Takuya Matsuzaki
Univ. of Tokyo

Takashi Ninomiya
Univ. of Tokyo

Kenji Sagae
Univ. of Tokyo

Tadayoshi Hara
Univ. of Tokyo

Yusuke Miyao
Univ. of Tokyo

Combining Statistical Models with Symbolic Grammar in Parsing

Research on Advanced Natural Language Processing and Text Mining: aNT
Grant-in-Aid for specially promoted research, MEXT (2006-2011)

Junichi TSUJII
Univ. of Tokyo
Univ. of Manchester

IEEE ASRU Kyoto, Dec 11th, 2007

Sentence Parsing

Nerd

IT Businessman

Parsing based on a proper linguistic formalism is one of the core research fields in CL and NLP.

It was considered as a monolithic, esoteric and inward looking field,
largely dissociated from real world application.

The field has matured, ready to be used by applications.

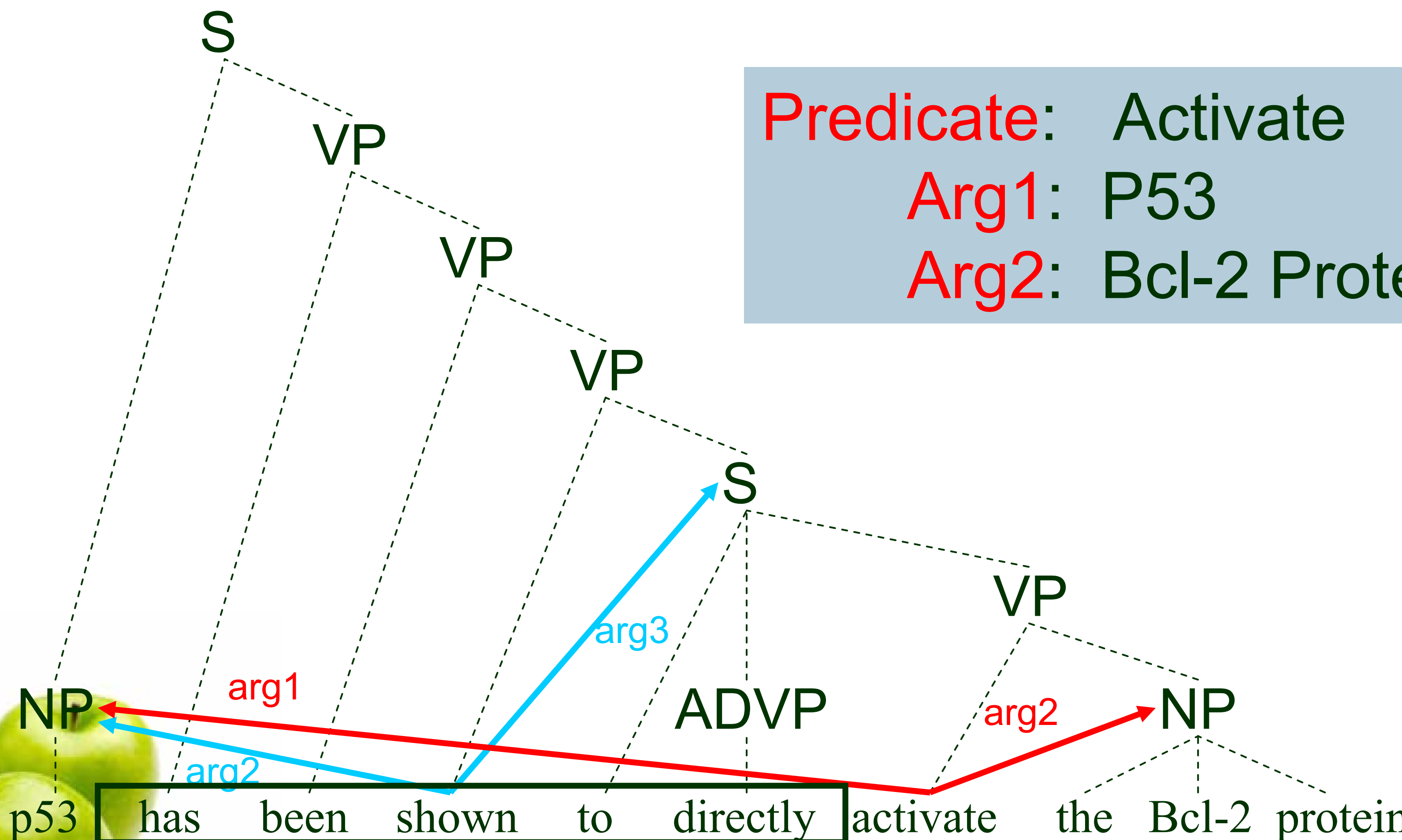
Integration of linguistic grammar formalisms with statistical models.

Robust, efficient and
open to eclectic sources of information other than syntactic ones

……. Speech Understanding
Speech/Text Retrieval ………

Deep parser

which produces semantic representation



Predicate: Activate
Arg1: P53
Arg2: Bcl-2 Protein

http://www-tsujii.is.s.u-tokyo.ac.jp/medie/search.cgi?search_type=semantic_search&subject=p53&verb=activate&base_form=verb&ontology=

Semantic Search | Keyword Search | GCL Search

subject: p53 | verb: activate | object: | Search! | Clear | Help

»Advanced search

Sentence Retrieval System Using Semantic Representation MEDIE

Results **1-50** for **p53 activate** »Show next »Show query

1. PMID: 15446548 »XML

   The molecules activated by p53 induce apoptosis , cell cycle arrest , and DNA repair to conserve genome .

2. PMID: 15273740 »XML

   In this report , we demonstrated that human AMID gene promoter was activated by p53 in reporter gene assays

Passive

3. PMID: 15020844 »XML

   Recently , p53 has been shown to directly activate the pro-apoptotic Bcl-2 protein ... apoptosis .

Passive and Infinitival Clause

4. PMID: 15105421 »XML

   Electrophoretic mobility shift assays reveal that both transcription factors are capable of binding to putative consensus sites , and luciferase reporter assays reveal that E2F1 and p53 can activate transcription from the SIVA promoter .

5. PMID: 15247038 »XML

   Although the role of the nuclear factor-kappa B (NF-kappa B) signaling cascade is crucial in ICAM-1 activation , we have shown that p53 directly activates the expression of ICAM-1 in an NF-kappa B-independent manner .

6. PMID: 15021899 »XML

   Because the MDM2 gene is transcriptionally activated by p53 , it forms part of an autoregulatory feedback loop that directly links the transcriptional activity of p53 with its degradation .

7. PMID: 15064739 »XML

ル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り

ス(D) http://nactem2.mc.man.ac.uk/medie/search.cgi?search_type=semantic_search&subject=p53&verb=activate&base_form=verb&ontology= 移動 リン

MEDIE — See what causes cancer?

MEDIE is a demo system presented by Tsujii Laborator

emantic Search | **Keyword Search** | **GCL Search**

subject	*verb*	*object*			
p53	activate		Search!	Clear	Help

lvanced search

Show	50 results	Help
Output format	sentence article table	Help
Keywords		Help
Modifiers	not	Help
Base form	*subject* *verb* *object* *keyword*	Help
Ontology	*subject* *verb* *object*	Help
Category	*subject* any *object* any	Help

sults **1-50** for **p53 activate** »Show next »Show query 35.82 seconds (5.37% finished

. PMID: 11212267 »XML

KAI1/CD82 has been shown to be a metastasis suppressor for several human cancers , and a recent study revealed that wild-type tumor suppressor p53 can directly activate KAI1/CD82 gene expression .

2. PMID: 11162500 »XML

However , in an in vitro transcription assay with partially purified basal transcription factors , p53 only partially activated transcription from the same binding site and required PAb421 for full activation .

3. PMID: 10521394 »XML

インターネット

ル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り

http://nactem2.mc.man.ac.uk/medie/search.cgi?search_type=semantic_search&subject=p53&verb=activate&base_form=verb&ontology= 移動

Category subject any object any Help

sults **1-11** for **p53 activate** »Show query 100.67 seconds (100.00% finished

1. PMID: 11483599 »XML

We demonstrated that mutant p53 did not activate either the MRP1 promoter or the endogenous gene .

2. PMID: 12019172 »XML

However , luciferase constructs driven by the HDAC5 promoter containing three to six potential binding sites were not activated by p53 , nor was the expression of HDAC5 mRNA induced by p53-activating agents .

3. PMID: 12048243 »XML

This activation occurred by a phosphorylation-independent mechanism involving direct binding of GSK3beta to p53 , which was confined to the nucleus where p53 is localized , and mutated p53 (R175H) bound but did not activate GSK3beta .

4. PMID: 14557665 »XML

Thus , it is likely that the E1B 55-kDa protein sequesters Daxx and p53 in specific nuclear locations , where p53 can not activate transcription .

5. PMID: 14517211 »XML

The DDATHF-stabilized p53 bound to the p21 promoter in vitro and in vivo but did not activate histone acetylation over the p53 binding sites in the p21 promoter that is an integral part of the transcriptional response mediated by the DNA damage pathway .

6. PMID: 8632013 »XML

Two monoclonal antibodies to the N terminus of p53 , PAb1801 and DO-1 , do not activate the latent DNA binding function of p53 but can protect the p53 wild-type conformation at 37 degrees C .

7. PMID: 9159467 »XML

RNA polymerase II transcriptional activators , like GAL4 , VP16 or p53 , fused to GAL4 DNA-binding domain , did not activate the UAS (G) -SNR6 gene .

8. PMID: 9360984 »XML

Internet Explorer ブラウザの起動
インターネット上の情報および Web サイトを検索して表示します。

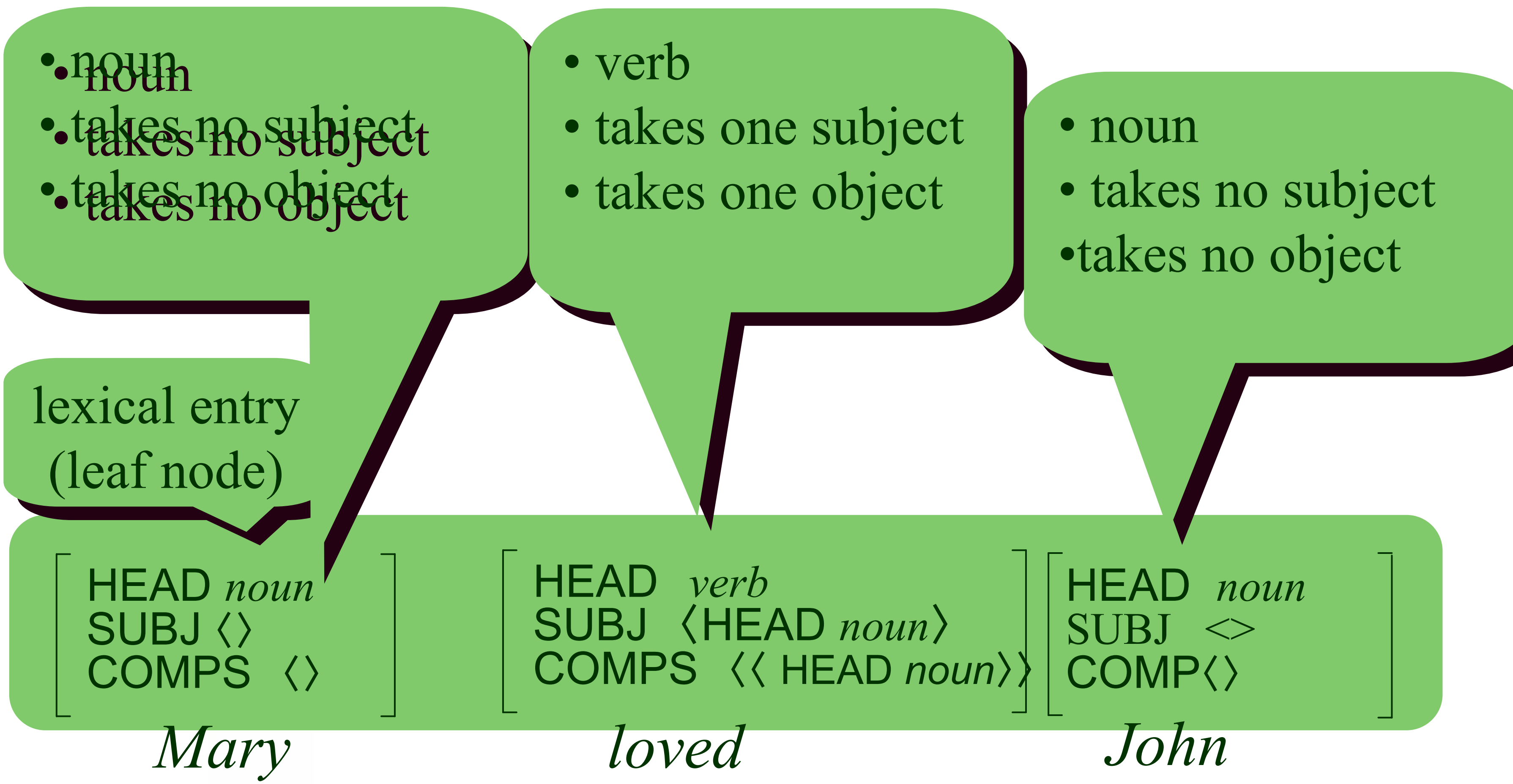
インターネット

Grammar Formalism: HPSG

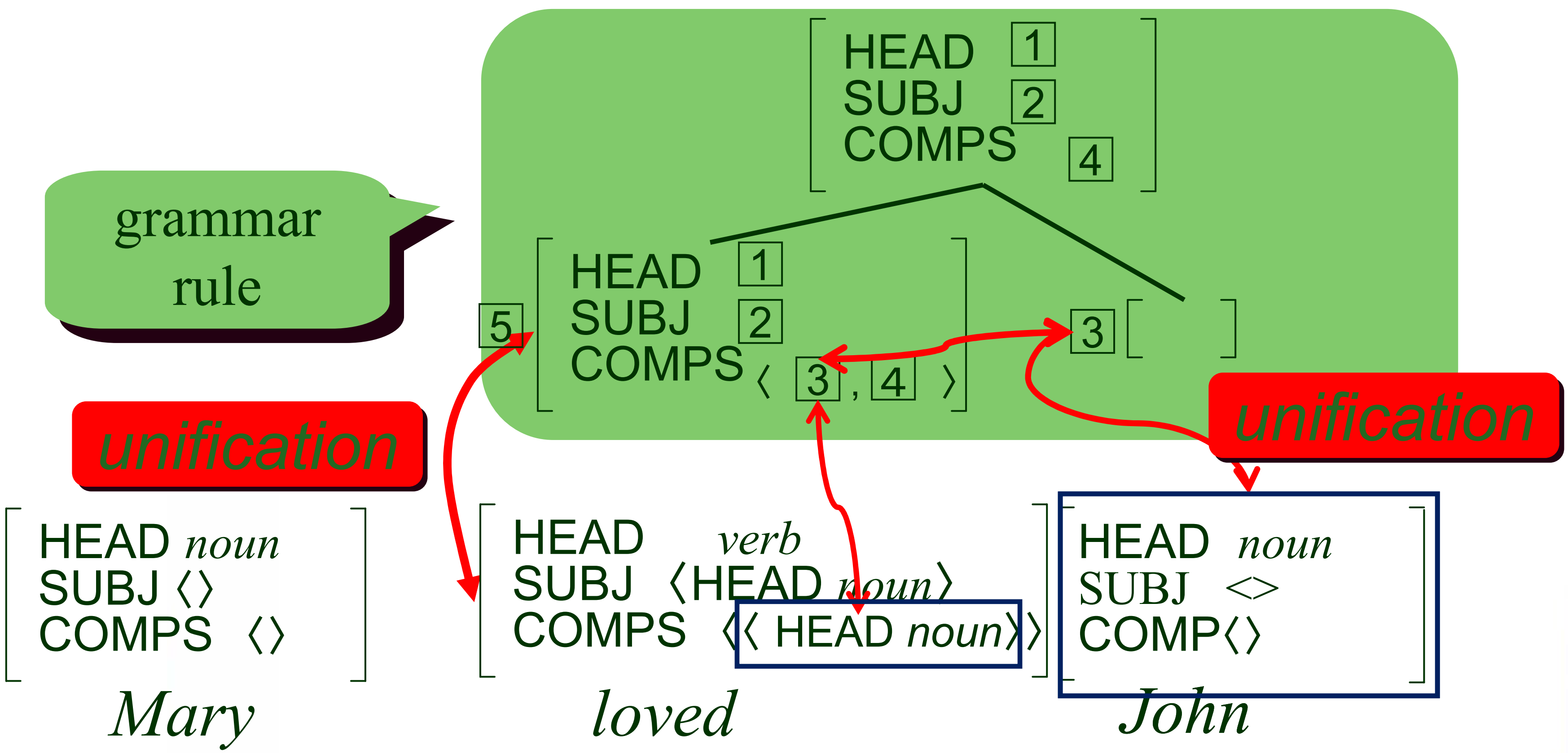
HPSG

- HPSG = Lexical entries + Grammar rules
  - Lexical entries: syntactic and semantic descriptions of word-specific behaviors
    - c.f. Enju grammar (Miyao et al 2004) has 3797 lexical entries for 10,536 words
  - Grammar rules: non-word-specific syntactic and semantic configurations
    - c.f. Enju grammar has 12 grammar rules

