EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Philosophische Fakultät
FB Neuphilologie
Seminar für Sprachwissenschaft

EVOLAEMP
LANGUAGE EVOLUTION:
THE EMPIRICAL TURN

Evaluating the Potential of a Large-Scale Polysemy Network as a Model of Plausible Semantic Shifts

Alla Münch

Johannes Dellert

This research has been supported by the ERC Advanced Grant 324246 EVOLAEMP.

Polysemy Networks

A **polysemy network** is

- a graph over concepts (expressed by glosses)
- where each link represents the fact that at least one language has a lexical item which can denote both concepts
  (or rather: can be translated by both glosses)
- if we count the number of instances in different language families, link strength measures cross-linguistic **colexification** between concepts

Polysemy Networks

Polysemy networks are a potential source of evidence for plausibility of semantic shifts:

- intuition: every instance of semantic shift needs to pass an intermediate stage where the word in question is polysemous
- a massively cross-linguistic polysemy network provides us with a snapshot of semantic evolution in action

$\Rightarrow$ Polysemy networks might be a good model for modeling semantic change in computational historical linguistics!

Our study

- **Question**: How well does a polysemy network model attested semantic shifts?
- **Data**:
  - a massively cross-linguistic polysemy network (TUE) developed in Tübingen within EVOLAEMP (Dellert, 2014)
  - the Catalogue of Semantic Shifts (ZAL) maintained by the Russian academy of sciences (Zalizniak, 2008; Zalizniak et al., 2012)
- **Method**: for each attested change, determine the length of the shortest paths between the relevant concepts in the network

Table of Contents

Tübingen Polysemy Network (TUE)

- is derived from a large German-based dictionary database containing more than 750,000 entries in 114 languages
- has near-complete coverage of a 1,000-item list of basic concepts in 88 languages
- is built on more than 10,000 translations each for 31 languages from 10 primary language families of Eurasia
- is available to other researchers upon request

Tübingen Polysemy Network: Some Data

- 32,653 concepts and 47,647 links
- a central connected component of 13,073 nodes (!)
- 1,640 smaller components and 14,391 unconnected islands
- an average of 2.4 neighbors for each node
- a typical shortest path length of 8

The Catalogue of Semantic Shifts

- crosslinguistically recurring shifts collected manually by experts
- no restrictions or preferences in the range of meanings or languages (de facto: focus on Eurasian)
- contained 3,650 semantic shifts at the time of publication
- classified into six types, each supported by up to 40 realizations
- publicly available via a web interface:
  `http://semshifts.iling-ran.ru/`

Table of Contents

Procedure

- English version of the catalogue was extracted from the website
- considerable semi-automated cleanup work
- the total number of shift pairs in the result: 6,174

Procedure

Preparatory steps:

- metalanguages of the catalogue: English and Russian
- primary language of the Tübingen Polysemy Network: German
- electronic English-German dictionary needed to be useed as an intermediary for comparing the links in both resources

Experiment:

- evaluate pairs of English glosses in the catalogue against the network by determining the **shortest paths in the network between any pair of German translation equivalents**
- do this separately for each type, and compare recall

Table of Contents

Results: No Translation

No translation was found in the dictionary or the network:

- for highly specialized or culture-specific concepts
- for hypernyms
- for multi-word metalinguistic descriptions

ZAL: hawk[Accipiter], hawkweed[Hieracium]

Results: No Translation

No translation was found in the dictionary or the network:

- for highly specialized or culture-specific concepts
- for hypernyms
- for multi-word metalinguistic descriptions

ZAL: hawk[Accipiter], hawkweed[Hieracium]

ZAL: [foreigner], fantastic monster

Examples:
Ossetian *Rum* ‘Rome’ $\rightarrow$ *Rujmon* ‘a mythical monster’
Buryat *mangad* ‘Russian; many-headed monster’

Types of Semantic Shifts in the Catalogue:

- Polysemy
- Diachronic Semantic Evolution
- Syncretism
- Cognates
- Borrowings
- Morphological Derivation

Only Diachronic Semantic Evolution, Cognates, and Borrowings contain instances of semantic change in the stricter sense.

