Combining EM Training and the MDL Principle for an Automatic Verb Classification Incorporating Selectional Preferences

Sabine Schulte im Walde, Christian Scheible, Christian Hying, Helmut Schmid
{schulte,scheibcn,hyingcn,schmid}@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

Main Goals

- A **Statistical model** for verb-argument tuples: *p(<read, subj:obj, student, book>)*
- Induction of a **semantic verb classification** with clustering techniques
- Learning of verbal **selectional restrictions** which are represented with WordNet concepts

Features of the Model

- Statistical **soft clustering** approach**:** verbs are assigned to one or more verb classes
- Representation of **verbal polysemy** by the assignment to multiple classes
- Training on verb-argument tuples with the **Expectation Maximization** algorithm
- Generalization of selectional restrictions with **Minimum Description Length** principle
- The model is **smooth** because it generalizes over
  - the verbs of a cluster
  - the nouns instantiating the WordNet concepts representing the selectional restrictions

Semantic Verb Classification

- Grouping of verbs according to semantic properties (Levin 1993)
  Break a Solid Surface with an Instrument: *break, crush, fracture, smash*, etc.
- Goals:
  - Organisation of verbs wrt. shared properties
  - Generalization over verbs to counter sparse-data problems
- 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.

Probabilistic Verb Class Model

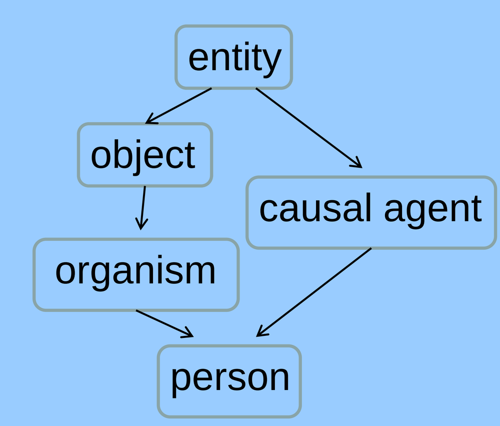
p(speak, subj-pp.to, professor, audience)

$$p(v,f,a_1,...,a_{n_f}) = \sum_c p(c)p(v|c)p(f|c)\prod_{i=1}^{n_f}\sum_r p(r|c,f,i)p(a_i|r)$$

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*)*

Conversion of WordNet into a Markov Model (Abney/Light)

- Additional node for each word
- Additional hyponym links from each concept to the members of its synset
- A probability for each hyponym link



