Relation Extraction
Part 2

Feiyu Xu

Machine Learning for Relation Extraction

Motivations of ML

- Porting to new domains or applications is expensive
- Current technology requires IE experts
  - Expertise difficult to find on the market
  - SME cannot afford IE experts
- Machine learning approaches
  - Domain portability is relatively straightforward
  - System expertise is not required for customization
  - “Data driven” rule acquisition ensures full coverage of examples

Problems

- Training data may not exist, and may be very expensive to acquire
- Large volume of training data may be required
- Changes to specifications may require reannotation of large quantities of training data
- Understanding and control of a domain adaptive system is not always easy for non-experts

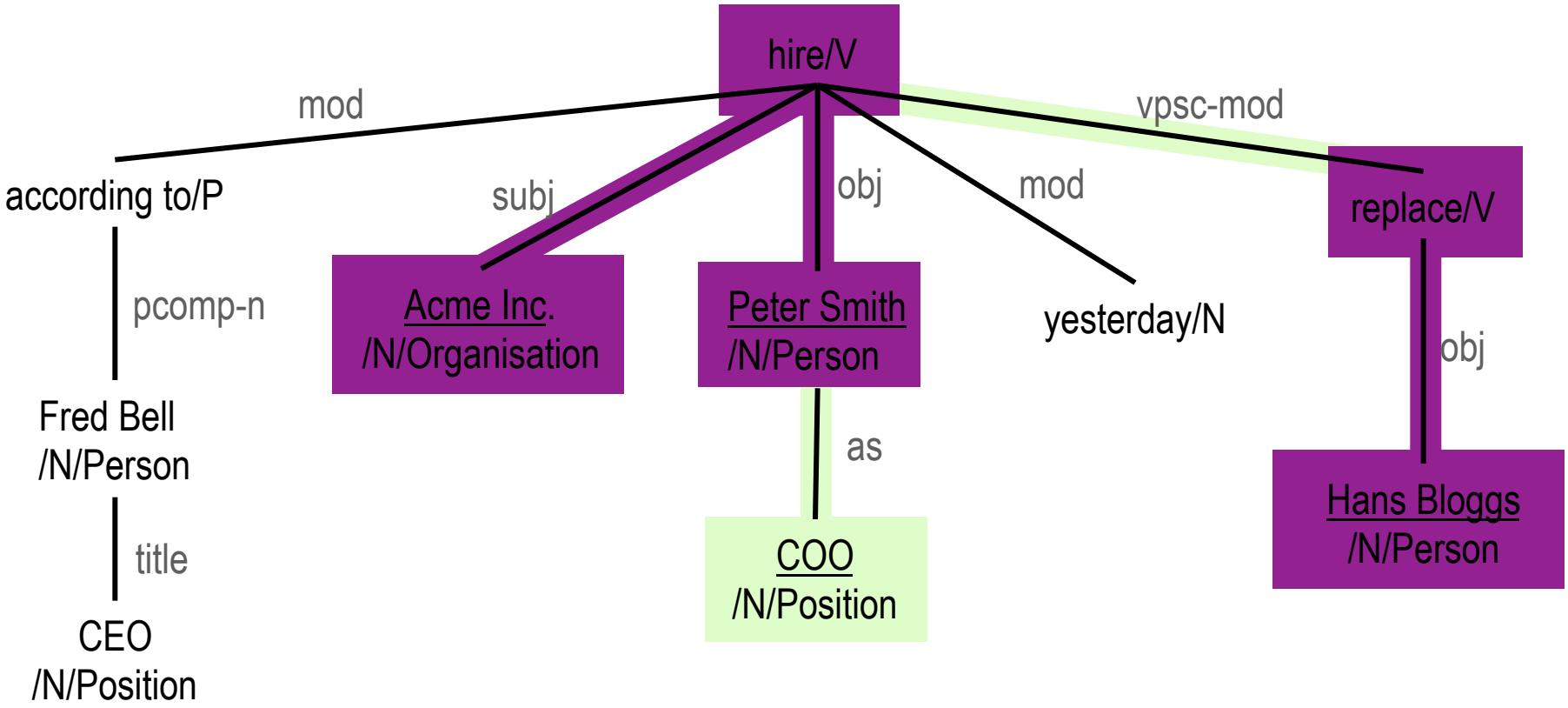
Parameters of IE Real-World Tasks

- Document structure
  - Free text
  - Semi-structured
  - Structured
- Linguistic annotation
  - Shallow NLP
  - Deep NLP
- Complexity and specificity of relation
  - Unary
  - N-ary
- Depth of extraction
  - Recognition
  - Classification
  - Semantic role labelling
- Degree of automation
  - Semi-automatic
  - Supervised
  - Semi-Supervised
  - Minimally-Supervised
  - Distant Supervision
  - Unsupervised
- Human interaction/contribution
- Data properties
  - Domain relevance
  - Redundancy
  - Connectivity
- Evaluation/validation
  - With/without gold standard
  - Performance: recall & precision
  - Interaction among parameters

State of the Art

Binary Relation Only Approaches

- Extraction of binary relations only such as
  - *author-book*
  - *company-location*
- Do not employ the existing syntactic and semantic structures among n > 2 arguments and rely on a later component to merge binary relations into complex relations
- Approaches
  - Ravichandran and Hovy, 2002
  - Pantel et al., 2004
  - Pasca et al., 2006a; Pasca et al., 2006b


hire/V
mod
vpsc-mod
according to/P
subj
obj
mod
replace/V
pcomp-n
Acme Inc.
/N/Organisation
Peter Smith
/N/Person
yesterday/N
obj
Fred Bell
/N/Person
as
Hans Bloggs
/N/Person
title
COO
/N/Position
CEO
/N/Position

Surface-oriented Rule Representation

- Shallow linguistic analyses
- These formalisms are robust and efficient but only handle binary relations.
- Work best for relations whose arguments usually co-occur in close proximity within a sentence and whose mentions exhibit limited linguistic variation
- Approaches
  - Pasca et al., 2006a; Pasca et al.,2006b;
  - Kozareva et al., 2008; Hovy et al., 2009; Kozareva and Hovy, 2010

Minimally Supervised (Bootstrapping)

- Based on iterative learning with limited initial knowledge
  - Start a small number of initial examples of relation instances (or patterns)
  - Label the free texts during iterations (e.g., Agichtein and Gravano, 2000; Yangarber et al.,2000; Ravichandran and Hovy, 2002; Stevenson and Greenwood, 2005).
- Often suffer from semantic drift or the propagation of errors occurring during iterations
- Performance depends on data properties

Distant Supervision

- A massive seed-based, one step version of bootstrapping
- Rely on a large amount of trustworthy facts
- Their performance does not hinge on corpus data properties such as redundancy
- Approaches
  - (Mintz et al., 2009)
  - Others: (Wu and Weld, 2007); (Wu et al., 2008); (Weld et al., 2008); (Hoffmann et al., 2010); (Xu et al., 2011(; (Nguyen and Moschitti, 2011)

Open IE

- They do not target given relations.
- They are very useful for applications continuously faced with new relation or event types, e.g., online social media monitoring
- However, the results of these systems cannot be directly taken for filling knowledge databases, because the semantics of the new relations including the roles of the entities remains unknown.
- Example: TextRunner (Banko et al., 2007; Yates et al., 2007)

Reality in IE Projects

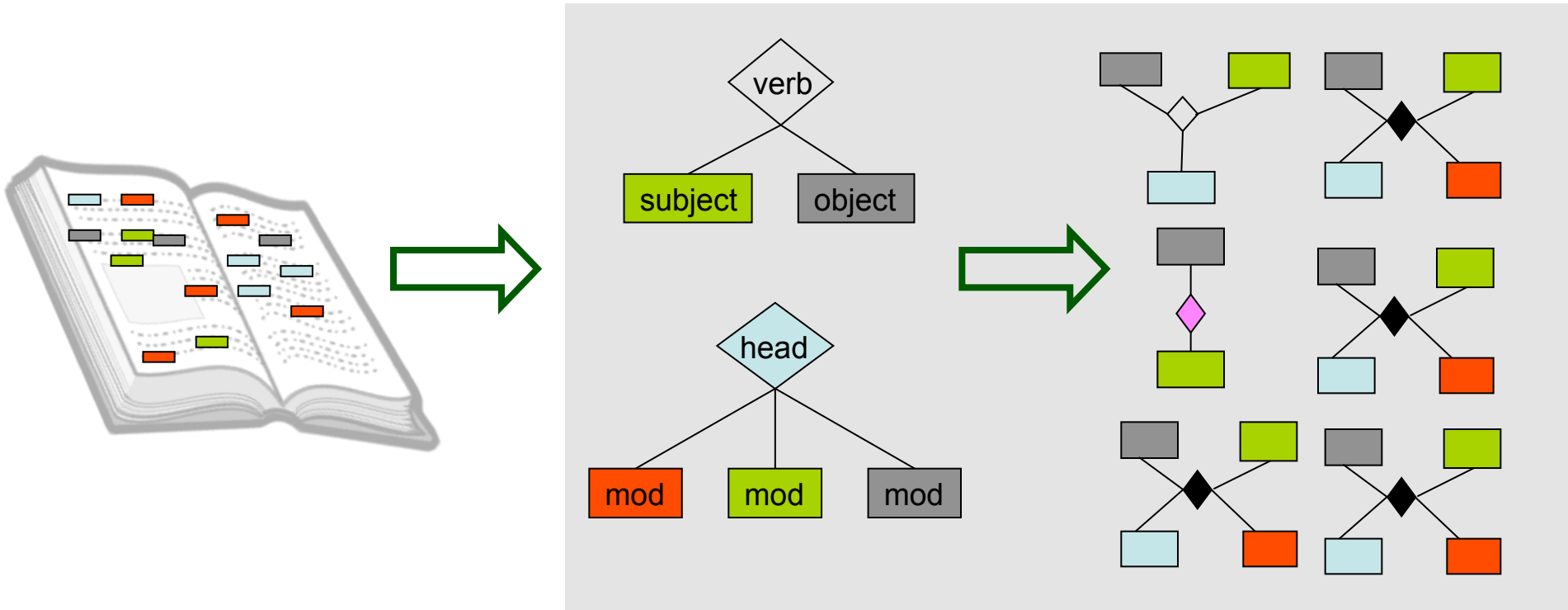
- Our IE users are often not domain experts
- IE experts have to develop methods and strategies for
  - Prospecting a domain
  - Proposing relevant relations
  - Finding relevant and suitable data

DARE:
Domain Adaptive Relation Extraction

http://dare.dfki.de

Research Goal

Development of a general framework for automatically learning mappings between linguistic analyses and target semantic relations, with minimal human intervention.



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

Example

A relation extraction task in the domain *management succession* (MUC-6)

< person_in, person_out, position, organisation>

- *person_in*: the person who obtained the position
- *person_out*: the person who left the position
- *position*: the job position that the two persons were involved in
- *organisation*: the organisation where the position was located

According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

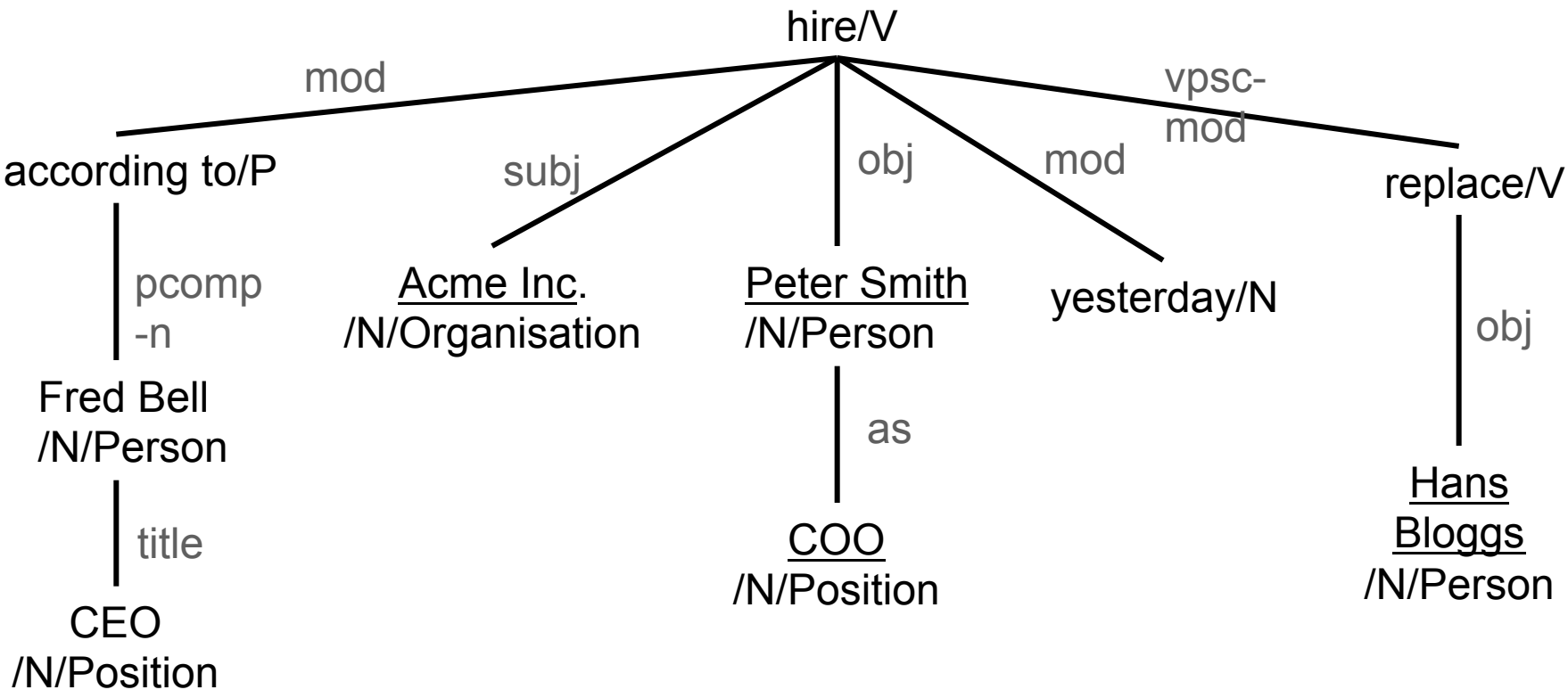
<person_in, person_out, position, organisation>

According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

<person_in, person_out, position, organisation>

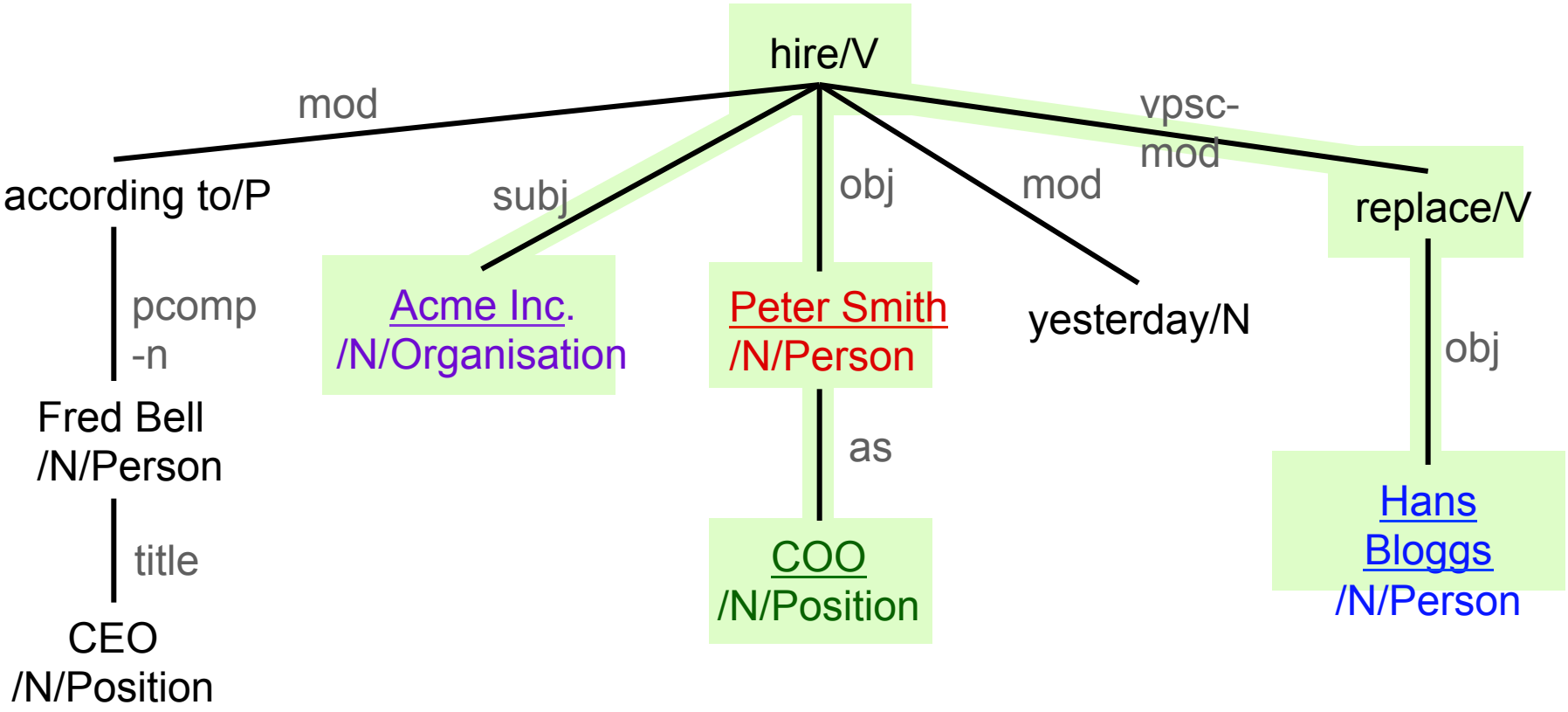
According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

<person_in, person_out, position, organisation>

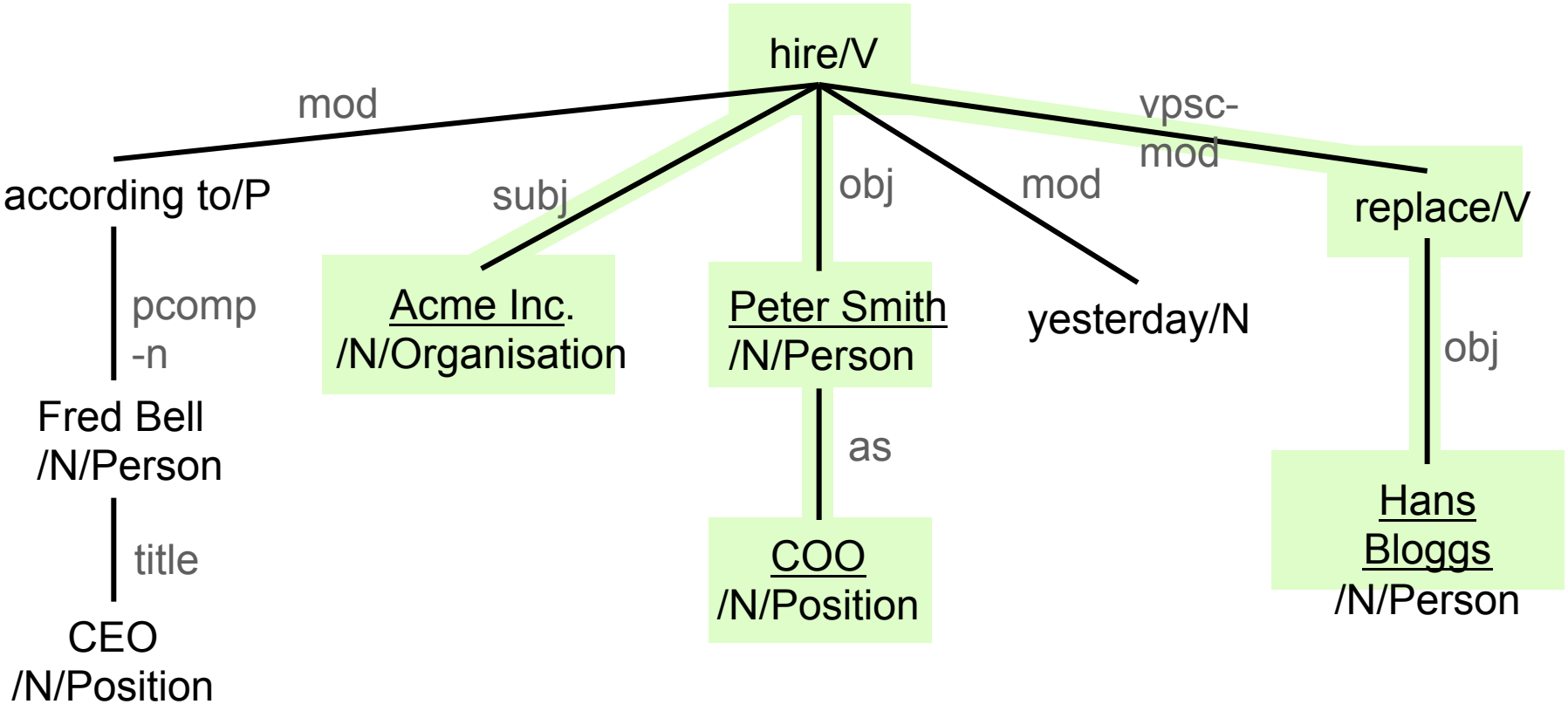


According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

<person_in, person_out, position, organisation>



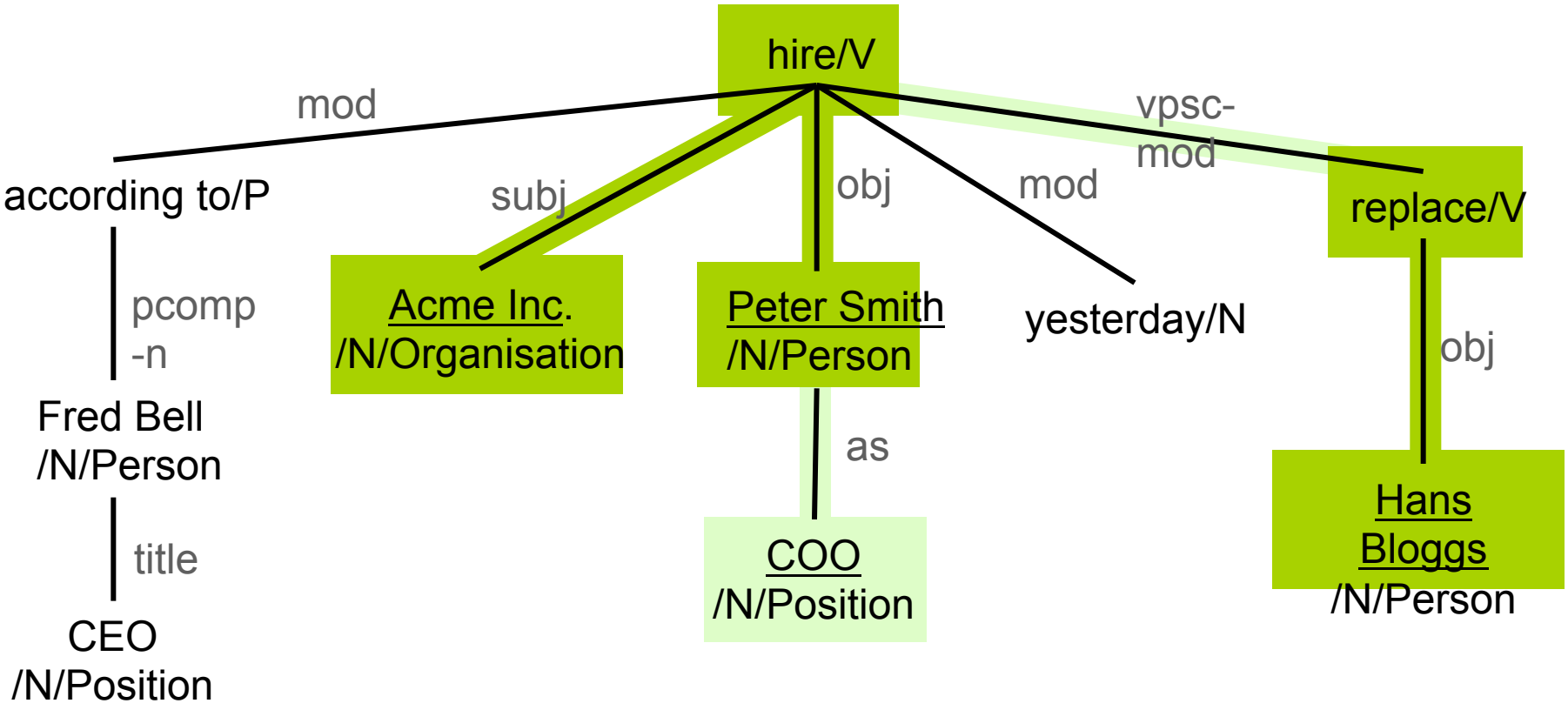
Ideal Target Pattern



Previous Work: SVO Model

Yangarber (2001)

