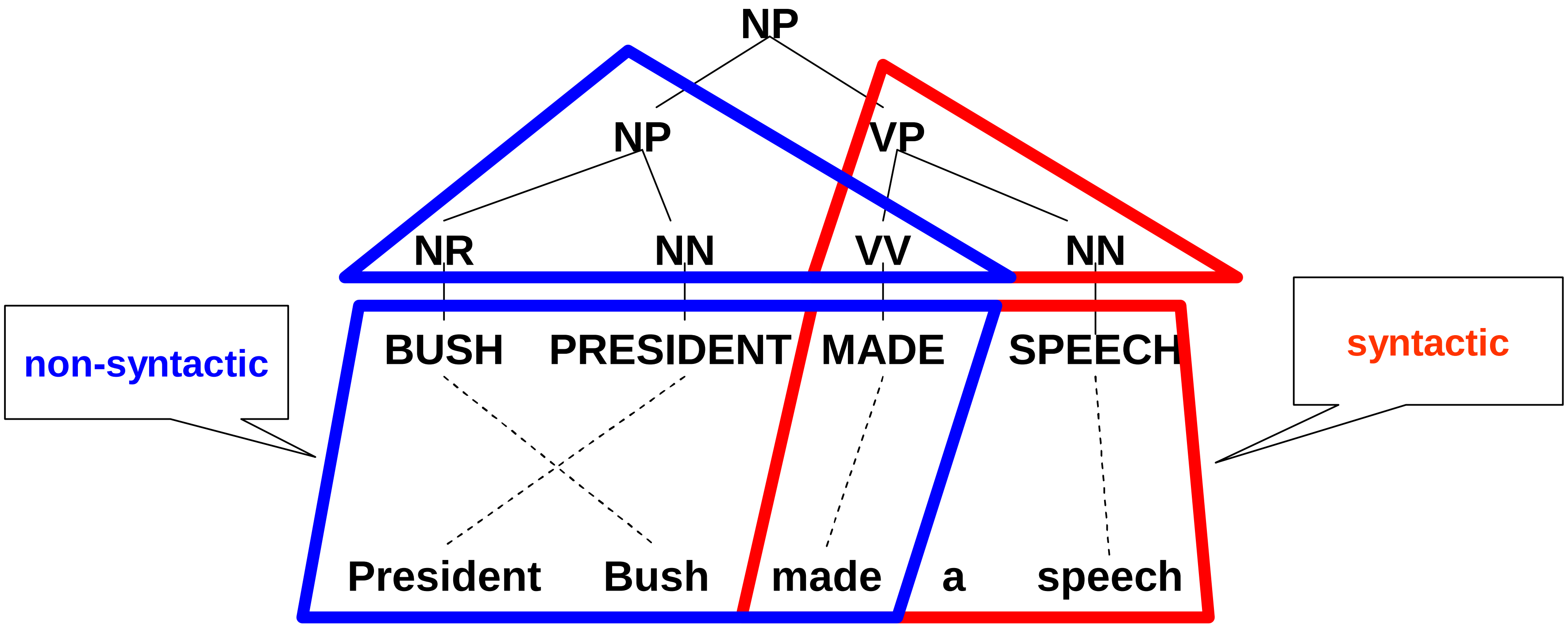
Forest-to-String Statistical Translation Rules

Yang Liu, Qun Liu, and Shouxun Lin
Institute of Computing Technology
Chinese Academy of Sciences

Outline

- Introduction
- Forest-to-String Translation Rules
- Training
- Decoding
- Experiments
- Conclusion

Syntactic and Non-syntactic Bilingual Phrases

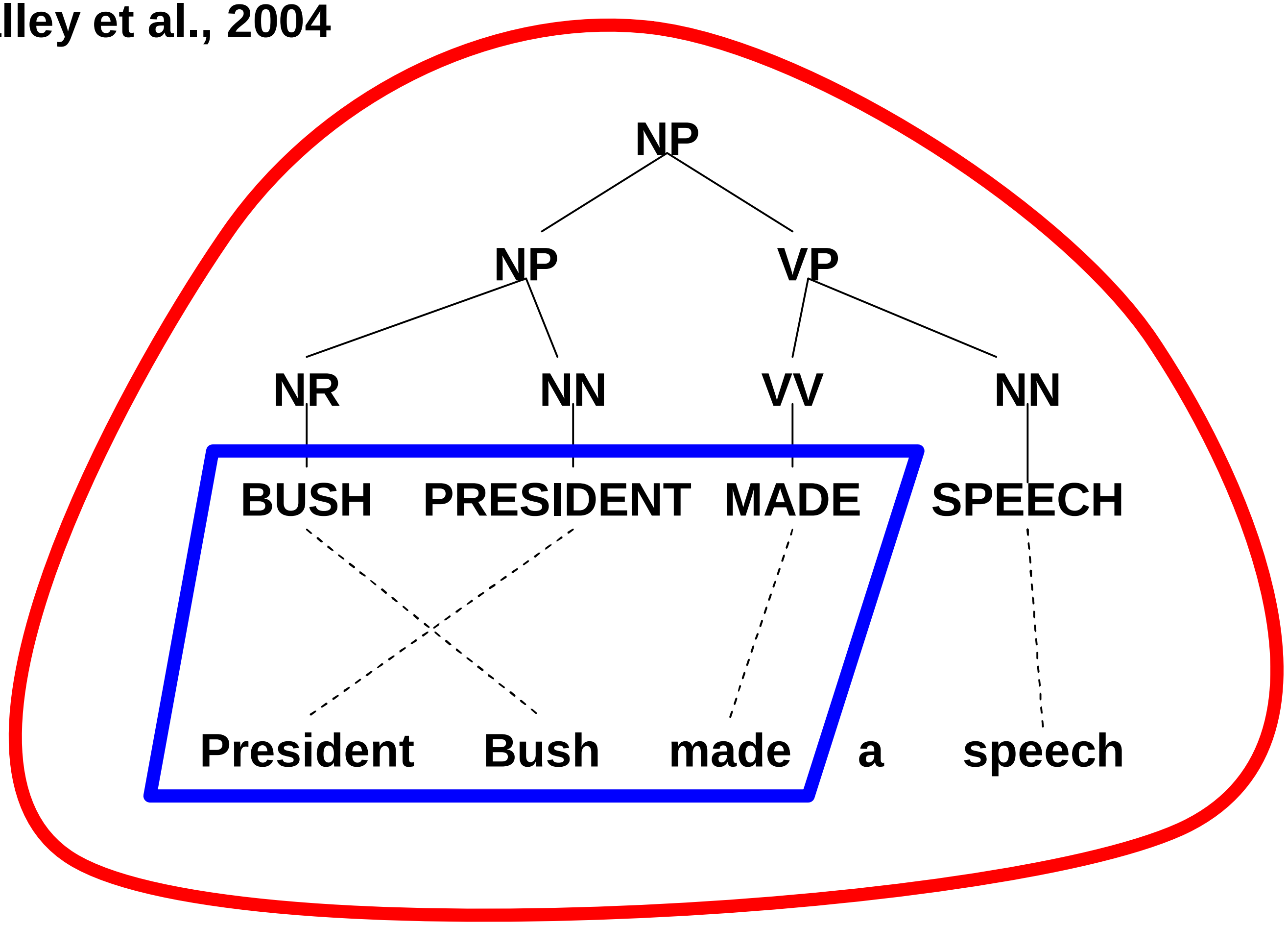


Importance of Non-syntactic Bilingual Phrases

- About 28% of bilingual phrases are non-syntactic on a English-Chinese corpus (***Marcu et al., 2006***).
- Requiring bilingual phrases to be syntactically motivated will lose a good amount of valuable knowledge (***Koehn et al., 2003***).
- Keeping the strengths of phrases while incorporating syntax into statistical translation results in significant improvements (***Chiang, 2005***) .