Results: Polysemy

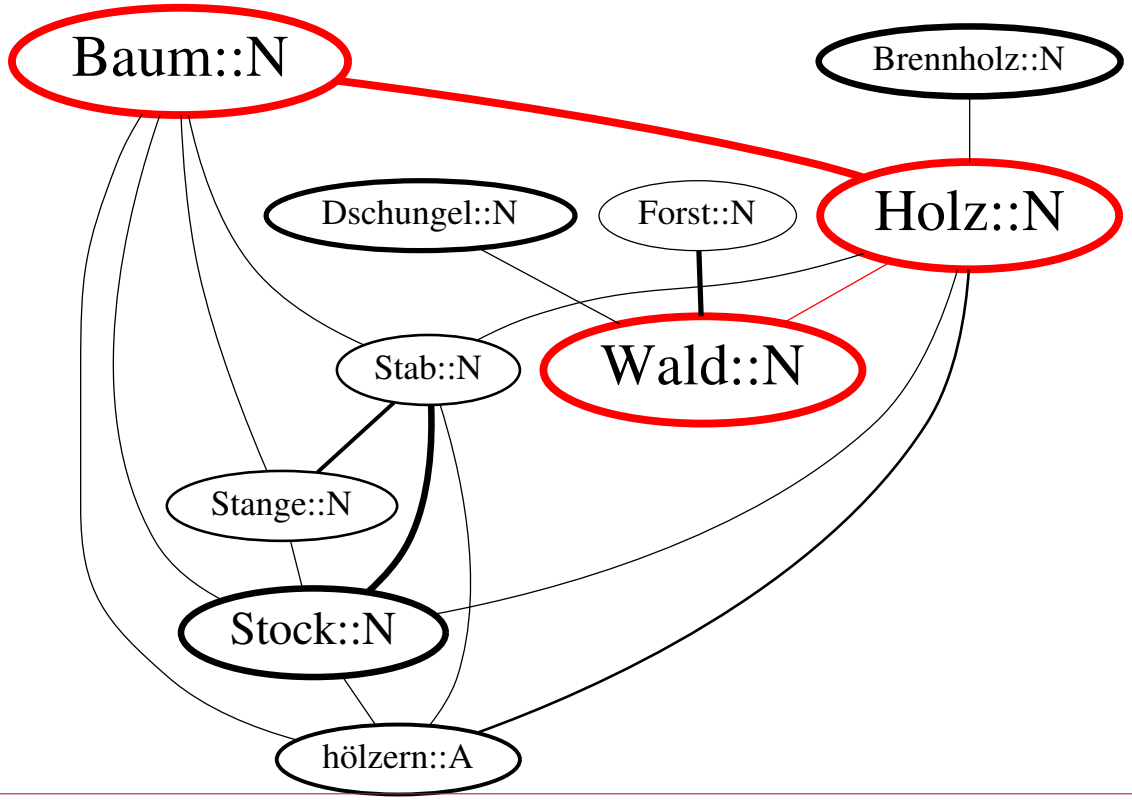
Synchronic Polysemy:
A and B are meanings of a polysemous word

ZAL: tree – forest
TUE: Baum “tree” – Holz “wood” – Wald “forest”

Example:
Tuvan *yjash* ‘tree; forest’

Results: Polysemy

ZAL: tree – forest
TUE: Baum “tree” – Holz “wood” – Wald “forest”