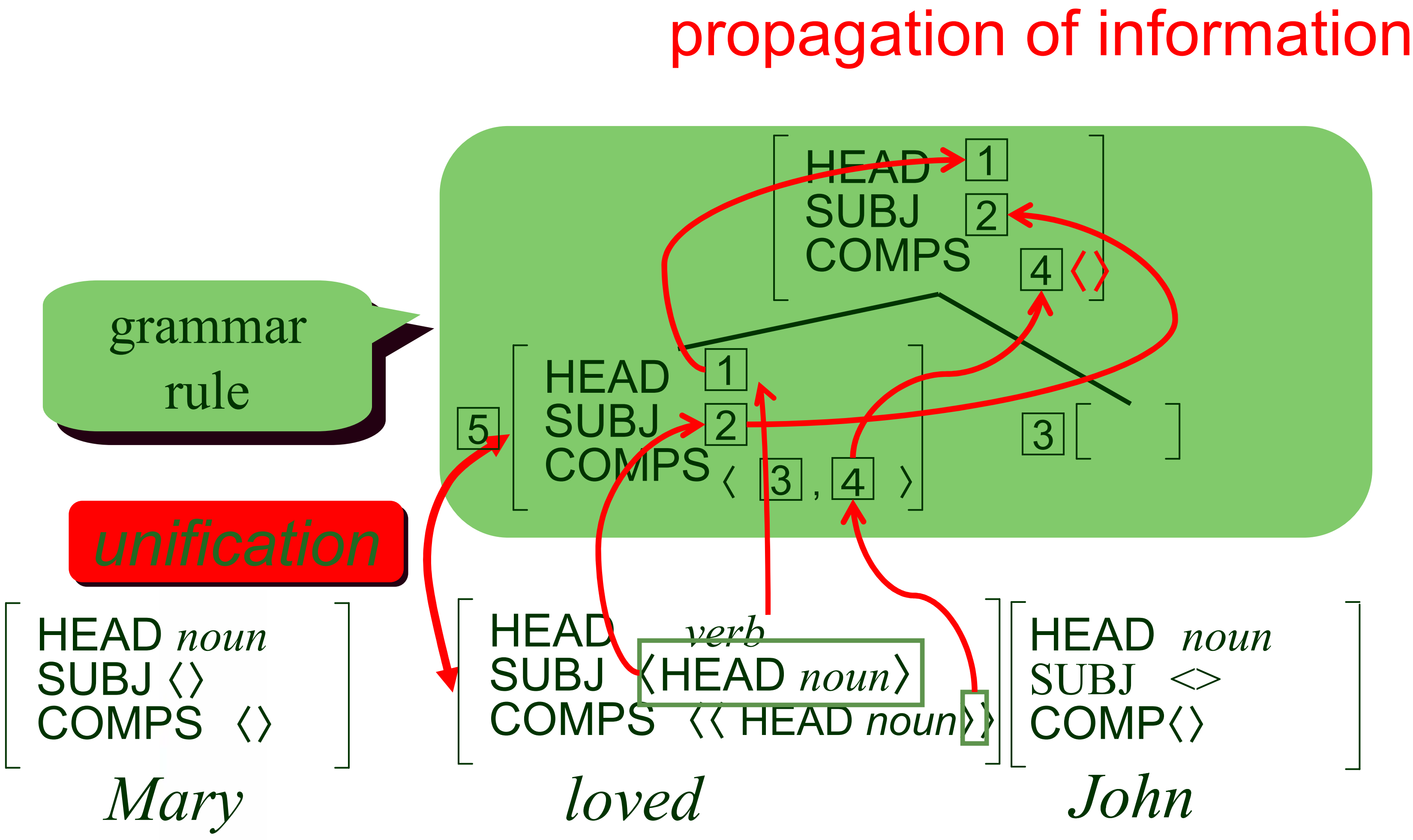
HPSG: Parsing



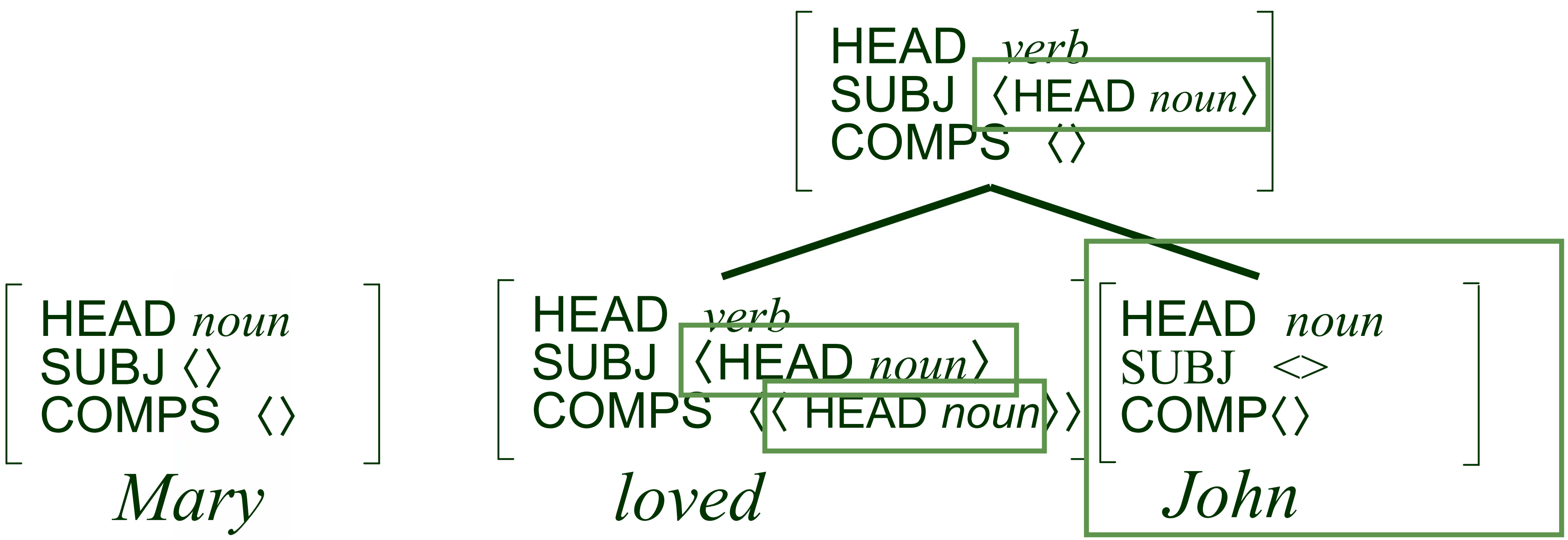
HPSG: Parsing



HPSG: Parsing

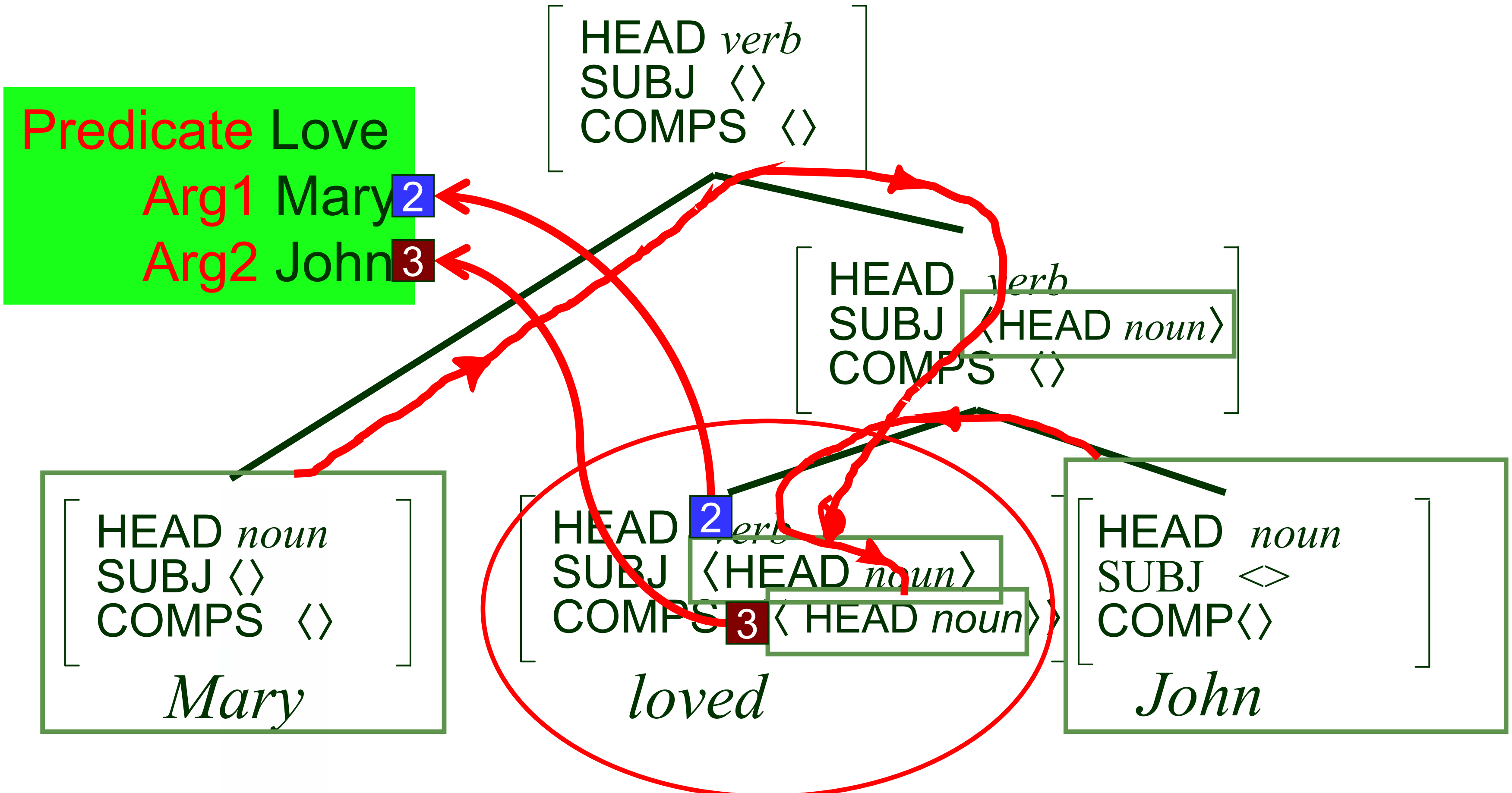


HPSG: Parsing



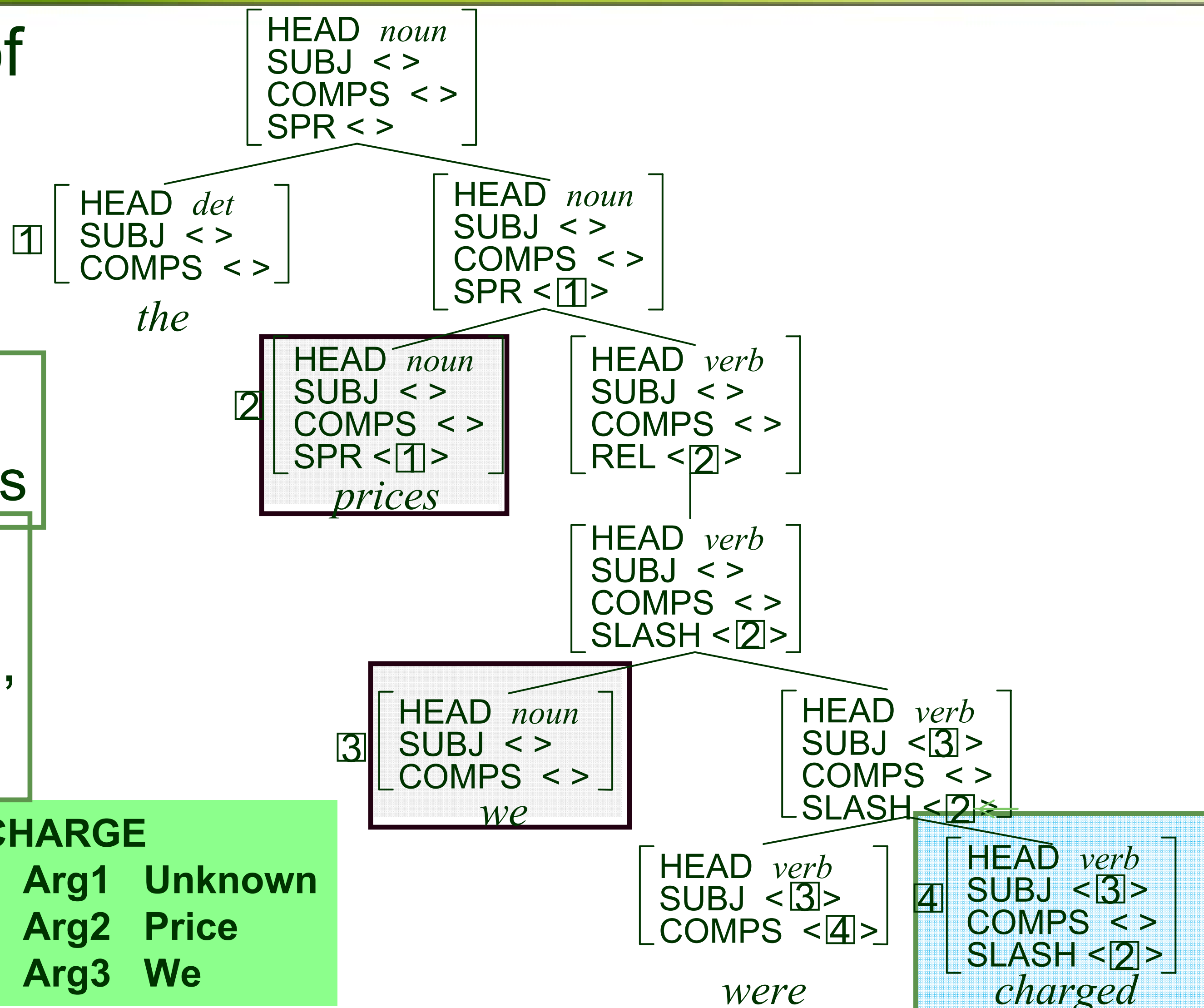
HPSG: Parsing

- A parse tree is derived by applying grammar rules recursively



HPSG: Parsing

- An example of a complex syntactic tree
  - SLASH, REL features explain non-local dependencies
  - WH movement, topicalization, relative clauses



CHARGE

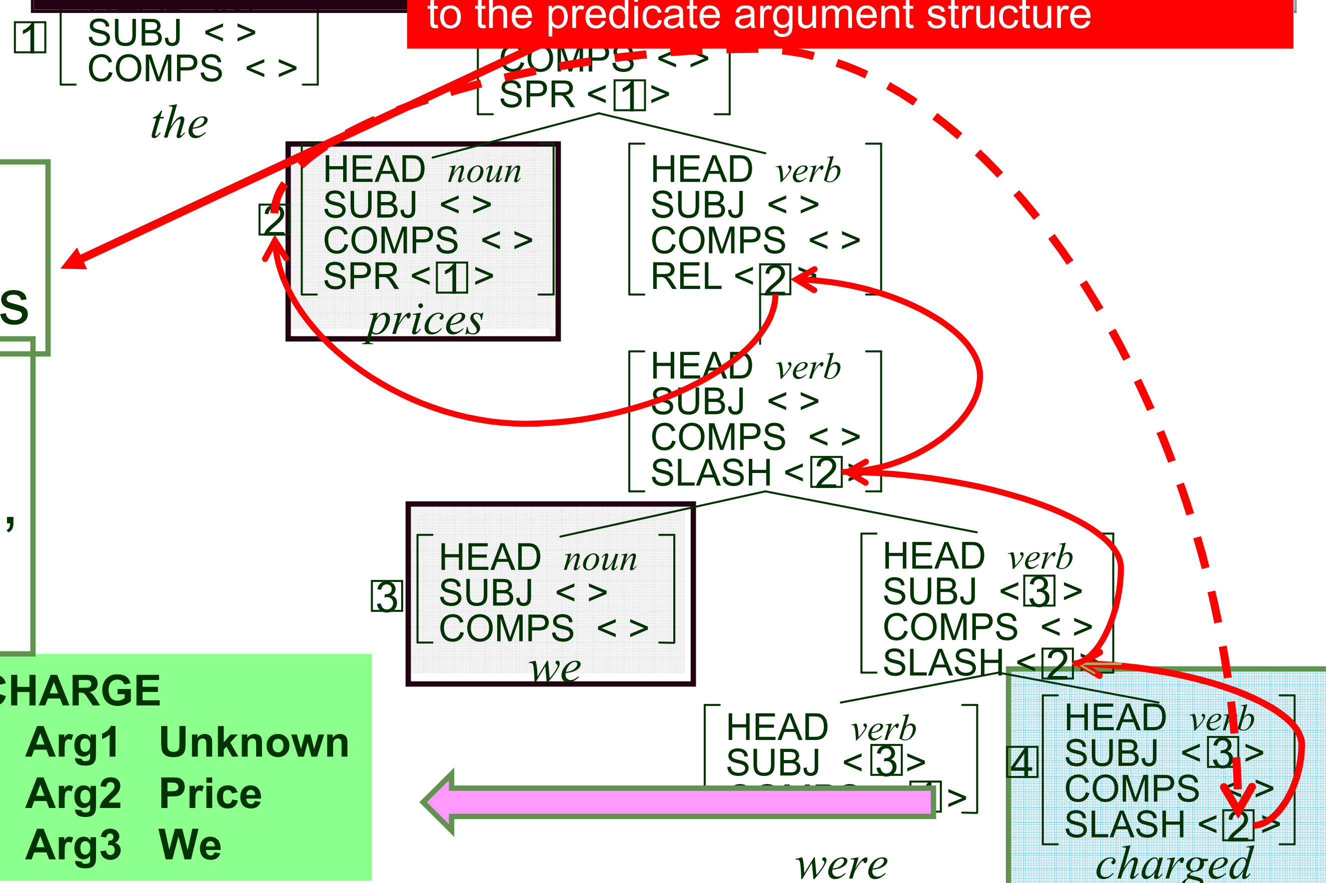
Arg1	**Unknown**
Arg2	**Price**
Arg3	**We**