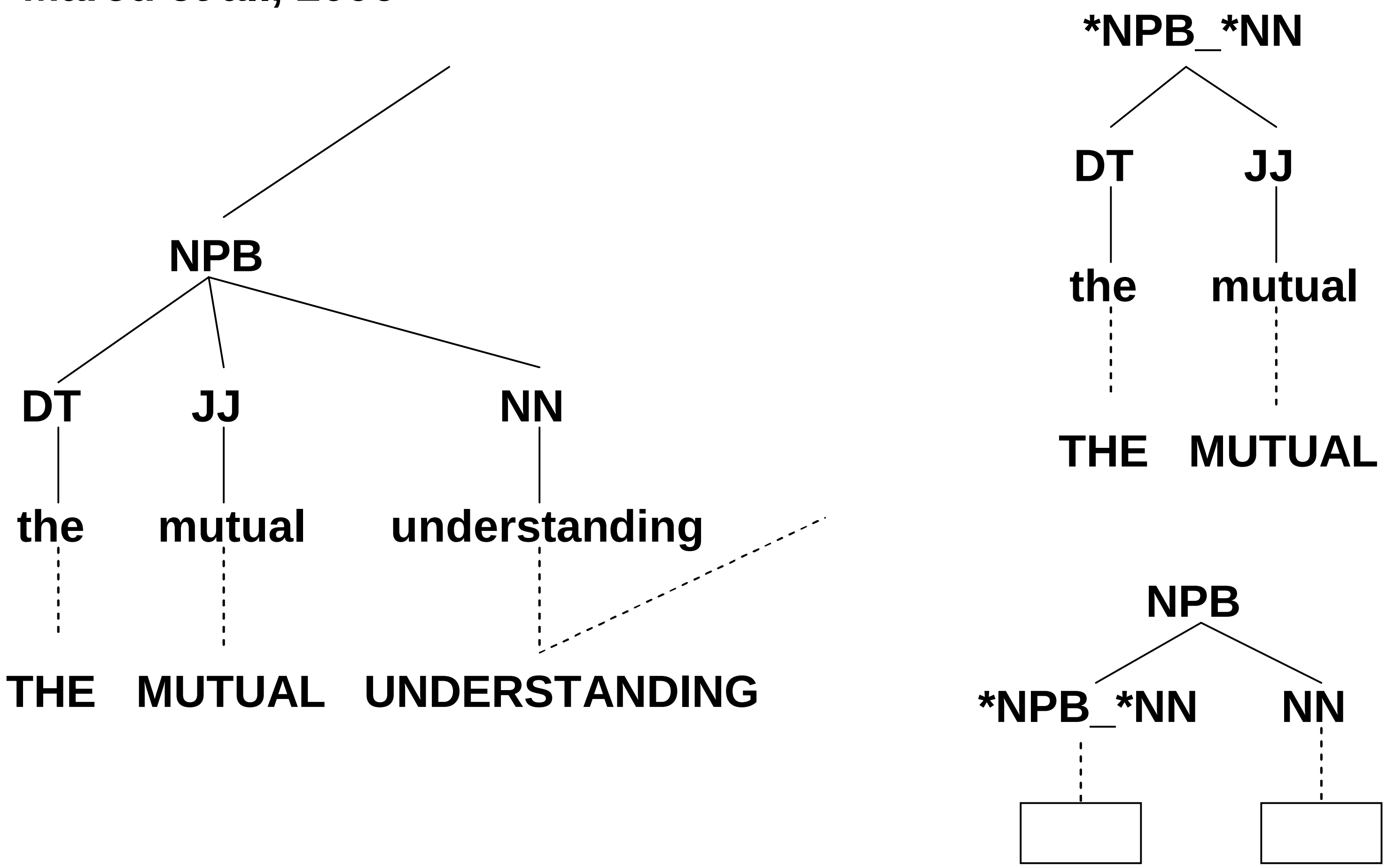
Previous Work

Galley et al., 2004



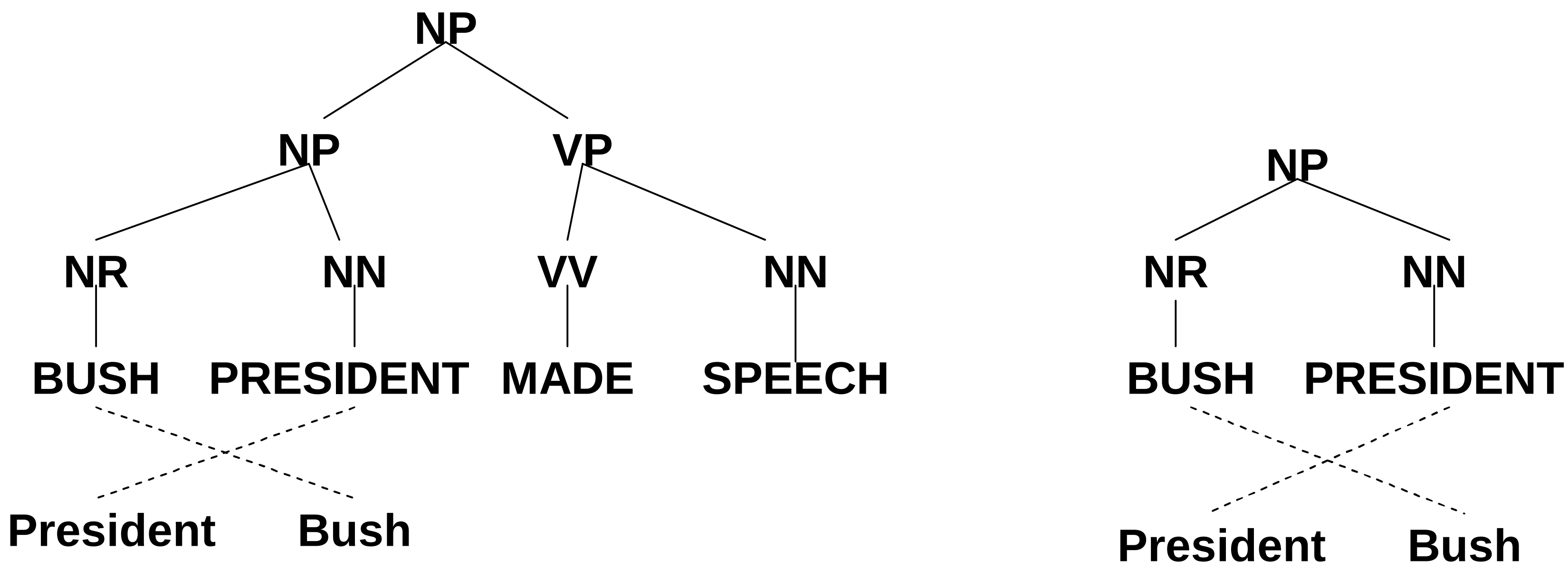
Previous Work

Marcu et al., 2006



Previous Work

Liu et al., 2006



Our Work

INSTITUTE OF COMPUTING TECHNOLOGY

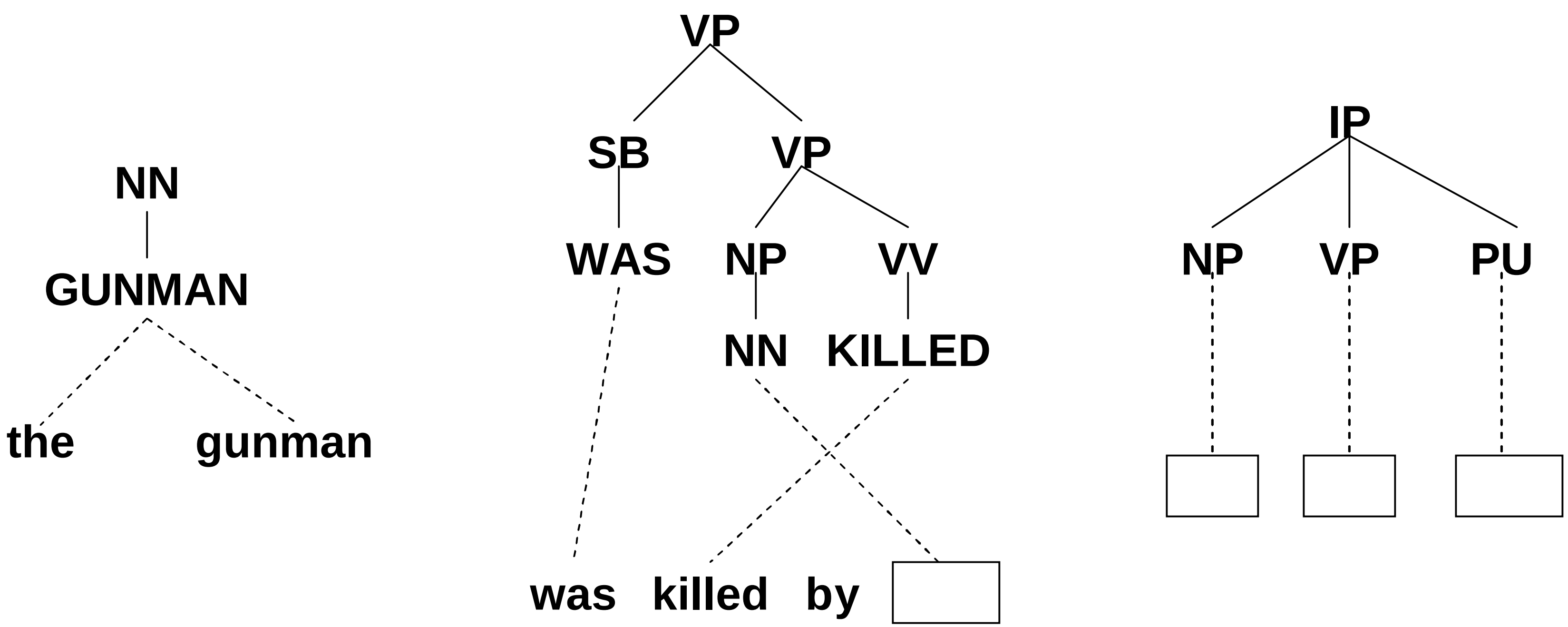
- We augment the tree-to-string translation model with
  - ***forest-to-string rules*** that capture non-syntactic phrase pairs
  - ***auxiliary rules*** that help integrate forest-to-string rules into the tree-to-string model

