# Relation Extraction with Hybrid NLP


***Feiyu Xu & Peter Adolphs***

***DFKI LT Lab***

Joint research with

Hans Uszkoreit and Hong Li


DELPH-IN Paris, 2010

# Parameters in Real-World IE Tasks

- Document structure
  - Free text
  - Semi-structured
  - Structured

- Linguistic annotation
  - Shallow NLP
  - Deep NLP

- Complexity, Distinctiveness and specificity of relation
  - Unary
  - N-ary

- Depth of extraction
  - Recognition
  - Classification
  - Semantic role labelling

- Degree of automation of rule contruction
  - Semi-automatic
  - Supervised
  - Semi-Supervised
  - Minimally-Supervised
  - Unsupervised

- Human interaction/contribution

- Data properties
  - Domain relevance
  - Redundancy
  - Connectivity

- Evaluation/validation
  - With/without gold standard
  - Performance: recall & precision
  - Interaction among parameters

# Domain Adaptive Relation Extraction (DARE)

- Xu, Feiyu. 2007. Bootstrapping Relation Extraction from Semantic Seeds.

  PhD-thesis, Saarland University
  - http://dare.dfki.de

- Xu, Feiyu, Hans Uszkoreit, and Hong Li. 2007. A seed-driven bottom-up machine learning framework for extracting relations of various complexity.

  ACL 2007.

# Extensions of DARE

- Feiyu Xu, Hans Uszkoreit, Hong Li. Task driven coreference resolution for relation extraction. ECAI 2008.
- Xu, Feiyu, Hans Uszkoreit, Hong Li, and Niko Felger. Adaptation of relation extraction rules to new domains. LREC 2008.
- Hans Uszkoreit, Feiyu Xu, Hong Li. Analysis and Improvement of Minimally Supervised Machine Learning for Relation Extraction. NLDB 2009. Keynote.
- Xu, Feiyu, Hans Uszkoreit Sebastian Krause and Hong Li. Boosting relation extraction with limited closed-world knowledge. COLING 2010.

# DREAM

- Automatically learn relation extraction grammars for each application domain on demand, with minimal human intervention

- Learn a group of reusable grammars for various relation types

IE is an approximation of language understanding

# Challenges

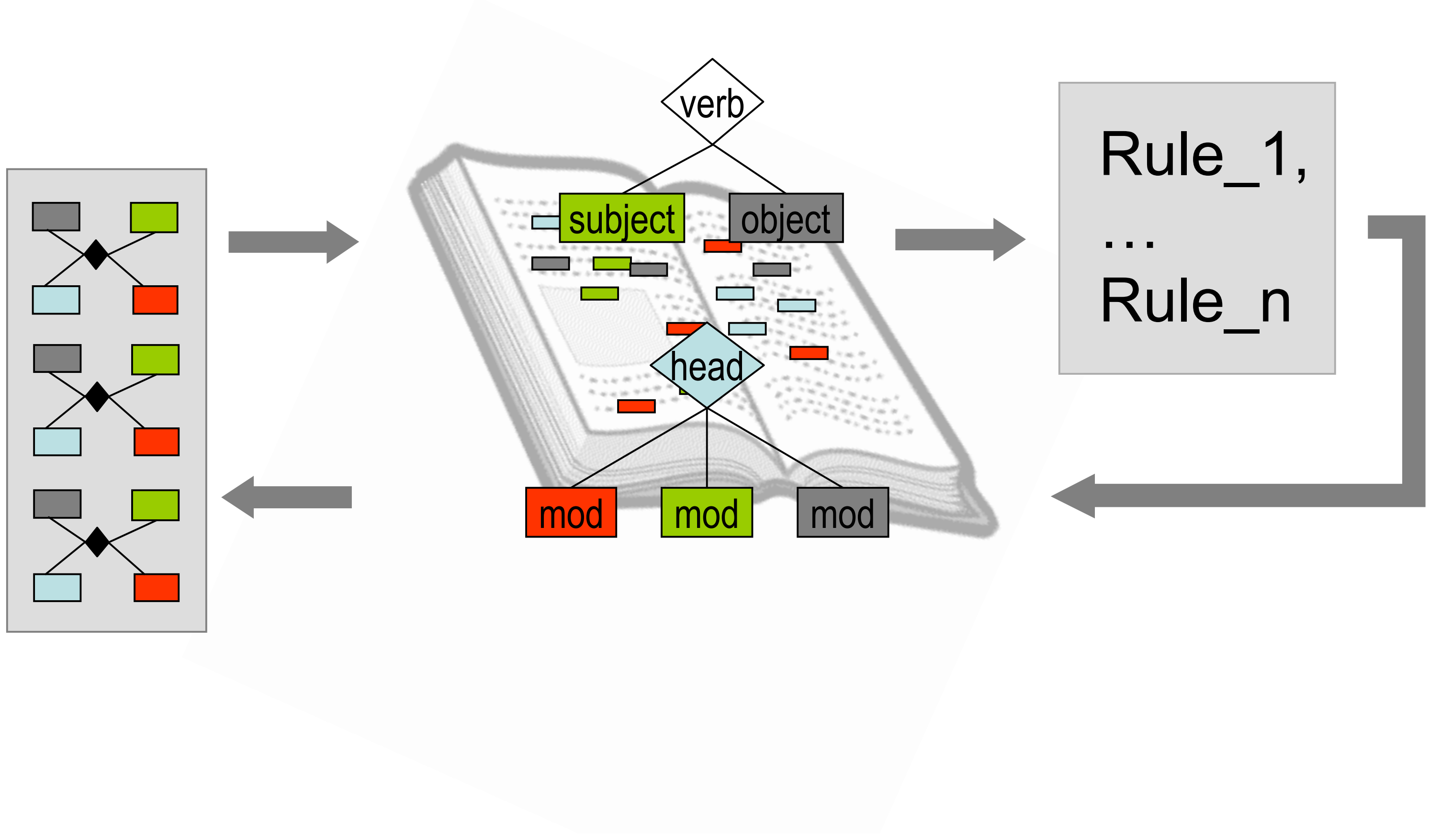
- Easy adaptation to new relation types with varied complexity
- Automatic learning without annotated corpus
- Exhaustive discovery of relevant linguistic patterns
- Integration of semantic role information into linguistic patterns

# DARE:
# Bootstrapping Relation Extraction with Semantic Seed

Adapted from

DIPRE (Brin, 1998) and Snowball (Agichtein & Gravano, 2000)

but extended and enriched with linguistic analysis



# **Novel** Properties of DARE

- Samples of target relation instances serve as semantic seed
- Systematic treatment of n-ary relations and their projections
- Exploitation of relation projections for pattern discovery
- Bottom-up compositional pattern discovery
- A recursive linguistic rule representation
- Rules contain semantic roles w.r.t. to target relation

