

The conceptual structure of Roget's Thesaurus, a conceptual hierarchy

John Old

Napier University, Edinburgh

Concept Analysis and Concept-based Retrieval

- Brief History
- Comparison:
 - Thesauri, Dictionaries, WordNet and Word Association Data
- Explicit Organization (Book Structure)
 - Conceptual Hierarchy/Ontology
 - Concept (>Notion, >sense, >image schema>)*
 - Body of the book & Index
 - Senses/Synonym sets (SynSets)
- Local, implicit Organization
 - Synonymy versus Polysemy
 - Lexical (words) vs conceptual/semantic (senses) relations
- Antonymy

Brief History

Doctor Peter Mark Roget, MD

Family--Belgian Huguenot refugees

Recorded and classified useful words throughout his life. On retirement, made them into a book, first published in 1852.

Thesaurus of English Words & Phrases, Classified & Arranged so as to Facilitate the Expression of Ideas & Assist in Literary Composition

Also invented the slide rule, public health clinics, the London sewerage system, and cinema!

Comparison

Thesaurus

Sense 1

- *word A*
- *word B*
- *word C*

Sense 2

- *word A*
- *word D*
- *word F*

Dictionary

Word A

- sense 1
- sense 2
- sense 3

Word B

- sense 1
- sense 6
- sense 7

Word Association

Word A

- *word B*
- *word C*
- *word D*
- *word F*

Word B

- *word A*
- *word C*
- ...

Explicit Structure

Roget Top Level Classes (1962 Edition--RIT)

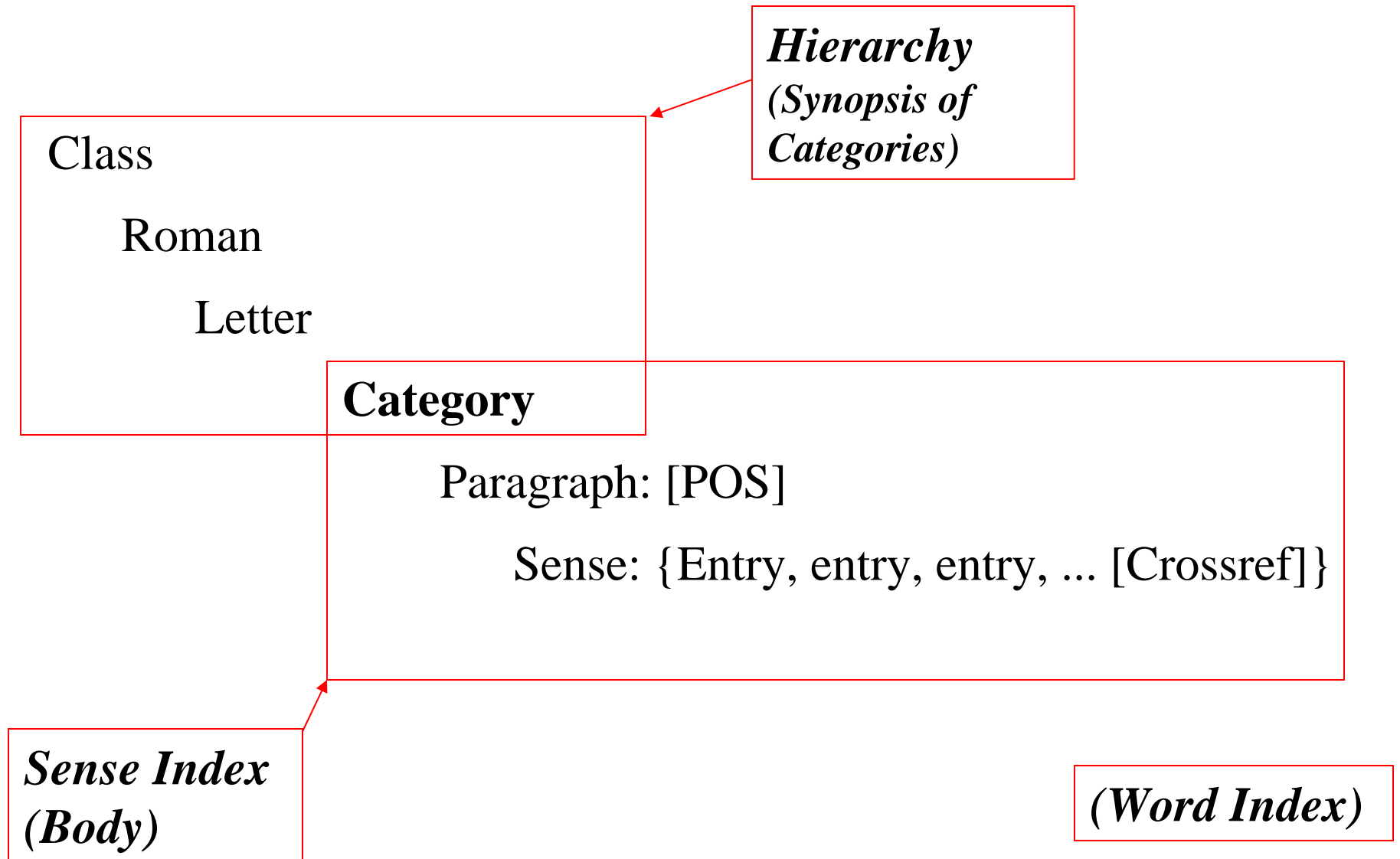
- 1. Abstract Relations**
- 2. Space**
- 3. Physics**
- 4. Matter**
- 5. Sensation**
- 6. Intellect**
- 7. Volition**
- 8. Affections**

Explicit Structure

Entries look something like this (each word or phrase is an entry):

thesaurus, synonym finder, wordbook, cyclopedia, glossary, language reference book, onomasticon, sourcebook, storehouse of words, vocabulary word list, word treasury, conceptual dictionary, semantic dictionary, synonymicon.

