Representing Underspecification by Semantic Verb Classes incorporating Selectional Preferences

Sabine Schulte im Walde, Helmut Schmid, Christian Scheible, Christian Hying
{schulte,schmid,scheibcn,hyingcn}@ims.uni-stuttgart.de

SFB 732 "Incremental Specification in Context"
Project D4 "Modular Lexicalisation of Probabilistic Context-Free Grammars"

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Semantic Verb Classifications

- Groupings of verbs according to semantic properties, such as
  Break a Solid Surface with an Instrument: *break, crush, fracture, smash*, etc.
- Classes refer to general semantic level;
  idiosyncratic lexical semantic properties are underspecified
- Goals:
  - » organise verbs with respect to common properties (Levin 1993; Koenig & Davis 2001)
  - » generalisation over shared properties
    - → data sparseness in processing natural language
    - → applications: word sense disambiguation (Dorr & Jones 1996; Kohomban & Lee 2005), machine translation (Prescher et al. 2000; Koehn & Joang 2007), document classification (Klavans & Kan 1998), etc.

Automatic Class Induction

- **Verbs → classes**
- Verbs in common class: as similar as possible
- Verbs in different classes: as dissimilar as possible
- Parameters in automatic induction (among others): verbs, verb properties, algorithm
- Algorithm: Expectation-Maximisation
- Soft clustering → model polysemy of verbs
- Verb properties: selectional preferences
- Source: WordNet; find WordNet concept(s) that best describe the selectional preferences for a verb-frame function
- Example: *drink tea, drink coffee, drink beer*, etc. → *drink a beverage* (→ *drink a substance*)

Verb Class Probabilistic Model

p(speak, subj-pp.to, professor, audience)

$$p(v,f,a_{f1},...,a_{fn}) = \sum_{c\in C} p(c)\; p(v\mid c)\; p(f\mid c) \prod_{i=1}^{n} \sum_{r\in R} p(r\mid c,f,i)\; p(a_i\mid r)$$

$p(c)$	probability of verb class *c*
$p(v\|c)$	probability of verb *v* in class *c*
$p(f\|c)$	probability of frame *f* in class *c*
$p(r\|c,f,i)$	probability that i^{th} argument of frame *f* in class *c* is realised by WordNet concept *r* e.g., *p(*person *| c3, subj-pp.to, 1)*
$p(a\|r)$	probability that WordNet concept *r* is realised by argument head *a* e.g., *p(professor |* person*)*

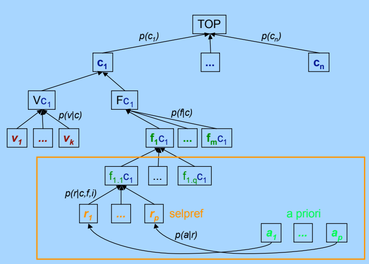
Steps

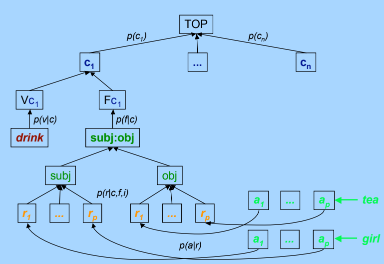
1. Input: verb-frame-argument tuples $<v,f,a_{f1},...,a_{fn}>$
   - » verb v,
   - » subcategorisation frame f,
   - » list of argument heads $a_{f1},...,a_{fn}$

   examples: <*drink subj-obj girl tea*> *43*
   <*speak subj-pp.to professor audience*> *27*
2. Training: Estimation-Maximisation algorithm; Minimum-Description Length principle
3. Output: cluster analysis with two dimensions (verbs and frames)

EM & MDL

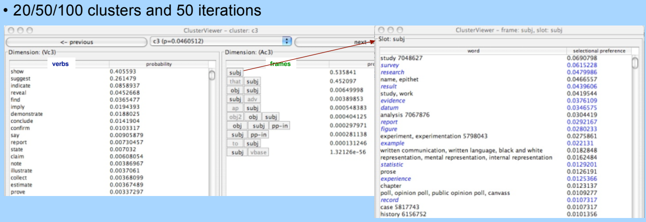
- Random initial assignment of frequencies/probabilities
- Initialisation of MDL cuts by WordNet top level *entity*
- Expansion of MDL cuts by next lower level
- Estimation of graph frequencies, using input tuples
- MDL cuts: leave or prune (recursively)
- Maximisation of graph probabilities

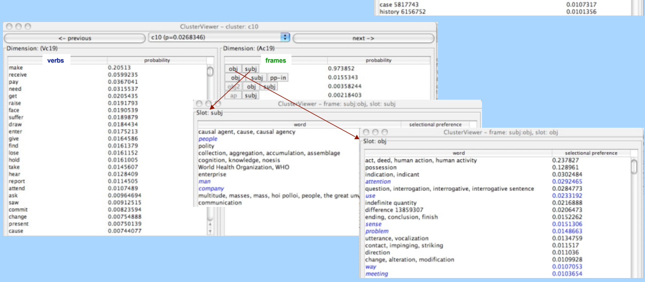




Experiments & Examples

- Tuples from BNC Viterbi parses (Carroll & Rooth, 1998)
- Only active clauses
- No auxiliary or modal verbs, no particle verbs, no personal pronouns
- 10/20 subcategorisation frame types
- Tuples with freq > 1 (51,569/55,980), freq ≥ 1 (671,461/815,553)
- 20/50/100 clusters and 50 iterations



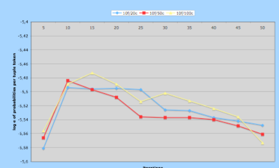


Evaluation

- Focus: statistical model of verb-argument tuples → model predicts tuple probabilities
- Comparison of verb class model predictions with baseline model
- Baseline model: product of conditional probabilities

$$p(v,f,a_{f_1}^{f_n}) = p(v)\; p(f\mid v) \prod_{i=1}^{n} p(a_i\mid a_i^{i-1},\langle v,f\rangle,f_i)$$

- Example: ⟨*speak, subj:pp-to, professor, audience*⟩
  p(speak)
  p(subj:pp-to | speak)
  p(professor | ⟨*speak, subj:pp-to*⟩*, subj)*
  p(audience | professor, ⟨*speak, subj:pp-to*⟩*, pp-to)*



Outlook

Parameters:

- Data sources: variation in domain, annotation and size
- Number of clusters and number of iterations: increase
- Initialisation of probabilities: variation
- Calculation of preferences against the a priori model
- MDL model: cut-based vs. synset-based

Goals:

- Modelling context by multi-dimensional soft clusters
- Induction of lexical information: verb senses and verb classes, subcategorisation and selectional restrictions, verb alternations
- Modular extension of unlexicalised PCFGs → statistical disambiguation for parse trees