Concept Neighbourhoods in Knowledge Organisation Systems

Uta Priss, L. John Old

School of Computing
Edinburgh Napier University
www.upriss.org.uk, j.old@napier.ac.uk

February, 2010

Since the 1950s ...

Margaret Masterman (1910-86), a pioneer of AI, NLP, Machine Translation and the founder of the Cambridge Language Research Unit (CLRU), suggested using thesauri and lattices in the area of "Semantic Transformations".

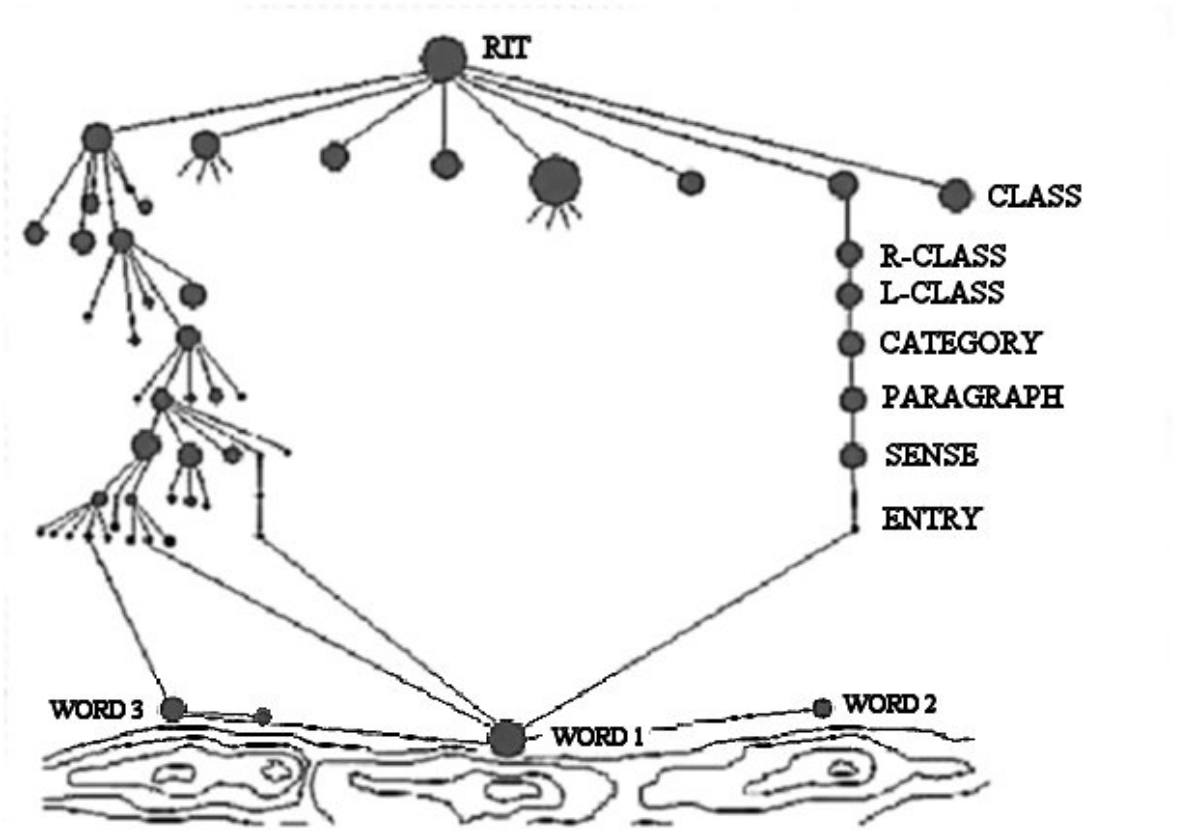
Her thesaurus-based notions are similar to modern research in conceptual structures and her lattices are essentially "neighbourhood lattices".

Four papers from the CLRU at the 1956 International Conference on Mechanical Translation (at MIT) about using a thesaurus as an interlingua in Mechanical Translation (MT).

Modern Versions of Masterman's Lattice-Thesaurus Ideas

- Yarowsky (1992) uses Roget's Thesaurus for word-sense disambiguation.
- Sally Yeates Sedelow's research in the 1960's on Automated Language Analysis uses a machine readable format Roget's Thesaurus (but no lattices). In the early 1990s the Sedelows met Rudolf Wille which lead to the invention of Neighbourhood Lattices.
- Helge Dyvik's (1998) Semantic Mirrors Method extracts semantic information from bilingual corpora. The "translational images" of words are computed. The resulting structures are lattices, (similar to neighbourhood lattices but independently invented).

Roget's Thesaurus (ROGET)