Roget's Thesaurus Conceptual Hierarchy



Roget's Thesaurus Conceptual Hierarchy Example

1. **1** Abstract Relations

2. **IV** Order

3. **A** Order in General

4. **60**:Arrangement

5. **1**: *Nn* Reduction to Order

6. **5**:{arrangement, ordering, “confusion unconfused” [Young], 59:0}

Category

```
graph TD; Category[Category] --> 60[60:Arrangement]; Entry[Entry] --> 59[59:0]; Sense["Sense - a.k.a Synset 60:1:5"] --> 59; Cross["Cross-reference"] --> 59;
```

Entry

Sense – a.k.a
Synset 60:1:5

Cross-
reference

Sense Index

CATEGORY

PARAGRAPH

40. Addition

SENSE

NOUNS 1.

...

ADVS. 10. and also, **additionally**, **in addition**, and all [coll.], yet, extra, on the side [slang], moreover, farther, further, furthermore, overplus, again, more, similarly, likewise, by the same sign, by the same token, item, **also**, then, **to boot**, and so, **as well**, **too**, else, au reste [F.], **besides**, plus, **into the bargain**, past, on top of, **over**, above, over and above, at that [coll.], beyond, beside; ...

PREPS. *etc.* 11 with, including ...

Index words

ENTRY=**40:10:1-over**

Word Index

WORD

over

Paragraph name - Index word

verbs

...

advs. additionally **40:10**

throughout 56.18

...

preps. concerning 9.12

exceeding 36.20

...

on 227.40

...

OVER AGAINST

...

overabound

...

PARAGRAPH

CATEGORY

ENTRY=40:10:1-over

Word-Sense Relation

WORDS



Synonymy

SENSES

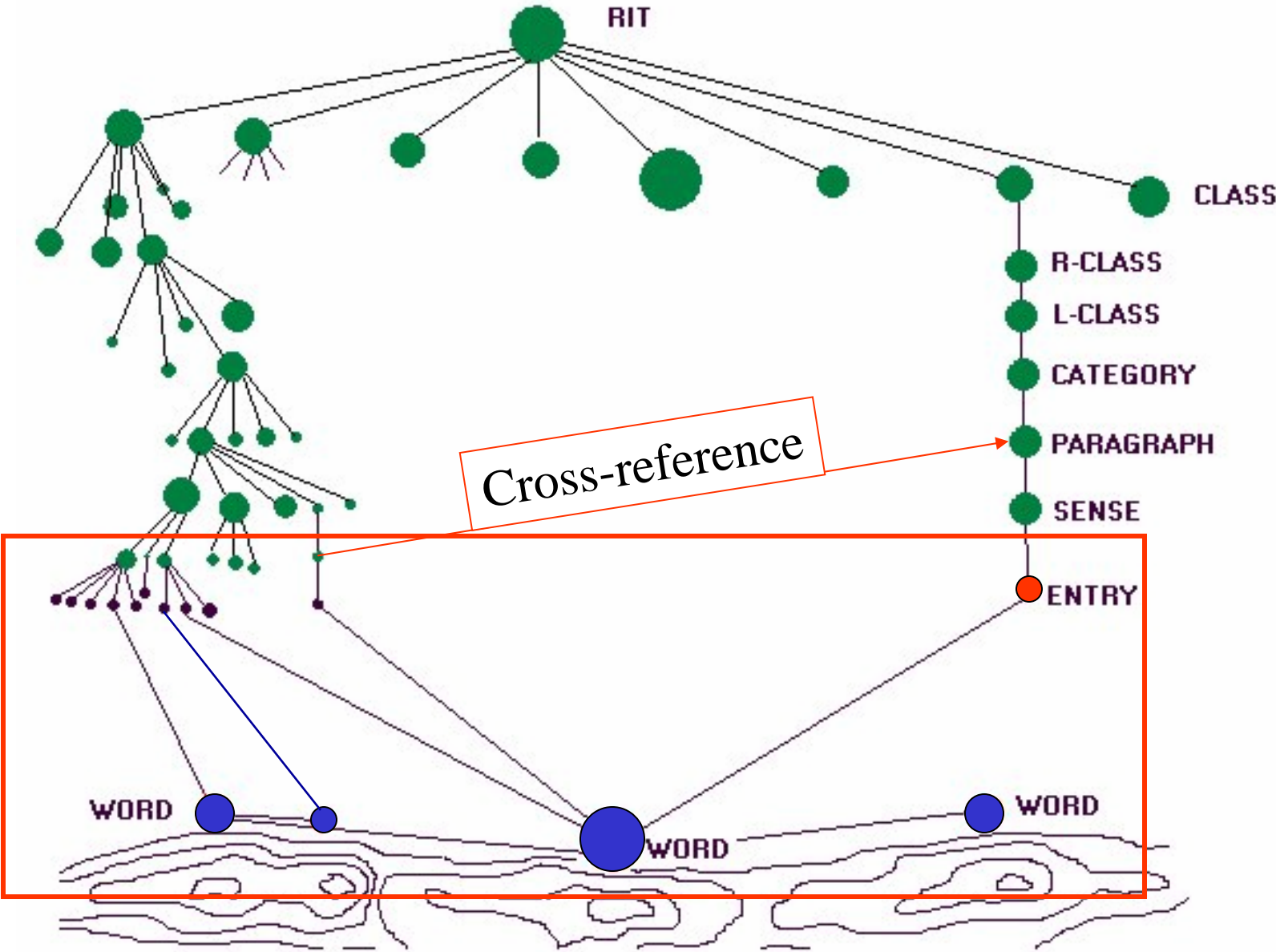


Polysemy

	over	above	beyond	across	past	in excess	additionally	...
9:12:1	x	x						
206:24:	x		x	x	x			
206:27:	x	x	x		x	x		
227:40:	x					x		
40:10:1	x	x	x	x	x	x	x	
...	x	x				x		