Results: Diachronic Semantic Evolution

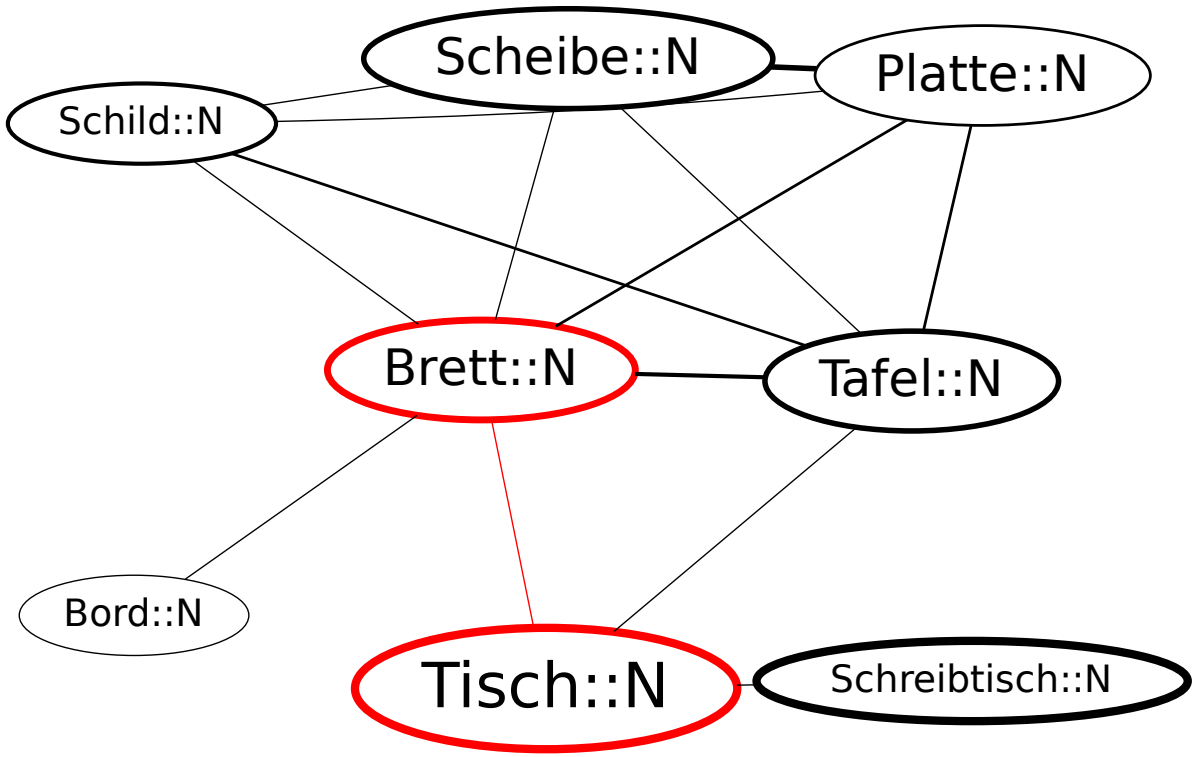
Diachronic semantic evolution of a word in one language or from an ancestor language to a descendant language

ZAL: board/plank - table/desk
TUE: Brett “board” - Tisch “table”

Example:
Latin *tabula* ‘plate, tablet’ $\rightarrow$ Italian *tavola* ‘table’

Results: Diachronic Semantic Evolution

ZAL: board/plank - table/desk
TUE: Brett “board” - Tisch “table”