http://www.roget.org

Words, Senses and Entries in ROGET

WORDS ⟶ SYNONYMY

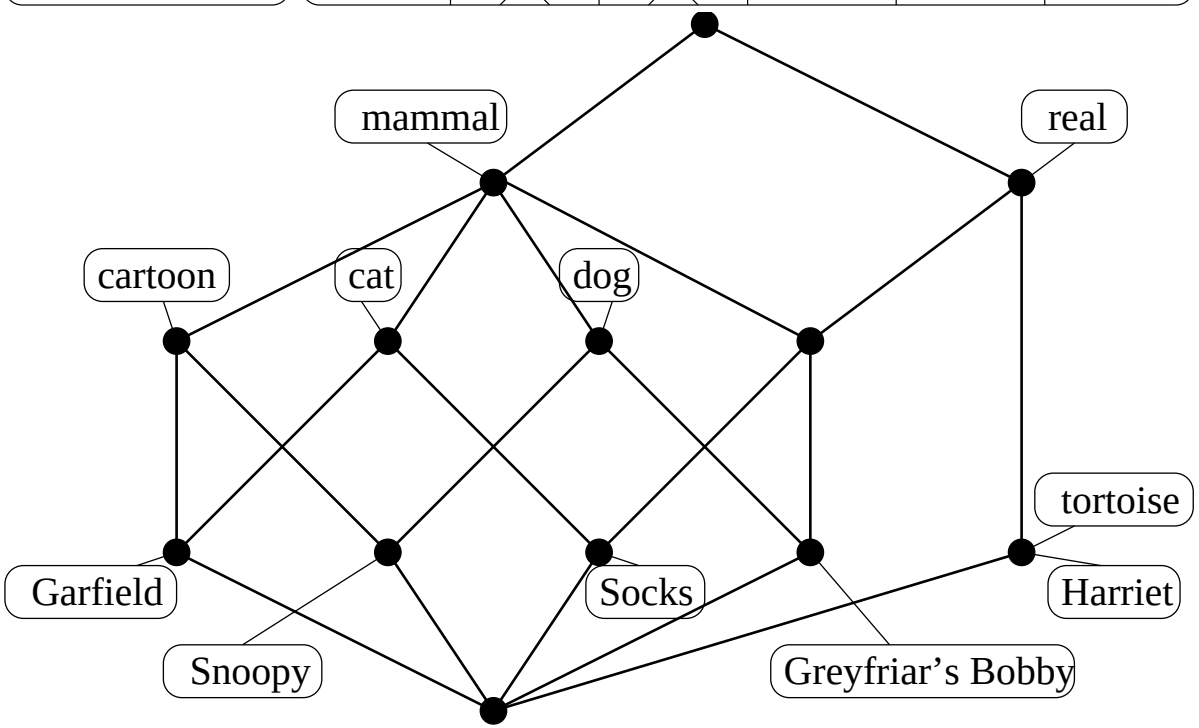
SENSES ↓

	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		

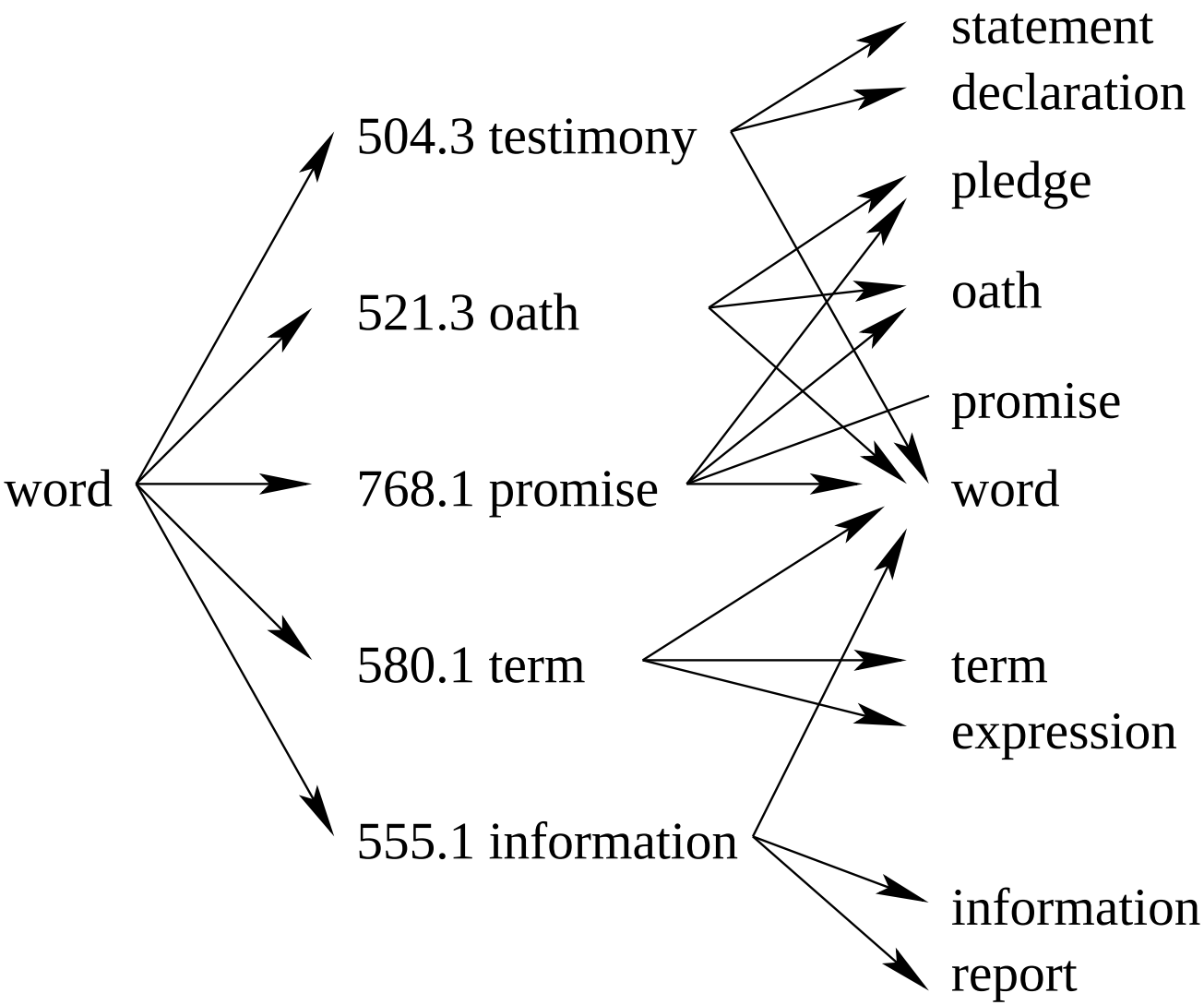
POLYSEMY

Formal Concept Analysis (FCA)