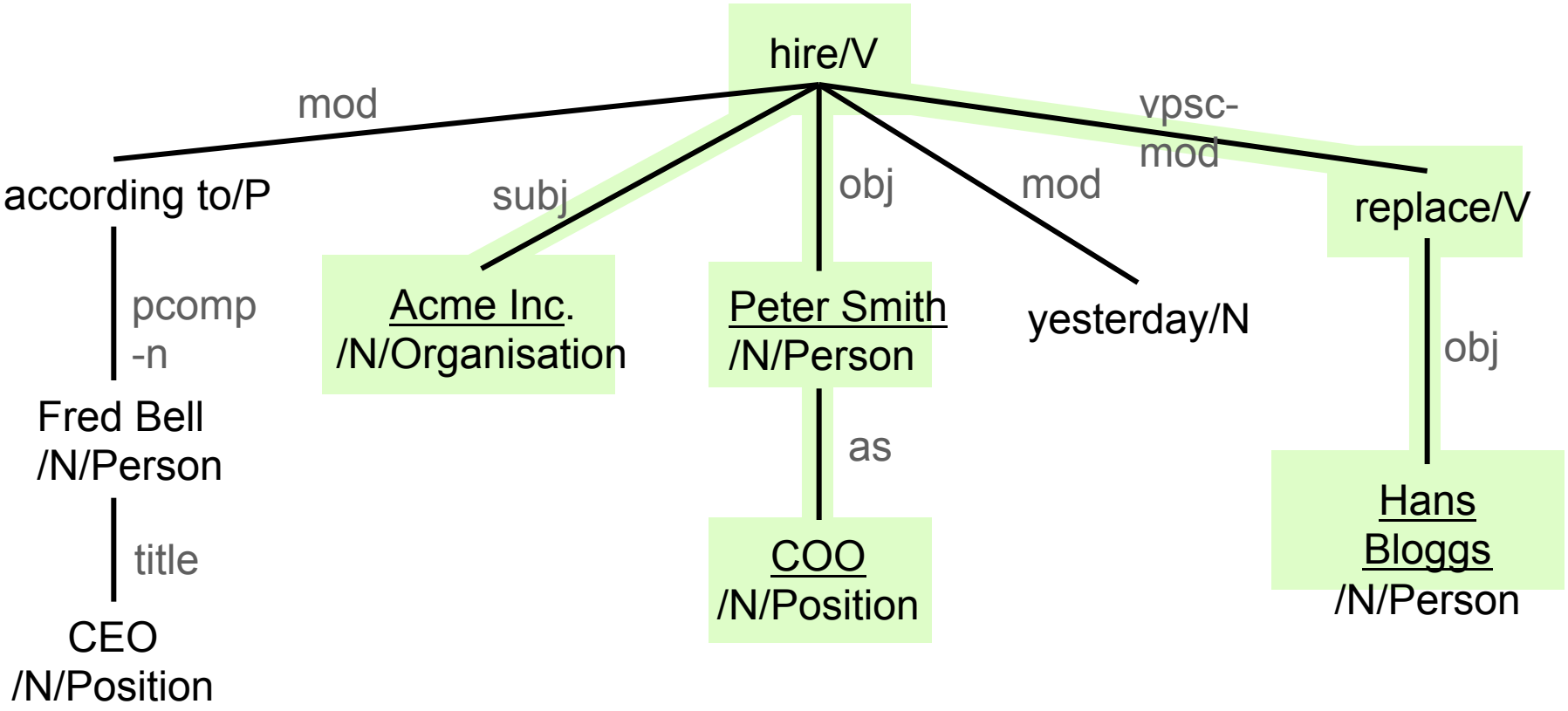
- Verb centered
- Direct relations between subject-verb-object
- Complex NP can not be extracted, e.g., the person and position relation
- The linguistic relations among patterns are not considered, e.g., hire and replace



Previous Work: Chain Model

Sudo et al. (2001)

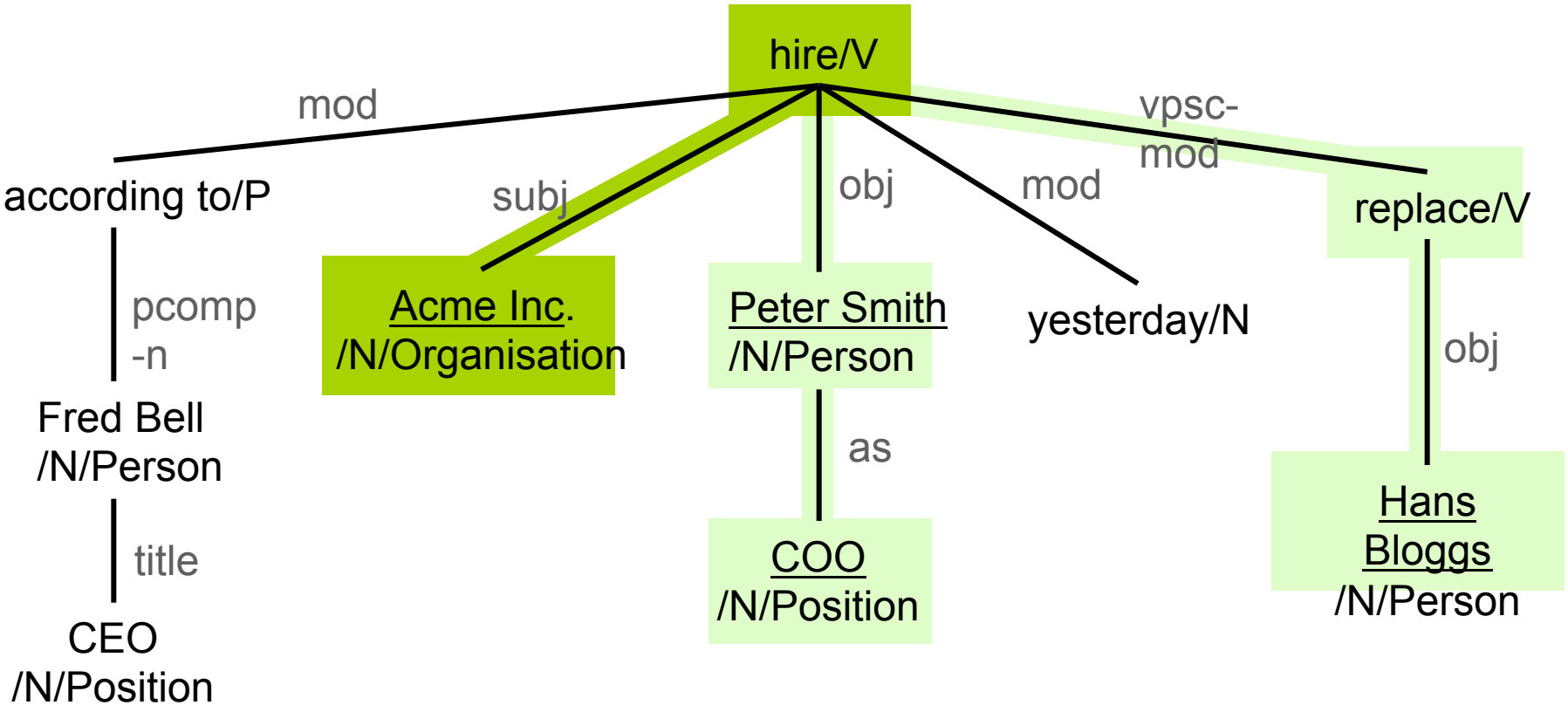
- Verb centered
- A single syntactic path dominated by a verb containing at least one relevant named entity concept



Previous Work: Chain Model

Sudo et al. (2001)

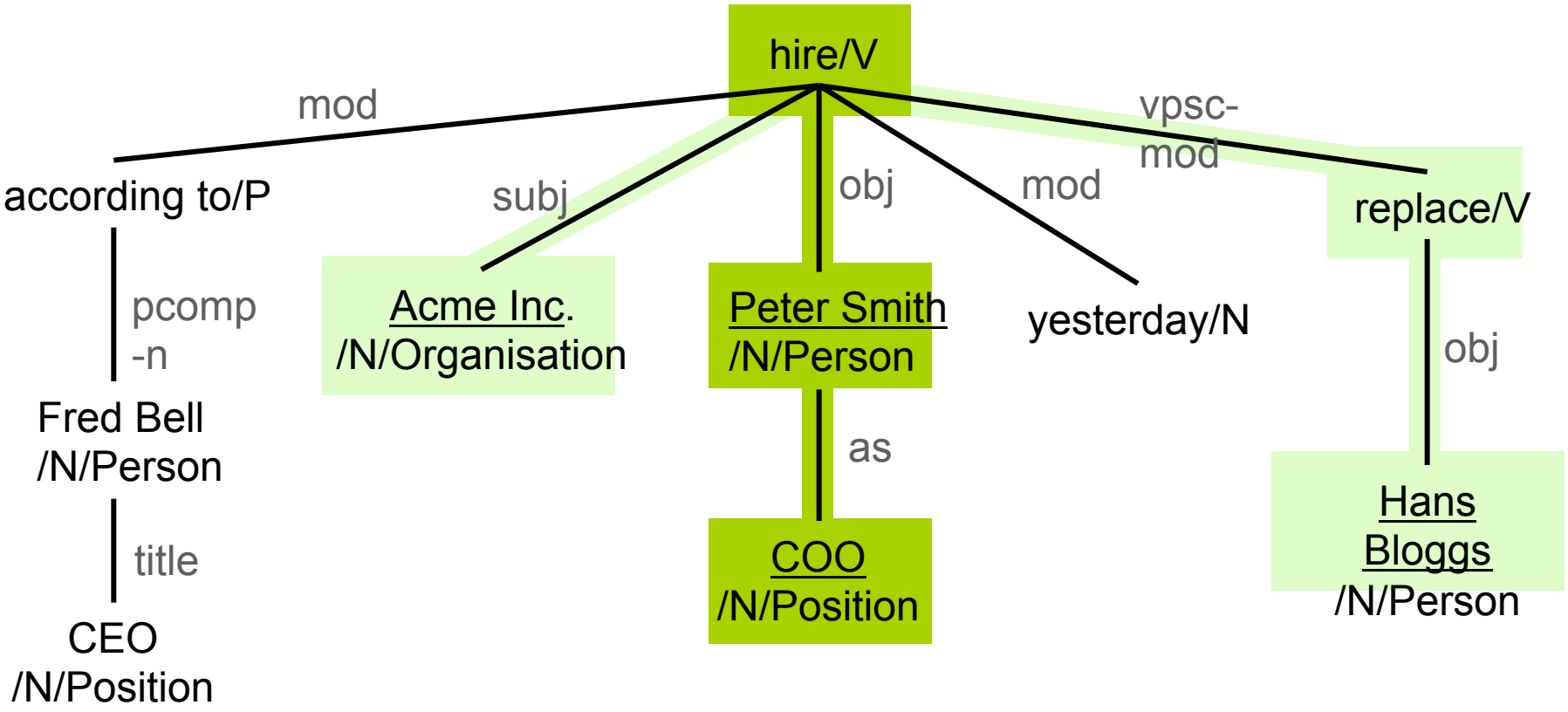
- Verb centered
- A single syntactic path dominated by a verb containing at least one relevant named entity concept



Previous Work: Chain Model

Sudo et al. (2001)

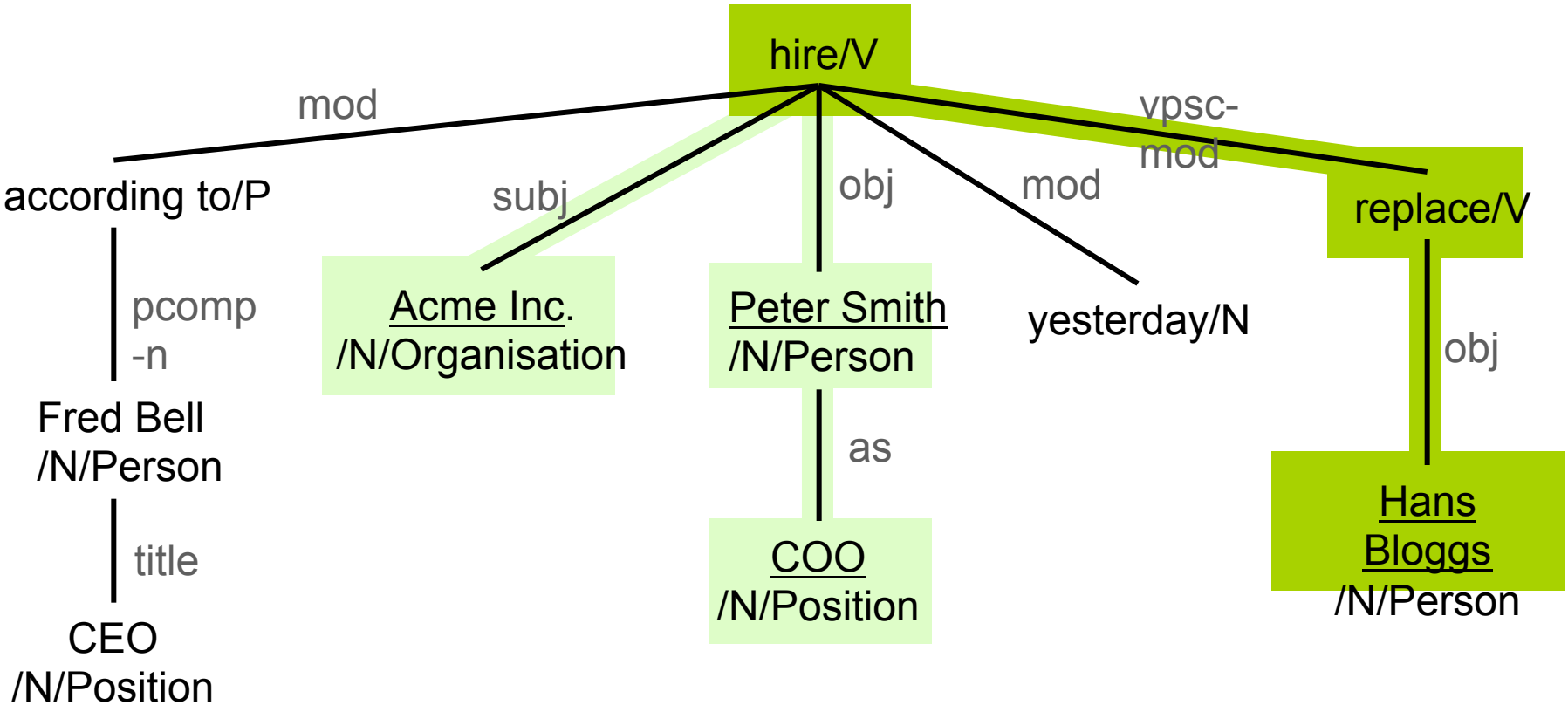
- Verb centered
- A single syntactic path dominated by a verb containing at least one relevant named entity concept



Previous Work: Chain Model

Sudo et al. (2001)

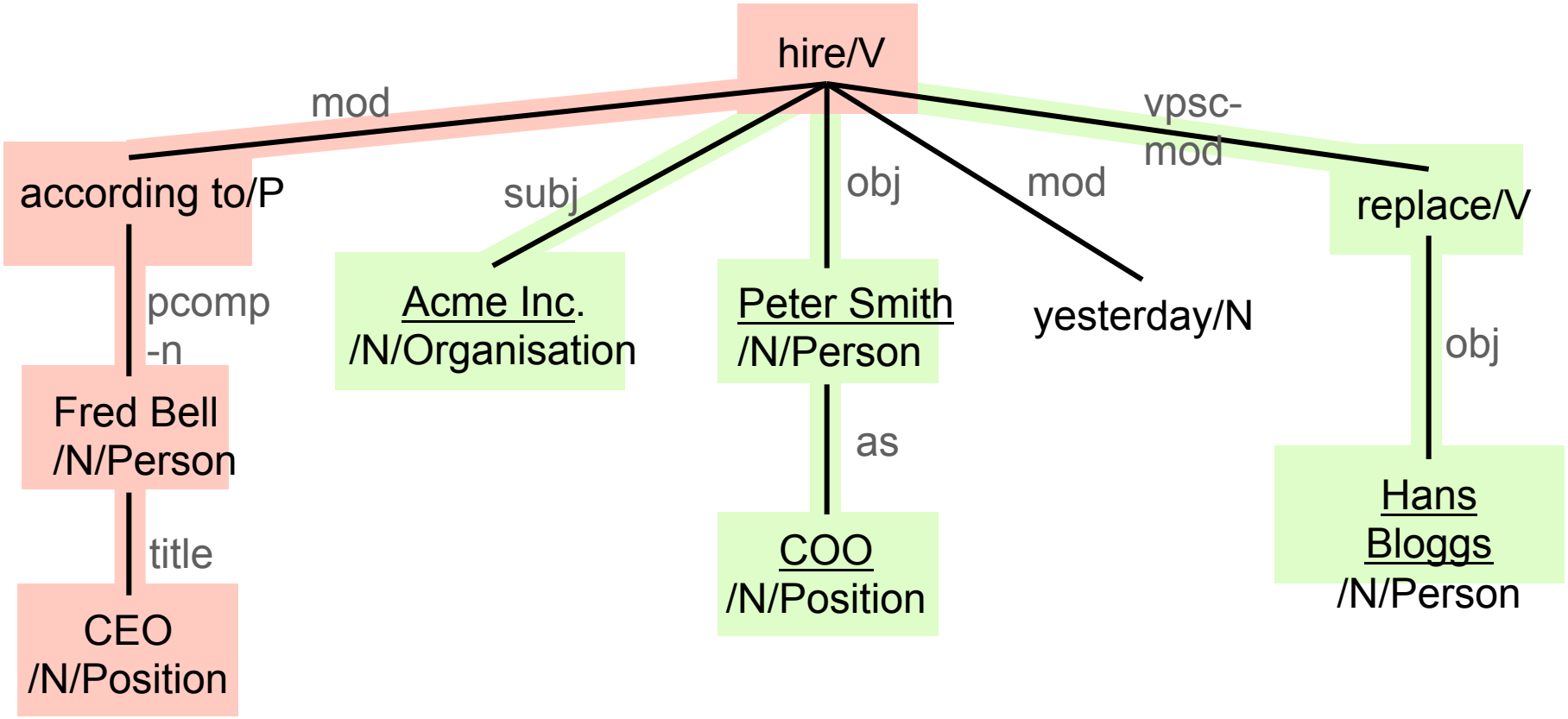
- Verb centered
- A single syntactic path dominated by a verb containing at least one relevant named entity concept



Previous Work: Chain Model

Sudo et al. (2001)

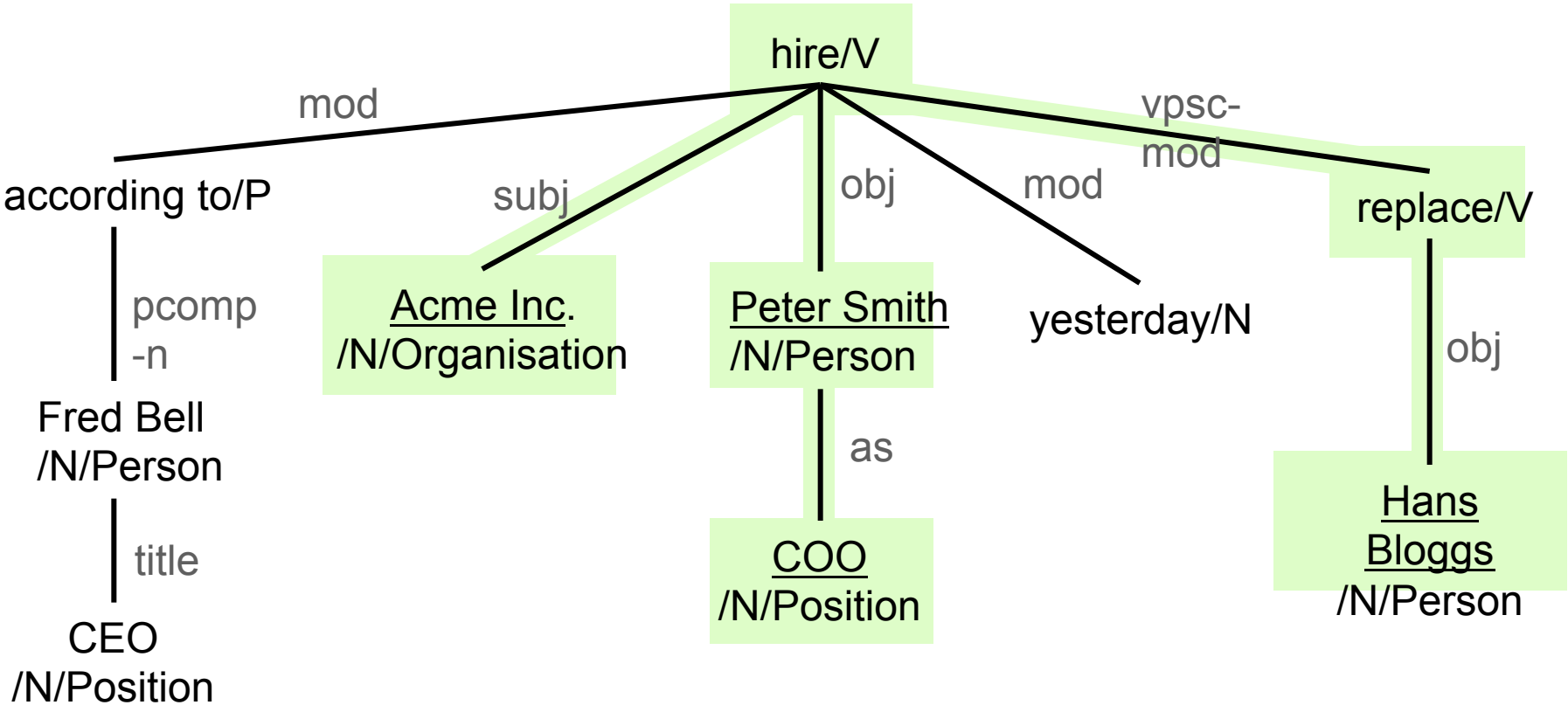
- Verb centered
- A single syntactic path dominated by a verb containing at least one relevant named entity concept



Previous Work: Linked Chain Model

Stevenson & Greenwood 2005

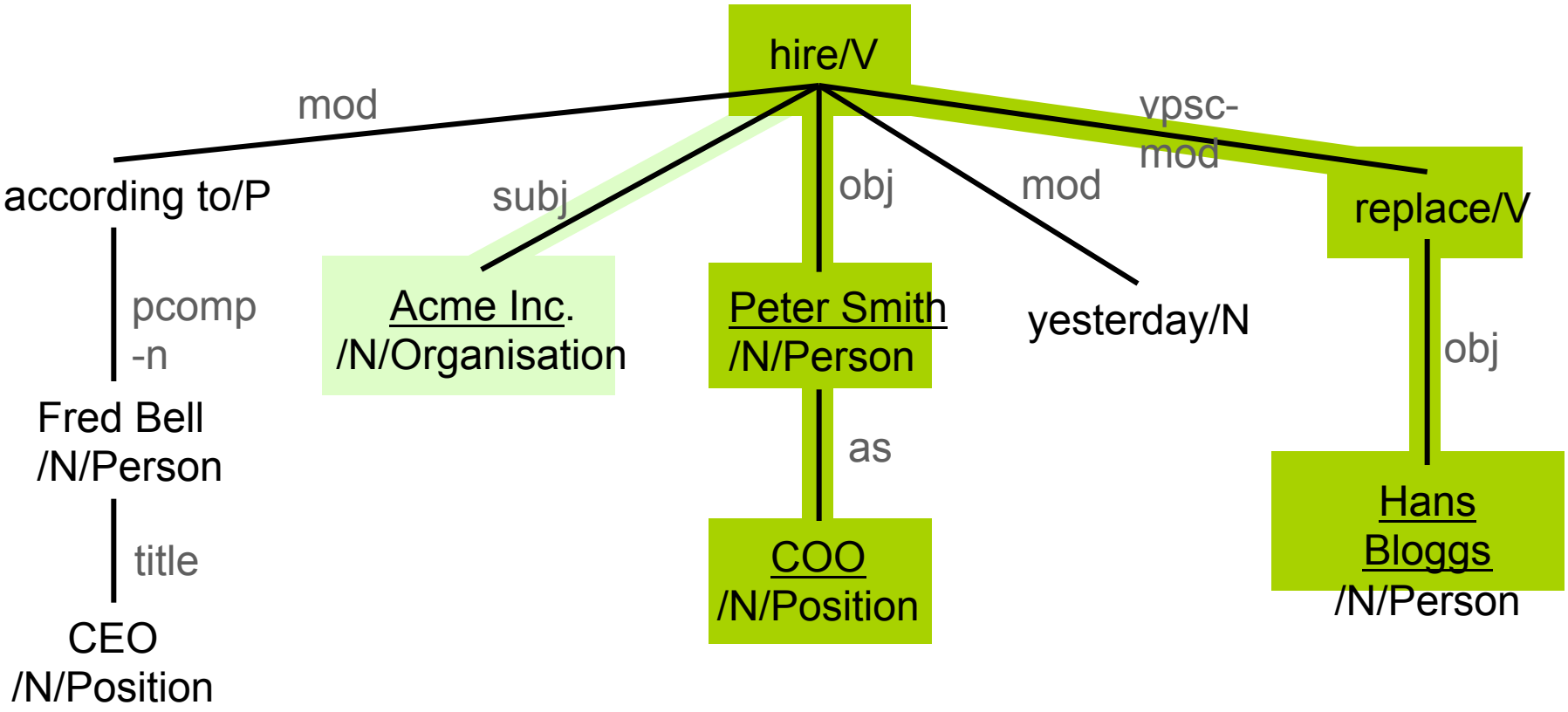
- verb centered
- pairs of chains instead of single paths