Two types of WordNet-HMMs

- Many partial WordNet HMMs for *p(r|c,f,i)*
- One a priori model for *p(a|r)*

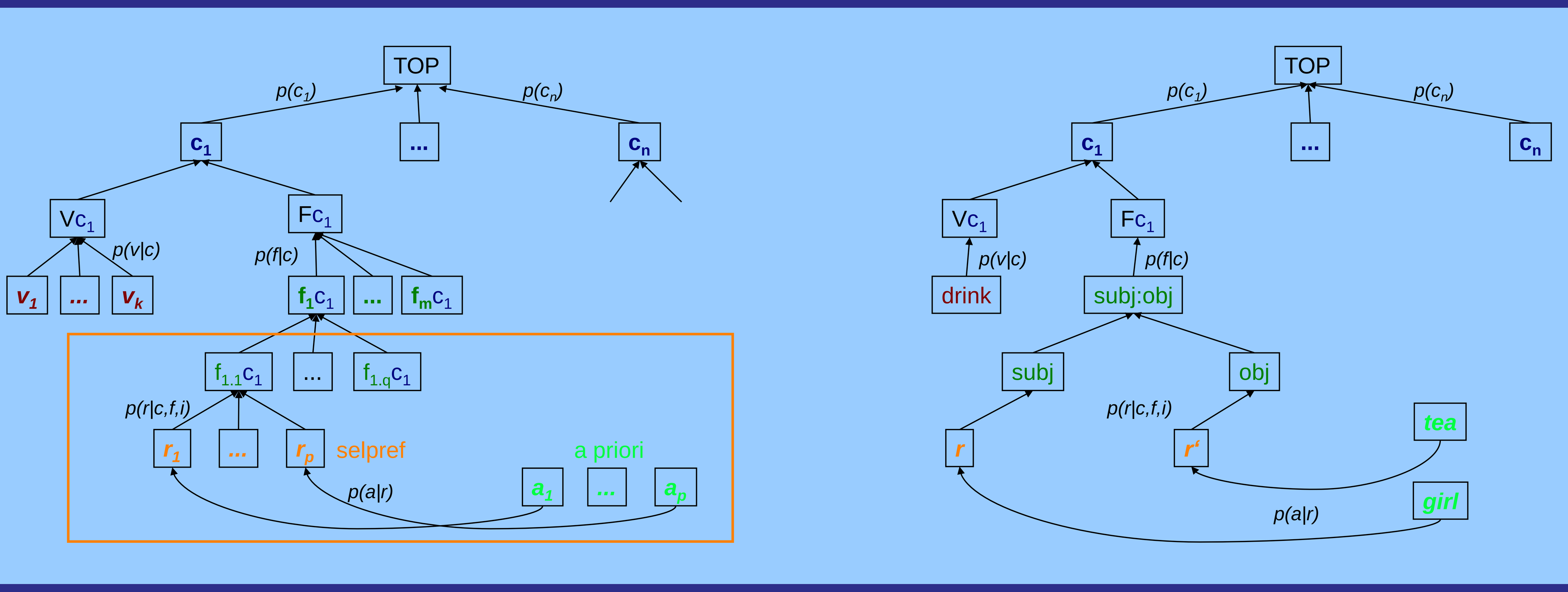
Path probabilities

- *p(a|r)* (and *p(r|c,f,i)*) is a sum of path probabilities
- Path probability = product of link probabilities

EM & MDL

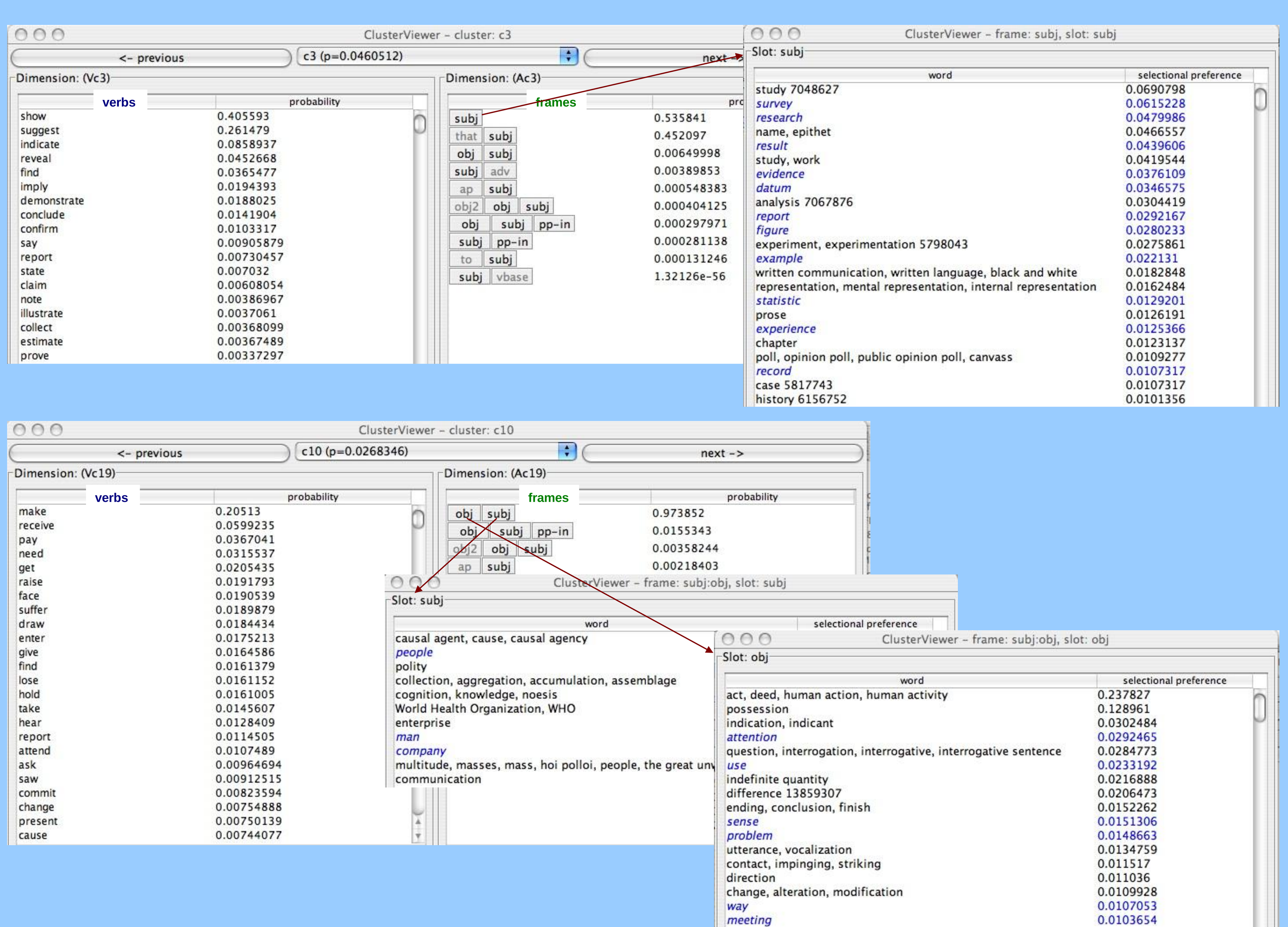
The whole model is represented as a large graph