HPSG: Parsing

- An example of a complex syntactic tree
  - SLASH, REL features explain non-local dependencies
  - WH movement, topicalization, relative clauses

The information is mostly written in a lexical entry

Mapping a syntactic tree
- passive in relative clause construction -
to the predicate argument structure



CHARGE
Arg1 Unknown
Arg2 Price
Arg3 We

- HPSG parsing (Pollard & Sag 1994)
  - Mathematically well-defined with sophisticated constraint-based system
  - Linguistically justified
  - Deep syntactic grammar that provides semantic analysis

10 years ago

Unrealistic solutions
for real-world text,
let alone real world speech

Combining HPSG with Statistical Models

Difficulties in HPSG Parsing

- Difficulty of developing a broad-coverage HPSG grammar
- Difficulty of disambiguation
  - No treebank for training an HPSG grammar
  - No probabilistic model for HPSG
- Efficiency
  - Very slow : CFG filtering, Efficient search, Feature Forest

Difficulties in HPSG Parsing

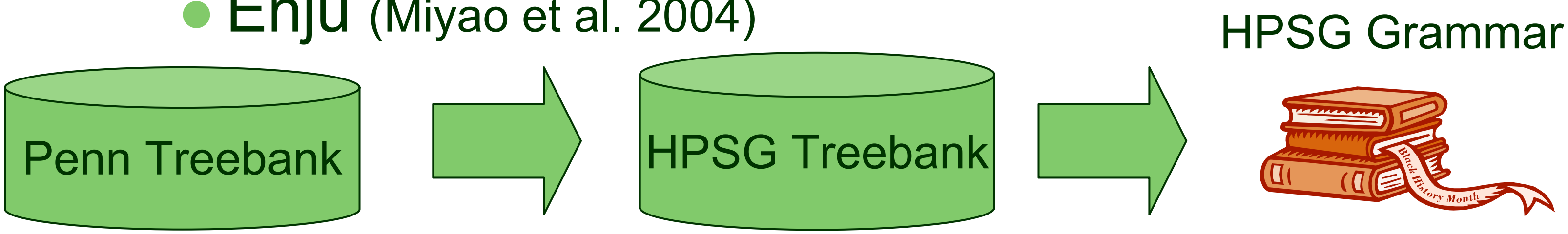
- Difficulty of developing a broad-coverage HPSG grammar

- Difficulty of disambiguation
  - No treebank for training an HPSG grammar
  - No probabilistic model for HPSG

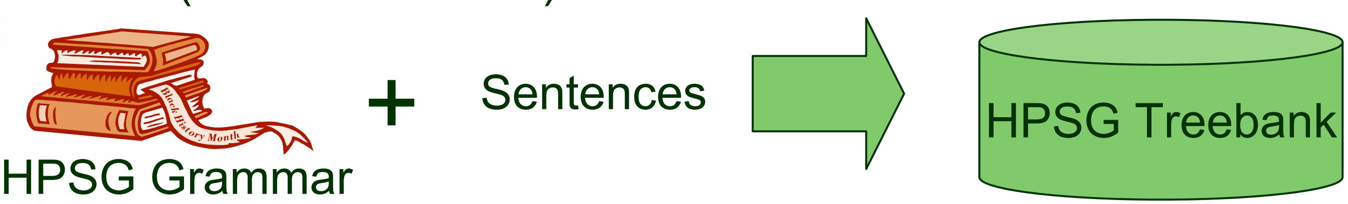
- Efficiency
  - Very slow : CFG filtering, Efficient search, Feature Forest

Grammar with Broad Coverage

- Treebank for Grammar development and evaluation
  - Treebank grammar
    - Enju (Miyao et al. 2004)



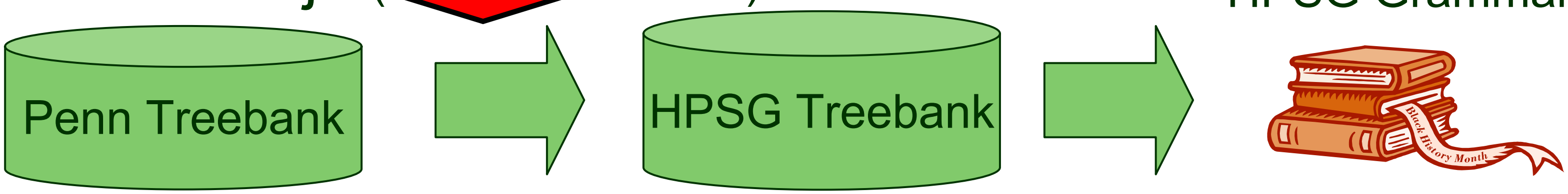
  - Treebank development
    - Redwood (Oepen et al. 2002)
    - Hinoki (Bond et al. 2004)



Grammar with Broad Coverage

- Treebank for Grammar development and evaluation
  - Tre ar
    - Enju (Mi et al. 2004)

Rule Application



Lexical Knowledge Acquisition

  - Treebank development
    - Redwood (Oepen et al. 2002)
    - Hinoki (Bond et al. 2004)

Performance of Semantic Parser

	Penn Treebank	GENIA
Coverage	[redacted]	
F-Value (PArelations)	87.4%	86.4%
Sentence Precison	39.2%	31.8%
Processing Time	0.68sec	1.00sec

Difficulties in HPSG Parsing

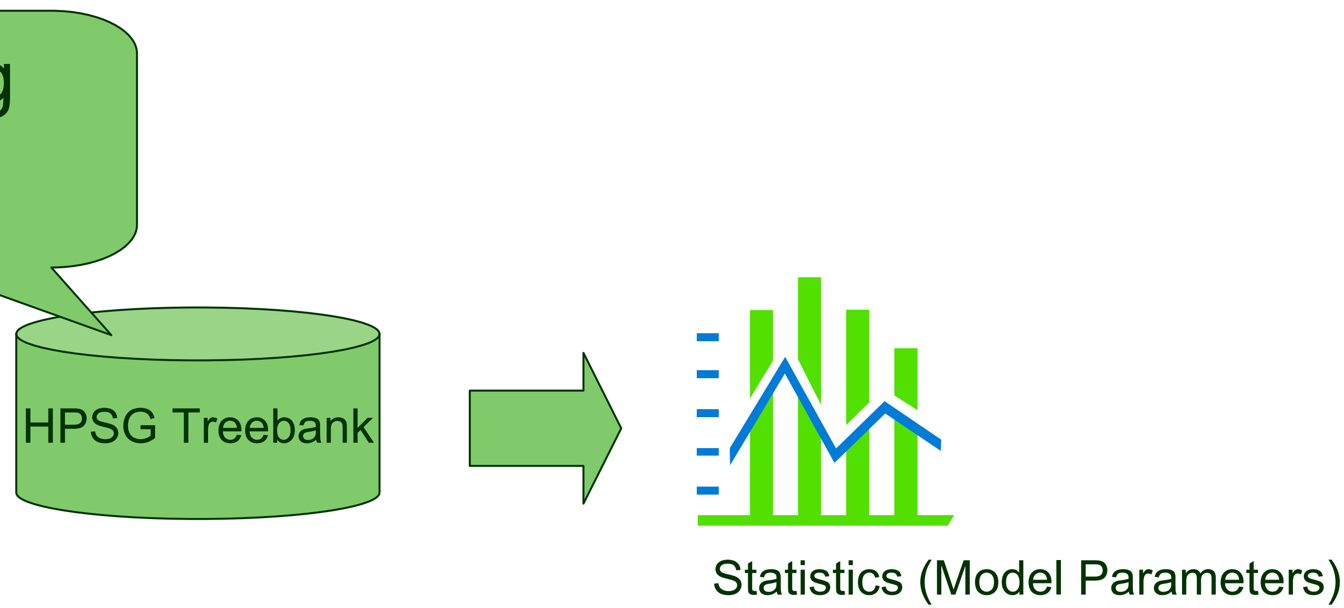
- Difficulty of developing a broad-coverage HPSG grammar

- Difficulty of disambiguation
  - No treebank for training an HPSG grammar
  - No probabilistic model for HPSG

- Efficiency
  - Very slow : CFG filtering, Efficient search, Feature Forest

Probabilistic Model and HPSG

- Probabilistic model
  - Log-linear model for unification-based grammars (Abney 1997, Johnson et al. 1999, Riezler et al. 2000, Miyao et al. 2003, Malouf and van Noord 2004, Kaplan et al. 2004, Miyao and Tsujii 2005)

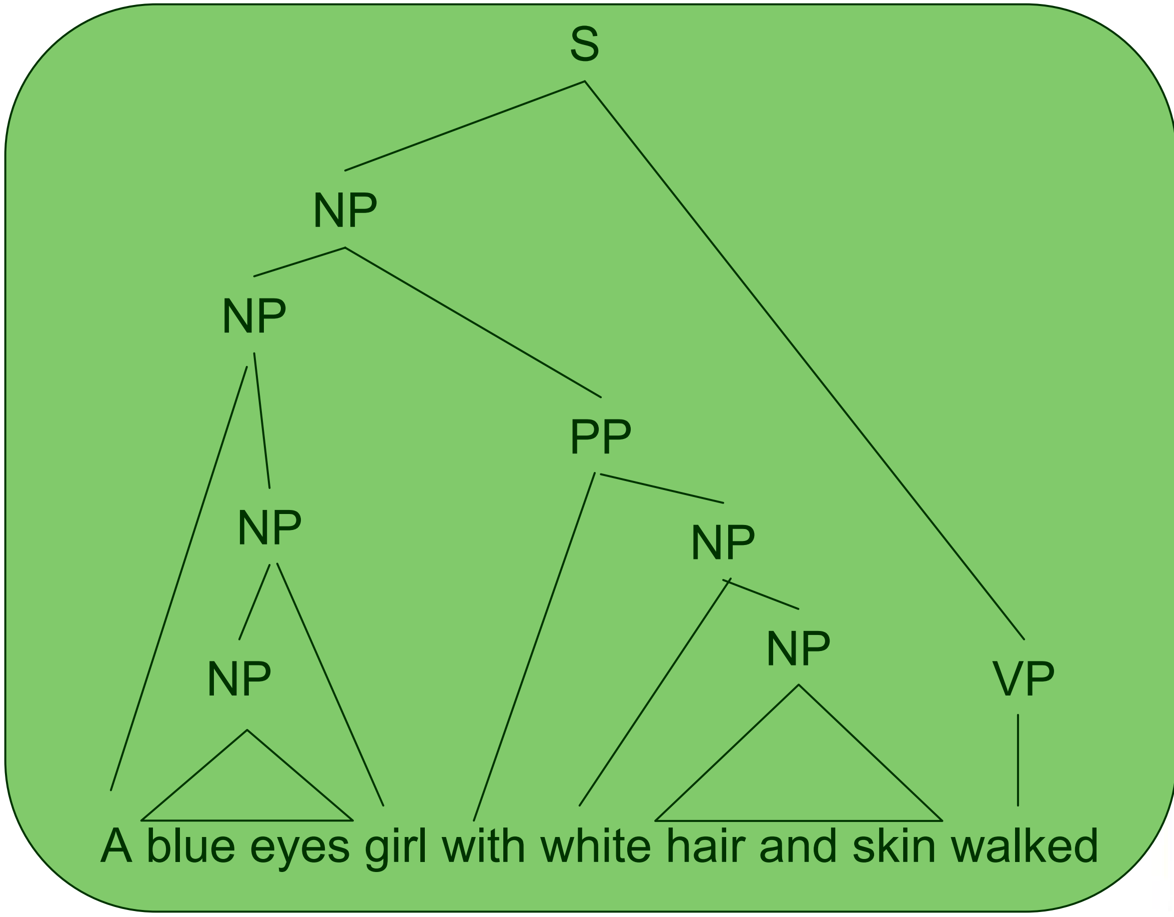


Probabilistic HPSG

$$p(T \mid \mathbf{w})$$

$\mathbf{w} =$ "A blue eyes girl with white hair and skin walked"

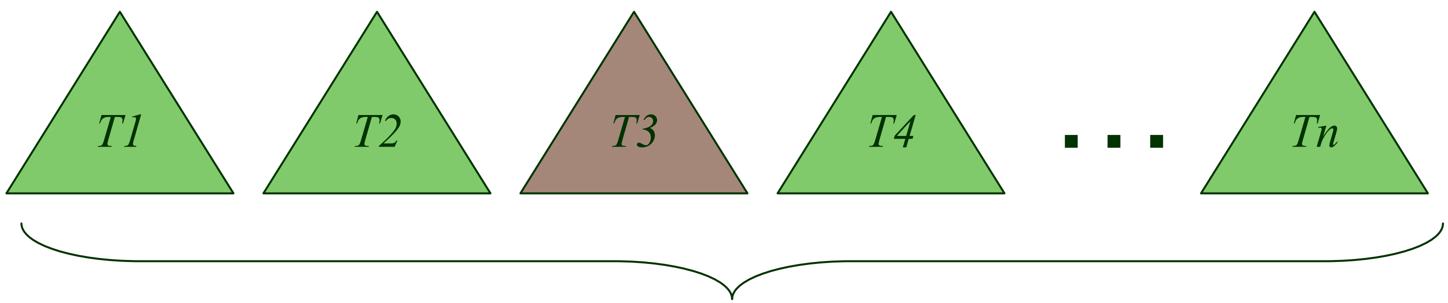
$T =$



Probabilistic HPSG

$$p(T \mid \mathbf{w})$$

$\mathbf{w} =$ "A blue eyes girl with white hair and skin walked"



All possible parse trees derived from $\mathbf{w}$ with a grammar

$p(T3|\mathbf{w})$ is the probability of selecting $T3$ from $T1$, $T2$, ..., and Tn.

Probabilistic HPSG

- Log-linear model for unification-based grammars
(Abney 1997, Johnson et al. 1999, Riezler et al. 2000, Miyao et al. 2003, Malouf and van Noord 2004, Kaplan et al. 2004, Miyao and Tsujii 2005)
  - Input: sentence $\mathbf{w}$
    - $\mathbf{w}=w_1/P_1, w_2/P_2, w_3/P_3,\ldots,w_n/P_n$
  - Output: parse tree T

word POS

feature function

$$p(T \mid \mathbf{w}) = \frac{1}{Z}\exp\Big(\sum_u \lambda_u f_u(T)\Big)$$

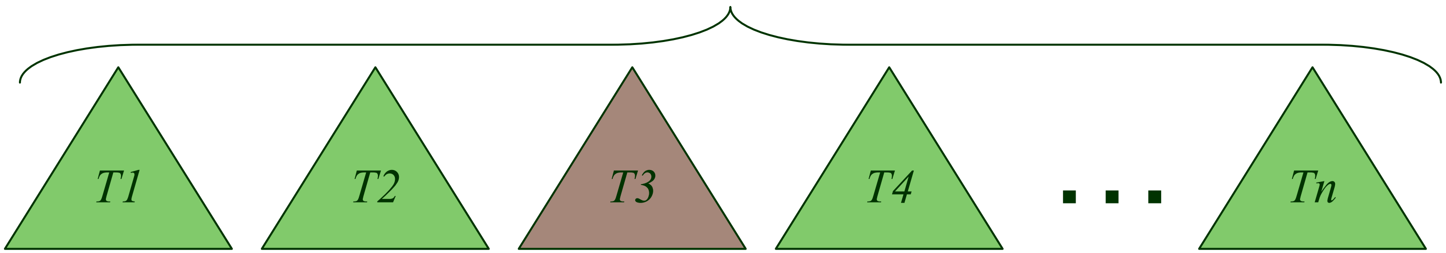
a weight for a feature function

normalization factor

Log-Linear Model
Maximum Entropy Model

$$p(T \mid \mathbf{w})$$ $\mathbf{w} =$ "A blue eyes girl with white hair and skin walked"

All parse trees derived from $\mathbf{w}$ with a grammar



$f1(T1)=1$	$f1(T2)=1$	$f1(T3)=1$	$f1(T4)=1$	$f1(Tn)=0$
$f2(T1)=0$	$f2(T2)=1$	$f2(T3)=1$	$f2(T4)=0$	$f2(Tn)=1$
$f3(T1)=0$	$f3(T2)=1$	$f3(T3)=0$	$f3(T4)=1$	$f3(Tn)=0$
…	…	…	…	…
$fm(T1)=1$	$fm(T2)=1$	$fm(T3)=0$	$fm(T4)=1$	$fm(Tn)=0$

feature functions are indicators that indicate the properties that the parse tree has.