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



The Plus Operator

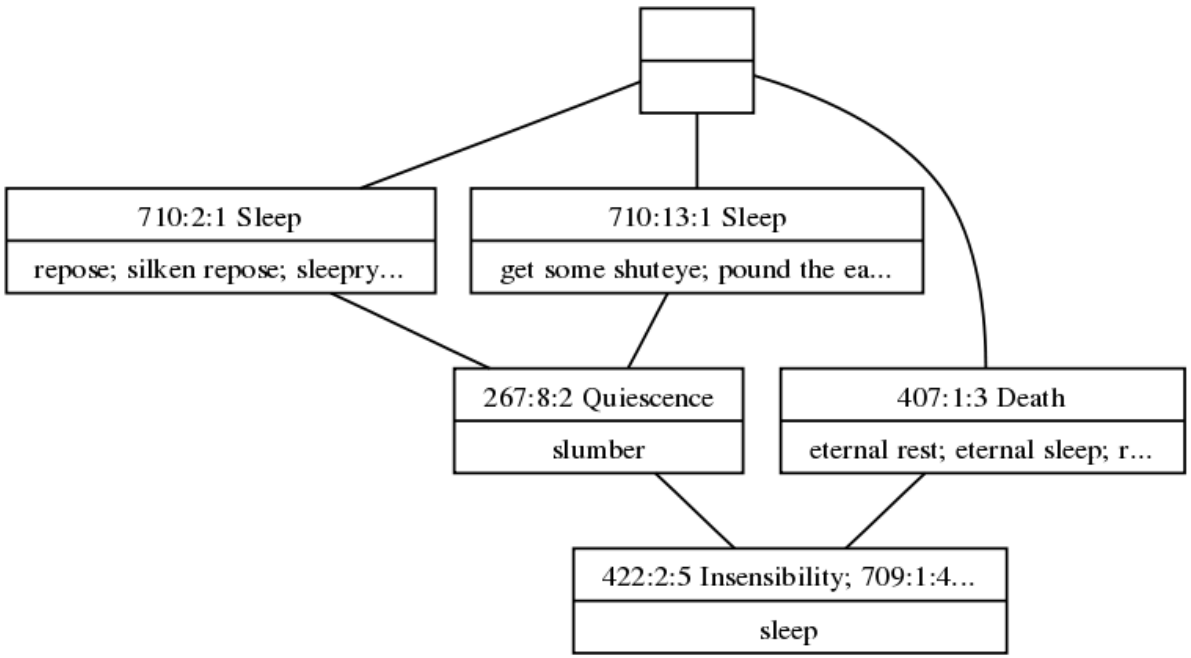


A Neighbourhood Lattice in ROGET

- Formal objects: the plus operator is used 0, 2, 4, ... times
- Formal attributes: the plus operator is used 1, 3, 5, ... times
- Relation: the word/sense relation