Previous Work: Linked Chain Model

Stevenson & Greenwood 2005

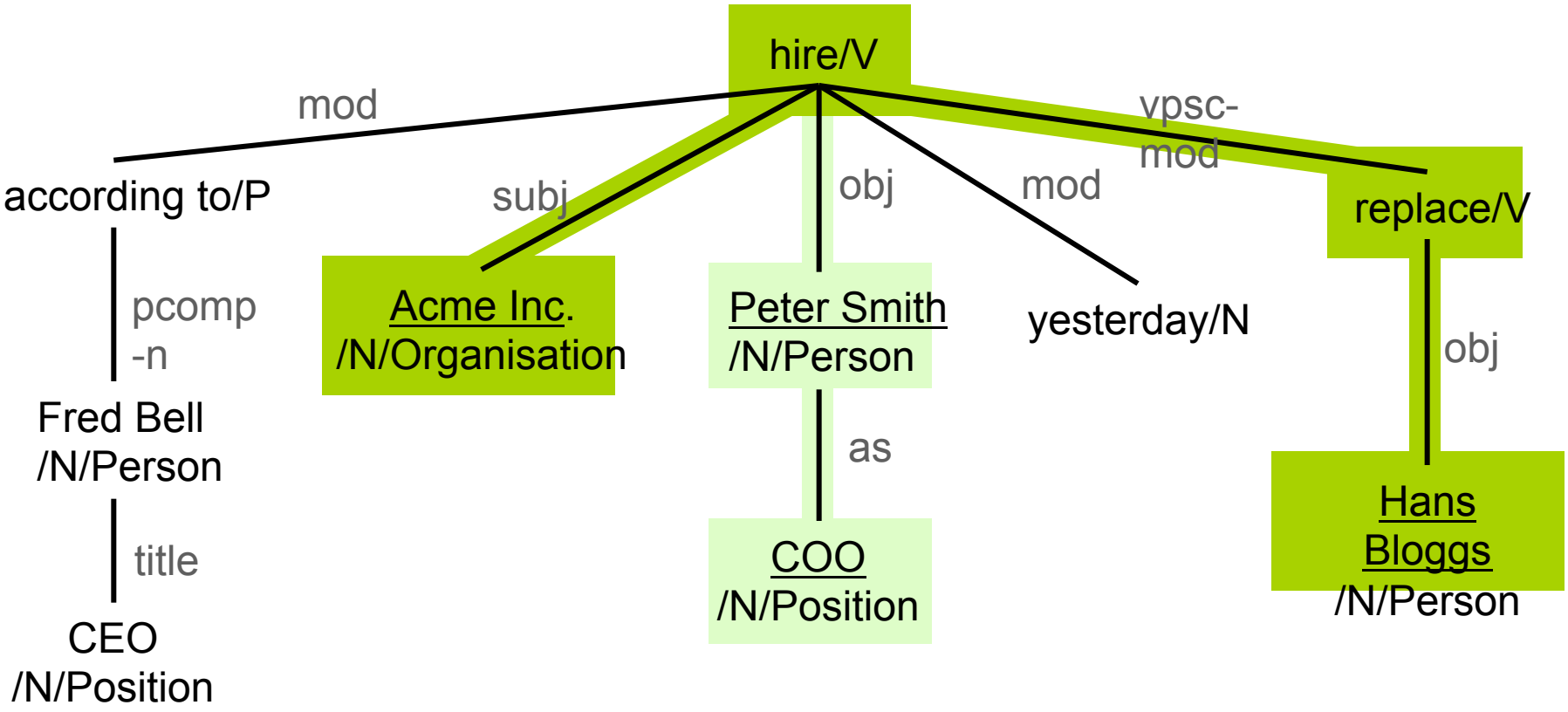
- verb centered
- pairs of chains instead of single paths



Previous Work: Linked Chain Model

Stevenson & Greenwood 2005

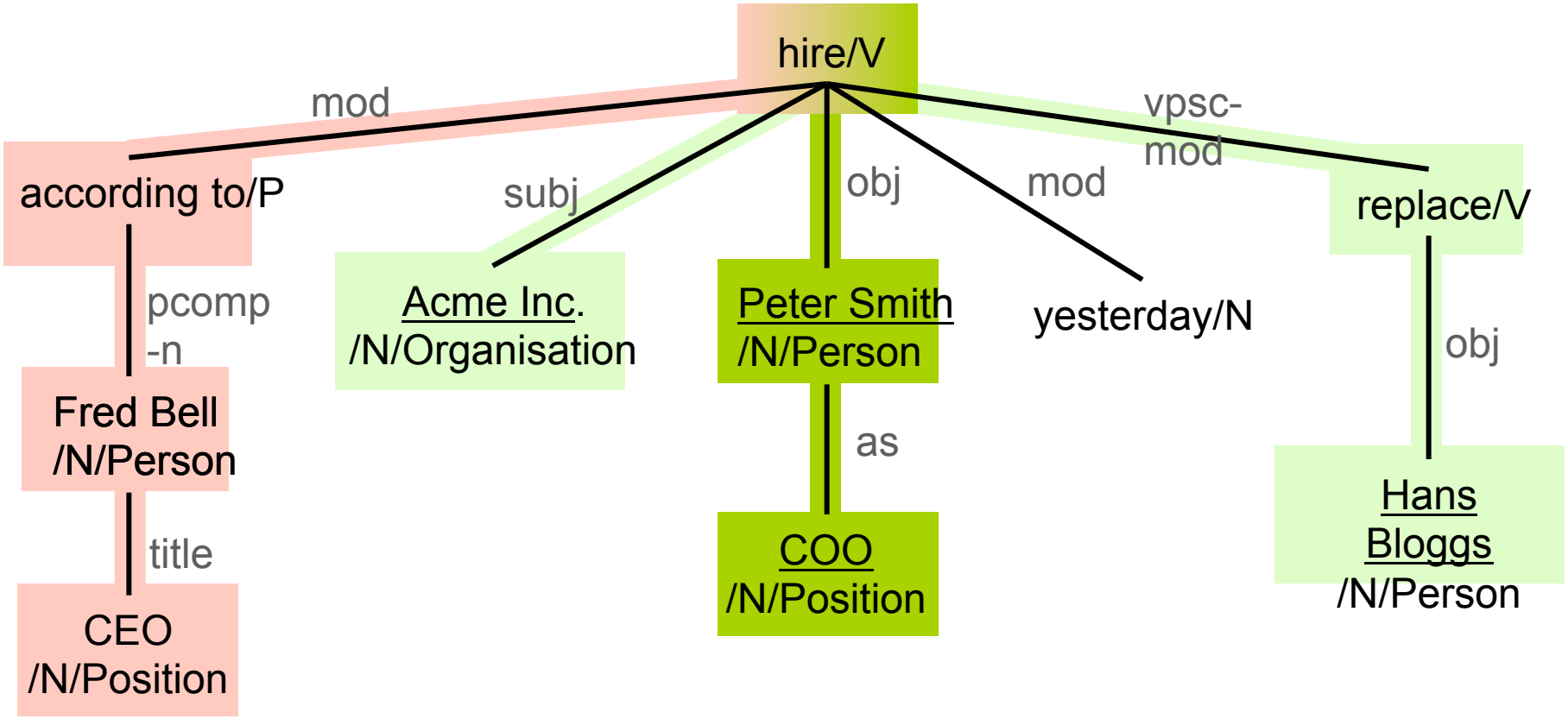
- verb centered
- pairs of chains instead of single paths



Previous Work: Linked Chain Model

Stevenson & Greenwood 2005

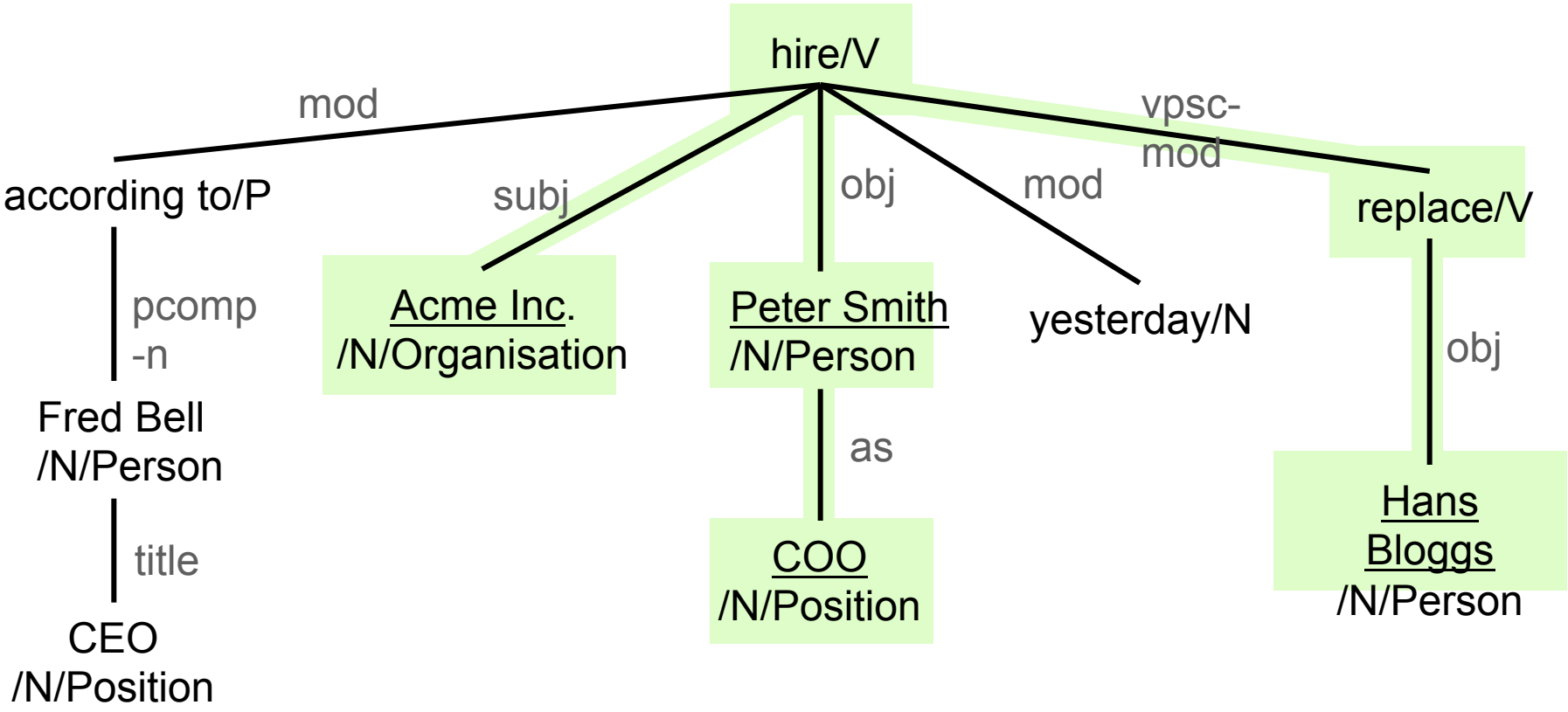
- verb centered
- pairs of chains instead of single paths



Previous Work: Subtree-Model

Sudo et al. (2003)

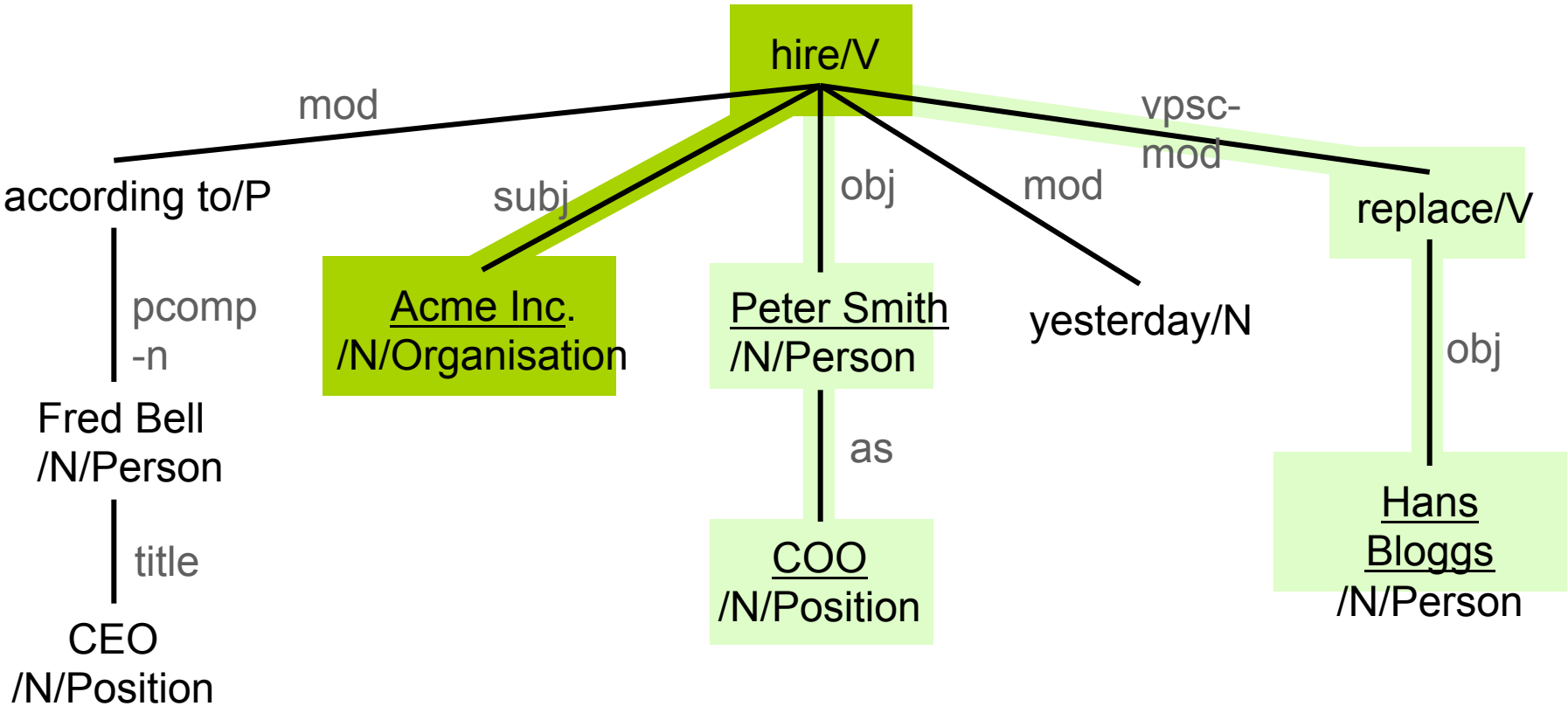
- verb centered
- All chains dominated by a verb, which contain at least one relevant named entity and their combinations



Previous Work: Subtree-Model

Sudo et al. (2003)

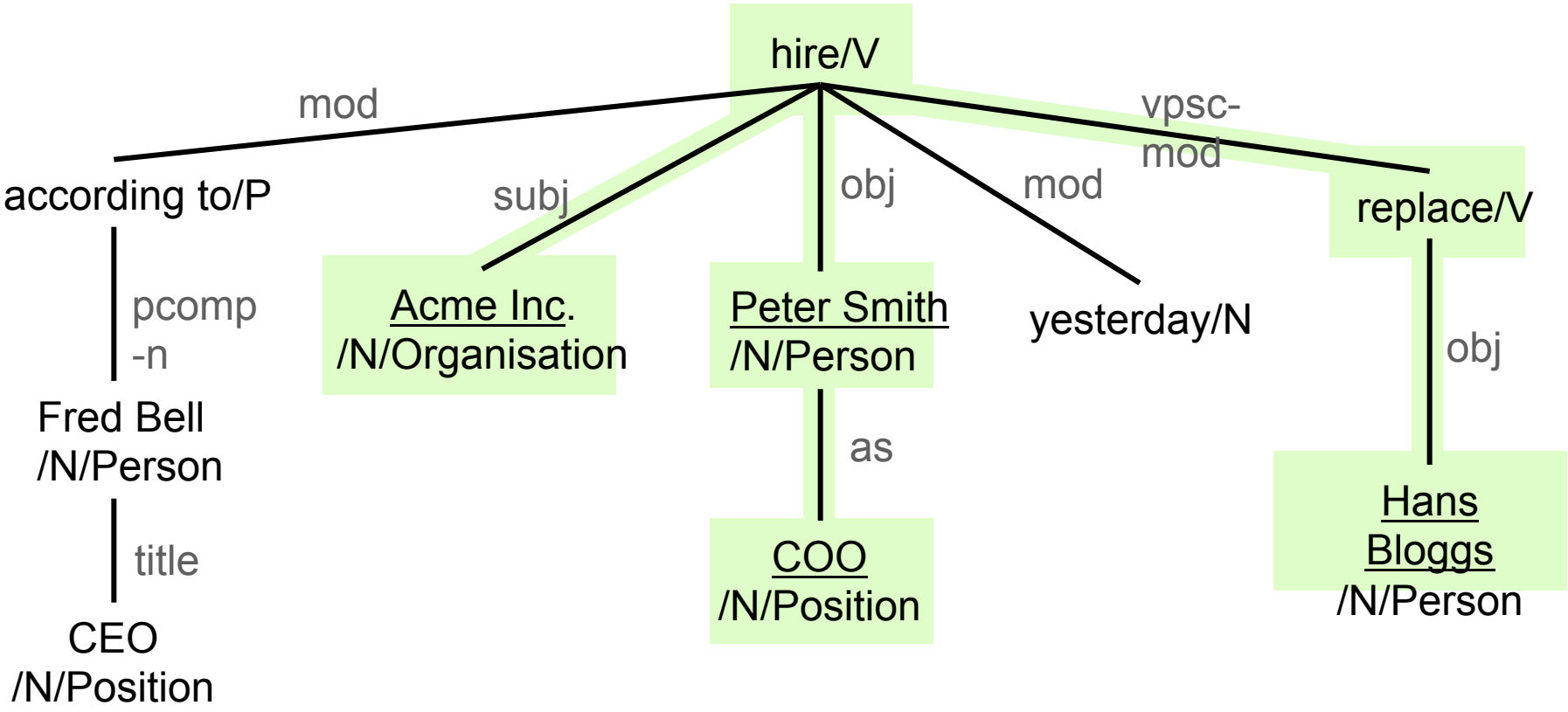
- verb centered
- All chains dominated by a verb, which contain at least one relevant named entity and their combinations



Previous Work: Subtree-Model

Sudo et al. (2003)

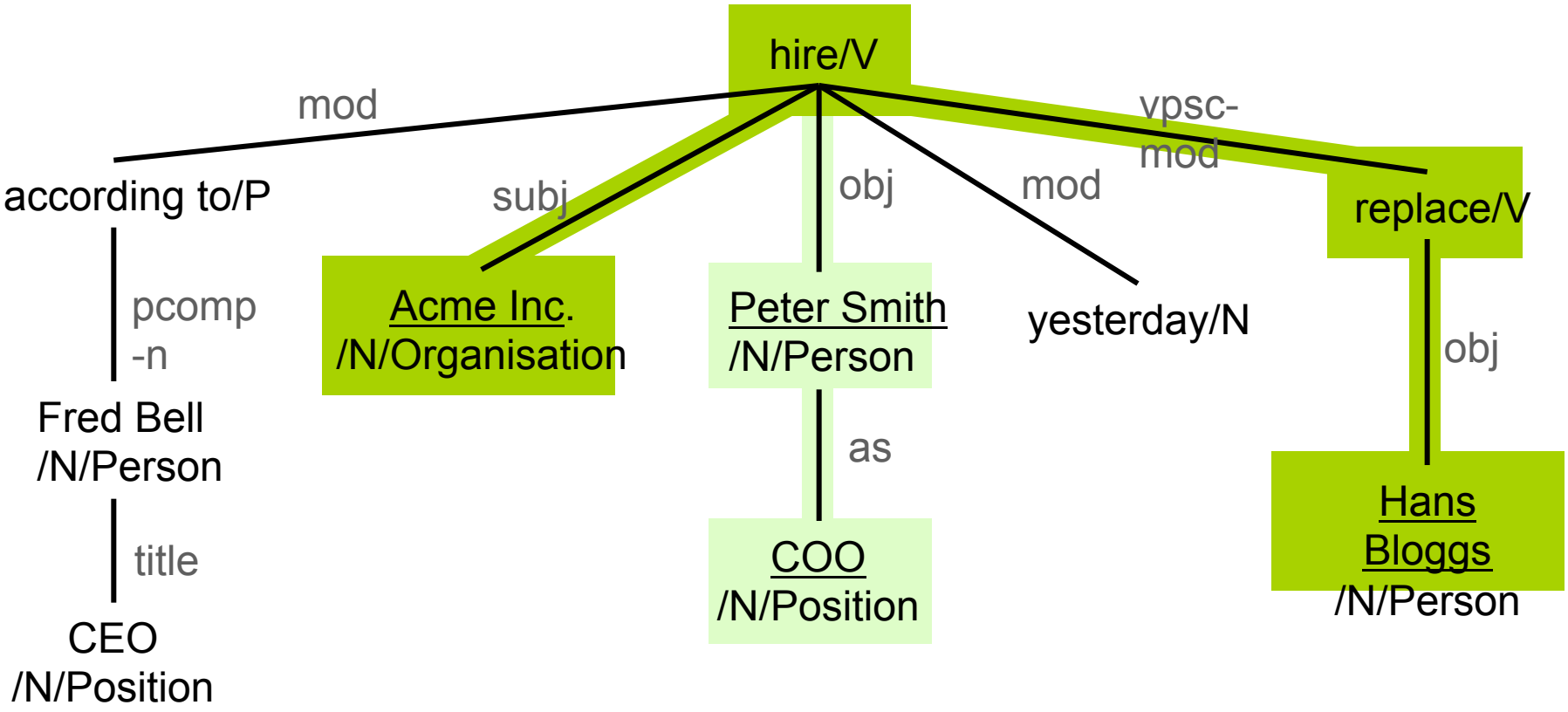
- verb centered
- All chains dominated by a verb, which contain at least one relevant named entity and their combinations



Previous Work: Subtree-Model

Sudo et al. (2003)

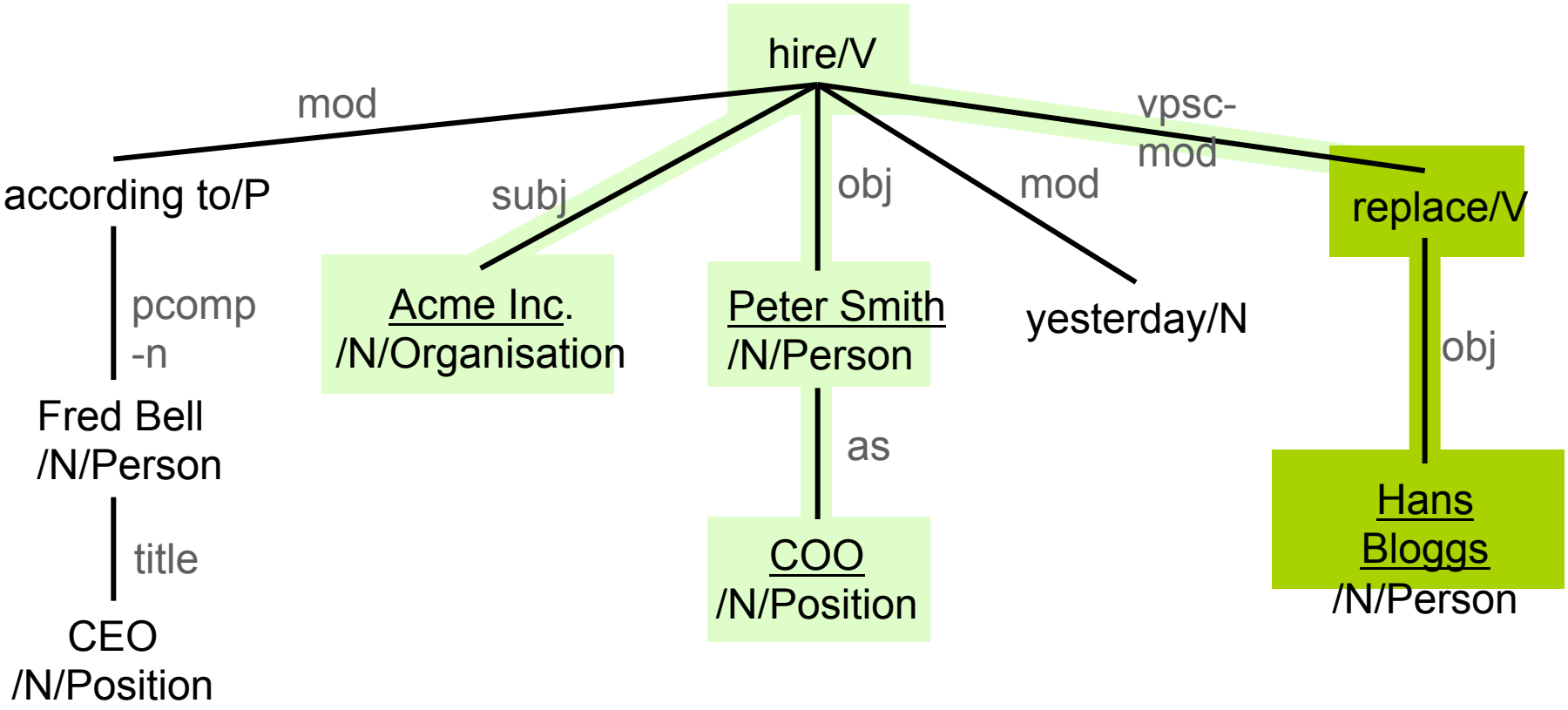
- verb centered
- All chains dominated by a verb, which contain at least one relevant named entity and their combinations



Previous Work: Subtree-Model

Sudo et al. (2003)