Log-Linear Model
Maximum Entropy Model

$$p(T \mid \mathbf{w})$$

$\mathbf{w} =$ "A blue eyes girl with white hair and skin walked"

All parse trees derived from $\mathbf{w}$ with a grammar

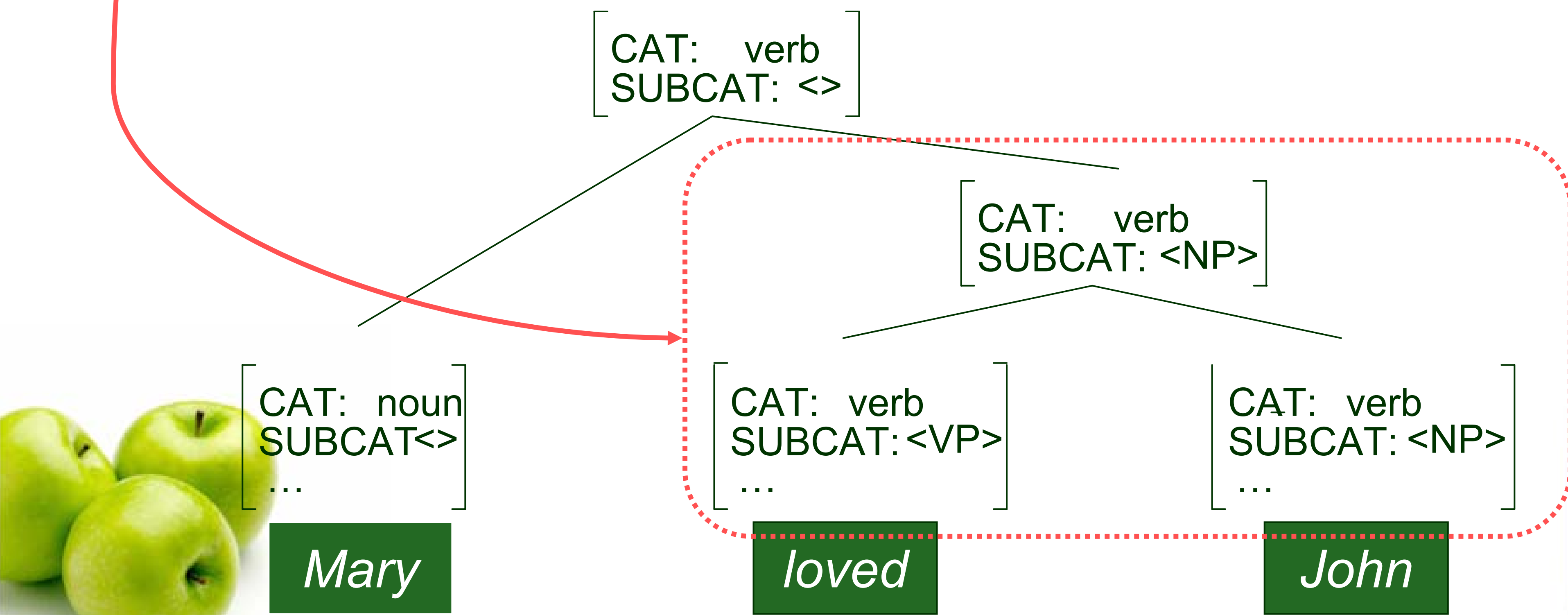
$$p(T \mid \mathbf{w}) = \frac{1}{Z}\exp\left(\sum_u \lambda_u f_u(T)\right)$$

$f1$				
$f2$				
$f3(T1)=0$	$f3(T2)=1$	$f3(T3)=0$	$f3(T4)=1$	$f3(Tn)=0$
...	...	...	...	...
$fm(T1)=1$	$fm(T2)=1$	$fm(T3)=0$	$fm(T4)=1$	$fm(Tn)=0$

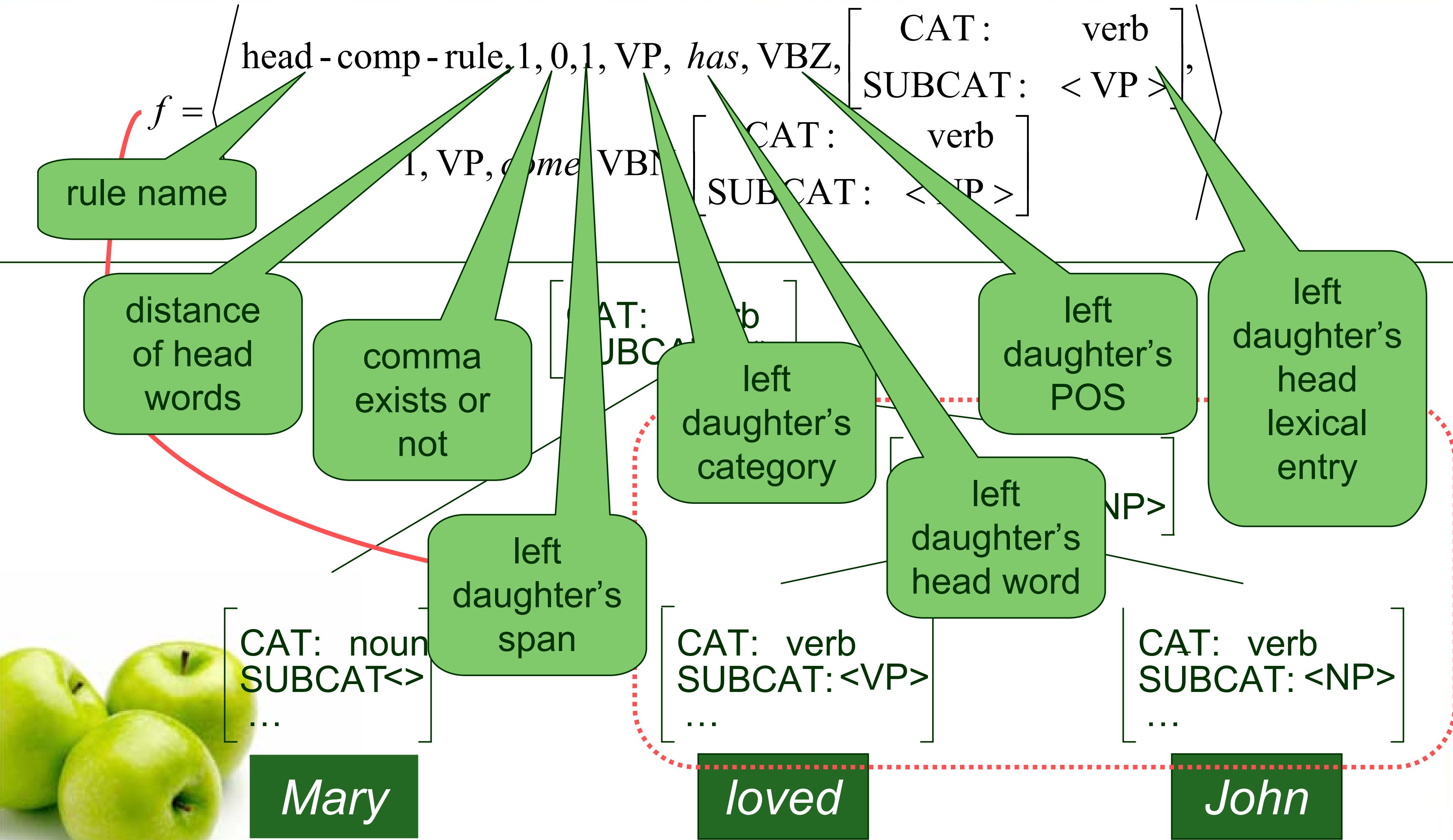
feature functions are indicators that indicate the properties that the parse tree has.

Example of Features in Probabilistic HPSG

$$f = \left\langle \begin{array}{c} \text{head-comp-rule}, 1, 0, 1, \text{VP}, \textit{has}, \text{VBZ}, \begin{bmatrix} \text{CAT}: & \text{verb} \\ \text{SUBCAT}: & <\text{VP}> \end{bmatrix}, \\ 1, \text{VP}, \textit{come}, \text{VBN}, \begin{bmatrix} \text{CAT}: & \text{verb} \\ \text{SUBCAT}: & <\text{NP}> \end{bmatrix} \end{array} \right\rangle$$



Example of Features in Probabilistic HPSG



Performance of Semantic Parser

	Penn Treebank	GENIA
Coverage	[redacted]	[redacted]
F-Value (PArelations)	87.4%	86.4%
Sentence Precison	39.2%	31.8%
Processing Time	0.68sec	1.00sec

Difficulties in HPSG Parsing

- Difficulty of developing a broad-coverage HPSG grammar
- Difficulty of disambiguation
  - No treebank for training an HPSG grammar
  - No probabilistic model for HPSG
- Efficiency
  - Very slow : CFG filtering, Efficient search, Feature Forest

lex_template
LEXEME_NAME "[npVPvp]_0"
LEXICAL_RULES
cons
hd singular3rd_verb_rule
tl nil

extracted for

word
INPUT string
SURFACE string
BASE "have"
INPUT_POS string
POS "VB"
BASE_POS string
POSITION integer

from

word
INPUT "has"
SURFACE "has"
BASE "have"
INPUT_POS [1] "VBZ"
POS [1]
BASE_POS "VB"
POSITION 3

hpsg_word
hpsg_synsem
hpsg_local
hpsg_cat
HEAD
hpsg_verb
AGR hpsg_3sg
ADJ hpsg_minus
VFORM verb_fin
AUX aux_have
PASSIVE hpsg_minus
TENSE tense_present
RELATIVE hpsg_binary
MODL hpsg_nolocal
MODR hpsg_nolocal
SUBJ
synsem_cons
hd
hpsg_synsem
LOCAL
[2] hpsg_local
CAT
hpsg_cat
HEAD
hpsg_noun
AGR hpsg_agreement
ADJ hpsg_binary
SPECIFIED hpsg_binary
MODL hpsg_nolocal
MODR hpsg_nolocal
SUBJ synsem_nil
COMPS synsem_nil
CONJ synsem_nil
WH hpsg_binary
CONT
hpsg_cont
HOOK [3] relation
RELS relation_list
MSG message
NONLOCAL hpsg_nonlocal
tl synsem_nil
CAT
synsem_cons

hpsg_lxm
LXM_SPEC
lxm_spec
DATIVE hpsg_minus
INSERT hpsg_minus
N_INV integer
N_PASSIVE integer
N_WH integer
hpsg_synsem
hpsg_local
hpsg_cat
HEAD
hpsg_verb
AGR hpsg_agreem
ADJ hpsg_minus
VFORM verb_base
AUX aux_have
PASSIVE hpsg_mi
TENSE hpsg_tens
RELATIVE hpsg_b
MODL hpsg_noloc
MODR hpsg_noloc
SUBJ
synsem_cons
hd
hpsg_synsem
LOCAL
[2] hp
CAT
CON

