PollyNett

***

Reconstructing Concept Networks on the Basis of Crosslinguistic Polysemy

JOHANN-MATTIS LIST, ANSELM TERHALLE

(Heinrich-Heine-University Düsseldorf)

Outline of the Talk

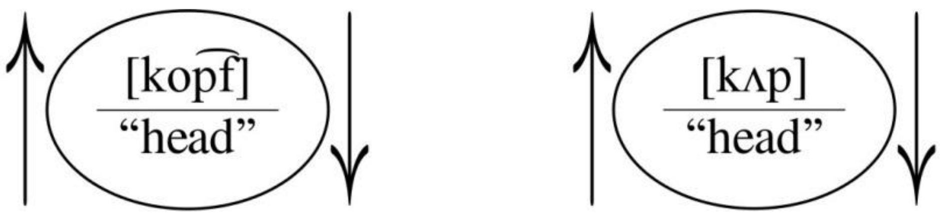
1. Conceptual Structures and Meaning Change
2. Cognitive Historical Semantics
3. PollyNett: Crosslinguistic Polysemy Network
4. The Semantic Potential of PollyNett
5. Concluding Remarks

Outline of the Talk

1. Conceptual Structures and Meaning Change
2. Cognitive Historical Semantics
3. PollyNett: Crosslinguistic Polysemy Network
4. The Semantic Potential of PollyNett
5. Concluding Remarks

Sign Model

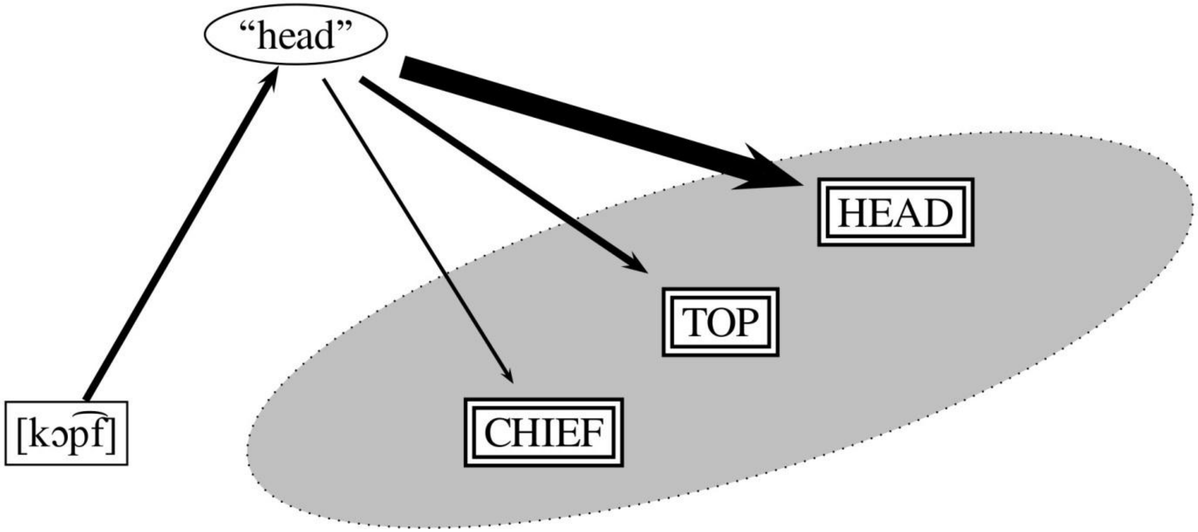
- Our aim is to reconstruct a conceptual network on the basis of polysemous words, i.e. combinations of sound chains with two or more meanings
- For this, we need a sign model which includes at least a sound-chain component and a meaning component



- Our study is based on linguistic data from 195 languages. As these data are semantically aligned, we disregard the fact that *meaning* and *conceptual frame* are different – even though strongly related, the meaning being some sort of abstraction from the frame (Locke 1690, Blank 1997, Löbner 2003) – and consider only the meaning component

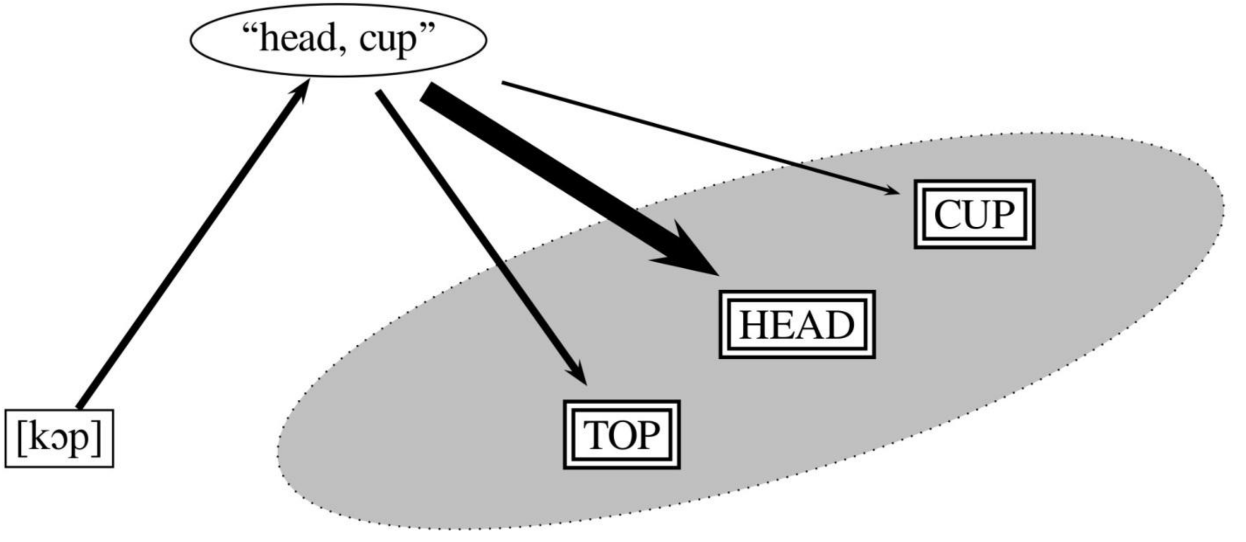
Sign Model

- To cover polysemy, it makes sense to add the notion of *reference potential* to our model: a given meaning allows speakers to refer to things which are associated in some way with the meaning of the word even if they are not instantiations of this meaning



Meaning Change

- Under certain circumstances, the intensive use of a word for members of its reference potential can change the word's meaning



- This kind of meaning change leads to polysemy

Meaning Change

- Meaning change leading to *polysemy* is assumed to be motivated by the conceptual relation between the meanings of the word
- Other possible cases of sound chains related to more than one meaning are *homonymy* (accidental correspondance between the sound chains of two words) and *underspecification* (no linguistic differenciation between two concept which are taxonomically related)
- As homonymy is relatively rare in comparison to polysemy and underspecification, we make the following – slightly simplifying – working assumption

Sound chains with two or more meanings strongly suggest that there is a conceptual relation between these meanings

Outline of the Talk

1. Conceptual Structures and Meaning Change
2. (Cognitive) Historical Semantics
3. PollyNett: Crosslinguistic Polysemy Network
4. The Semantic Potential of PollyNett
5. Concluding Remarks

Analysis of Meaning Change

- Available data
  - The data on which the analysis of meaning change is based consists of *semantic states*, i.e. pairs consisting of a sound chain and a meaning
- Relation between semantic states
  - Two semantic states are considered as related, if there is a genetic relation between the sound chains
  - Remark: These sound-chain relations have also been deduced from *sound states* and assumptions on sound change regularities

A pair of semantic states is then analyzed with respect to a possible relation between the involved meanings (or the related conceptual frames) and possible triggers of the meaning change

Traditional and Cognitive Historical Semantics

- Antiquity:
  - Tropes and their habitualization (Quintilian, Cicero, but also Lausberg, 1960)
- Transfer between rhetoric tropes and meaning change regularities (Reisig 1972)
- Traditional historical semantics (Paul 1880, Bréal 1897, Nyrop 1913)
  - typologies of semantic change based mainly on rhetoric and logic categories
  - mainly aiming at facilitating etymological research (Blank 1997)
  - first appearances of psychological criteria (Wundt 1900, Roudet 1921)
- Structuralist historical semantics (Trier 1931, Dornseiff 1954)
- Cognitive historical semantics
  - foundation of typologies on cognitive principles (Ullmann 1951, Traugott 1985, Santos Domínguez & Espinoza Elorza 1996)
  - influence of prototype semantics (Geeraerts 1983, 1992)

Traditional and cognitive historical semantics rely on the study of individual cases of semantic change which are classified according to rhetoric, logic and/or cognitive criteria.