Outline

中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

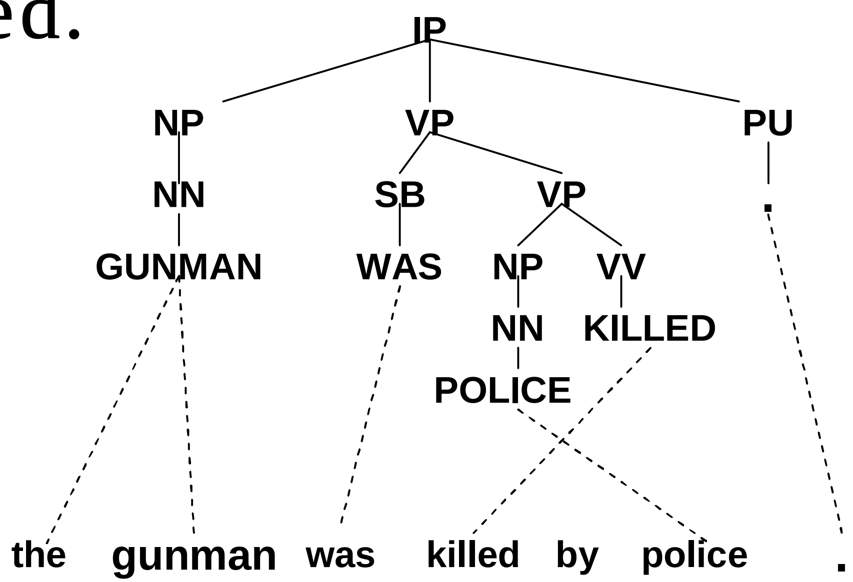
- *Introduction*
- Forest-to-String Translation Rules
- Training
- Decoding
- Experiments
- Conclusion

Tree-to-String Rules



Derivation

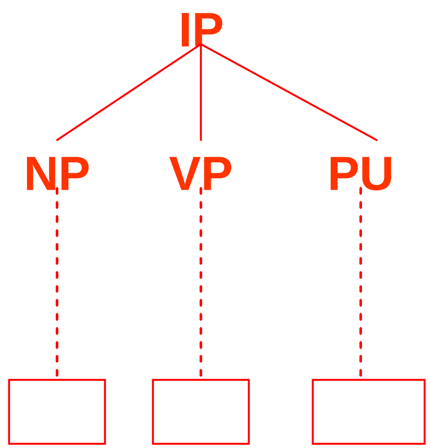
- A ***derivation*** is a left-most composition of translation rules that explains how a source parse tree, a target sentence, and the word alignment between them are synchronously generated.