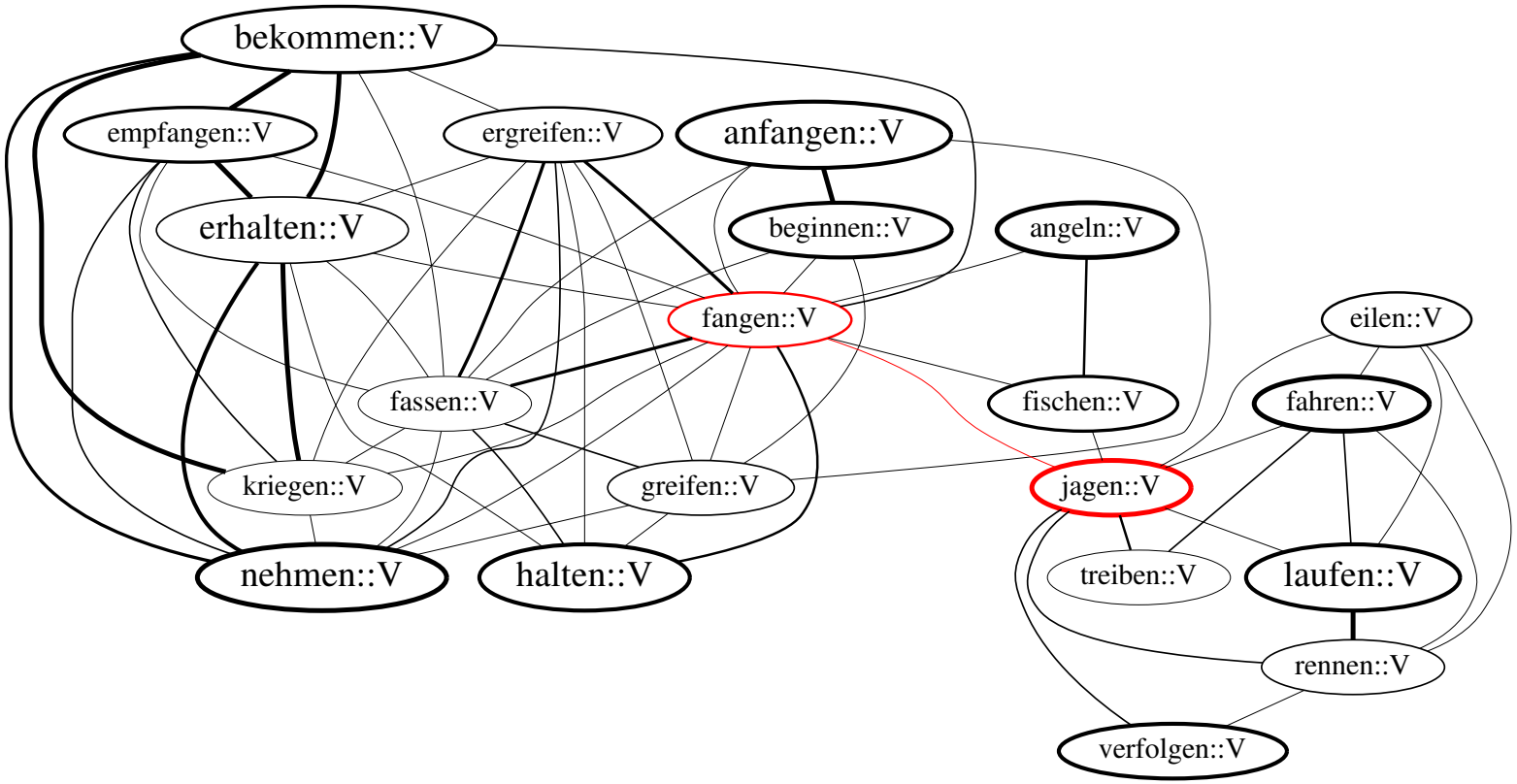
Results: Diachronic Semantic Evolution

ZAL: catch - hunt
TUE: jagen “hunt” - fangen “catch”

Example:
Latin *capto* ’various meanings related to catching’
$\rightarrow$ Italian *cacciare* ‘to hunt’

Results: Diachronic Semantic Evolution

ZAL: catch - hunt
TUE: jagen “hunt” - fangen “catch”