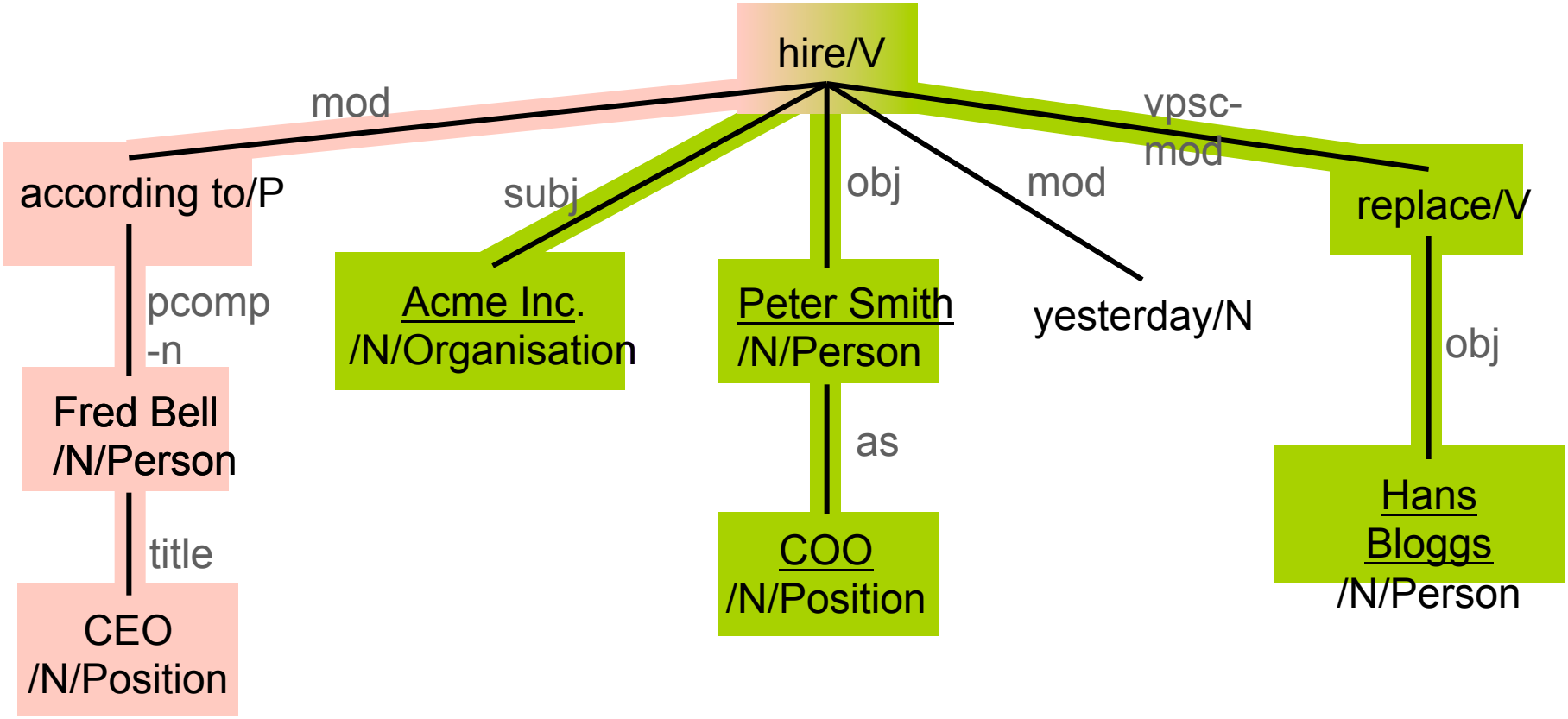
- verb centered
- All chains dominated by a verb, which contain at least one relevant named entity and their combinations



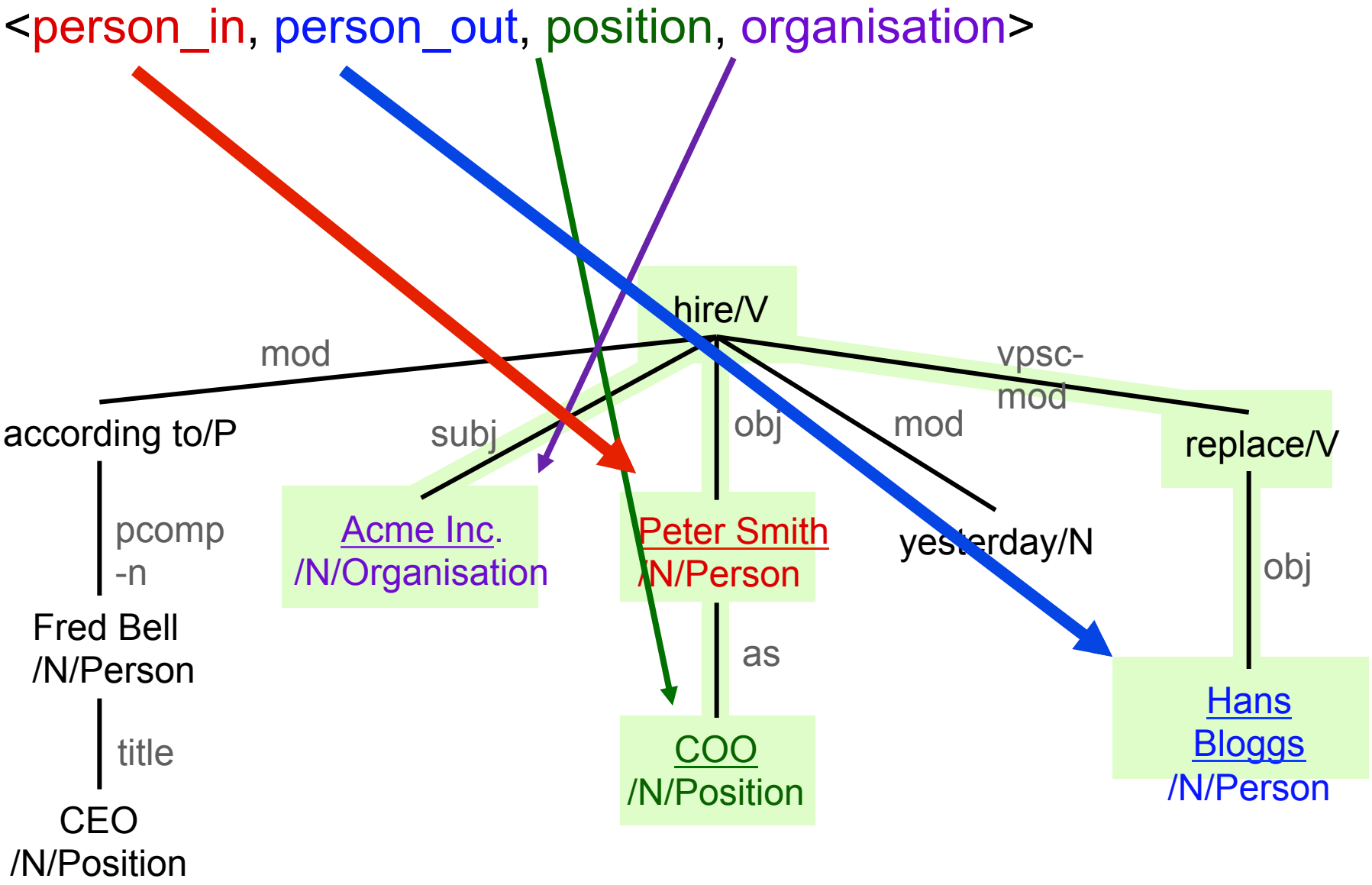
Previous Work: Subtree-Model

Sudo et al. (2003)

- verb centered
- All chains dominated by a verb, which contain at least one relevant named entity and their combinations



None of the existing models links the detected slot-filling candidates with their respective semantic roles



Properties of DARE

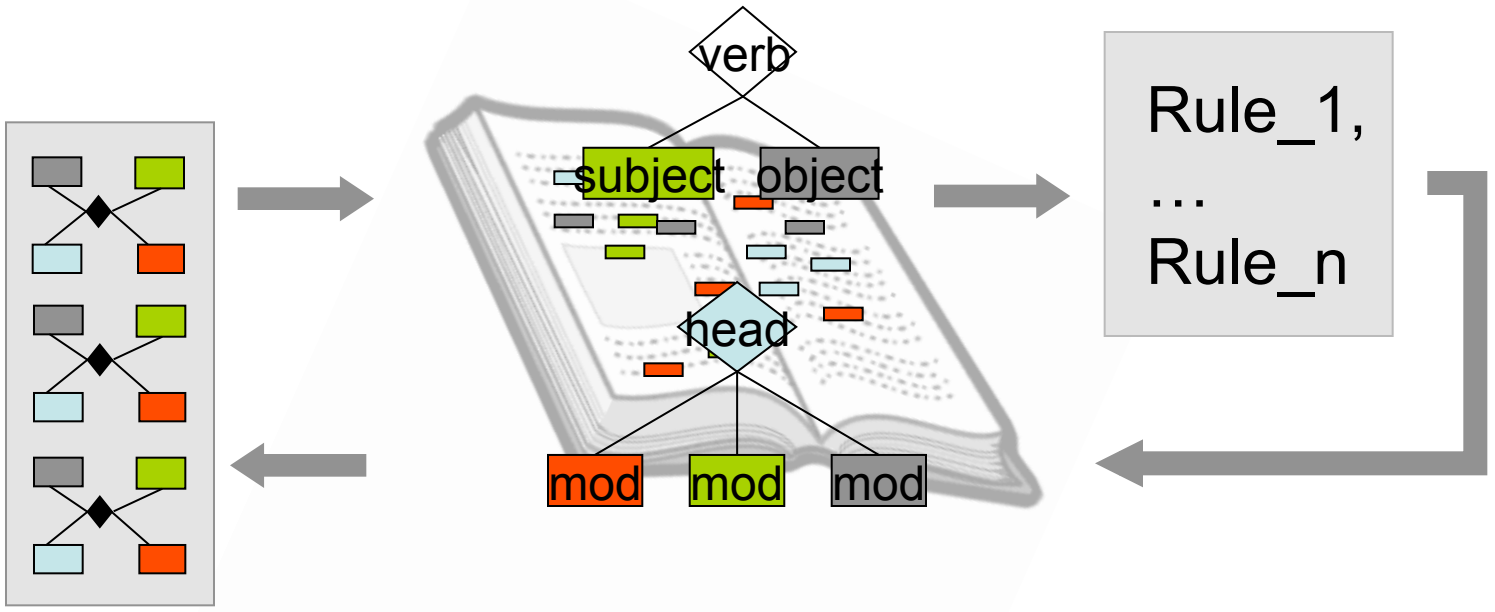
- Samples of target relation instances serve as semantic seed
- Systematic treatment of n-ary relations and their projections
- Exploitation of relation projections for pattern discovery
- Bottom-up compositional pattern discovery
- A recursive linguistic rule representation
- Rules contain semantic roles w.r.t. to target relation
- Bottom-up compression method to generalize rules
- Filtering of rule candidates by “domain relevance”

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
- Bottom-up compression method to generalize rules
- Filtering of rule candidates by “domain relevance”

Bootstrapping Relation Extraction with Semantic Seed

Adapted from
DIPRE (Brin, 1998) and Snowball (Agichtein & Gravano, 2000)
but extended and enriched with linguistic analysis

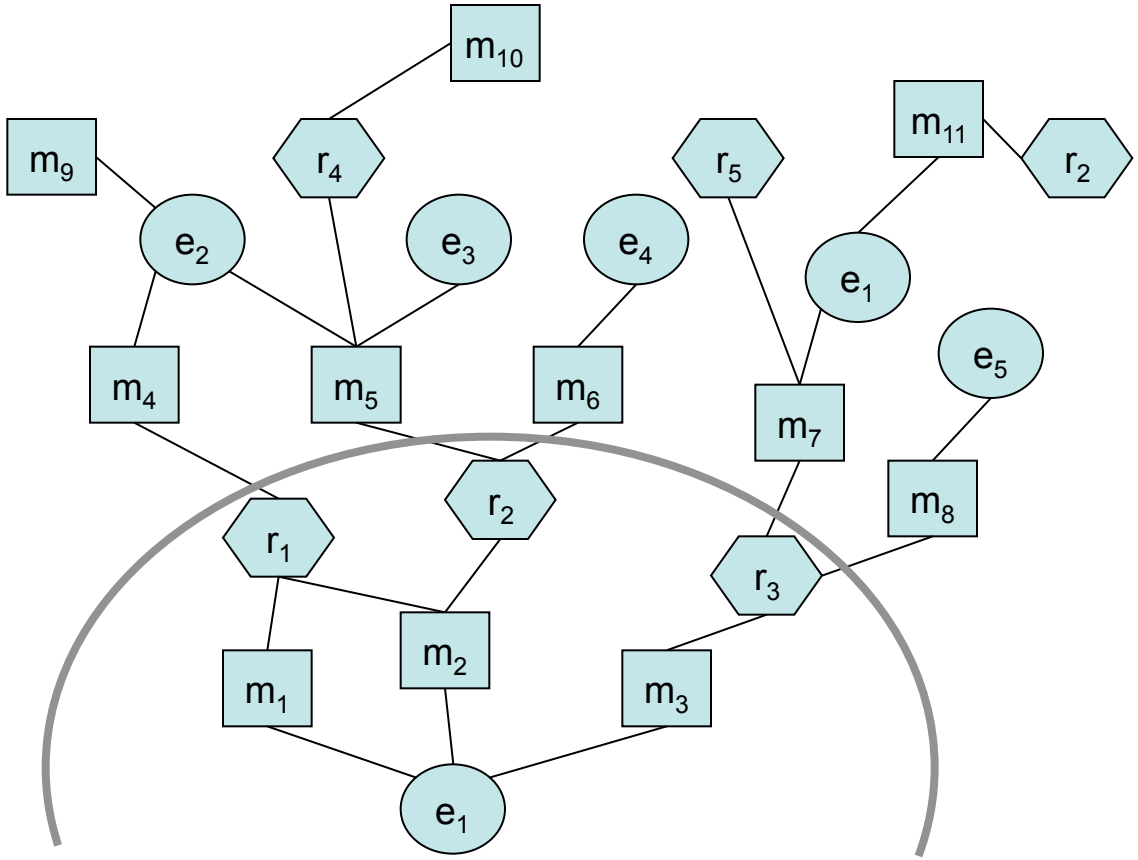


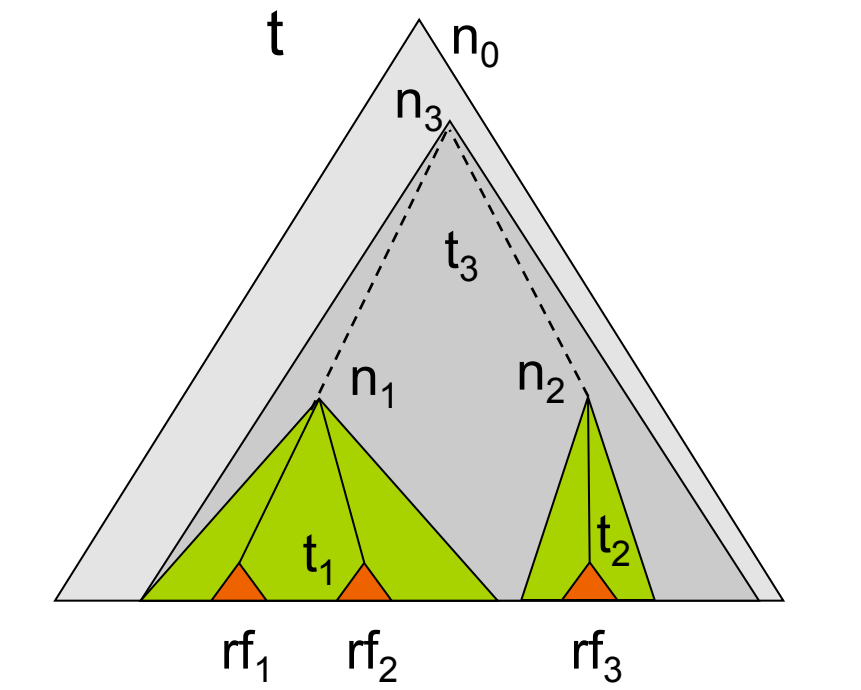
Bootstrapping Relation Extraction with Semantic Seed

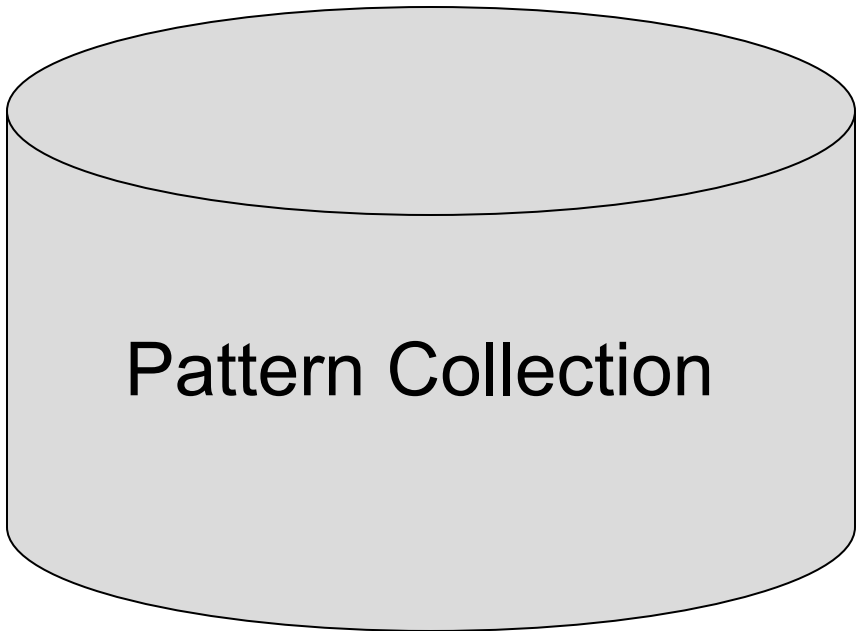
- DIPRE and Snowball
  - binary relations only, no projections, no linguistic analysis
- DARE
  - n-ary relations and their projections, deep linguistic analysis

(in the experiments I use MINIPAR by Dekan Lin 1999)

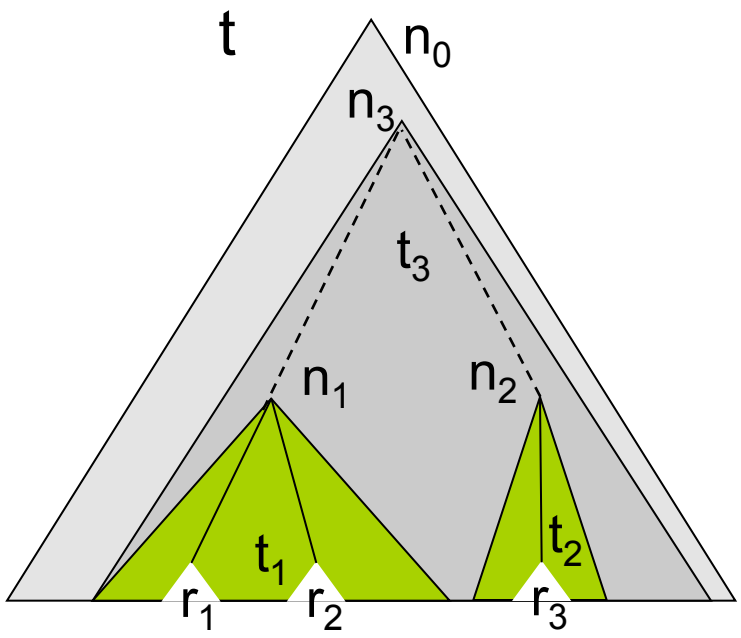
Start of Bootstrapping (simplified)

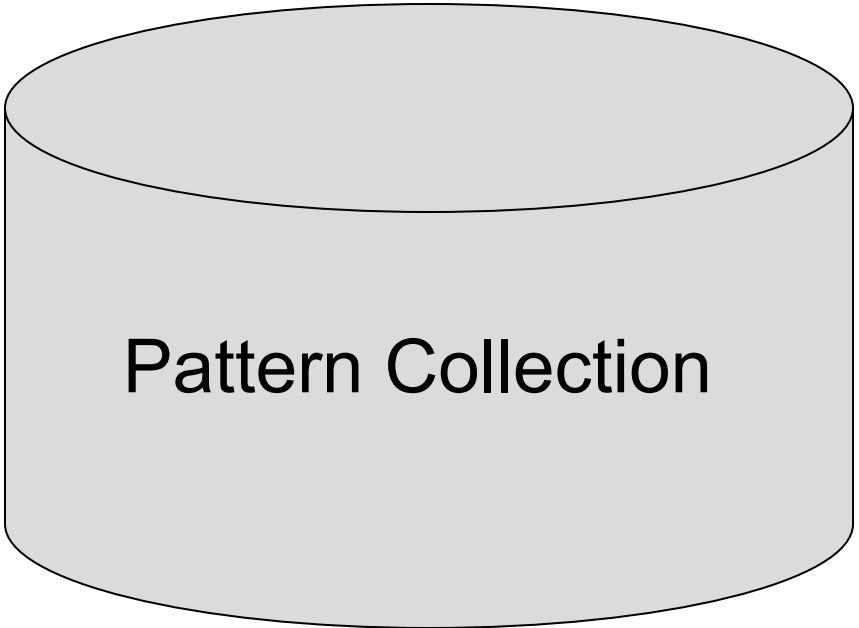




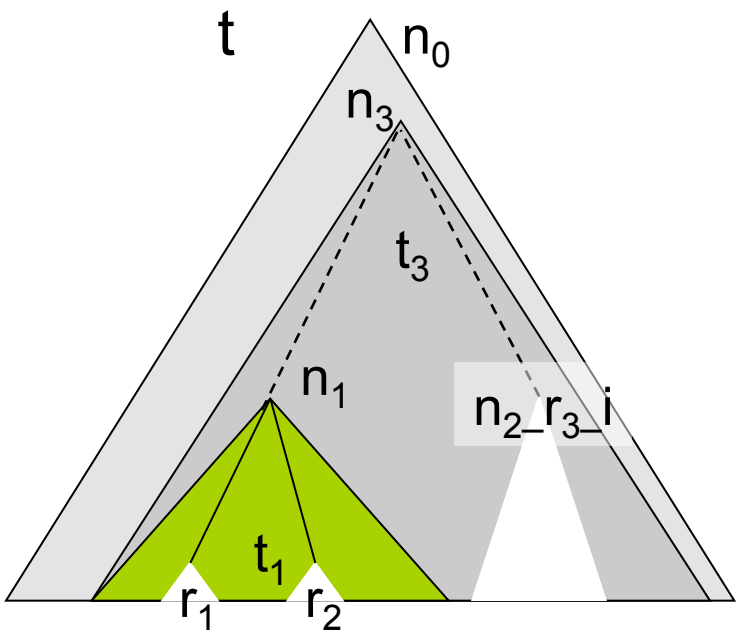


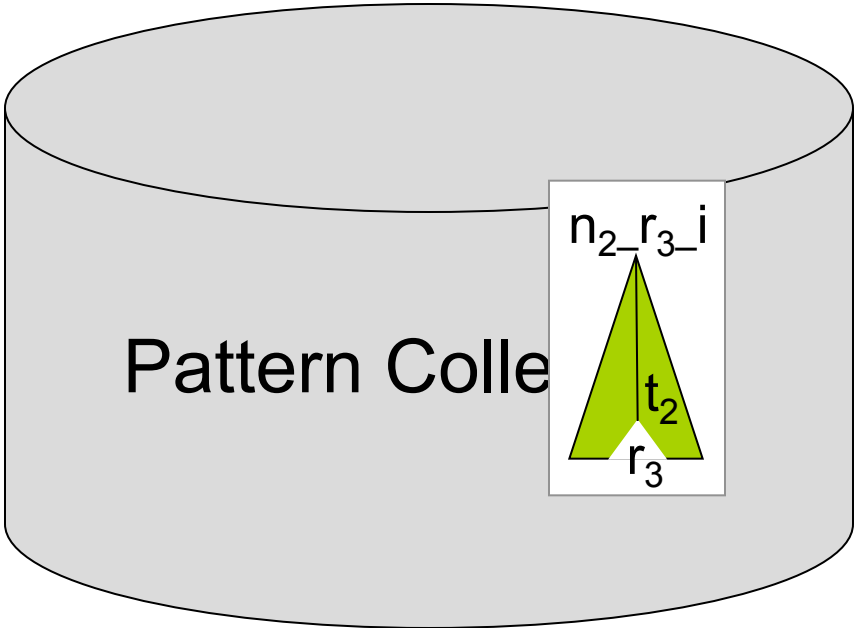
0. replace all nodes that are instantiated with the seed arguments by new nodes. Label these new nodes with the seed argument roles and their entity classes;





0. replace all nodes that are instantiated with the seed arguments by new nodes. Label these new nodes with the seed argument roles and their entity classes;





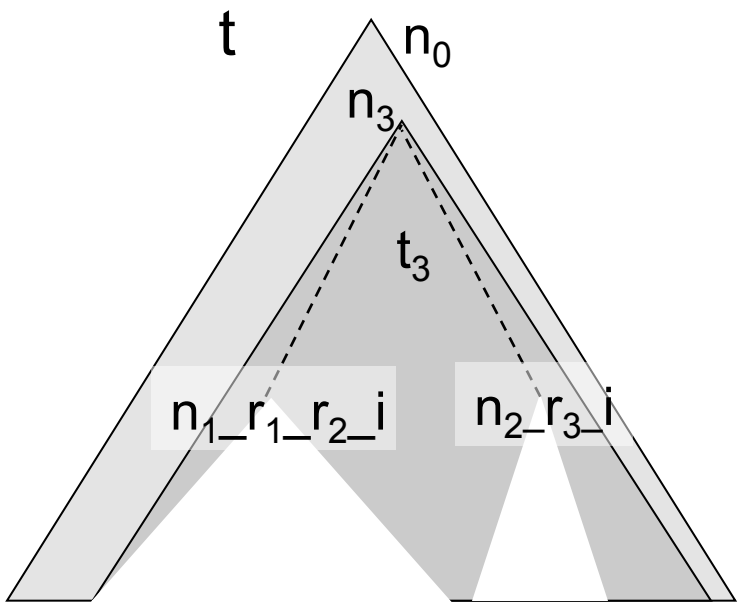
0. replace all nodes that are instantiated with the seed arguments by new nodes. Label these new nodes with the seed argument roles and their entity classes;

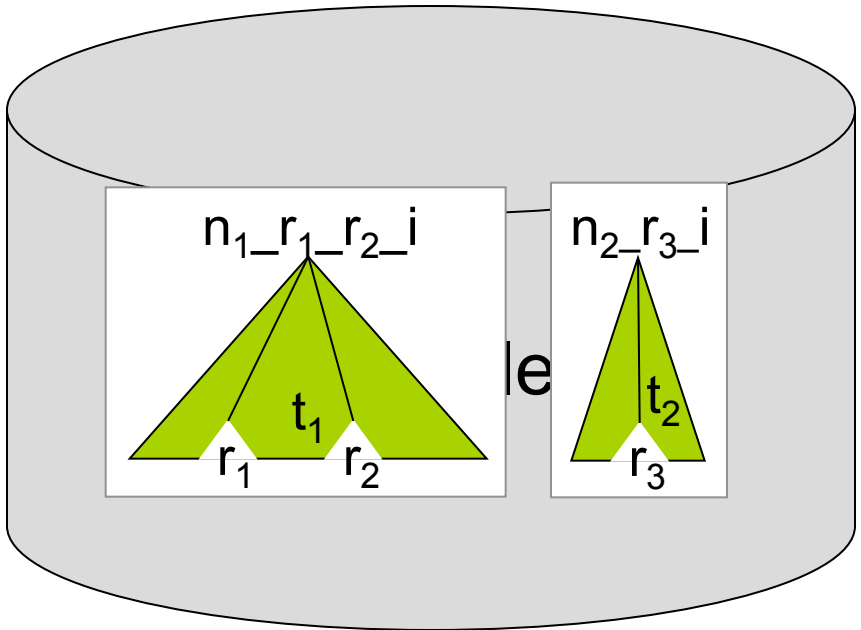
for i=1 to n

1. identify the set of the lowest non-terminal nodes N_1 in t that dominate i arguments (possibly among other nodes).

