Improving Trees and Alignments for Syntax-Based Machine Translation

Kevin Knight

USC/Information Sciences Institute

joint work with Steven DeNeefe, Daniel Marcu,
Wei Wang, and Jonathan May

SRI, July 12, 2007

Syntactic Approaches to MT

- Use of syntactic information (noun, verb, etc) in the translation process:
  - Manually constructed rule-based systems
  - Statistical systems
    - Wu & Wong, 1998
    - Yamada & Knight, 2001-2002
    - Galley et al, 2004
  - Contrast with phrase-based statistical approaches

Phrase-Based Output

枪手 被 警方 击毙 .

Gunman of police killed .

Decoder Hypothesis #1

Phrase-Based Output

枪手 被 警方 击毙 .

Gunman of police attack .

Decoder Hypothesis #7

Phrase-Based Output

枪手 被 警方 击毙 .

Gunman by police killed .

Decoder Hypothesis #12

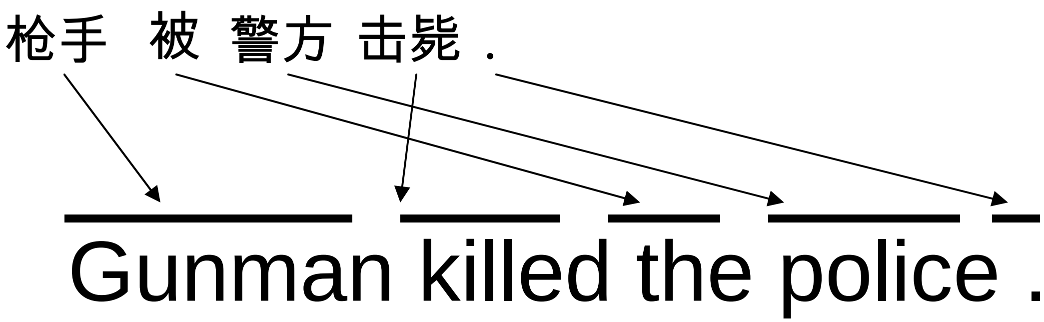
Phrase-Based Output

枪手 被 警方 击毙 .

Killed gunman by police .

Decoder
Hypothesis #134

Phrase-Based Output



Decoder Hypothesis #9,329

Phrase-Based Output

枪手 被 警方 击毙 .

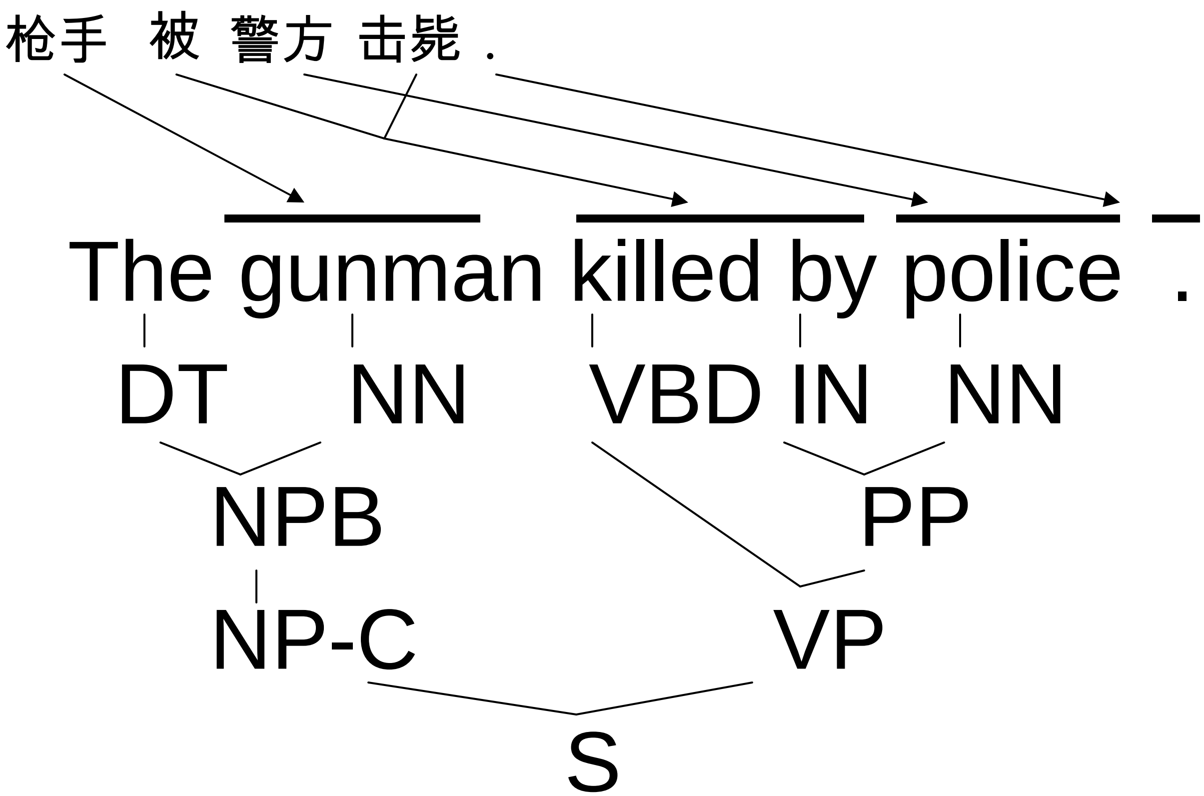
Gunman killed by police .

Decoder Hypothesis #50,654

Problematic –

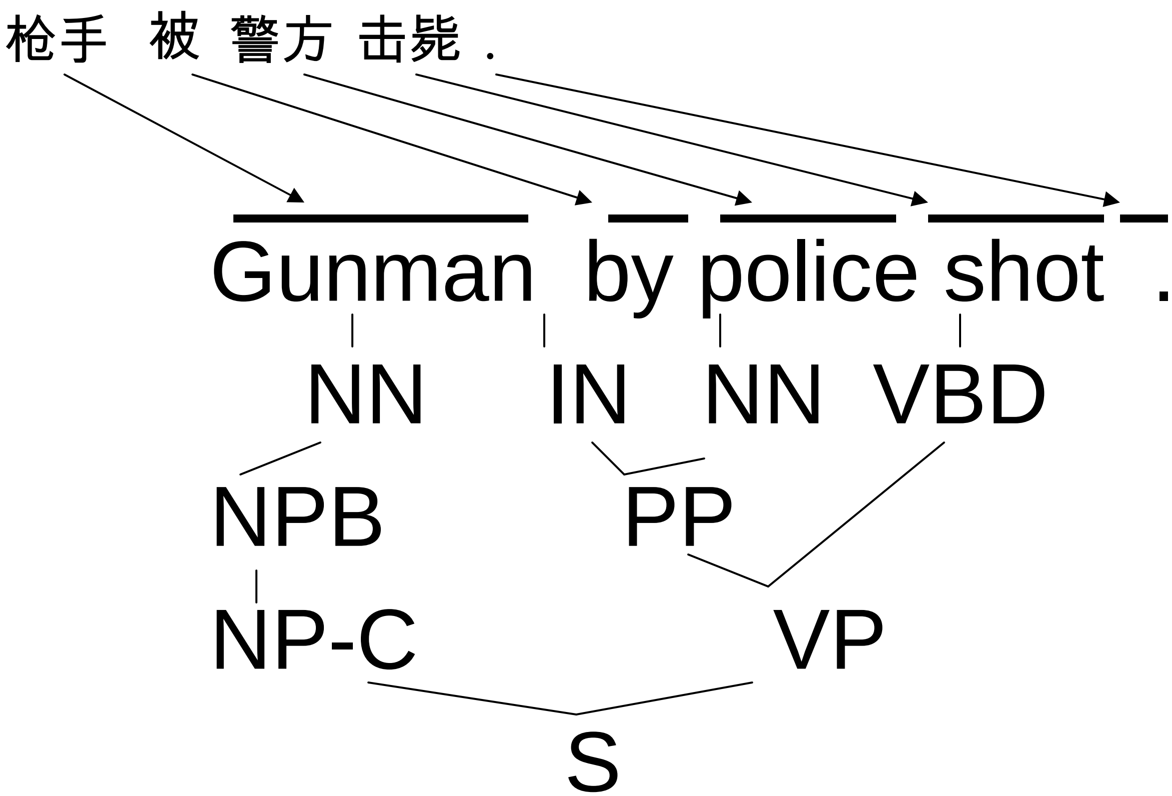
- Output lacks English **auxiliary** and **determiner**
- Re-ordering relies on luck, instead of on Chinese passive marker

Syntax-Based Output



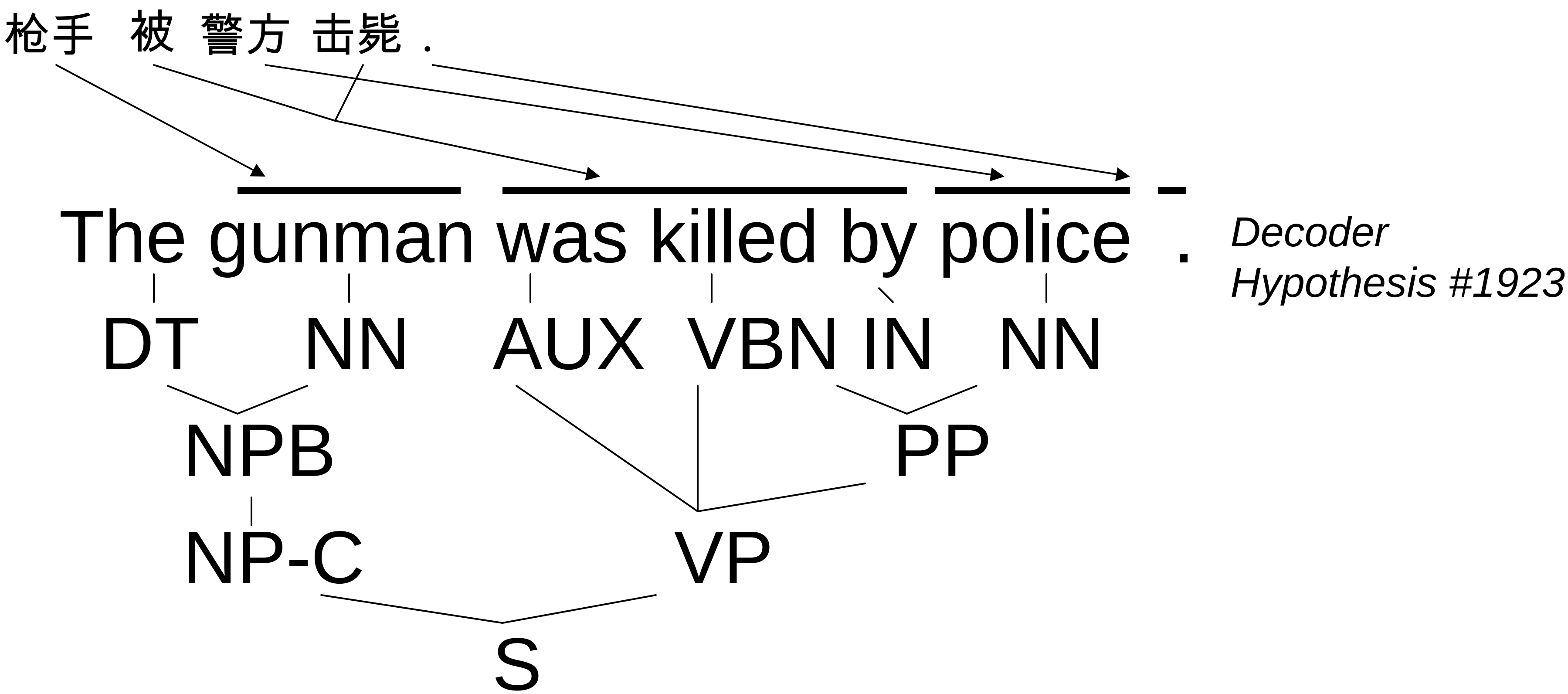
Decoder Hypothesis #1

Syntax-Based Output



Decoder Hypothesis #16

Syntax-Based Output



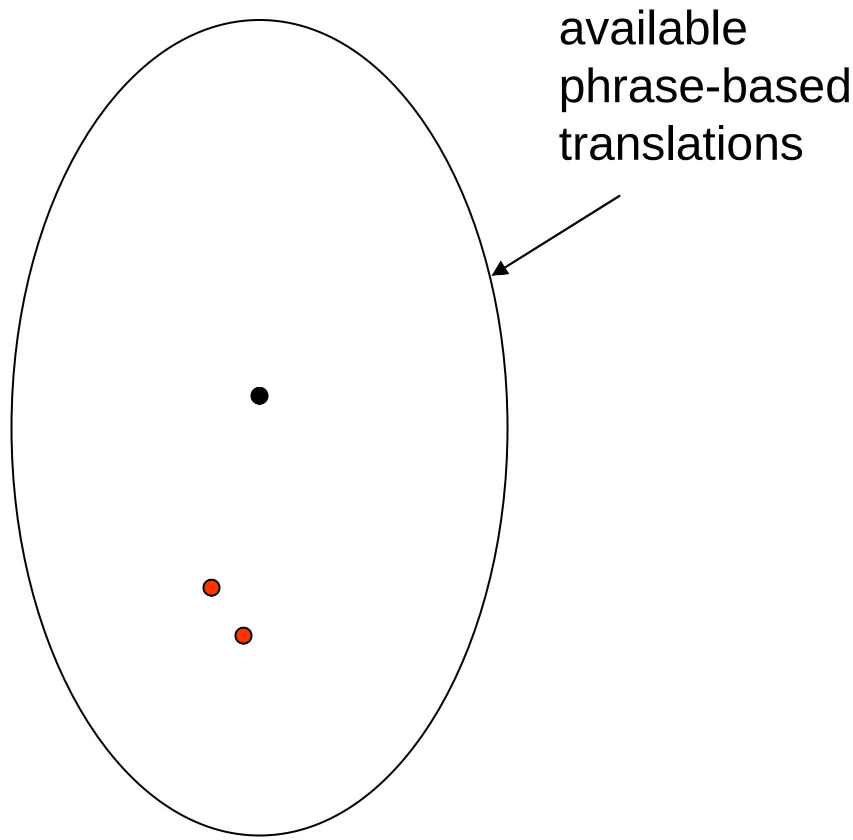
Why Might Syntax Help?

- Phrase-based MT output is “n-grammatical”, not grammatical
  - Every sentence needs a subject and a verb
- Re-ordering is poorly explained as “distortion” -- better explained as syntactic transformation
  - Arabic to English, VSO → SVO
- Function words have syntactic effects even if they are not themselves translated

Why Might Syntax Hurt?

