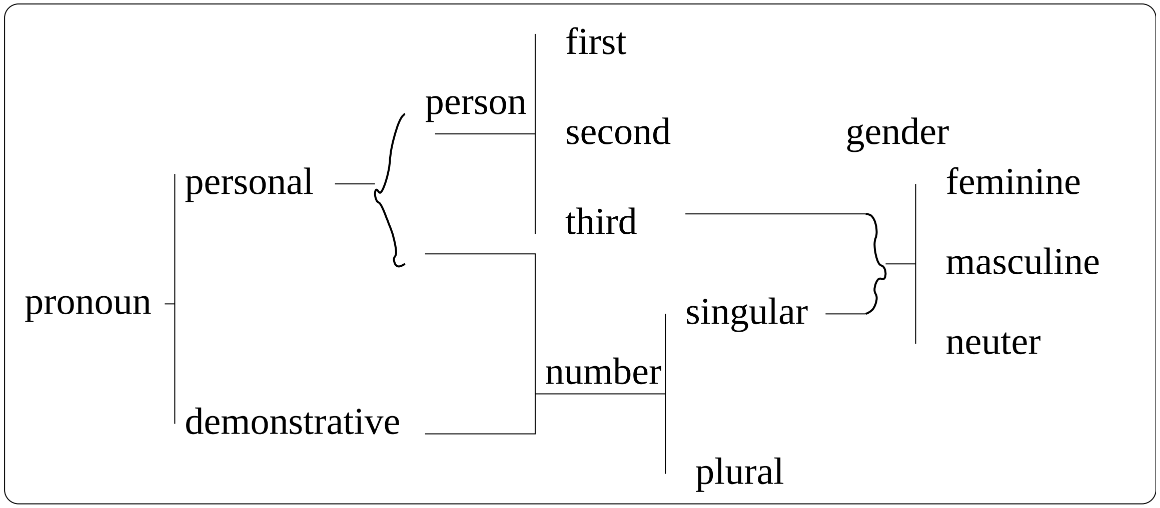
3) Very Large Lexical Databases

4) Applications of Lexical Databases

5) Future Research

1.1 Analysing Linguistic Structures

Systemic Classification of English Pronouns



Terry Winograd (1983): Language as Cognitive Process

quoted from Osswald &Petersen (2002): Induction of Classifications from Linguistic Data

1.2 Paradigms, Speech Act Verbs and Componential Analysis

- Großkopf, A., (1996). *Formal concept analysis of verb paradigms in linguistics.* In: Diday; Lechevallier & Opitz (Eds.) Ordinal and Symbolic Data Analysis.

- Großkopf, A.; Harras, G. (1999). *Begriffliche Erkundung semantischer Strukturen von Sprechaktverben.* In: Stumme & Wille (Eds.) Begriffliche Wissensverarbeitung: Methoden und Anwendungen.

- Kipke, U.; Wille, R. (1987). *Formale Begriffsanalyse erläutert an einem Wortfeld.* LDV–Forum, 5.

2.1 Lexical Semantics: Lexicon = Ontology

HPSG

$$
\left[\begin{array}{ll}
\text{cake} & \\
\text{ARGSTR} = & \left[\begin{array}{l} \text{ARG1 = x: food} \\ \text{D−ARG1 = y:mass} \end{array}\right] \\
\text{QUALIA} = & \left[\begin{array}{l} \text{CONST = y} \\ \text{FORMAL = x} \\ \text{TELIC = eat(e2,z,x)} \\ \text{AGENTIVE = bake−act(e1,w,y)} \end{array}\right]
\end{array}\right]
$$

2.2 The Lexicon is not Enough

"In any randomly chosen newspaper paragraph, each sentence will be likely to have an extended sense of at least one word, usually a verb, in the sense of a use that breaks conventional preferences and which might be considered extended or metaphorical use, and quite likely not in a standard lexicon." (Basili et. al, 2000)

(cf. Chaitin (1999). *The Unknowable* $\rightarrow$ Almost everything in mathematics is random.)