ENTRY=40:10:1-over

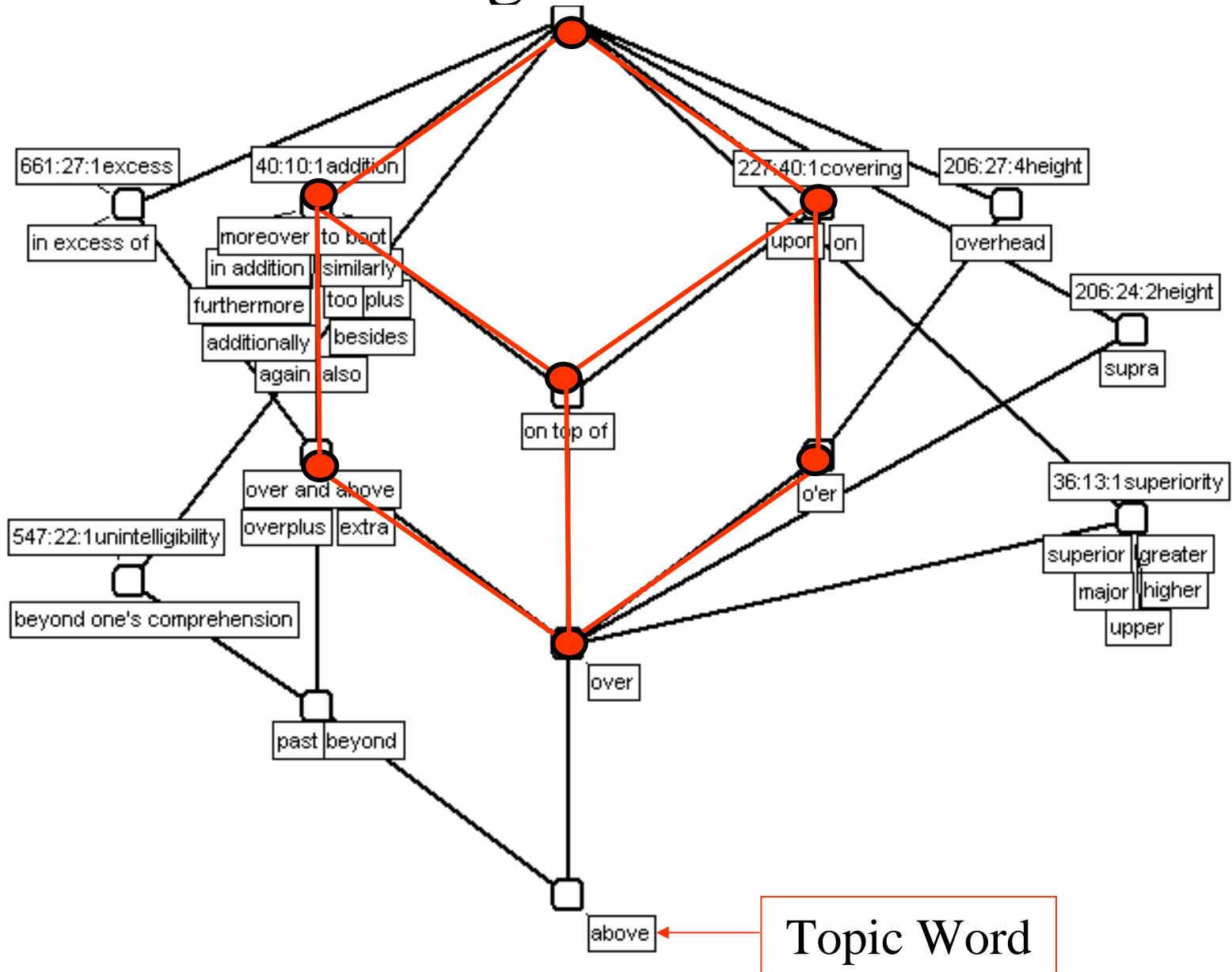
Implicit Structure



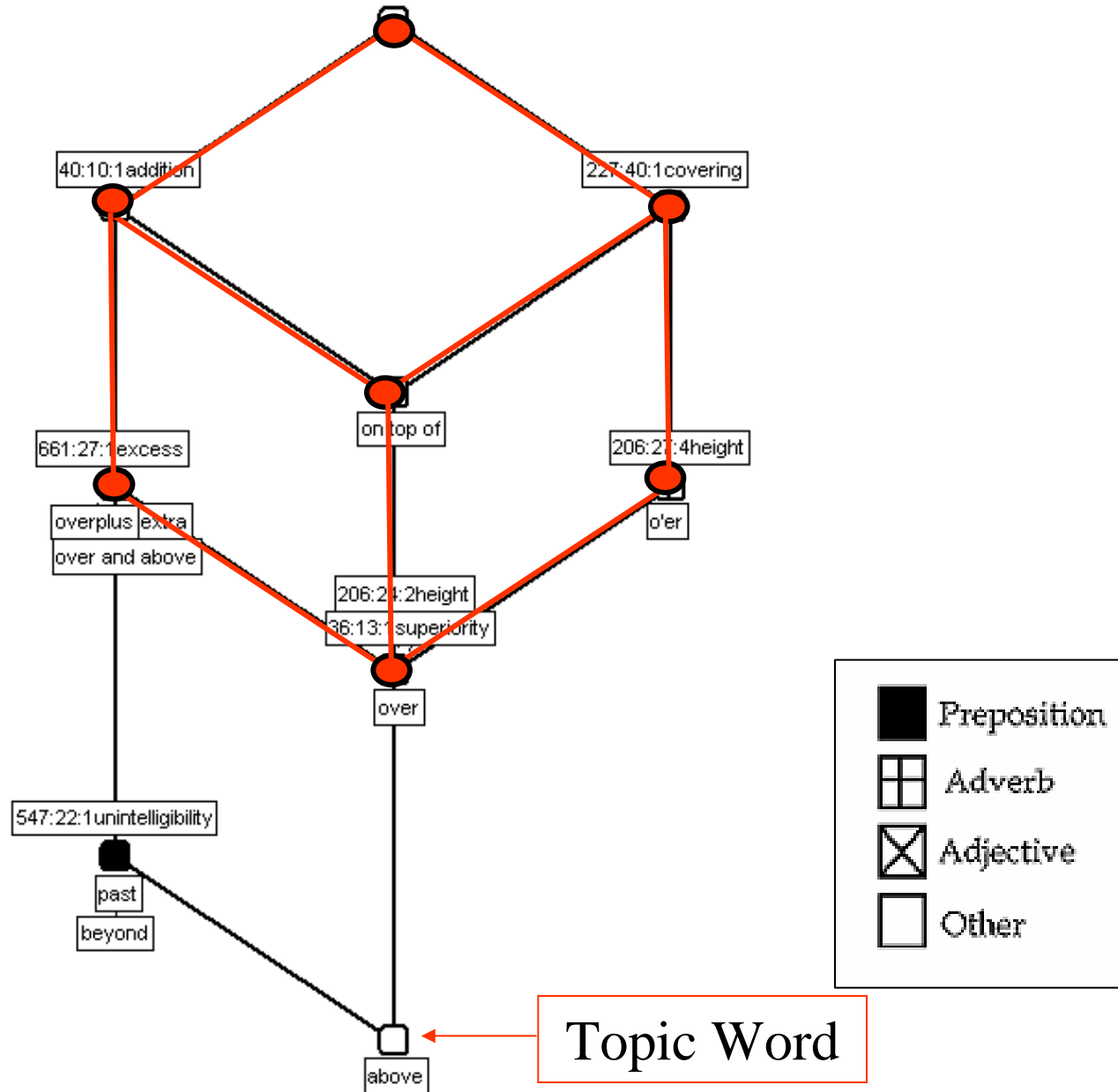