# Example in Prize Award Domain

- Target relation

  <recipient, prize, area, year>

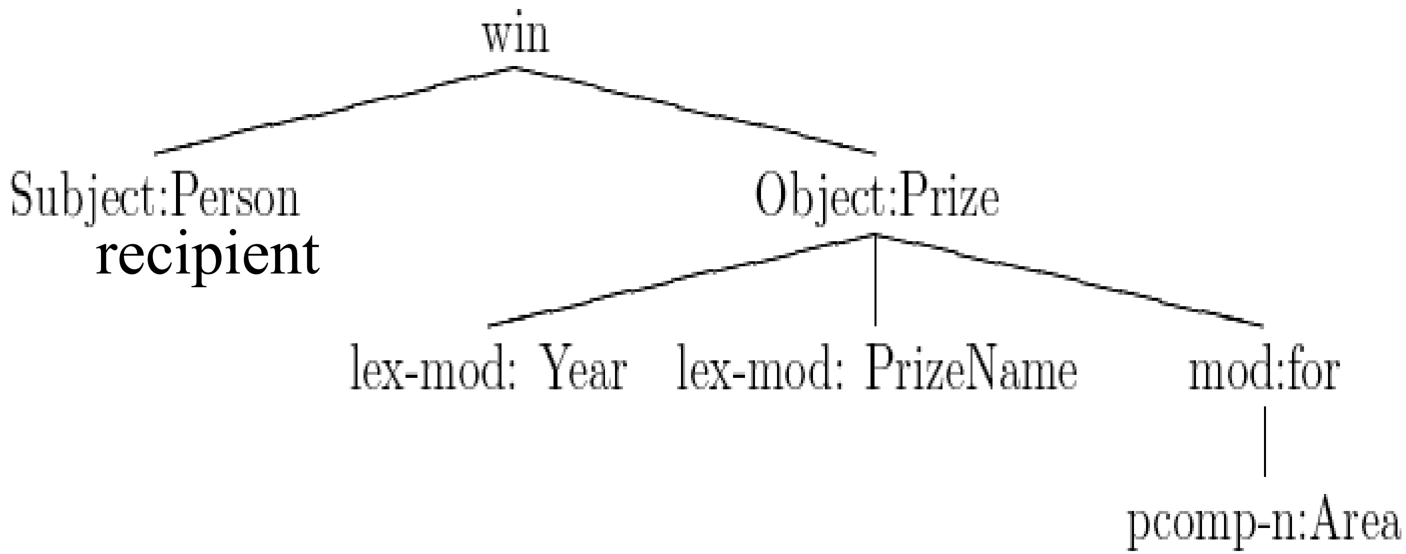
- Seed example

  <Mohamed ElBaradei, Nobel, Peace, 2005>

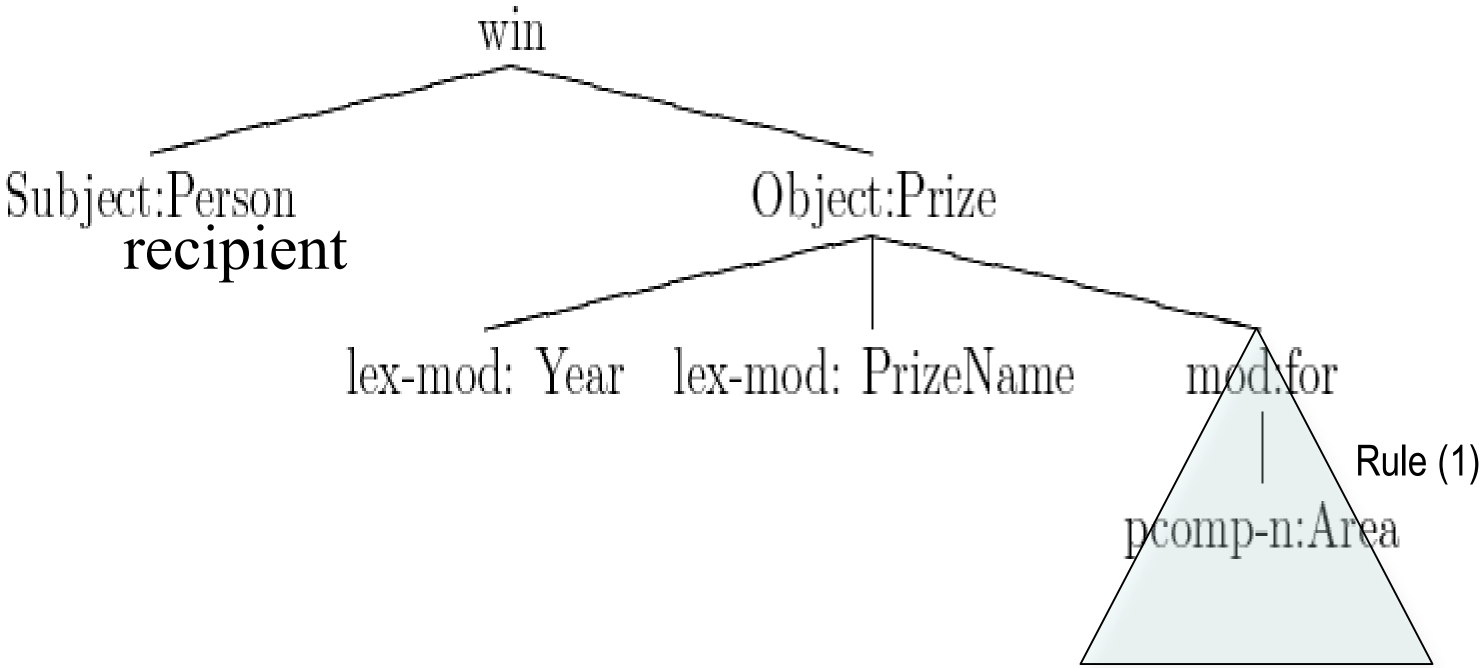
- Sentence matched with the seed

  Mohamed ElBaradei won the 2005 Nobel Prize for Peace on Friday for his efforts to limit the spread of atomic weapons.

