German Semantic Verb Classes

Theoretical, Computational and Cognitive Approaches

Sabine Schulte im Walde

Computational Linguistics
Saarland University, Saarbrücken

Interview Osnabrück

September 18, 2004

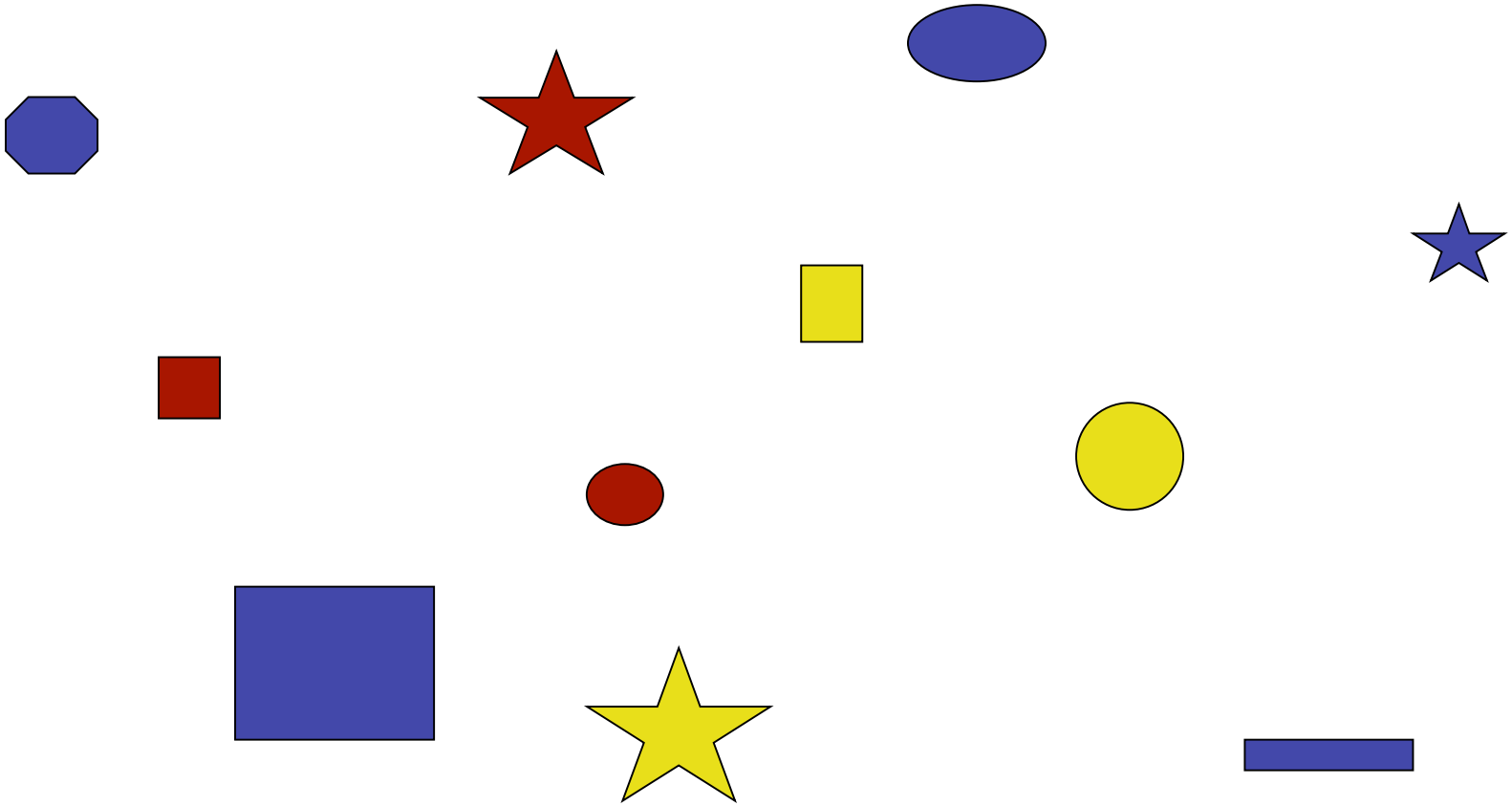
Overview

1. Classifications: theory and practise
2. German semantic verb classes
3. Clustering experiments and results
4. Comparison: theoretical and computational classifications
5. Cognitive approach: web-experiment on verb associations
6. Outlook