2.3 Associative and Formal Concepts

"The woman finished her beer." $\rightarrow$ "stopped drinking"

"The man finished his cigarette." $\rightarrow$ "extinguished"

"She looked through the window." $\rightarrow$ "the glass of the window"

"He painted the window." $\rightarrow$ "the frame of the window"

- Subtle distinctions in meaning are based on language external or *associative concepts* and not on lexical or *formal concepts*. (Priss, 2002)

2.4 A Vicious Cycle

- Machine learning helps building a lexicon but a lexicon is needed for machine learning.

- Selectional restrictions help word sense disambiguation but word sense disambiguation is needed for selectional restrictions (Basili et al., 2000).

- Ontologies help NLP but NLP is needed for ontologies.

Reason: Ontologies, NLP and others attempt to capture associative structures, which can only be approximated by formal structures (or Turing Machines).

2.5 A Partial Solution: Bootstrapping

- Basili, R.; Pazienza, M.; Vindigni, M. (1997) *Corpus-driven unsupervised learning of verb subcategorization frames.* AI*IA-97.

- Basili, R.; Catizone, R.; Pazienza, M.; Stevenson, M.; Velardi, P.; Vindigni, M.; Wilks, Y. (2000). *An empirical approach to Lexical Tuning.*

2.6 Lexical Tuning (Basili et al., 1997)

Start with some lexicon and then fine-tune with patterns found in a corpus.

Build an event hierarchy as a concept lattice of verb patterns using linguistic contexts (as objects) and syntagmatic arguments (as attributes). Each node in the lattice represents certain linguistic contexts or word senses of the verbs.

Use semi-automated machine learning algorithms for conceptual clustering of the patterns.

Search lattice top down based on size of extension.

2.7 Lexical Inheritance Hierarchies and Machine Learning

- Petersen, Wiebke (2001). *A Set-Theoretical Approach for the Induction of Inheritance Hierarchies.* Electronic Notes in Theoretical Computer Science 51.

- Osswald and Petersen (2002). *Induction of Classifications from Linguistic Data.*

- Sporleder, Caroline (2002). *A Galois Lattice based Approach to Lexical Inheritance Learning.* ECAI Workshop on ML and NLP for Ontology Engineering.

3.1 Very Large Lexical Databases

- Thesauri, Roget's Thesaurus $\rightarrow$ tree-hierarchy, subconcept-superconcept relation

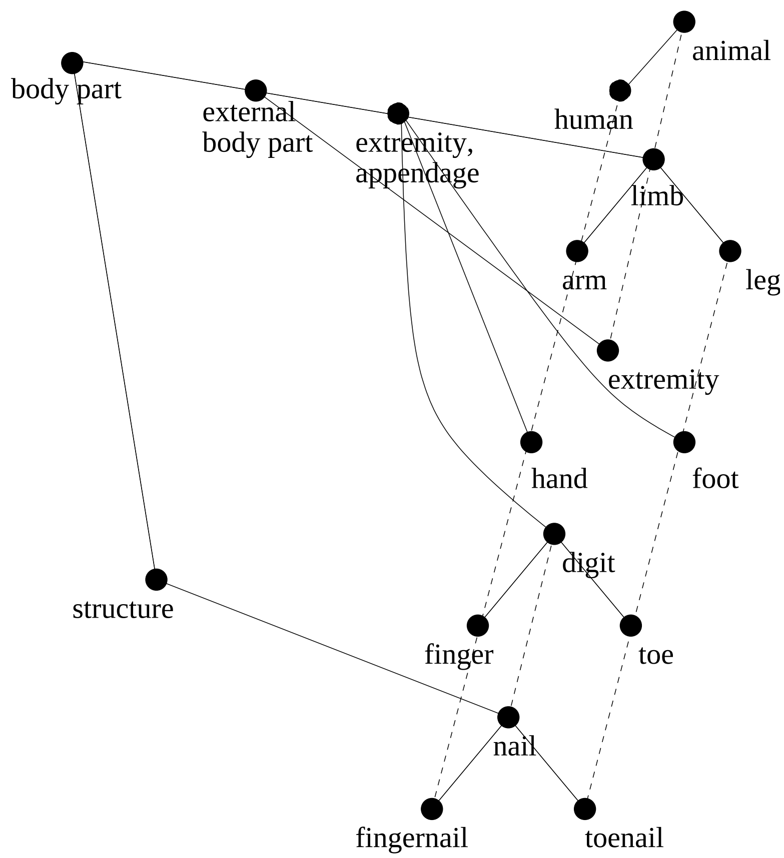
- WordNet, EuroNet $\rightarrow$ several semantic relations