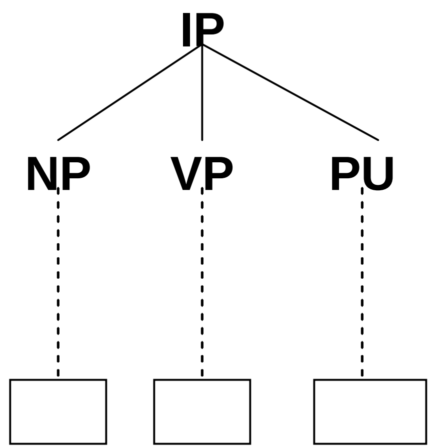


A Derivation Composed of TRs

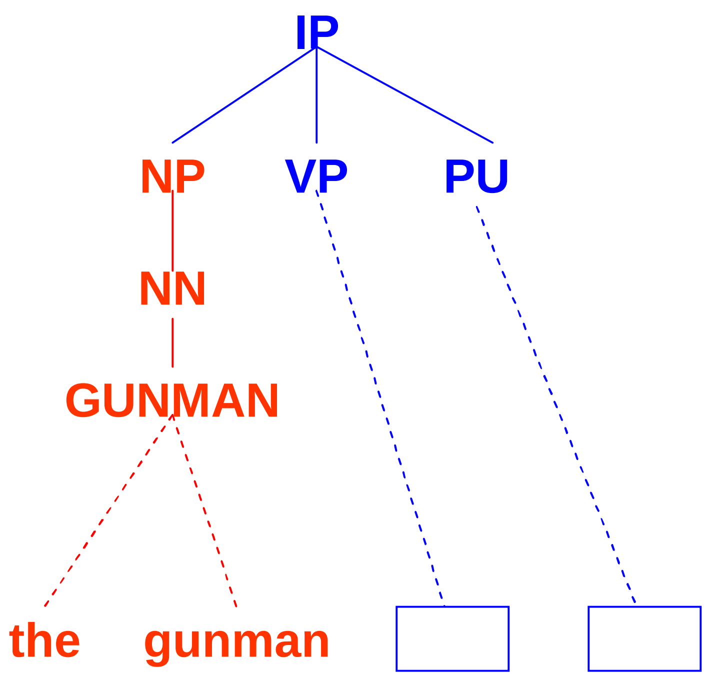
A Derivation Composed of TRs

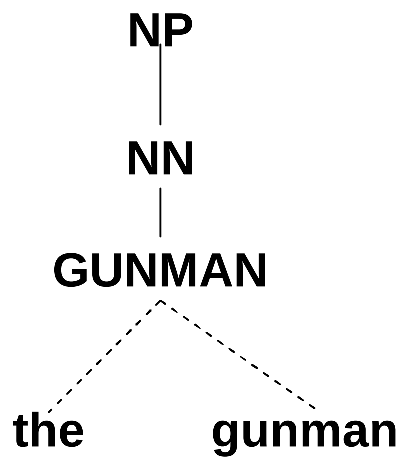
INSTITUTE OF COMPUTING TECHNOLOGY



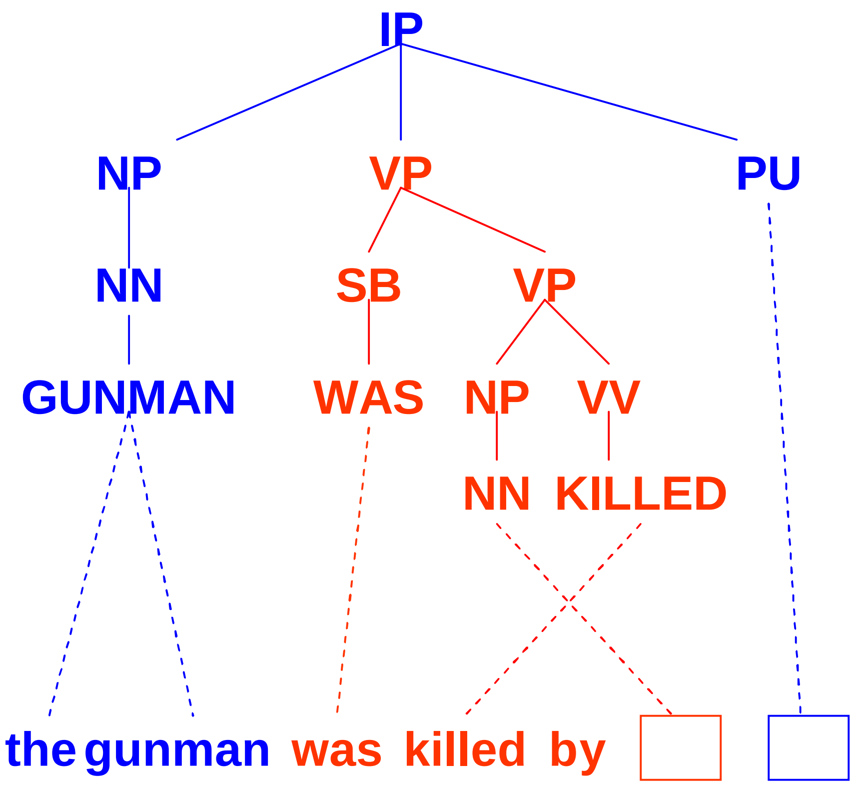


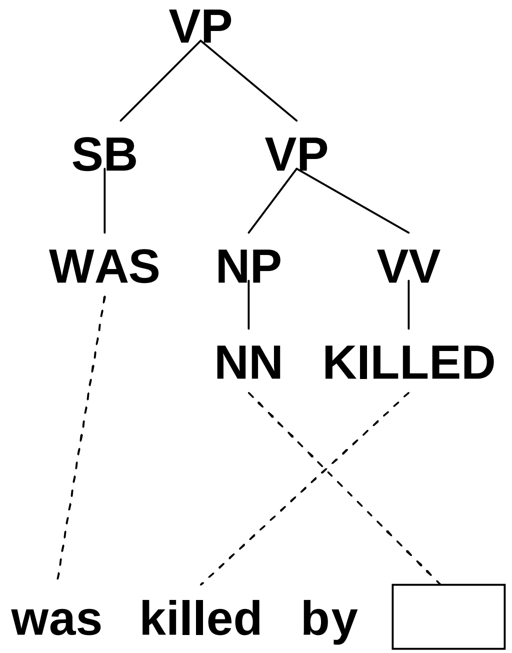
A Derivation Composed of TRs



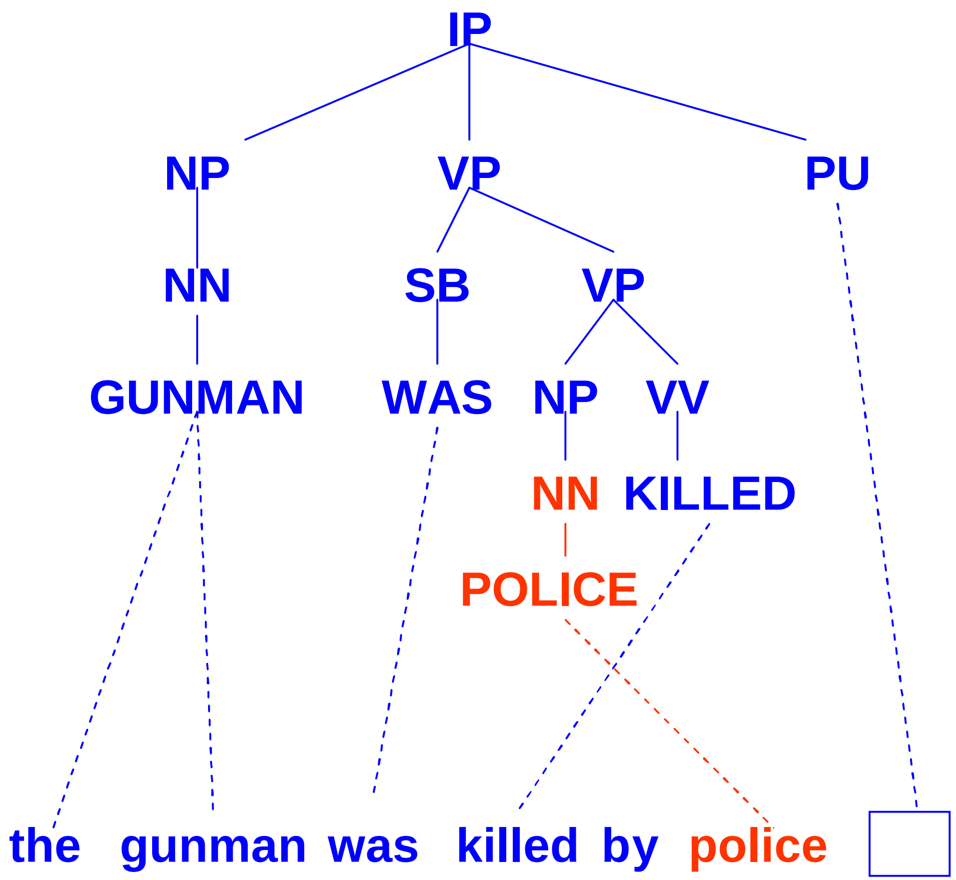


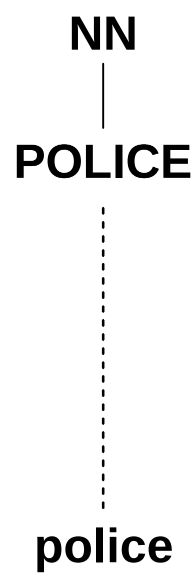
A Derivation Composed of TRs



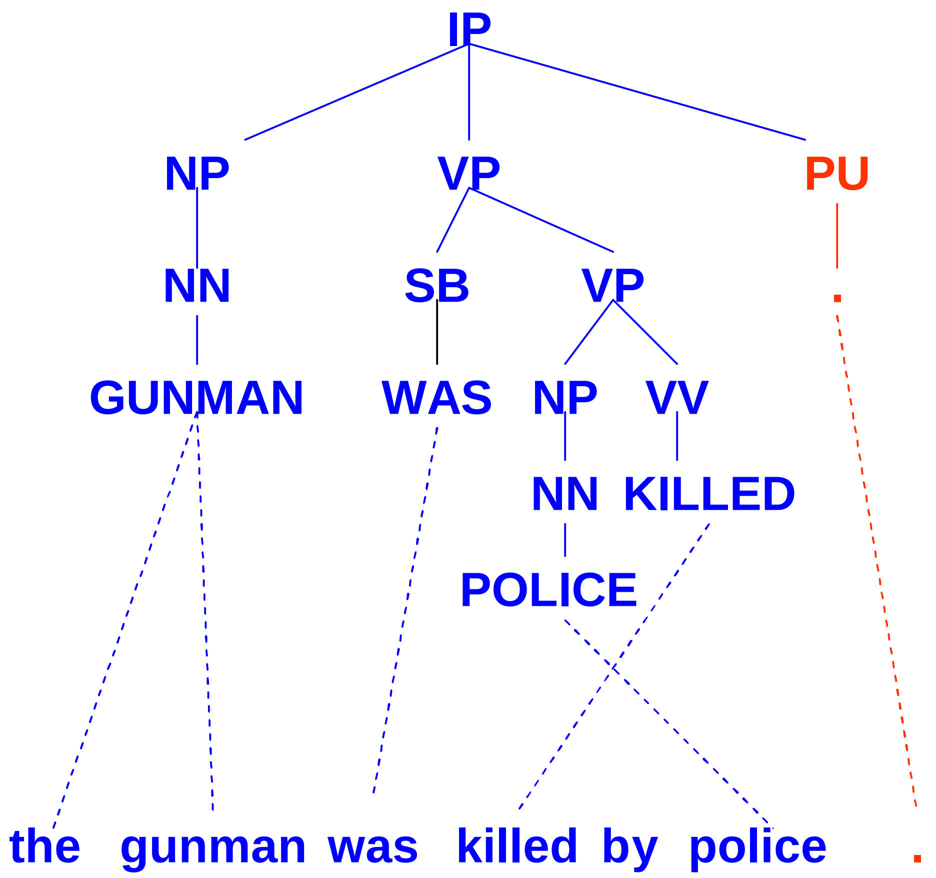


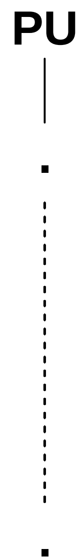