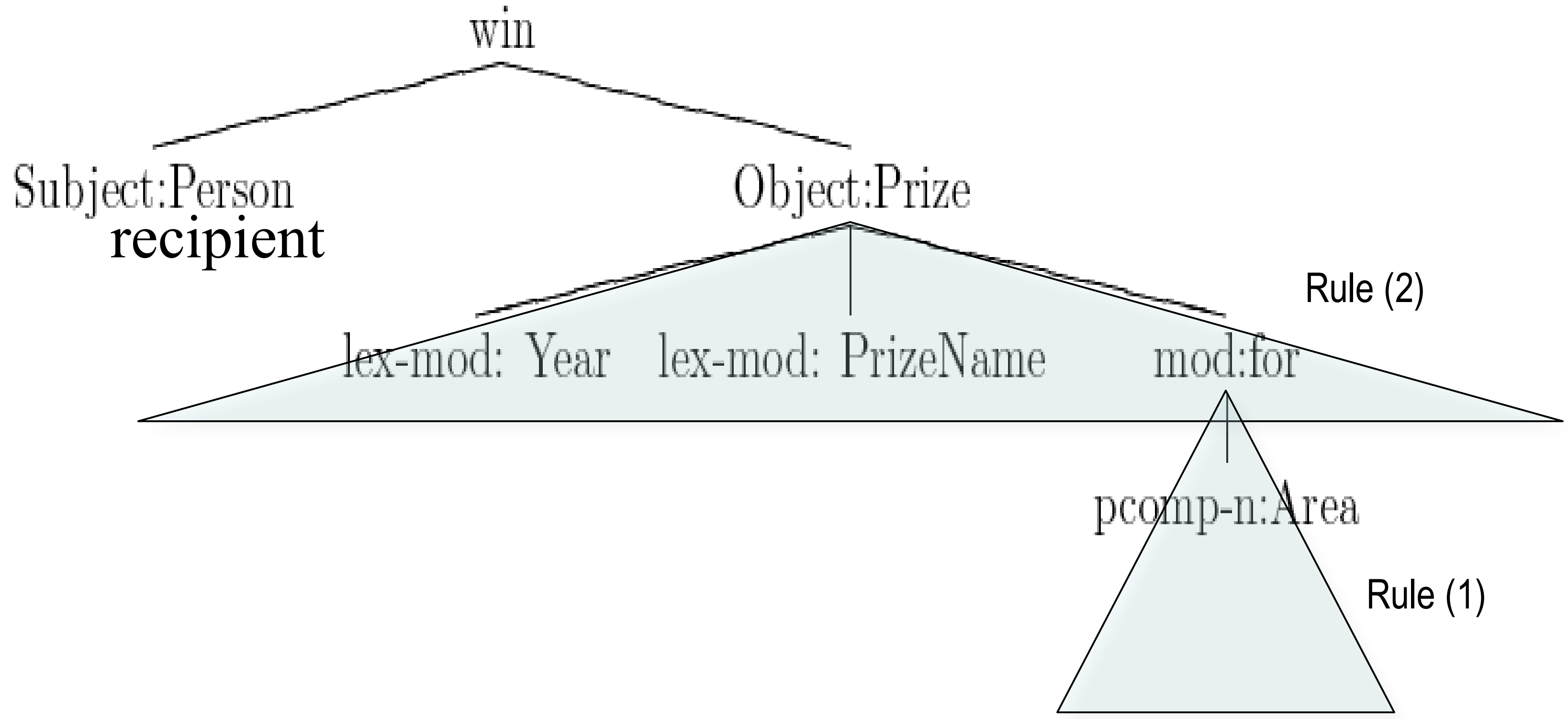
# Matched Dependency Tree mit Semantic Roles