- Initialisation of selectional restrictions (SR) with top concept *entity*
- Random initial assignment of probabilities
- Expansion of SR by the next lower level
- Estimation of graph frequencies from training tuples using the Inside-Outside algorithm
- Re-estimation of the probabilities
- MDL pruning of the selectional restrictions



Experiments & Examples

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

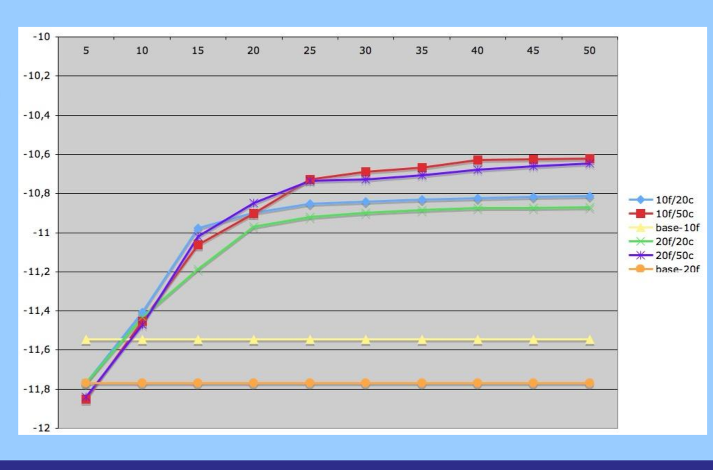


Evaluation

- Focus: statistical model of verb-argument tuples → model predicts tuple probabilities
- Comparison of verb class model predictions with baseline model
- Baseline model without hidden variables

$$p(v,f,a_1,...,a_{n_f}) = p(v)p(f|v)\prod_{i=1}^{n_f} p(a_i|a_1^{i-1},v,f,f_i)$$

- Example:
  p(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

- Experiments with other languages and corpora
- Refinement of the model (representation of alternations and collocations, etc.)
- Refinement of the training (split and merge clusters, training on data slices)
- Applications
  - Induction of verb classes, subcategorisation, and selectional restrictions
  - Detection of verbal polysemy, verb alternation, and collocations
  - Automatic assignment of new nouns to WordNet synsets
  - Refinement of a PCFG parser with verb-argument association scores