- Less freedom to glue pieces of output together -- search space has fewer output strings
- Search space is more difficult to navigate
- Rule extraction from bilingual text has limitations

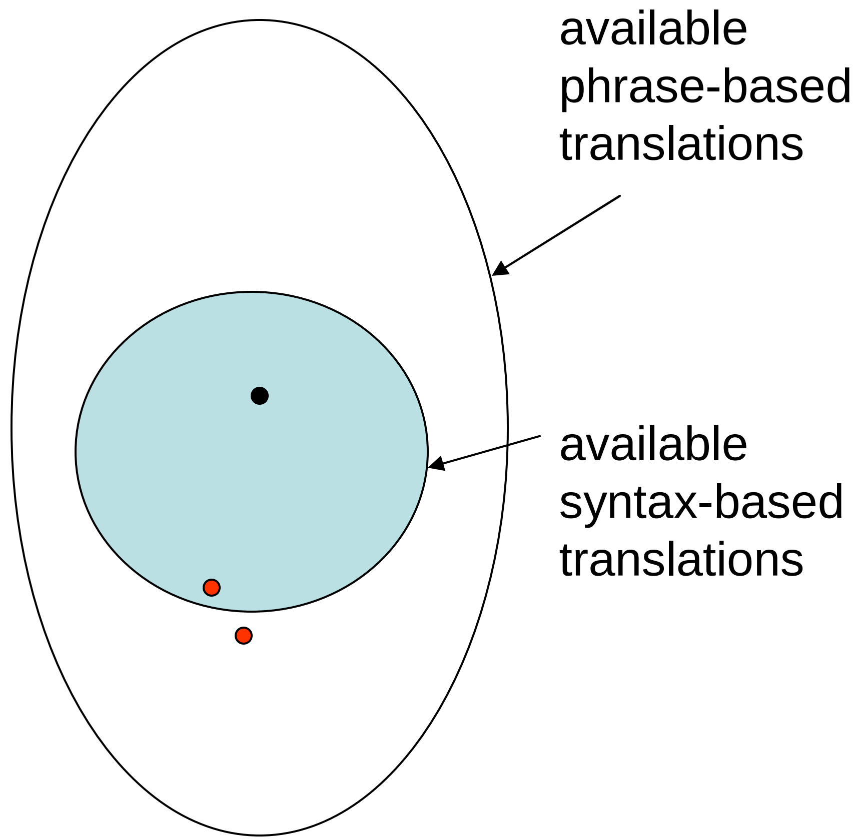
this talk



Why Might Syntax Hurt?

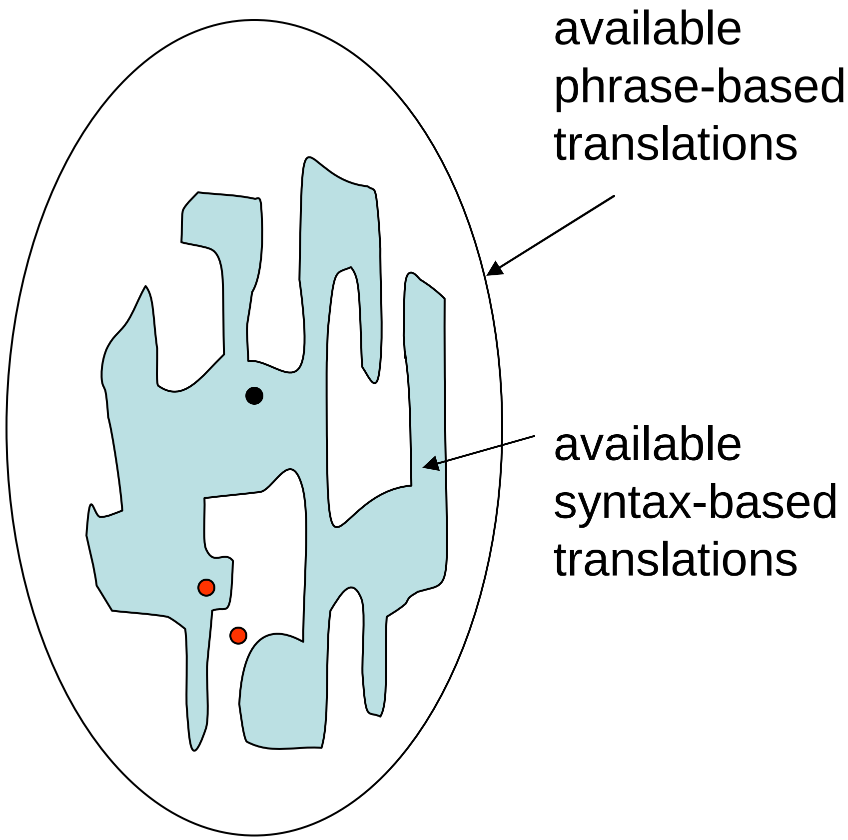
- Less freedom to glue pieces of output together -- search space has fewer output strings
- Search space is more difficult to navigate
- Rule extraction from bilingual text has limitations

this talk



Why Might Syntax Hurt?

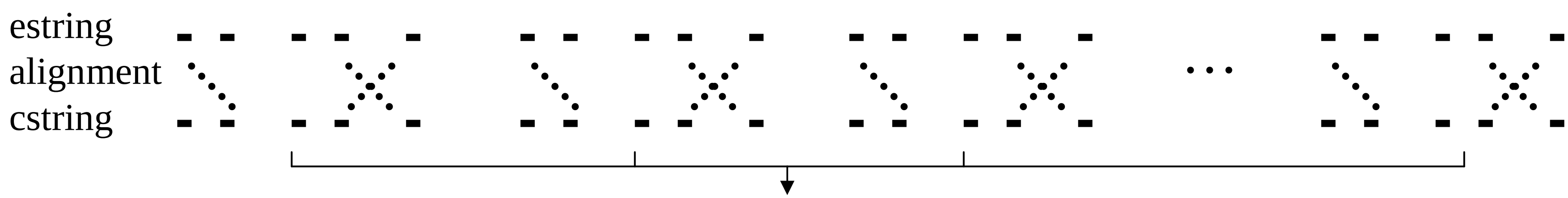
- Less freedom to glue pieces of output together -- search space has fewer output strings
- Search space is more difficult to navigate
- Rule extraction from bilingual text has limitations



Comparing Phrase-Based Extraction with Syntax-Based Extraction

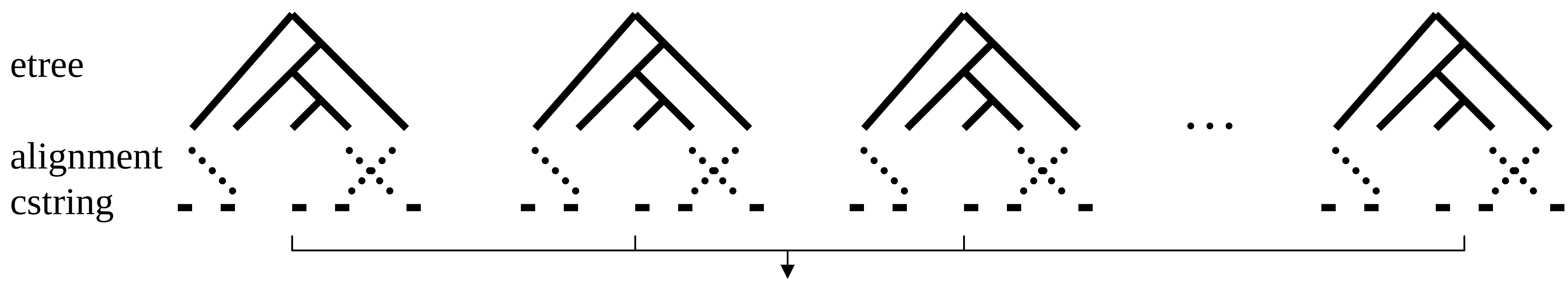
- Quantitatively compare
  - A typical phrase-based bilingual extraction algorithm (**ATS**, Och & Ney 2004)
  - A typical syntax-based bilingual extraction algorithm (**GHKM**, Galley et al 2004)
  - These algorithms picked from two good-scoring NIST-06 systems
- Identify areas of improvement for syntax-based rule coverage

Phrase-Based and Syntax-Based Pattern Extraction



ATS [Och & Ney, 2004]

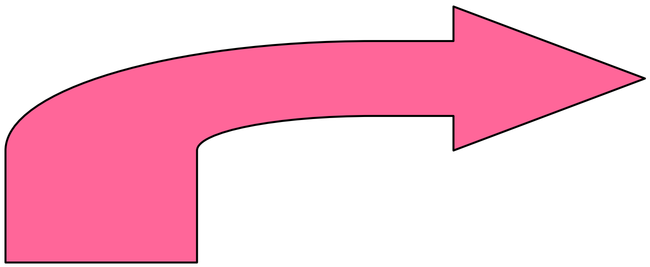
phrase pairs consistent with word alignment



GHKM [Galley et al 2004]

syntax transformation rules consistent with word alignment

ATS (Och & Ney, 2004)



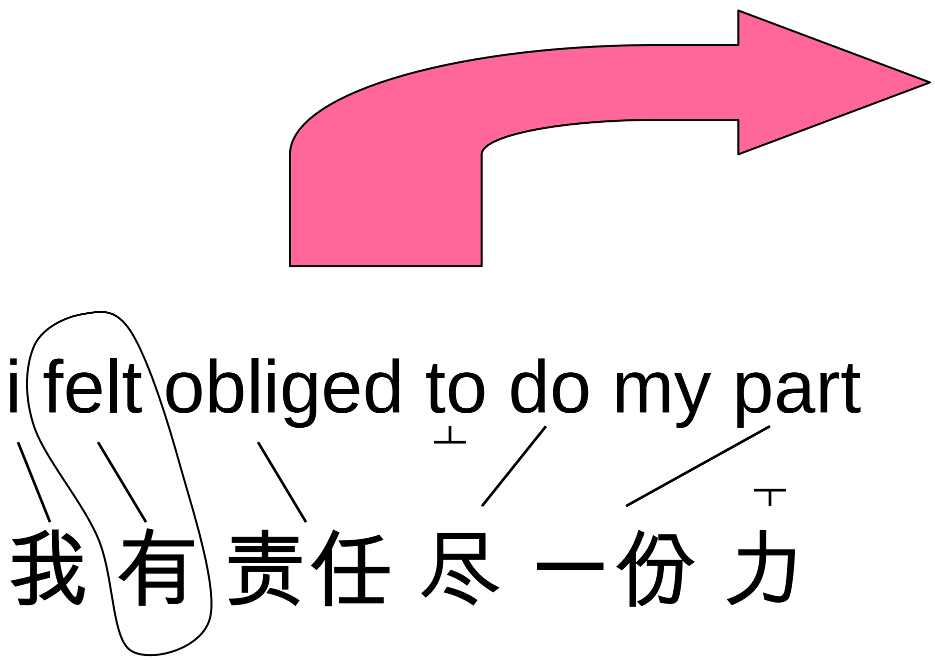
i felt obliged to do my part

我 有 责任 尽 一份 力

PHRASE PAIRS ACQUIRED:

felt	→	有
felt obliged	→	有 责任
felt obliged to do	→	有 责任 尽
obliged	→	责任
obliged to do	→	责任 尽
do	→	尽
part	→	一份
part	→	一份 力

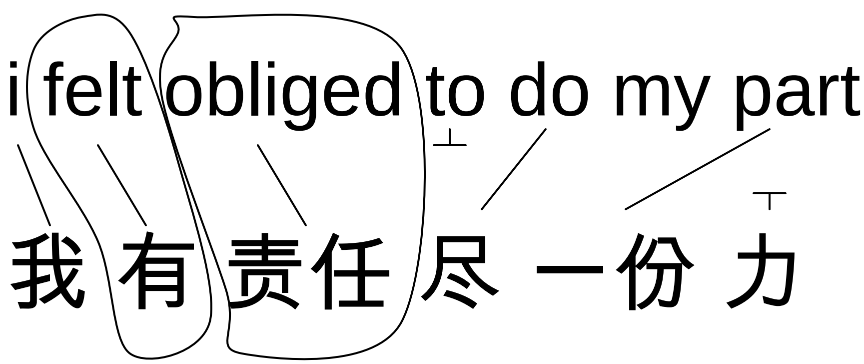
ATS (Och & Ney, 2004)



PHRASE PAIRS ACQUIRED:

felt → 有
felt obliged → 有 责任
felt obliged to do → 有 责任 尽
obliged → 责任
obliged to do → 责任 尽
do → 尽
part → 一份
part → 一份 力

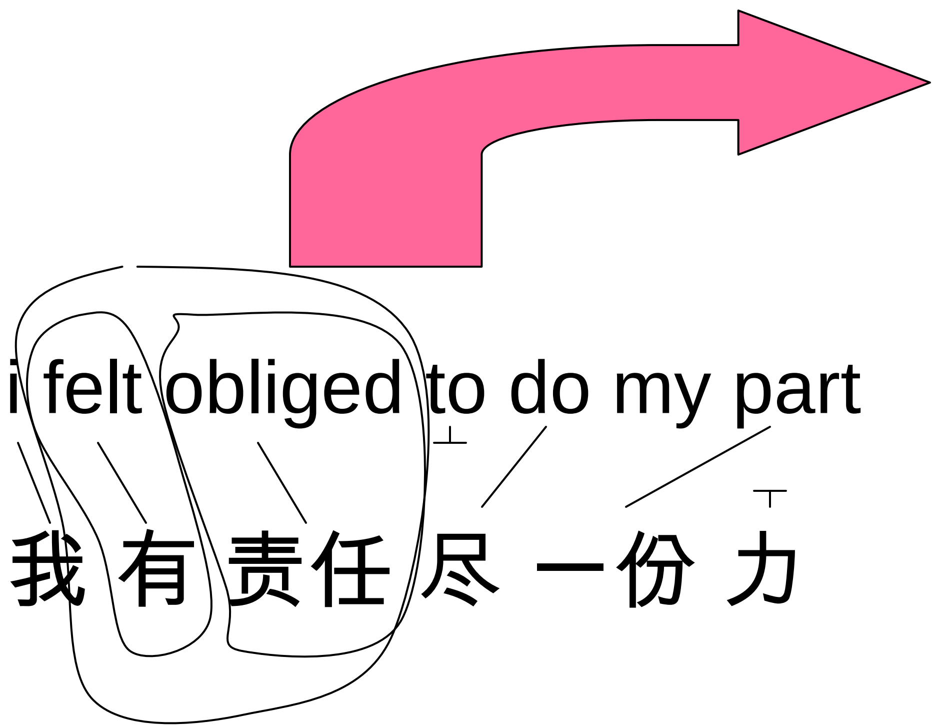
ATS (Och & Ney, 2004)



PHRASE PAIRS ACQUIRED:

felt	→	有
felt obliged	→	有 责任
felt obliged to do	→	有 责任 尽
obliged	→	责任
obliged to do	→	责任 尽
do	→	尽
part	→	一份
part	→	一份 力