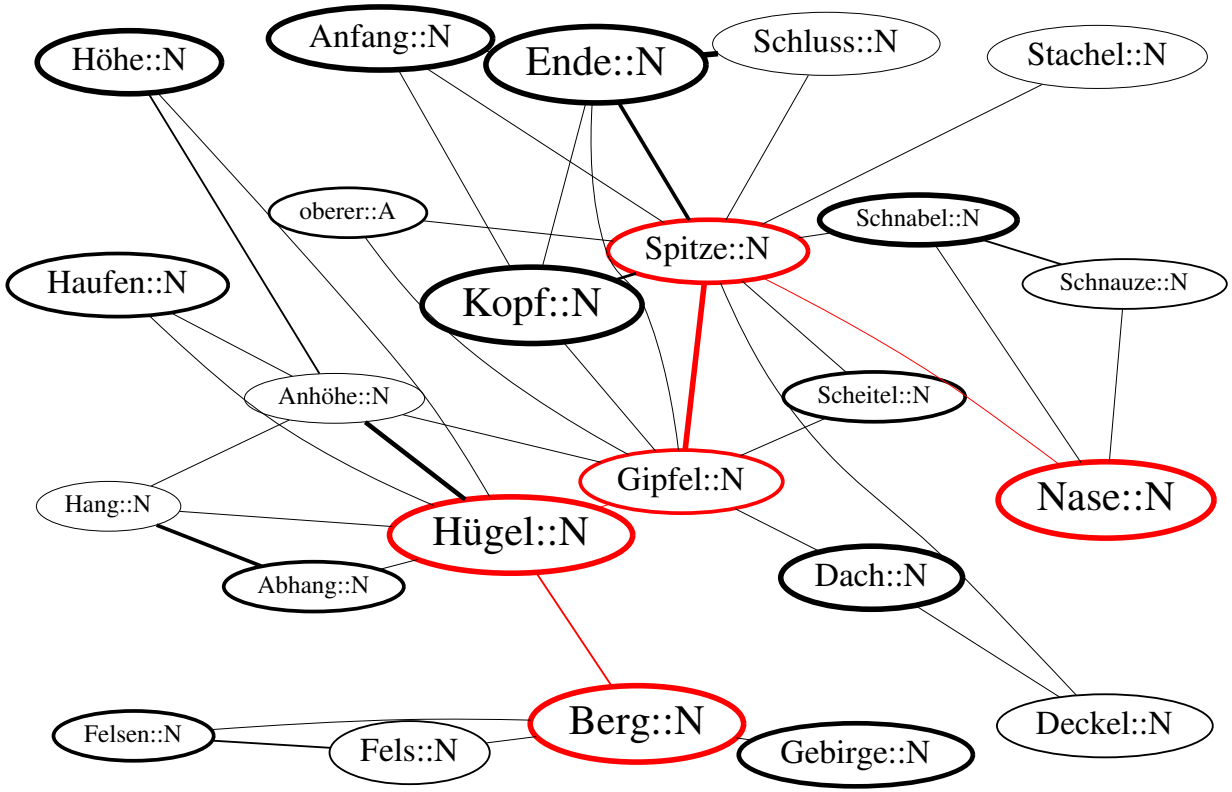
Results: Diachronic Semantic Evolution

ZAL: mountain - nose
TUE: Berg “mountain” – Hügel “hill” – Gipfel “summit” – Spitze “tip” – Nase “nose”

Example:
Hungarian *orr* ‘nose’ ← ‘top, mountain’

Results: Diachronic Semantic Evolution

ZAL: mountain - nose
TUE: Berg “mountain” – Hügel “hill” – Gipfel “summit” – Spitze “tip” – Nase “nose”



Results: Syncretism

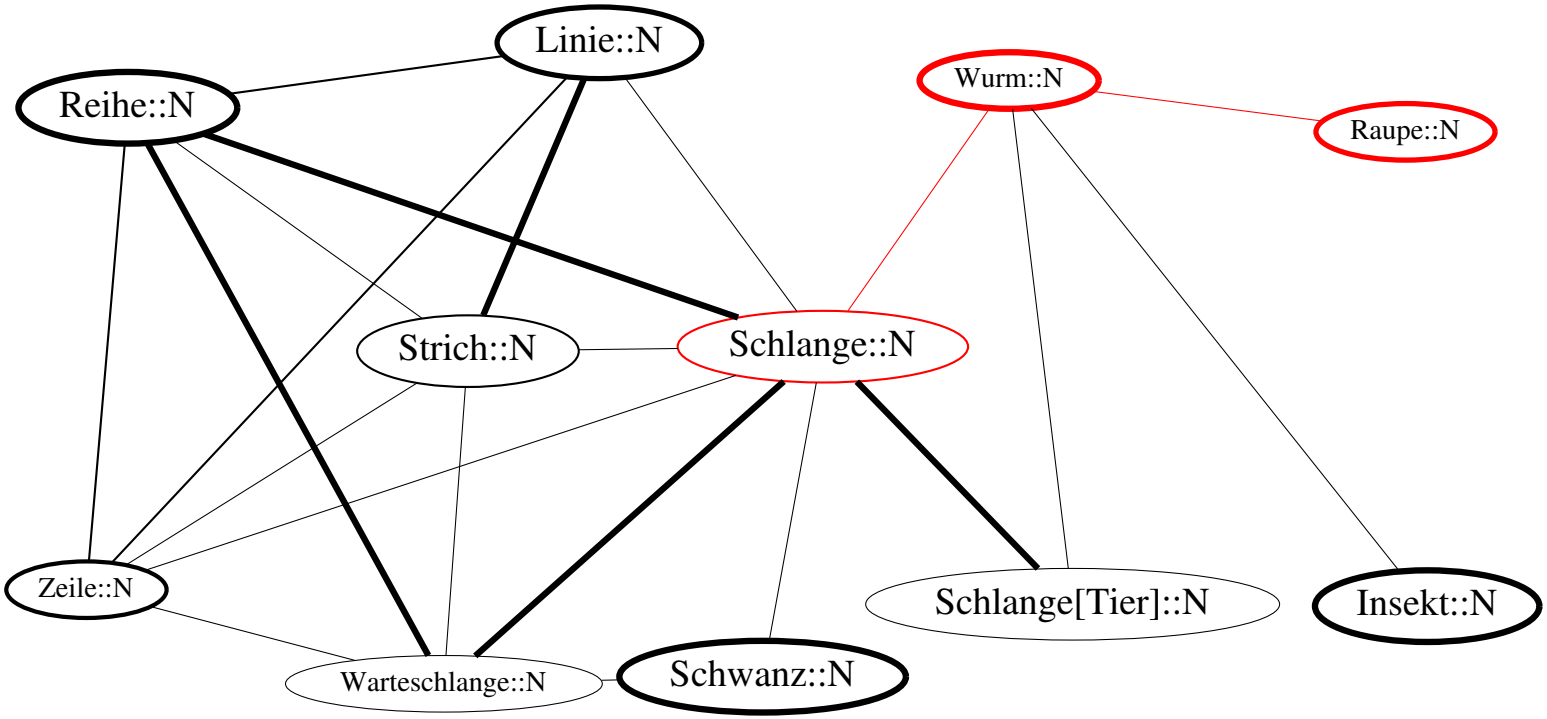
Syncretism: the concepts are connected by being subconcepts of a concept which is denoted by a single word in the language (polysemy-vs-vagueness problem)