Chart parsing

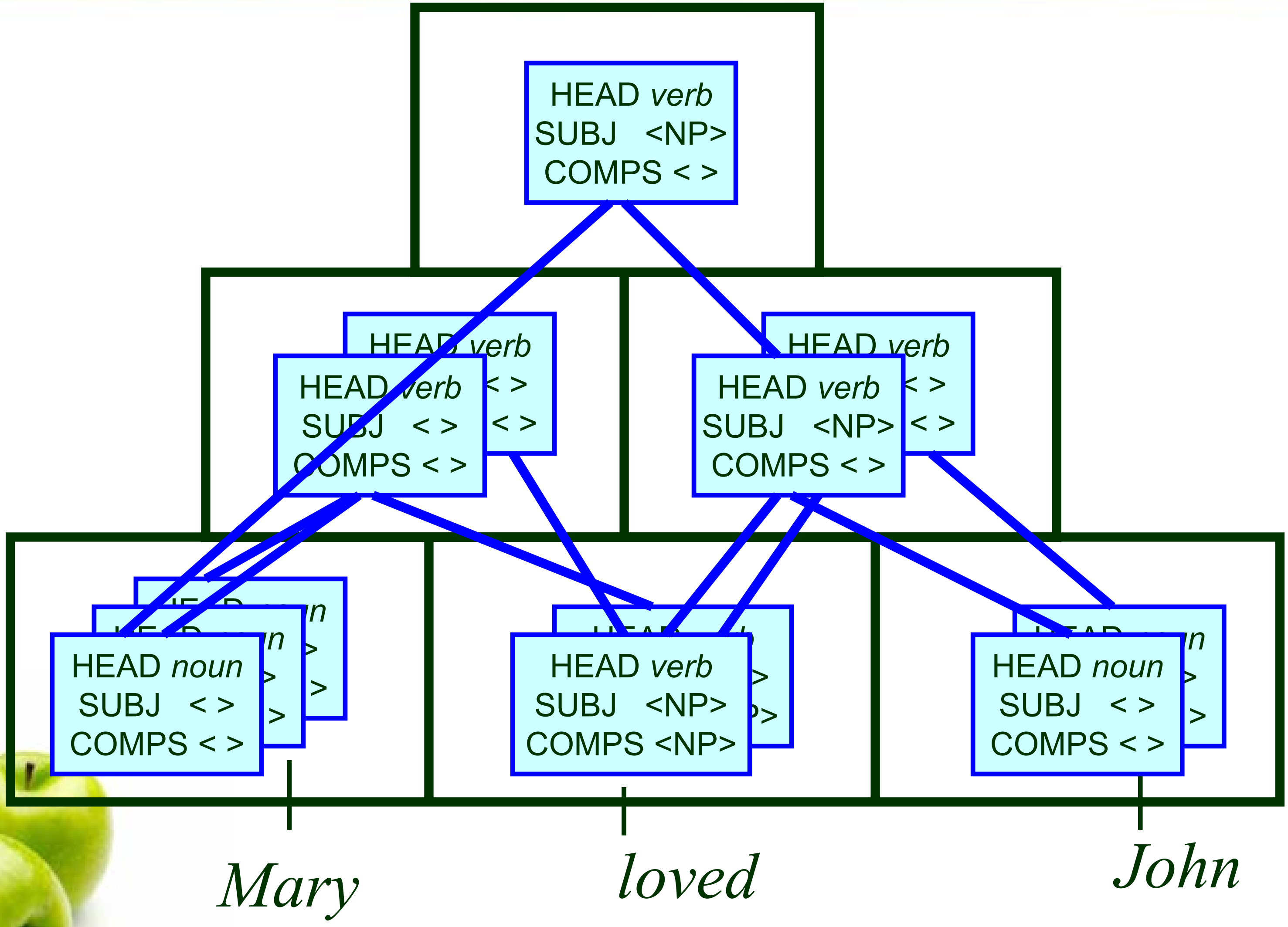
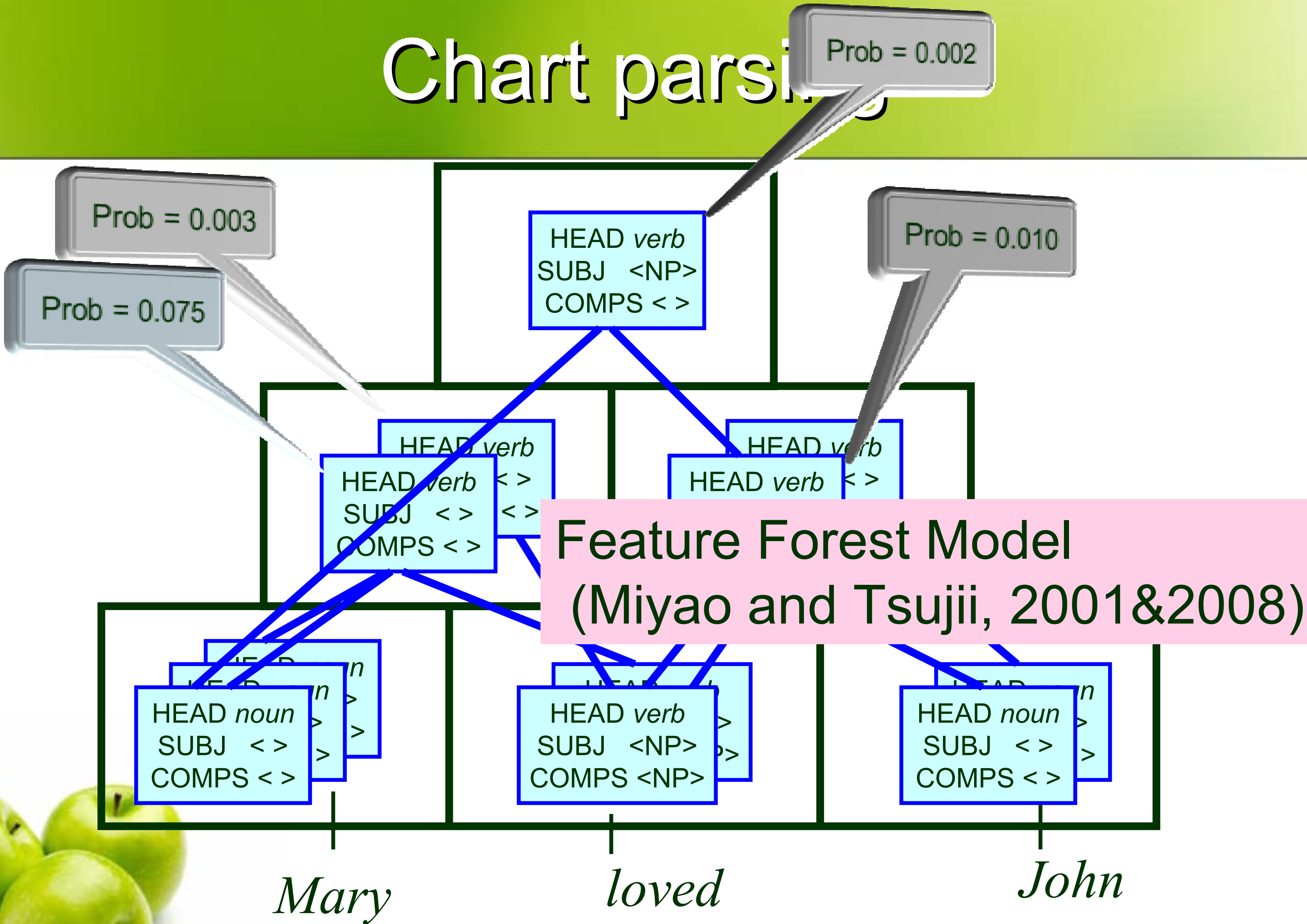
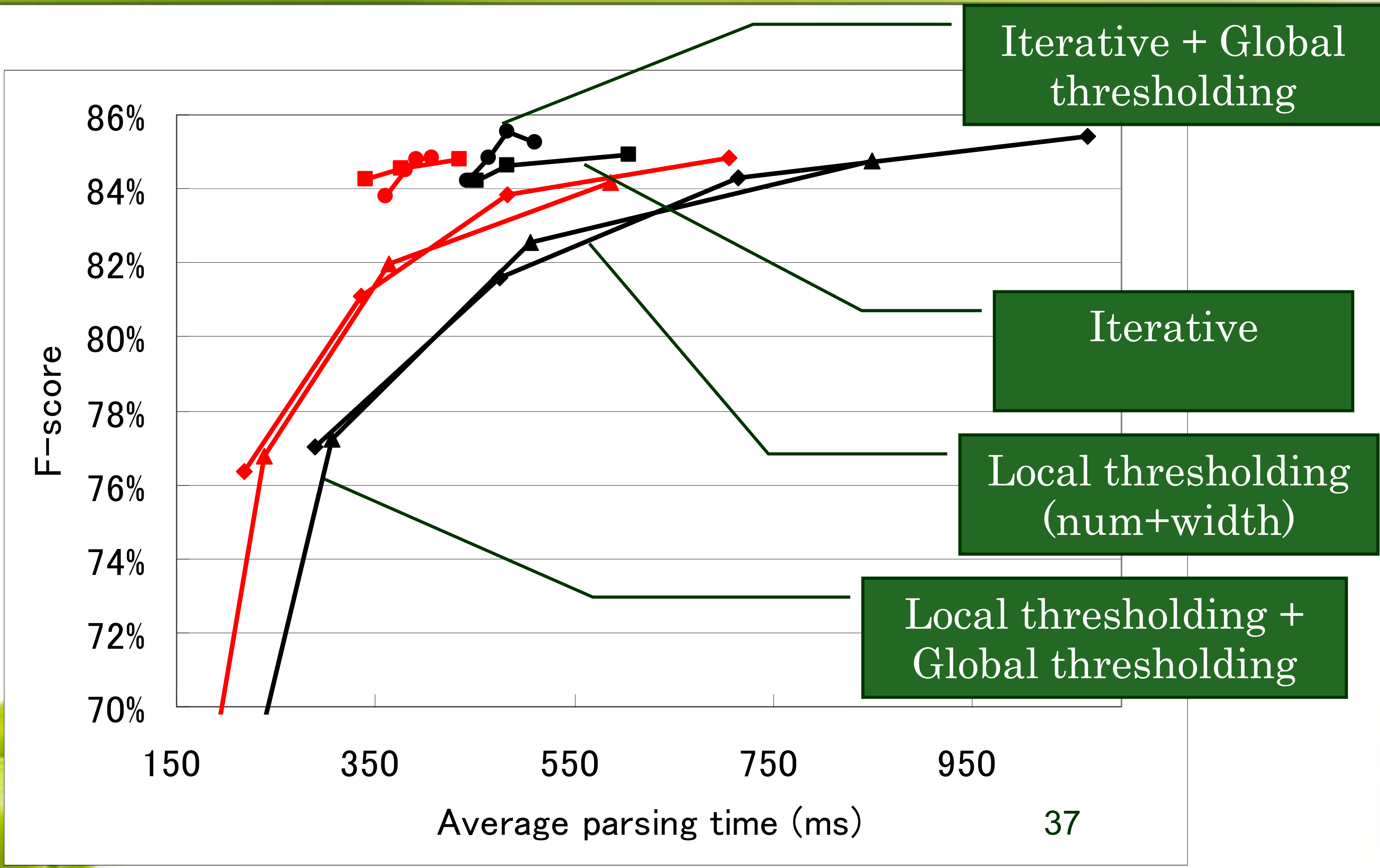


Chart parsing



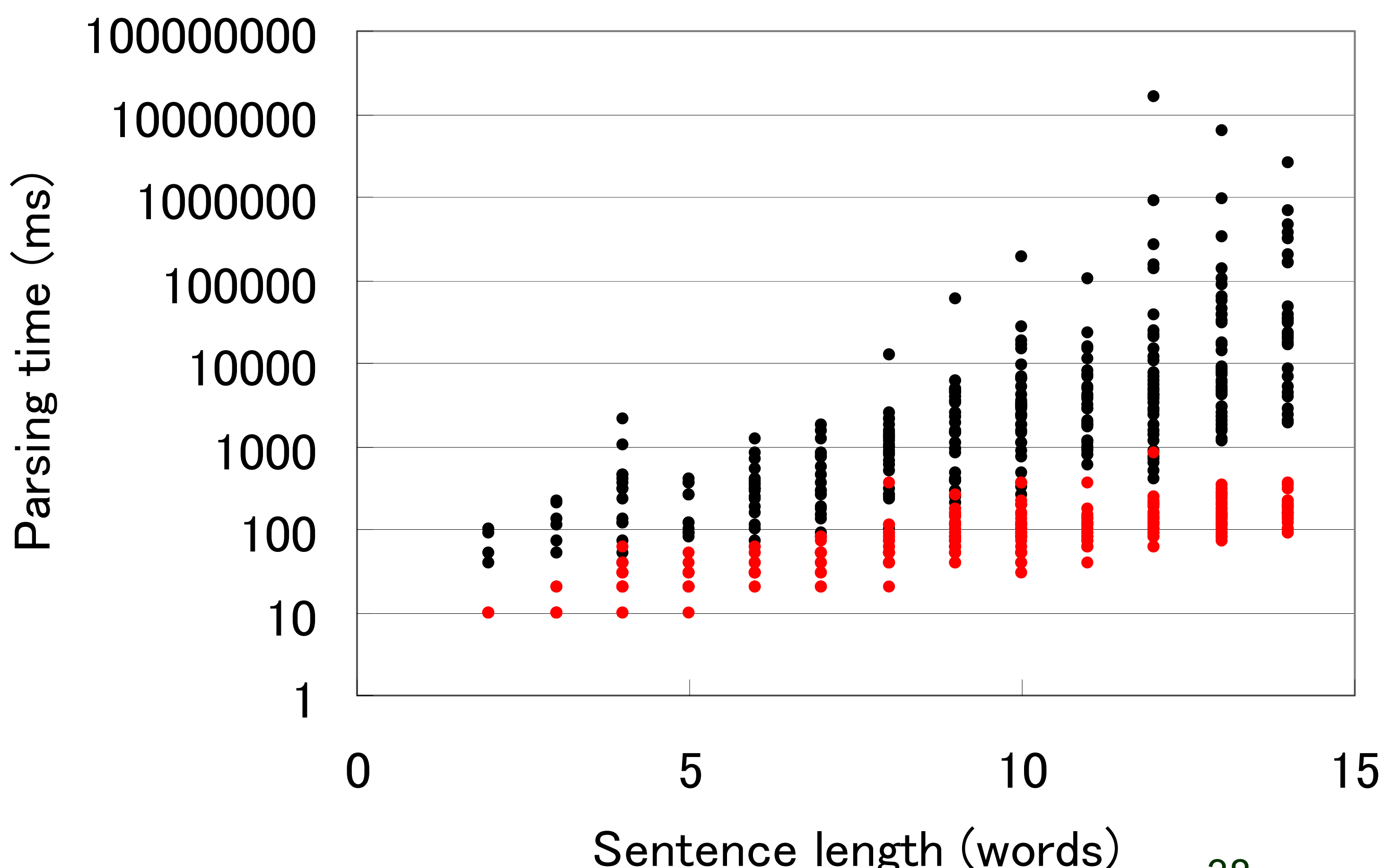
Feature Forest Model
(Miyao and Tsujii, 2001&2008)

Beam Search and Iterative Widening
Ninomiya 2005



Distribution of Parsing time for Sentence Length

(Black … none) (Red … Iterative Parsing)



Performance of Semantic Parser

	Penn Treebank	GENIA
Coverage	99.7%	99.2%
F-Value (PArelations)	87.4%	86.4%
Sentence Precison	39.2%	31.8%
Processing Time	[illegible]	[illegible]

Scalability of TM Tools - MEDIE

Target Corpus: MEDLINE corpus

The number of papers	14,792,890
The number of abstracts	7,434,879
The number of sentences	70,815,480
The number of words	1,418,949,650
Compressed data size	3.2GB
Uncompressed data size	10GB

Scalability of TM Tools - MEDIE

Target Corpus: MEDLINE corpus

The number of papers	14,792,890
The number of abstracts	7,434,879
The number of se	,480
The number of w	,650
Compressed data size	3.2GB
Uncompressed data size	10GB

Suppose, for example,
that it takes one
second for parsing
one sentence....

Scalability of TM Tools - MEDIE

Target Corpus: MEDLINE corpus

The number of papers	14,792,890
The number of abstracts	7,434,879
The number of se	,480
The number of w	,650
Comp	3.2GB
Un	10GB

Suppose, for example, that it takes one second for parsing one sentence….

70 million seconds, that is, about 2 years

TM and GRID
[Ninomiya 2006, Taura 2004]

- Solution
  - The entire MEDLINE were parsed by distributed PC clusters consisting of 340 CPUs
  - Parallel processing was managed by grid platform GXP
- Experiments
  - The entire MEDLINE was parsed in 8 days
- Output
  - Syntactic parse trees and predicate argument structures in XML format
  - The data sizes of compressed/uncompressed output were 42.5GB/260GB.

More Accurate and Efficient Parser

- Current Research -

Research on Advanced Natural Language Processing and Text Mining: aNT
Grant-in-Aid for Specially promoted research, MEXT (2006-2011)

Selection of Lexical Entries

- Reference distribution of unigram lexical entry selection (Miyao & Tsujii 2005)
  - Filtering unlikely lexical entries during parameter estimation

$$p(T \mid \mathbf{w}) = p_{uni}(T \mid \mathbf{w}) \boxed{\frac{1}{Z} \exp\left(\sum_u \lambda_u f_u(T)\right)}$$

reference distribution

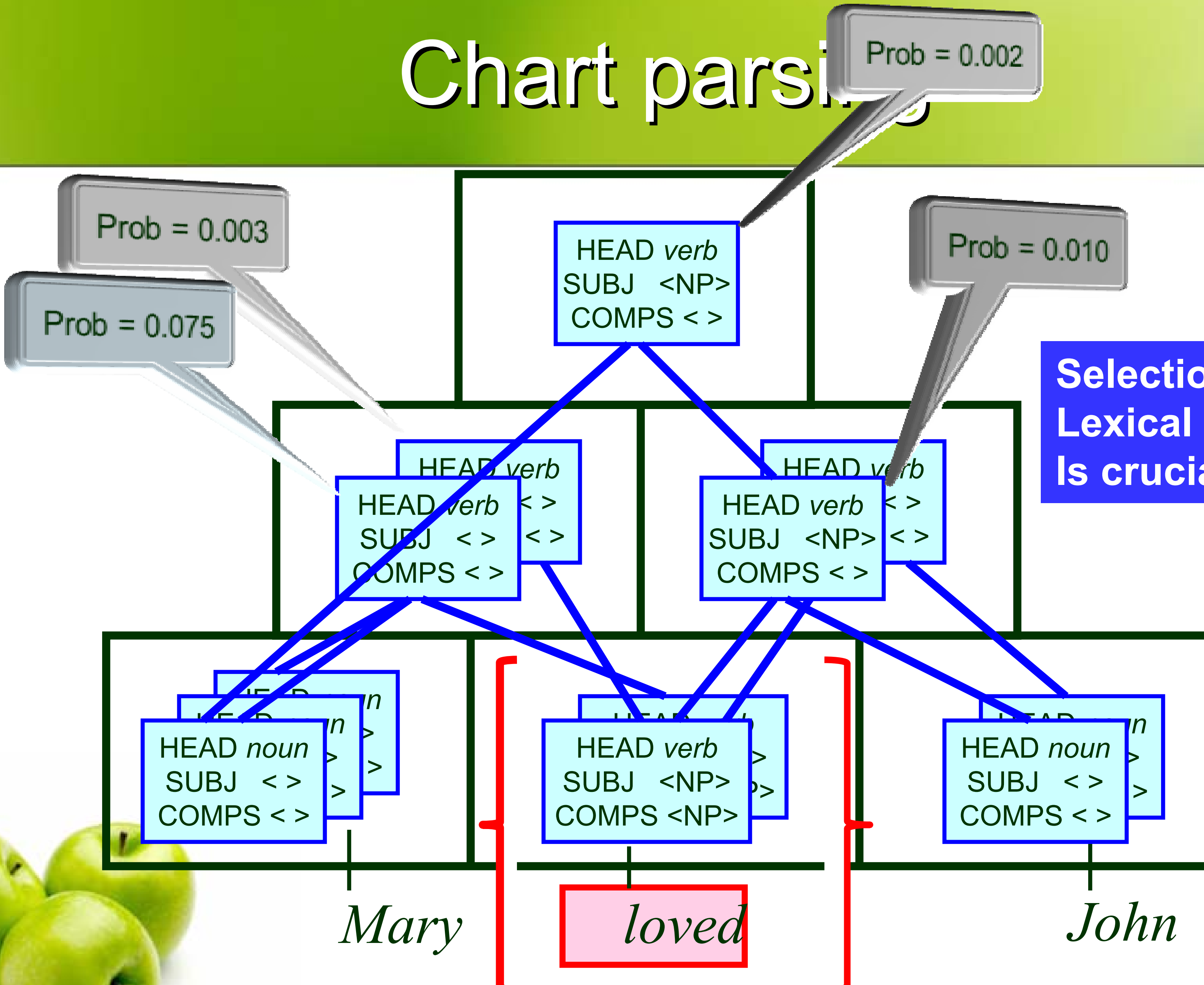
  - Unigram lexical entry selection

lexical entry

$$p_{uni}(T \mid \mathbf{w}) = \prod_{i=1}^{n} p(l_i \mid w_i, P_i)$$

word POS

Chart parsing



Selection of Lexical Entries Is crucial.