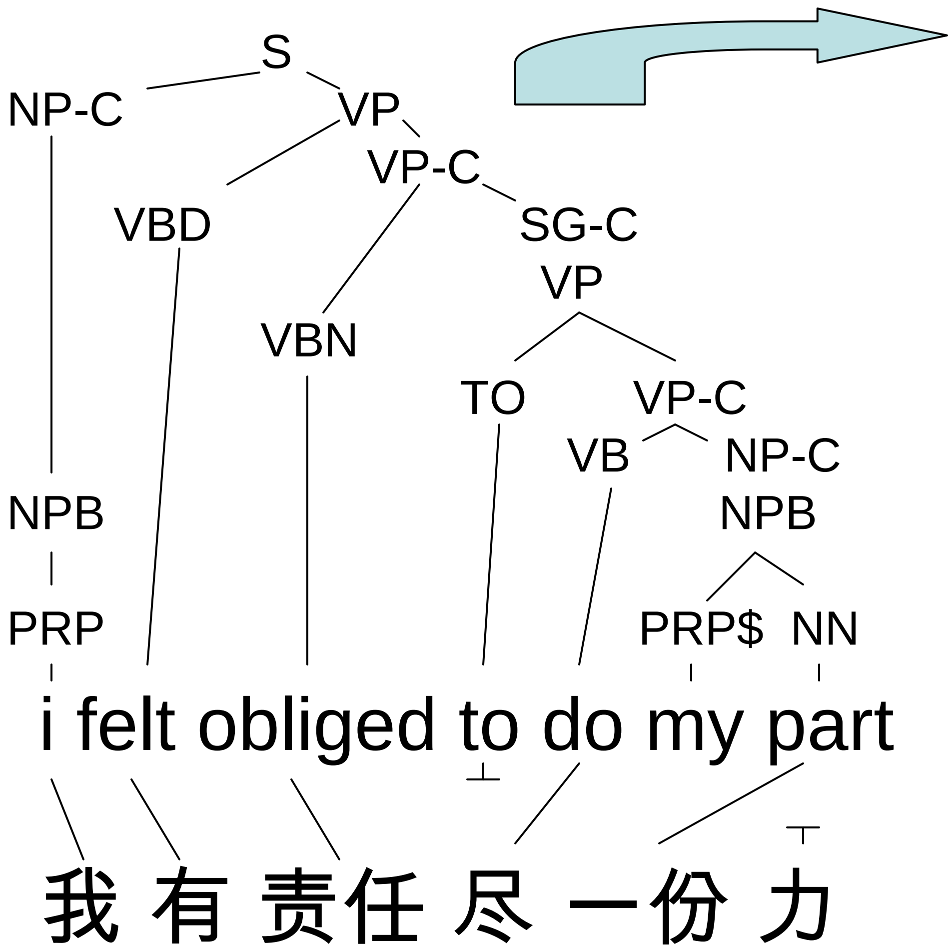
ATS (Och & Ney, 2004)



PHRASE PAIRS ACQUIRED:

felt	→	有
felt obliged	→	有 责任
felt obliged to do	→	有 责任 尽
obliged	→	责任
obliged to do	→	责任 尽
do	→	尽
part	→	一份
part	→	一份 力

GHKM (Galley et al, 2004)



RULES ACQUIRED:

VBD(felt) → 有

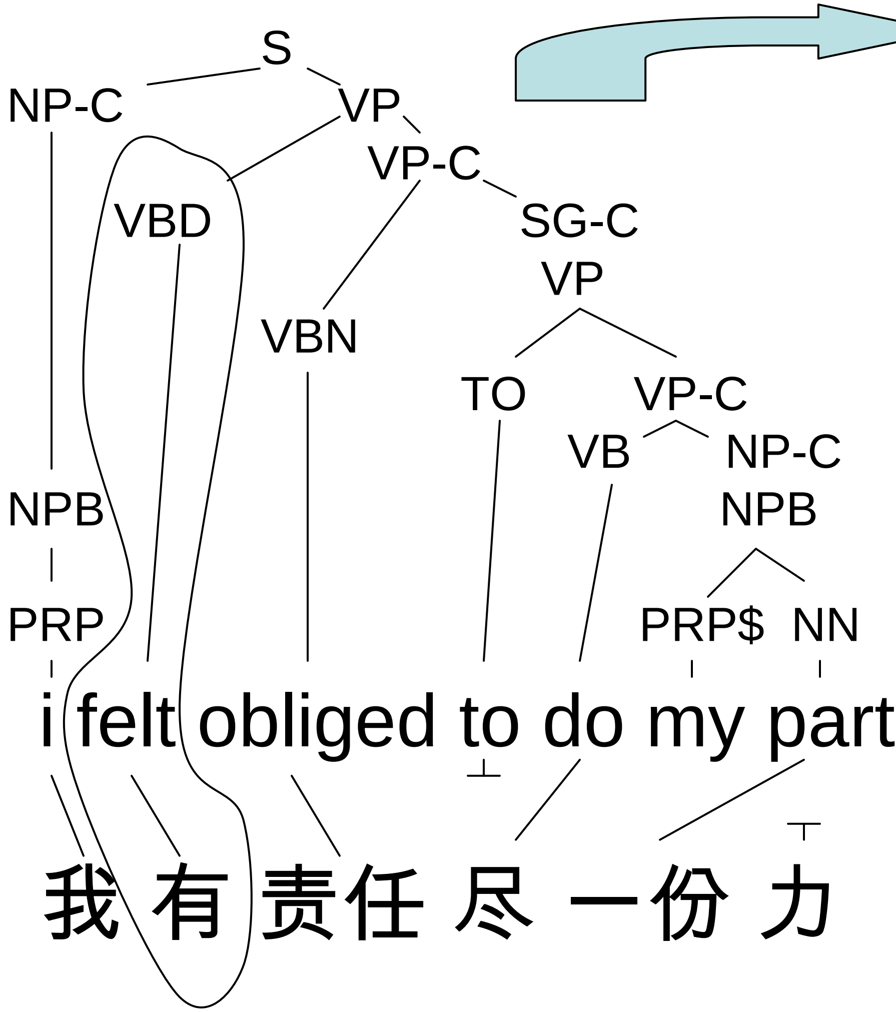
VBN(obliged) → 责任

VP(x0:VBD
VP-C(x1:VBN
x2:SG-C) → x0 x1 x2

VP(VBD(felt)
VP-C(VBN(obliged))
x0:SG-C) → 有 责任 x0

S(x0:NP-C x1:VP) → x0 x1

GHKM (Galley et al, 2004)



RULES ACQUIRED:

VBD(felt) → 有

VBN(obliged) → 责任

VP(x0:VBD
VP-C(x1:VBN
x2:SG-C) → x0 x1 x2

VP(VBD(felt)
VP-C(VBN(obliged))
x0:SG-C) → 有 责任 x0

S(x0:NP-C x1:VP) → x0 x1

GHKM (Galley et al, 2004)

RULES ACQUIRED:

VBD(felt) → 有

VBN(obliged) → 责任

VP(x0:VBD
VP-C(x1:VBN
x2:SG-C) → x0 x1 x2

VP(VBD(felt)
VP-C(VBN(obliged))
x0:SG-C) → 有 责任 x0

S(x0:NP-C x1:VP) → x0 x1

GHKM (Galley et al, 2004)

RULES ACQUIRED:

VBD(felt) → 有

VBN(obliged) → 责任

VP(x0:VBD
VP-C(x1:VBN
x2:SG-C) → x0 x1 x2

VP(VBD(felt)
VP-C(VBN(obliged))
x0:SG-C) → 有 责任 x0

S(x0:NP-C x1:VP) → x0 x1

GHKM (Galley et al, 2004)

RULES ACQUIRED:

VBD(felt) → 有

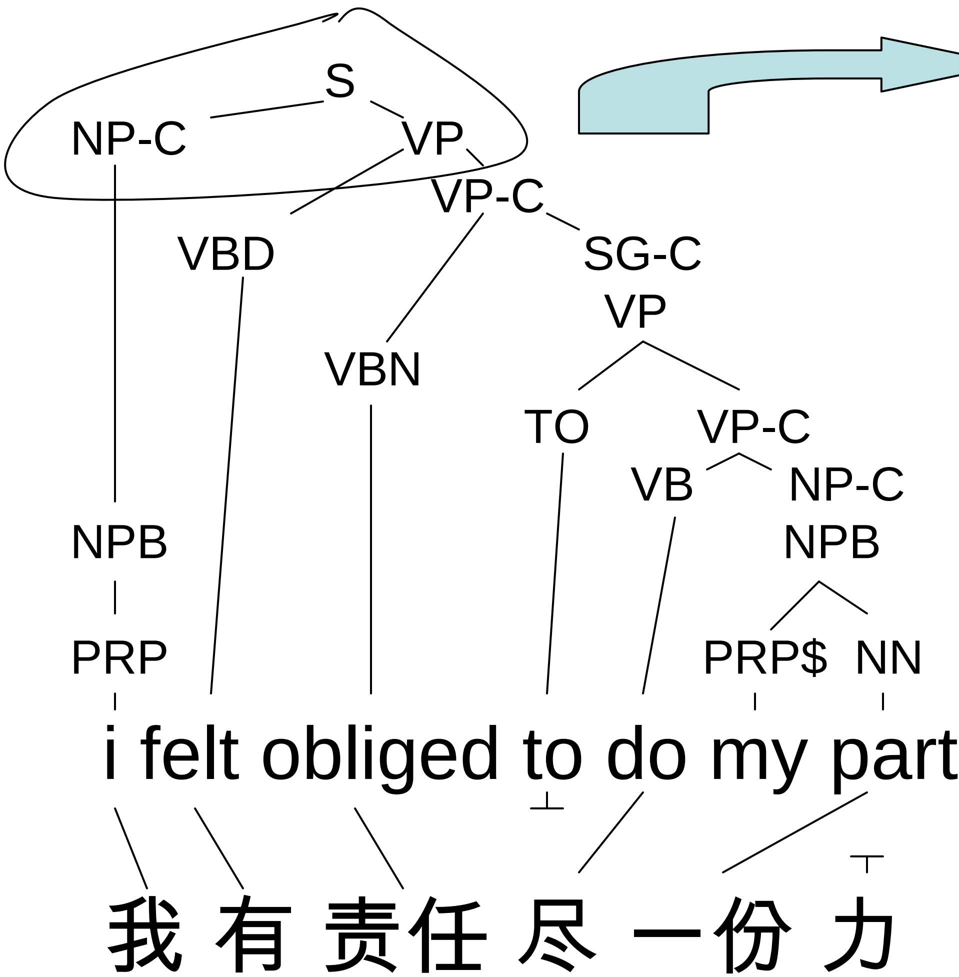
VBN(obliged) → 责任

VP(x0:VBD
VP-C(x1:VBN
x2:SG-C) → x0 x1 x2

VP(VBD(felt)
VP-C(VBN(obliged))
x0:SG-C) → 有 责任 x0

S(x0:NP-C x1:VP) → x0 x1

GHKM (Galley et al, 2004)



RULES ACQUIRED:

VBD(felt) → 有

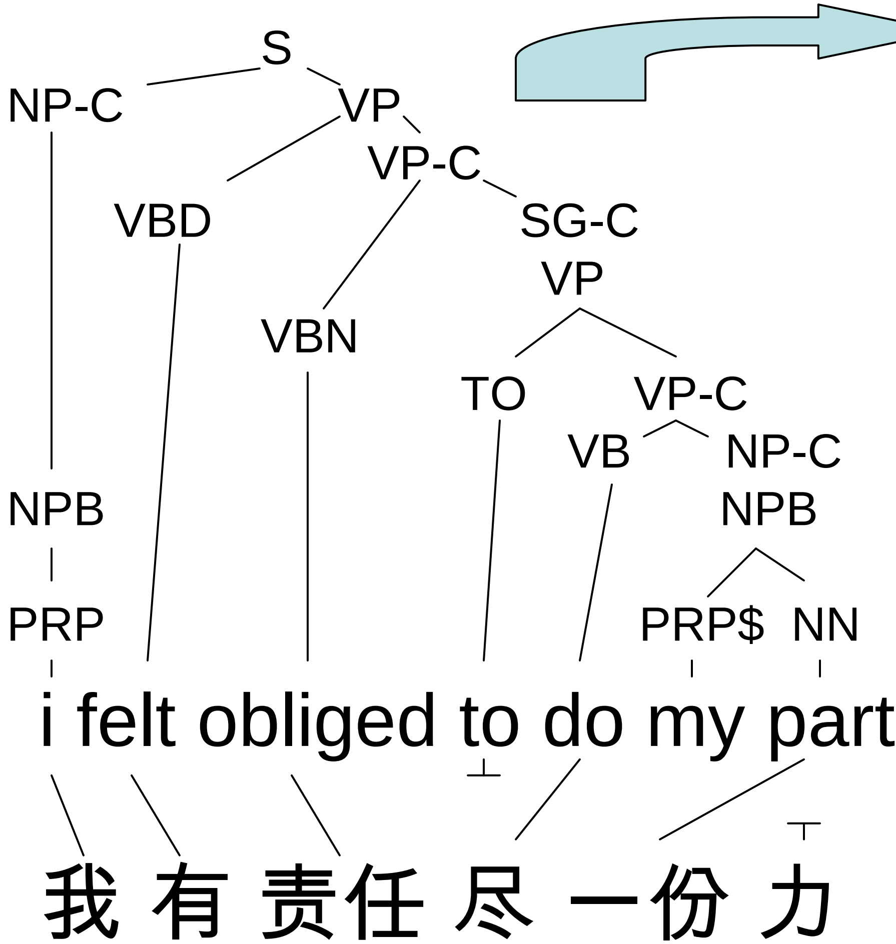
VBN(obliged) → 责任

VP(x0:VBD
VP-C(x1:VBN
x2:SG-C) → x0 x1 x2

VP(VBD(felt)
VP-C(VBN(obliged))
x0:SG-C) → 有 责任 x0

S(x0:NP-C x1:VP) → x0 x1

GHKM (Galley et al, 2004)



RULES ACQUIRED:

VBD(felt) → 有

VBN(obliged) → 责任

VP(x0:VBD
VP-C(x1:VBN
x2:SG-C) → x0 x1 x2

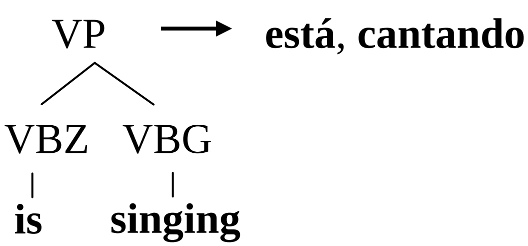
VP(VBD(felt)
VP-C(VBN(obliged))
x0:SG-C) → 有 责任 x0

S(x0:NP-C x1:VP) → x0 x1

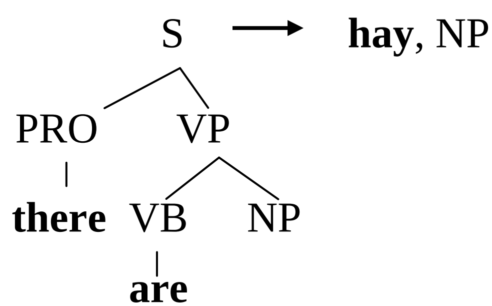
minimal rules tile the tree/string/alignment triple.
composed rules are made by combining those tiles.