Quantitative Historical Semantics

- **The semantic-map approach in typology (Cysouw 2010)**

  "**[C]ross-linguistic variation in the expression of meaning can be used as a proxy to the investigation of meaning itself**. […] Thus, the assumption is that when the expression of two meanings is similar in language after language, then the two meanings themselves are similar. Individual languages might (and will) deviate from any general pattern, but when combining many languages, overall the cross-linguistic regularities will overshadow such aberrant cases." (Cysouw 2010: 74)

- **Semantic-map approach as a heuristic device in automatic cognate detection (Steiner et al. 2011)**

  "[S]imilar meanings have a larger probability to be expressed similarly in human language than different meanings. Individual languages might (and will) deviate strongly from general trends, but on average across many languages the formal similarity in the linguistic expression of meaning will reflect the similarity in meaning itself." (Steiner et al. 2011: 12f)

Our approach basically follows up this idea, but it is based on a dramatically increased data basis that allows us to fully exploit the semantic potential of cross-linguistic polysemy networks (PollyNets).

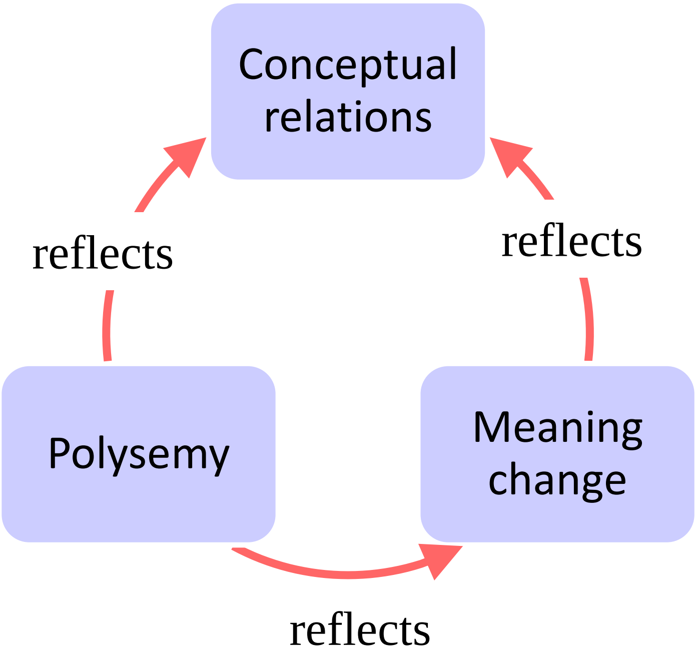
Outline of the Talk

1. Conceptual Structures and Meaning Change
2. Cognitive Historical Semantics
3. PollyNett: Crosslinguistic Polysemy Network
4. The Semantic Potential of PollyNett
5. Concluding Remarks

PollyNett

Basic idea of large polysemy-based networks (PollyNett)

- Meaning change is assumed to be based on relations between concepts
- Thus, **meaning change is a symptom of conceptual relations**
- Meaning change leads to polysemy
- Thus, **polysemy is a symptom of meaning change**
- Polysemy is a universal linguistic phenomenon
- Thus, the analysis of polysemy tells us something about universal, **language family-specific** or **language specific** meaning changes and conceptual structures

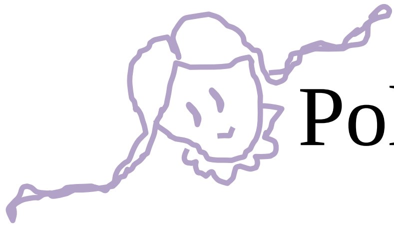


PollyNett

Data preparation

1. Data Basis
   - 195 languages (44 families) from three different sources:
     - IDS: 133 languages (Key and Comrie 2009)
     - WOLD: 30 languages (Haspelmath and Tadmor 2010), and
     - LOGOS: Logos Group (2008)
   - 946 semantic items (meanings)
     - Extracted as the most frequent semantic items from the 1310 items used in the IDS
2. Data Conversion
   - Cleaning the data with help of specifically written Python scripts
   - Identifying similar patterns of polysemy and storing them in networks with help of Python scripts
3. Data Enrichment
   - Tagging (for specific semantic items, part of speech, etc.)
4. Data Analysis
   - using Python Networkx (Hagberg et al.2008) for internal creation and manipulation of networks
   - using Cytoscape (Smoot et al. 2011) for visualization and extended network operations

Data (input data, scripts, and network representation) is not yet published online but we gladly share it upon request…



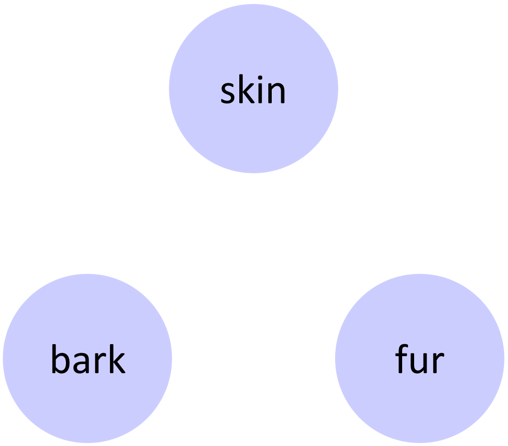
PollyNett

Structure of the data PollyNet is based on

Key	Meaning	Russian	German
1.1	world	mir, svet	Welt
1.21	earth, land	zemlja	Erde, Land
1.212	ground, soil	počva	Erde, Boden
1.213	dust	pyl	Staub
1.214	mud	grjaz	Dreck
1.420	tree	derevo	Baum
1.430	wood	derevo	Wald
…	…	…	…