Selection of Lexical Entries
Super-Tagging

- Reference distribution of unigram lexical entry selection (Miyao & Tsujii 2005)
  - Filtering unlikely lexical entries during parameter estimation

$$p(T \mid \mathbf{w}) = \boxed{p_{uni}(T \mid \mathbf{w})}\boxed{\frac{1}{Z}\exp(\sum_{u} \lambda_u f_u(T))}$$

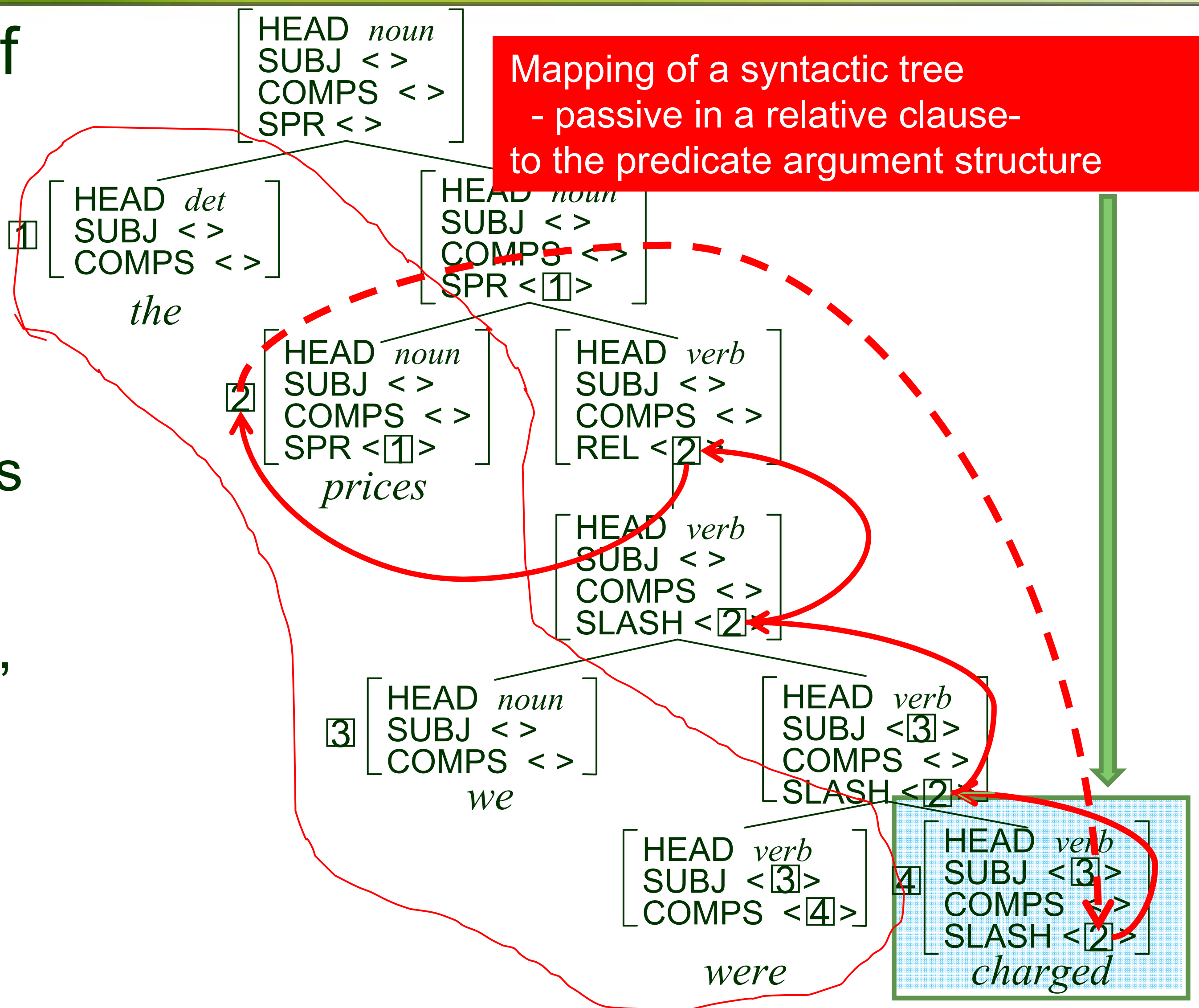
reference distribution

  - Super-tagger

$$p_{suptag}(T \mid \mathbf{w}) = \prod_{i=1}^{n} p(l_i \mid w_{i-1}, w_i, w_{i+1}, P_{i-2}, P_{i-1}, P_i, P_{i+1}, P_{i+2})$$

Super-tagging and HPSG

- An example of a complex syntactic tree
  - SLASH, REL features explain non-local dependencies
  - WH movement, topicalization, relative clauses



Deep Parser with Super-Tagging

Accuracy of predicate-argument dependencies and parsing time (Section 23 ≦ 100 words, Gold POS)

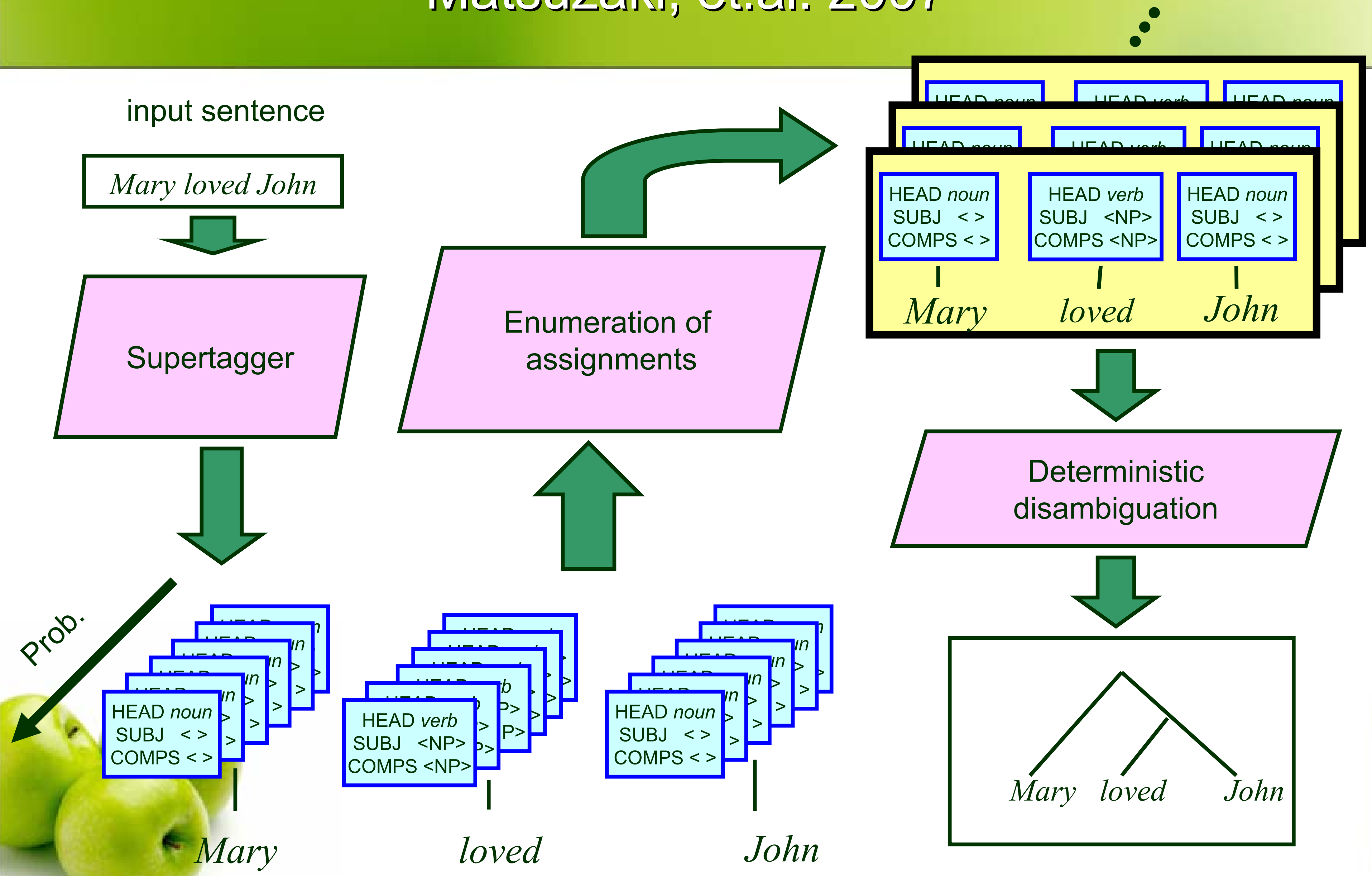
Model	Precision	Recall	F-Score	Avg. Time (ms/sentence)
Miyao & Tsujii (2005) (=unigram ref)	87.3%	86.5%	86.9%	604
Ninomiya et al. (2006) (=n-gram multi)	89.5%	88.6%	89.0%	152
Ninomiya et al.1 (2007) (=n-gram ref, fast and accurate)	89.8%	89.3%	89.5%	234
Ninomiya et al. 2 (2007) (=n-gram ref, slow but accurate)	90.3%	89.6%	89.8%	1379

Integrated Model vs. Staged Model

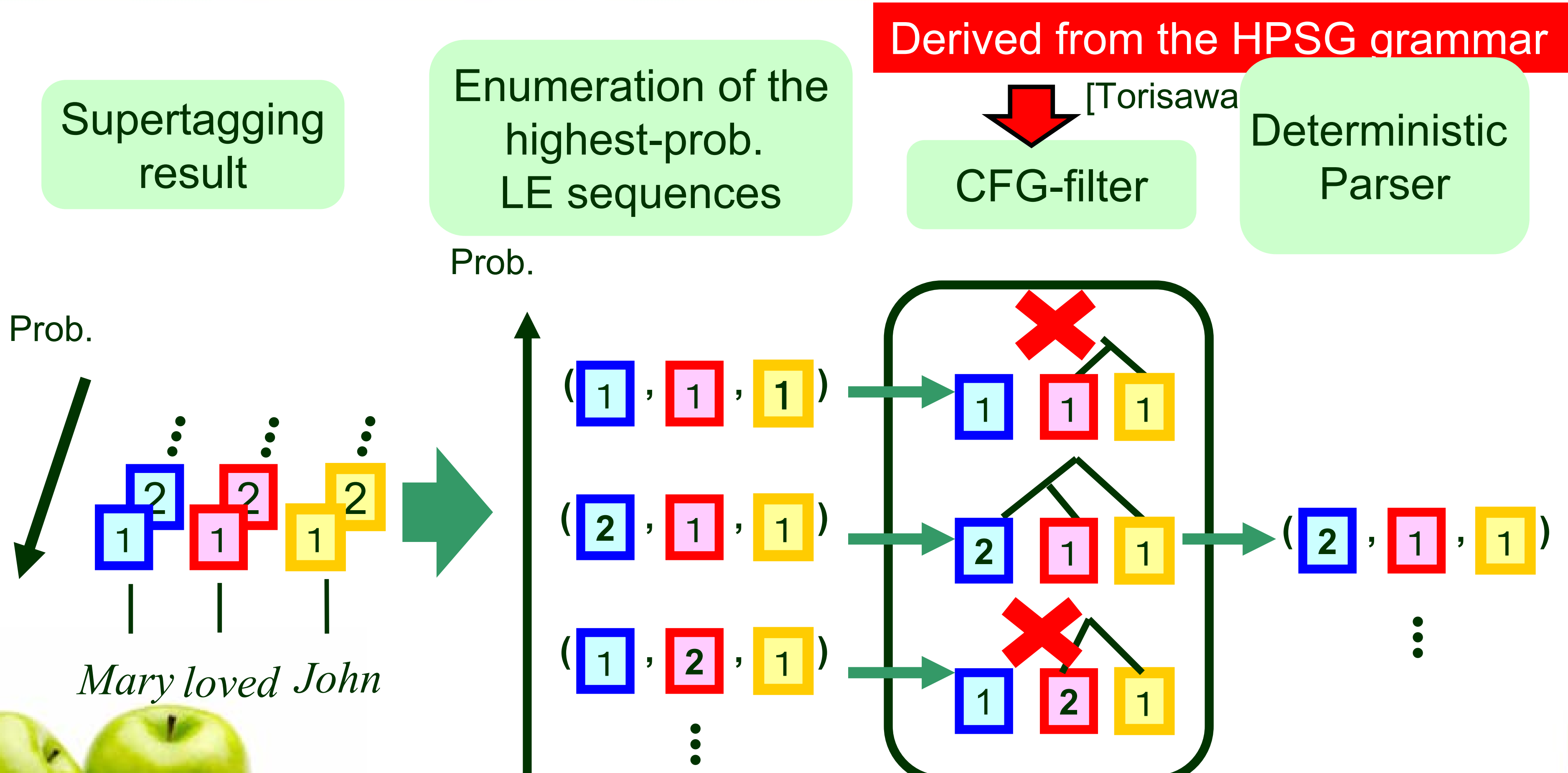
$$p(T \mid \mathbf{w}) = \boxed{p_{suptag}(T \mid \mathbf{w})}\ \boxed{\frac{1}{Z}\exp\Big(\sum_u \lambda_u f_u(T)\Big)}$$