2. substitute N_1 by nodes labelled with the seed argument roles and their entity classes

3. prune the subtrees dominated by N_1 from t and add these subtrees into the pattern collection. These subtrees are assigned the argument role information and a unique id.





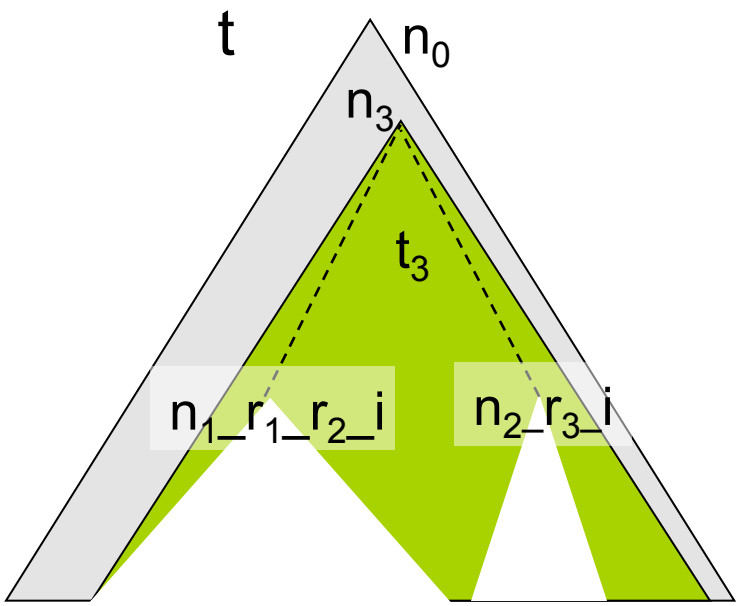
0. replace all nodes that are instantiated with the seed arguments by new nodes. Label these new nodes with the seed argument roles and their entity classes;

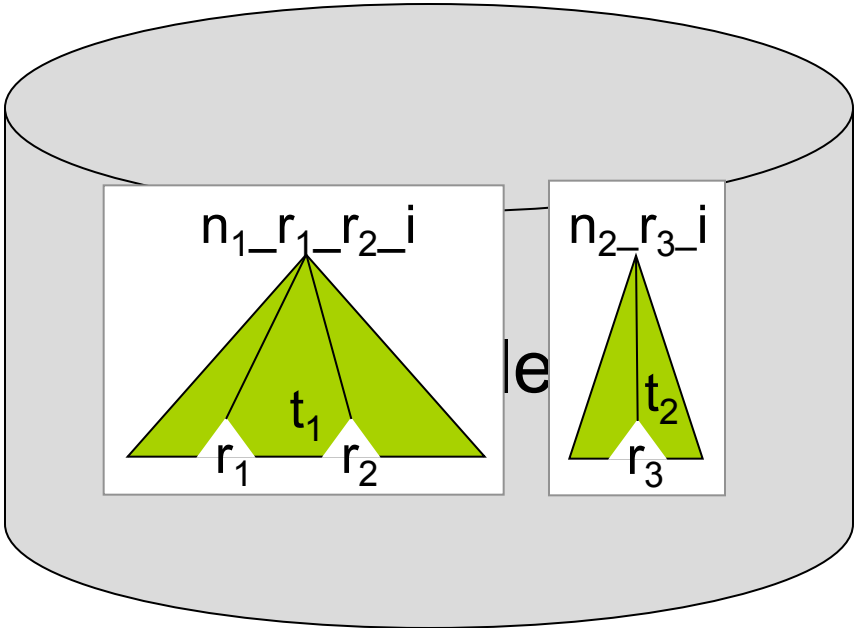
for i=1 to n

1. identify the set of the lowest non-terminal nodes N_1 in t that dominate i arguments (possibly among other nodes).

2. substitute N_1 by nodes labelled with the seed argument roles and their entity classes

3. prune the subtrees dominated by N_1 from t and add these subtrees into the pattern collection. These subtrees are assigned the argument role information and a unique id.





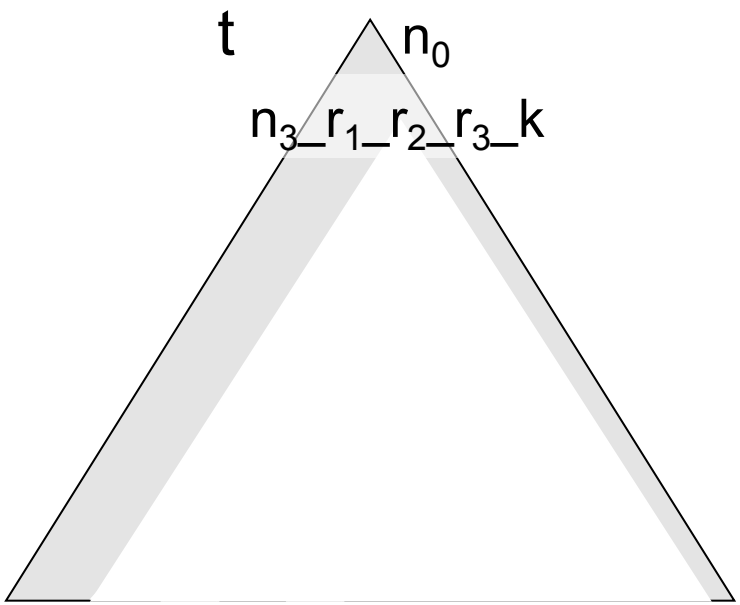
0. replace all nodes that are instantiated with the seed arguments by new nodes. Label these new nodes with the seed argument roles and their entity classes;

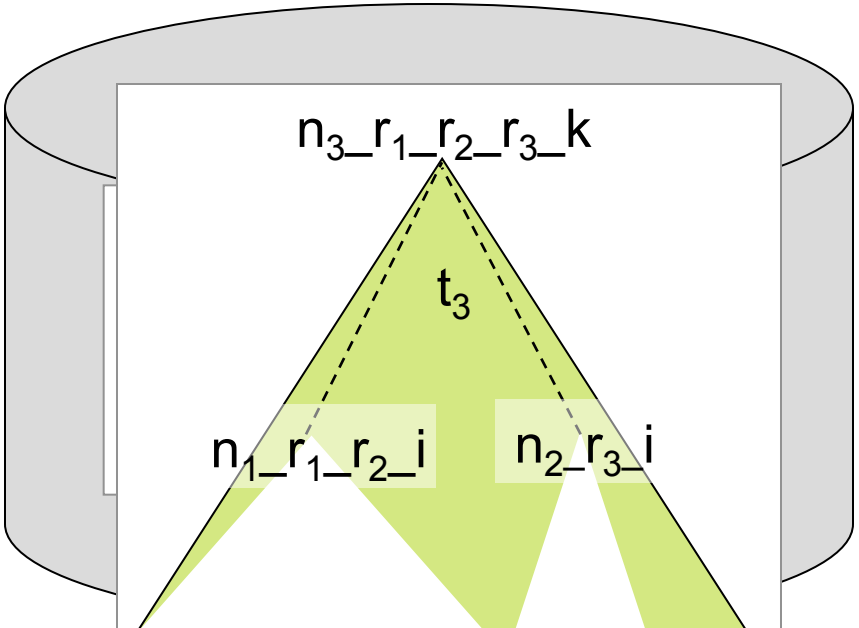
for i=1 to n

1. identify the set of the lowest non-terminal nodes N_1 in t that dominate i arguments (possibly among other nodes).

2. substitute N_1 by nodes labelled with the seed argument roles and their entity classes

3. prune the subtrees dominated by N_1 from t and add these subtrees into the pattern collection. These subtrees are assigned the argument role information and a unique id.





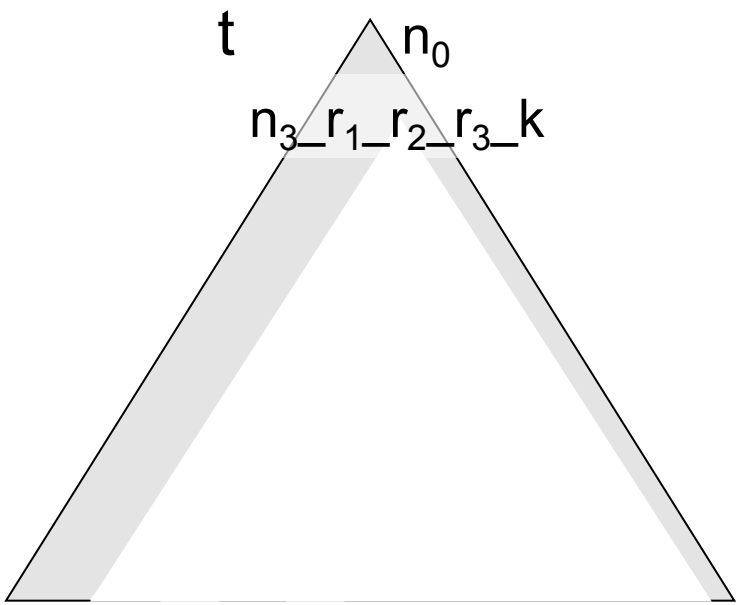
0. replace all nodes that are instantiated with the seed arguments by new nodes. Label these new nodes with the seed argument roles and their entity classes;

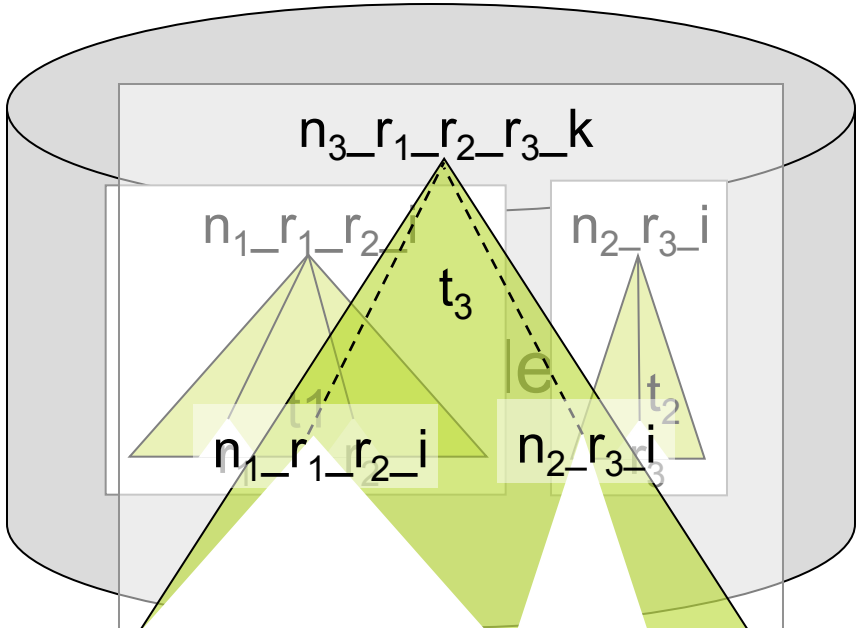
for i=1 to n

1. identify the set of the lowest non-terminal nodes N_1 in t that dominate i arguments (possibly among other nodes).

2. substitute N_1 by nodes labelled with the seed argument roles and their entity classes

3. prune the subtrees dominated by N_1 from t and add these subtrees into the pattern collection. These subtrees are assigned the argument role information and a unique id.





0. replace all nodes that are instantiated with the seed arguments by new nodes. Label these new nodes with the seed argument roles and their entity classes;

for i=1 to n

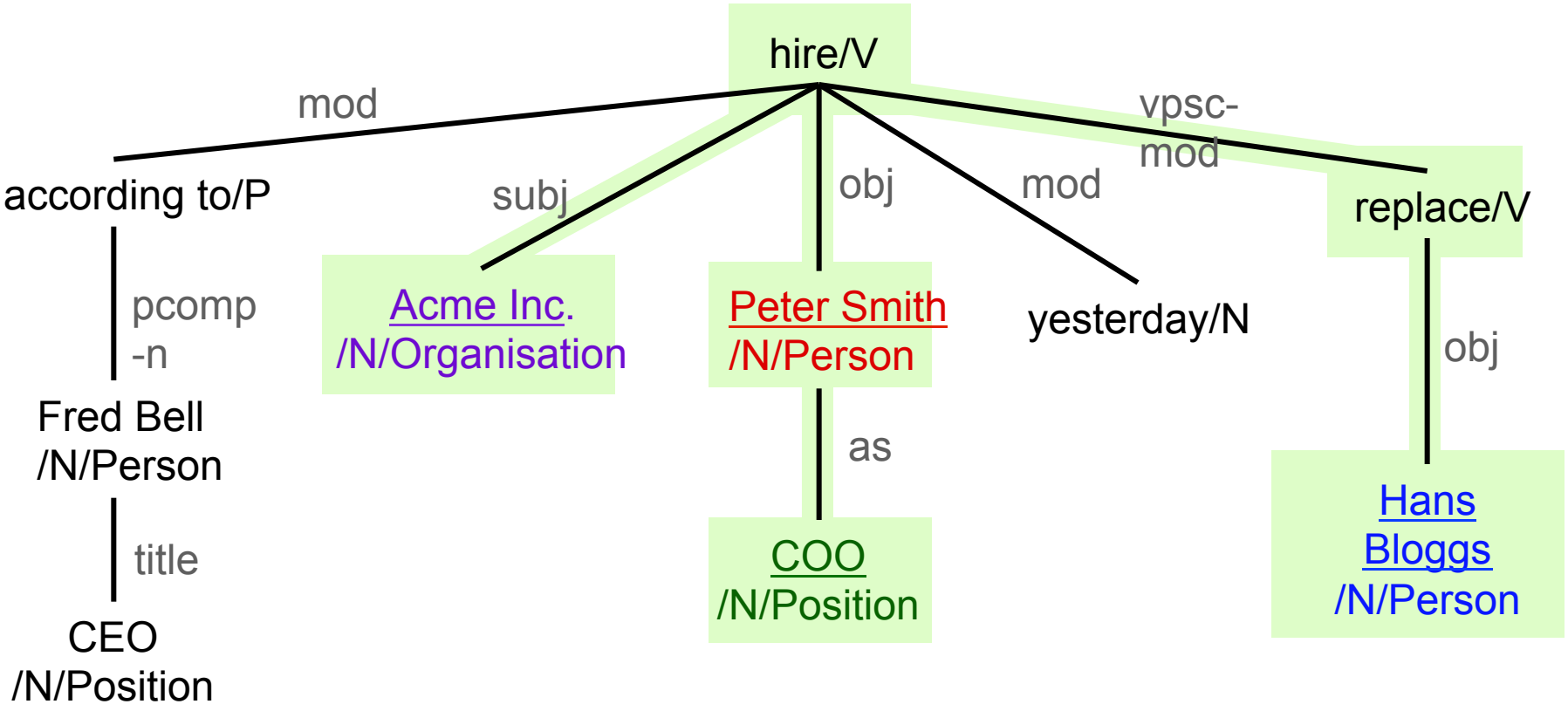
1. identify the set of the lowest non-terminal nodes N_1 in t that dominate i arguments (possibly among other nodes).

2. substitute N_1 by nodes labelled with the seed argument roles and their entity classes

3. prune the subtrees dominated by N_1 from t and add these subtrees into the pattern collection. These subtrees are assigned the argument role information and a unique id.

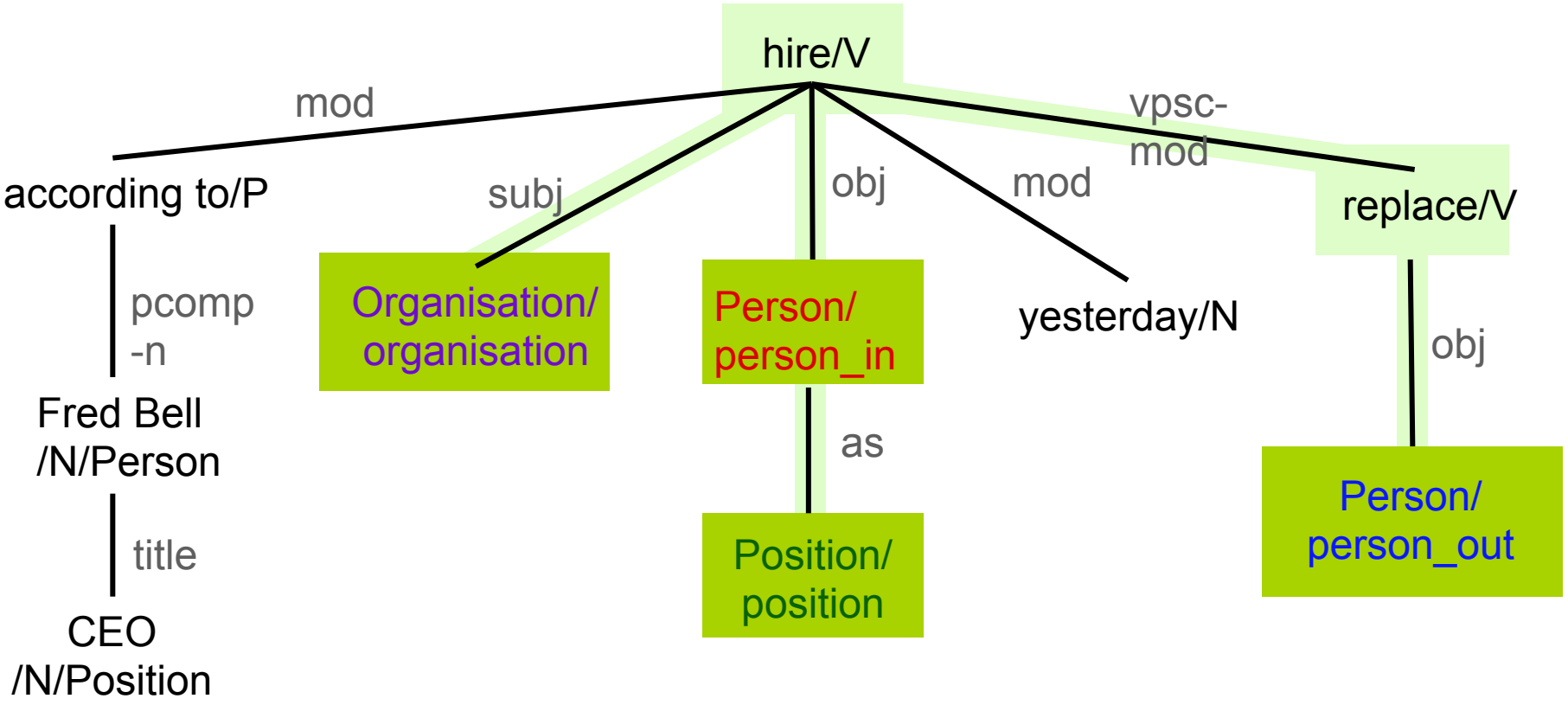
According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

<Peter Smith/person_in, Hans Bloggs/person_out, COO /position, Acme Inc. /organisation>



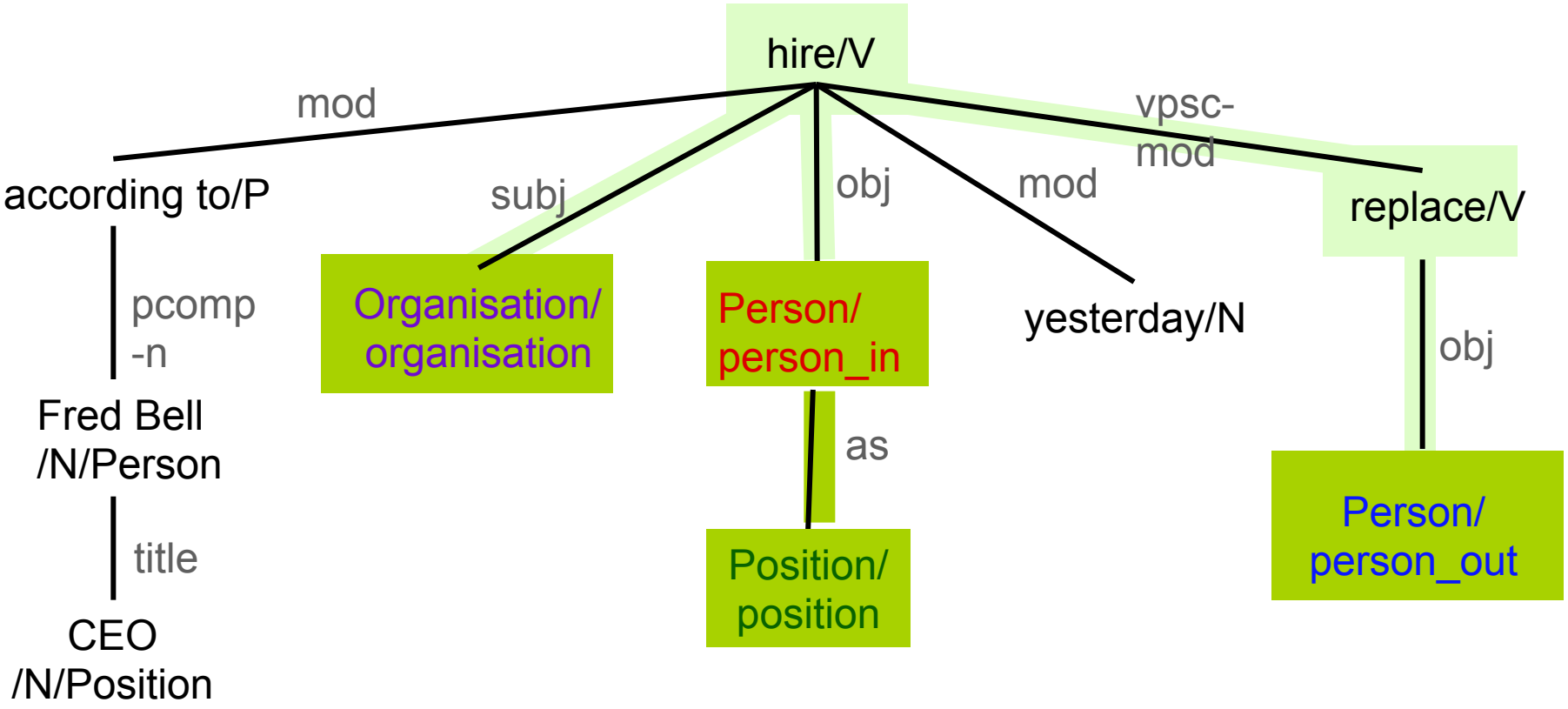
According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

<Peter Smith/person_in, Hans Bloggs/person_out, COO /position, Acme Inc. /organisation>



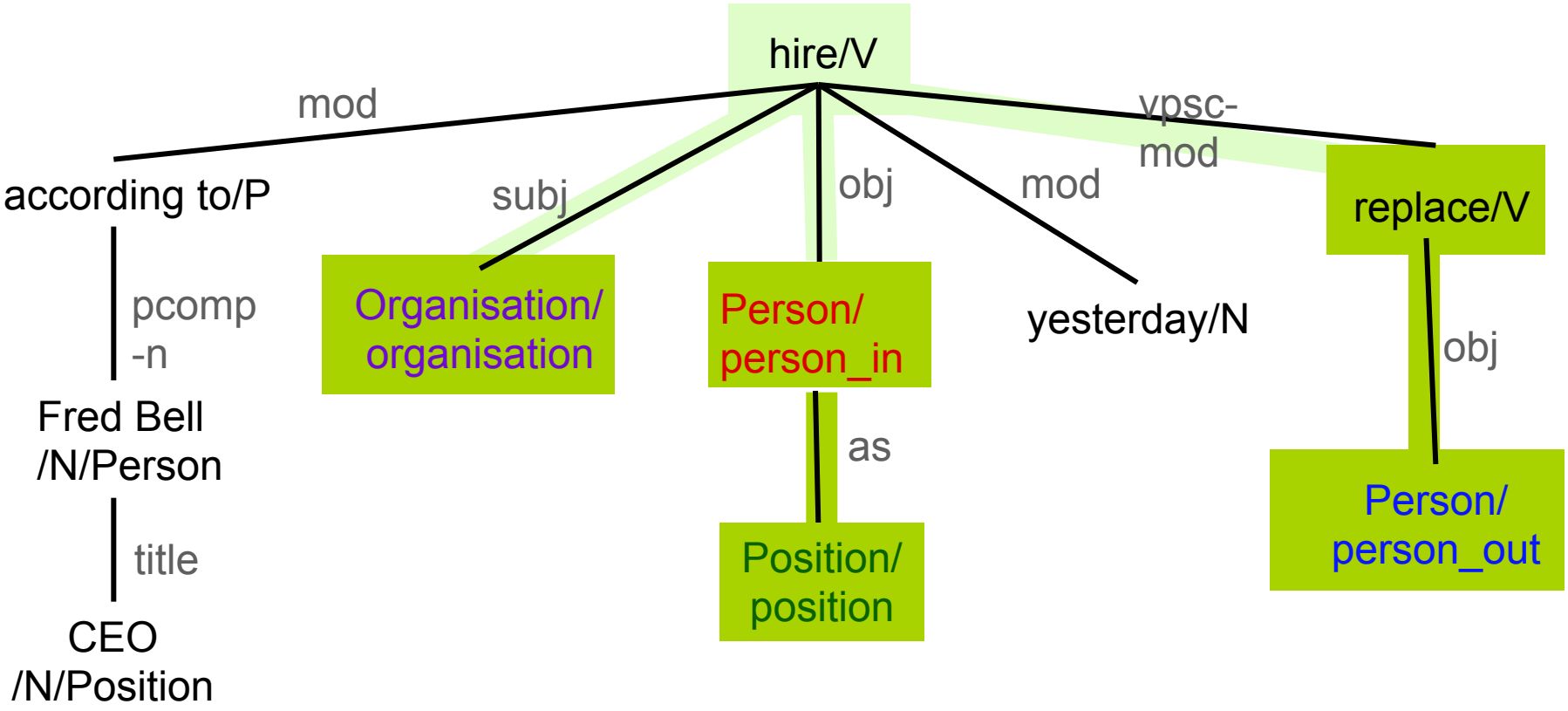
According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