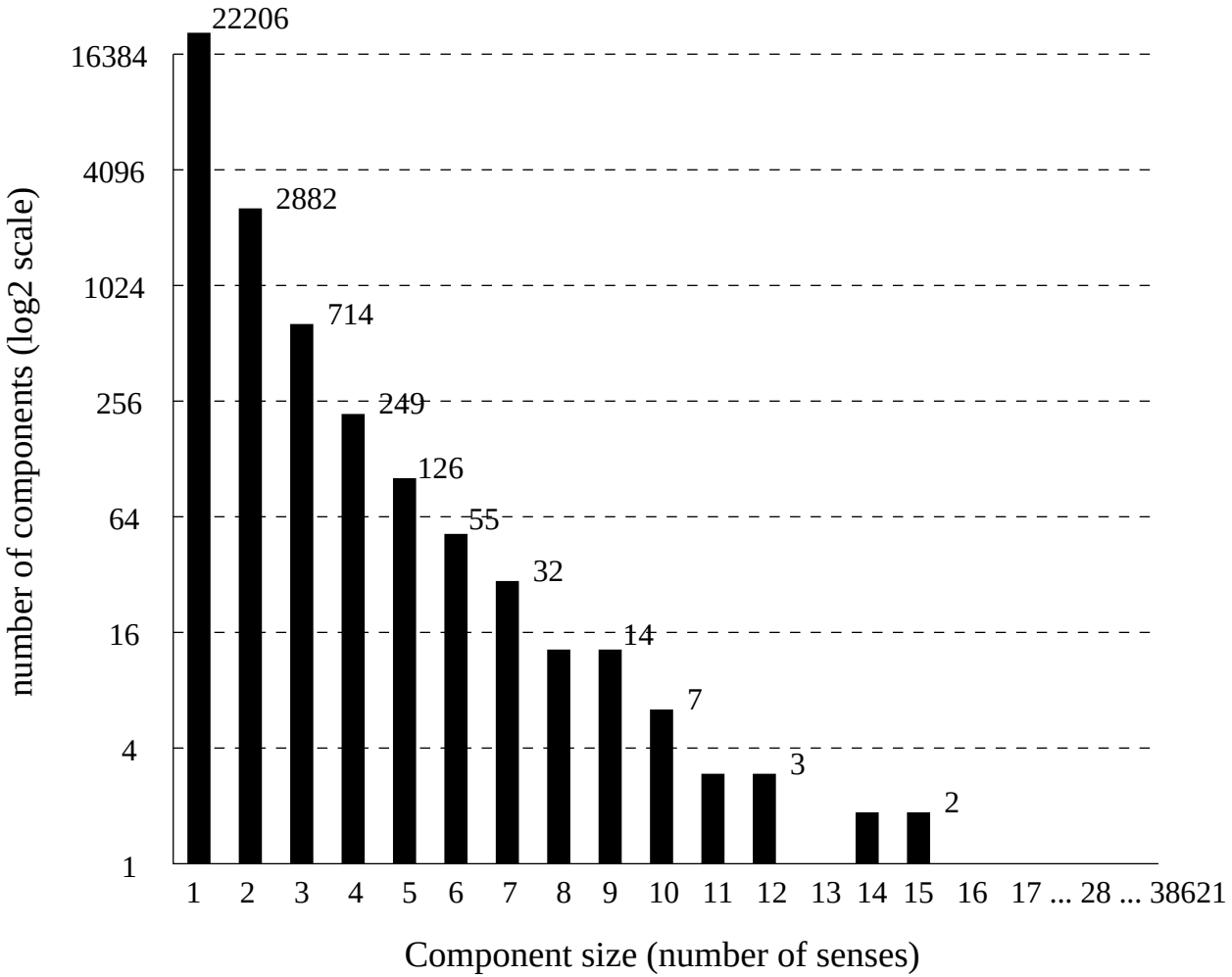
(Use of restriction if the lattice becomes too large.)

A Neighbourhood Lattice in ROGET



`http://www.ketlab.org.uk/roget.html`

Neighbourhood Closure Lattices in ROGET

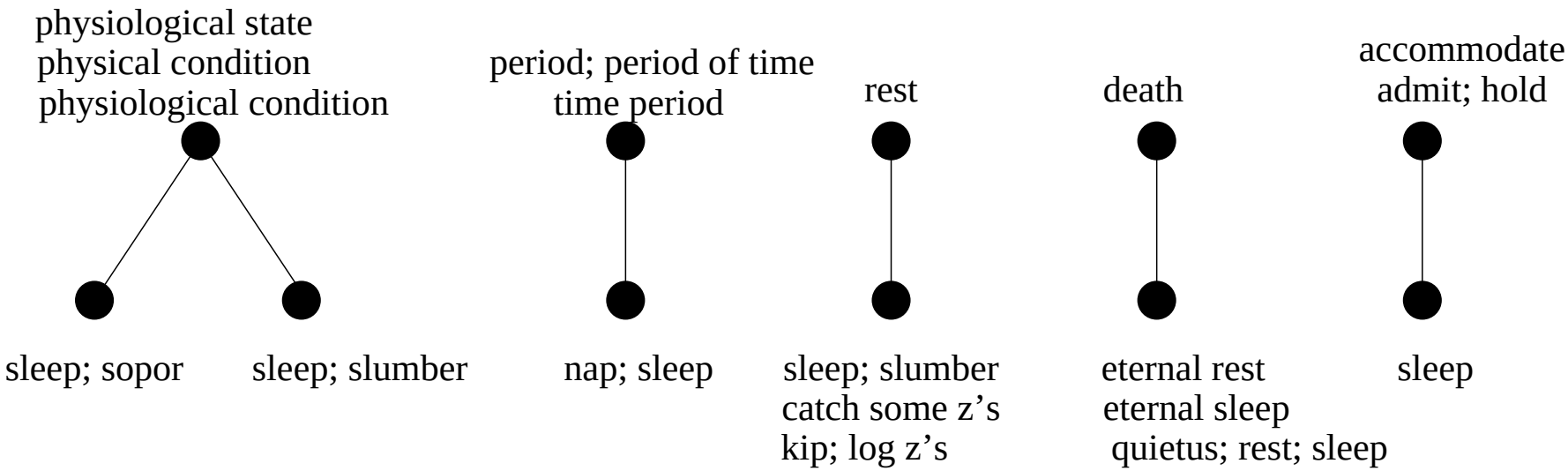


If one keeps applying the plus operator, when does it stop?