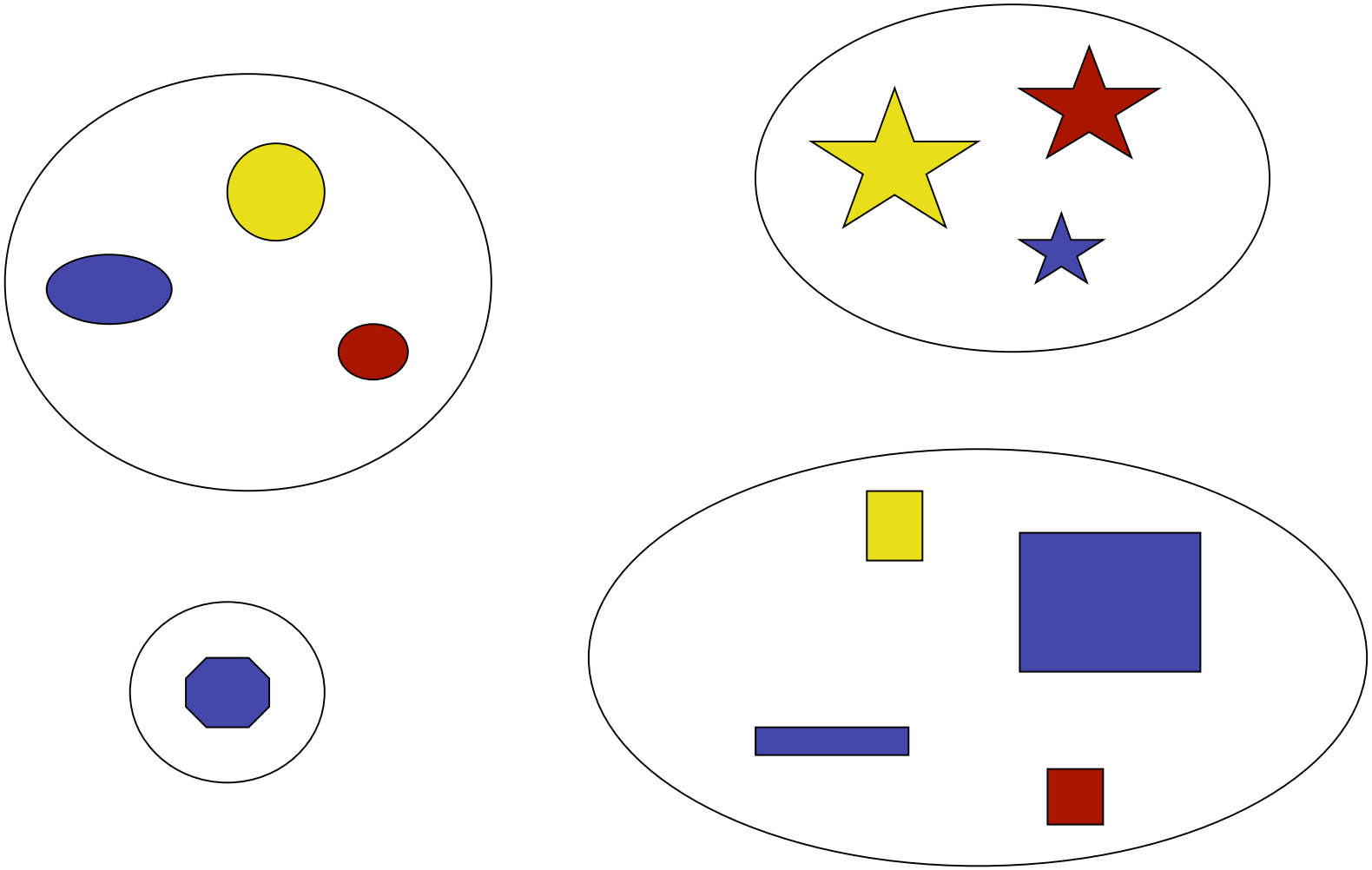
Classification: Definition

- Object → Class
- Objects in the same class: as similar as possible
- Objects in different classes: as dissimilar as possible
- Exploration of inherent data structure
- Generalisation over objects and their features