<Peter Smith/person_in, Hans Bloggs/person_out, COO /position, Acme Inc. /organisation>



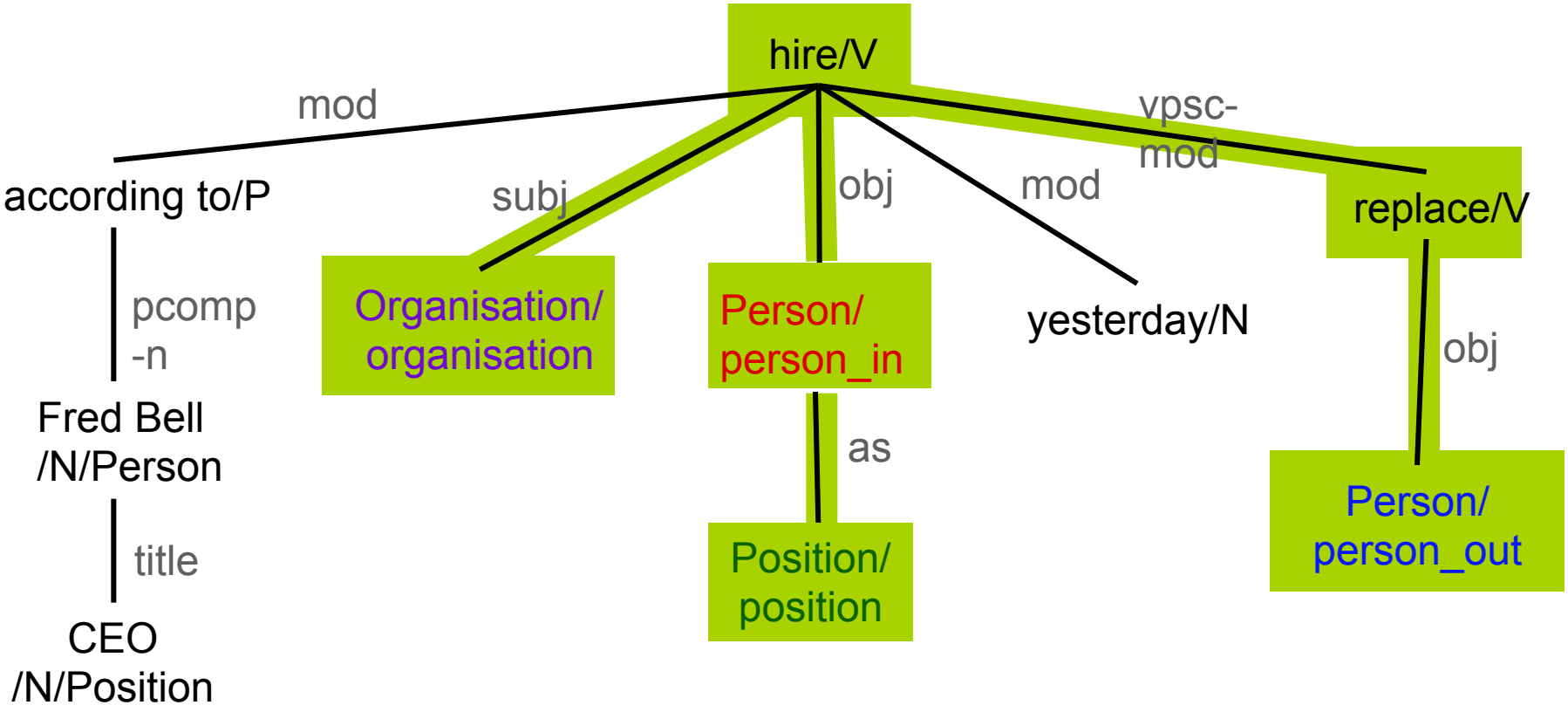
According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

<Peter Smith/person_in, Hans Bloggs/person_out, COO /position, Acme Inc. /organisation>



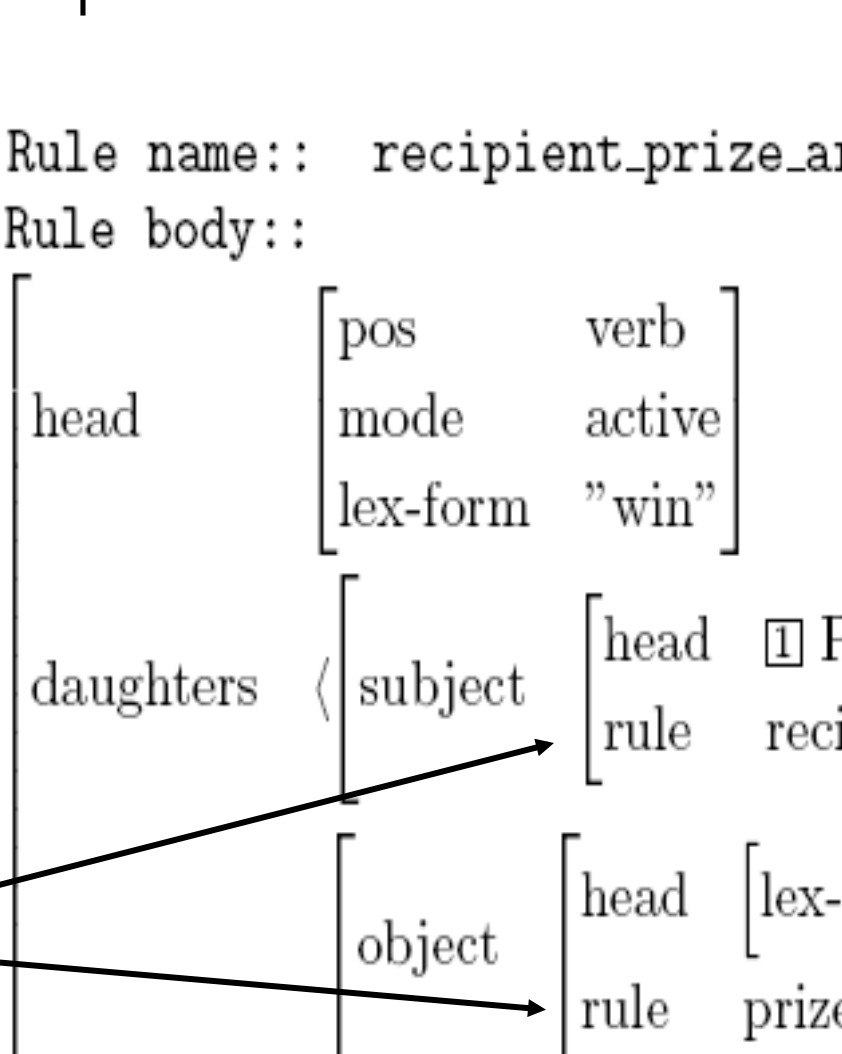
According to CEO Fred Bell, Acme Inc. hired Peter Smith as COO yesterday, replacing Hans Bloggs.

<Peter Smith/person_in, Hans Bloggs/person_out, COO /position, Acme Inc. /organisation>



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



DARE Rule Components

Rule name:: recipient_prize_area_year_1

Rule body::

$$\left[\begin{array}{ll} \text{head} & \left[\begin{array}{ll} \text{pos} & \text{verb} \\ \text{mode} & \text{active} \\ \text{lex-form} & \text{"win"} \end{array}\right] \\ \text{daughters} & \left\langle \left[\begin{array}{ll} \text{subject} & \left[\begin{array}{ll} \text{head} & \boxed{1}\ \text{Person} \\ \text{rule} & \text{recipient\_1::}\ \langle \boxed{1} Person \rangle \end{array}\right] \end{array}\right], \right. \\ & \left. \left[\begin{array}{ll} \text{object} & \left[\begin{array}{ll} \text{head} & \left[\text{lex-form} \quad \text{"prize"}\right] \\ \text{rule} & \text{prize\_area\_year\_1::}\ \langle \boxed{2} Prize, \boxed{3} Area, \boxed{4} Year \rangle \end{array}\right] \end{array}\right] \right\rangle \end{array}\right]$$

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

prize_area_year_1

Rule name:: prize_area_year_1

Rule body::

$$\begin{bmatrix} \text{head} & \begin{bmatrix} \text{pos} & \text{noun} \\ \text{lex-form} & \text{"prize"} \end{bmatrix} \\ \text{daughters} & \left\langle \begin{bmatrix} \text{lex-mod} & \begin{bmatrix} \text{head} & \boxed{3}\ \text{Year} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{lex-mod} & \begin{bmatrix} \text{head} & \boxed{1}\ \text{Prize} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \text{lex-mod} & \begin{bmatrix} \text{head} & \boxed{2}\ \text{Area} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

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

Two Domains

- Award Events (start with subdomain Nobel Prizes)

  reasons: good news coverage
  complete list of all award events
  good starting point for other award domains

- Management Succession Events

  reason: comparison with previous work

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 A (1981-1998)	1032	5.8 MB
Nobel Prize B (1999-2005)	2296	12.6 MB
Nobel Prize A+B	3328	18.4 MB
MUC-6	199	1MB

Evaluation Against Ideal Tables

Data Set	Seed	Precision	Recall
Nobel Prize A	<[Sen, Amartya], nobel, economics, 1998>	**87.3%**	**31.0%**
Nobel Prize A	<[Arias, Oscar], nobel, peace, 1987>	**83.8%**	**32.0%**
Nobel Prize B	<[Zewail, Ahmed H], nobel, chemistry,1999>	**71.6%**	**50.7%**
A+B	<[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%**

Comparison

Our result with 20 seeds (after 4 iterations)

- precision: 48.4%
- recall: 34.2%

compares well with the best result reported so far by (Greenwood and Stevenson, 2006) with the linked chain model starting with 7 hand-crafted patterns (after 190 iterations)

- precision: 43.4%
- recall: 26.5%

Reusability of Rules

- Prize award patterns
  - Detection of other Prizes such as *Pulitzer Prize*, *Turner Prize*
  - Precision: 86.2%
- Management succession
  - Domain independent binary pattern rules:
    Person-Organisation, *Person-Position*
  - Evaluation of top 100 relation instances
    Precision: 98%

Research Questions

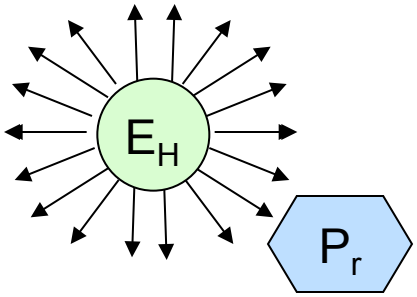
As scientists we want to know

- Why does it work for some tasks?
- Why doesn‘t it work for all tasks?
- How can we estimate the suitability of domains?
- How can we deal with less suitable domains?

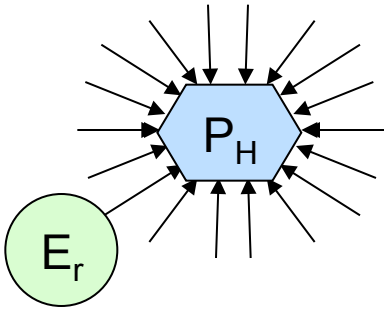
Careful analysis confirmed the following assumption:

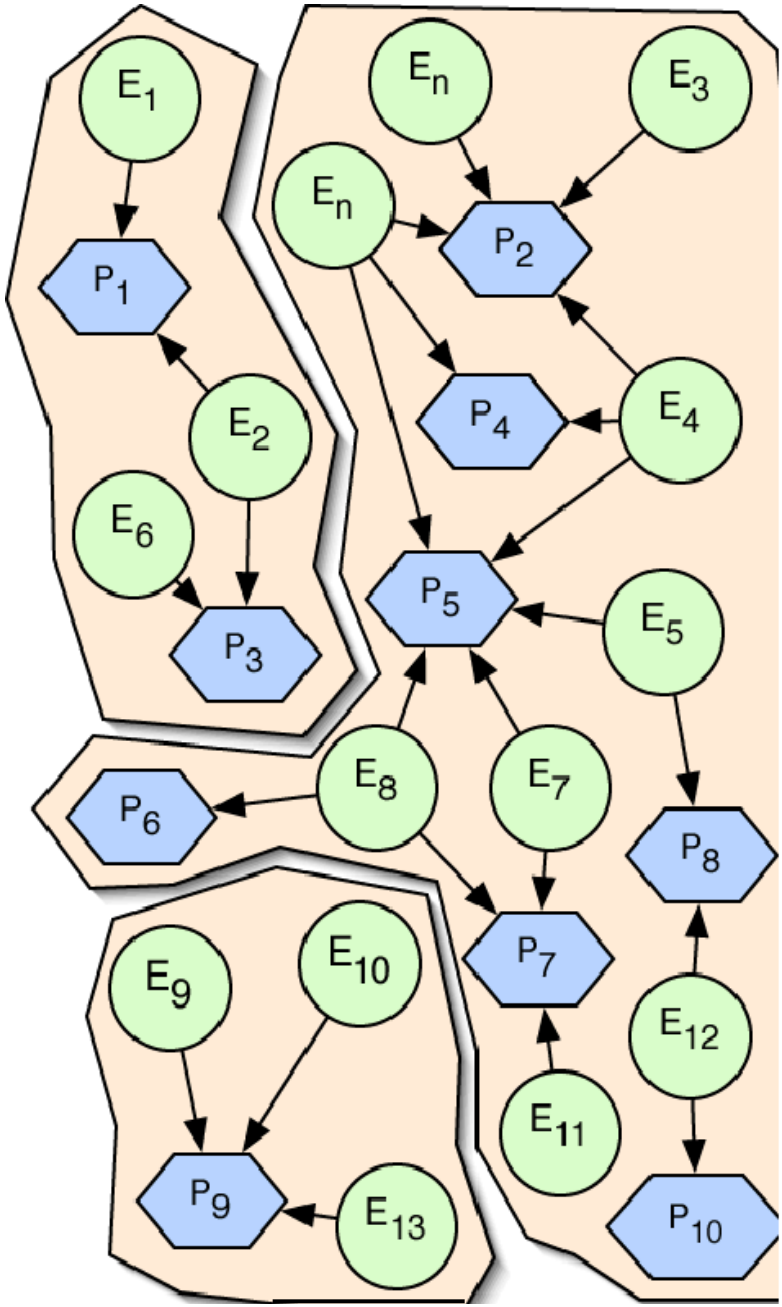
redundancy, both on patterns and event mentions, helps.

Frequently reported events make rare patterns reachable

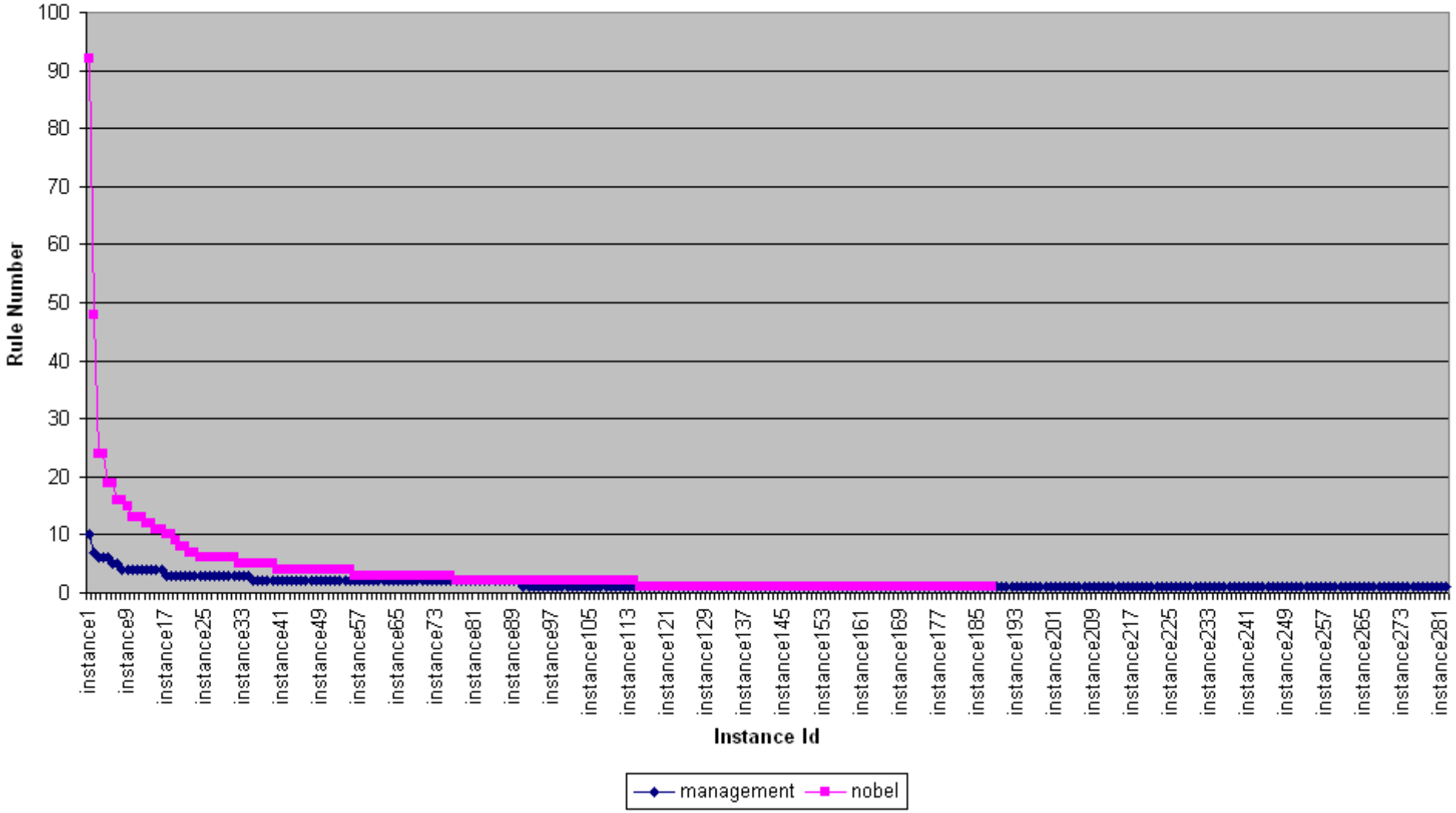


Popular patterns help to reach rarely mentioned events

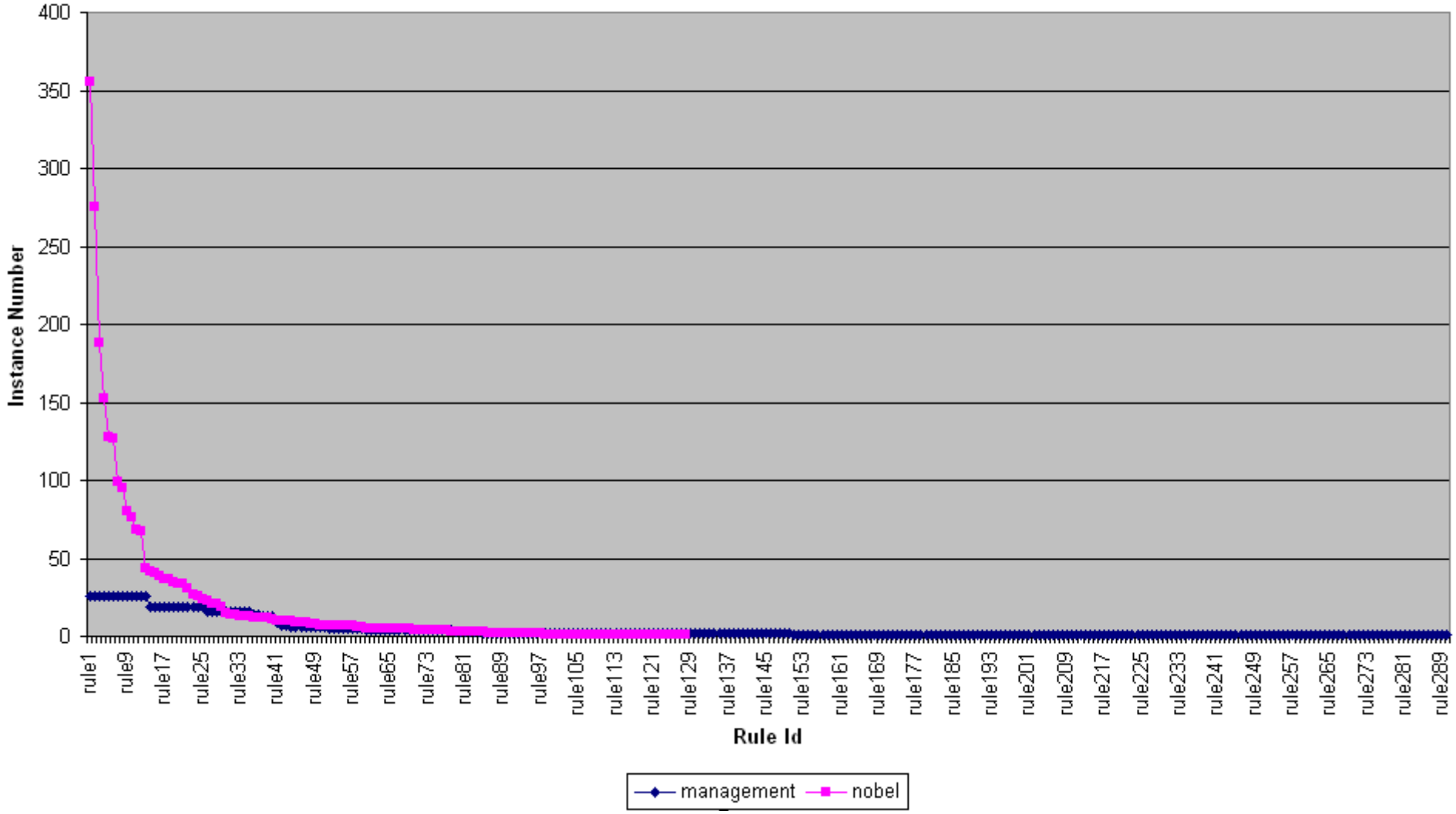




Instance to Pattern
Nobel Prize vs. Management Succession



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)

Conclusion

- 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

- In addition to the obvious practical advantages, the approach offers theoretical benefits: It supports a view of IE as a systematic gradual approximation of language understanding.

Domain-Adaptation for Big Text Data Analytics

Text Analytics for Big Textual Data

- Three main features of big data
  - *Volume: large-scale* in volume
  - *Variety:* with respect to heterogeneous domains and formats
  - *Velocity:* because of its rapid and steady growing.

- Requirements of text analytics technologies for big data
  - *efficient*
  - *robust*
  - *scalable*
  - *domain-adaptive*

Domain Adaptation is Essential for Big Data!

- Among the three big data features, variety and velocity are even more challenging than the sheer size volume

Reasons:

- New domains have been constantly emerging, rapidly growing in size.
- Domains can differ in
  - topics (e.g., medicine, chemistry or mechanics)
  - genres (e.g., news, novels, blogs, scientific publications or patents)
  - targets (e.g., different relations such as marriage, person-parent relation, disease-symptom relation)
  - data internal properties (e.g., size or redundancy or connectivity).
- Systems, methods or strategies developed or trained for so-called general purpose or one specific domain can often not be directly taken over by other domains, because
  - each domain needs its own domain knowledge and
  - each application data has its own special properties.

Relevant Strategies for Domain Adaptation

- Minimally dependent on the labeled training data
  - Minimally or weakly supervised machine learning methods
- Strategies for
  - confidence estimation of automatically learned information and knowledge
  - filtering of irrelevant and wrong information
- Domain adaptation of generic systems for specific applications

Our solutions (1)

- minimally supervised and distantly supervised automatic learning of domain-specific grammar-based pattern rules for n-ary RE: DARE and Web-DARE Systems

  - Feiyu Xu, Hans Uszkoreit, Hong Li, "A Seed-driven Bottom-up Machine Learning Framework for Extracting Relations of Various Complexity (2007)". In ACL 2007.

  - Hans Uszkoreit, Feiyu Xu, Hong Li. "Analysis and Improvement of Minimally Supervised Machine Learning for Relation Extraction". In NLDB 2009.

  - Sebastian Krause, Hong Li, Hans Uszkoreit, Feiyu Xu, "Large-Scale Learning of Relation-Extraction Rules with Distant Supervision from the Web". In Proceedings of the 11th International Semantic Web Conference (ISWC 2012).

Our solutions (2)

- Various filtering and confidence estimation methods for high-performance and large-scale relation extraction

  - Sebastian Krause, Hong Li, Hans Uszkoreit, Feiyu Xu, "Large-Scale Learning of Relation-Extraction Rules with Distant Supervision from the Web". In Proceedings of the 11th International Semantic Web Conference (ISWC 2012)

  - Andrea Moro, Hong Li, Sebastian Krause, Feiyu Xu, Roberto Navigli, Hans Uszkoreit, "Semantic rule filtering for web-scale relation extraction". In Proceeding of International Semantic Web Conference (ISWC 2013).

  - Feiyu Xu, Hans Uszkoreit, Sebastian Krause, Hong Li. Boosting Relation Extraction with Limited Closed-World Knowledge. COLING 2009, Poster.

Our solutions (3)

- Automatic adaptation and improvement of generic parsing results for specific domains

  - Peter Adolphs, Feiyu Xu, Hans Uszkoreit, Hong Li, "Dependency Graphs as a Generic Interface between Parsers and Relation Extraction Rule Learning". In Proceedings of KI 2011, pp. 50-62, 2011.