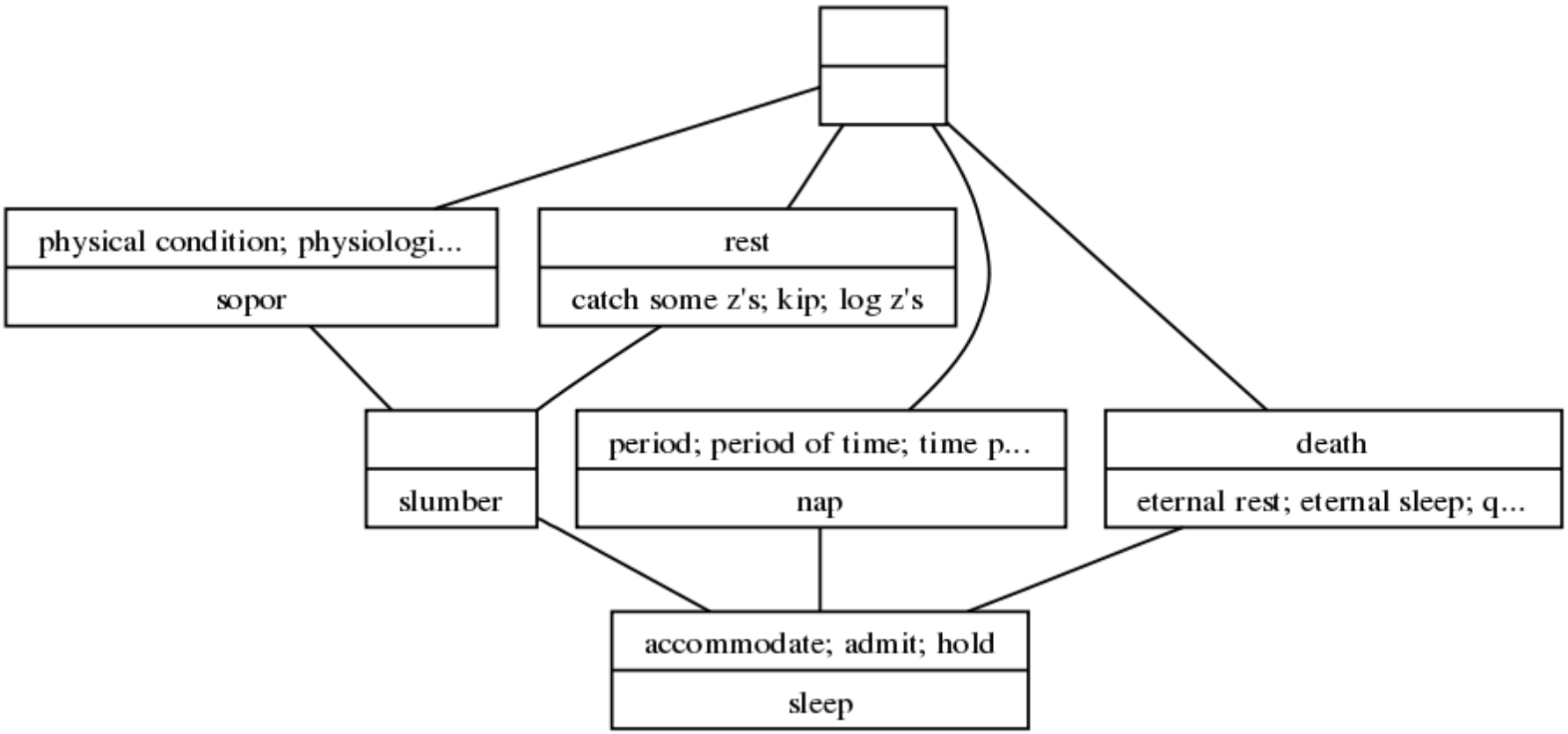
Neighbourhood lattices in WordNet

- WordNet is more finely grained than ROGET
- WordNet contains many different types of relations (therefore there are many possibilities for creating lattices)