Implicit Structure

Local Semantic Structure

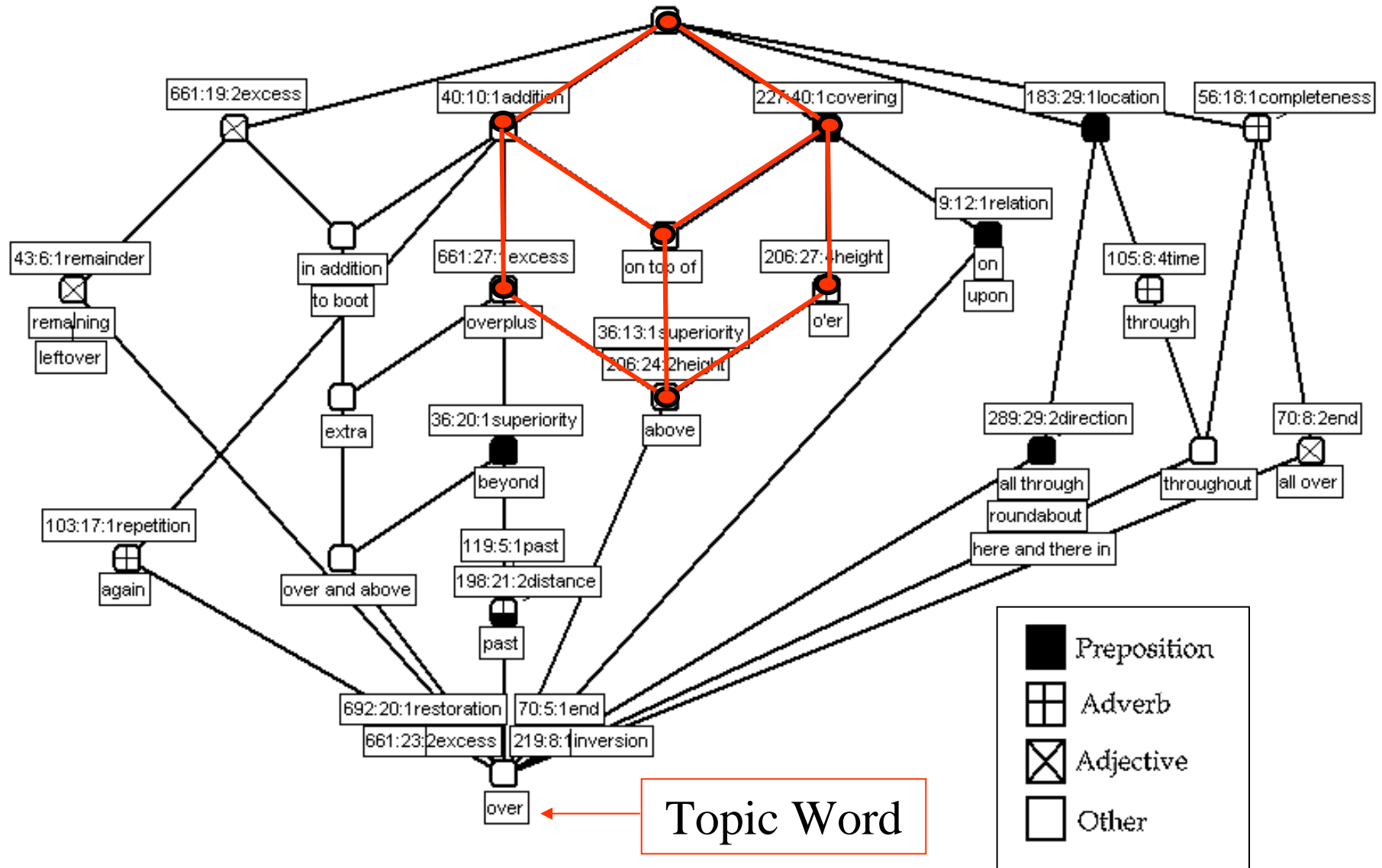
Semantic Neighborhood of *above*