A Derivation Composed of TRs



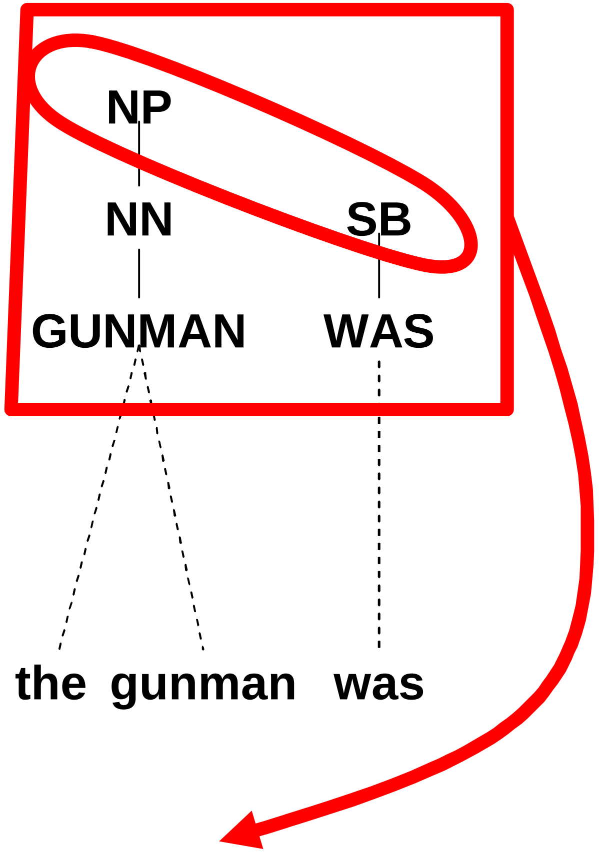


A Derivation Composed of TRs

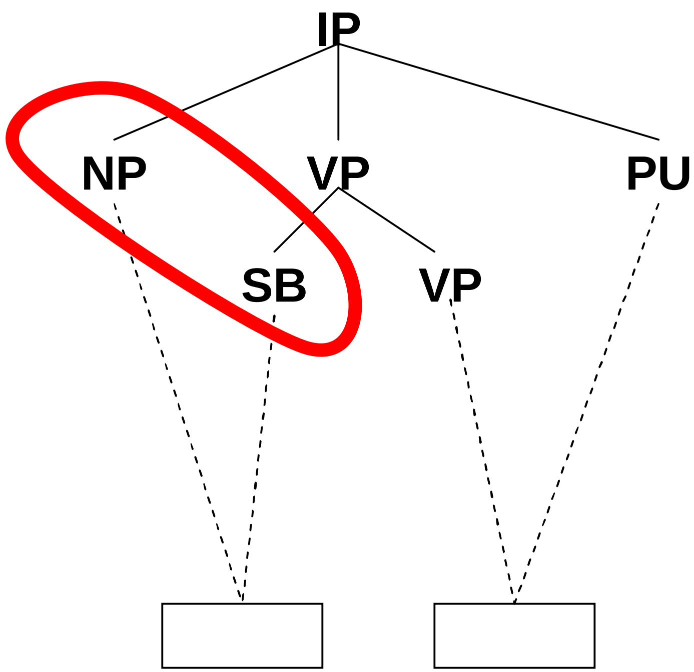




Forest-to-String and Auxiliary Rules



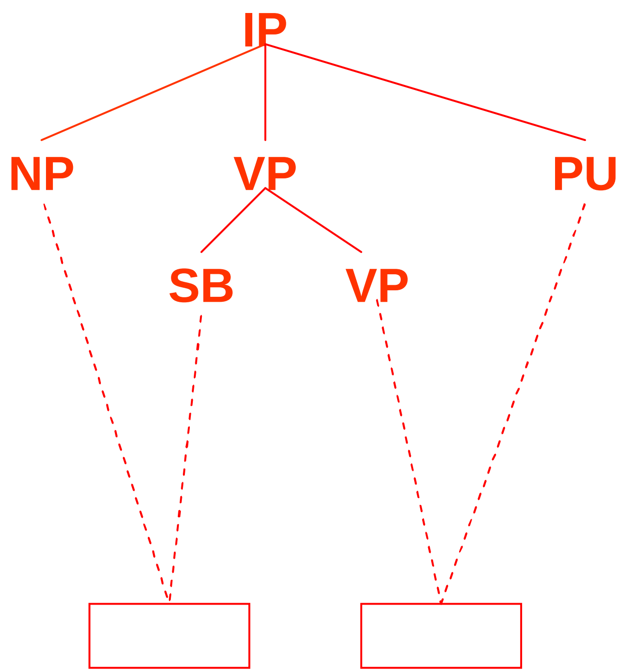
forest = tree sequence !

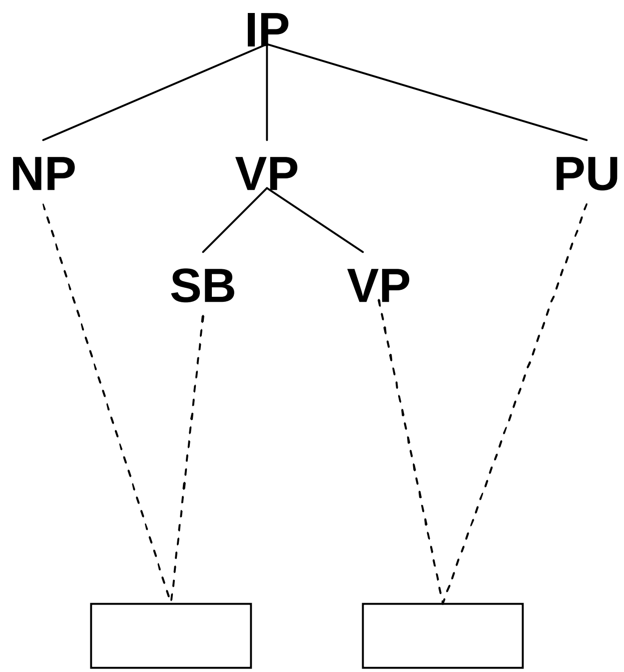


care about only root sequence while incorporating forest rules

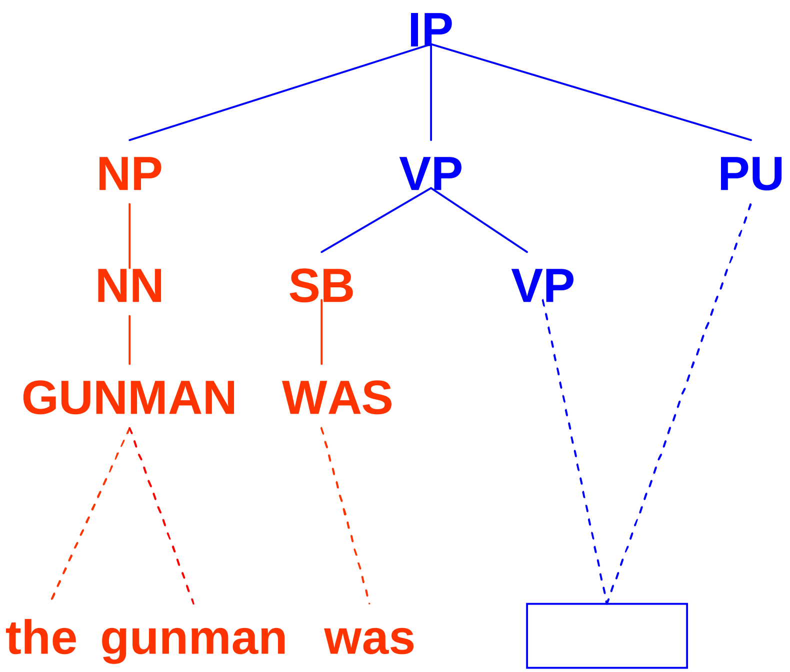
A Derivation Composed of TRs, FRs, and ARs

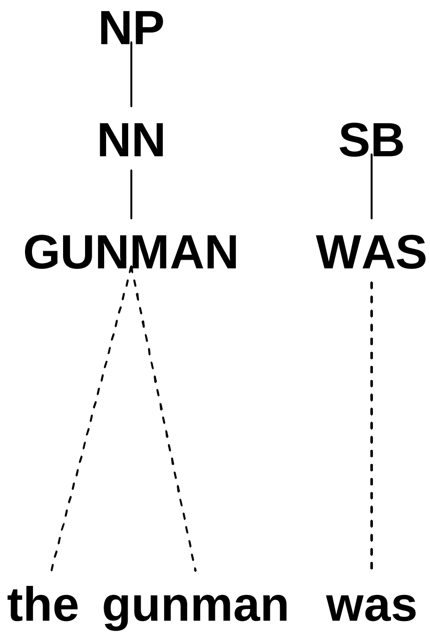
A Derivation Composed of TRs, FRs, and ARs