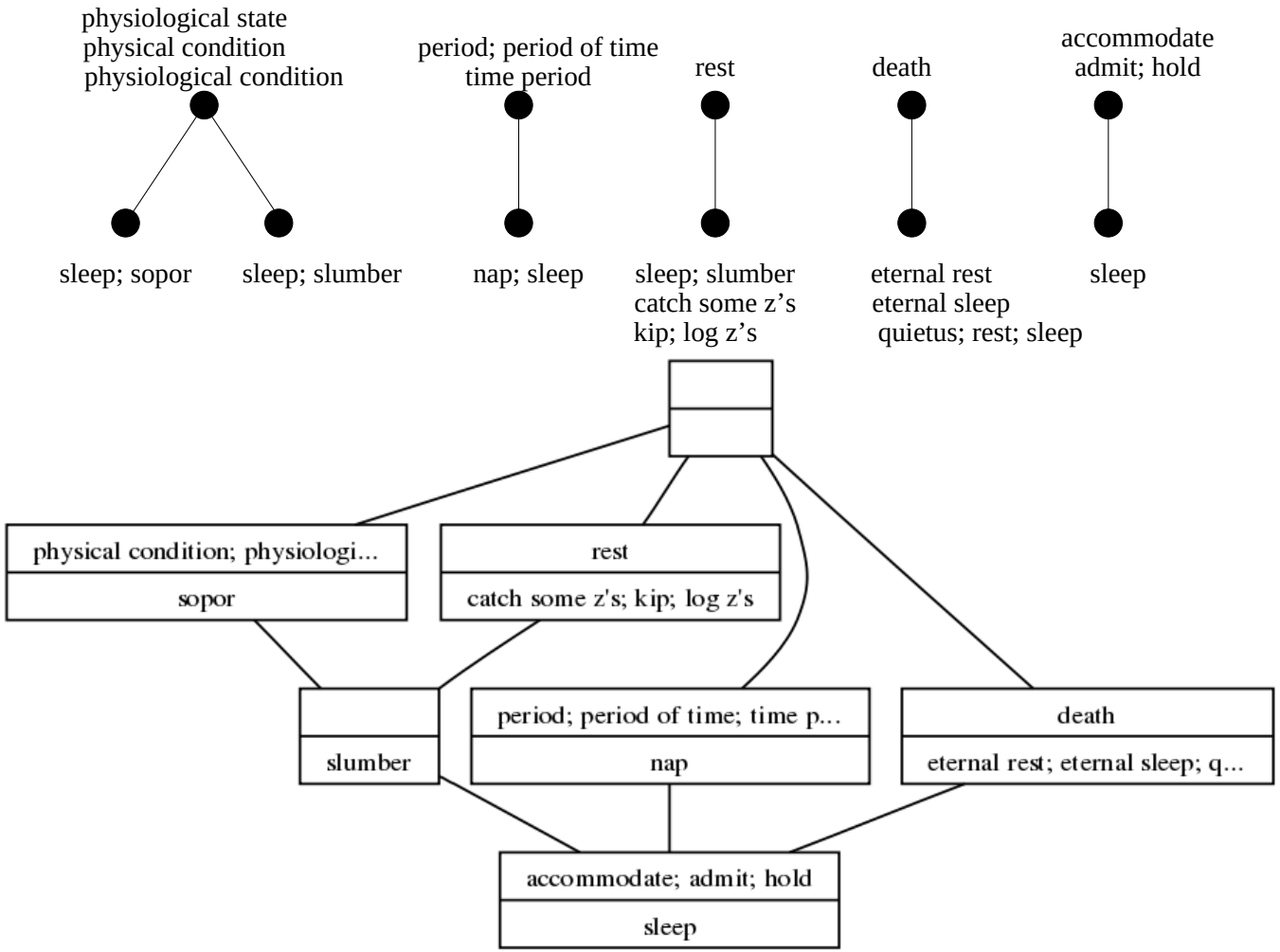
WordNet's Hypernymy Relation



A Neighbourhood Lattice in WordNet



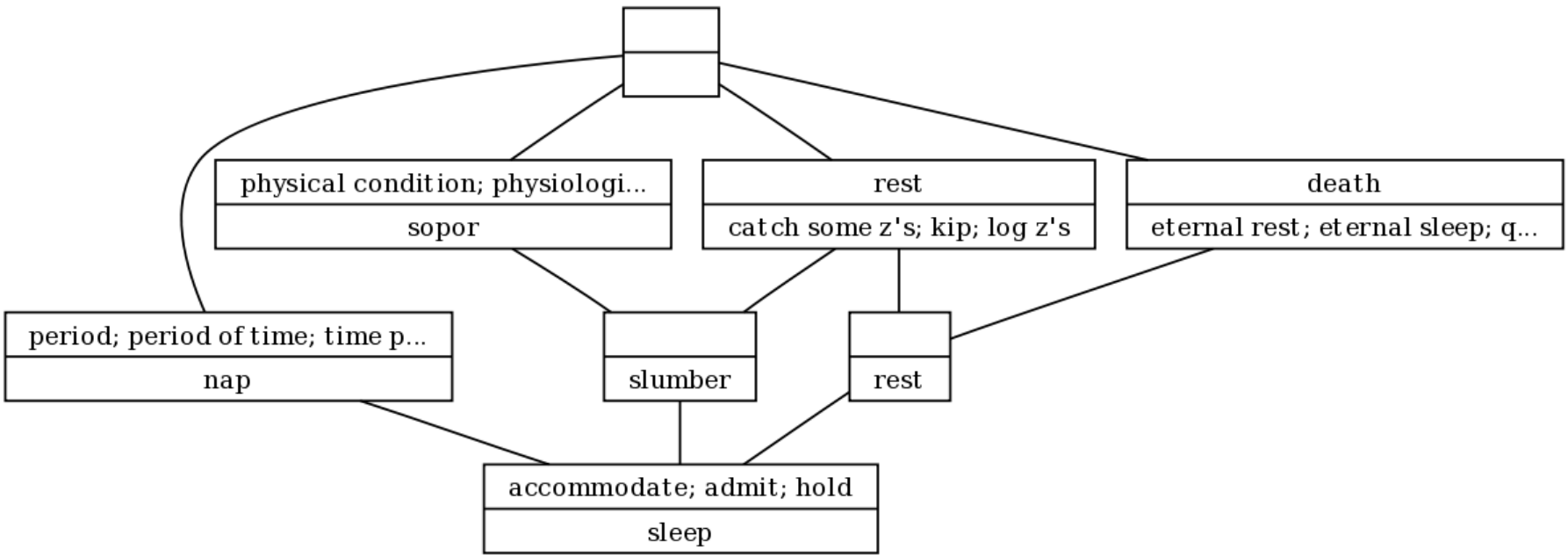
Comparing Hypernymy and Neighbourhoods in WordNet