- Corpora $\rightarrow$ unstructured

3.2 Using Formal Concept Analysis to ...

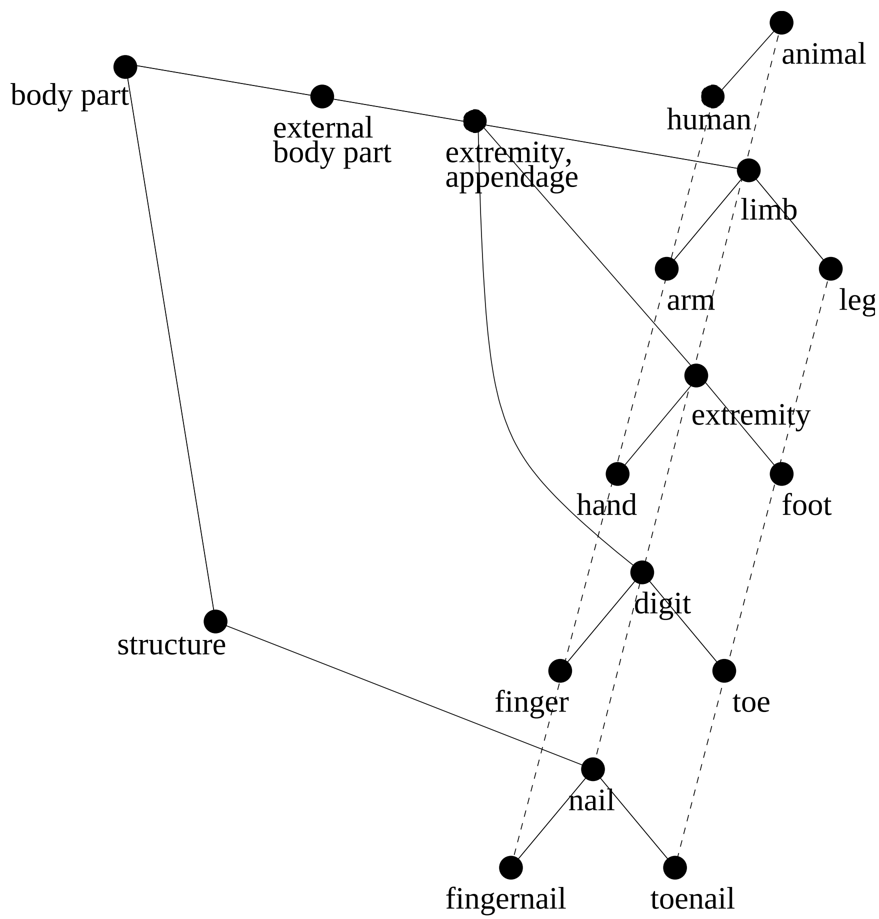
- build a lexical database, thesaurus or ontology
- visualise conceptual structures in a lexical databases
- analyse semantic relations and identify inconsistencies among semantic relations in a lexical database

Meronymy



animal
body part
human
external
body part
extremity,
appendage
limb
arm
leg
extremity
hand
foot
digit
structure
finger
toe
nail
fingernail
toenail

Meronymy (continued)