GHKM Syntax Rules

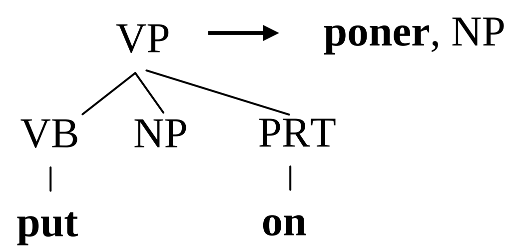
Phrasal Translation



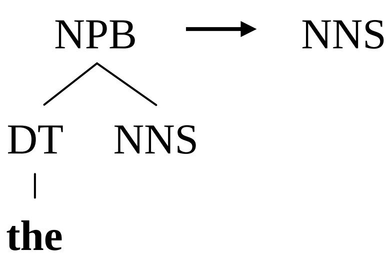
Non-constituent Phrases



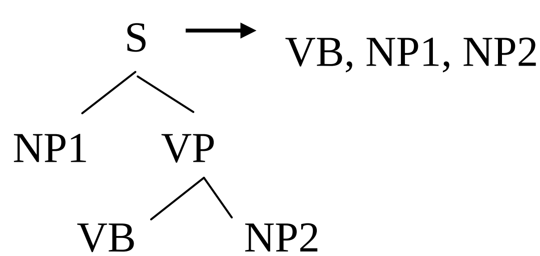
Non-contiguous Phrases



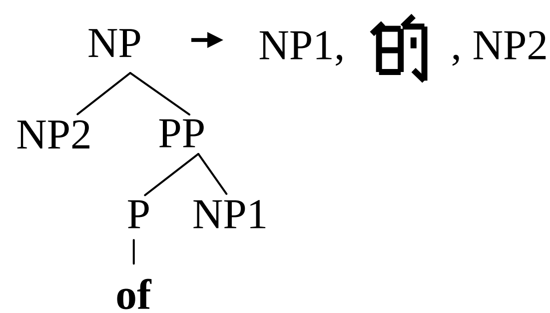
Context-Sensitive Word Insertion



Multilevel Re-Ordering

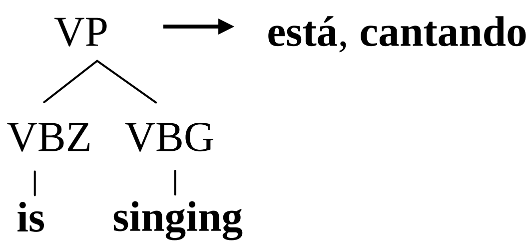


Lexicalized Re-Ordering

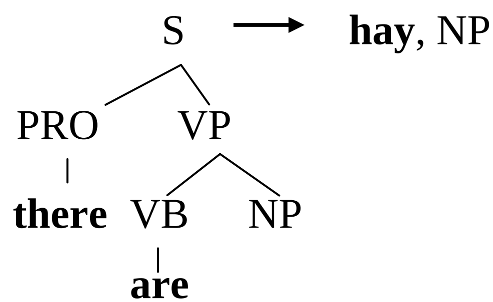


GHKM Syntax Rules

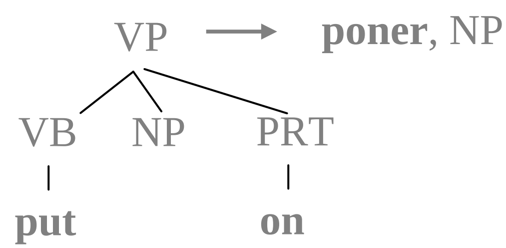
Phrasal Translation



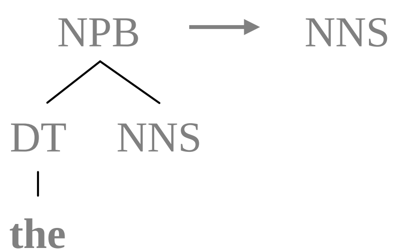
Non-constituent Phrases



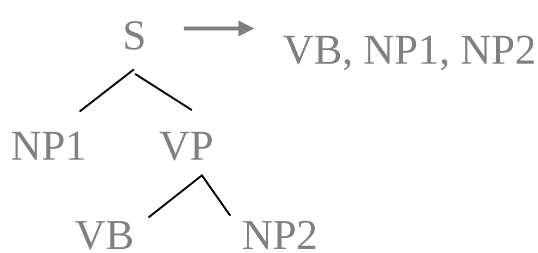
Non-contiguous Phrases



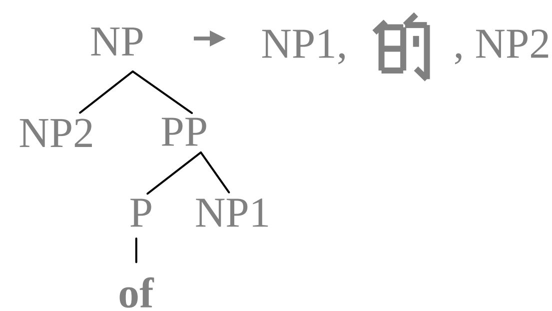
Context-Sensitive Word Insertion



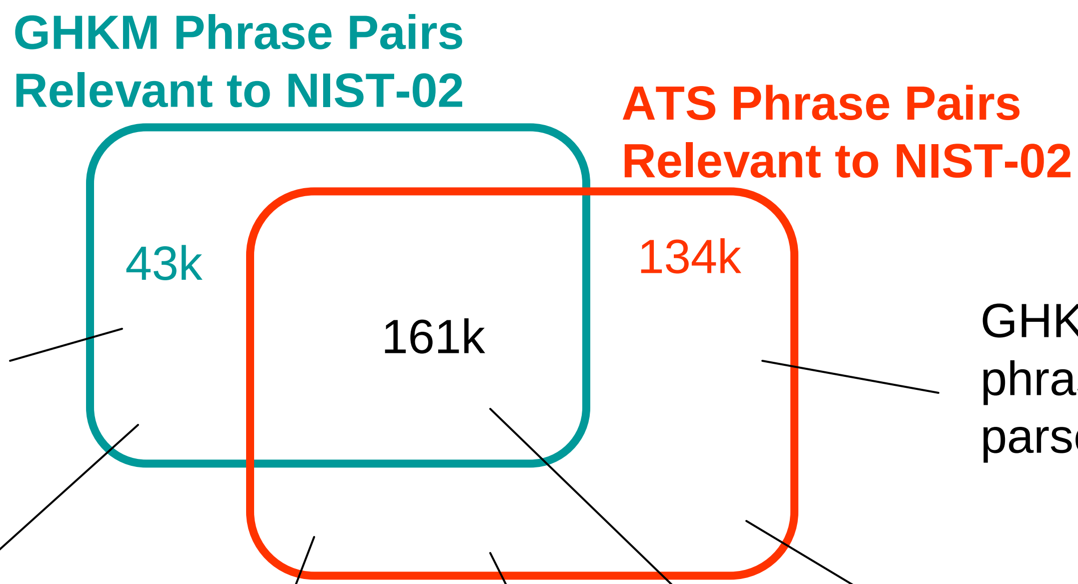
Multilevel Re-Ordering



Lexicalized Re-Ordering



ATS and GHKM Methods Do Not Coincide



GHKM has no built-in phrase size limit -- ATS does.

GHKM pulls unaligned English words into phrases.

GHKM only gets *minimal* rules to explain each segment pair.

GHKM forced to incorporate unaligned English words into phrases.

GHKM misses phrases due to parse failures.

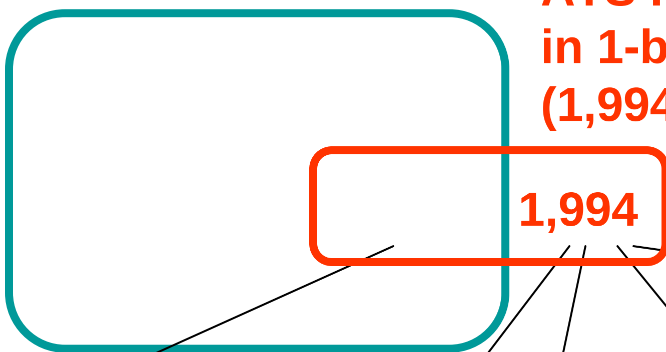
GHKM forced to incorporate some unaligned foreign words into phrases.

GHKM phrases come with applicability conditions.

ATS and GHKM Methods Overlap

GHKM Phrase Pairs Relevant to NIST-02

ATS Phrase Pairs <u>actually used</u> in 1-best decodings of NIST-02 (1,994 = 2 per sentence).



GHKM misses phrases due to parse failures.

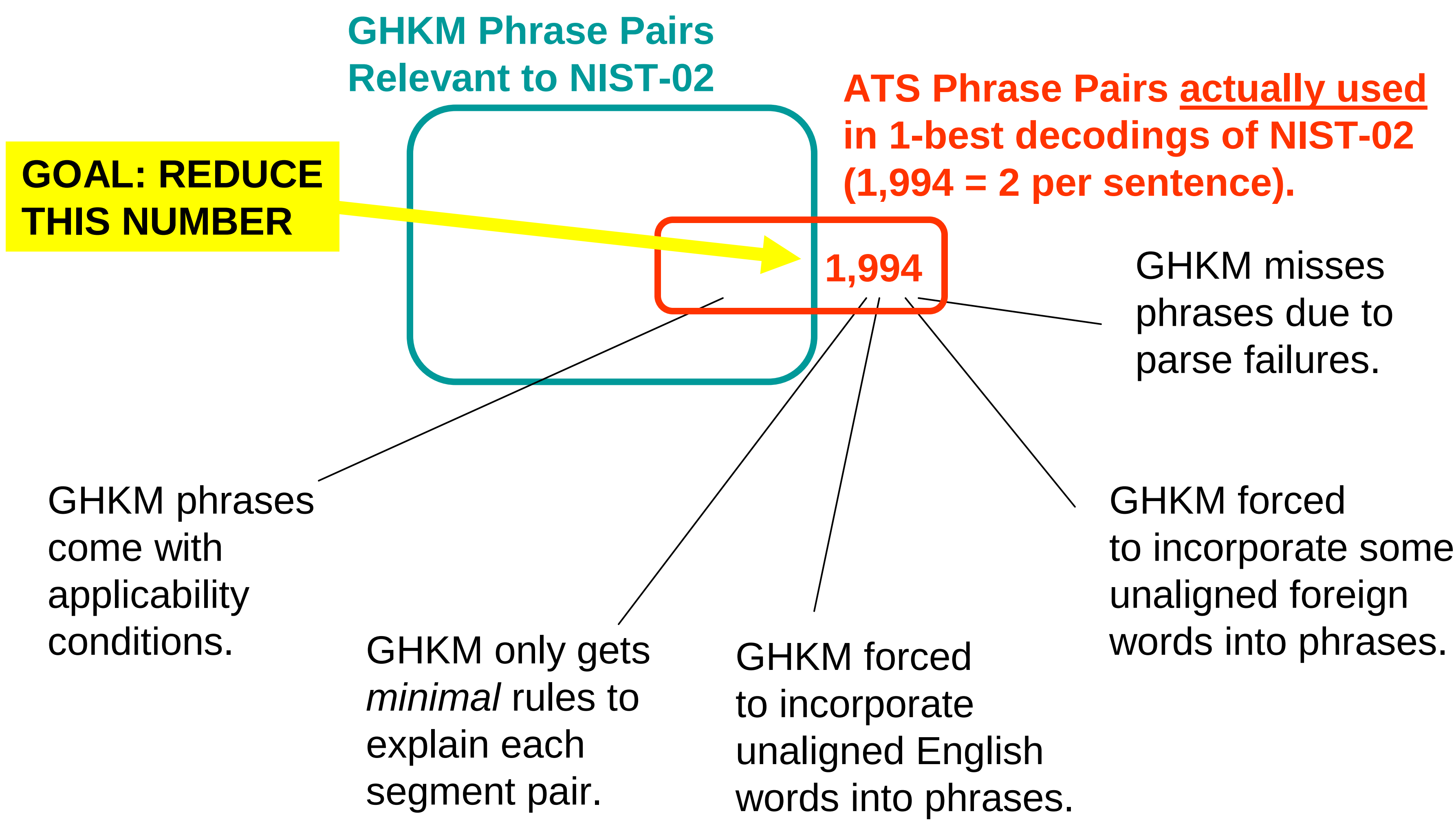
GHKM phrases come with applicability conditions.

GHKM only gets *minimal* rules to explain each segment pair.

GHKM forced to incorporate unaligned English words into phrases.

GHKM forced to incorporate some unaligned foreign words into phrases.

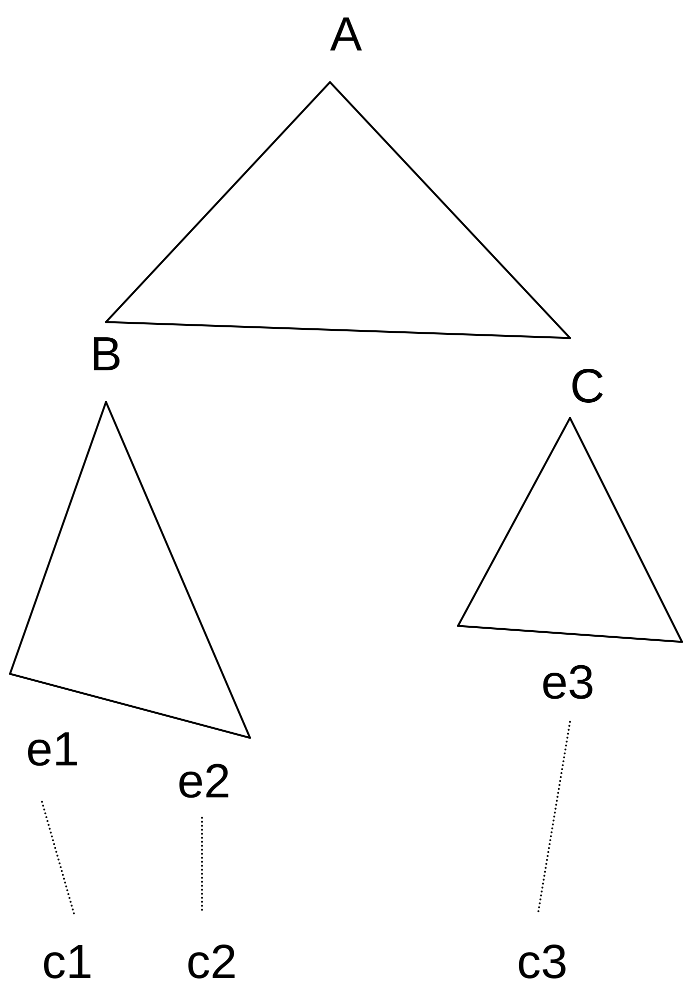
ATS and GHKM Methods Overlap



Four Ideas for Improving Syntax-Based Rule Extraction

- Acquire larger rules
  - Composed rules (Galley et al, 06)
  - Phrasal rules (Marcu et al, 06)
- Acquire more general rules
  - Re-structure English trees (Wang et al, 07)
  - Re-align tree/string pairs (May & Knight, 07)

Larger, Composed Rules



Minimal GHKM Rules:

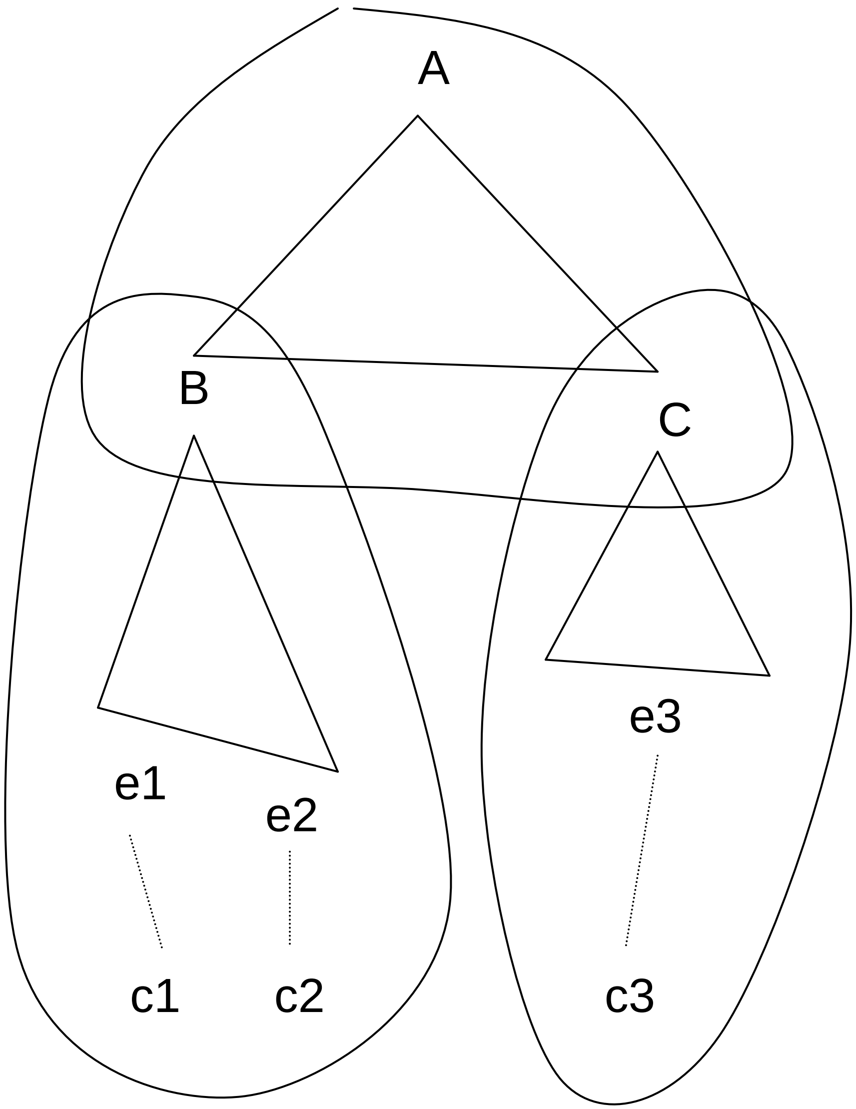
B(e1 e2) → c1 c2
C(e3) → c3
A(x0:B x1:C) → x0 x1

Additional Composed Rules:

A(B(e1 e2) x0:C) -> c1 c2 x0
A(x0:B C(e3)) -> x0 c3
A(B(e1 e2) C(e3)) -> c1 c2 c3

"big phrasal rule"

Larger, Composed Rules



Minimal GHKM Rules:

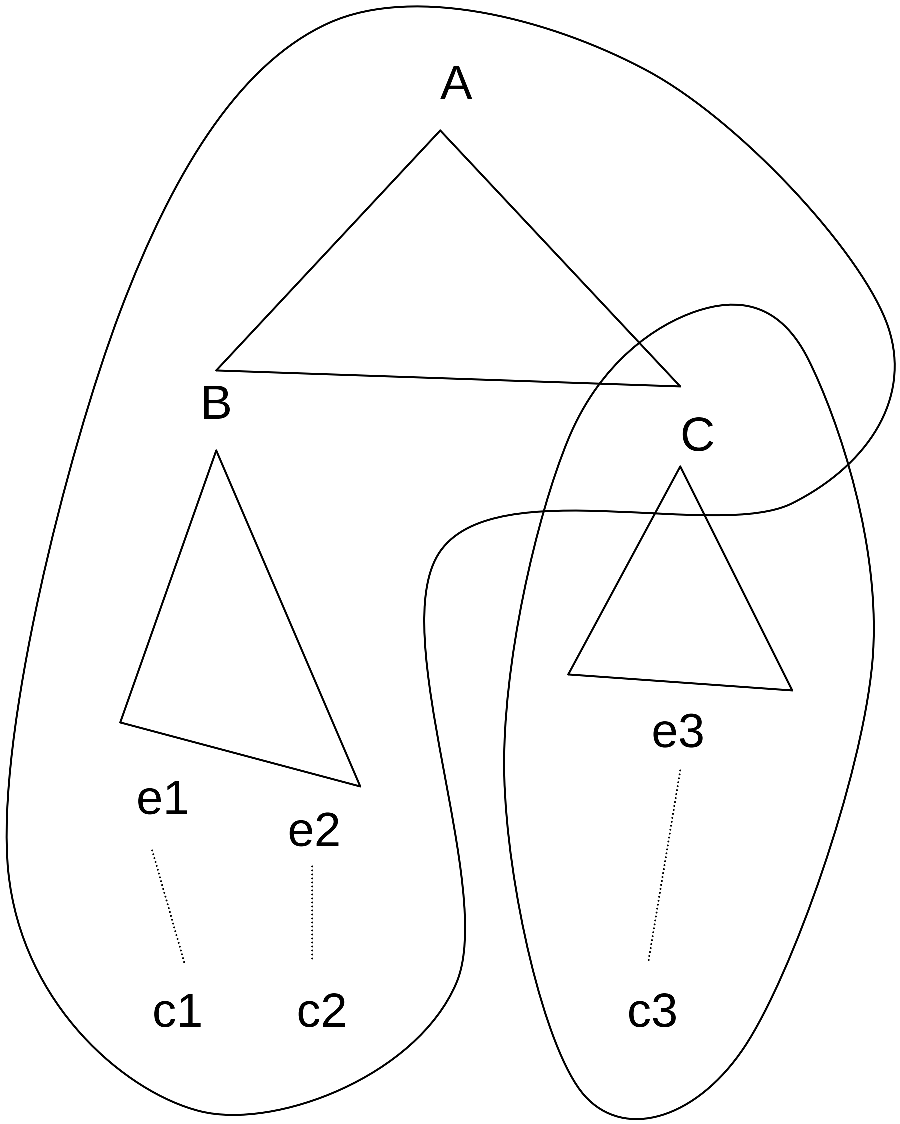
B(e1 e2) → c1 c2
C(e3) → c3
A(x0:B x1:C) → x0 x1

Additional Composed Rules:

A(B(e1 e2) x0:C) -> c1 c2 x0
A(x0:B C(e3)) -> x0 c3
A(B(e1 e2) C(e3)) -> c1 c2 c3

"big phrasal rule"

Larger, Composed Rules



Minimal GHKM Rules:

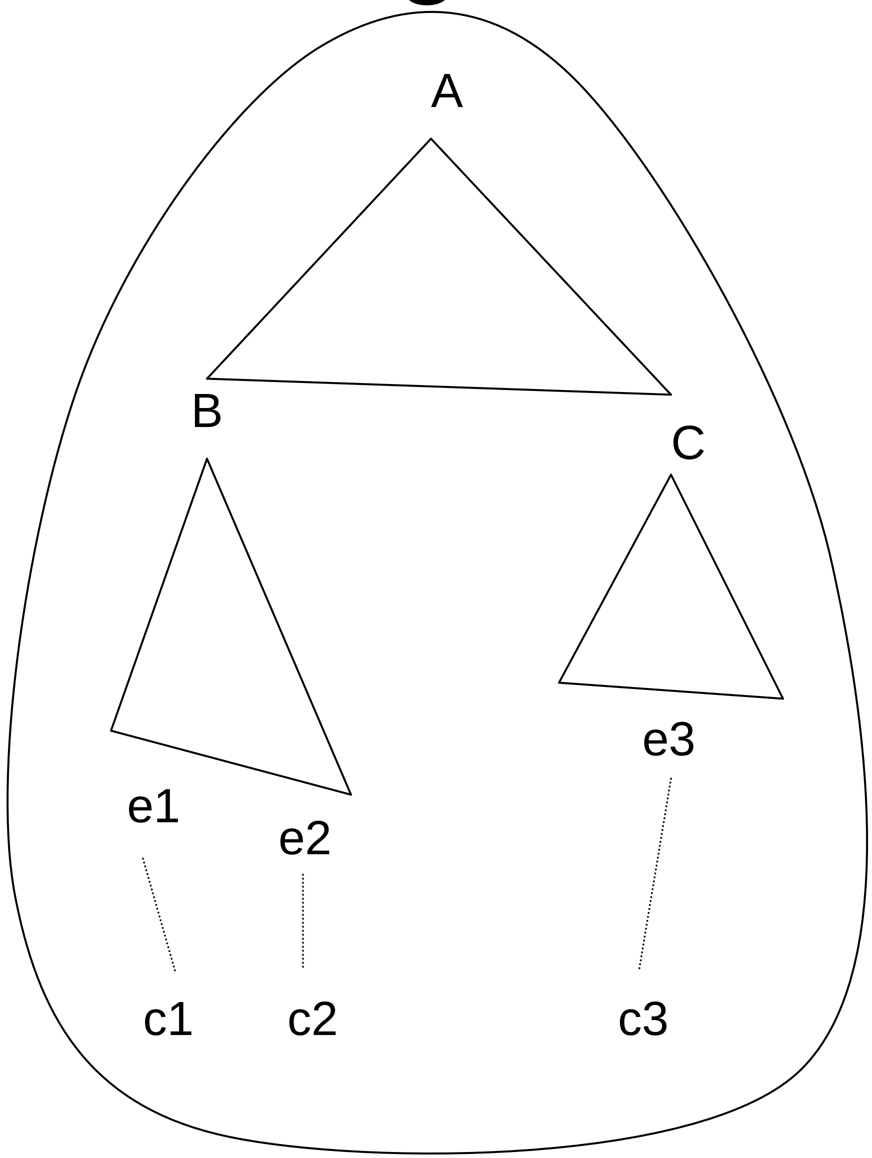
B(e1 e2) → c1 c2
C(e3) → c3
A(x0:B x1:C) → x0 x1

Additional Composed Rules:

A(B(e1 e2) x0:C) -> c1 c2 x0
A(x0:B C(e3)) -> x0 c3
A(B(e1 e2) C(e3)) -> c1 c2 c3

“big phrasal rule”

Larger, Composed Rules



Minimal GHKM Rules:

B(e1 e2) → c1 c2
C(e3) → c3
A(x0:B x1:C) → x0 x1

Additional Composed Rules:

A(B(e1 e2) x0:C) -> c1 c2 x0
A(x0:B C(e3)) -> x0 c3
A(B(e1 e2) C(e3)) -> c1 c2 c3

“big phrasal rule”

Larger, Composed Rules

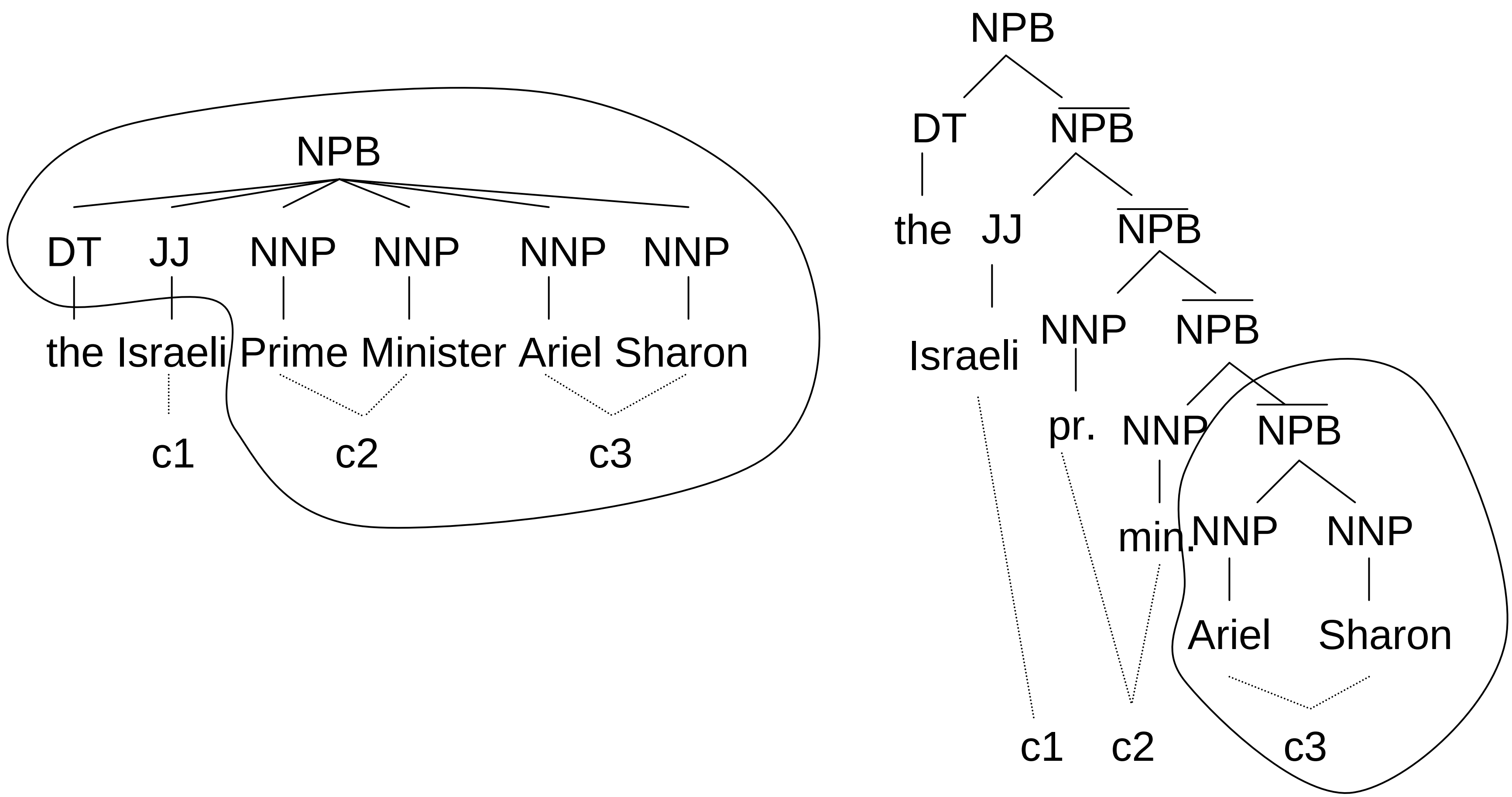
Composed limit (internal nodes in composed rule)	# of rules acquired	Unacquired phrase pairs used in ATS 1-best decodings
0 = minimal	2.5m	1994
2	12.4m	1478
3	26.9m	1096
4	**55.8m**	**900**

“Phrasal” Syntax Rules

- SPMT Model 1 (Marcu et al 2006)
  - consider each foreign phrase up to length L
  - extract smallest possible syntax rule that does not violate alignments

Method	Unacquired ATS Phrase Pairs
Minimal	1994
Composed 4	900
SPMT M1	**676**
Both	**663**

Restructuring English Training Trees



Restructuring English Training Trees

Method	Unacquired ATS Phrase Pairs
Minimal	1994
+ Composed 4	900
+ SPMT M1	663
+ Restructuring	**458**