  - Feiyu Xu, Hong Li, Yi Zhang, Hans Uszkoreit, Sebastian Krause, "Parse reranking for domain-adaptive relation extraction". Journal of Logic and Computation, doi: 10.1093/logcom/exs055, Oxford University Press, 2012.

Our solutions (4)

- Automatic generation of domain-specific linguistic knowledge resources

  - Hans Uszkoreit and Feiyu Xu, "From Strings to Things, SAR-Graphs: A New Type of Resource for Connecting Knowledge and Language". In Proceedings of 1st International Workshop on NLP and DBpedia volume 1064, Sydney, NSW, Australia, CEUR Workshop Proceedings, 10/2013

  - Open source: sargraph.dfki.de

Web-DARE
Distant-supervised Web-scale RE

Web-DARE: Distant Supervision based RE

- Large number of RE rules are automatically learned by using Freebase as seed knowledge and Web as training corpus
- Goal:
  - covering most linguistic variants for expressing a relation
  - thus solving the notorious long-tail problem of real-world NL applications

Data Set

- rules learned for 39 relations
  - n-ary relations n>=2
- three domains: business, awards and people
- 2.8 million relation instances retrieved from Freebase as seed
- 20 million web documents as training corpus

Example in Nobel Prize Award Domain

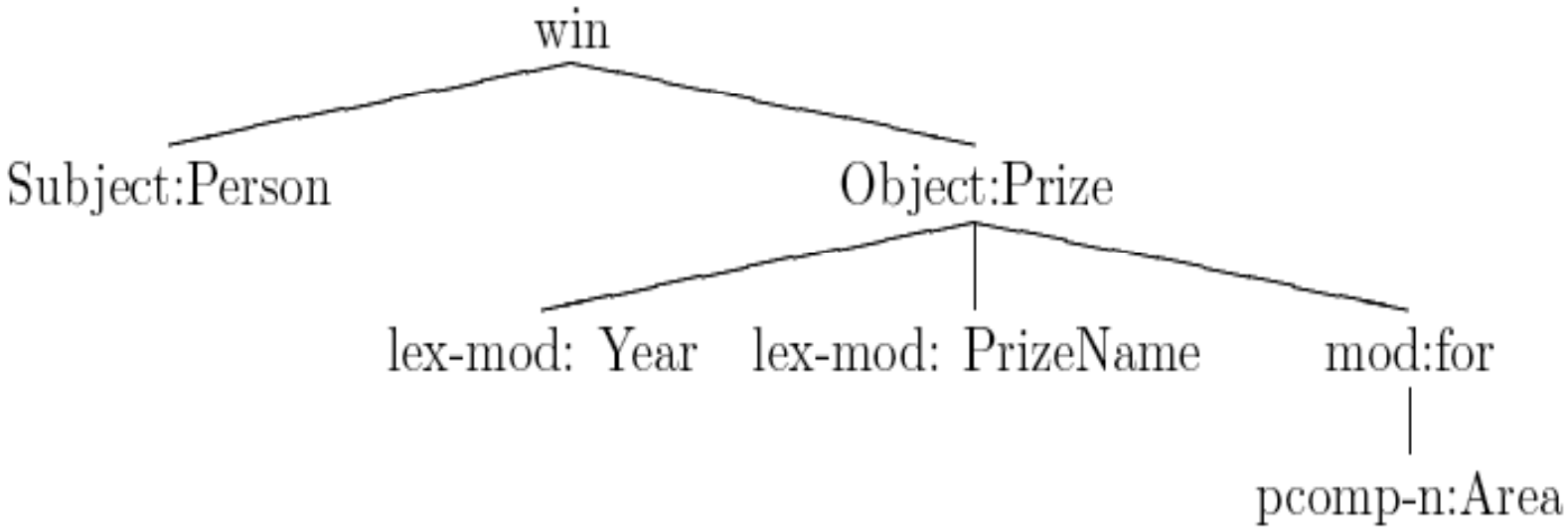
- Seed example

> *<Mohamed ElBaradei/Person, Nobel/Prize, Peace/Area, 2005/Year>*

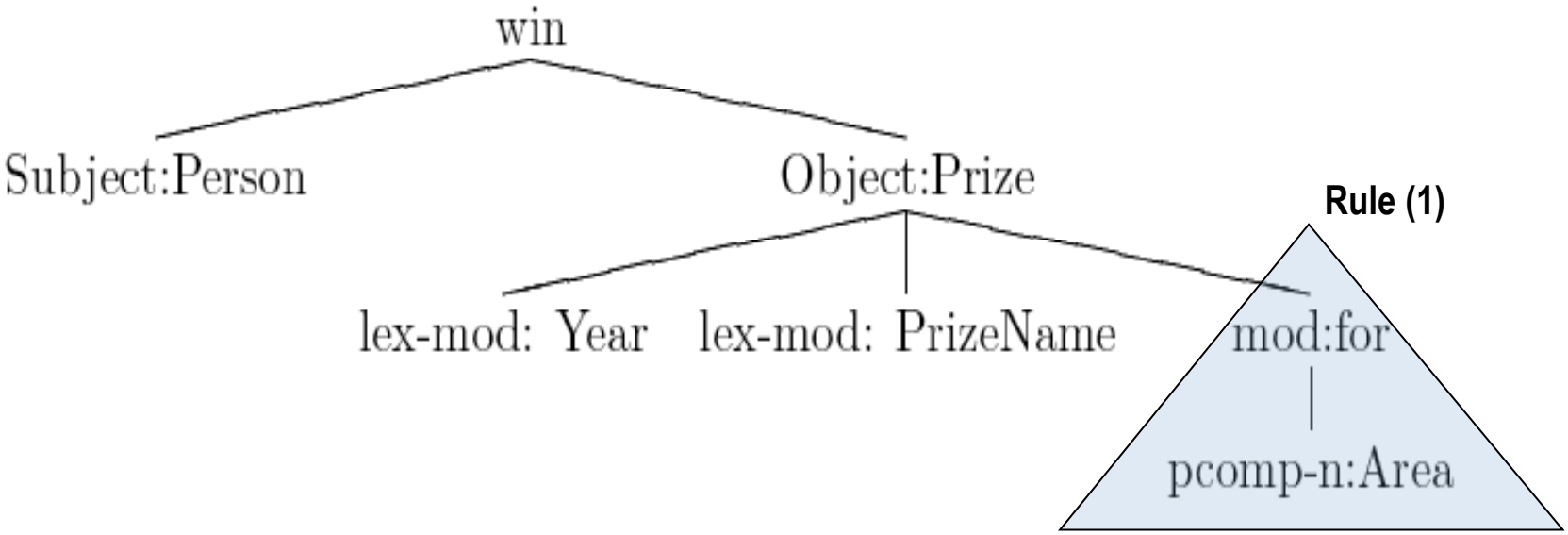
- Sentence matched with the seed

> Mohamed ElBaradei won the 2005 Nobel Prize for Peace on Friday ...

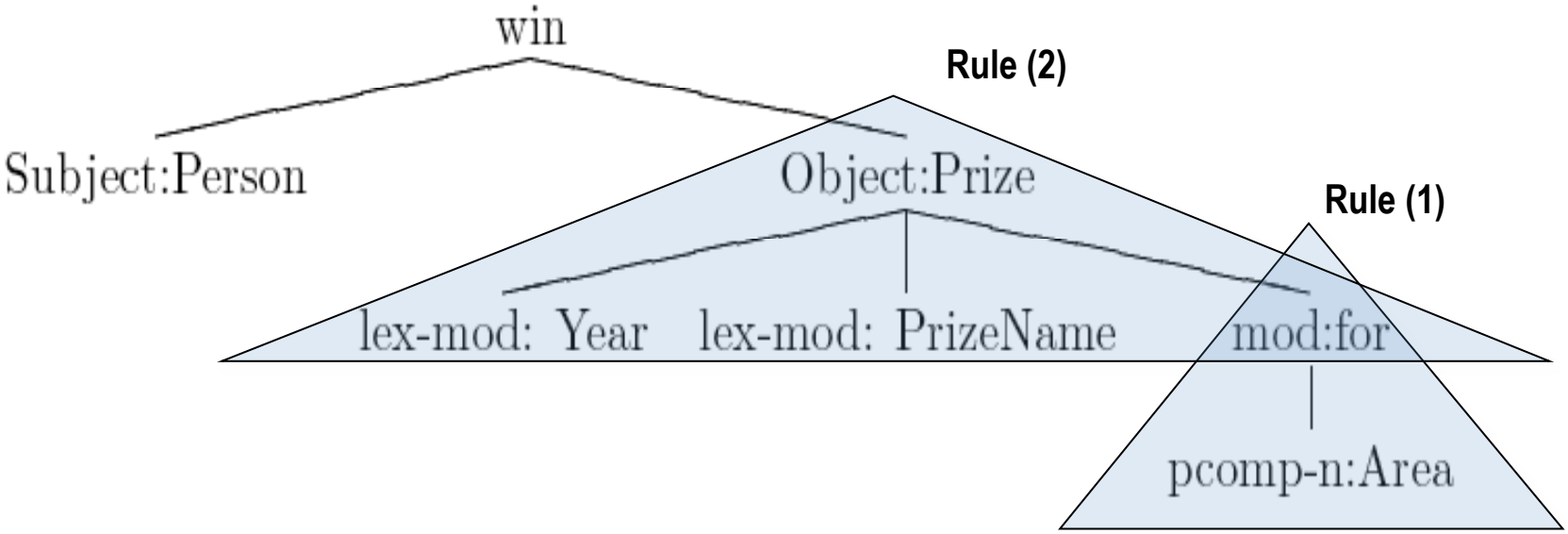
Dependency Parse Result



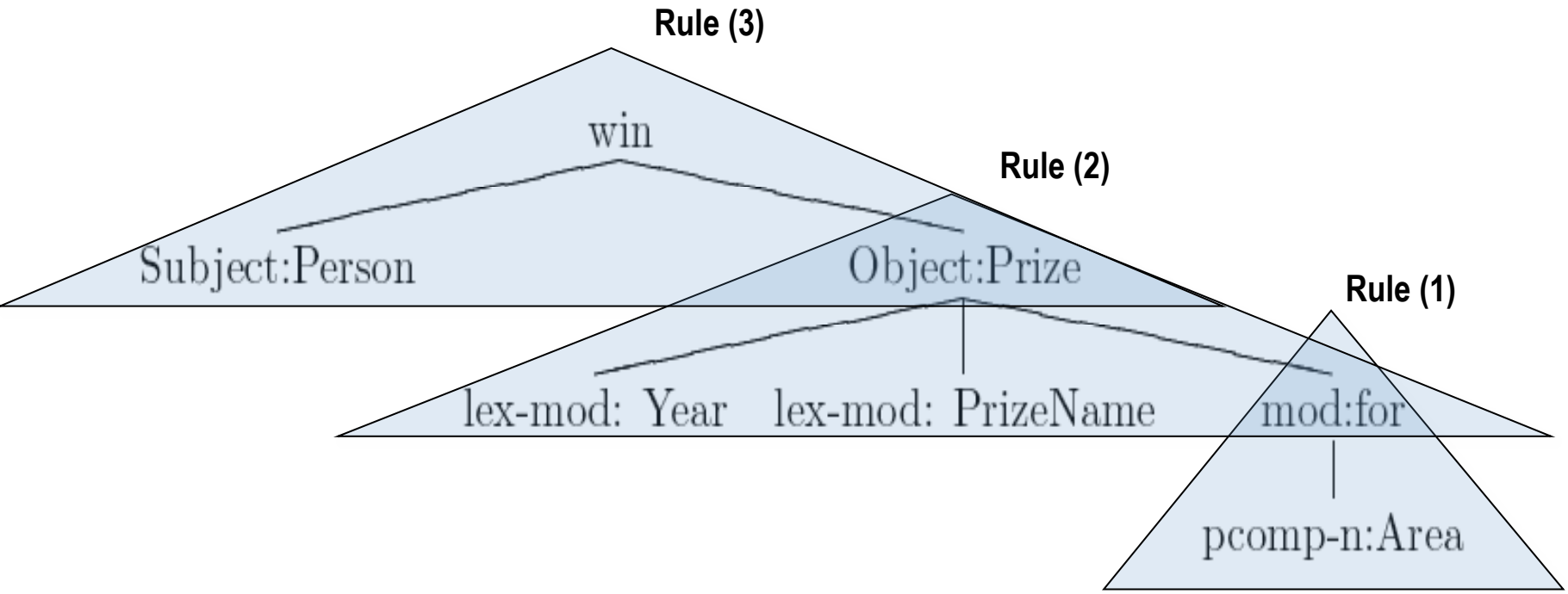
Bottom Up Rule Learning



Bottom Up Rule Learning



Bottom Up Rule Learning



Web-DARE Architecture

Google
bing

Queries

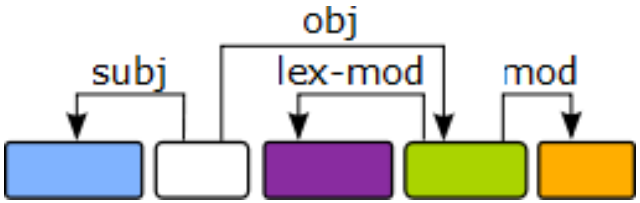
Facts

Web pages

Filtered Patterns

Sentence Mentions of Facts

Patterns



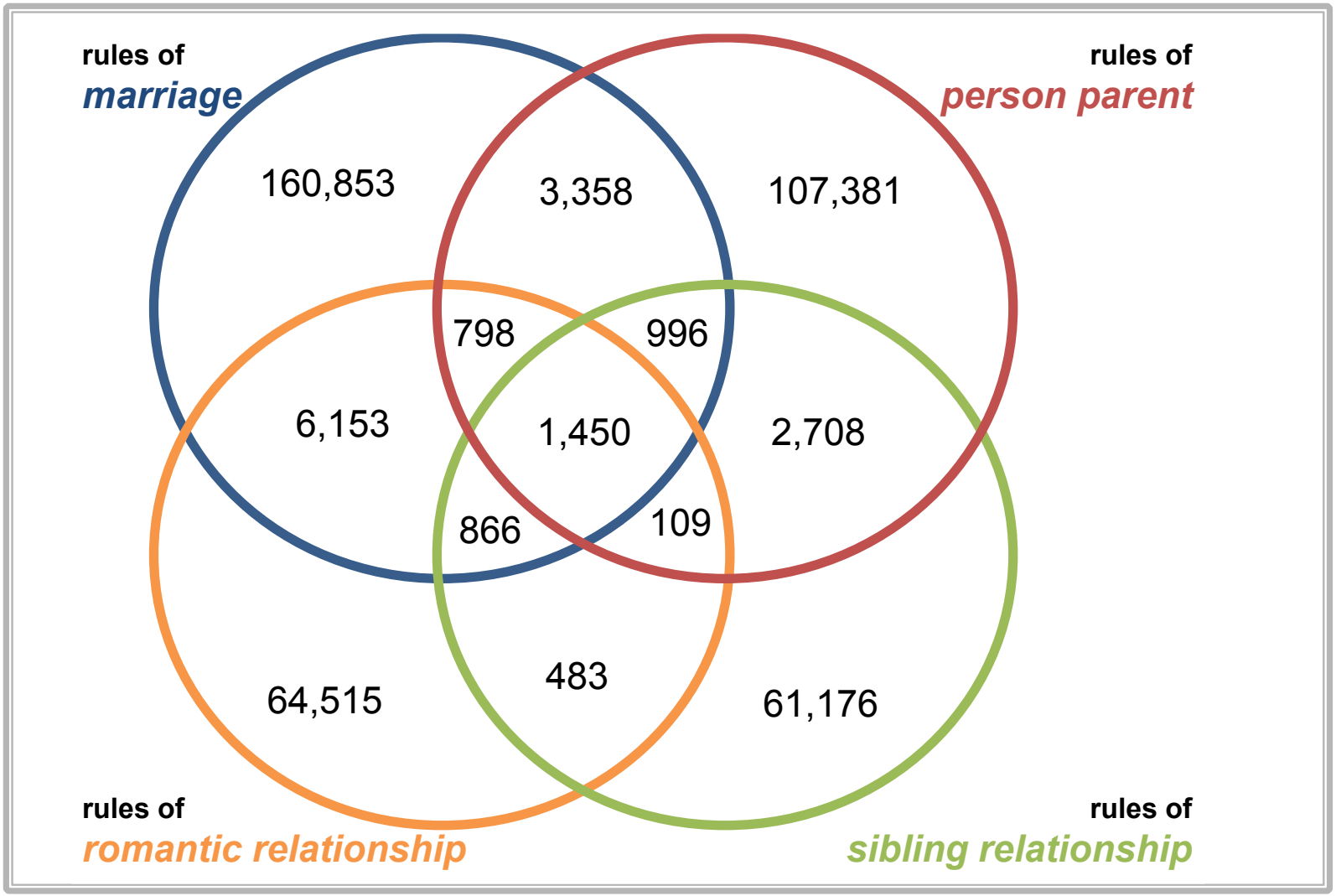
Some Statistics of Web-DARE Rules

Relation	# Sentences used	# Sentences w/ a learned rule	# Rules	# Rules w/o duplicates
award nomination	13,966	13,149	23,987	7,800
award honor	50,550	49,001	106,550	40,578
hall of fame induction	31,244	28,278	44,920	17,450
organization relationship	46,331	42,824	60,379	28,903
acquisition	63,967	60,903	96,747	50,544
organization merger	2,996	1,521	3,243	1,758
company name change	9,433	9,132	15,619	6,910
spin off	5,247	5,094	8,319	4,798
marriage	342,895	335,313	557,478	176,949
sibling relationship	167,611	160,893	255,788	69,596
romantic relationship	155,335	152,878	229,393	74,895
person parent	192,610	186,834	390,878	119,238
average of 39 relations	66,545	66,509	109,435	41,620

Problems of Large-Scale Approach

- Very low precision
  - a lot of noisy rules
  - many rules are learned from more than one relation

Euler Diagram for Four People-Relations



Various Filtering Strategies for High-Performance Web-Scale RE

Frequency-Driven Rule Filters

- Merged Filter:

$$valid_m^{\mathcal{R}}(r) = valid_{freq}^{\mathcal{R}}(r) \wedge valid_{inter}^{\mathcal{R}}(r)$$

1) **absolute frequency filtering**: a threshold to exclude rules with low occurrency

Rule Frequency Driven Filters

- Merged Filter:

$$valid_m^{\mathcal{R}}(r) = valid_{freq}^{\mathcal{R}}(r) \wedge valid_{inter}^{\mathcal{R}}(r)$$

1) **absolute frequency filtering**: a threshold to exclude rules with low occurrency
2) **inter-relation filter (Overlap Filter – FO Filter):**
   - based on mutual exclusiveness of relations with similar entity-type signatures.
   - a rule is only valid for a relation, if its relative frequency is higher than any other relations with similar entity type signatures.

$$valid_{inter}^{\mathcal{R}}(r) = \begin{cases} true & \text{if } \forall \mathcal{R}' \in \mathbb{R} \backslash \{\mathcal{R}\} : rf_{r,\mathcal{R}} > rf_{r,\mathcal{R}'} \\ false & \text{otherwise} \end{cases}$$

Weakness of Filtering with Frequency

- Undetected low-quality patterns:
  - high frequency in target relation, low frequency in coupled relations

Weakness of Filtering with Rule Frequency

- Undetected low-quality patterns:
  - high frequency in target relation, low frequency in coupled relations

✗ ?

✔ ?

Weakness of Filtering with Rule Frequency

- Undetected low-quality patterns:
  - high frequency in target relation, low frequency in coupled relations

✗ ?

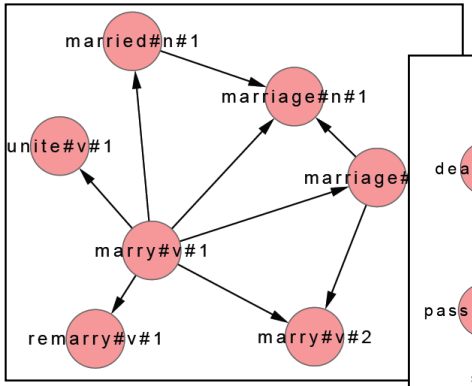
- Erroneously-deleted good patterns:
  - infrequent patterns

✔ ?

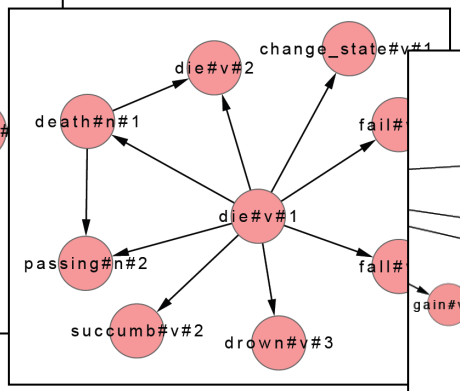
Lexical Semantics can help!

World Wide Web

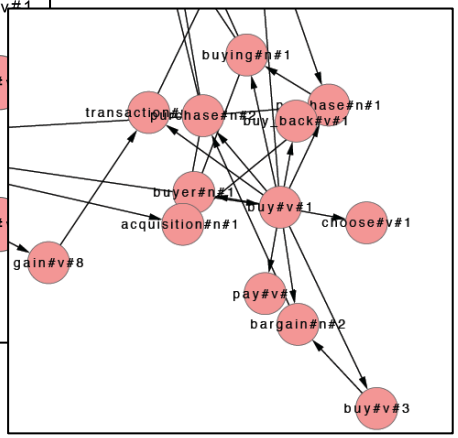
marriage



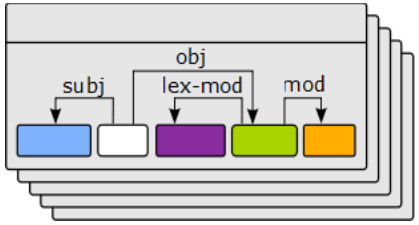
person-death



acquisition

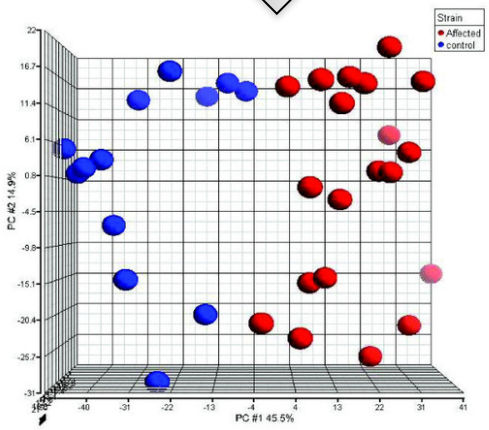


Relation-specific lexical semantic graphs

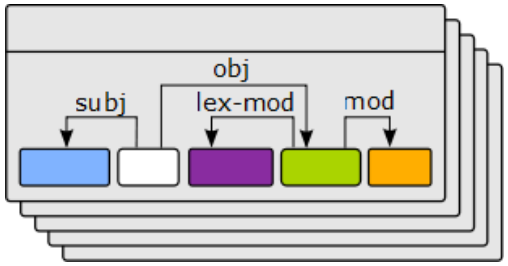


✔ ? ✗ ?

Candidate RE Patterns



Unsupervised Classification

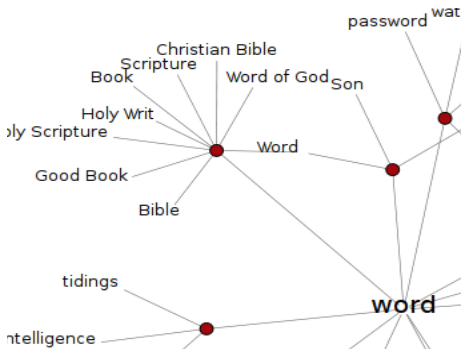


✔ !!!