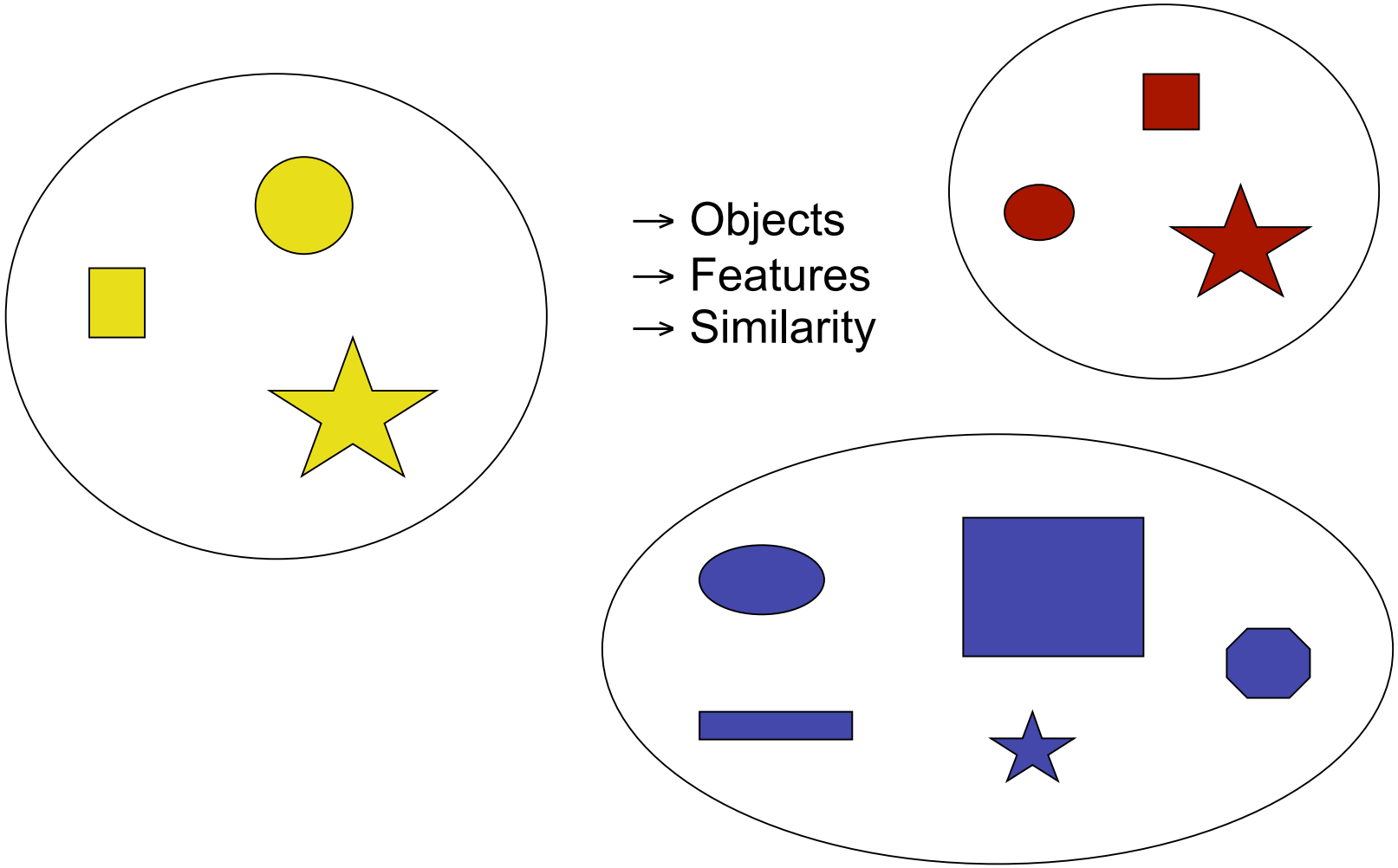
Classification: Illustration



Classification: Illustration



Classification: Illustration



Semantic Verb Classes in CL: Motivation

- Lexical verb resources in Natural Language Processing
- Semantic verb classes
  - → Lexical and conceptual semantic verb information
  - → Redundancy reduction in verb description
  - → Prediction and refinement of vague properties
- Standard procedure in multivariate data analysis:
  clustering methodology for automatic class induction