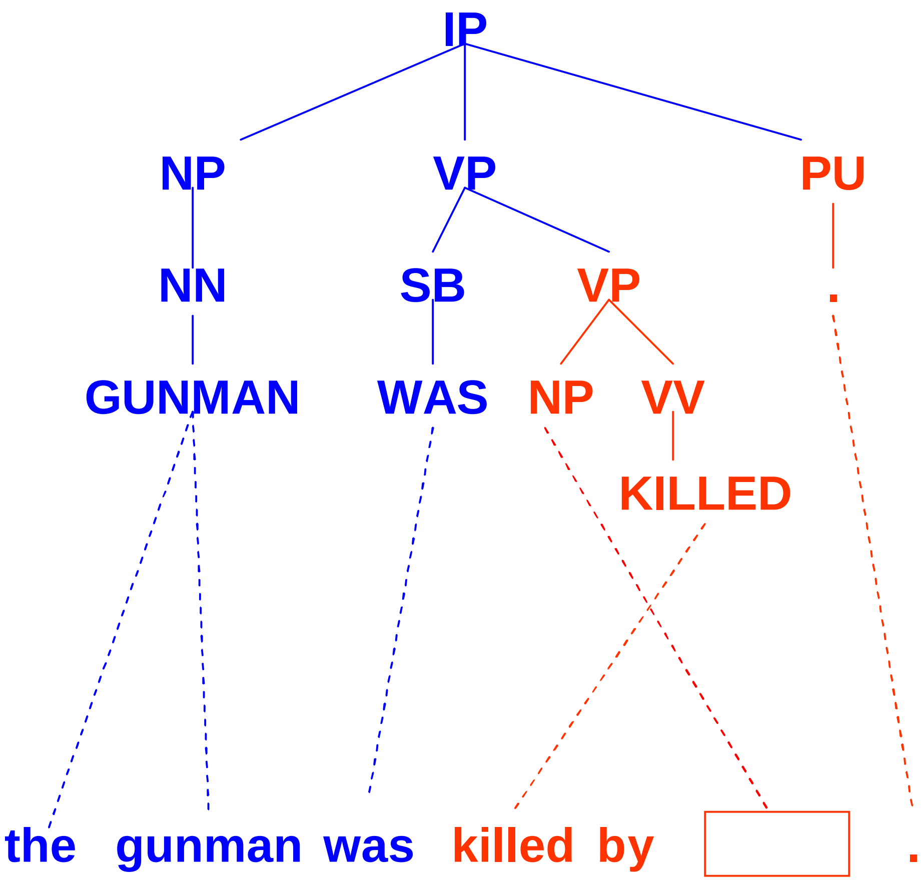


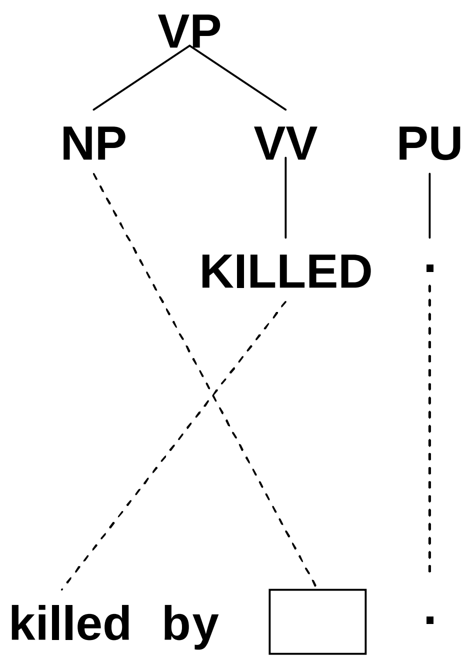
A Derivation Composed of TRs, FRs, and ARs



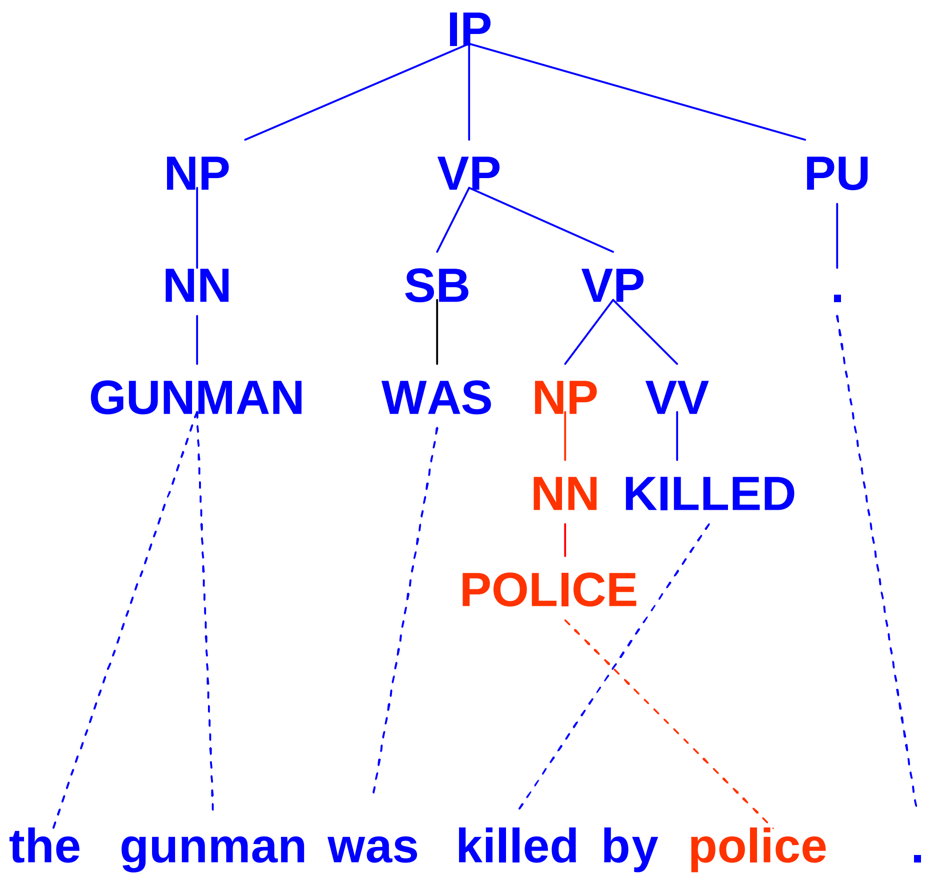


A Derivation Composed of TRs, FRs, and ARs





A Derivation Composed of TRs, FRs, and ARs



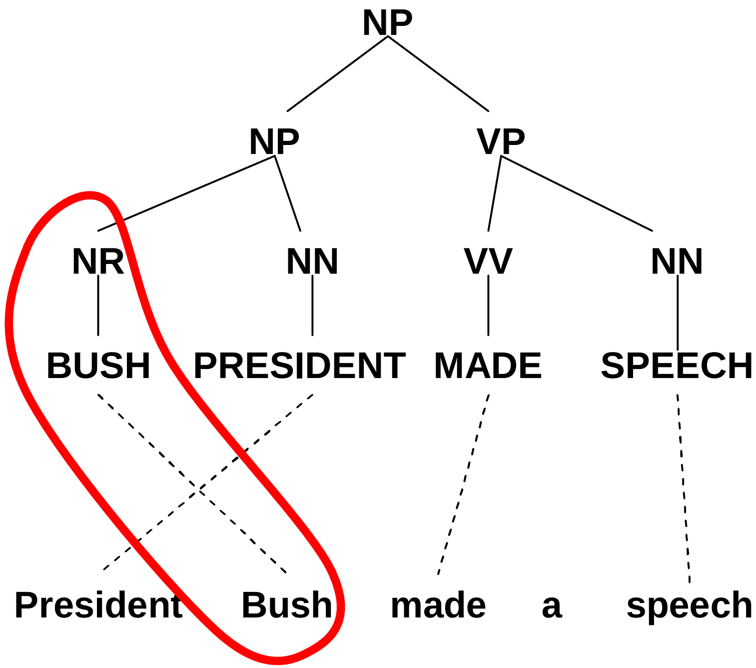
Outline

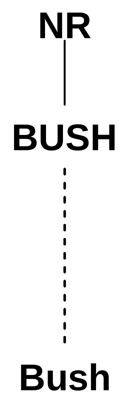
- *Introduction*
- *Forest-to-String Translation Rules*
- Training
- Decoding
- Experiments
- Conclusion