PollyNett: Construction Principle

Examplary conceptual subspace

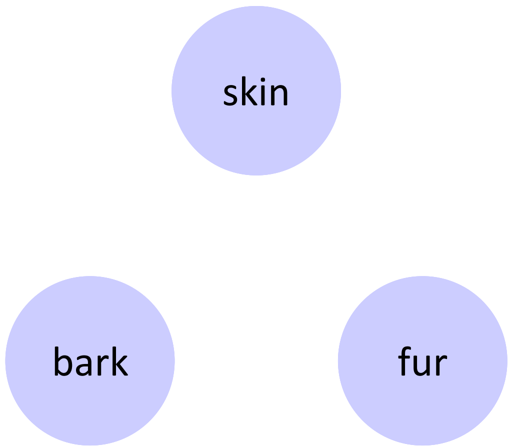


PollyNett: Construction Principle

Three languages which verbalize these concepts

deu

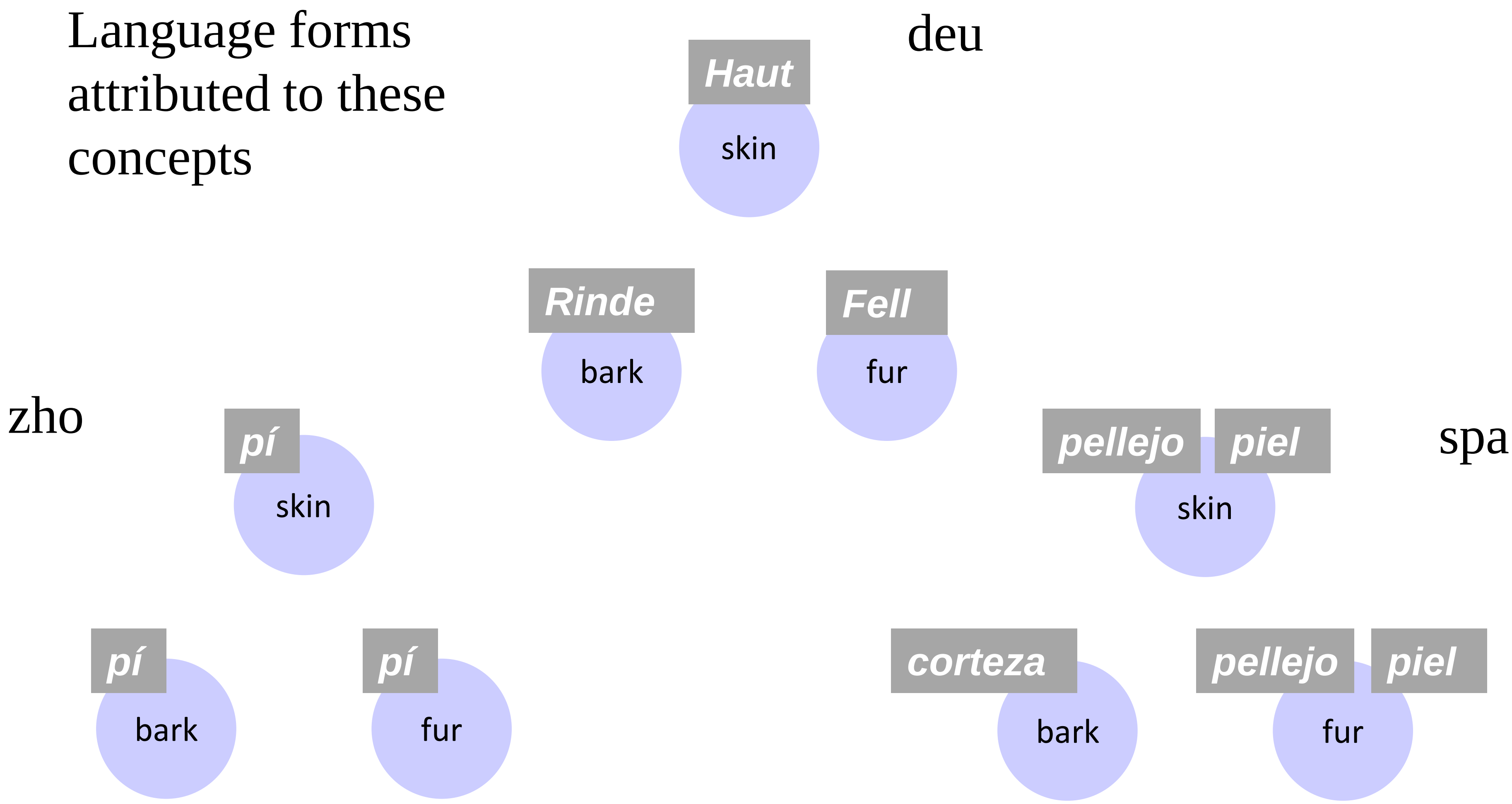
zho



spa

PollyNett: Construction Principle

Language forms attributed to these concepts

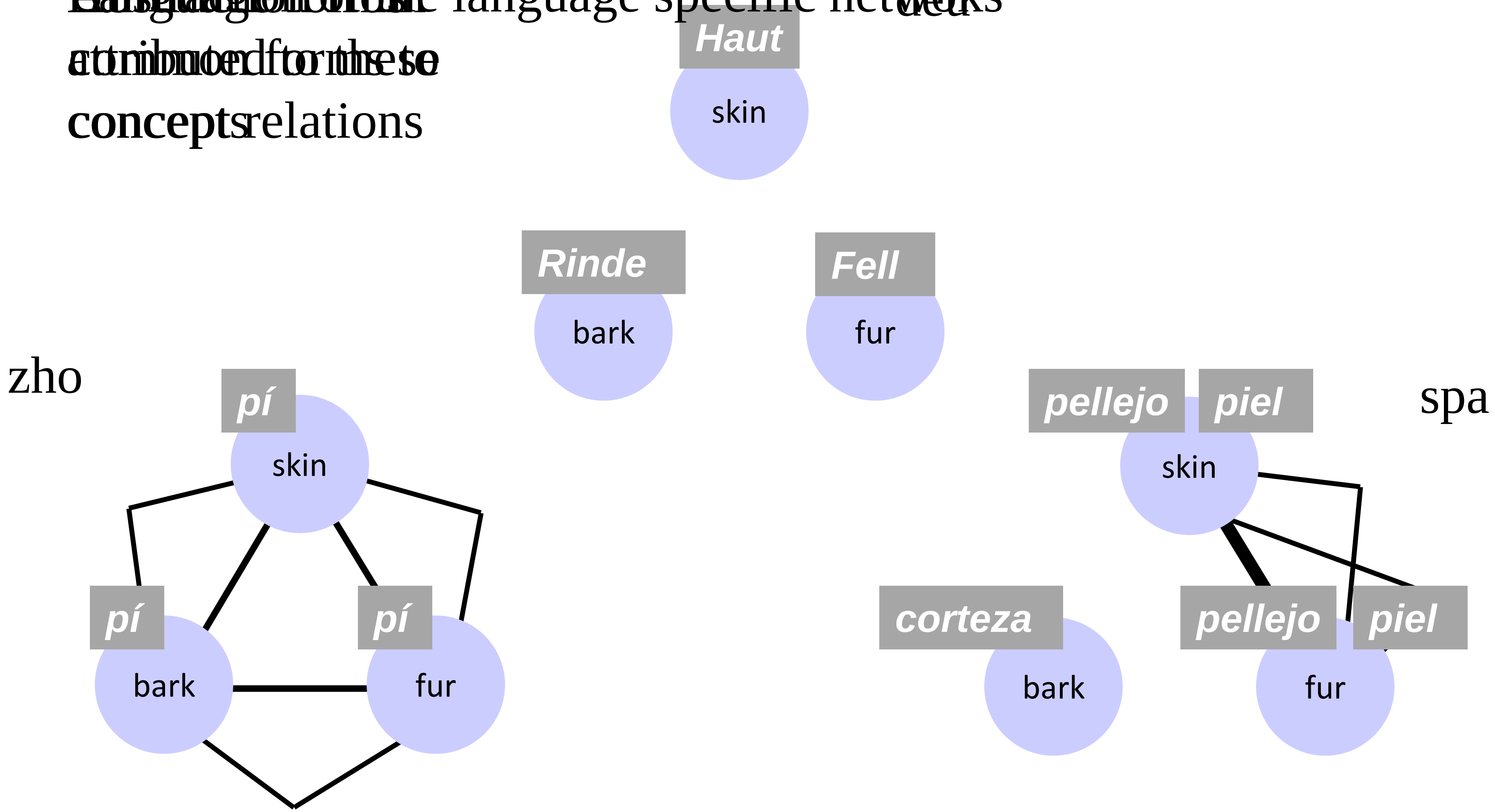


PollyNett: Construction Principle

Unification of the language specific networks

Language forms attributed to the concepts

Abstraction from combined to these relations

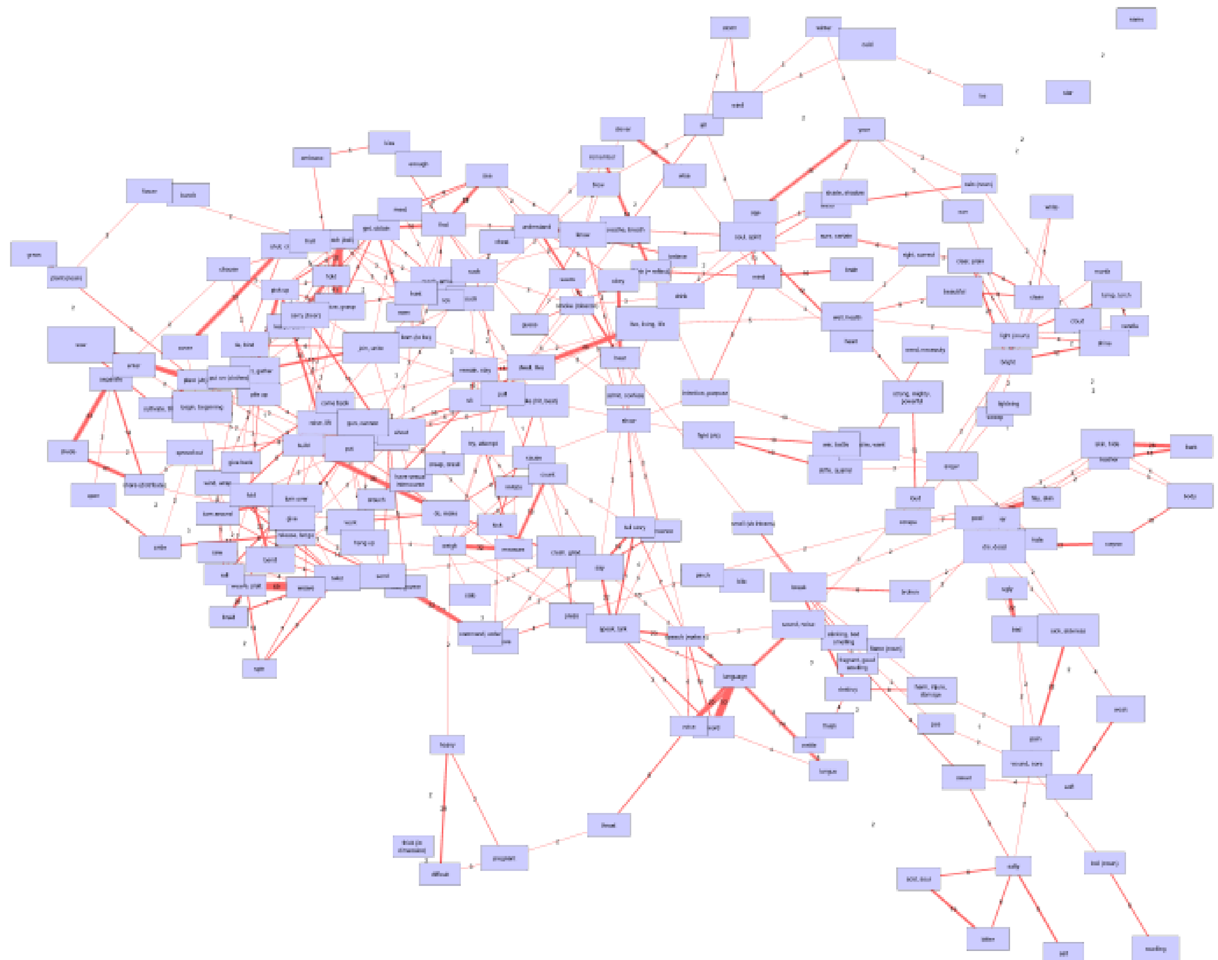


Outline of the Talk

1. Conceptual Structures and Meaning Change
2. Cognitive Historical Semantics
3. PollyNett: Crosslinguistic Polysemy Network
4. The Semantic Potential of PollyNett
5. Concluding Remarks

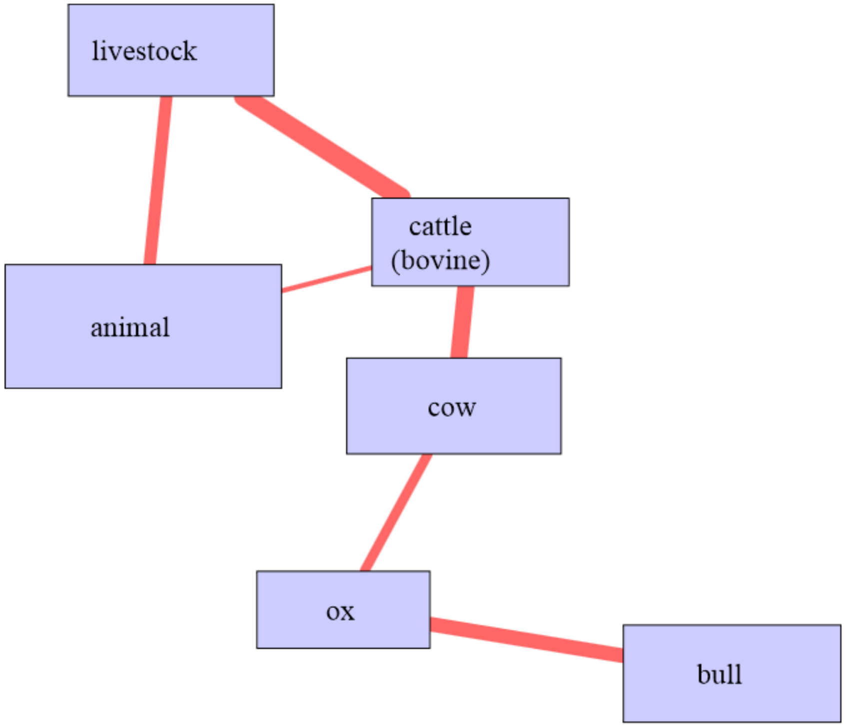
PollyNett network structure