Super-Tagger → Deterministic Parser

System Overview
Matsuzaki, et.al. 2007

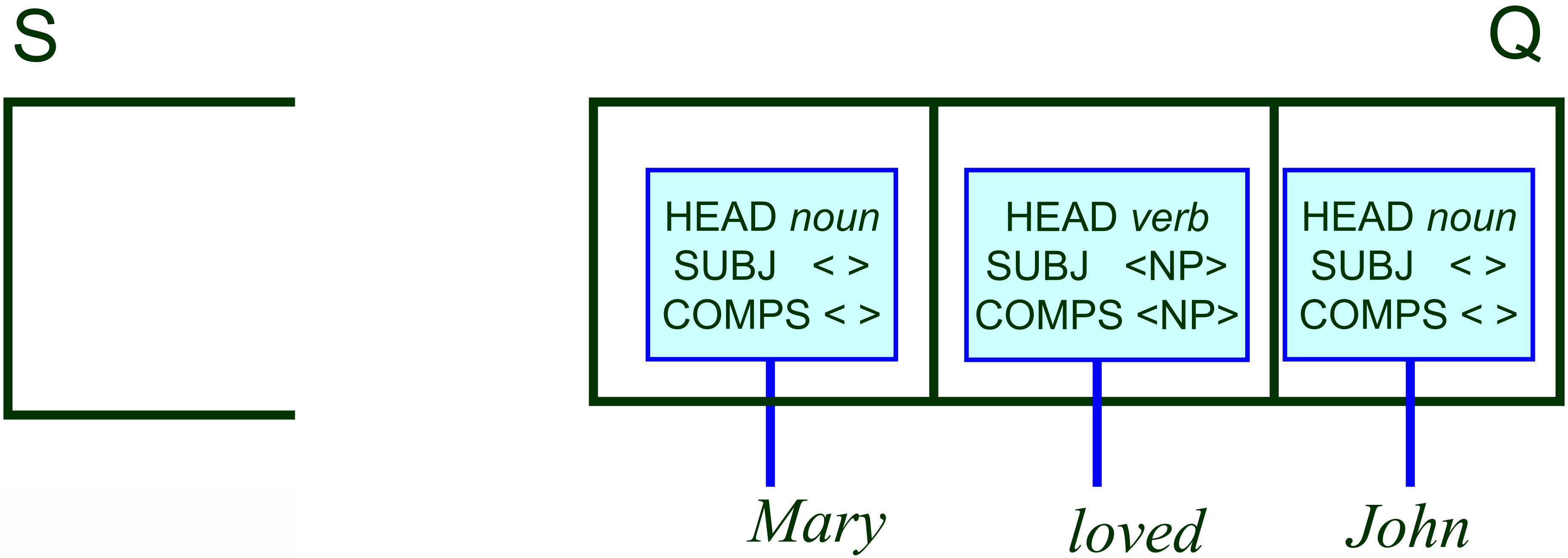


Enumaration of the maybe-parsable LE assignments

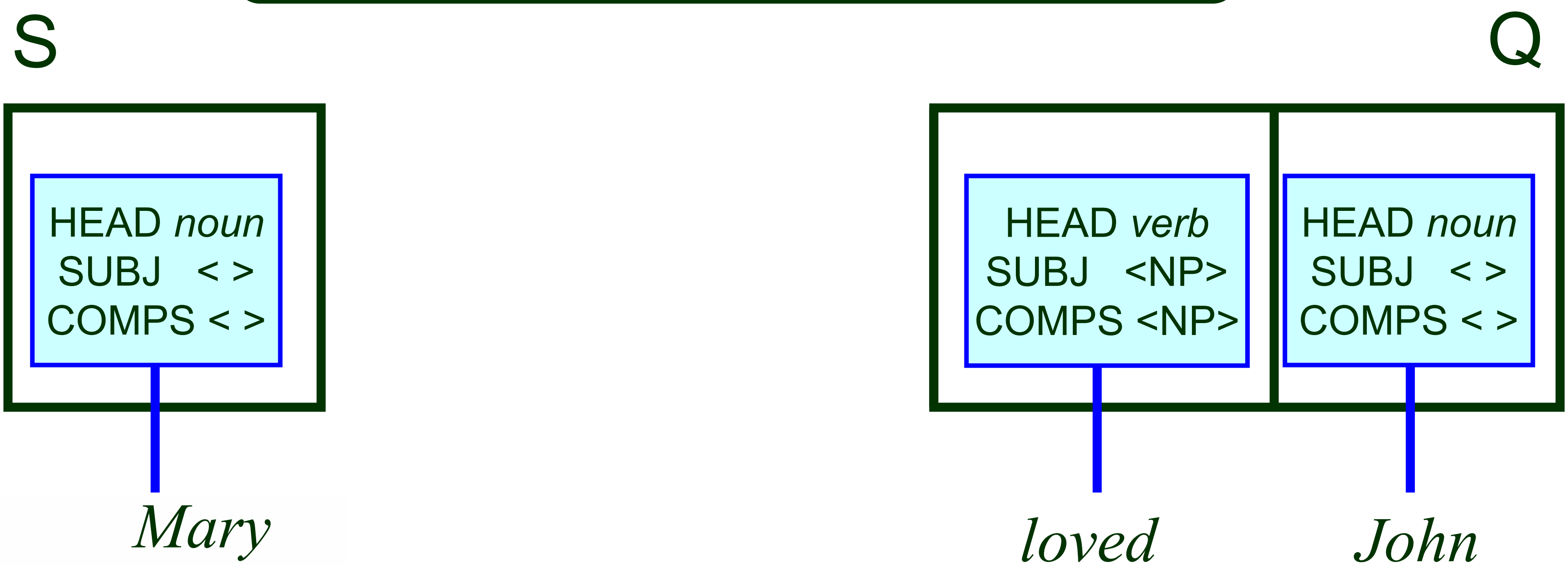


Deterministic S-R Parser

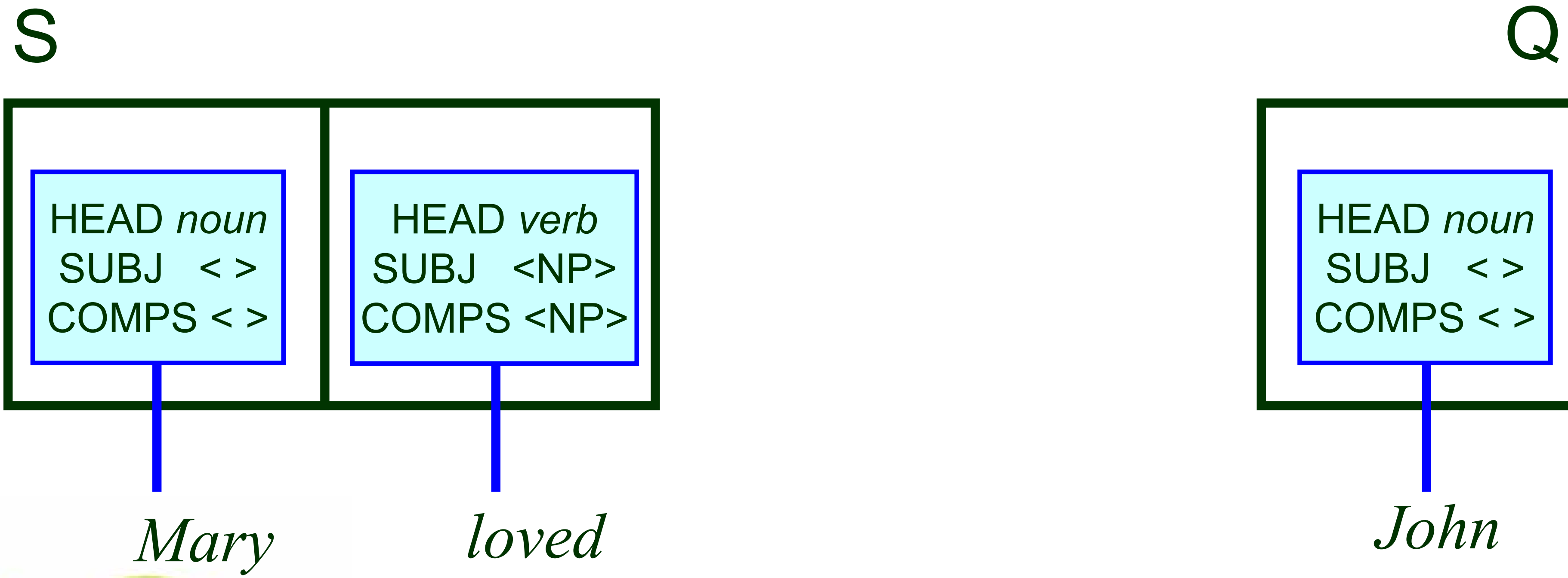
Initial state



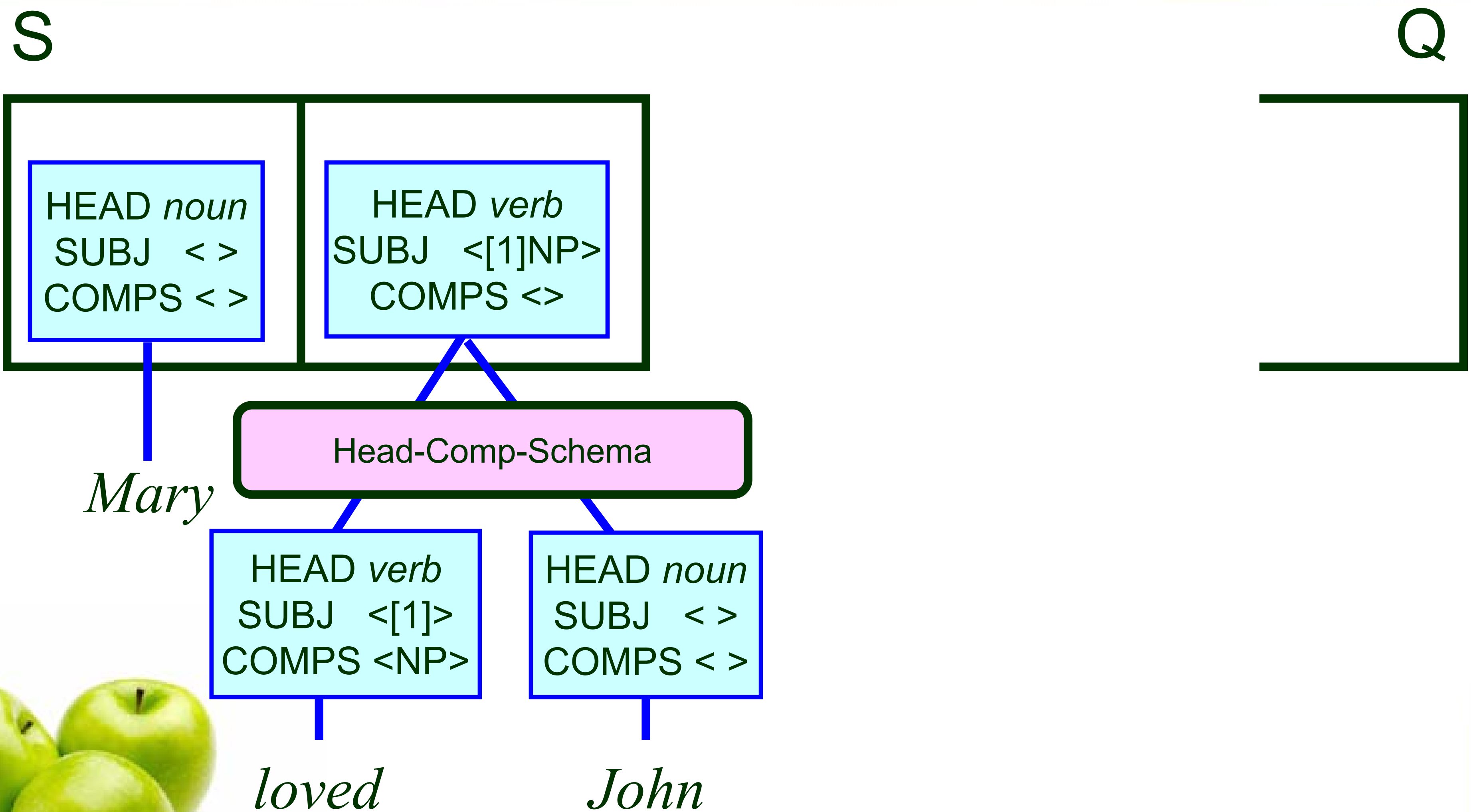
argmax F(a, S, Q) = SHIFT



argmax F(a, S, Q) = SHIFT



argmax F(a, S, Q) = REDUCE(Head_Comp)



Experiment Results

	LP(%)	LR(%)	F1(%)	Avg. time
Staged/Deterministic model	86.93	86.47	86.70	**30ms/snt**
Previous method 1 （Supertagger+ChartParser）	87.35	86.29	86.81	183ms/snt
Previous method 2 （Unigram + ChartParser）	84.96	84.25	84.60	674ms/snt

6 times faster

20 times faster than the initial model

Richer Models
Domain Adaptation

- Low parsing accuracy for different domains

  Ex.) Enju: trained on the Penn Treebank
  - *Penn Treebank*: 89.81 (F-score)
  - *GENIA** (biomedical domain): 86.39 (F-score)
- Re-training a probabilistic model on the domain
- Small training data for the target domain
  - *Penn Treebank*: 39,832 sentences
  - *GENIA**: 10,848 sentences (>> other domains)

---

* Kim et al., 1998

Adaptation with Reference Distribution

Lexical Assignment

Syntactic Preference

$$p_E(t \mid \mathbf{w}) = \frac{1}{Z_{\mathbf{w}}} \prod_{w_i \in \mathbf{w}} p_{lex}(l_i \mid w_i) \cdot q_{syn}(t \mid \mathbf{l}),$$

$$Z_{\mathbf{w}} = \sum_{t \in T(\mathbf{w})} \prod_{w_i \in \mathbf{w}} p_{lex}(l_i \mid w_i) \cdot q_{syn}(t \mid \mathbf{l})$$

Feature function

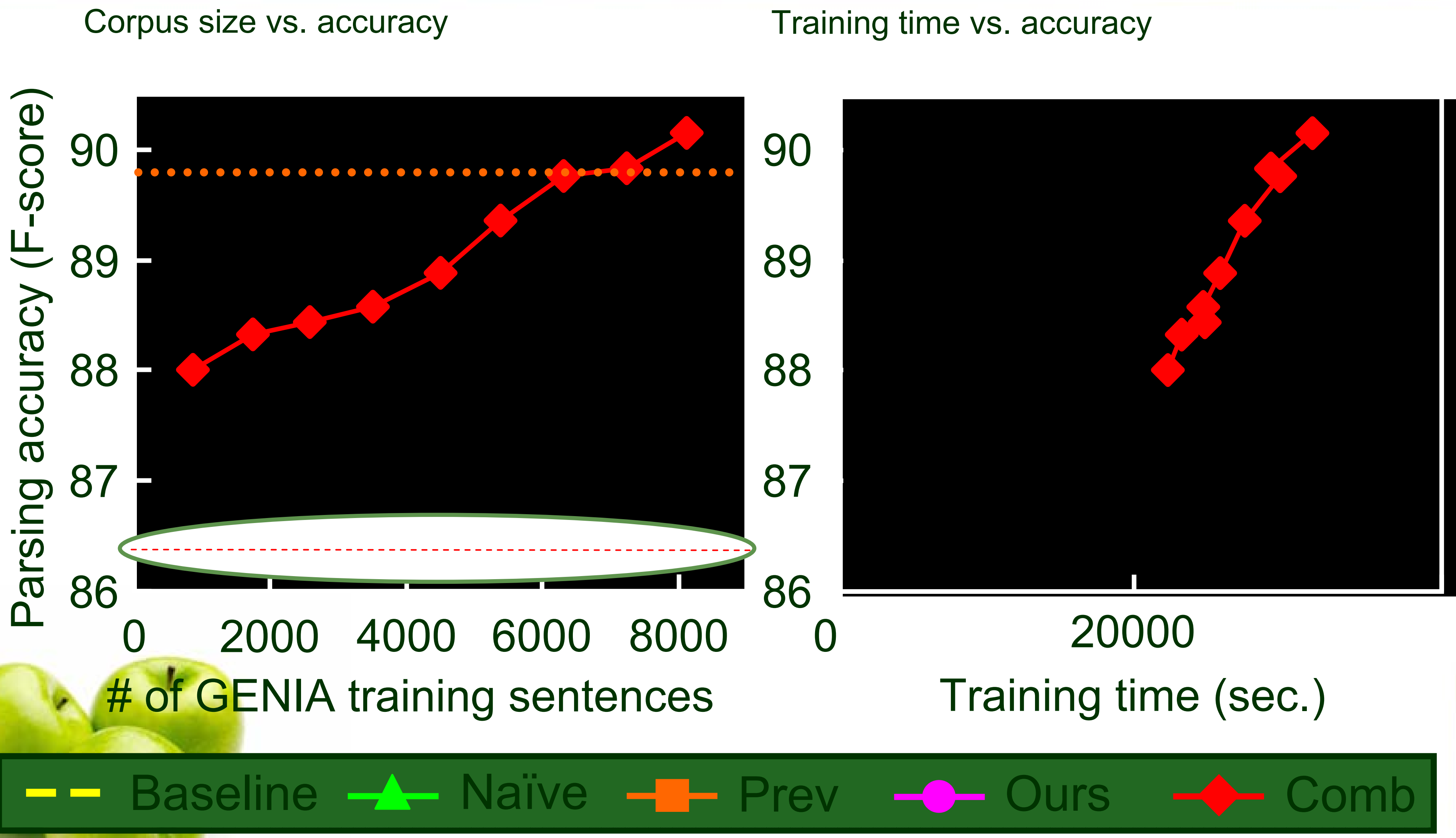
Feature weight

Original model

$$p_M(t \mid s) = \frac{1}{Z'_s} \boxed{p_0(t \mid s)} \exp\left(\sum_j \rho_j g_j(t \mid s) \right)$$

Performance of Adaptation Models

Hara 2007

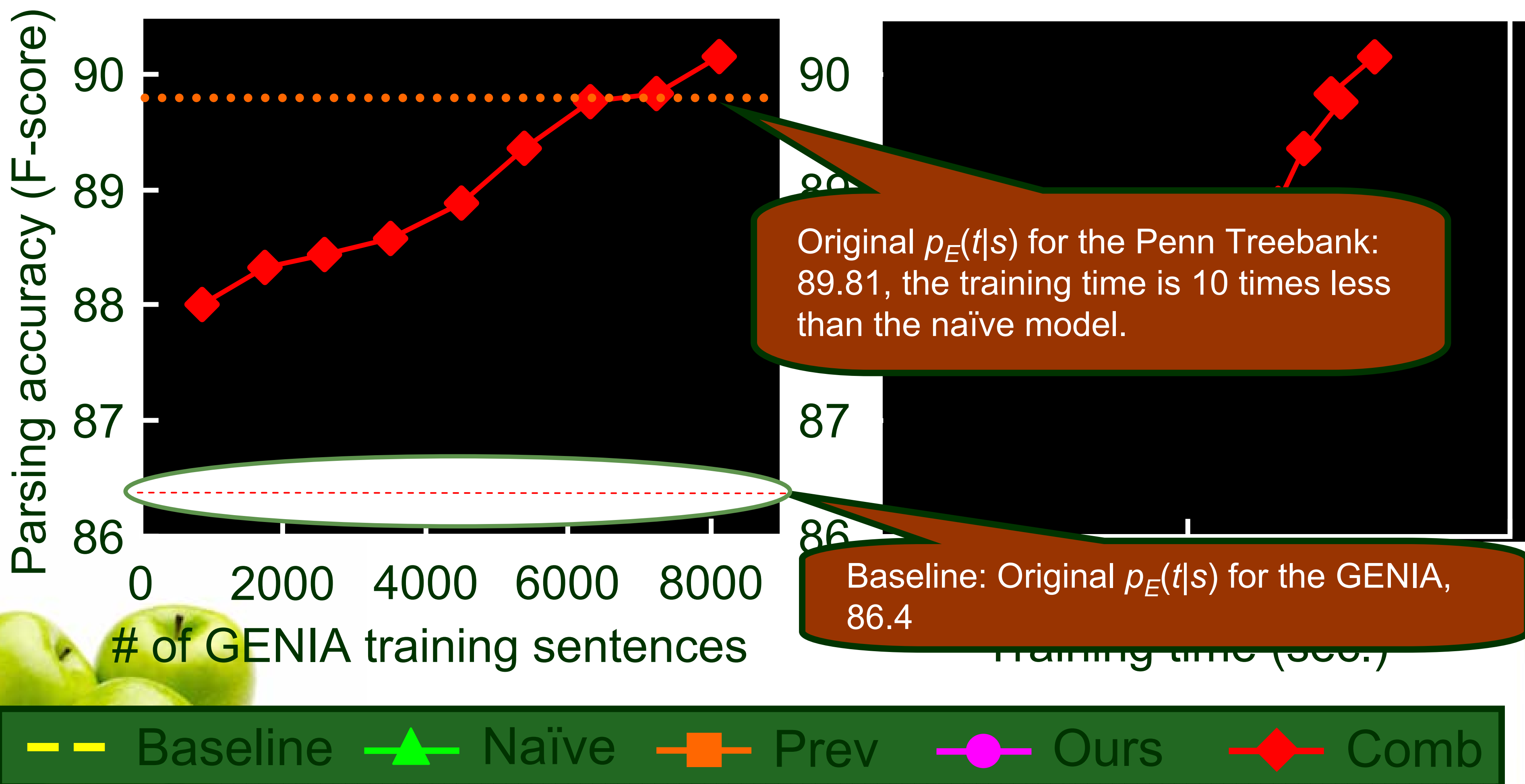


Performance of Adaptation Models

Hara 2007

Corpus size vs. accuracy

Training time vs. accuracy



Adaptation with Reference Distribution

Lexical Assignment

Syntactic Preference

$$p_E(t \mid \mathbf{w}) = \frac{1}{Z_{\mathbf{w}}} \prod_{w_i \in \mathbf{w}} p_{lex}(l_i \mid w_i) \cdot q_{syn}(t \mid \mathbf{l}),$$