German Semantic Verb Classes: Examples

- *Aspect:* anfangen, aufhören, beenden, beginnen, enden
- *Manner of Motion*
  - *Locomotion:* gehen, klettern, kriechen, laufen, rennen, ...
  - *Rotation:* drehen, rotieren
  - *Rush:* eilen, hasten
  - *Vehicle:* fahren, fliegen, rudern, segeln
  - *Flotation:* fließen, gleiten, treiben
- *Insistence:* beharren, bestehen$_1$, insistieren, pochen
- *Support:* dienen, folgen, helfen, unterstützen
- *Existence:* bestehen$_2$, existieren, leben

Hypothesis: Verb Meaning and Behaviour

verb meaning components ↔ verb behaviour

[...] the behavior of a verb, particularly with respect to the expression and interpretation of its arguments, is to a large extent determined by its meaning.

Beth Levin (1993)

Clustering Methodology

1. Statistical acquisition of lexical verb information at the syntax-semantics interface
2. Association of verbs with distributional frame vectors
3. Automatic verb clustering by standard technique *k-Means*
4. Clustering evaluation against manual verb classification
5. Clustering interpretation

Experiment Feature Choice

D1 Purely syntactic definition of verb subcategorisation

D2 Syntactico-semantic definition of subcategorisation with prepositional preferences

D3 Syntactico-semantic definition of subcategorisation with prepositional and selectional preferences

Subcategorisation Frame Elements

n	noun phrase (case: nominative)
a	noun phrase (case: accusative)
d	noun phrase (case: dative)
r	reflexive pronoun
p	prepositional phrase
x	expletive *es*
i	non-finite clause
s-2	finite verb second clause
s-dass	finite *dass*-clause
s-ob	finite *ob*-clause
s-w	indirect *wh*-question
k	copula construction

Examples:

- na
- np
- npr
- nds-dass

Prepositional Phrase Types

- Akk: an, auf, bis, durch, für, gegen, in, ohne, um, unter, vgl, über
- Dat: ab, an, auf, aus, bei, in, mit, nach, seit, unter, von, vor, zu, zwischen, über
- Gen: wegen, während
- Nom: vgl

Examples: Akk.an, Dat.nach, Gen.wegen, Nom.vgl

GermaNet Top Level Nodes

• Lebewesen	`creature´
• Sache	`thing´
• Besitz	`property´
• Substanz	`substance´
• Nahrung	`food´
• Mittel	`means´
• Situation	`situation´
• Zustand	`state´
• Struktur	`structure´
• Physis	`physis´
• Zeit	`time´
• Ort	`space´
• Attribut	`attribute´
• Kognitives Objekt	`cognitive object´
• Kognitiver Prozess	`cognitive process´

k-Means Clustering

- k-Means algorithm (Forgy, 1965)
- Unsupervised hard clustering
- Iterative re-organisation of cluster membership

- Number of clusters k = 43
- Frame distribution:
  - frequencies / probabilities / binaries
  - original / smoothed / noisy
- Initial cluster assignment:
  - random clusters
  - agglomerative hierarchical clusters
- Similarity measure

Clustering Examples

ahnen	`to guess´
bedauern	`to regret´
befürchten	`to fear´
bezweifeln	`to doubt´
merken	`to notice´
vermuten	`to assume´
weißen	`to whiten´
wissen	`to know

basieren	`to be based on´
beruhen	`to be based on´
resultieren	`to result from´
stammen	`to stem from´

beziffern	`to amount to´
schätzen	`to estimate´
veranschlagen	`to estimate´

befragen	`to interrogate´
entlassen	`to release´
ermorden	`to assassinate´
erschießen	`to shoot´
festnehmen	`to arrest´
töten	`to kill´
verhaften	`to arrest´

Taking a Step Back …

What's relevant for a classification, again?