- Pollynetts can be visualized and analyzed with the help of Cytoscape (Smoot et al. 2011), a software originally designed for network analysis in biology, especially genetics
- example: arbitrary subgraph (208 nodes, 460 edges out of 946 nodes, 2034 edges)



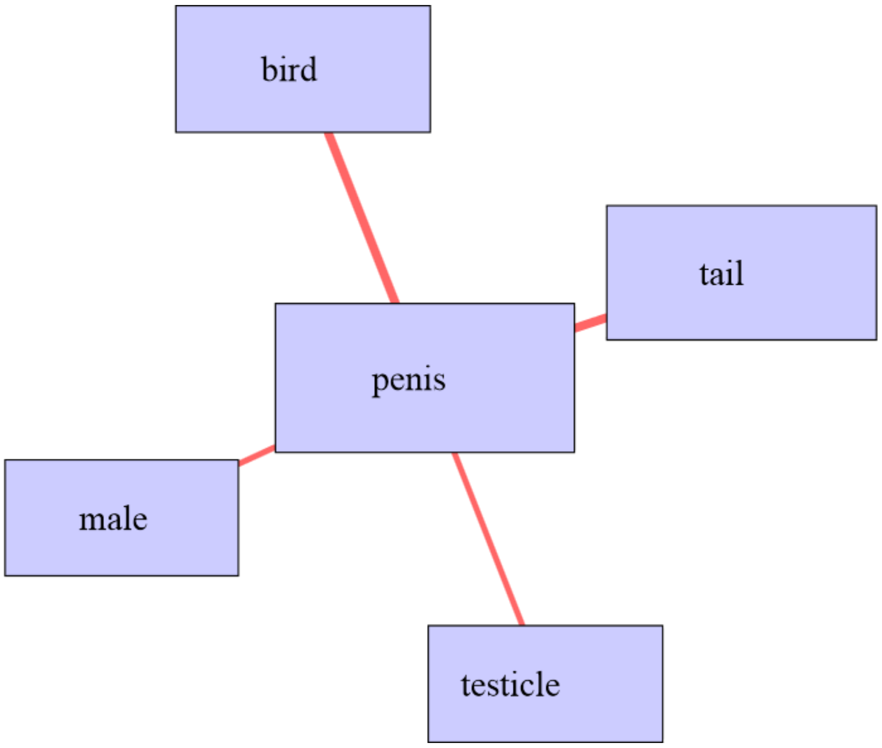
Conceptual Relations

- Steiner et al. (2011) indicate that “similar meanings have a larger probability to be expressed similarly in human language than different meanings”
- Even though the *similarity of sound chains* is the structural base of PollyNetts, the meanings which are linked are not only similar
- Taxonomic relations



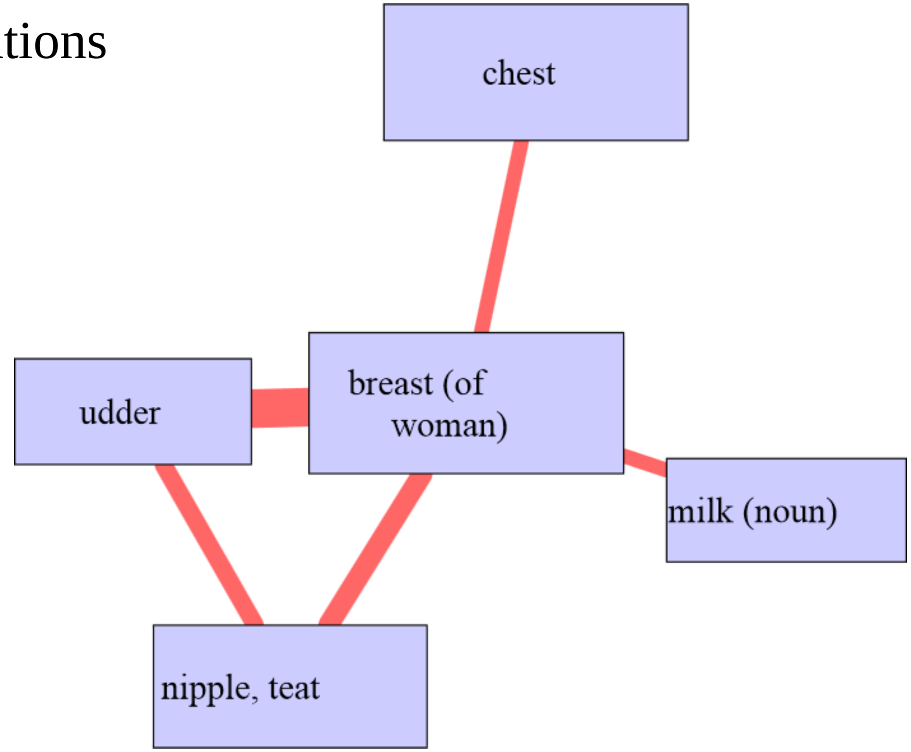
Conceptual Relations

- Steiner et al. (2011) indicate that “similar meanings have a larger probability to be expressed similarly in human language than different meanings”
- Even though the *similarity of sound chains* is the structural base of PollyNetts, the meanings which are linked are not only similar
- Similarity-based relations