Training

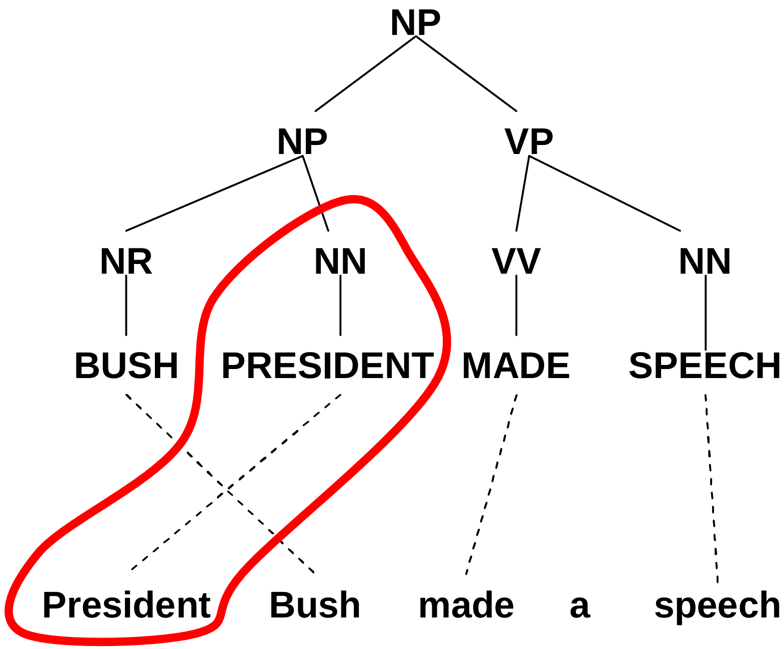
- Extract both ***tree-to-string*** and ***forest-to-string*** rules from word-aligned, source-side parsed bilingual corpus
- Bottom-up strategy
- ***Auxiliary*** rules are NOT learnt from real-world data