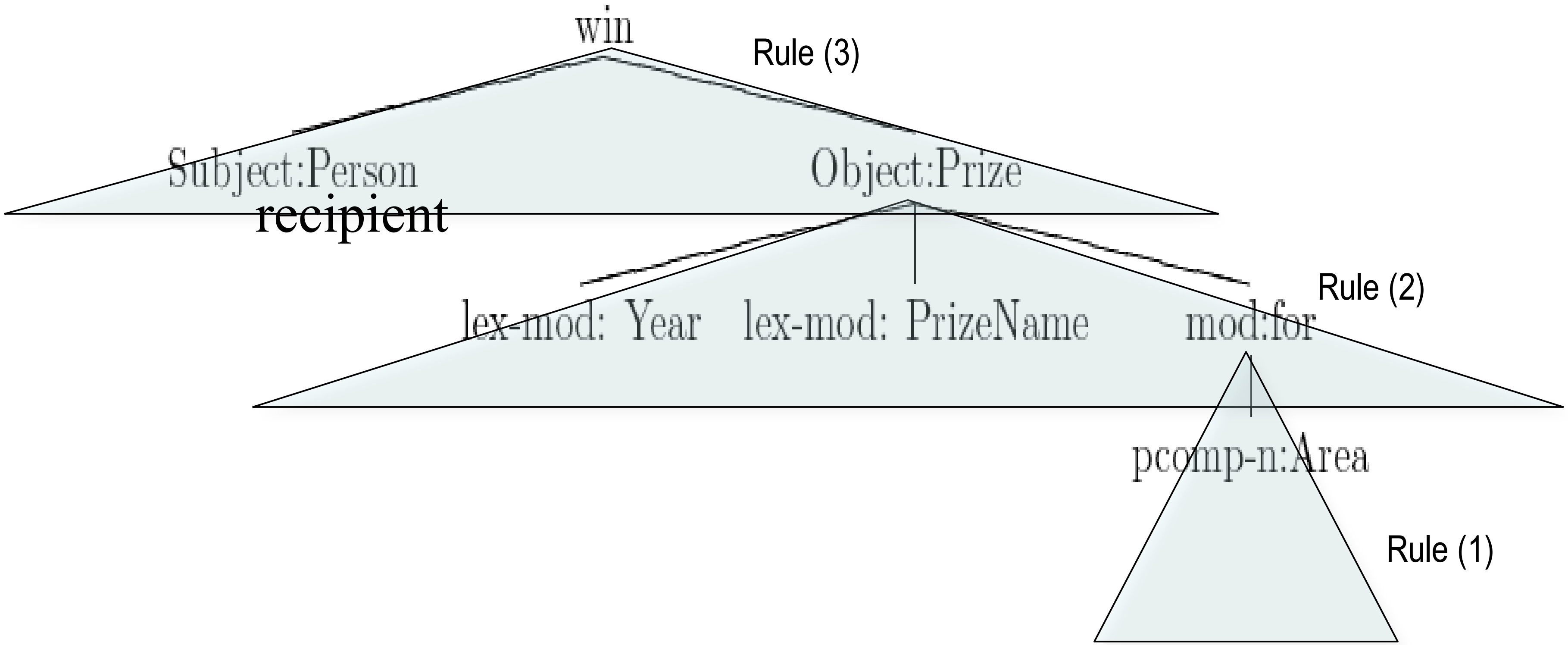
# Bottom Up Rule Learning



# Bottom Up Rule Learning



# Bottom Up Rule Learning



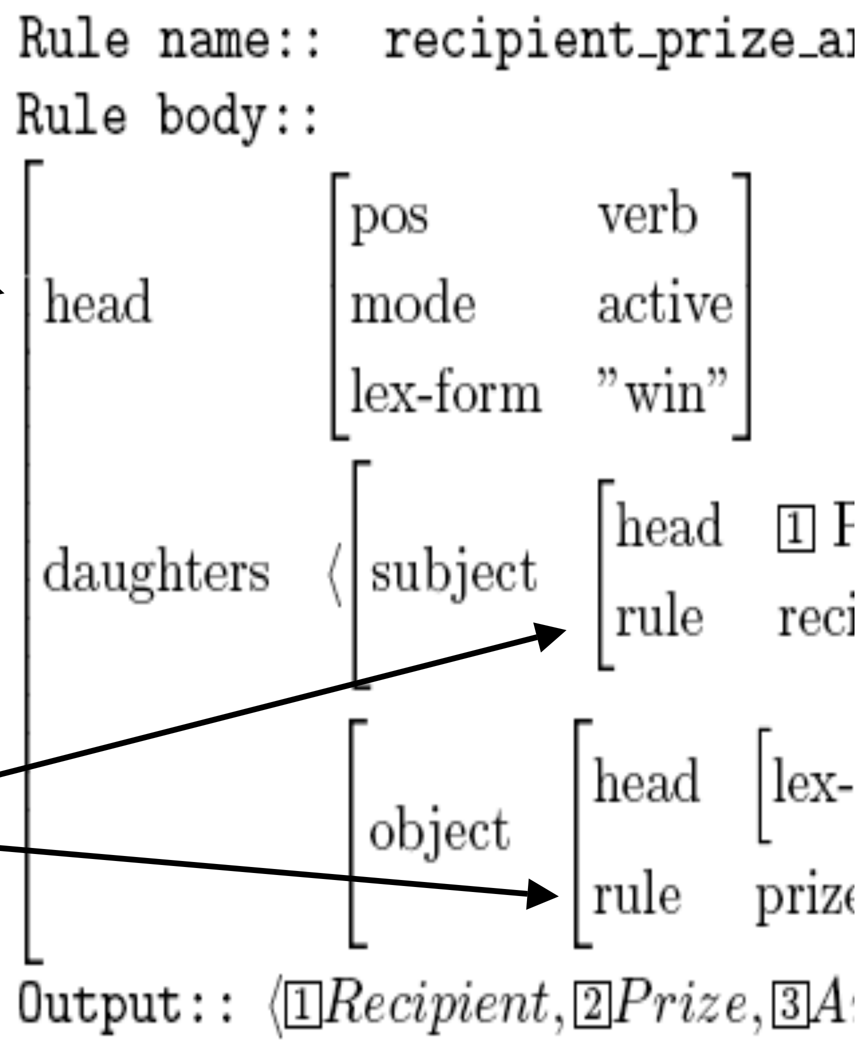
# DARE Rule Components

**1. rule name**: ***$r_i$***;

**2. rule body:** in AVM format containing:

- **head**: the linguistic annotation of the top node of the linguistic structure;
- **daughters**: its value is a list of specific linguistic structures (e.g., subject, object, head, mod), derived from the linguistic analysis, e.g., dependency structures and the named entity information;
- **rules**: its value is a DARE rule which extracts a subset of arguments of the target relation.

**3. Output:** n-tupel of arguments with their roles



# Rule (1)

*2005 Nobel Prize <u>for Peace</u>*

`Rule name::` `area_1`

`Rule body::`

$$\begin{bmatrix} \text{head} & \begin{bmatrix} \text{pos} & \text{preposition} \\ \text{lex-form} & \text{“for”} \end{bmatrix} \\ \text{daughters} & \left\langle \begin{bmatrix} \text{pcomp-n} & \begin{bmatrix} \text{head} & \boxed{1}\ \text{Area} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

`Output::` $\langle \boxed{1} Area \rangle$

# Rule (2)

*2005 Nobel Prize <for Peace>*

`Rule name::` `year_prize_area_1`

`Rule body::`

$$\begin{bmatrix} \text{head} & \begin{bmatrix} \text{pos} & \text{noun} \\ \text{lex-form} & \text{“prize”} \end{bmatrix} \\ \text{daughters} & \left\langle \begin{bmatrix} \text{lex-mod} & \begin{bmatrix} \text{head} & \boxed{1}\ \text{Year} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{lex-mod} & \begin{bmatrix} \text{head} & \boxed{2}\ \text{Prize} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{mod} & \begin{bmatrix} \text{head} & \begin{bmatrix} \text{pos} & \text{preposition} \\ \text{lex-form} & \text{“for”} \end{bmatrix} \\ \text{rule} & \text{area\_1::}\ \langle \boxed{3} \mathit{Area} \rangle \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

`Output::` $\langle \boxed{1} \mathit{Year}, \boxed{2} \mathit{Prize}, \boxed{3} \mathit{Area} \rangle$

# Rule (3)

`Rule name::  recipient_prize_area_year_1`

`Rule body::`

$$\begin{bmatrix} \text{head} & \begin{bmatrix} \text{pos} & \text{verb} \\ \text{mode} & \text{active} \\ \text{lex-form} & \text{``win''} \end{bmatrix} \\ \text{daughters} & \left\langle \begin{bmatrix} \text{subject} & \begin{bmatrix} \text{head} & \boxed{1}\ \text{Person} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{object} & \begin{bmatrix} \text{head} & \begin{bmatrix} \text{pos} & \text{noun} \\ \text{lex-form} & \text{``prize''} \end{bmatrix} \\ \text{rule} & \text{year\_prize\_area\_1::}\ \langle \boxed{4} Year, \boxed{2} Prize, \boxed{3} Area \rangle \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

`Output::` $\langle \boxed{1} Recipient, \boxed{2} Prize, \boxed{3} Area, \boxed{4} Year \rangle$

# Experiments

- Two domains
  - Nobel Prize Awards: <recipient, prize, area, year>
  - Management Succession: <person_in, person_out, position, organisation>

- Test data sets

| Data Set Name | Doc Number | Data Amount |
|---|---|---|
| Nobel Prize Corpus | 3328 | 18.4 MB |
| MUC-6 Corpus | 199 | 1MB |

