Inventory, Induction and Disambiguation of German Preposition Senses

Annual Meeting SFB 732
Sabine Schulte im Walde

Blaubeuren, November 14-16, 2008
Overview

1. **Sense inventory** of (German) prepositions
2. **Automatic induction** of preposition senses
3. **Disambiguation** of preposition senses
4. Preposition senses and *particle verb compositionality*
5. Preposition senses and *PP argumenthood*
6. Preposition senses in the SFB 732
Ambiguity of Prepositions

Central issues and motivation:

• Prepositions are notoriously ambiguous.
  Examples:  *by category / two o’clock / Mr. Jones*
  *nach drei Stunden / Berlin / Meinung*

• Number and distinction of senses → sense inventory

• Issues (theoretical ↔ computational):
  » definitions and agreement on sense inventory
  » properties of preposition senses
  » automatic induction of preposition senses
  » application of sense inventories
Sense Inventory of Prepositions

- Manual definitions (e.g., in DRT)
- Grammar books (e.g., Helbig/Buscha, Duden)
- Lexicography (e.g., frame semantic roles in FrameNet)
- CL projects (e.g., The Preposition Project)
- Parallel corpora
- Human judgements
CL-related projects:
• Bonnie Dorr’s Lexical Conceptual Structures
• Penn Treebank
• FrameNet/SALSA
• The Preposition Project (TPP) and SemEval
• Patrick Saint-Dizier‘s PrepNet
• ACL-SIGSEM Workshops on Prepositions
• Cyc
Parallel corpora:

- Take various translations of prepositions in parallel corpus as sense inventory
- Combine evidence from several languages (to identify coincidences of underspecified categories)
- Example:
  
  \[ \text{nach drei Stunden} \quad \text{after three hours} \]
  \[ \text{nach Berlin} \quad \text{to Berlin} \]
  \[ \text{nach Meinung} \quad \text{according to the opinion} \]
Sense Inventory of Prepositions

Collect human judgements on preposition senses:

- **Direct** (i.e., labelling prepositions with existing categories) vs. **indirect** (i.e., asking for sense properties) collection

- Use sense inventories to **annotate preposition senses** *(web experiment and corpus data) → validation*

- Ask for **example usages** (e.g., subcategorised words or example sentences) → source of preposition properties

- **Judgement of prototypical features** (e.g., preposition-noun pairs)

- **Existing work**: SemEval 2007
Automatic Induction of Senses

• Unsupervised learning
  » no agreed classification
    → explore data structure
    → use resulting clusters as senses

• Supervised learning
  » use existing classifications
    → validate and compare
Unsupervised Induction of Senses

- Unsupervised learning
  - Latent Semantic Classes (LSC):
    IMS approach (Rooth): two-dimensional clusters; useful for binary/n-ary lexical dependencies
  - LSC with selectional preferences (LSCpref):
    similar to LSC but generalisation over nouns
  - Self-Organising Maps (SOM):
    artificial neural network (Kohonen); topology preserving maps; visualisation of high-dimensional data
Supervised Induction of Senses

• Supervised learning
  » Decision Trees:
    validate various approaches to sense inventories;
    intuitive interpretation;
    facilitate exploration of preposition properties
    (via features)
- Learned features point to properties of prep senses, e.g.:
  - subcategorising word and word class
  - part-of-speech of subcategorised word (class)
  - subcategorised word (lemma and case)

  *relying on annotated/parsed data:*
  - PP function and semantic role
  - surrounding words and parts-of-speech
  - parse tree features

  *relying on parallel corpora:*
  - respective features from parallel corpus
Evaluation of Induced Senses

- Annotation of corpus data with preposition senses (based on manual inventories as well as clusters)
- **Labelling** of preposition senses (supervised approach: category names; unsupervised approach: clusters)
- **Model human judgements** on preposition senses (e.g., via preposition-noun pairs)
- Application of induced senses
Application of Preposition Senses

• **Theoretical definitions** of preposition senses and preposition sense properties

• **Disambiguation** of preposition senses in corpus data (raw corpus data and annotated/parsed data, e.g., LFG parses)

• Particle verb compositionality

• PP argumenthood
Prep Senses and Particle Verbs

• Preposition sense inventories useful as particle senses? → e.g., anbauen, anfassen, anlächeln, annehmen, etc.

• Focus: particle verb (pv) compositionality

• Challenge: syntax-semantics interface to induce pv compositionality from corpus data does not work
  Example: *Sie lächelt ihre Mutter. vs. Sie lächelt ihre Mutter an.

• Assumption: syntax-semantics mapping plus subcategorisation transfer works per particle verb sense
  Example: intrans(base_{an-x}) → trans(pv_{an-x})
Prep Senses and Particle Verbs

• Task:
  1. mapping between preposition senses and pv senses and
  2. decision about degree of pv compositionality

• Scenario: supervised experiments relying on pvs annotated with prep senses

Example: *an* has *x* senses; relying on pvs annotated with preposition senses, we can distinguish properties and transfer patterns with respect to the preposition/particle sense; for each new pv, determine similarity of its transfer patterns to *an* transfer patterns; decide about particle sense and compositionality
• Assumption:
  PP argumenthood requires knowledge about preposition senses:
  *beteiligen* requires an PP (of a certain type);
  *enden* requires PP of type temporal, location, etc.;
  *schneiden* does not subcategorise PP but is often complemented by instrumental *mit* PP

• Examples:
  *Er beteiligte sich schon am Anfang an der Wette.*
  *Das Spiel in Rom endete um 15 Uhr mit 1:0.*
• Experiment:

Does the distinction between preposition senses influence automatic judgements on PP argumenthood?

\[ \Rightarrow \text{use data annotated with preposition senses and calculate} \ 1. \text{agreement for humans and} \ 2. \text{correctness of learning approach based on prepositions vs. preposition senses} \]

» collec\[\text{t human judgements on PP argumenthood} \]

» supervised experiment learns and applies PP argumenthood decisions to parsing decisions
• Prepositions are notoriously ambiguous. Examples: *by category / two o’clock / Mr. Jones*  
  *nach drei Stunden / Berlin / Meinung*

• Number and distinction of senses → **sense inventory** as systematic semantic classification

• Parts:
  1. **Explore sense inventory of German prepositions** (manual, grammar books, CL, parallel corpora, human judgements) from **theoretical and computational perspectives**  
    → **complementary argumentation and properties** of sense discriminations
2. Validate approaches by supervised and unsupervised classification techniques

   \[\rightarrow\] explore data, create cluster analyses, validation and comparison, induce preposition properties

3. Apply sense inventories and sense properties: theoretical definitions, disambiguation of LFG parses, nominalisations

• Side tracks:
  a) Particle verb compositionality: Are preposition sense inventories useful as particle senses?
     \[\rightarrow\] e.g., anbauen, anfassen, anlächeln, annehmen, etc.
  b) PP argumenthood: Does the distinction between preposition senses influence automatic judgements?
Preposition Senses in the SFB 732

- Underspecification of preposition senses by preposition type/lemma
- Sense inventory expresses extent and mode of underspecification
- Specification of preposition senses by contextual features

Collaboration with other projects:
- Prepositions and particle verbs
- PP argumenthood learned by LSCpref
- LFG parse disambiguation
- Preposition senses for nominalisations