Conceptual Relations

- Steiner et al. (2011) indicate that "similar meanings have a larger probability to be expressed similarly in human language than different meanings"
- Even though the *similarity of sound chains* is the structural base of PollyNetts, the meanings which are linked are not only similar
- Contiguity-based relations



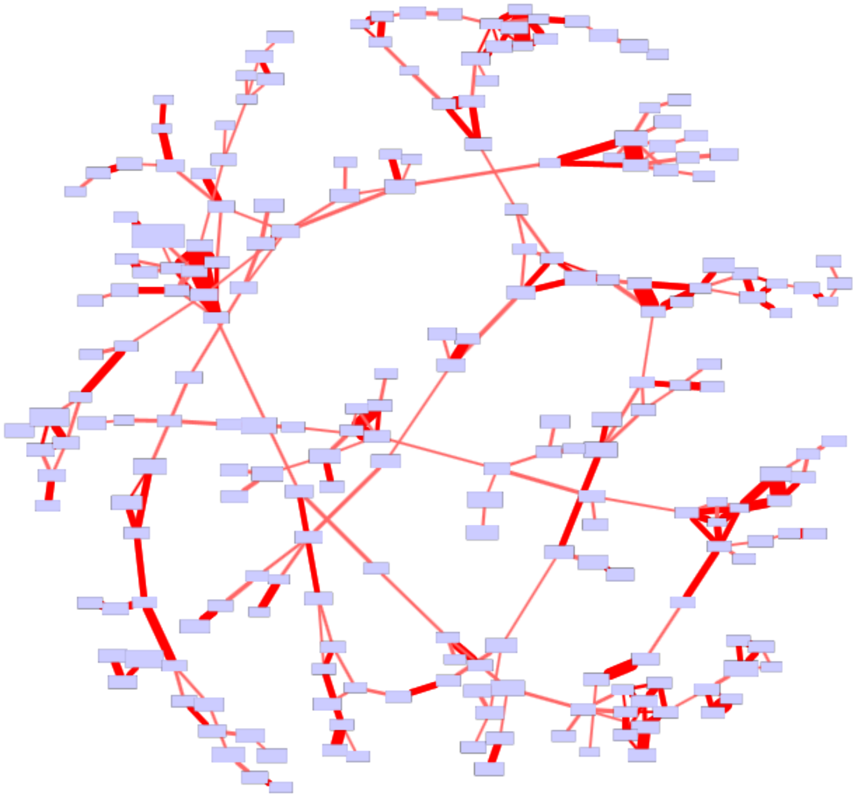
Network Analysis

Cluster analysis

- Statistical accounts on cross-linguistic polysemy retrieved from semantically aligned word lists make it possible to define the similarity between concepts on an item-to-item basis
- Cluster analyses, however, make it possible to assign several items to specific groups of items (communities) that share a high similarity among themselves while being less similar to items outside the group
- For our initial tests we restrict ourself to simple *Connected Components* (CC) *cluster* analyses*:*
  - Nodes that are directly or indirectly connected are assigned to the same group
  - Unconnected nodes are assigned to different groups
  - Varying the thresholds that define which items are assumed to be connected or not allow the representation of clusters in different levels of abstraction

Network Analysis

- Subnetwork with no edge cut-off …   … and a cuf-off beneath 5

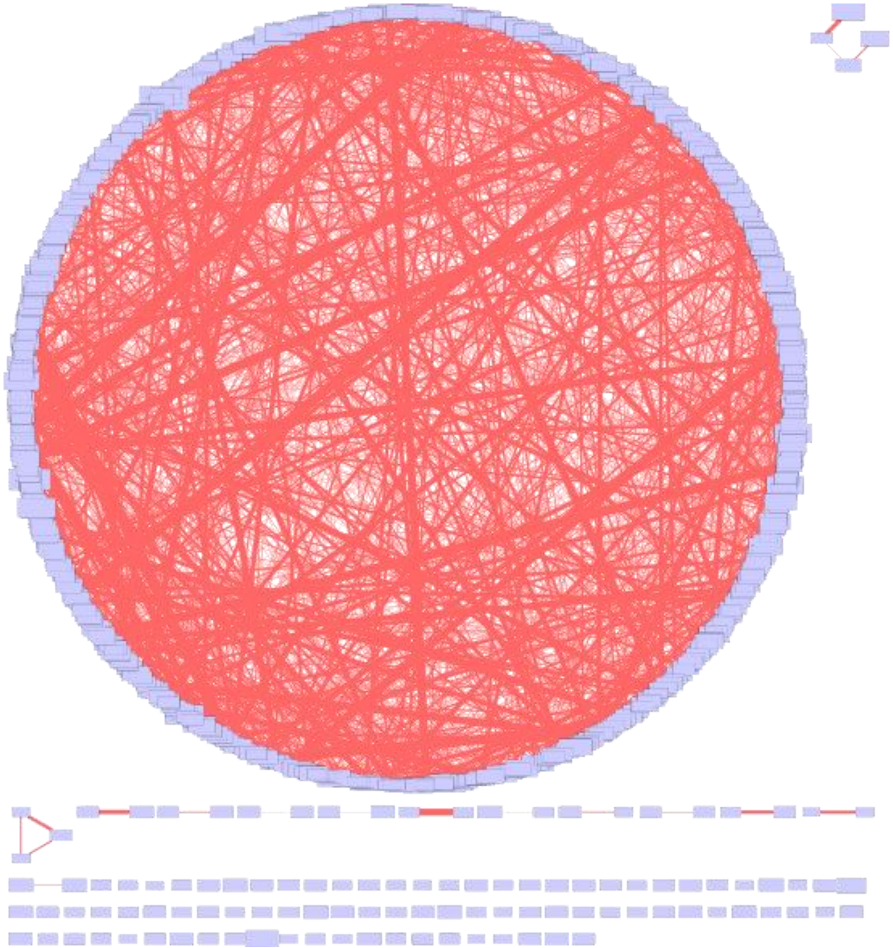


Abstraction Levels

- PollyNett depends on the level of abstraction which is applied to the links between the concepts.
- In the *language version* (PollyNett^{lang}), two concepts are connected if there are two or more **languages** which verbalize these concepts by the same word
- However, it is not visible to which degree the closeness of the conceptual connection is universal: all languages displaying the same polysemy might be part of the same language family. In consequence, the conceptual connection would be restricted to a certain cultural background
- In the *language family version* (PollyNett^{fam}), two concepts are connected if there are two or more **language families** which verbalize these concepts by the same word. Thus, strongly connected concepts imply that their relation has a certain degree of universality