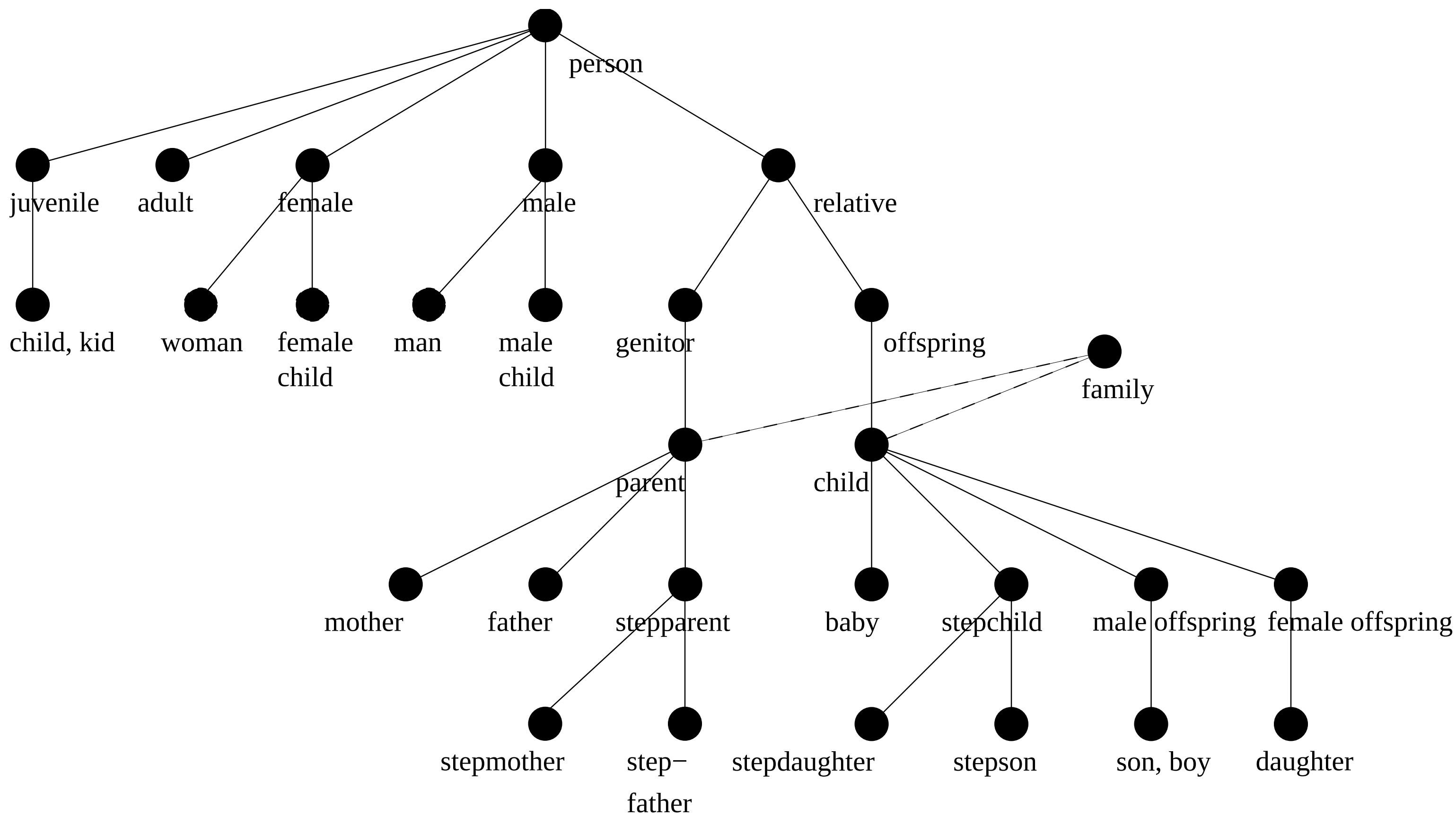
Visualisation of lexical databases

Problem: how to extract sublattices.