Effects of Coverage Improvements on Syntax-Based MT Accuracy

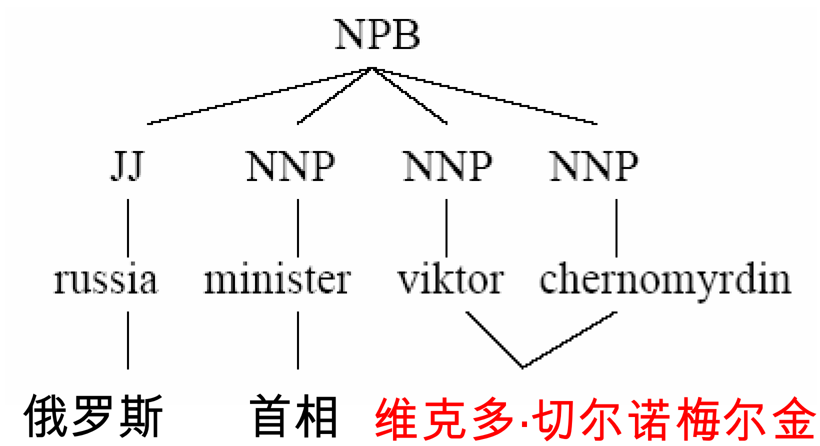
	Chinese/English Trained on 9.8m words		Arabic/English Trained on 4.1m words	
	Dev-02	Test-03	Dev-02	Test-03
ATS	36.00	34.31	50.88	51.04
GHKM minimal	39.11	38.85	49.81	50.46
GHKM composed 2	41.59	40.90	51.18	51.52
GHKM composed 3	42.28	41.62	51.96	52.04
GHKM composed 4	42.63	41.82	52.05	52.26
GHKM minimal + SPMT	41.01	40.34	50.74	51.81
GHKM composed 4 + SPMT	43.30	42.17	52.15	52.12
+ Left binarization of etrees	43.45	42.41	52.86	52.42

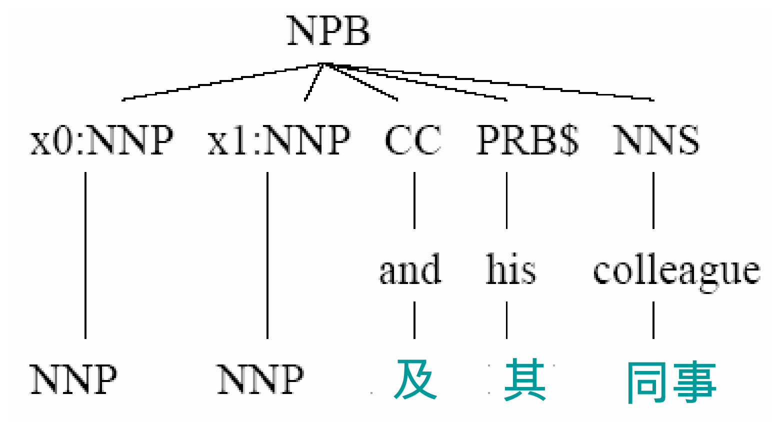
NIST Bleu r4n4

Can We Do Better?

- Improved binarization methods
- Improved word alignment of tree/string pairs

Why are Penn Treebank Trees Problematic?

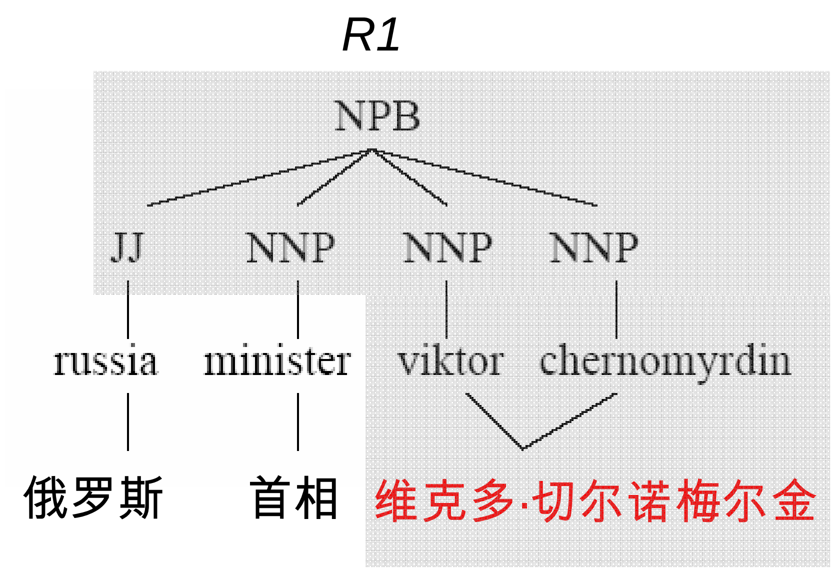


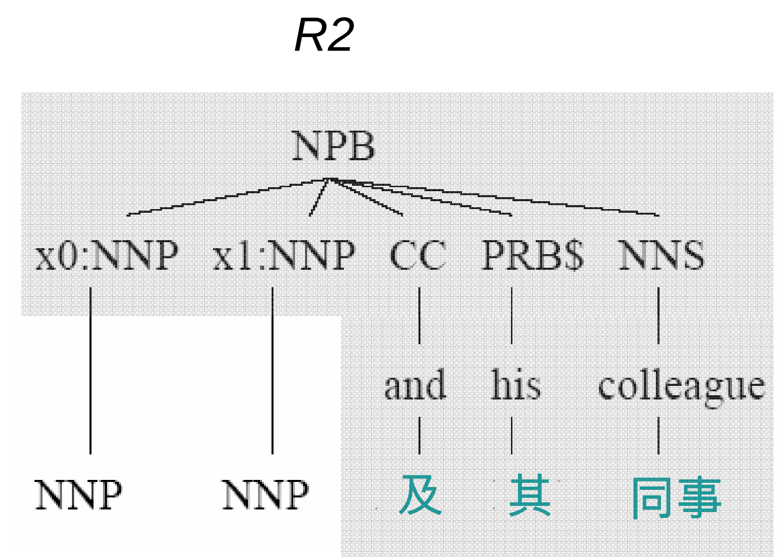


?

维克多·切尔诺梅尔金 及 其 同事

Why are Penn Treebank Trees Problematic?

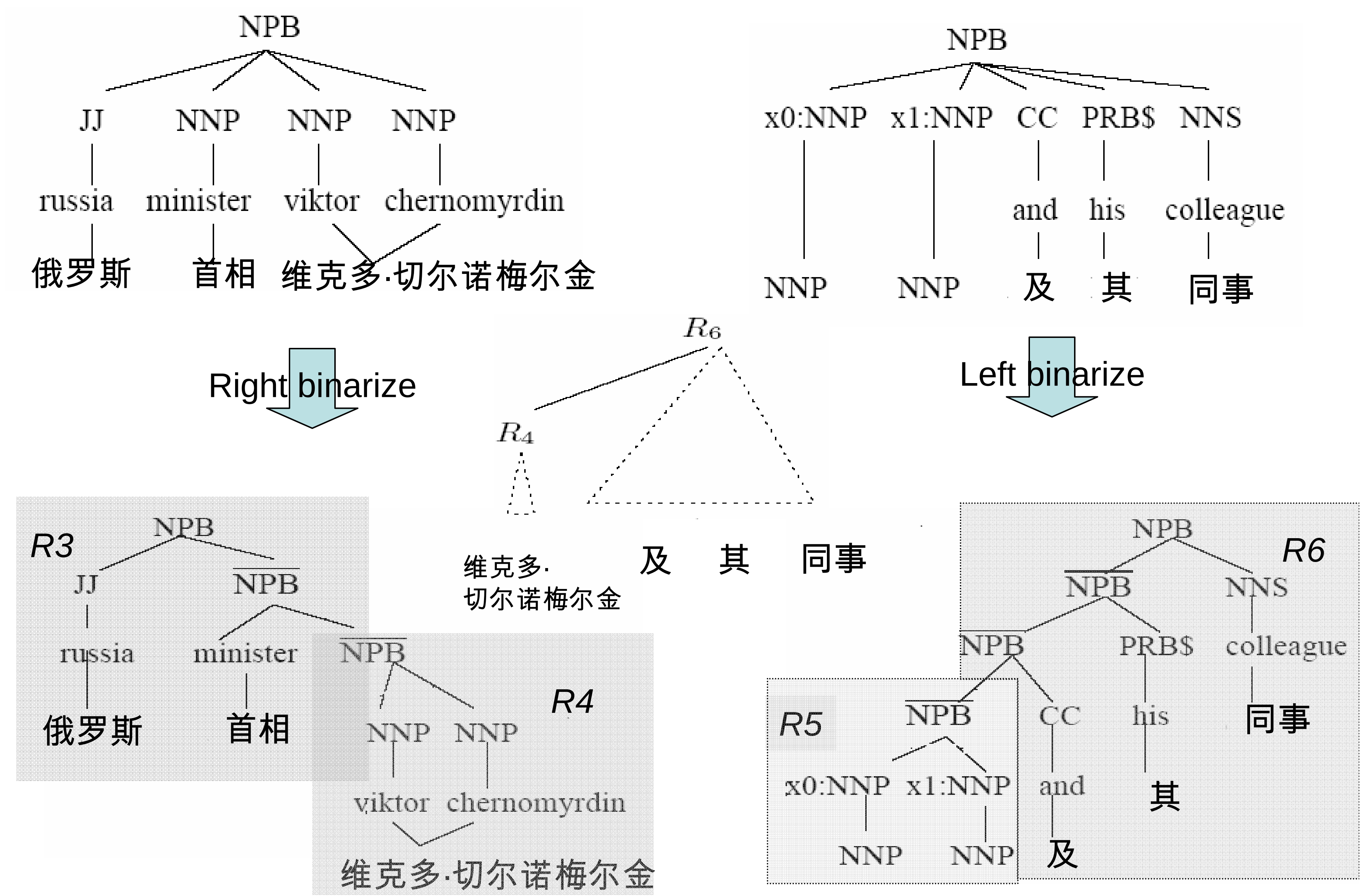




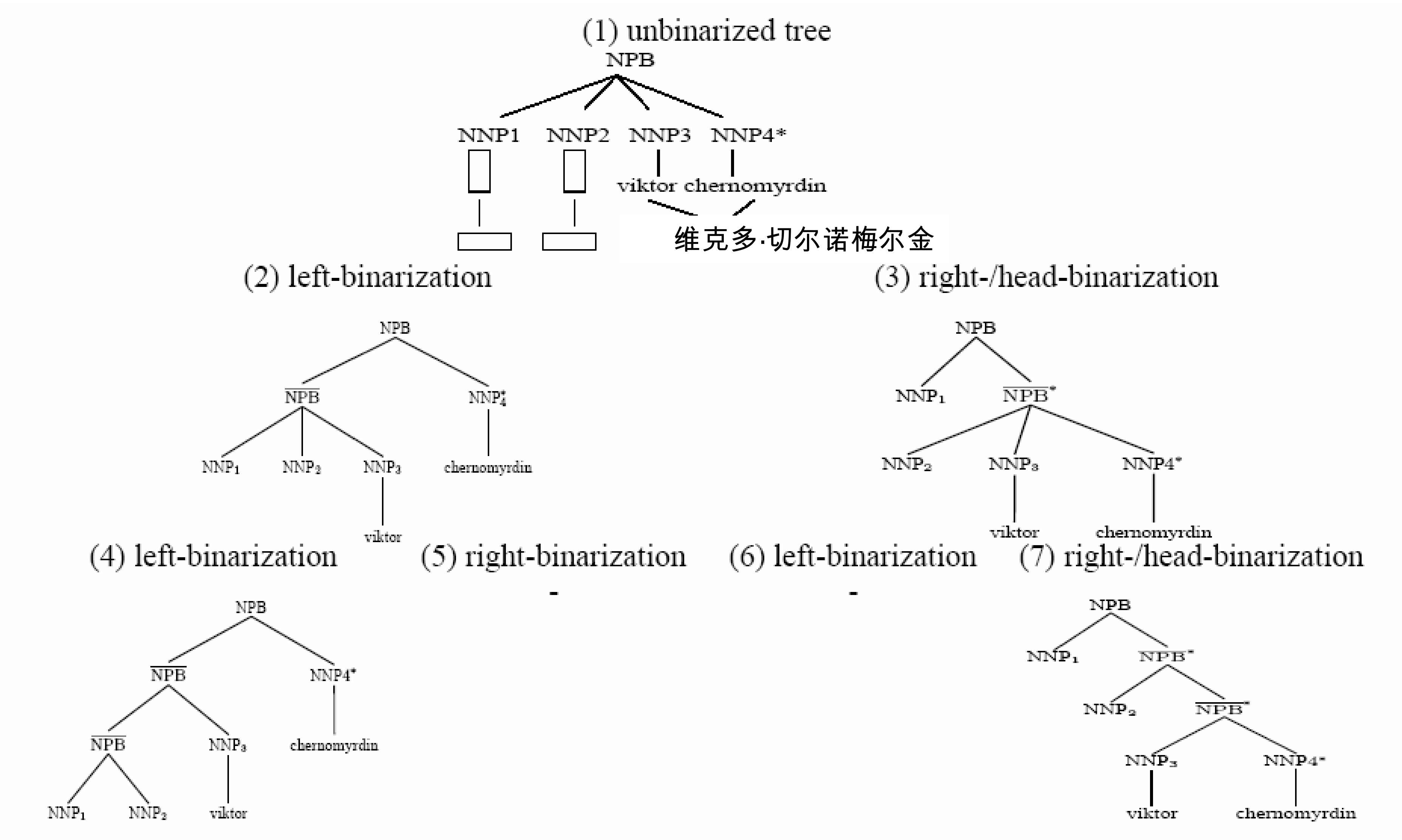
?