Comparison of German Semantic Verb Classes

Background:

- Motivation
- Goals

Organisation and Design:

- Classification structure and class relations
- Choice of verbs
- Distinction of verb senses
- Choice of verb features

(joint work with Katrin Erk)

Examples of German Semantic Verb Classes

- Semantic Fields (Schumacher, 1986)
- Verb Process Classification (Ballmer/Brennenstuhl, 1986)
- WordNet → GermaNet (Fellbaum, 1998; Kunze, 2000)
- FrameNet → Salsa (Baker et al., 1998; Erk et al., 2003)
- Induced verb classes (Schulte im Walde, 2003)

→ Case study on *Manner of Motion (MOM)* verbs

Verb Process Classification

- Goal: complete thesaurus of German verbs
- Choice: 8,000 common, non-prefixed German verbs
- Organisation: *process models*
- Model: initial situation, transition, end situation;
  preconditions, results, consequences
- Model phases → verb classes
- Class relations: causal, temporal, implicative, etc.
- Verb meaning ← process ← paraphrase

Verb Process Model: MOM Example

R	**ruhen**	`to rest´
B	sich fortbewegen wollen	`to want to move´
B	sich erheben	`to raise´
B	sich orientieren	`to get the orientation´
B	**aufbrechen**	`to start moving´
B	sich entfernen	`to move away´
A	**sich fortbewegen**	`to move forward´
A	sich im Kreis fortbewegen	`to move in circles´
A	als Passagier mitfahren	`to move as passenger´
A	sich verirren	`to get lost´
A	begleiten	`to accompany´
E	**eintreffen**	`to arrive´
U	**rasten**	`to rest´

Fortbewegungsmodell
`forward movement´

als Passagier mitfahren

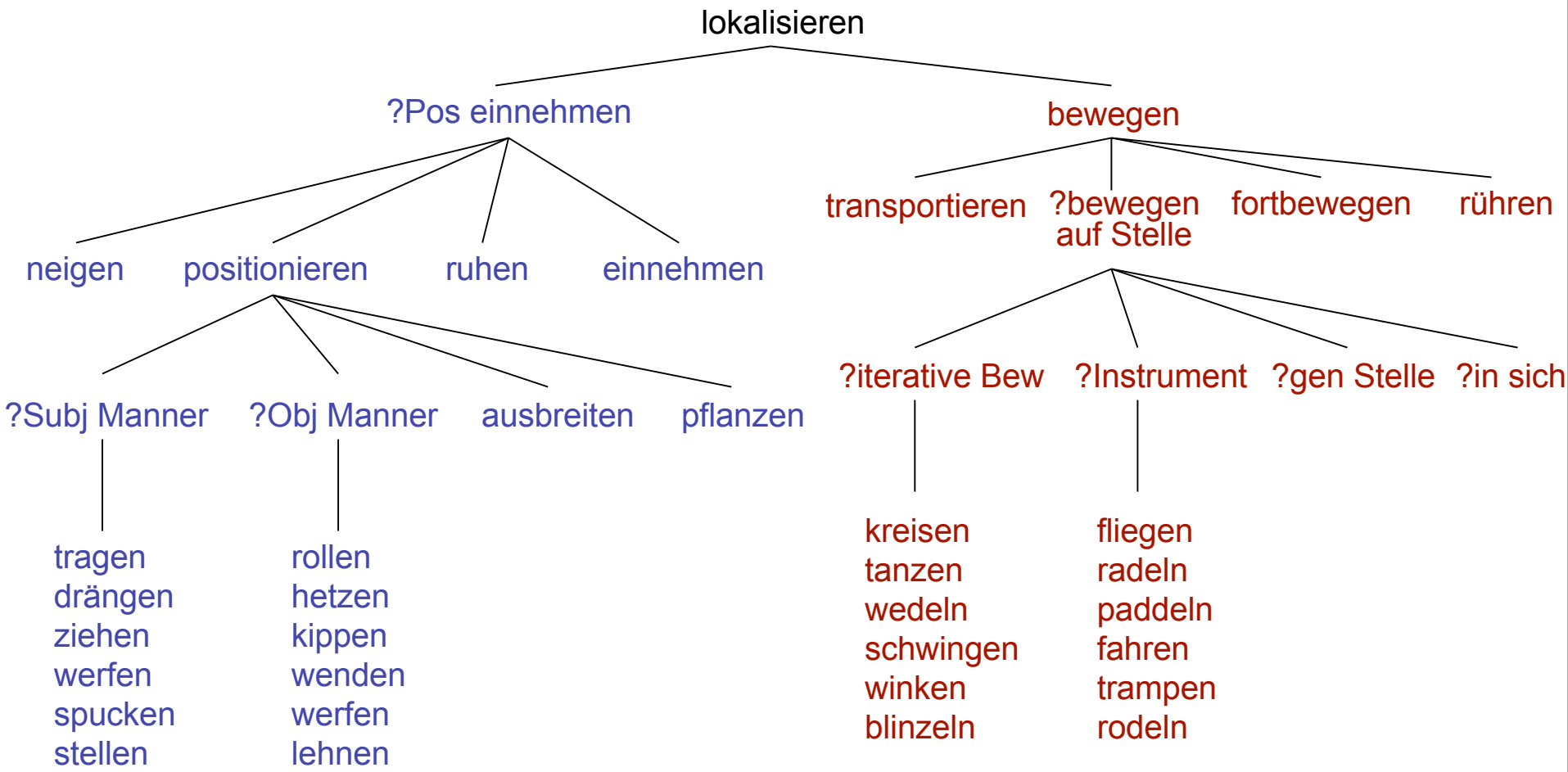
fahren	`to drive´
fliegen	`to fly´
gondeln	`to go in a carriage´
reisen	`to travel´