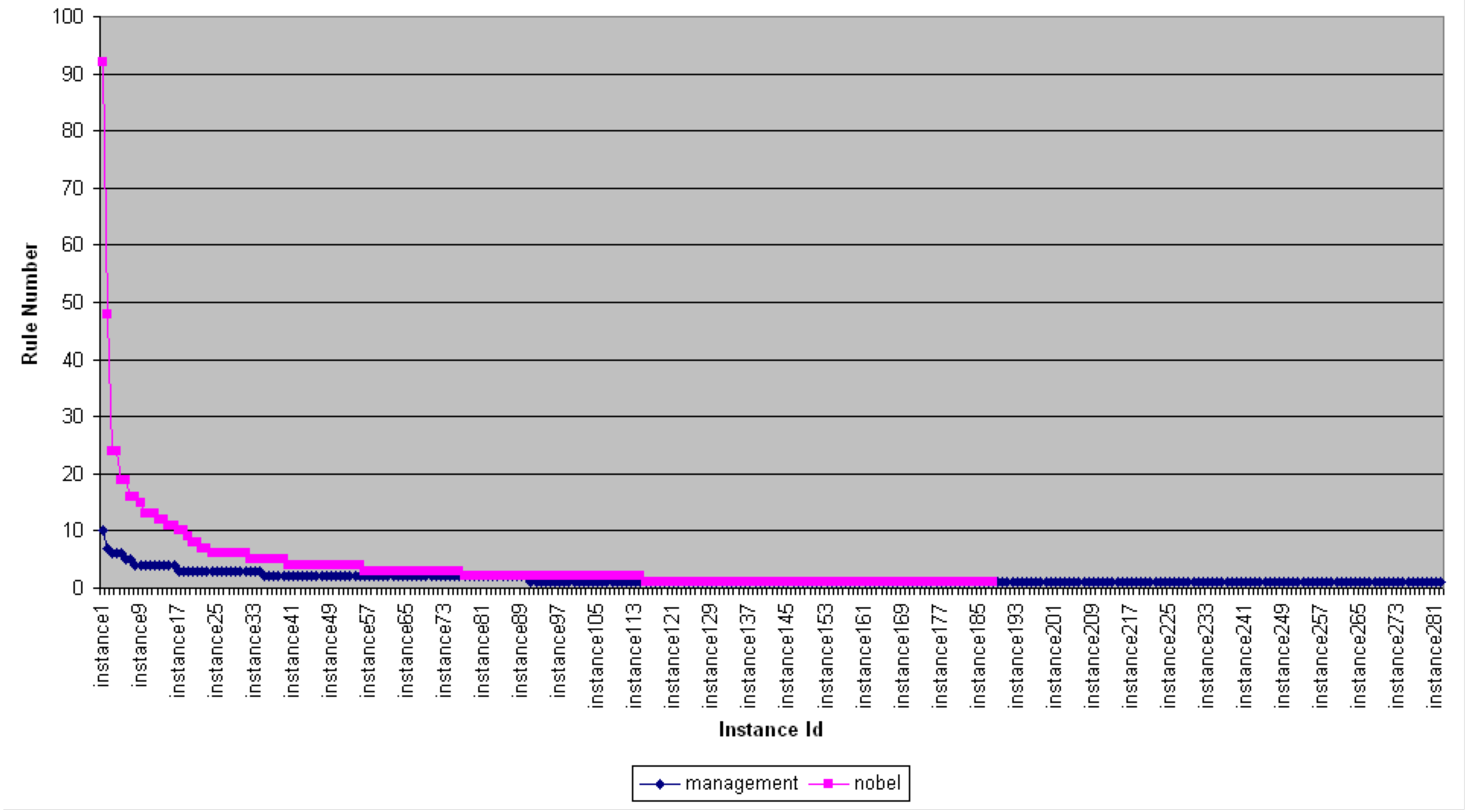
# Evaluation Against Ideal Tables

| Data Set | Seed | Precision | Recall |
|---|---|---|---|
| Nobel Prize | <[Zewail, Ahmed H], nobel, chemistry,1999> | **80.6%** | **62.9%** |

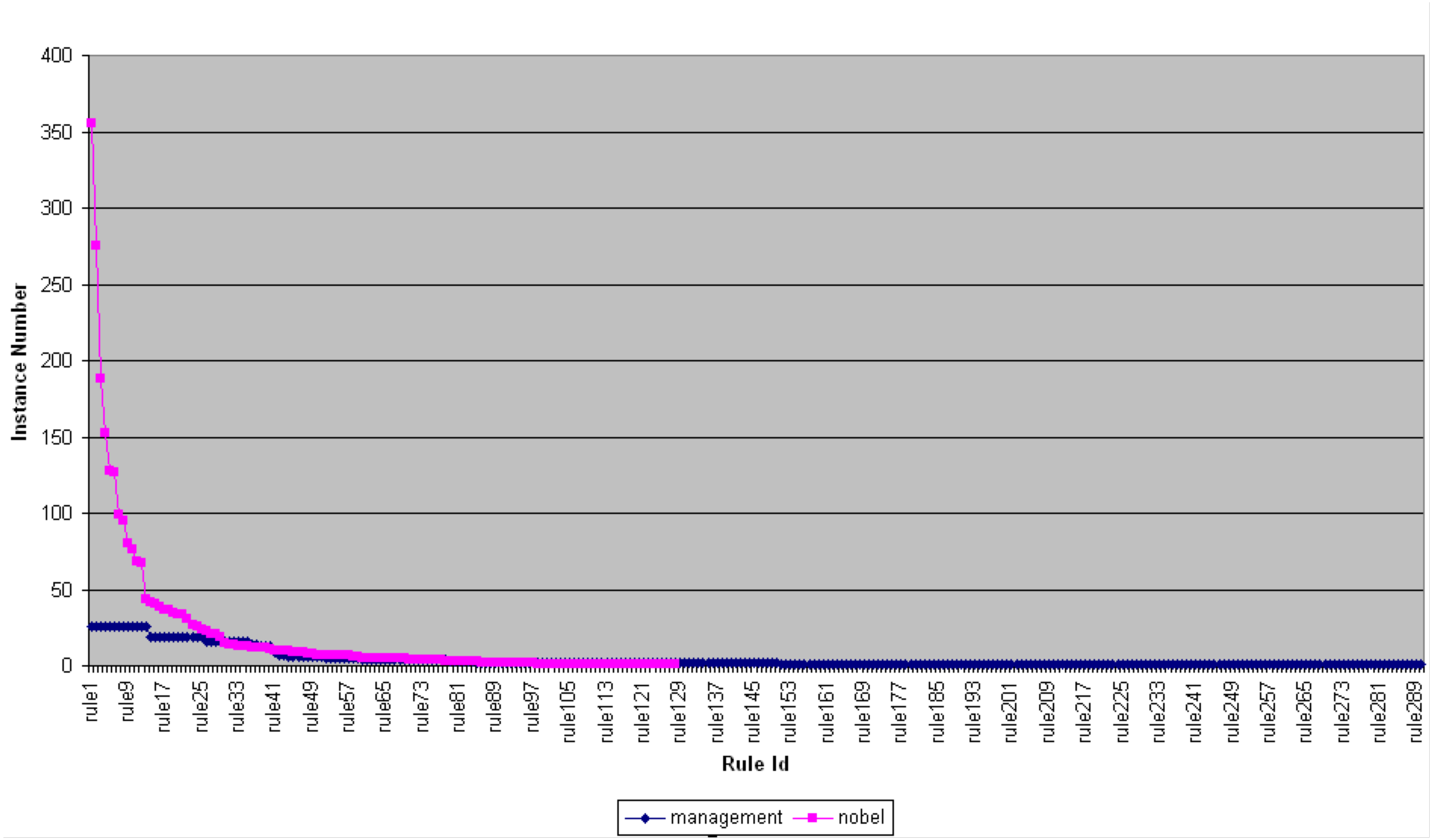
# Management Succession Domain

| Initial Seed # | Precision | Recall |
|---|---|---|
| **1** | **12.6%** | **7.0%** |
| **1** | **15.1%** | **21.8%** |
| **20** | **48.4%** | **34.2%** |
| **55** | **62.0%** | **48.0%** |

Instance to Pattern
Nobel Prize vs. Management Succession



# Rule to Instances
# (Nobel Prize vs. Management Succession)



# Insights

- Results from graph theory help to understand the requirements on data.