ZAL: caterpillar – snake
TUE: Raupe “caterpillar” – Wurm “worm” – Schlange “snake”

Examples:
Melpa *kimbukla* ‘snake’ — ‘caterpillar’
Kowiai *aroi* ‘snake’ — ‘caterpillar’

Results: Syncretism

ZAL: caterpillar – snake
TUE: Raupe “caterpillar” – Wurm “worm” – Schlange “snake”



Results: Cognates

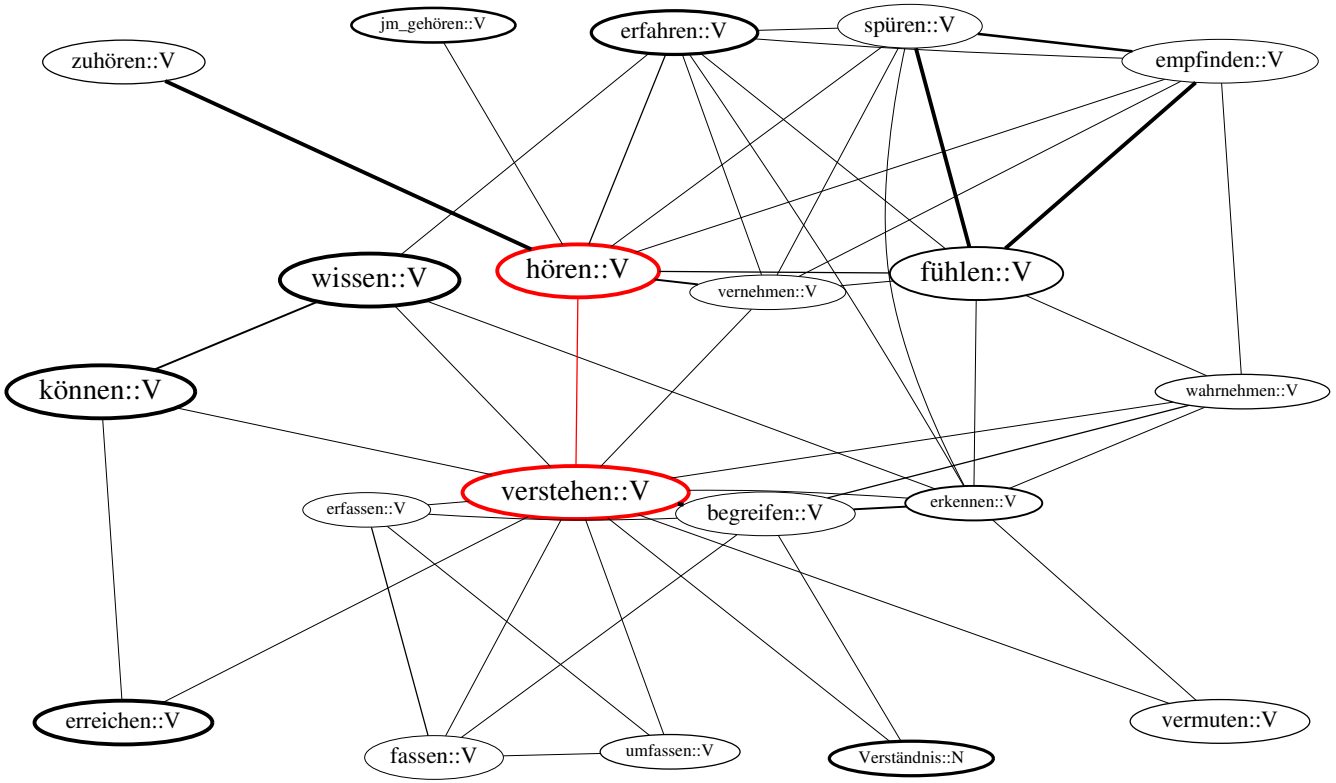
Cognates: meanings A and B belong to words of two sister languages which have developed from the same root in their common ancestor