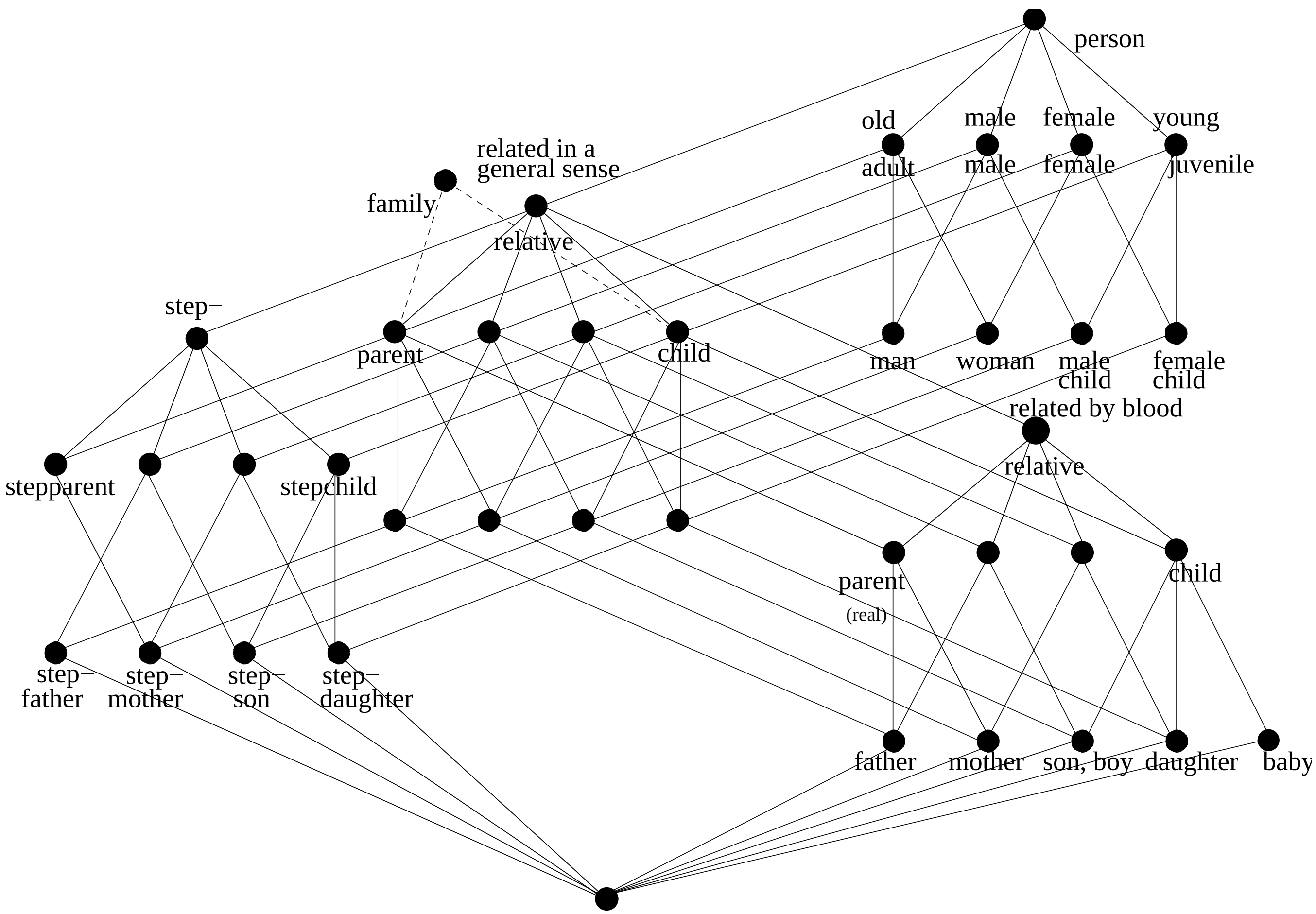
Solutions:

- Neighbourhood lattices: selecting sets of words with similar meanings
- Scaling: manual or semi-automated
- Facets, regular polysemy