Restricted Neighborhood of *above*



Restricted Neighborhood of *over*

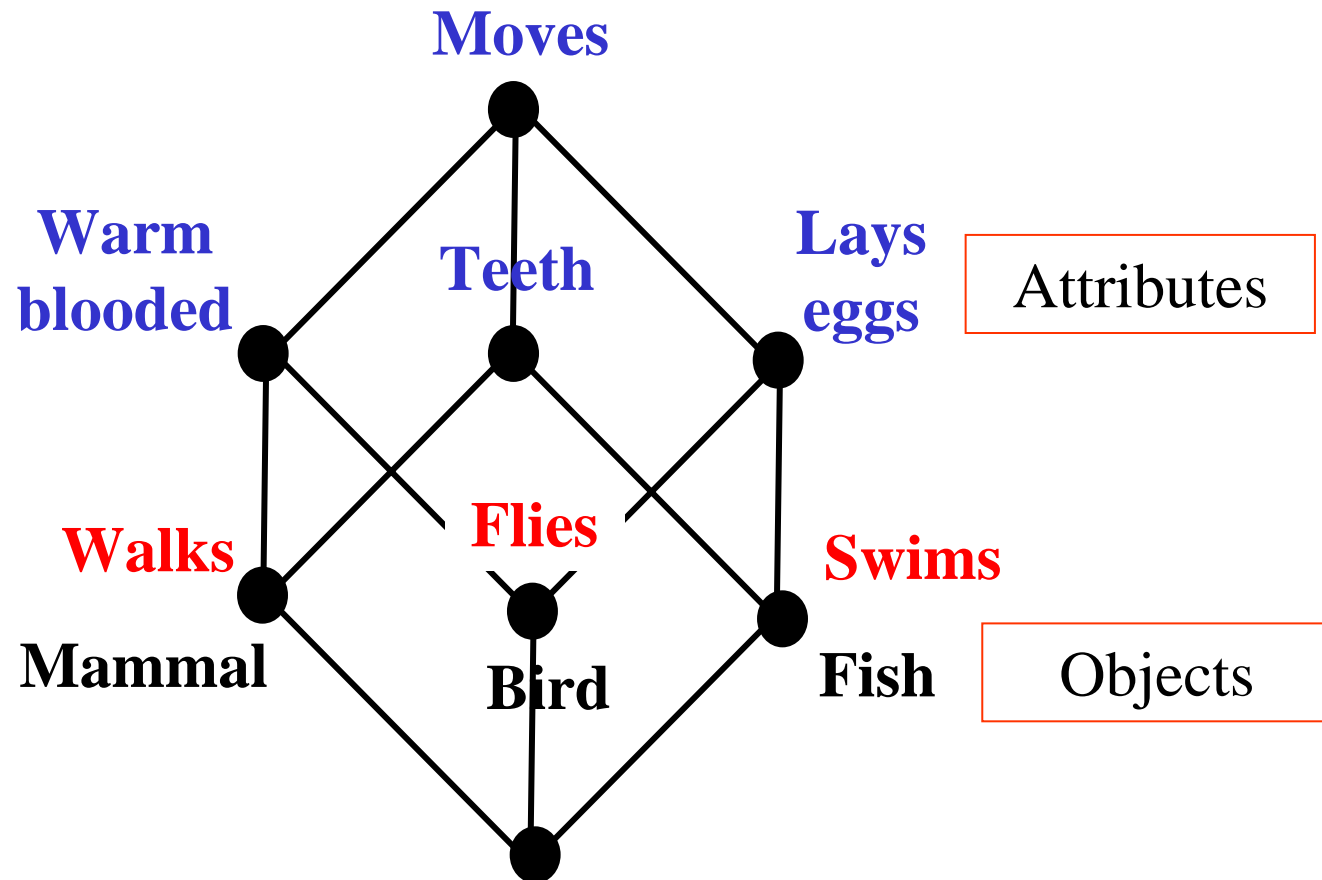


Formal Context