ZAL: hear - understand
TUE: hören “hear” - verstehen “understand”

Example:
French *entendre* ‘to hear’ -
Spanish *entender* ‘to understand, to be knowledgeable’

Results: Cognates

ZAL: hear - understand
TUE: hören “hear” - verstehen “understand”



Results: Borrowings

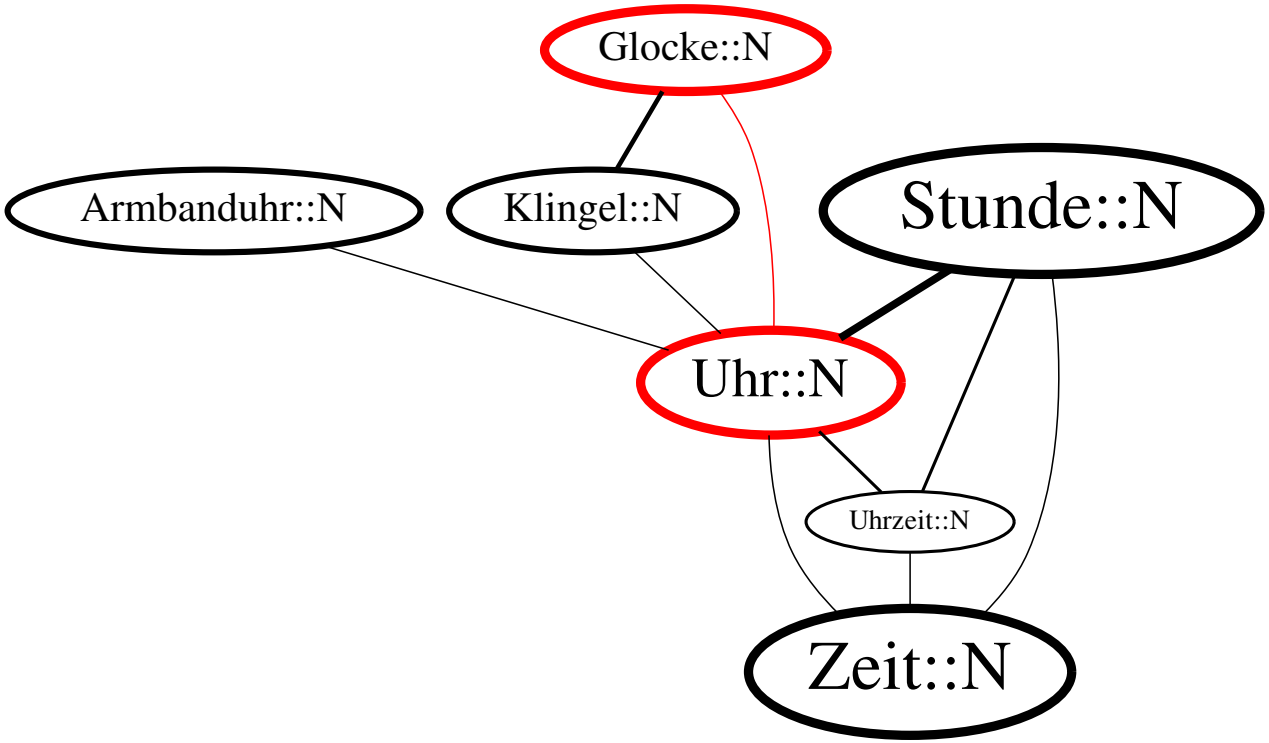
Borrowing: B is the meaning of a borrowed word, while A is the meaning of its source in the donor language

ZAL: bell - watch/clock
TUE: Glocke “bell” - Uhr “watch/clock”

Example:
Medieval Latin *clocca* ‘bell’ → English *clock* ‘clock, watch’

Results: Borrowings

ZAL: bell - watch/clock
TUE: Glocke “bell” - Uhr “watch/clock”