  **Example:** small world property

- For data sets with continents and islands, we can sometimes exploit additional data or auxiliary domains to bridge the islands by learning rare patterns.

  **Example:** use of Nobel prize domain for learning patterns for events concerning less popular prizes (many other prizes could be detected)

# Error analysis

- *content*: Wrong facts are expressed by the corpus sentences
- *modality*: The facts or events are embedded in a scope of a modality, which either denies or weakens the truth value of the facts or events, e.g, negation or wish
- *NLP annotations*: the NLP components deliver a wrong analysis or cannot analyse the sentence
- *rule*: the learned rules lead to wrong seeds prizes (many other prizes could be detected)

| content % | modality % | *SProUT* % | MINIPAR % | *SProUT* & MINIPAR % | rule % |
|---|---|---|---|---|---|
| 11.8 | 17.6 | 5.9 | 38.2 | 11.8 | 14.7 |

Table 6.14: Distribution of error types

NLP 55.9%

| content % | modality % | *SProUT* % | MINIPAR % | *SProUT* & MINIPAR % | rule % |
|---|---|---|---|---|---|
| 11.8 | 17.6 | 5.9 | 38.2 | 11.8 | 14.7 |

Table 6.14: Distribution of error types

## combination of coordination and apposition is very challenging

**Minipar is too eager.**

William Crowe, former chairman of the joint chiefs of staff; Hans Bethe, [*the Nobel Prize-winning physicist, and Herbert York*], a former founding director of the Livermore National Laboratories sent letters to the Senate urging action on the treaty now.

# Quality Analysis of Rules

| good | bad | useless | dangerous |
|---|---|---|---|
| 11.7% | 1.6% | 83% | 3.7% |

# Motivation for HPSG in DARE

- Precision
  - Precise and deeper understanding
    - Reduction of “bad rules” or “dangerous rules”
    - Identification of modality context
- Recall
  - Type hierarchy
    - Generalization of learned rules with the help of type hierarchy
  - More general linguistic rules

# Possible Side Result

- Learning domain-specific subgrammar of ERG
  - Learning domain-specific/relevant reading
    - Active learning

# Hybrid NLP for DARE

- Parallel processing
  - (weighted or qualified) voting
  - Merging (of results)

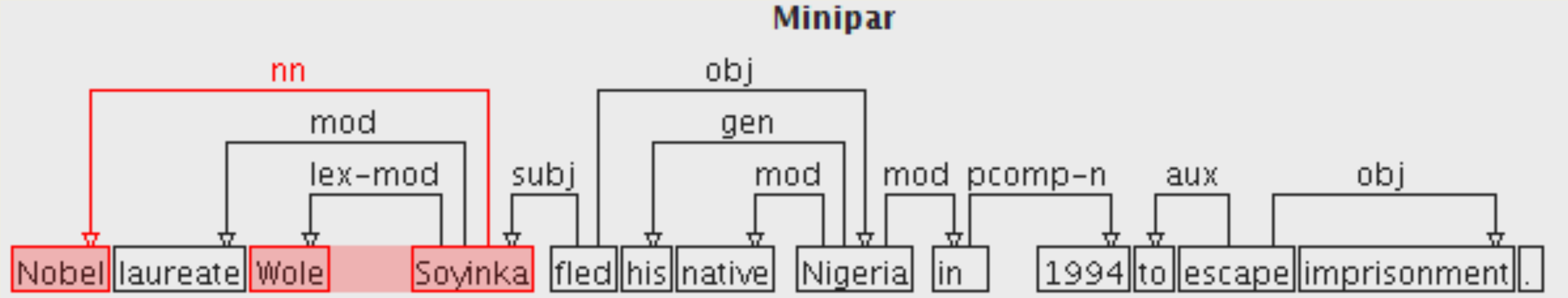
- Interleaved processing
  - DARE-rule composition from various parsers
    - Global structure vs. local structure: e.g. VPs or NPs

# What we want to evaluation

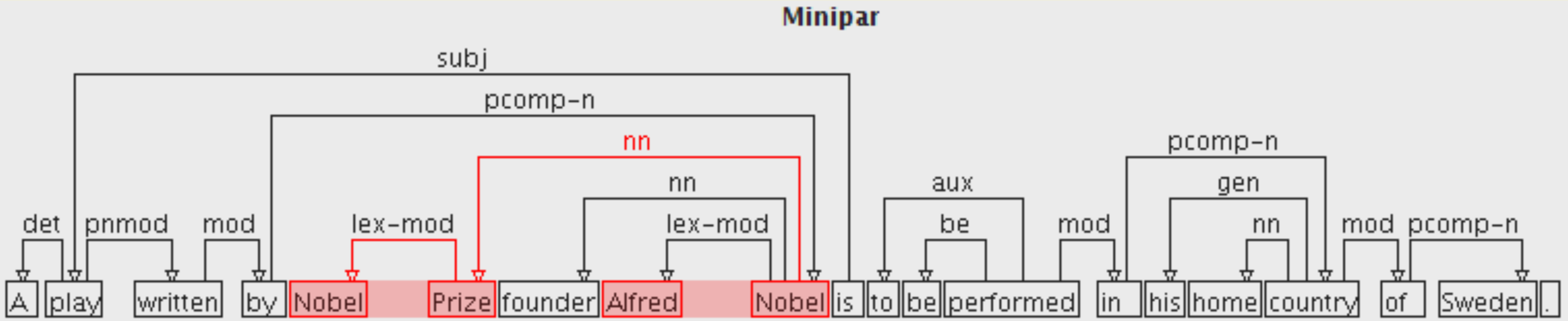
- Strength of ERG with respect to linguistic phenomena in comparison to dependency parsers
- Performance
  - Recall
  - Precision
- Comparison of rules learned from ERG and other dependency parsers
  - Overlap
  - Richness
  - Generality

# Error Analysis for MINIPAR – Example

- Compound structure is broken in MINIPAR
- Rule learned from this example:




- „laureate“not part of the rule, though crucial
- When applied to extract relation instances:



# Error Analysis for MINIPAR – Example

- Appositions and conjunctions are very often confused

- Rule learned from sentence:

  Because last year's Nobel laureate in literature, the Italian playwright Dario Fo, proved a surprise, speculation is cautious this year.