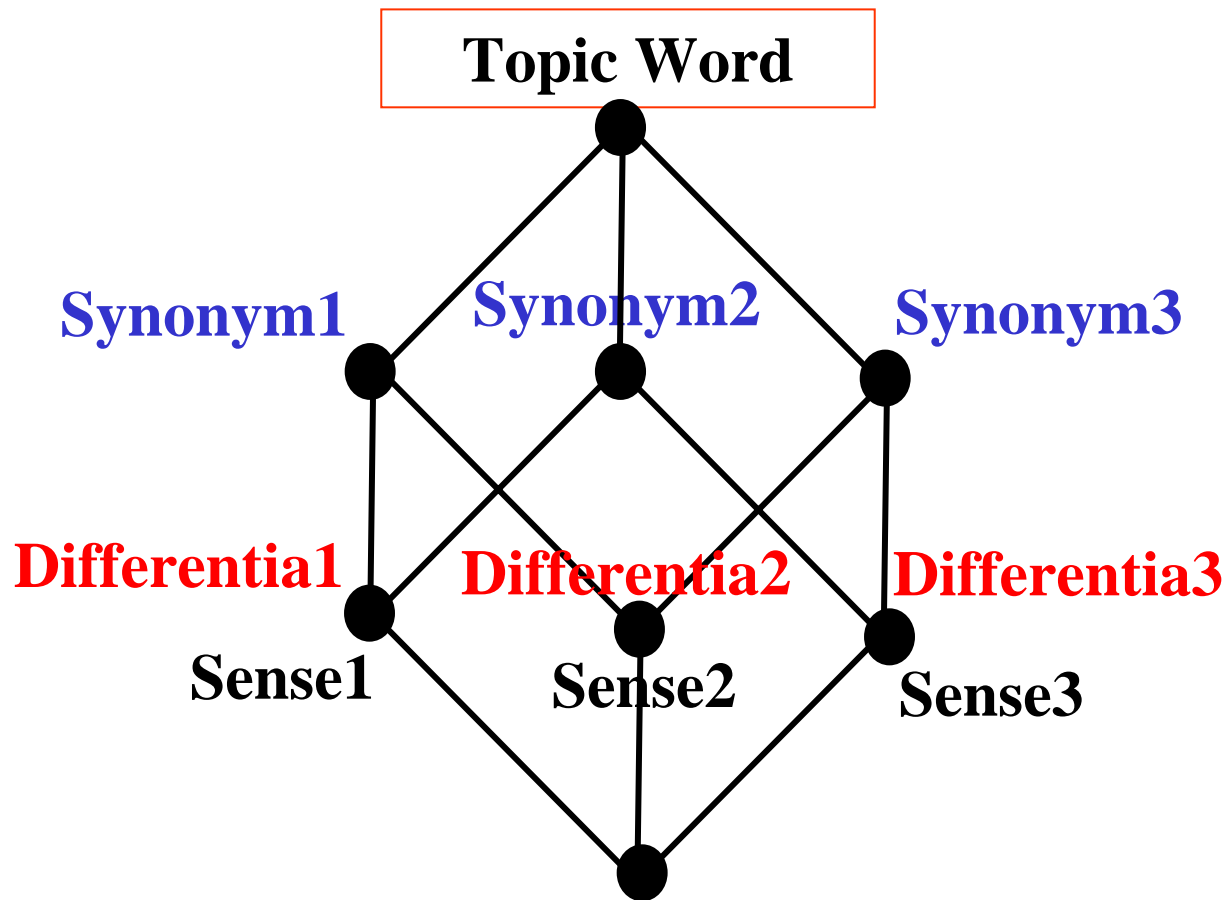
	Bird	Fish	Mammal
Moves	X	X	X
Lays Eggs	X	X	
Warm-blooded	X		X
Has Teeth		X	X
Flies	X		
Swims		X	
Walks			X