Independent of the original model

$$Z_{\mathbf{w}} = \sum_{t \in T(\mathbf{w})} \prod_{w_i \in \mathbf{w}} p_{lex}(l_i \mid w_i) \cdot q_{syn}(t \mid \mathbf{l})$$

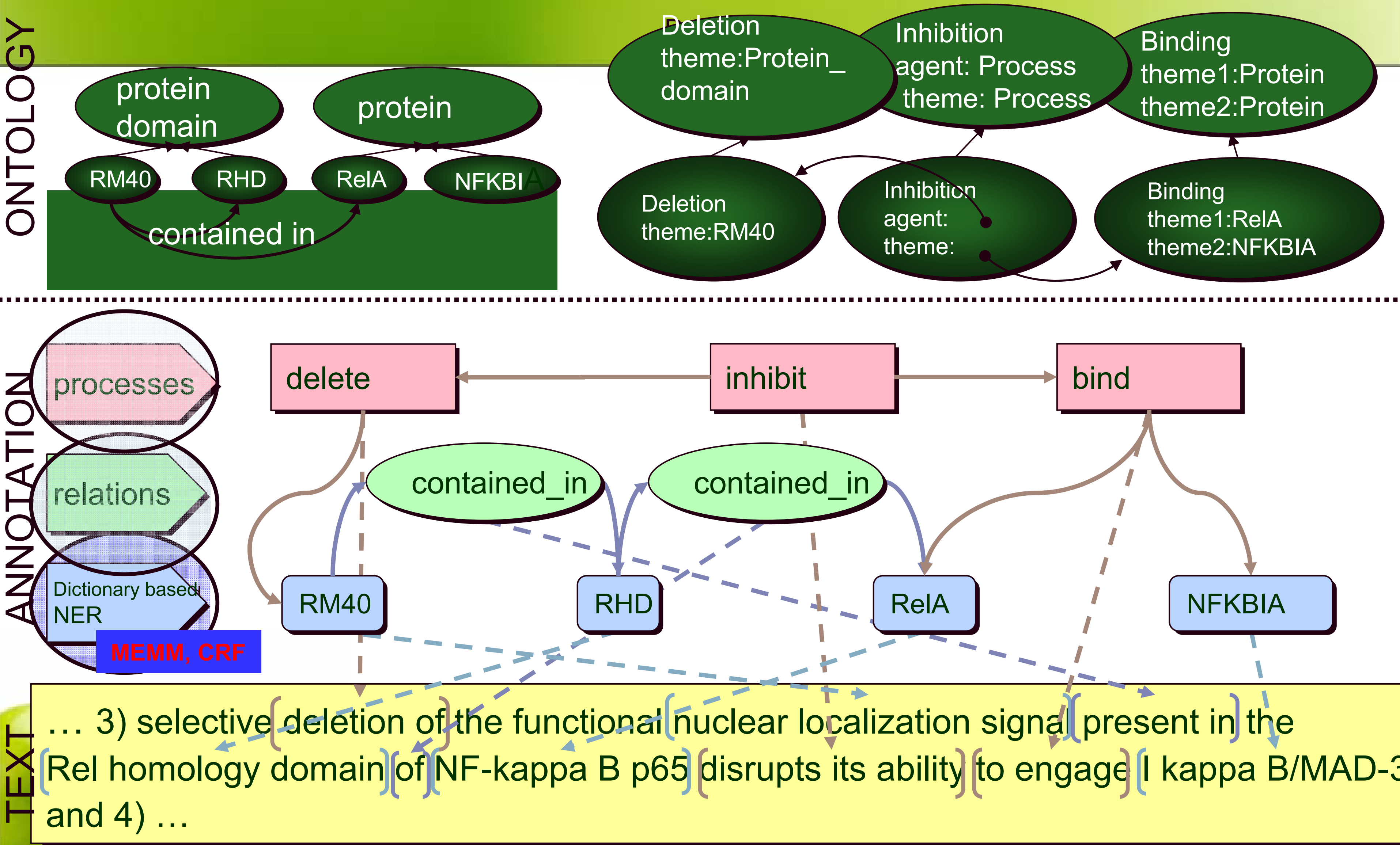
Feature function

Feature weight

Original model

$$p_M(t \mid s) = \frac{1}{Z'_s} \boxed{p_0(t \mid s)} \exp\left(\sum_j \rho_j g_j(t \mid s) \right)$$

NER and Knowledge-based Processing



Adaptation with Reference Distribution

Lexical Assignment

Syntactic Preference

Relation, Event Recognition

$$p_E(t \mid \mathbf{w}) = \frac{1}{Z_{\mathbf{w}}} \prod_{w_i \in \mathbf{w}} p_{lex}(l_i \mid w_i) \cdot q_{syn}(t \mid \mathbf{l}),$$

NER results as soft constraints

$$Z_{\mathbf{w}} = \sum_{t \in T(\mathbf{w})} \prod_{w_i \in \mathbf{w}} p_{lex}(l_i \mid w_i) \cdot q_{syn}(t \mid \mathbf{l})$$

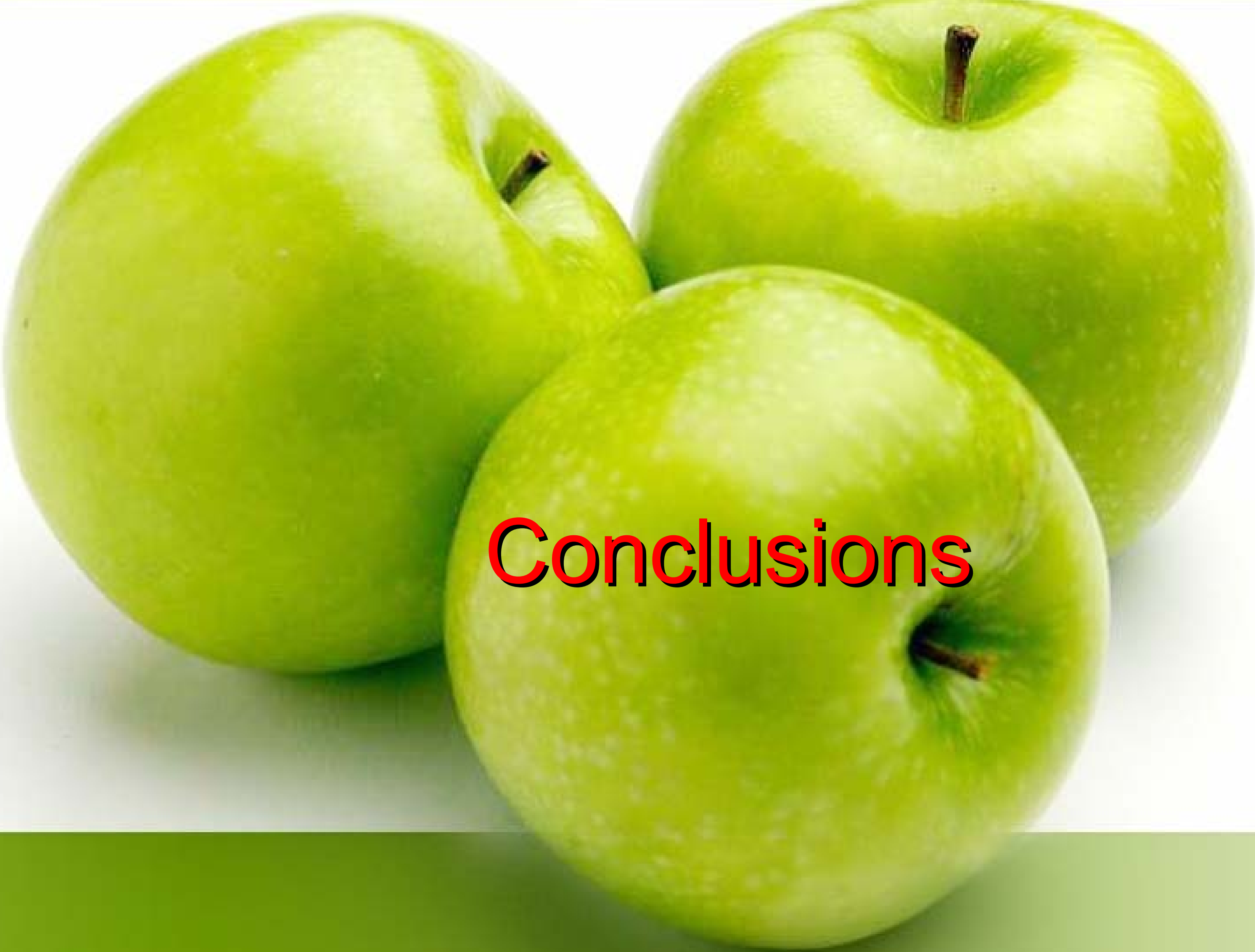
Feature function

Feature weight

$$p_M(t \mid s) = \frac{1}{Z'_s} \boxed{p_0(t \mid s)} \exp\left(\sum_j \rho_j g_j(t \mid s)\right)$$

Original model

Conclusions



Conclusions: Lessons

- A Deep Parser, which produces semantic representation, has become a practical option
- Integrated Model to Staged Model, lower level processings with rich context

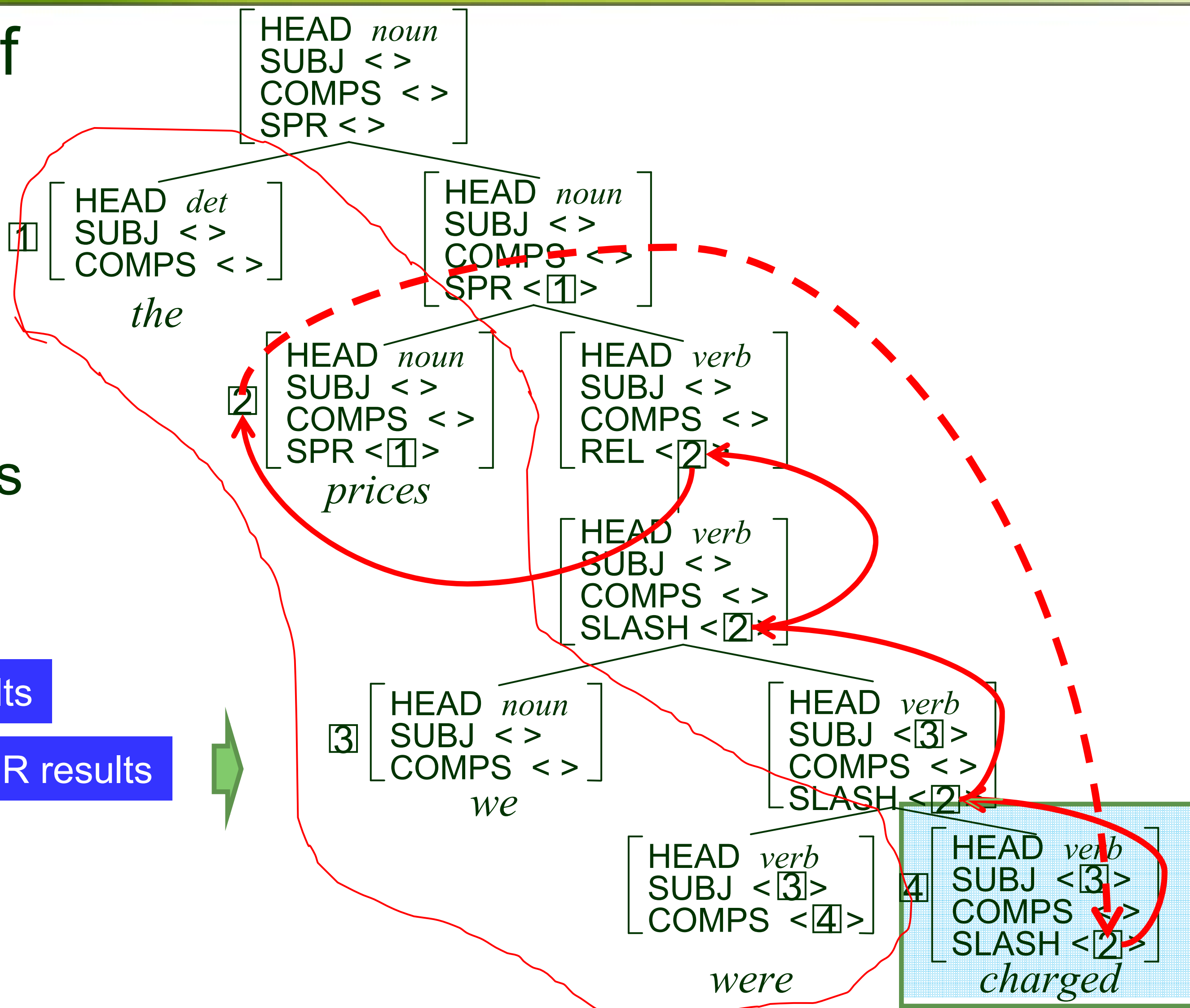
Super-tagging and HPSG

- An example of a complex syntactic tree
  - SLASH, REL features explain non-local dependencies
  - WH movement, topicalization, relative clauses

RR results

NER results

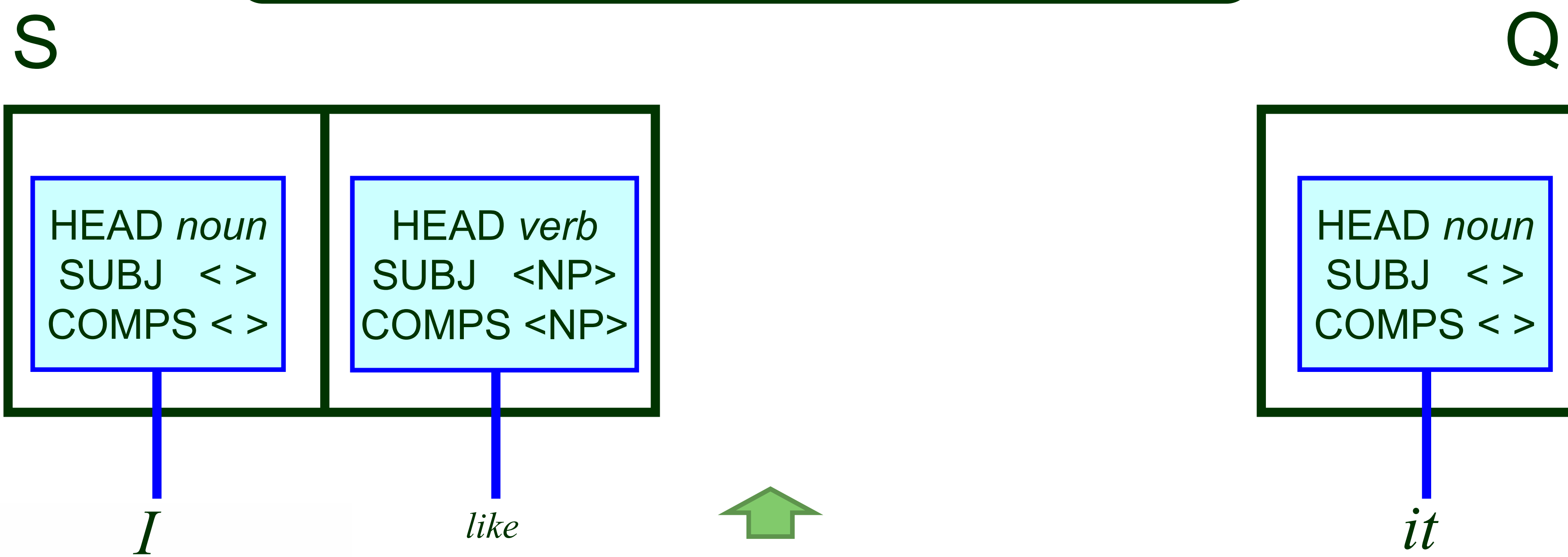
ER results



Conclusions: Lessons

- A Deep Parser, which produces semantic representation, has become a practical option
- Integrated Model to Staged Model, lower level Processing with rich context
- Deterministic Parser with classifiers based on rich linguistic and extra-linguistic information

argmax F(a, S, Q) = SHIFT



NER results

RR results

ER results

Conclusions: Lessons

- A Deep Parser, which produces semantic representation, has become a practical option
- Integrated Model to Staged Model, lower level Processing with rich context
- Deterministic Parser with classifiers based on rich linguistic and extra-linguistic information
- Combination of Constraints & Preferences, more robust parsers

Thank You !

The field has matured, ready to be used by applications.

Integration of linguistic grammar formalisms with statistical models.

Robust, efficient and
open to eclectic sources of information
other than syntactic ones

……. Speech Understanding
Speech/Text Retrieval ………