Abstraction Levels

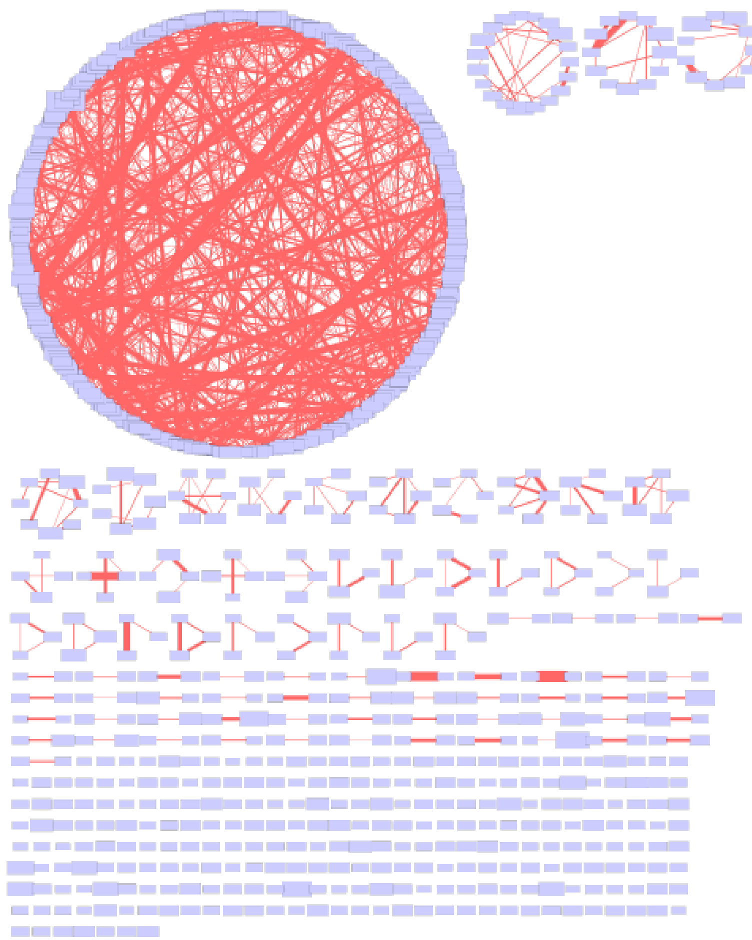
Comparison of different levels of abstraction applied to **PollyNett as whole**



PollyNett^lang:
no cut-off

Abstraction Levels

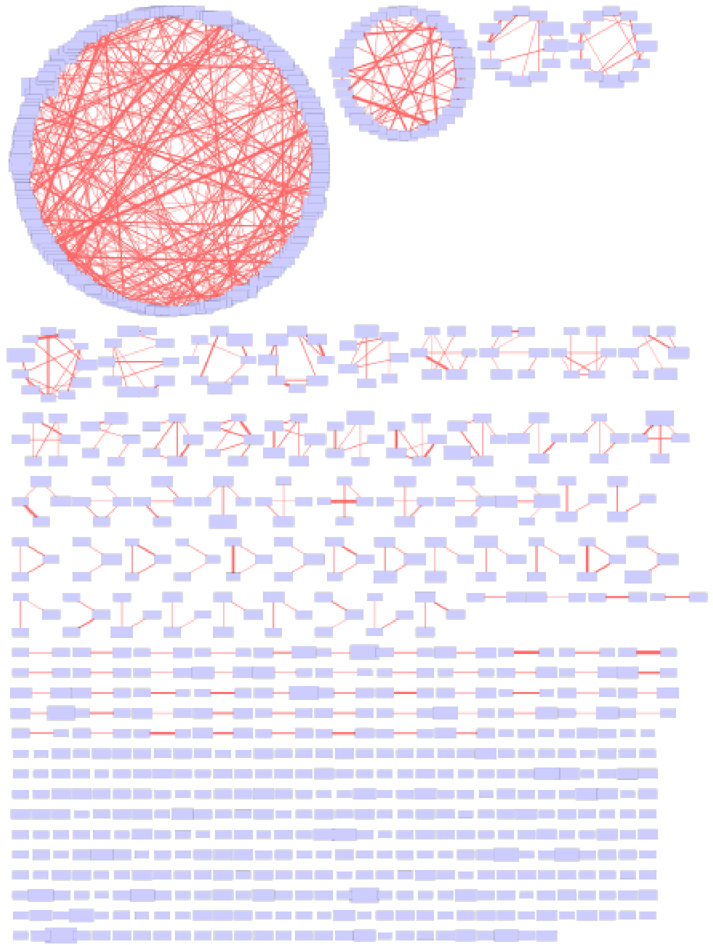
Comparison of different levels of abstraction applied to **PollyNett as whole**



PollyNett[lang]:
cut-off beneath 4

Abstraction Levels

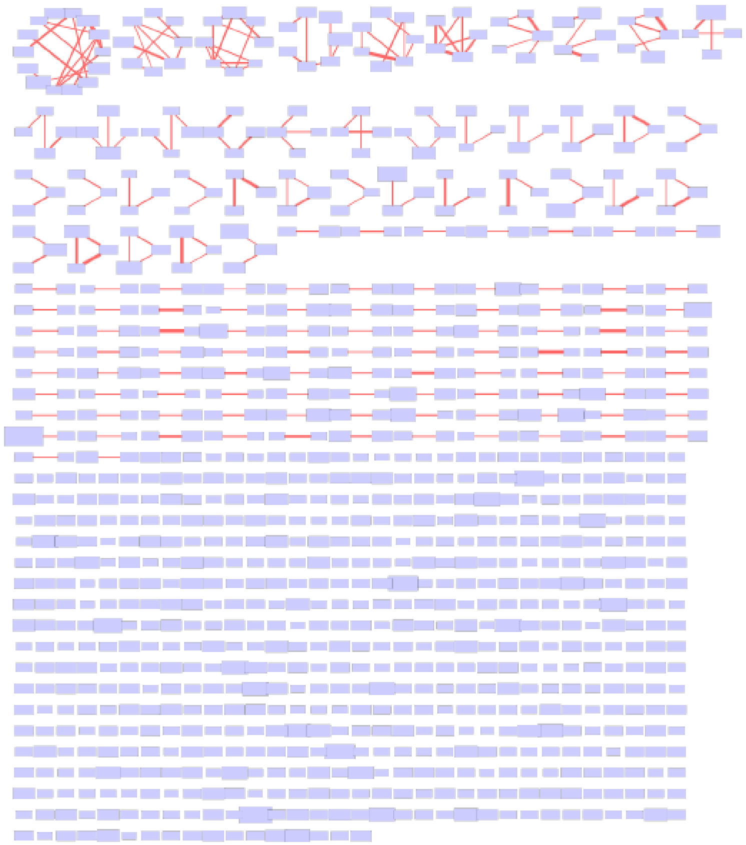
Comparison of different levels of abstraction applied to **PollyNett as whole**



PollyNett^{fam}:
cut-off beneath 4

Abstraction Levels

Comparison of different levels of abstraction applied to **PollyNett as whole**

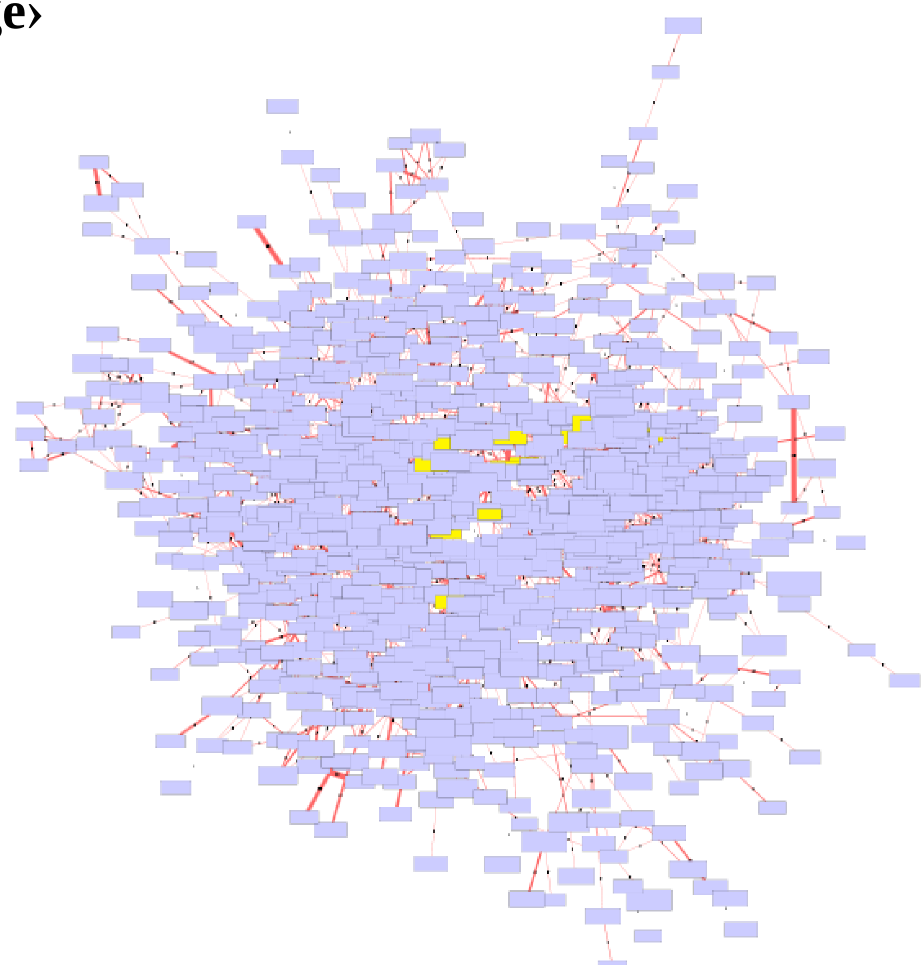


PollyNett^{fam}:

cut-off beneath 9

Abstraction Levels

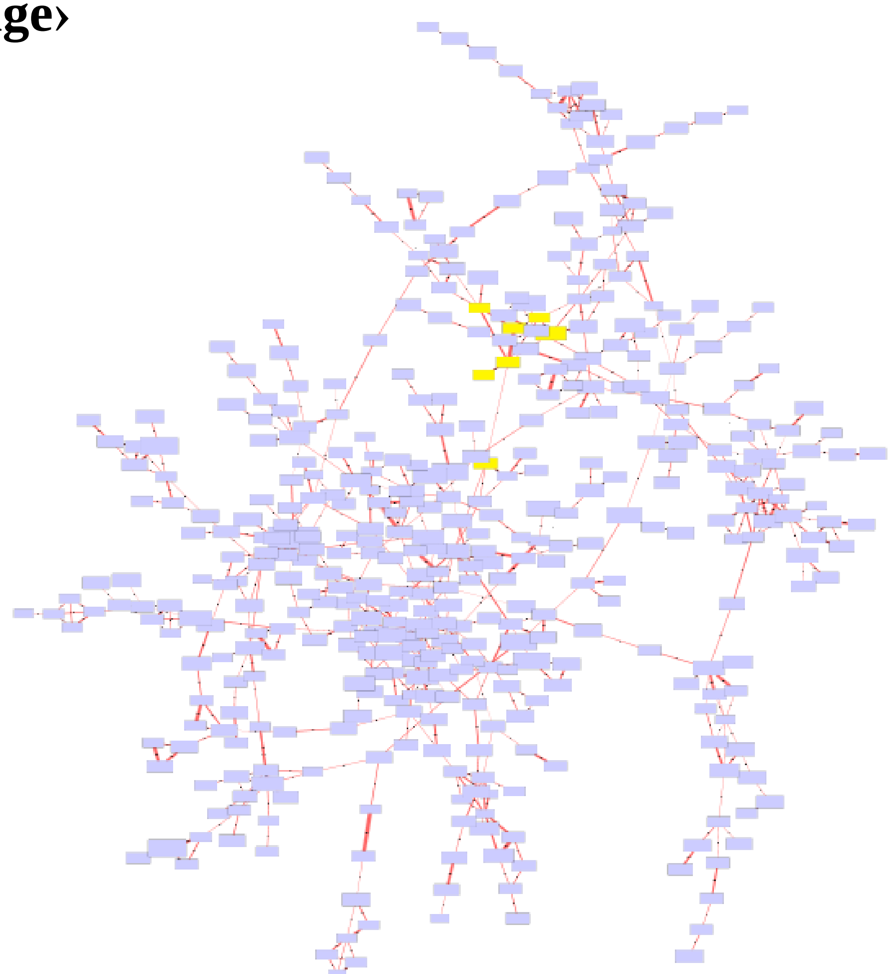
Comparison of different levels of abstraction applied to the **cluster around the concept ‹language›**



PollyNett^{lang}(‹language›):
no cut-off

Abstraction Levels

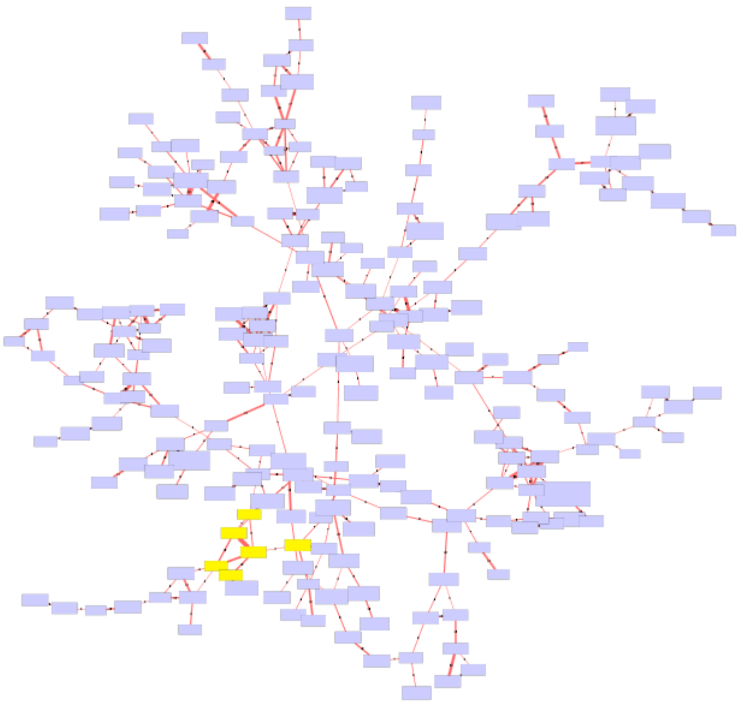
Comparison of different levels of abstraction applied to the **cluster around the concept ‹language›**



PollyNett[lang](‹language›):
cut-off beneath 4

Abstraction Levels

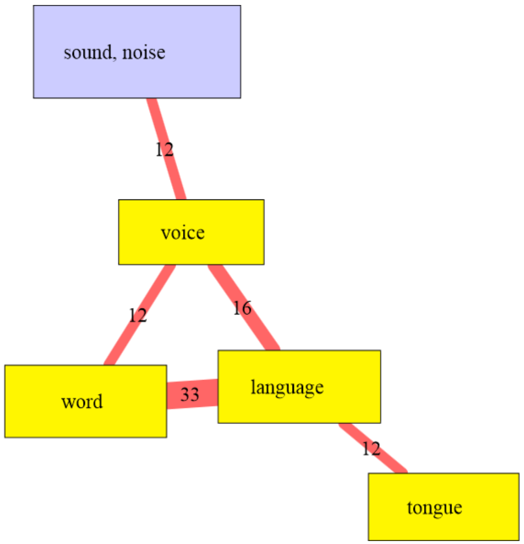
Comparison of different levels of abstraction applied to the **cluster around the concept ‹language›**

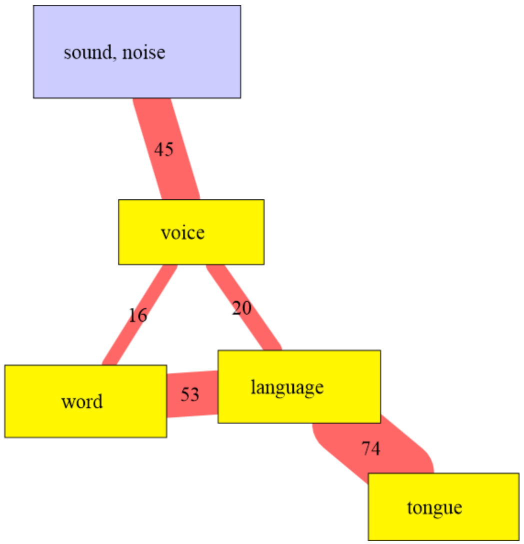


PollyNett^{fam}(‹language›):
cut-off beneath 4

Abstraction Levels

Comparison of different levels of abstraction applied to the **cluster around the concept ‹language›**





PollyNett^{fam}(‹language›):
cut-off beneath 9

for comparison:
same cluster with *language* links instead of *language family* links