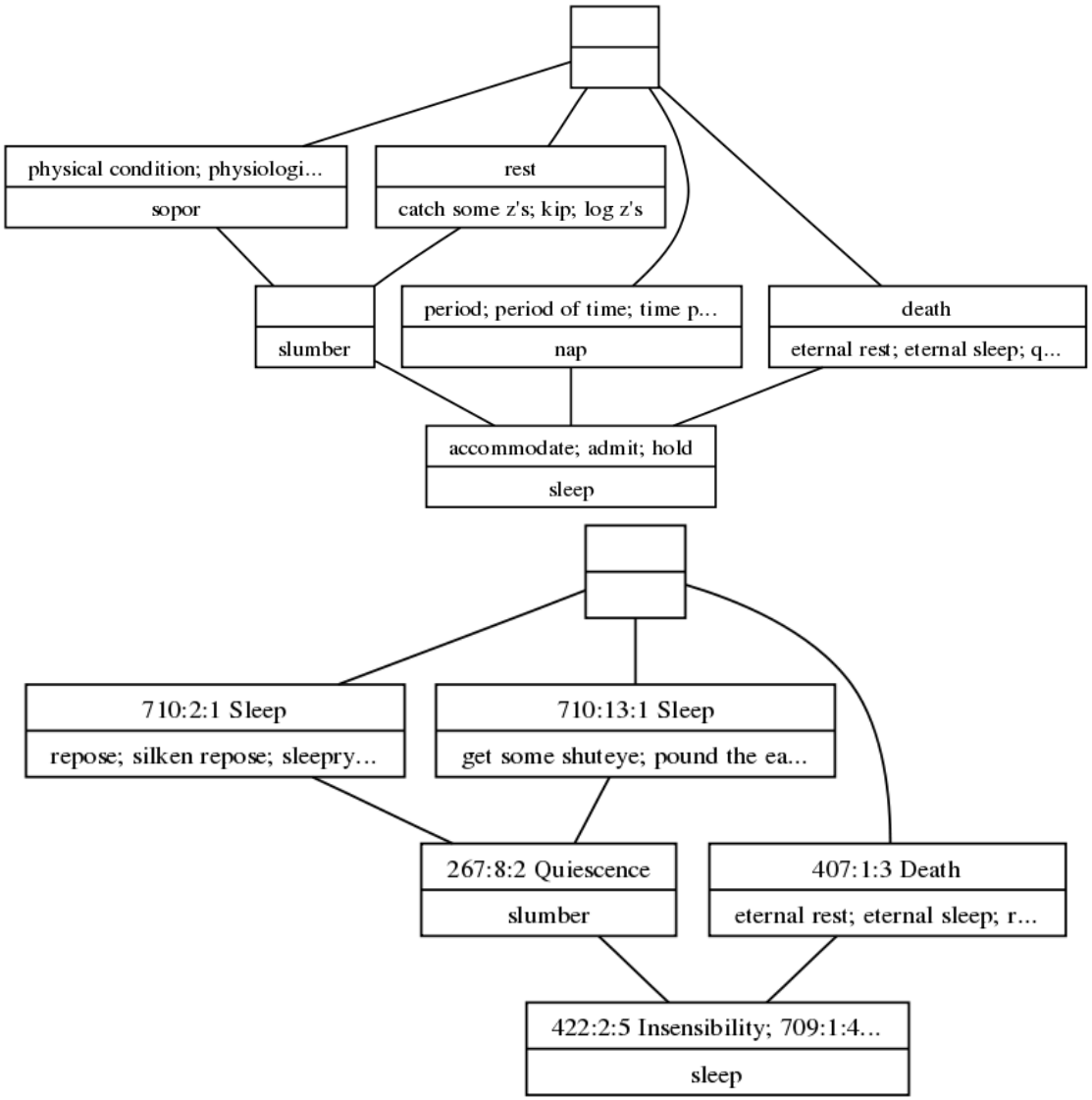
Formally: Neighbourhood Lattices in WordNet

- Formal objects: the words of the synsets belonging to all senses of a word
- Formal attributes: the words of the hypernymic synsets
- Relation: the semantic relation between synsets and their hypernymic synsets is treated as a (lexical) relation between the words in the synsets

