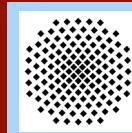


# Representing Underspecification by Semantic Verb Classes incorporating Selectional Preferences



Sabine Schulte im Walde, Helmut Schmid, Christian Scheible, Christian Hyng  
[{schulte,schmid,scheibcn,hyingcn}@ims.uni-stuttgart.de](mailto:{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

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

$p(c)$  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.,  $p(\text{person} \mid c3, \text{subj-pp.to}, 1)$

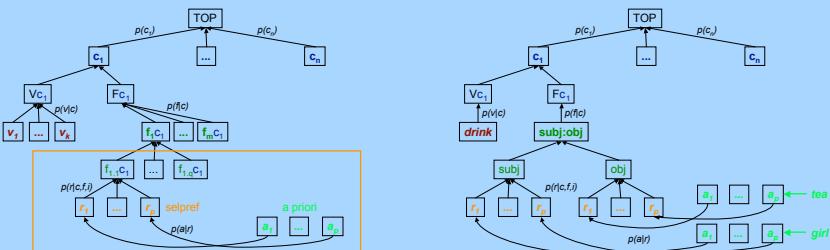
$p(a|r)$  probability that WordNet concept  $r$  is realised by argument head  $a$   
e.g.,  $p(\text{professor} | \text{person})$

## Steps

1. Input: verb-frame-argument tuples <v,f,a<sub>f1</sub>,...,a<sub>fn</sub>>
    - » verb v,
    - » subcategorisation frame f,
    - » list of argument heads a<sub>1</sub>,...,a<sub>fn</sub>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

Dimension: (V,19) ← previous    < 10 (p=0.0268440)    next → >

Dimension: (A,19)

	frames	probability
verb	act, do, human action, human activity	0.037862
make	action, aggregation, accumulation, assembly	0.035923
receive	causal, cause, causal agency	0.035923
give	causal, cause, causal agency	0.035923
need	causal, cause, causal agency	0.035157
buy	commercial, commerce, market	0.035157
raise	causal, cause, causal agency	0.031973
take	causal, cause, causal agency	0.031973
suffer	causal, cause, causal agency	0.031973
find	causal, cause, causal agency	0.031973
enter	causal, cause, causal agency	0.031973
get	causal, cause, causal agency	0.031973
hold	causal, cause, causal agency	0.031973
see	causal, cause, causal agency	0.031973
hear	causal, cause, causal agency	0.031849
attend	causal, cause, causal agency	0.031748
ask	causal, cause, causal agency	0.031748
comment	communication	0.031748
change	composition, masses, mass, hel pollit, people, the and us	0.031748
present	communication	0.031748
cause	communication	0.03174707

Slot: subj

Causal viewer - frame (subj, slot, obj)

selected preference

	word	selectional preference
act	act, do, human action, human activity	0.227827
assumption	assumption	0.0320484
indication	indication, indicate	0.0320484
question	question, interrogation, interrogative sentence	0.0284735
indefinite quantifier	indefinite quantifier	0.0284735
ending	ending, conclusion, finish	0.02152642
interference	interference, vocalization	0.01374159
impingement	impingement, intruding	0.01374159
direction	direction, inspiring, linking	0.010306
conversion	conversion, alteration, modification	0.010306
say	say, say, saying, speaking	0.010306

## 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
$$p(v, f, a_i^{f_i}) = p(v) \cdot p(f \mid v) \prod_{i=1}^n p(a_i \mid a_i^{i-1}, \langle v, f \rangle, f_i)$$
  - Example:  $\langle \text{speak}, \text{subj:pp-to}, \text{professor}, \text{audience} \rangle$



Verb Class	Arg Class 1	Arg Class 2	Arg Class 3	Arg Class 4
Verb Class 1	-0.50	-0.45	-0.40	-0.45
Verb Class 2	-0.50	-0.40	-0.35	-0.40
Verb Class 3	-0.50	-0.35	-0.30	-0.35
Verb Class 4	-0.50	-0.30	-0.25	-0.30



## 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. syntex-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