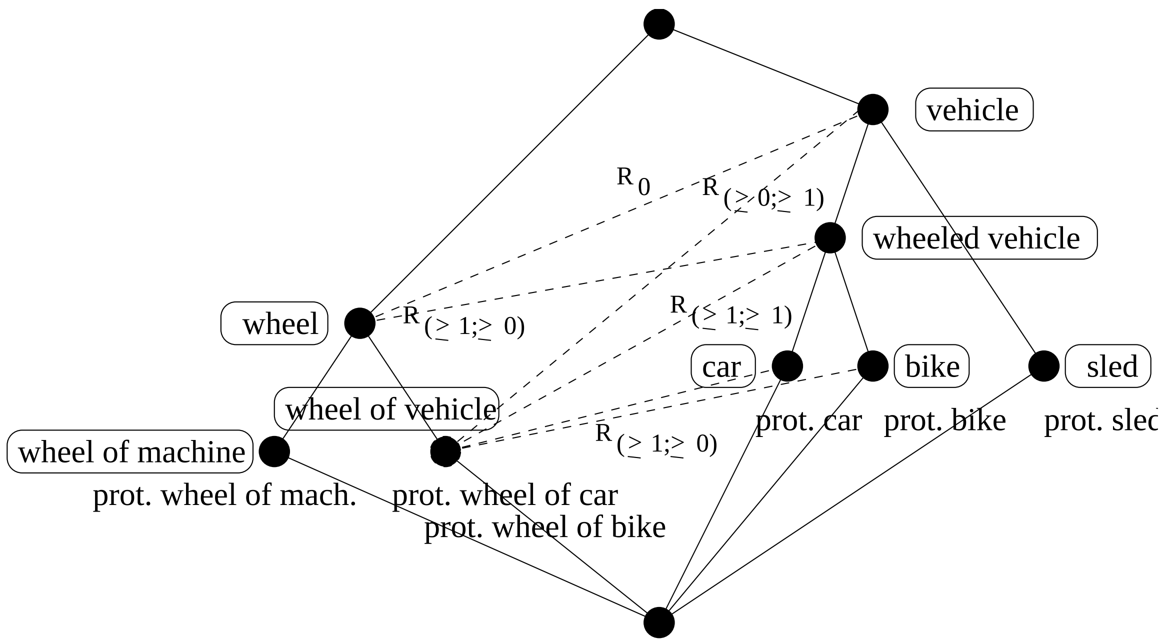
Nouns in Wordnet



Facets/Regular Polysemy

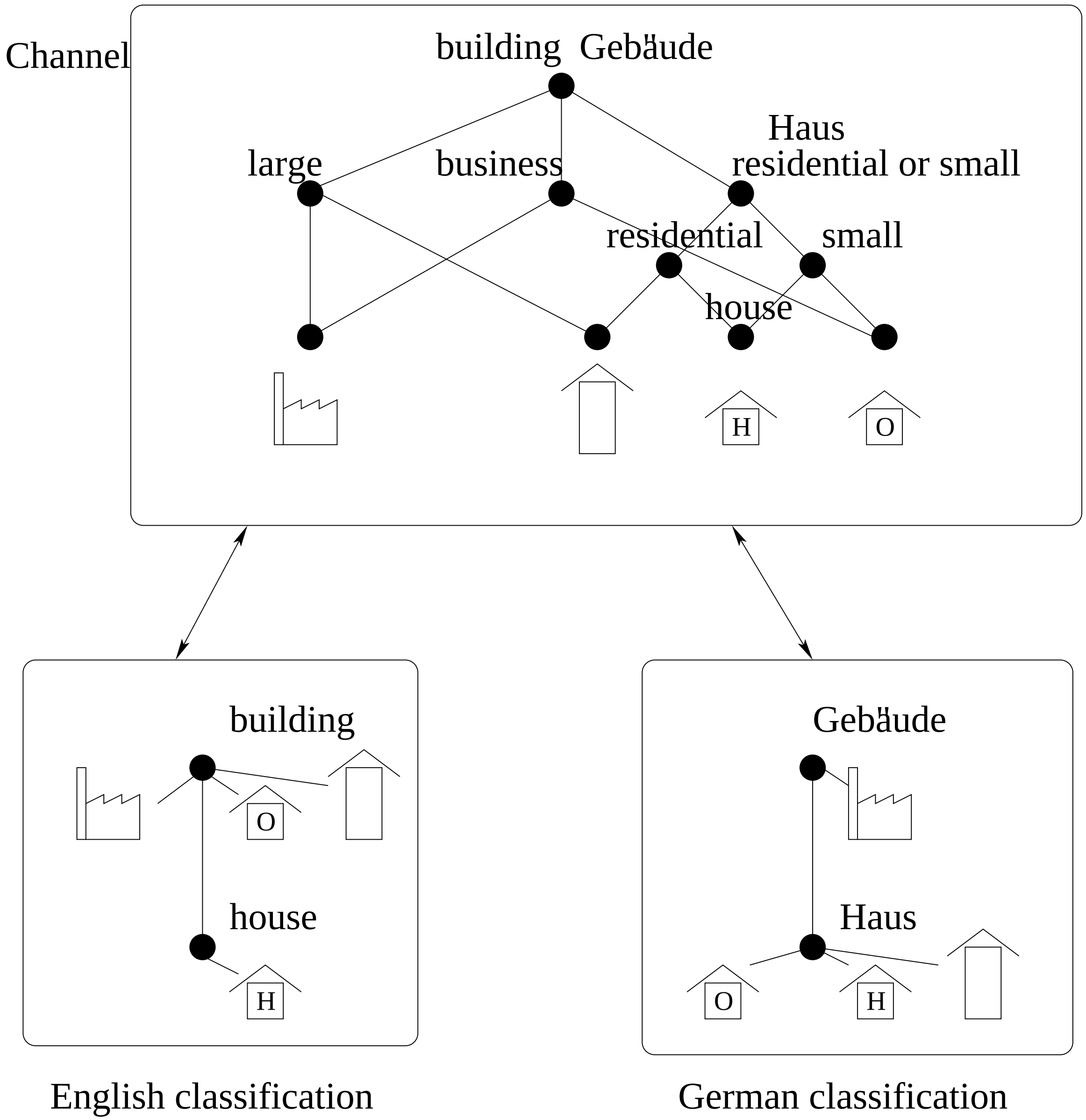


Bases of semantic relations

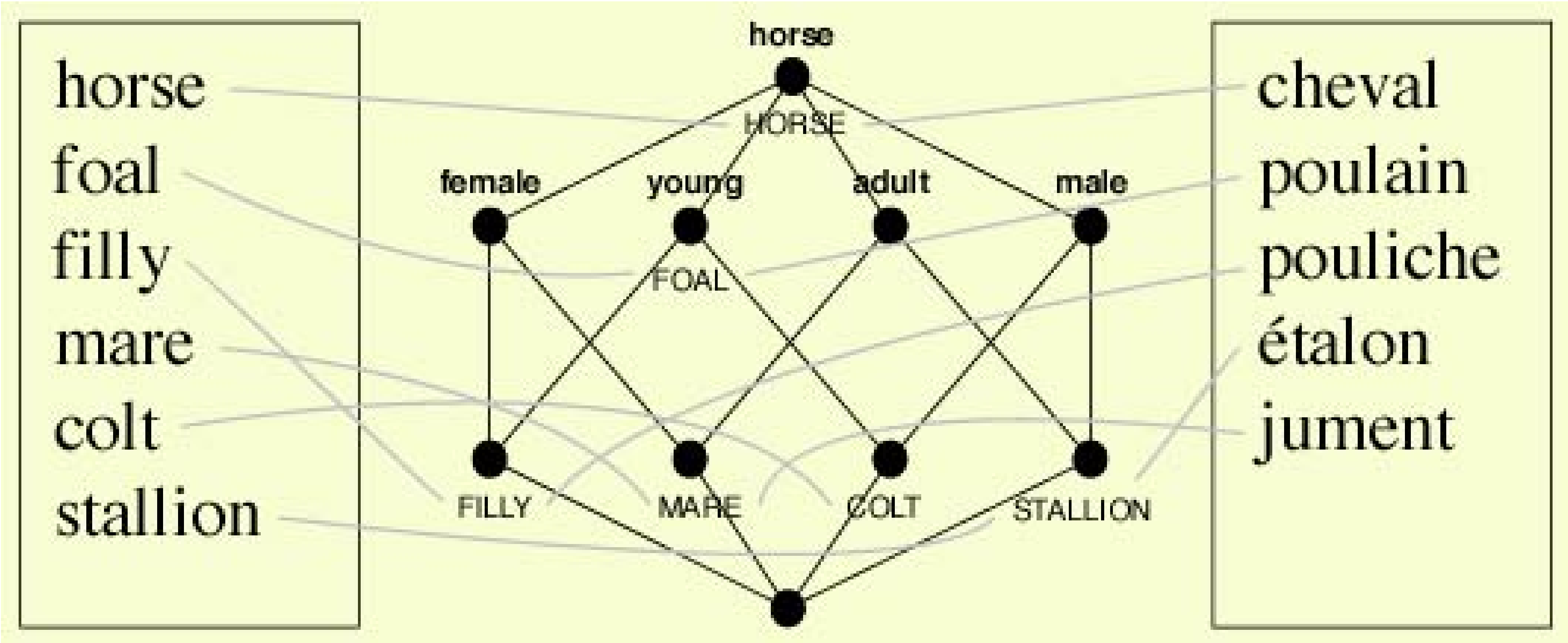


4.1 Applications of Lexical Databases

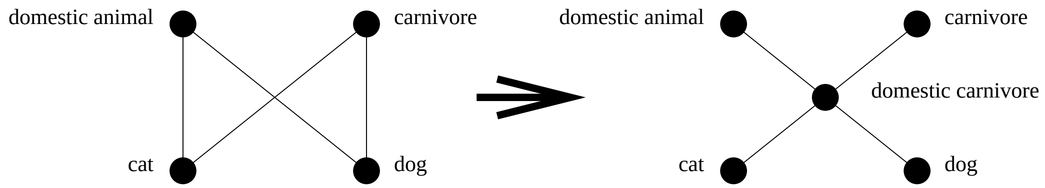
- Multilingual conceptual structures: lexical gaps, conceptual interlingua