Results: Morphological Derivation

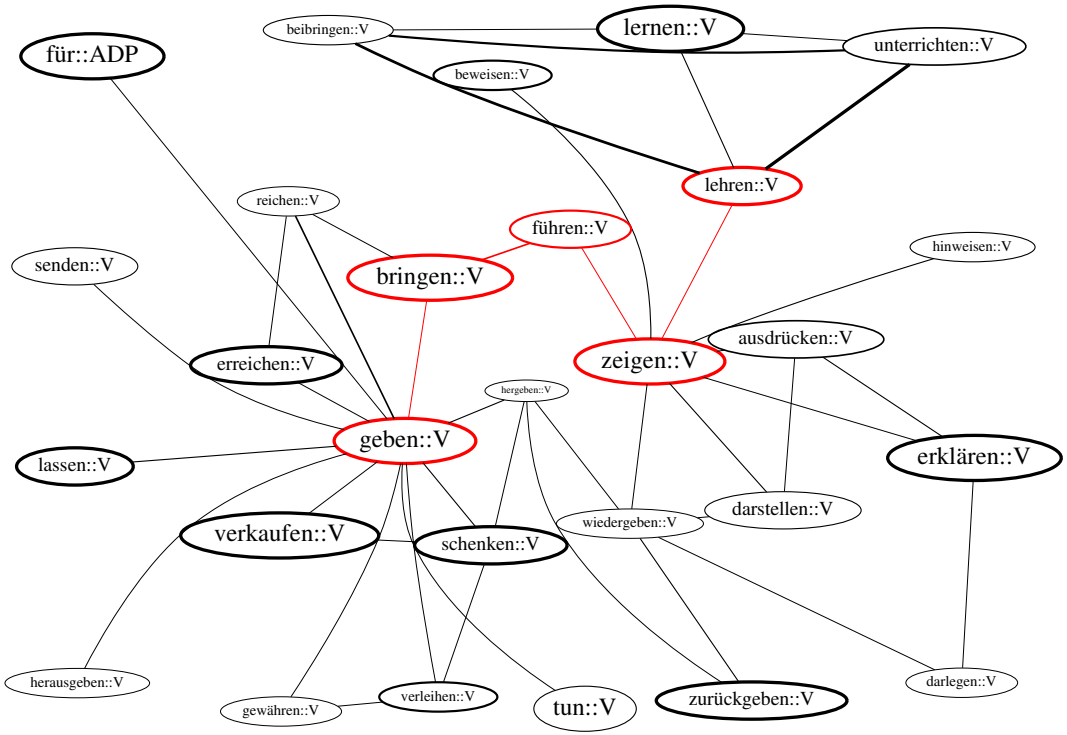
Morphological Derivation: meaning B is represented by a morphological derivative from the word which has meaning A

ZAL: give - teach
TUE: geben “give” – bringen “bring” – führen “guide” – zeigen “show” – lehren “teach”

Example:
Russian *davat’* ‘to give’ → *prepodavat’* ‘to teach’

Results: Morphological Derivation

ZAL: give - teach
TUE: geben “give” – bringen “bring” – führen “guide” – zeigen “show” – lehren “teach”



Results: Summary

Table: Shifts in the catalogue covered by the polysemy network.

Semantic shift type	Number of instances	No path	Path length 1 or 2	Path length 3 and more
Polysemy	2315	20.6 %	35.0 %	44.4 %
Semantic Evolution	107	26.2 %	33.6 %	40.2 %
Morphological Derivation	597	28.5 %	29.0 %	42.5 %
Syncretism	43	25.6 %	55.8 %	18.6 %
Borrowing	58	31.0 %	41.4 %	27.6 %
Cognates	393	23.7 %	37.9 %	38.4 %

Table of Contents

Conclusions

- high potential of polysemy networks for modelling the plausibility of semantic shifts
- a useful model of diachronic processes
  both within a language (semantic evolution) and
  across languages (cognates and borrowings)
- less useful as a model of shifts during morphological derivation
- overlap between resources shows there is a common core of frequent polysemies which can be used in computational models of semantic evolution

References

Johannes Dellert. Evaluating cross-linguistic polysemies as a model of semantic change for cognate finding. Proceedings of the Workshop on semantic technologies for research in the humanities and social sciences (STRiX). November 24-25, Gothenburg, Sweden, 2014.

Anna A. Zalizniak. A catalogue of semantic shifts: Towards a typology of semantic derivation. In Martine Vanhove, editor, *From polysemy to semantic change. Towards a typology of lexical semantic associations*, pages 217–232. John Benjamins, Amsterdam and Philadelphia, 2008.

Anna A. Zalizniak, Maria Bulakh, Dmitrij Ganenkov, Ilya Gruntov, Timur Maisak, and Maxim Russo. The catalogue of semantic shifts as a database for lexical semantic typology. *Linguistics*, 50(3):633–669, 2012.