Differentiating Attributes

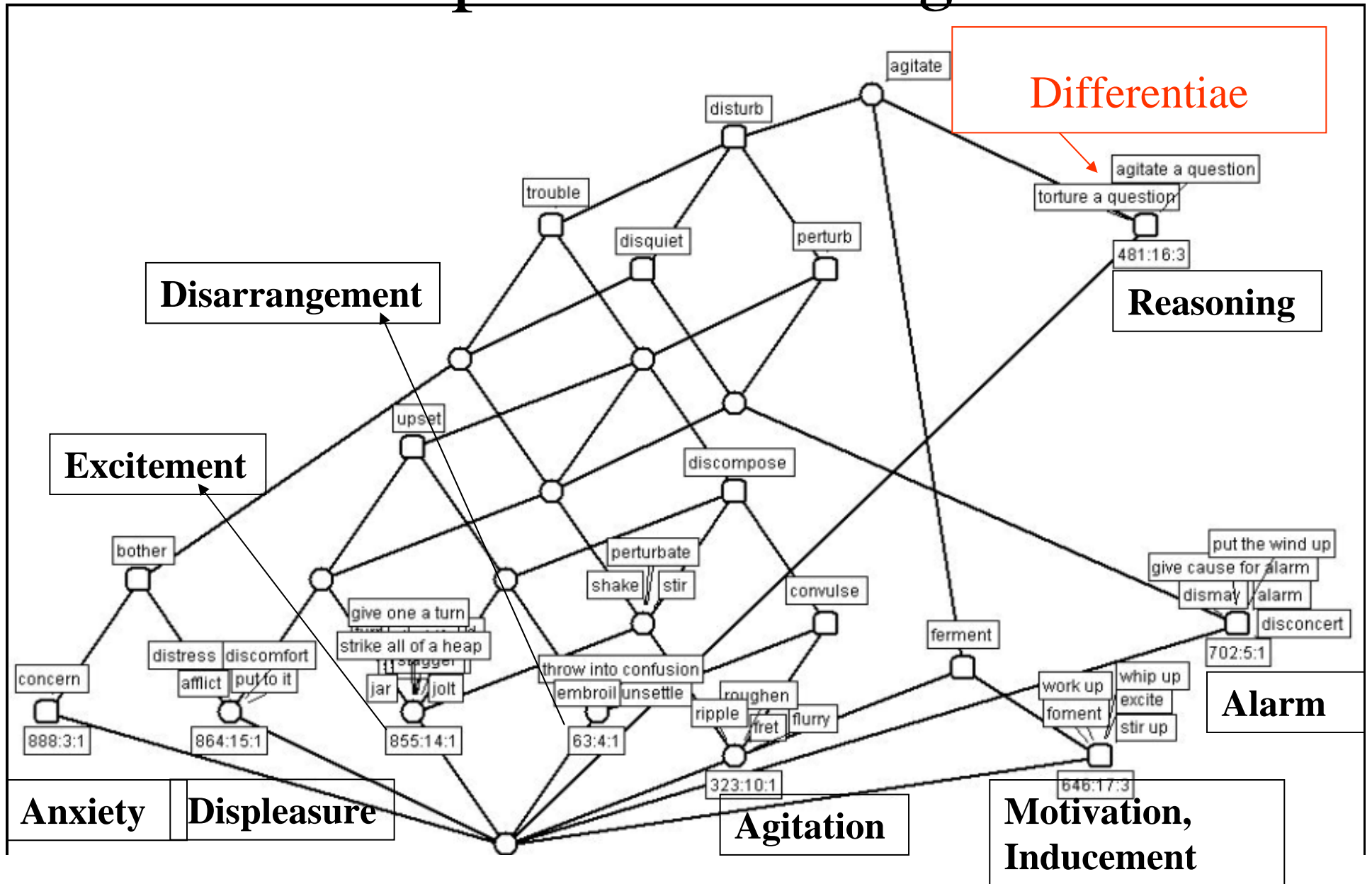
Formal Concept Lattice



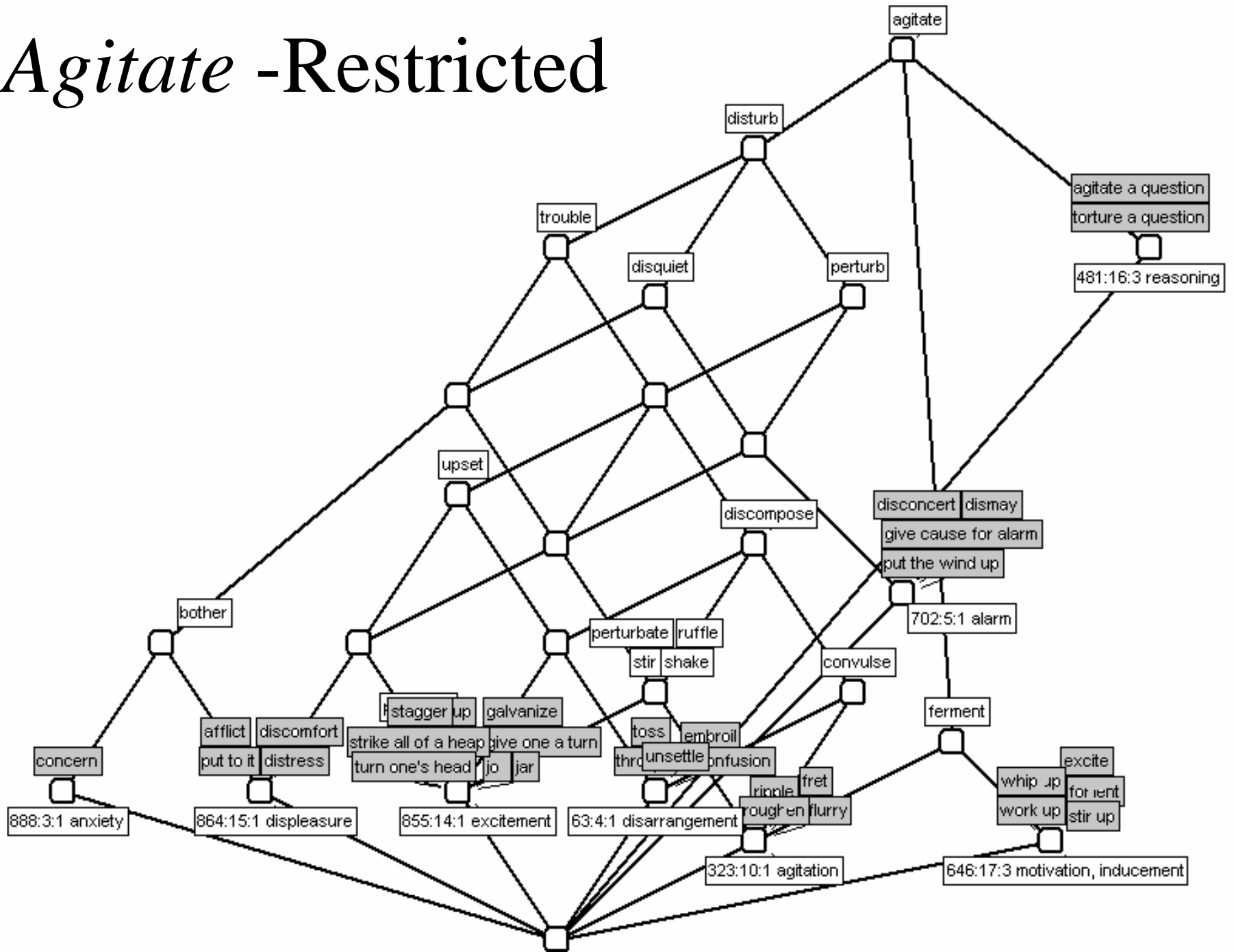
Formal Concept Lattice



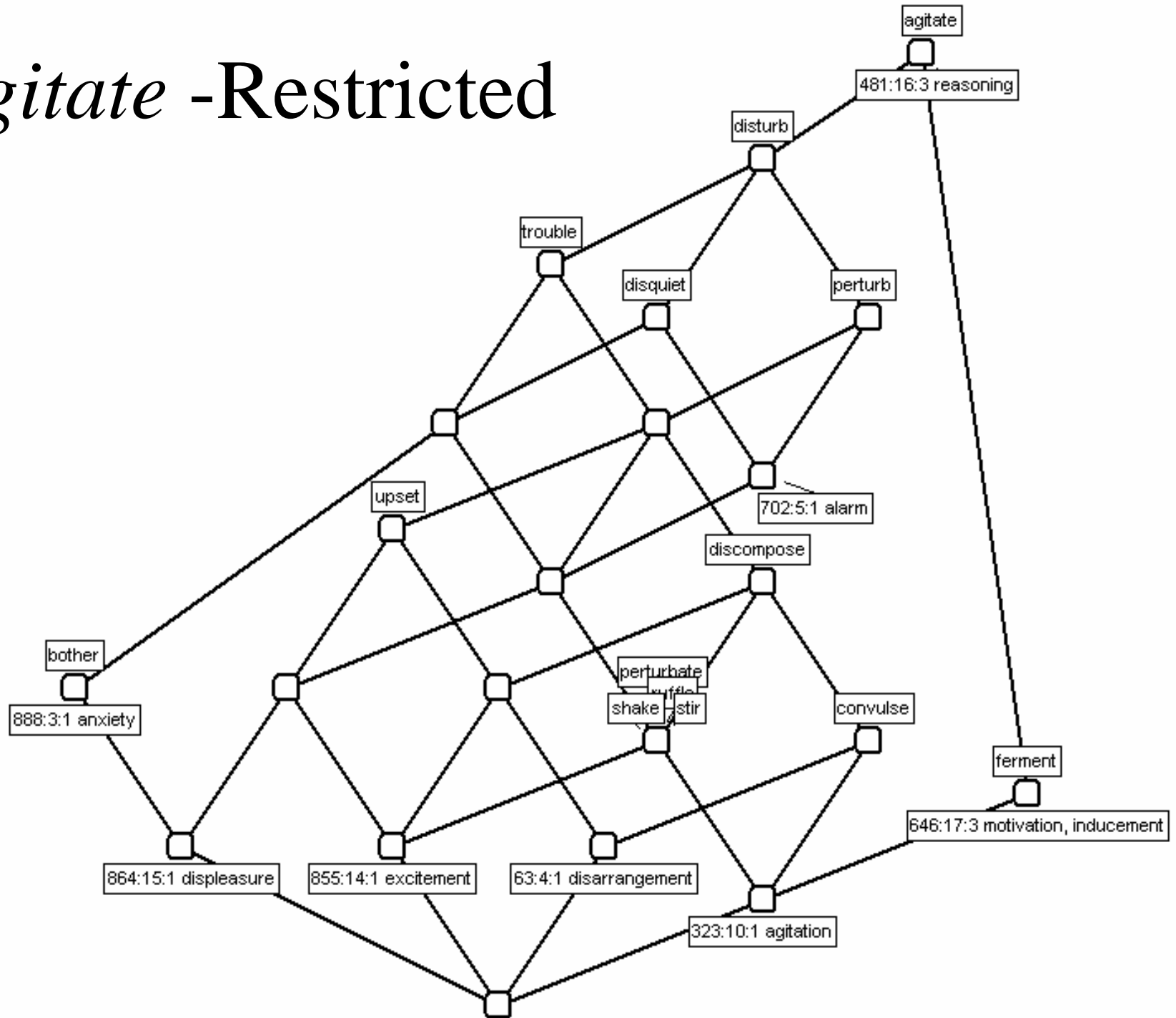
Concept Lattice of *agitate*



Agitate - Restricted



Agitate -Restricted



Relations

Relations

Structural Dimensions of Roget's Semantics

- **Word-Sense** relation (**entry**; chains; components)
- **Word-Word** relations (synonymy, part-of-speech, etymology, and antonymy.)
- **Sense-Sense** relations (polysemy and hypernymy (BT/NT))
- **Subset relations** (implicational structures among words, senses or cross-references)
- **Cross-references** (Synset=>Paragraph and Synset=>Category.)

Comparison

G. Salton (1968) on automatic thesaurus generation:

when words of unequal frequency are included in a thesaurus or represented on an association map... a hierarchical arrangement results almost inevitably, since frequent words can be made into categories, and words of lesser frequency into subcategories (p. 57).

supports the hypothesis that the thesaurus hierarchy, at least from the Category level down, may result naturally from the structure of word associations

Salton, G. (1968). *Automatic information organization and retrieval*. New York: McGraw-Hill.