维克多·切尔诺梅尔金　及　其　同事

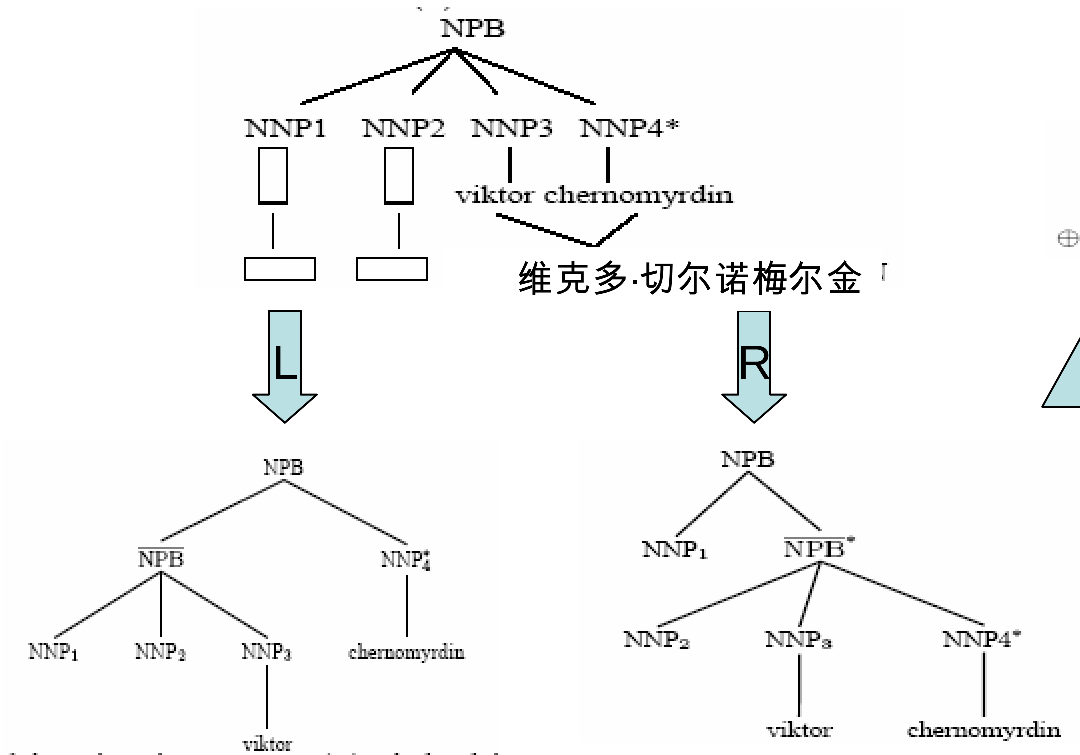
Binarizing English Trees

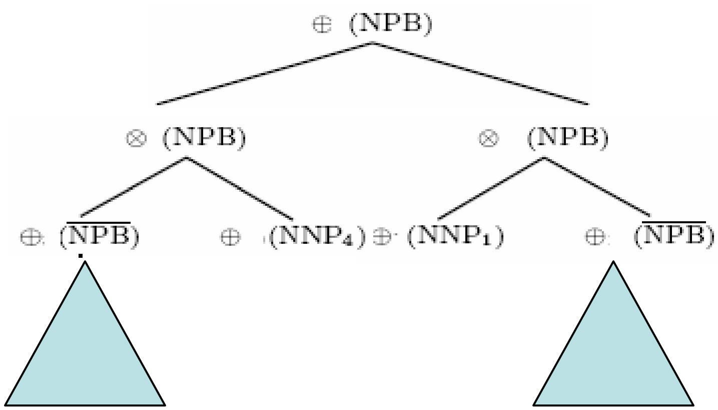


Simple Binarizations

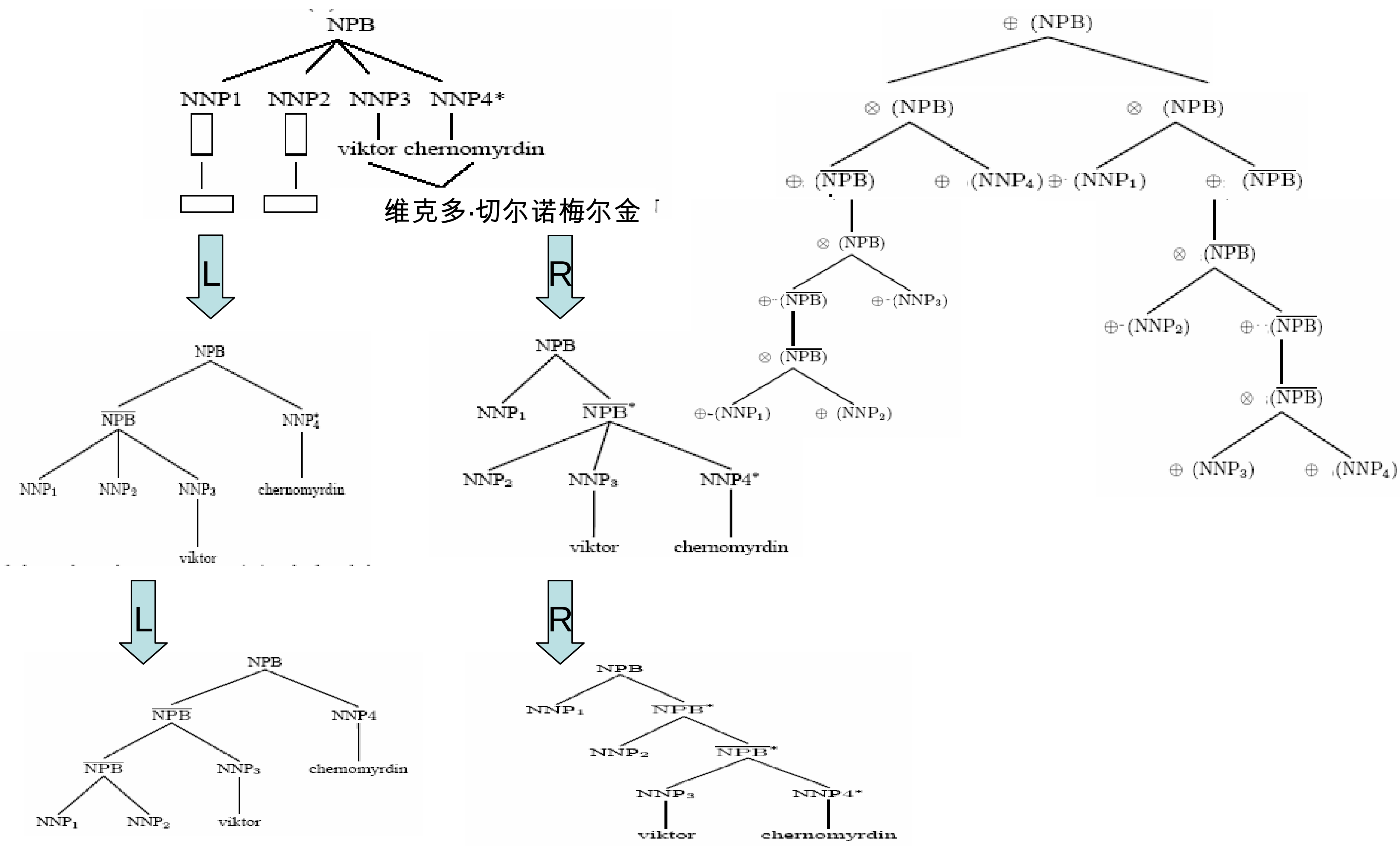


Parallel Binarization





Parallel Binarization



Forest-Based Rule Extraction

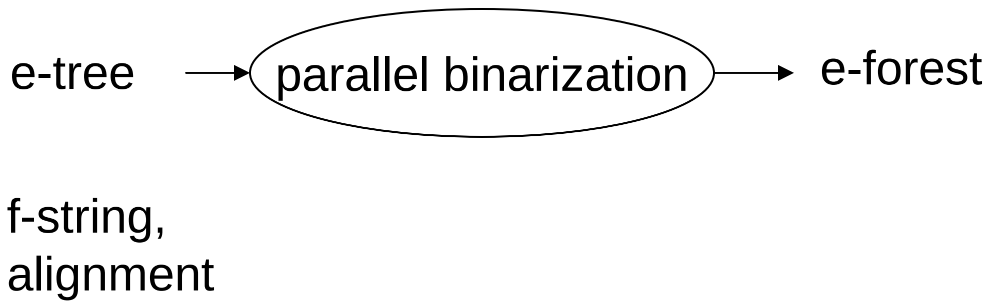
- Gets **all** minimal rules consistent with word alignment and **some** binarization
- Run EM algorithm to determine best binarization of each node in each tree