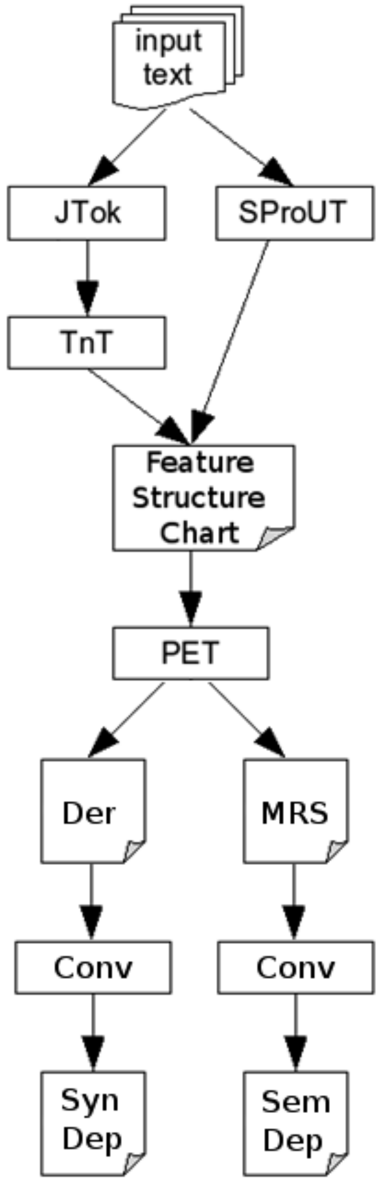
- Relation extracted from sentence:

  The elite partners of Long-Term Capital include Myron S. Scholes and Robert C. Merton, both Nobel laureates in economics, David W. Mullins Jr., a former vice chairman of the Federal Reserve Board, and Lawrence E. Hilibrand.

- NB: This is often very hard for the current parse disambiguation model of the ERG, too, but we do have the readings.

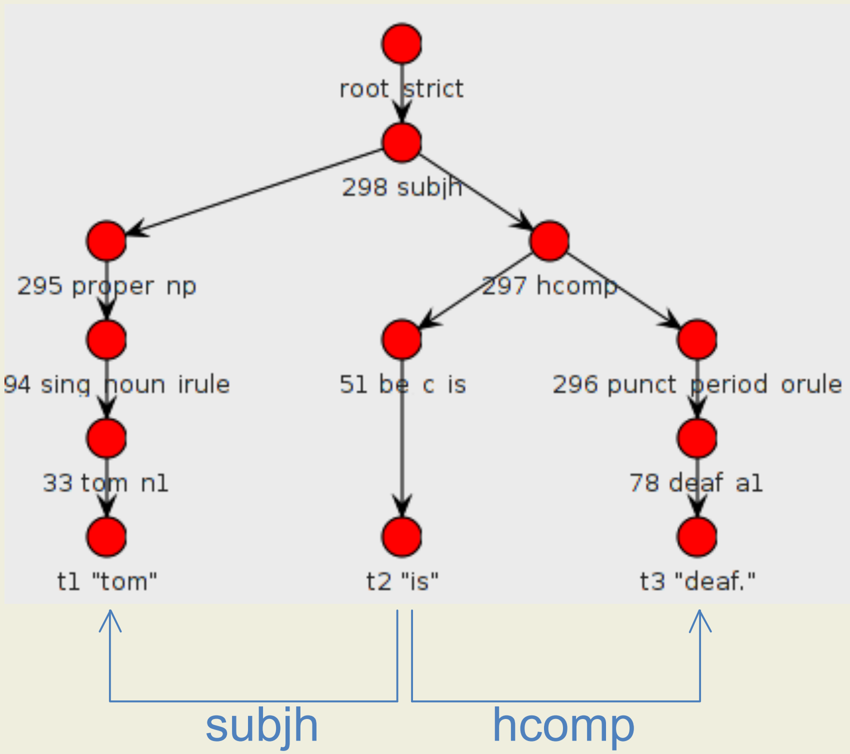
# Hybrid Pipeline

- Tokenization by jTok
- Part-of-speech tagging by TnT
- Named-entity recognition by SProUT
- Tokens with TnT annotations and tokens with SProUT annotations are merged into a single FS input chart for PET
- Finally, syntactic derivation and semantic MRS are converted to dependency structures



# Converting Derivations to Dependency Structures

- for every binary node, determine the heads $h_l$ and $h_r$ of the two daugthers
- $h_l$ is the head and $h_r$ the dependent iff the node is produced by a head-intial rule, otherwise: $h_l$ is the dependent and $h_r$ the head
- dependency label: rule name