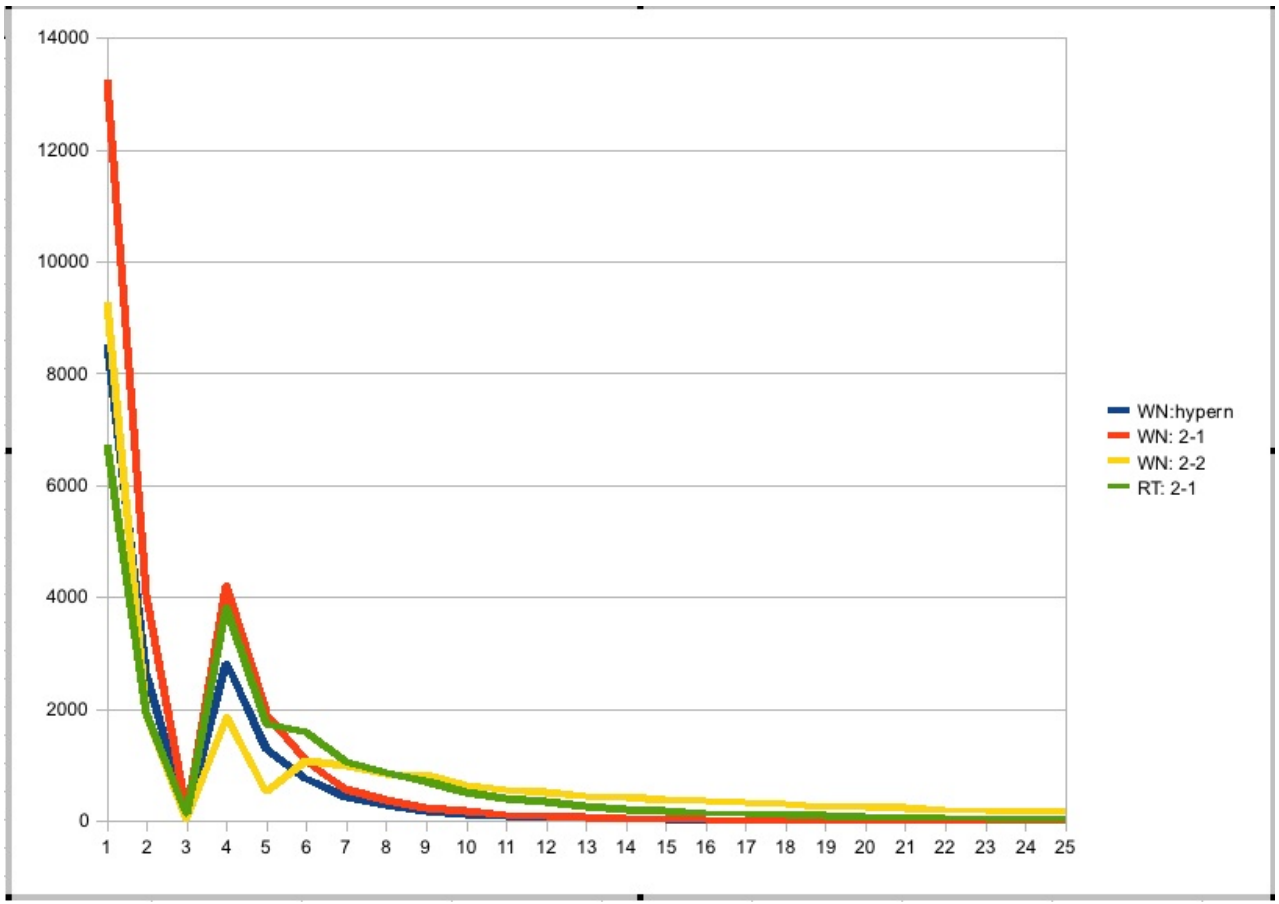
A Neighbourhood Lattice with Identity



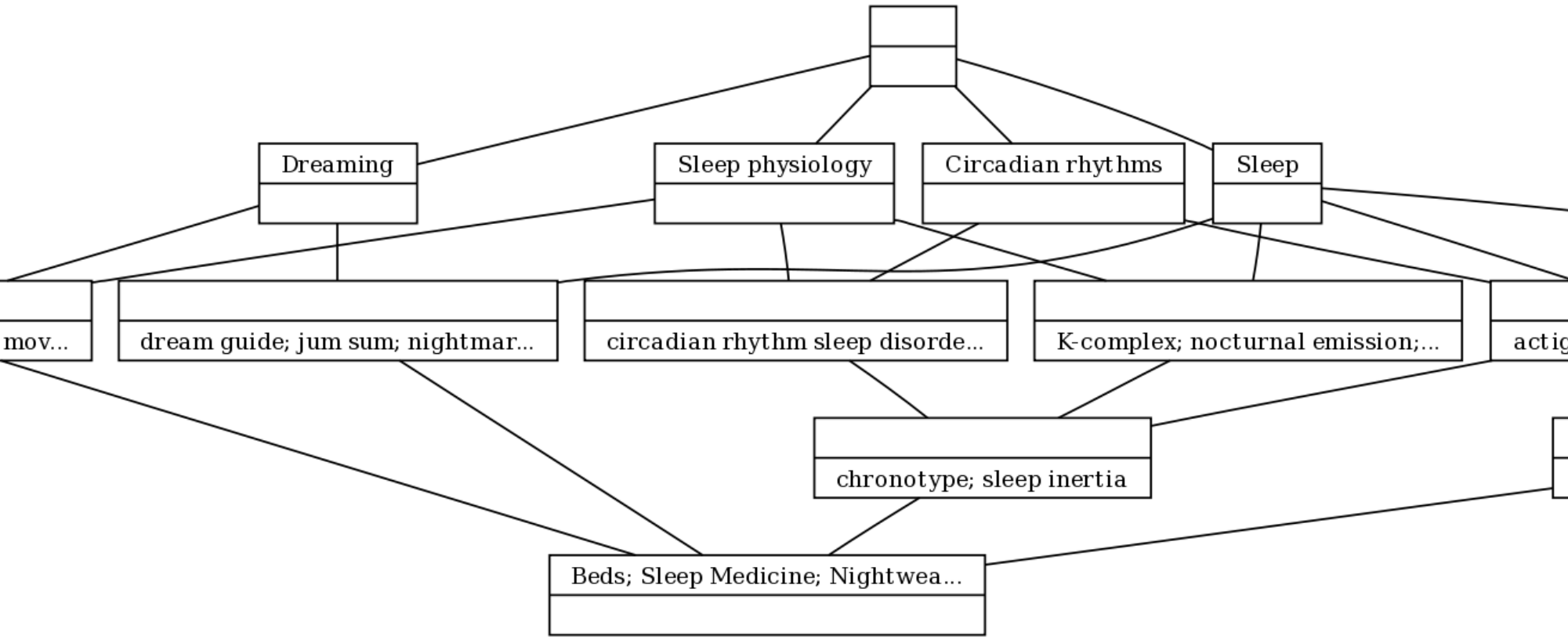
A Neighbourhood in WordNet (top) and ROGET



Sizes of neighbourhood lattices



Wikipedia: Sleep



Conclusion

- Work in Progress
- On-line exploration of neighbourhood lattices in ROGET and WordNet
- Other KOS?
- Connection to small-world networks?

Thank you!