- Joining different conceptual structures, information flow (Barwise)
- Information retrieval, query by navigation


Channel
building Gebäude
Haus
large
business
residential or small
residential
small
house
H
O
building
O
house
H
Gebäude
Haus
O
H
English classification
German classification

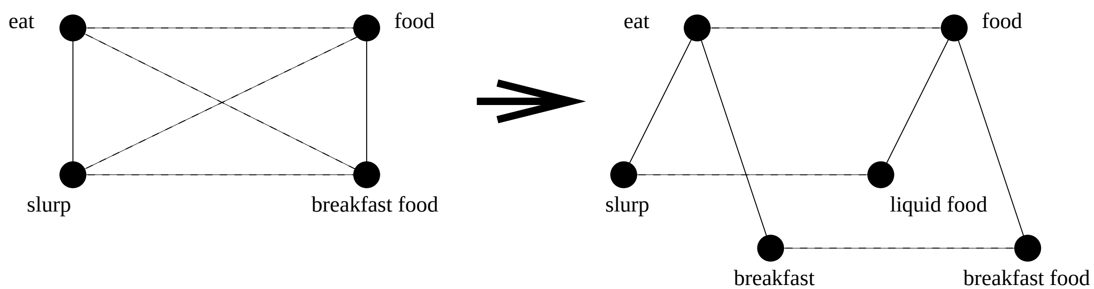
SIMuLLDA - Maarten Janssen (2002)



Lexical gaps: concepts generated by the conceptual hierarchy



Lexical gaps: concepts generated by semantic relations



5.1 Future Research Possibilities

- Use Basili's approach in information retrieval
- Visualisations of large lexical databases
- Large scale ontology construction using FCA
- Associative structures
- Information flow