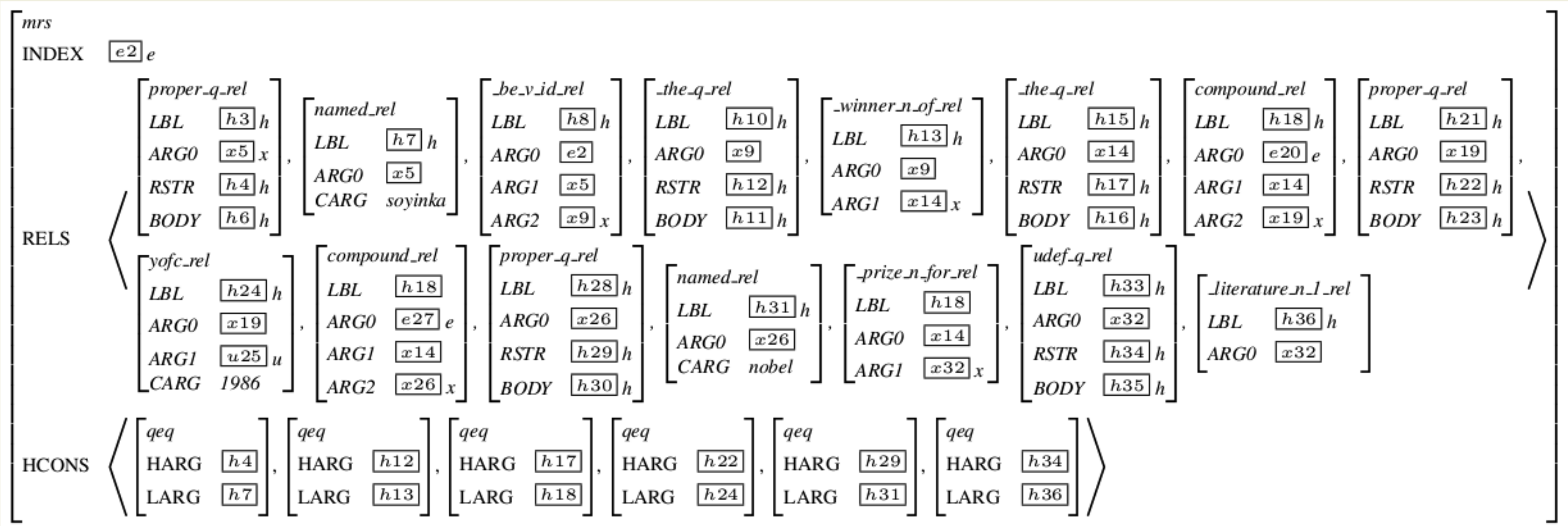


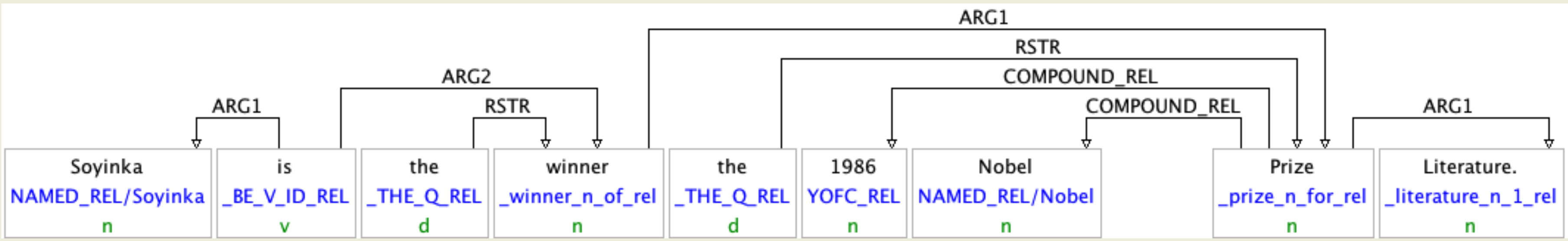
# Converting MRSes to Dependency Structures

- Simple first experiment:
  - conversion of MRS to classical token-to-token dependency structures
  - Conversion on top of the DMRS steps (Copestake, 2008 draft)

- Procedure:
  - Remove unshared variables and labels
  - Resolve handle constraints
  - Incorporate characteristic variables into owning EPs
  - Resolve ‚forwarding nodes‘ (e.g. nominalization_rel)
  - Resolve ‚edge nodes‘ (e.g. compound_rel)
  - Collapse overlapping token nodes

# MRS Dependency Structure Example



# Parsing the Corpus

- Corpus: Nobel corpus as used in our previous relation extraction experiments (cf. http://dare.dfki.de/)
- Newswire text
- Deep Parser: PET + ERG 0907, Tourism texts reading disambiguation model (jhpstg)

| Sentences | Share |
|---|---|
| Parsable | 71.2% |
| Out-of-Grammar | 9.1% |
| Errors (mostly resource limits) | 19.7% |

# Relation Extraction Experiments Setup

- Target Relation:
  *<recipient, award, area, year>*

- Seed:
  *<Ahmed Zewail, 1999, Chemistry, Nobel>*

- Shallow Dependency Parsers:
  – MINIPAR
  – Stanford Parser

- Deep Dependency Parsers:
  – SynDep-Extractor
  – SemDep-Extractor

# Evaluation Method

- Run bootstrapping process on the corpus
- Evaluation of extracted relation instances against the Ideal Table
- Ideal Table (Agichtein and Gravano, 2000): Filter a high-quality compilation of target relations with relation member mentionings in the corpus

# First Results

| Parser | Precision | Recall |
| --- | --- | --- |
| Minipar | 79,57% | 79,09% |
| Stanford | 76,5% | 85,76% |

Shallow Dependencies – All Sentences

| Parser | Precision | Recall |
| --- | --- | --- |
| Minipar | 85,04% | 78,55% |
| Stanford | 79,55% | 81,77% |

Shallow Dependencies – Parsable Sentences

| Parser | Precision | Recall $_{rel}$ | Recall $_{abs}$ |
| --- | --- | --- | --- |
| SynDep | 94,12% | 59,45% | 53,33% |
| SemDep | 88,42% | 56,76% | 50,90% |

Deep Dependencies (Best-Ranked Parse)

# Different Readings

- Ideal world: first reading is the best.
- But: grammar cannot resolve all ambiguities
- So can we exploit the numerous readings that the ERG offers us?
- Naive: always learn with reading *x* and always use reading *x* for relation extraction.

| Reading | Precision | $\text{Recall}_{rel}$ | $\text{Recall}_{abs}$ |
|---|---|---|---|
| SynDep 1 | 94,12% | 59,45% | 53,33% |
| SynDep 2 | 92,47% | 58,11% | 53,12% |
| SynDep 3 | 93,01% | 58,45% | 51,52% |
| SynDep 4 | 91,89% | 57,43% | 51,52% |
| SynDep 5 | 92,19% | 59,80% | 53,64% |

| Reading | Precision | $\text{Recall}_{rel}$ | $\text{Recall}_{abs}$ |
|---|---|---|---|
| SemDep 1 | 88,42% | 56,76% | 50,90% |
| SemDep2 | 90,53% | 58,11% | 52,12% |
| SemDep3 | 91,62% | 59,12% | 53,03% |
| SemDep4 | 90,56% | 55,07% | 49,39% |
| SemDep5 | 86,17% | 54,73% | 49,09% |

# Combining Readings

- Better: combine the sets of learned rules / the sets of extracted relation instances
- Learn rules for all *n* readings and either keep all of them or only those which are present in all readings
- Likewise for extracted relations

| | ∩ Relations Extracted from r. 1 – n | ∪ Relations Extracted from r. 1 – n |
|---|---|---|
| ∩ Rules, Learned from r. 1 – n | | |
| ∪ Rules, Learned from r. 1 – n | | |

# Results

| Rule Set | Reading | Number Rules | Prec. | ∩ Rel. $Recall_{rel}$ | $Recall_{abs}$ | Prec. | ∪ Rel. $Recall_{rel}$ | $Recall_{abs}$ |
|---|---|---|---|---|---|---|---|---|
| ∩ SynDep | 1-30 | 30 | 98,65% | 24,66% | 22,12% | 91,98% | 58,11% | 52,12% |
| ∩ SemDep | 1-30 | 20 | 97,26% | 23,99% | 21,52% | 88,42% | 56,76% | 50,91% |
| ∪ SynDep | **1-30** | **255** | **98,33%** | **39,86%** | **35,76%** | **82,97%** | **79,05%** | **70,91%** |
| ∪ SemDep | 1-30 | 254 | 89,17% | 36,14% | 32,42% | 76,97% | 82,43% | 73,94% |

# Conclusions

- DARE is the first approach to combine the idea of bootstrapping IE systems with a linguistic grammar

- This can be illustrated by a simple formula:

  reusable generic linguistic knowledge
  \+ raw data
  \+ a few examples (seed)
  
  ---
  = domain specific relation extraction grammar

- Deep DARE is a first and very promising step towards utilizing the wealth of information available in linguistic precision grammars in minimally supervised information extraction

# Conclusion

- We provide a flexible configuration scheme for trading of precision and recall in an RE grammar.
- We have achieved an automatic learning of complex semantic structures to a real-world interpretation for a target relation.
- We can confirm that relation extraction is a very suitable task for evaluating parsers in situ.

# Ongoing Research Work

- Learning DARE rules directly on top of RMRSs
- Challenges
  - Handle graphs instead of trees
  - Underspecifications in