High-quality RE Patterns

Automatic learning of relation-specific lexical semantic network

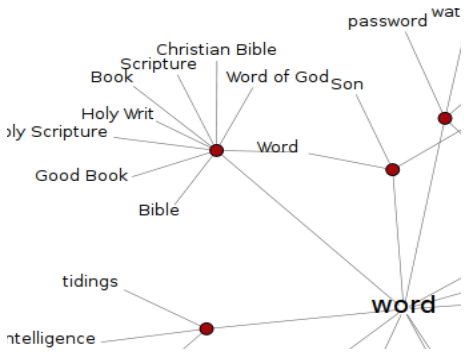
Generic Lexical Semantic Network (BabelNet)



automatically learned unfiltered RE rules and their mentions

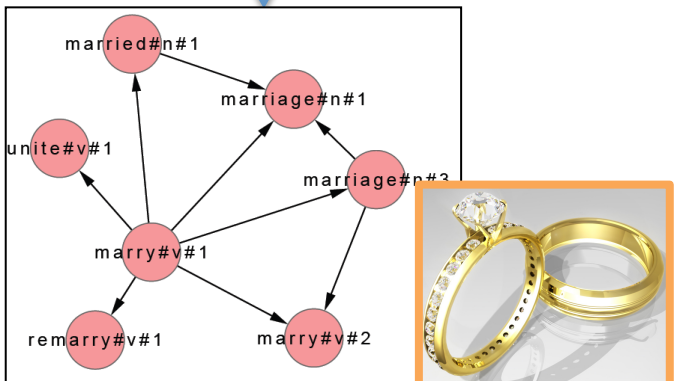
Automatic learning of relation-specific lexical semantic network

Generic Lexical Semantic Network (BabelNet)



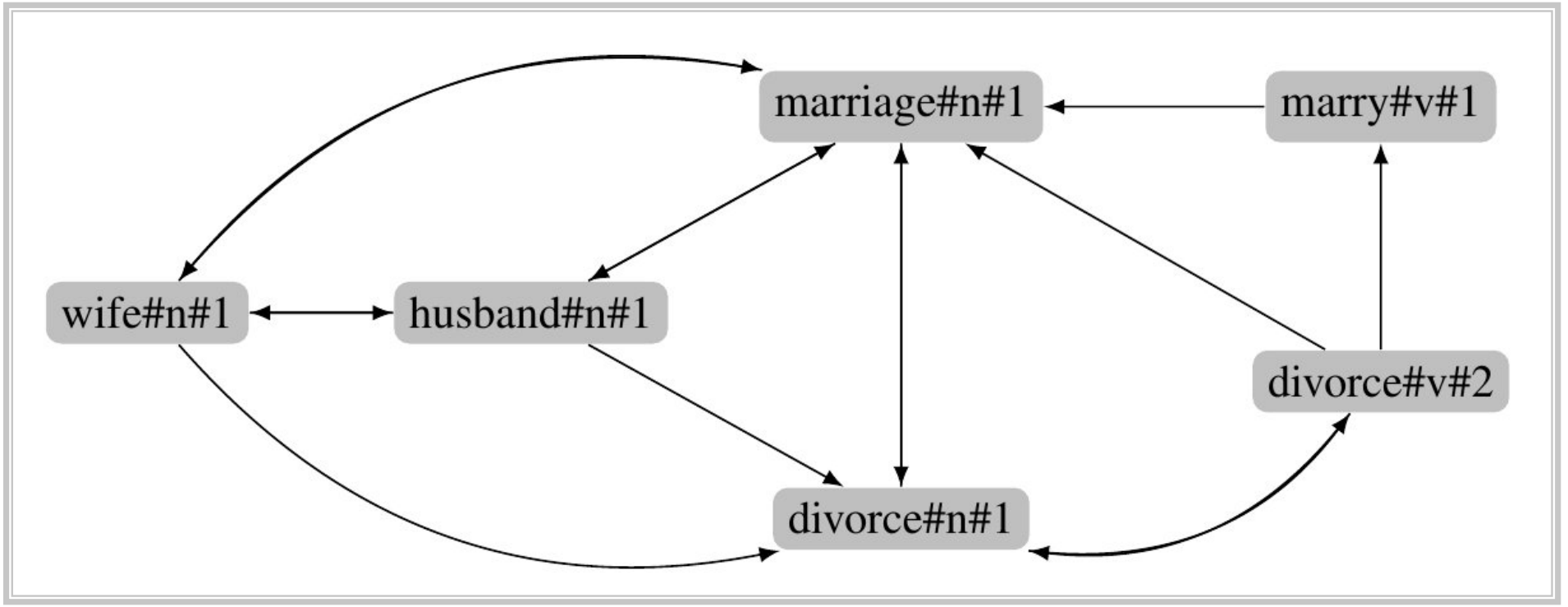
automatically learned unfiltered RE rules and their mentions

Word Sense Disambiguation

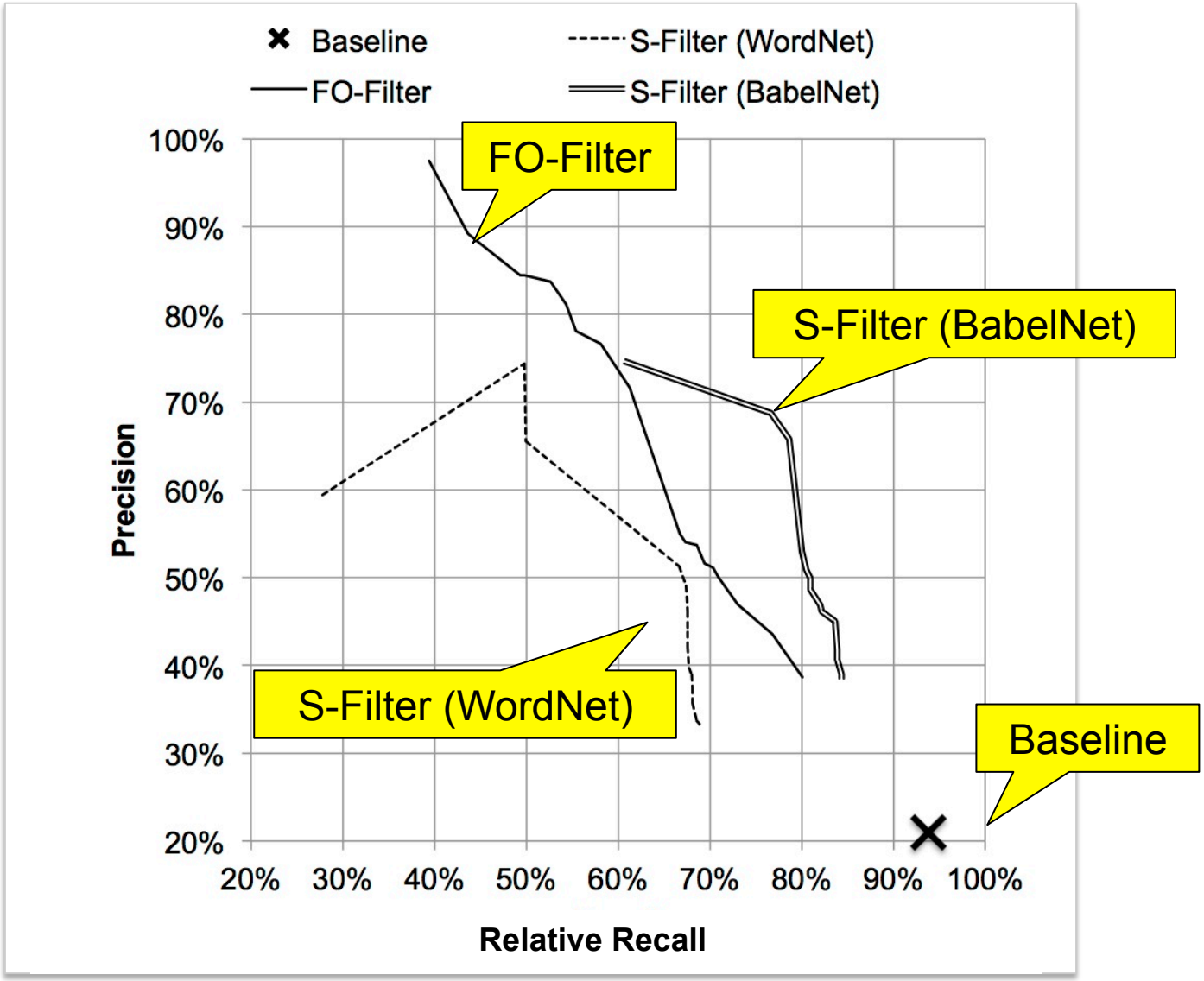


The Relation-Specific Semantic Graph

An excerpt of the semantic graph for the relation *marriage*



Extrinsic Eval. – Web-DARE



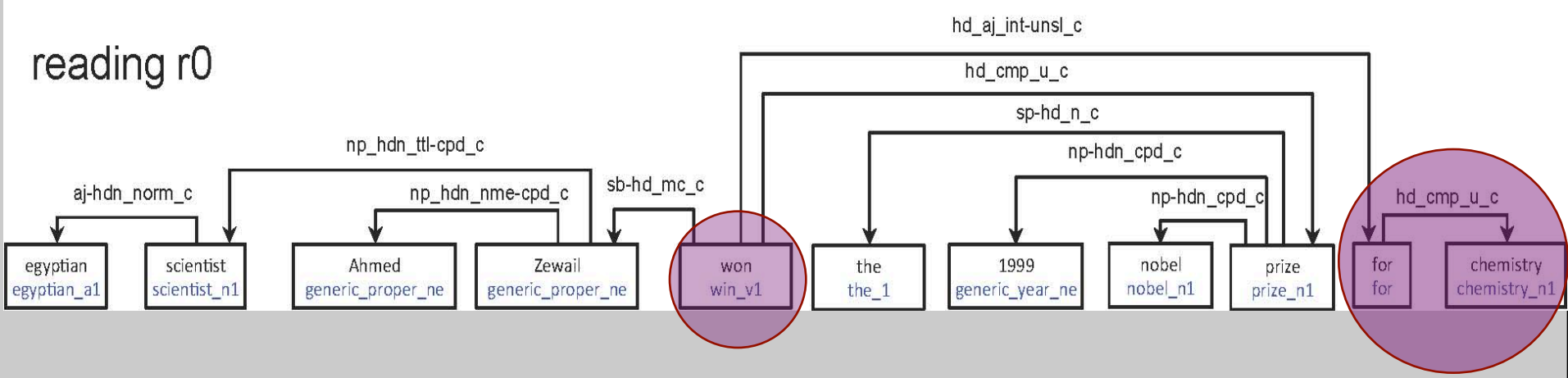
Parse-Reranking for Domain-adaptive RE

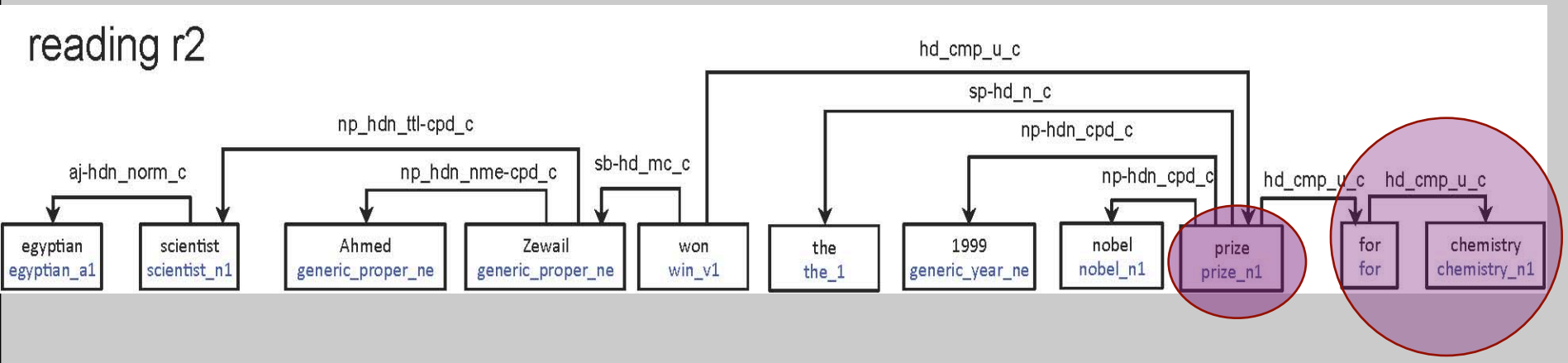
Error Types of Extracted Wrong Instances

Content	Modality	Named Entity Recognition (NER)	Parsing	NER & Parsing	DARE Rules
11.8%	17.6%	5.9%	38.2%	11.8%	14.7%

HPSG Parses: PP Attachment

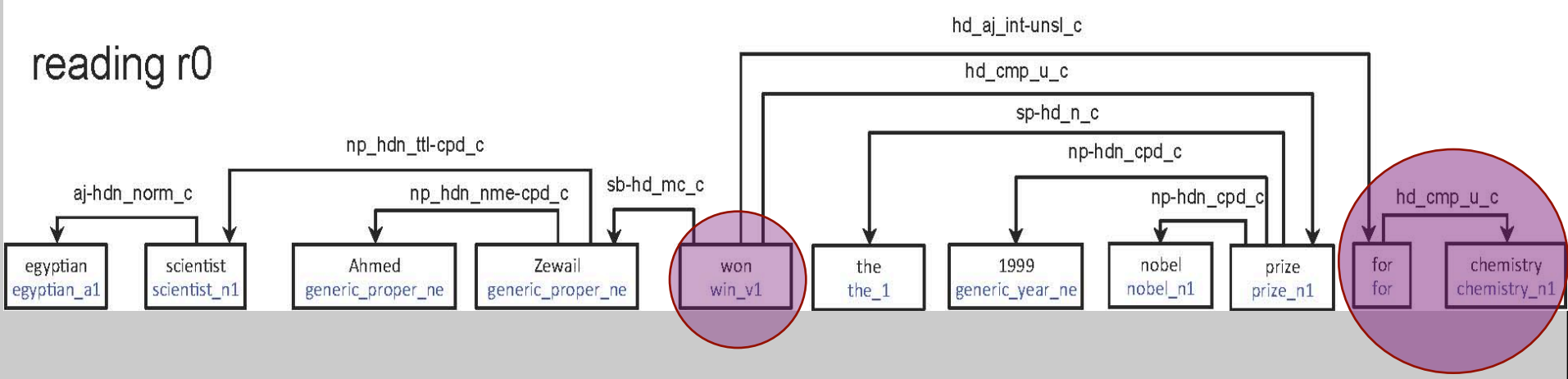
Egyptian scientist Ahmed Zewail won the 1999 Nobel Prize for chemistry

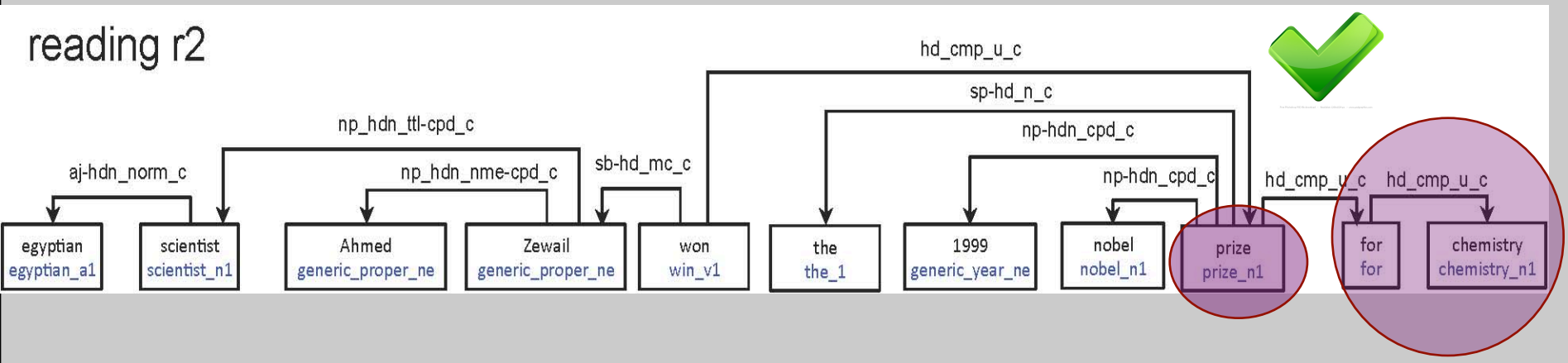




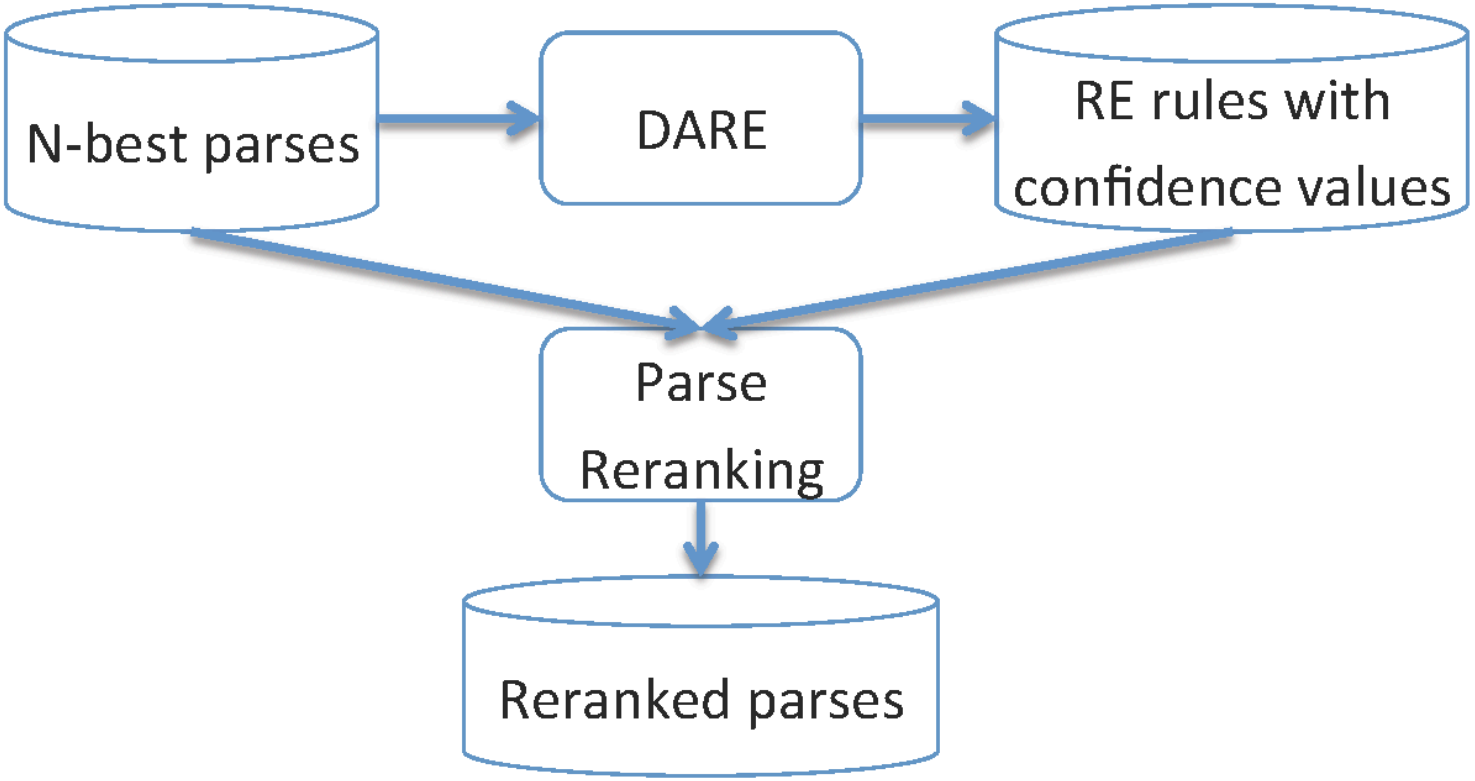
HPSG Parses: PP Attachment

Egyptian scientist Ahmed Zewail won the 1999 Nobel Prize for chemistry



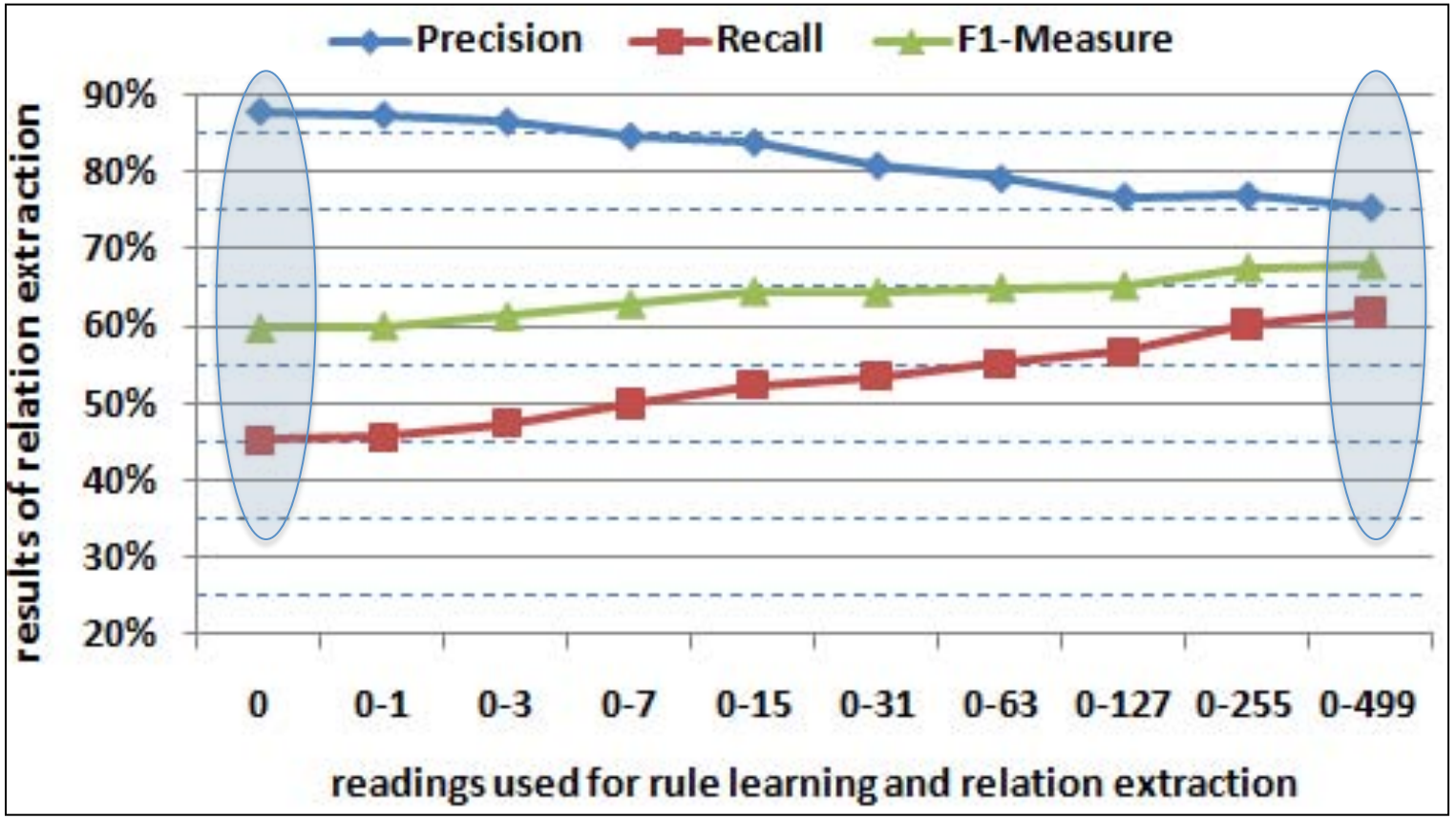


Reranking Architecture



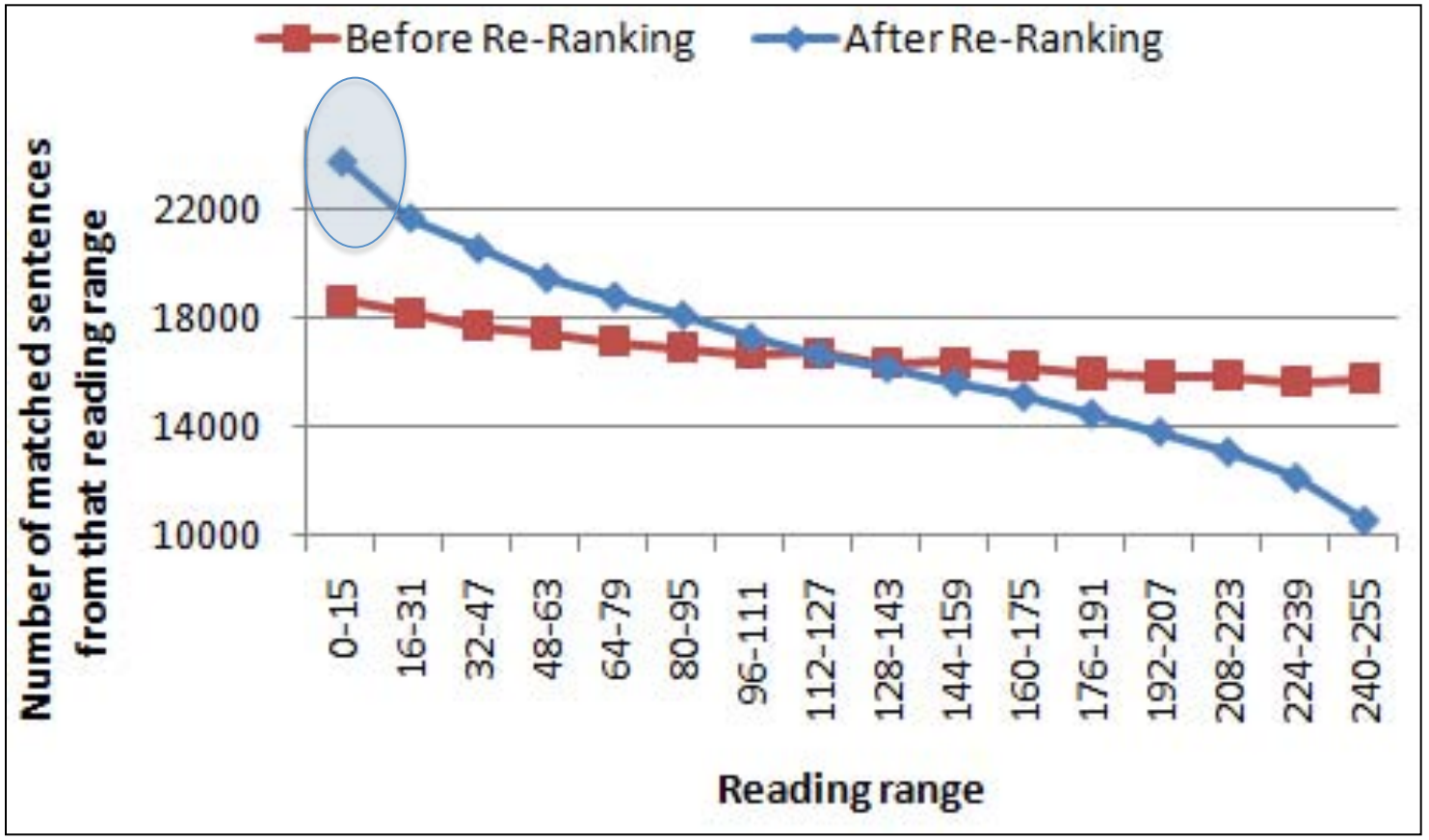
Baseline: before Re-ranking

- Best reading: high precision, low recall, low F-measure
- **500 readings**: lower precision, higher recall, higher F-measure



After Re-Ranking:

- Re-ranked top readings match more sentence mentions containing RE instances
- Improvements of Recall and F-Measure



Conclusion

- The performance of large-scale RE for each application is dependent on the performance of domain-adaptation methods
- Three original contributions (among others):
  - Extension of relation extraction to n-ary relations
  - Semantic filtering with large lexical knowledge bases
  - Parser improvement for the specific RE task by reranking
- For our work we received a Google Focused Research Award