Binarization Using EM

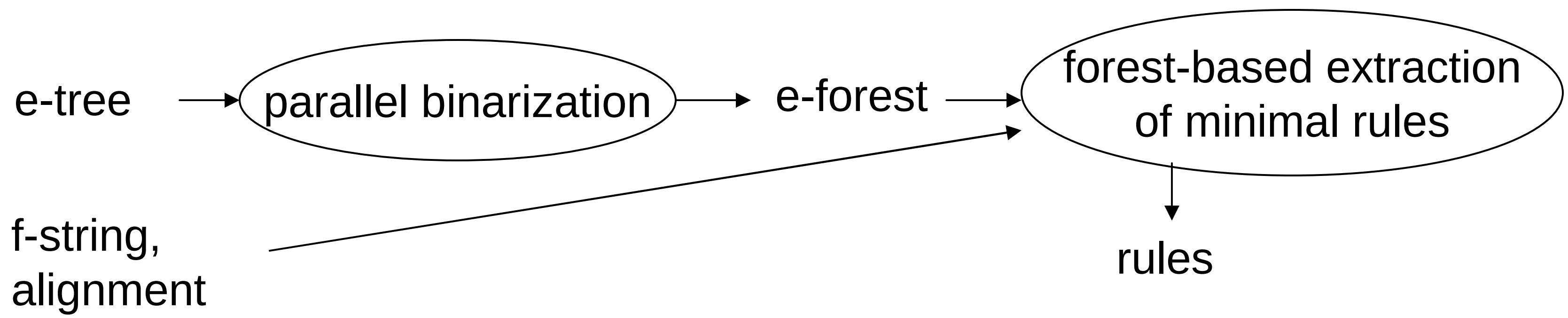
e-tree

f-string,
alignment

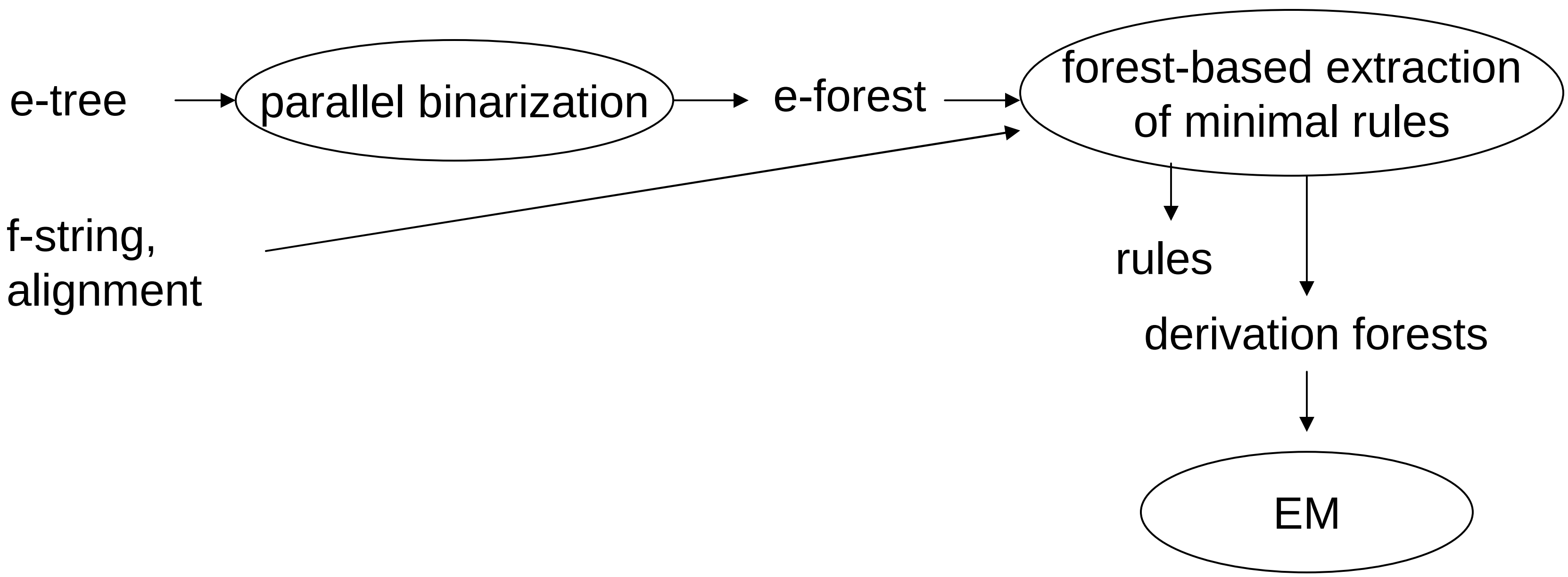
Binarization Using EM



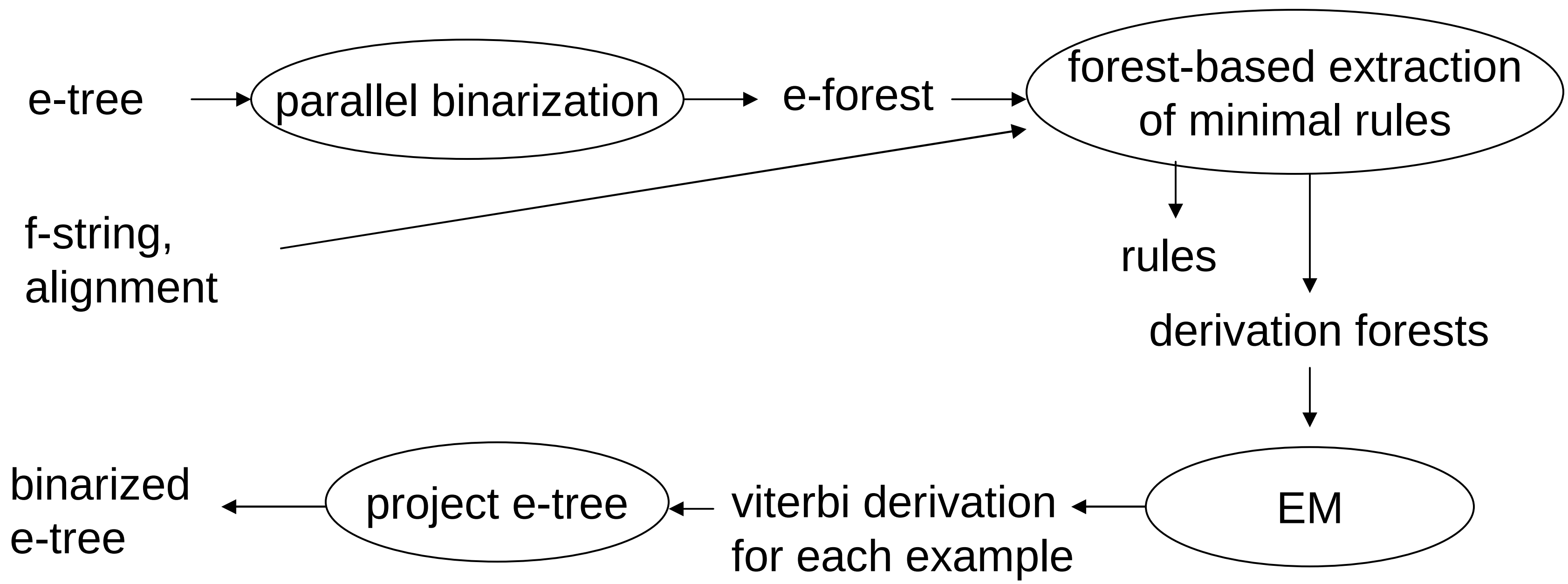
Binarization Using EM



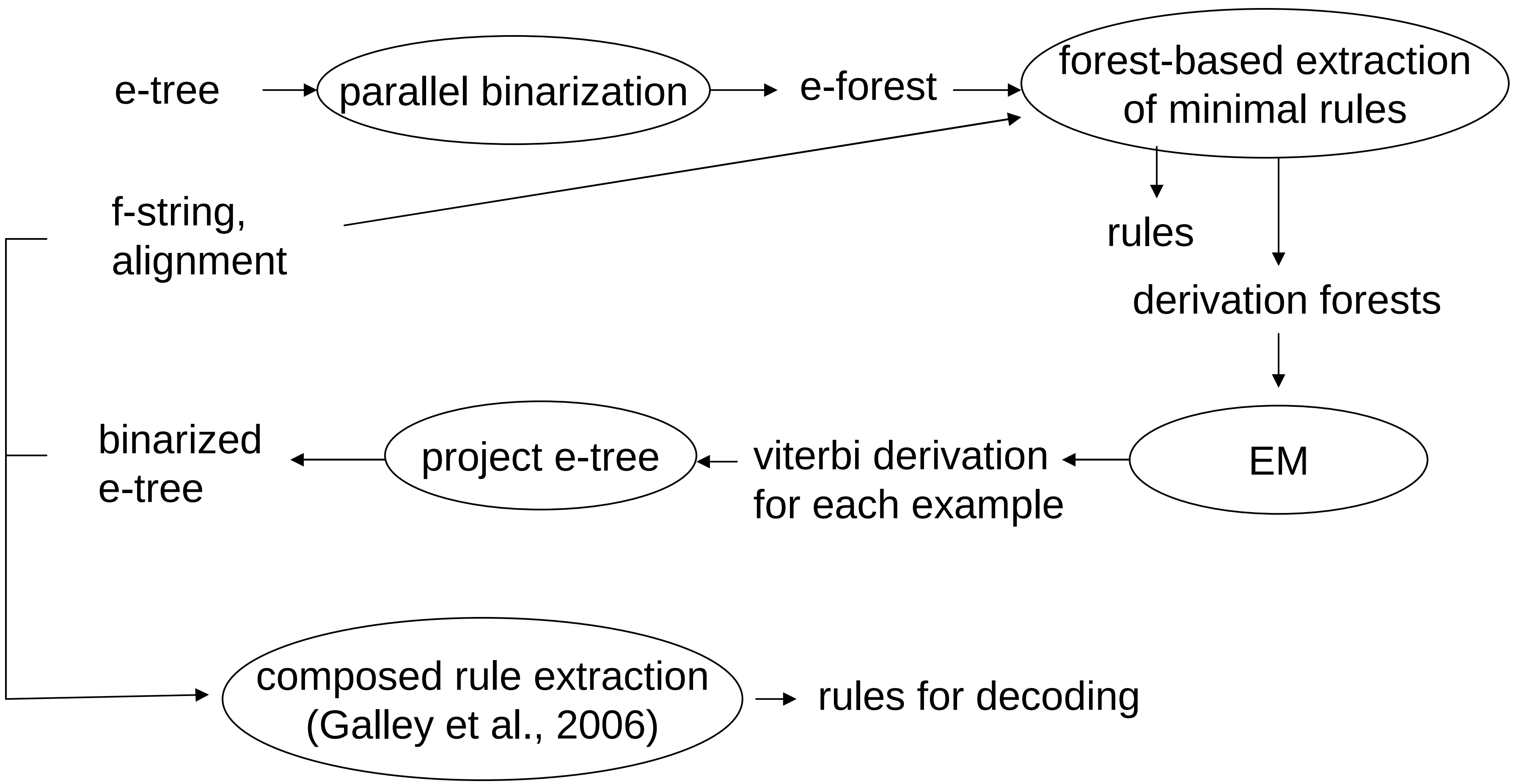
Binarization Using EM



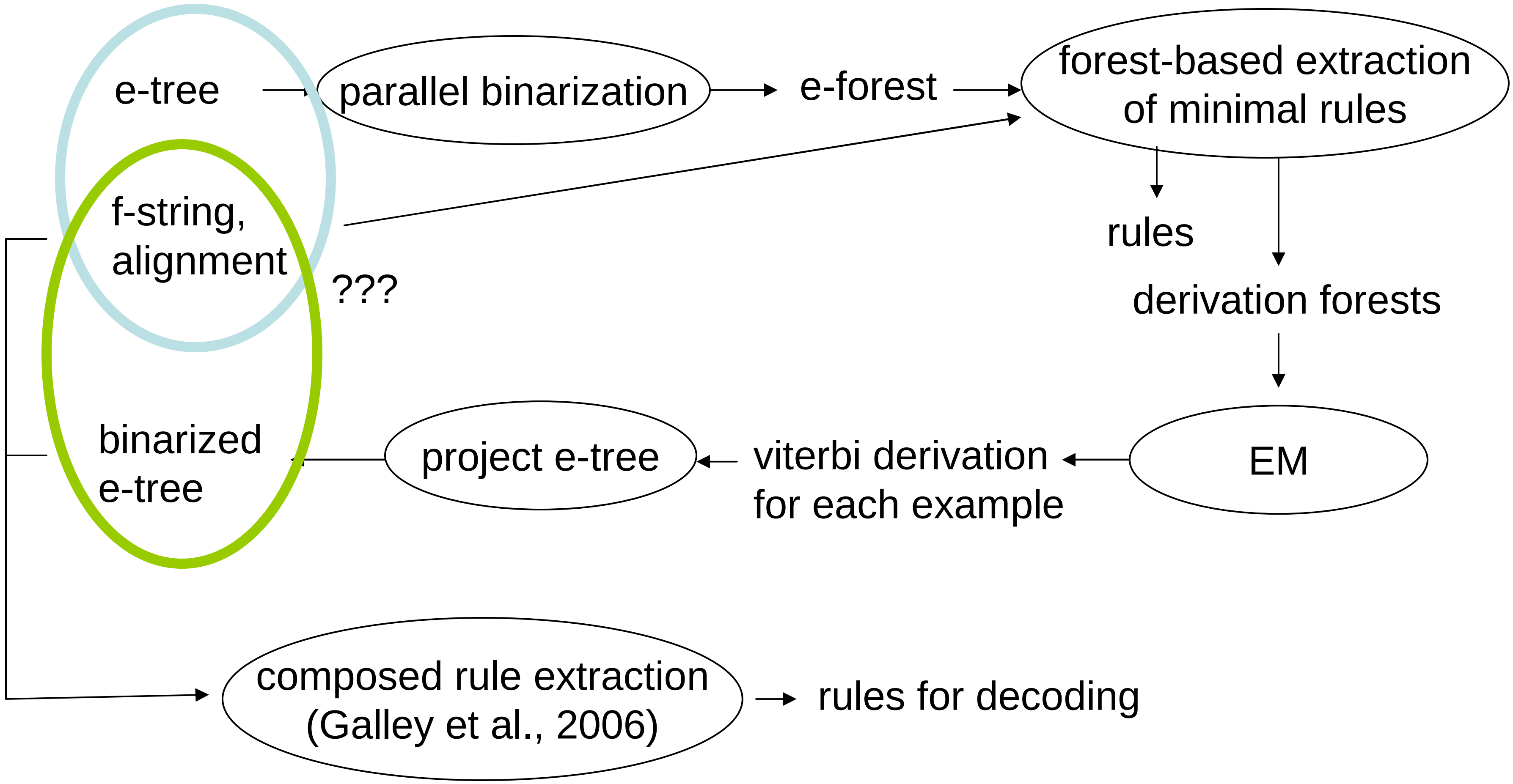
Binarization Using EM



Binarization Using EM



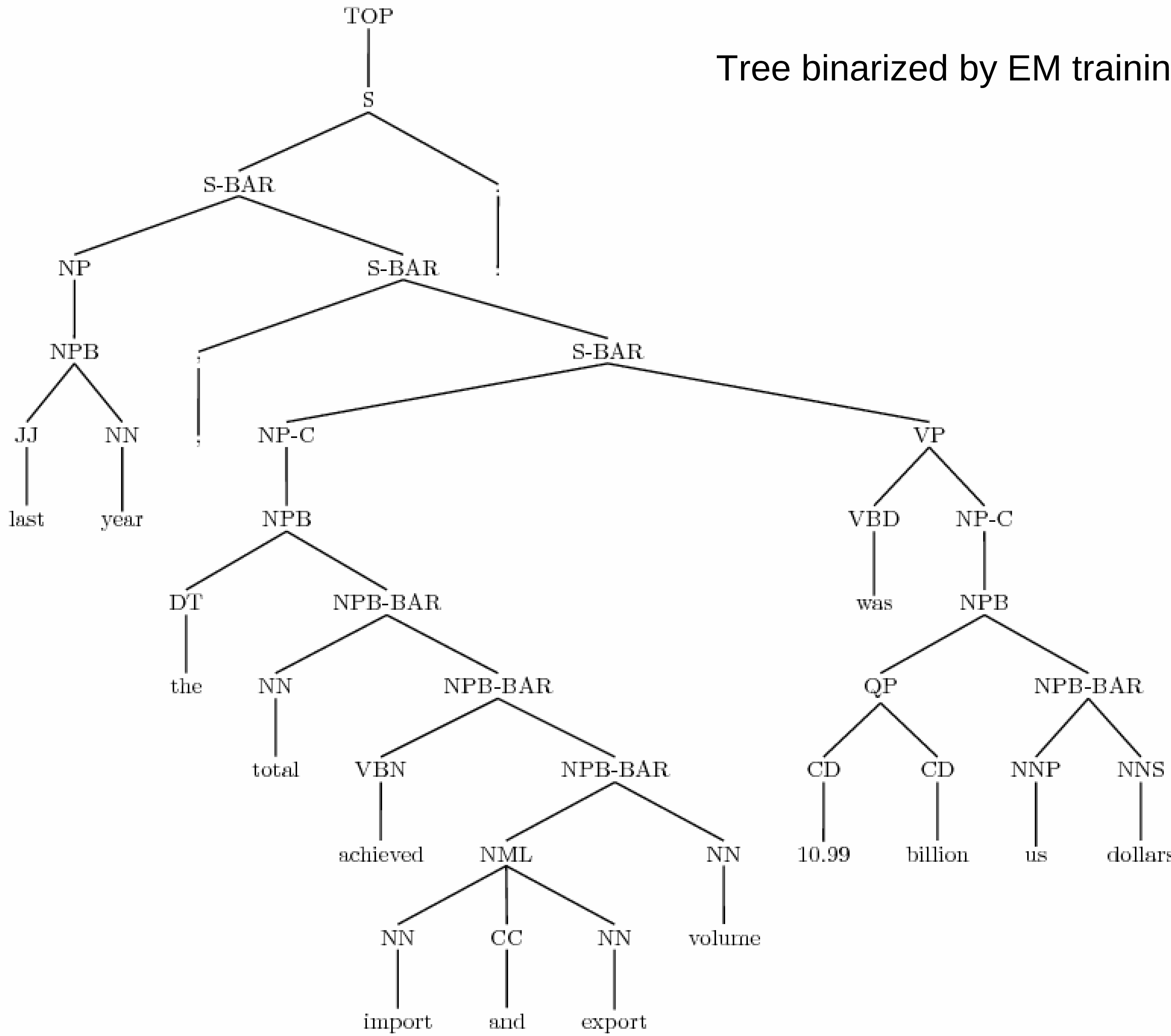
Binarization Using EM



Experimental Results

Type of Binarization	# of Rules Learned	Test Bleu (NIST-03)
None	63.4m	36.94
Left	114.0m	37.47 (p=0.047)
Right	113.0m	37.49 (p=0.044)
Head	113.8m	37.54 (p=0.086)
EM	115.6m	**37.94** (p=0.0047)

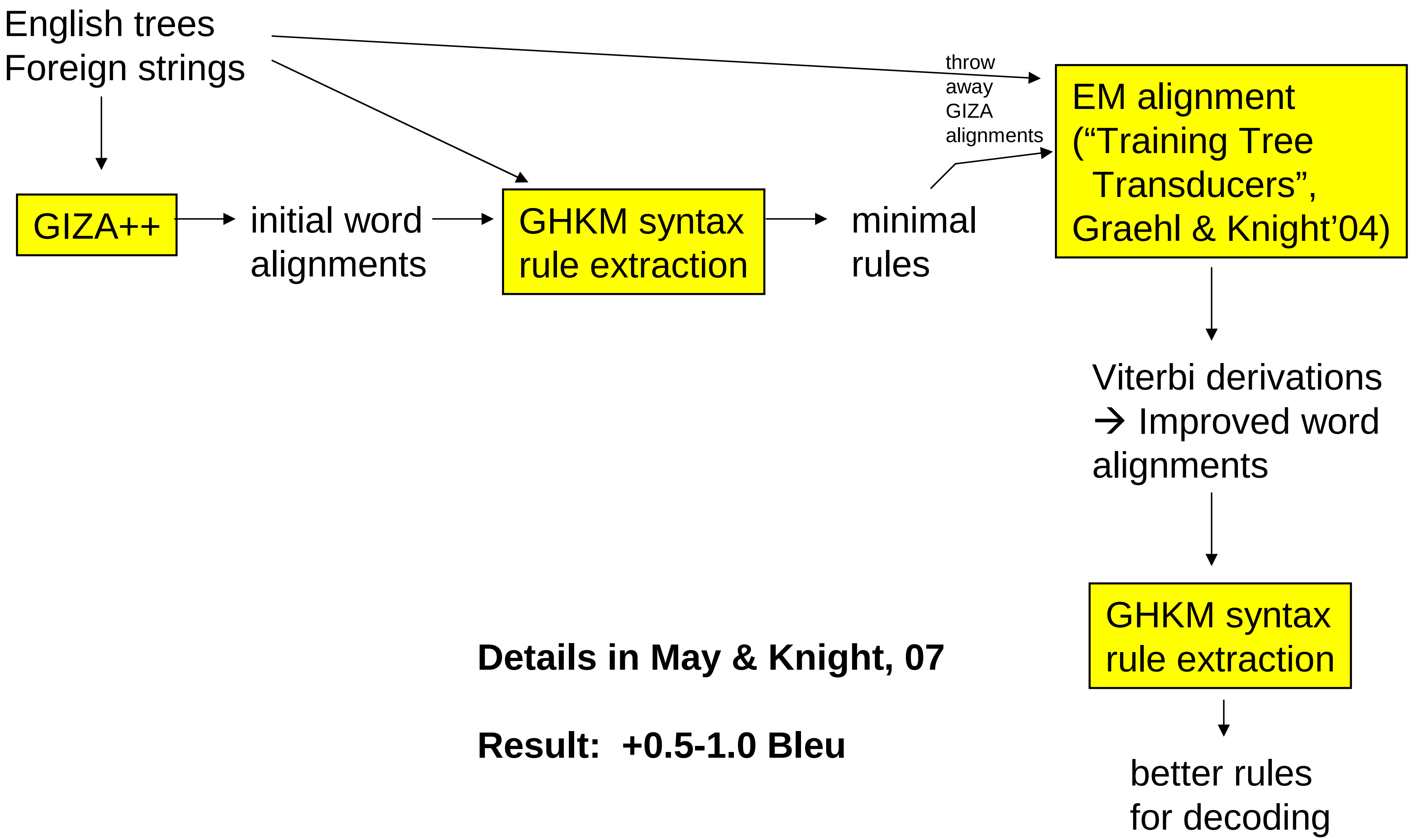
Tree binarized by EM training



Last Topic: Alignment

- GIZA++ string-based alignments
  - are errorful
  - don’t match our syntax-based MT system
- Would like to use the tree-based translation model to align data

Last Topic: Alignment



Details in May & Knight, 07

Result: +0.5-1.0 Bleu

Conclusions

- Phrase-based and syntax-based extraction algorithms have different coverage.
- Syntax-based coverage can be improved:
  - composed rules
  - phrasal rules
  - binarizing English trees with EM
  - re-aligning tree/string pairs with EM
- Improvements lead to better translation accuracy.

the end