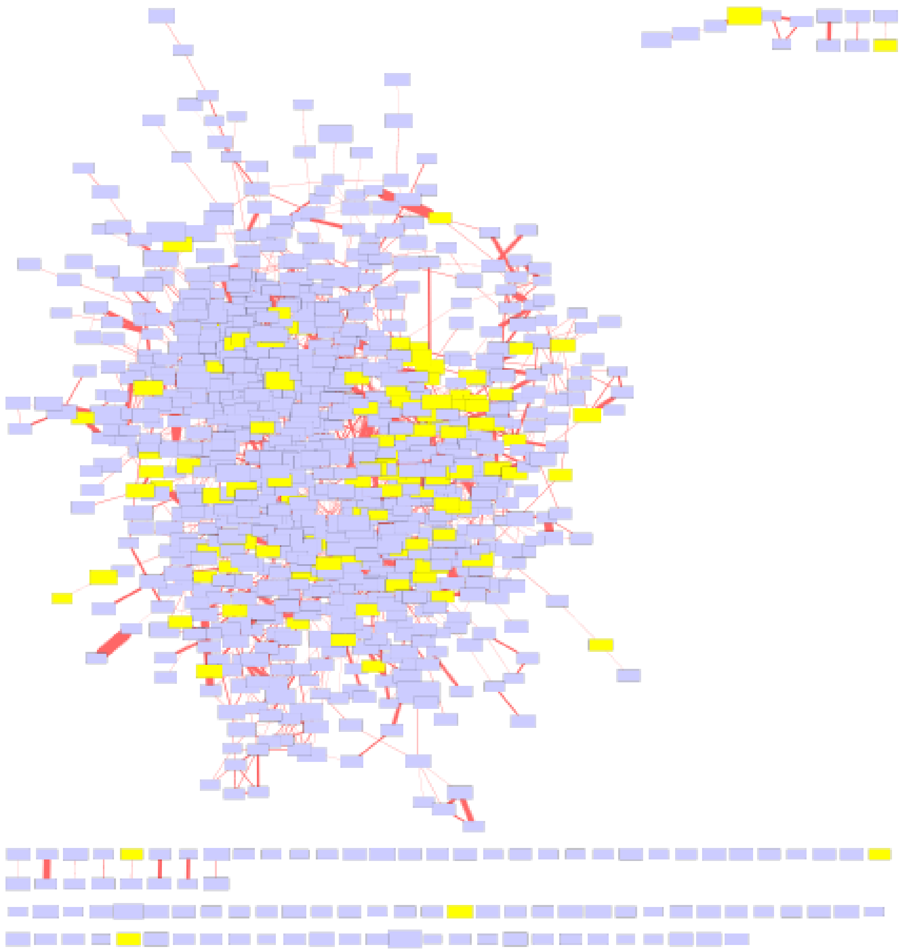
Swadesh subnetwork

Swadesh Lists and Basic Vocabulary Items

- Swadesh lists (named after Swadesh's publications from 1950, 1952, and 1955) are collections of semantic items traditionally glossed by English words that are supposed to reflect the *basic vocabulary* of all languages
- In theory, basic vocabulary refers to those meanings that are so basic (general, important, fundamental) that they are reflected by simple expressions in all languages of the world, independent of time and space (Sankoff 1969: 2f).
- Due to the basic character of these meanings, the words that express basic meanings are further expected to be rather prone to processes of lexical replacement due to semantic shift or borrowing.

Swadesh subnetwork

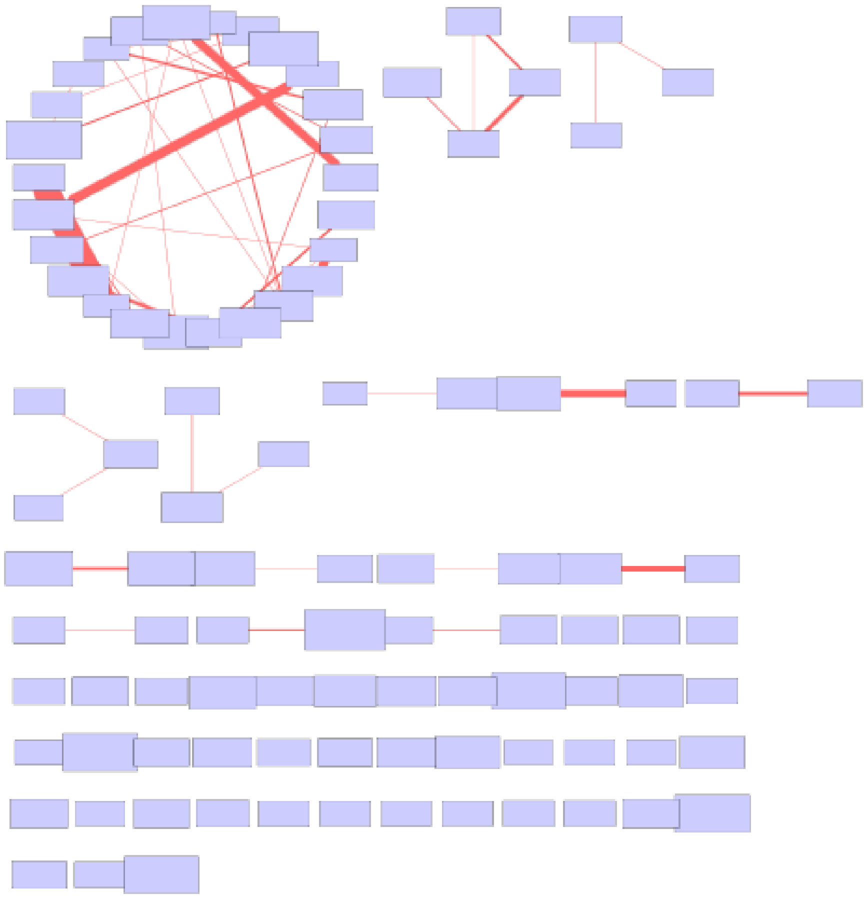
PollyNett contains a number of Swadesh items (101 concepts)



PollyNet(Swadesh marked):
no cut-off

Swadesh subnetwork

PollyNett contains a number of Swadesh items (101 concepts)



PollyNett^{swadesh}:

no cut-off

Swadesh subnetwork

Given the assumptions that are made about the Swadesh items, how should they be reflected in PollyNett or PollyNett^{swadesh}?

- **Universality**: „Many languages have a word for each item“
- **Possible reflex**: number of languages per concept is higher in PollyNet^{swadesh} than in PollyNet

Average number of languages that verbalize an item:

PollyNett:	89%, i.e. 174 out of 195
PollyNett^{swadesh}:	93%, i.e. 182 out of 195

Swadesh subnetwork

Given the assumptions that are made about the Swadesh items, how should they be reflected in PollyNett or PollyNett^{swadesh}?

- **Stability**: „The concepts are crucial for the functioning of the system and so basic, that they are not interconnected nor prone to lexical replacement“
- **Possible reflex 1**: average degree is lower than average degree of overall network

Average node degree (number of links per node):

PollyNett:	^{0}4.3	^{4}1.7
Swadesh-subnet of PollyNett:	^{0}4.9	^{4}2.2
PollyNett^{swadesh}:	^{0}0.9	^{4}0.4

Swadesh subnetwork

Given the assumptions that are made about the Swadesh items, how should they be reflected in PollyNett or PollyNett^{swadesh}?

- **Stability**: „The concepts are crucial for the functioning of the system and so basic, that they are not interconnected nor prone to lexical replacement“
- **Possible reflex 2**: average number of forms per concept is lower than in overall network

Average number of forms per concept:	
PollyNett^{0}:	1.28
Swadesh-subnet of PollyNett^{0}:	1.26
PollyNett^{swadesh_0}:	1.26

Swadesh subnetwork

Given the assumptions that are made about the Swadesh items, how should they be reflected in PollyNett or PollyNett^{swadesh}?

- **Stability**: „The concepts are crucial for the functioning of the system and so basic, that they are not interconnected nor prone to lexical replacement“
- **Possible reflex 3**: Swadesh-density is lower than density of overall network

Density (number of edges per number of possible edges):

PollyNett:	^{0}0.005	^{4}0.002
Swadesh-subnet of PollyNett:	^{0}0.049	^{4}0.022
PollyNett^{swadesh}:	^{0}0.009	^{4}0.004

Outline of the Talk

1. Conceptual Structures and Meaning Change
2. Cognitive Historical Semantics
3. PollyNett: Crosslinguistic Polysemy Network
4. The Semantic Potential of PollyNett
5. Concluding Remarks

Concluding Remarks

Open questions

- Is it possible to infer concrete patterns of semantic change, despite the fact that PollyNetts are indifferent regarding the processes that initially lead to polysemy?
- Do distance metrics derived from PollyNetts reflect the conceptual distances realistically?

Future Challenges

- We plan to enrich the data by including more meta-information (taxonomic relations from wordnet, ranked Swadesh-lists, etc.).
- We would like to find out whether it is possible to infer common (directional) change patterns from the undirected structure of PollyNetts.

Thanks to …

- the DFG for funding this research within the CRC991 (project C04)
- You, for listening

Bye, bye, Polly….