fortbewegt werden mit Mittel

WordNet / GermaNet

- Motivation: psycholinguistic research on lexical memory
- Goal: lexical semantic ontology
- Organisation: deep hierarchy
- Class relations: lexical and conceptual relations
- Class formation: synsets of synonymous verbs
- Basis for synonymy: substitution in prototypical context

GermaNet: MOM Examples



SALSA: German FrameNet

- Background: Fillmore‘s scenes-and-frames semantics
- Goal: lexicographic database with conceptual structures, participating elements and annotated example sentences
- Organisation: flat hierarchical classification
- Frame: conceptual structure describing a situation
- Frame element: role in frame
- Verb meaning: frame choice

SALSA: MOM Examples

- *Motion:* bewegen, fahren, fliegen, gehen, gleiten, heraus-springen, huschen, kreisen, rollen, springen, ...
- *Motion_Directional:* fallen, kippen, kreisen, schlingern, ...
- *Motion_Noise:* krachen, quietschen, trampeln
- *Self-Motion:* anspringen, aufspringen, besteigen, bewegen, bummeln, eilen, eintauchen, fliegen, gehen, gleiten, hasten, hineinspringen, klettern, kreisen, ...
- *Transfer:* gehen
- *Transportation:* bewegen, fahren, fliegen, rollen, schleichen
- *Travel:* abreisen, fahren, fliegen, gehen, in Urlaub gehen, reisen, zurückziehen

Automatic Induction: MOM Classes

Manner of Motion

Locomotion: gehen, klettern, kriechen, laufen, rennen, ...

Rotation: drehen, rotieren

Rush: eilen, hasten

Vehicle: fahren, fliegen, rudern, segeln

Flotation: fließen, gleiten, treiben

MOM Features across Classifications

- Central features, e.g. path, direction, self, noise, means
- Idiosyncratic features, e.g. deliberation, travel, preparation

→ There is no “true” semantic verb classification!

→ Still: learning about verb features for machine learning …

→ What’s the correlation with human judgements?

Web Experiment: Verb Associations

- Native German speakers
- Spontaneous associations on 200 German verbs
- Verbs varying in domain and frequency
- Random experiment setup: 50 verbs, 30 seconds each
- Types of semantic relationships and grammatical functions

→ Learning about verb features for machine learning!

(joint work with Katrin Erk and Alissa Melinger)

Outlook

- Clustering and classification experiments:
  explore algorithms, approach polysemy
- Clustering of particle verbs:
  specific treatment, transparency, semantic class
- Human intuition on semantic verb classes:
  associations, relations, functions, application
- NLP application for semantic verb classes