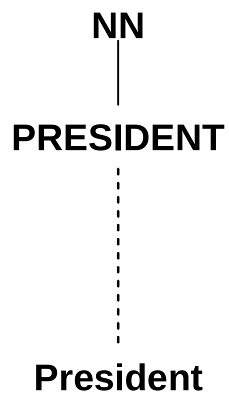
An Example


NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech

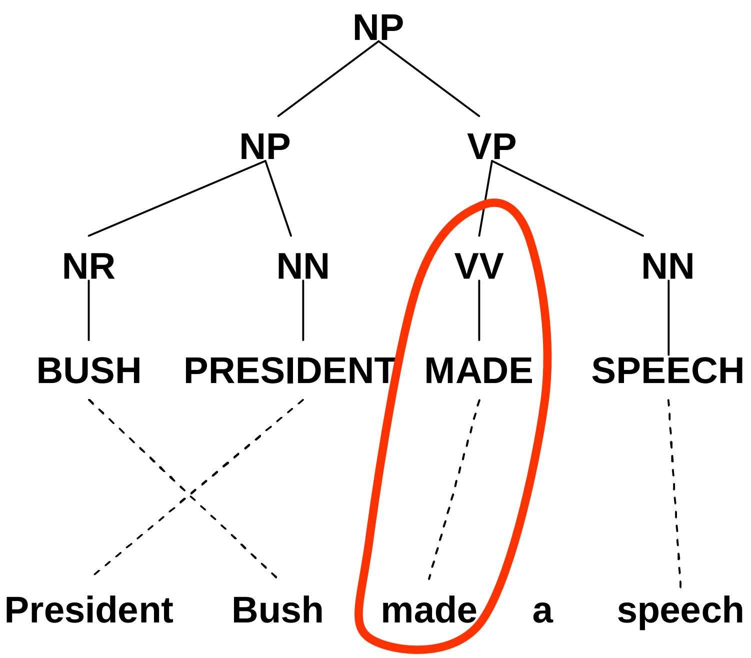

NR
BUSH
Bush

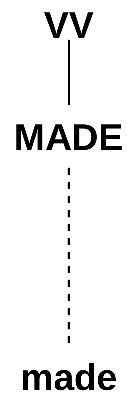
An Example


NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech

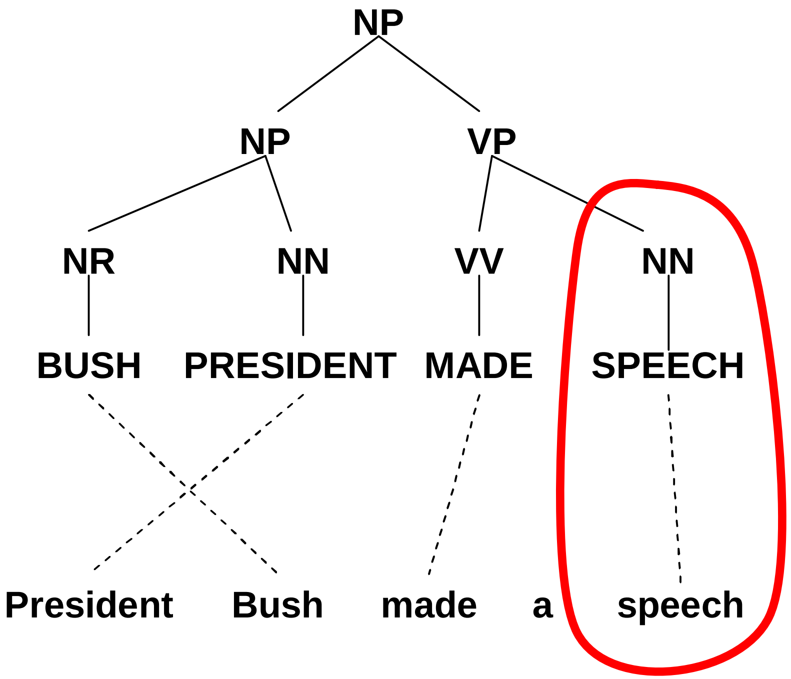

NN
PRESIDENT
President

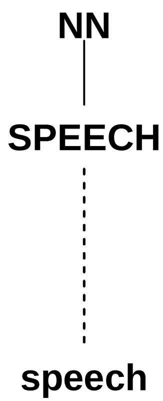
An Example


NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech

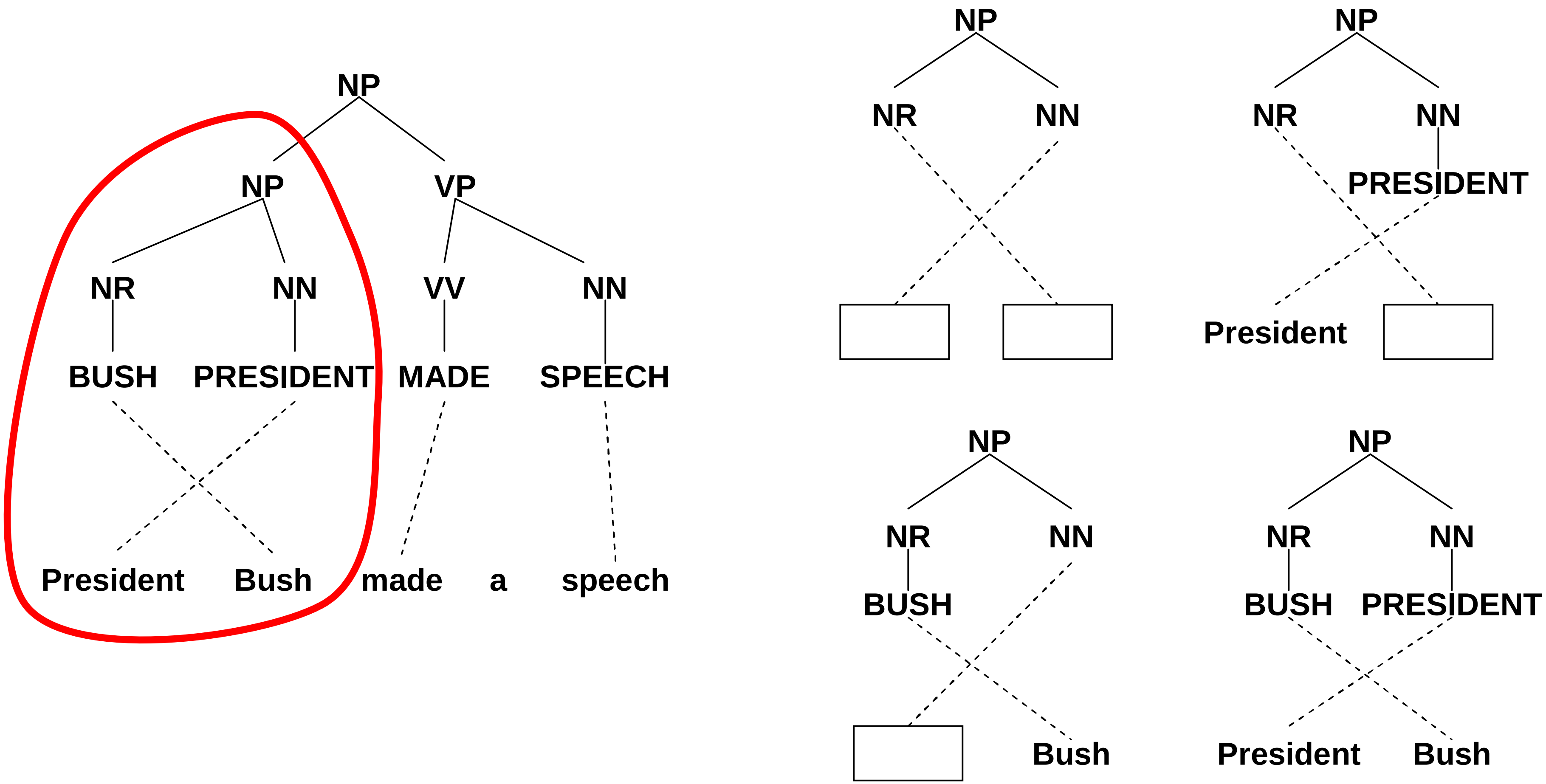

VV
MADE
made

An Example


NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech

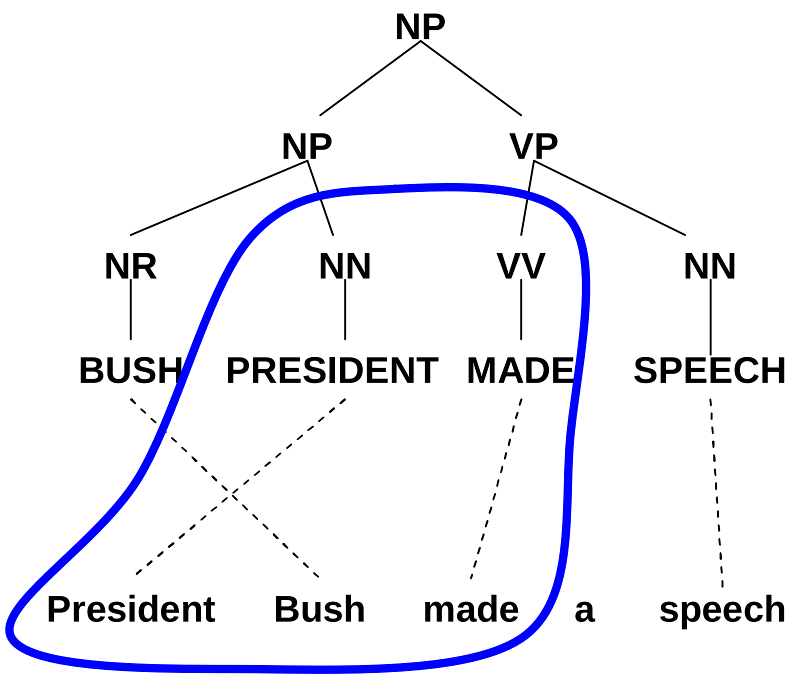

NN
SPEECH
speech

An Example

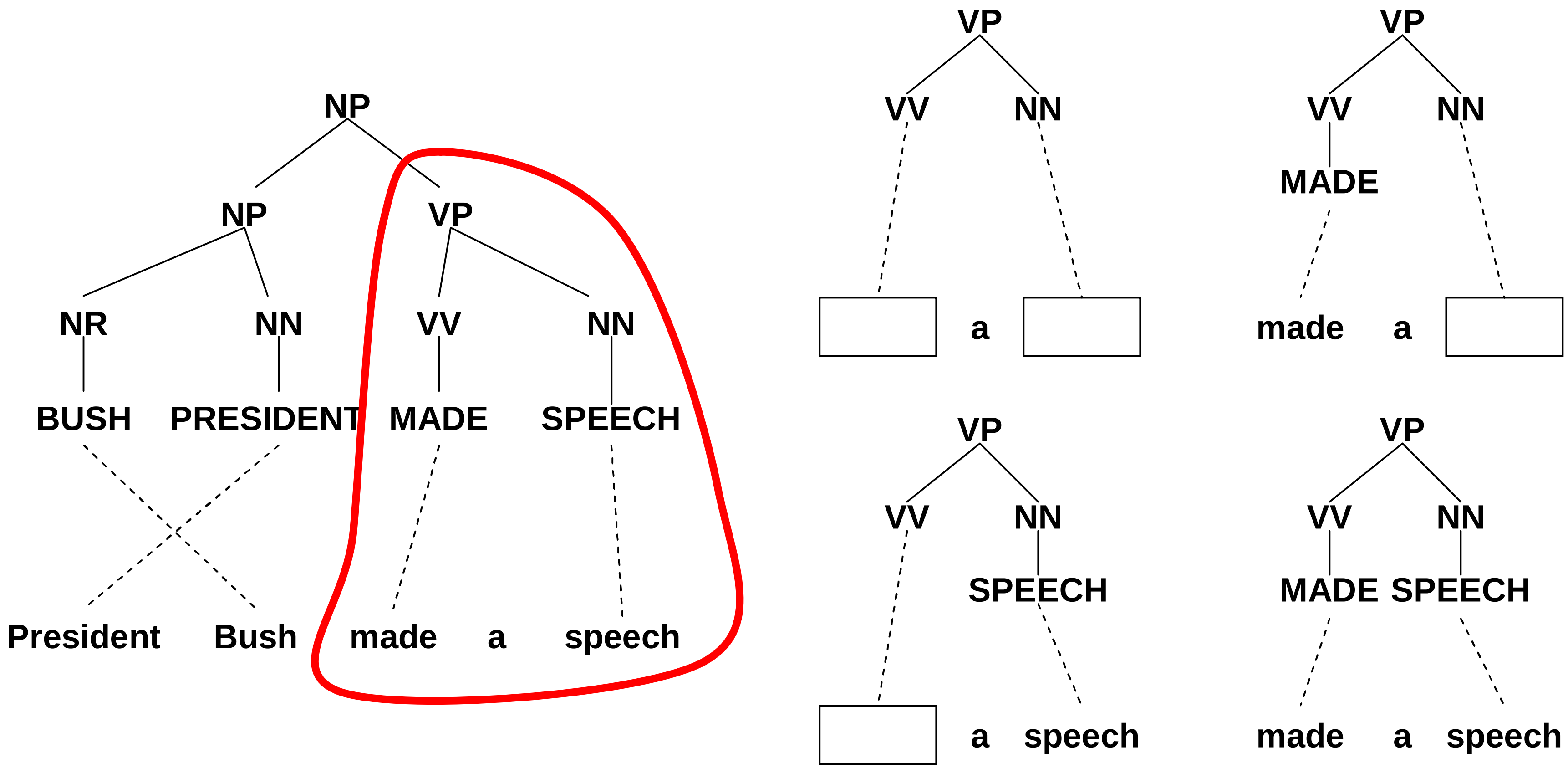


NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech
NP
NR
NN
NP
NR
NN
PRESIDENT
President
NP
NR
NN
BUSH
Bush
NP
NR
NN
BUSH
PRESIDENT
President
Bush

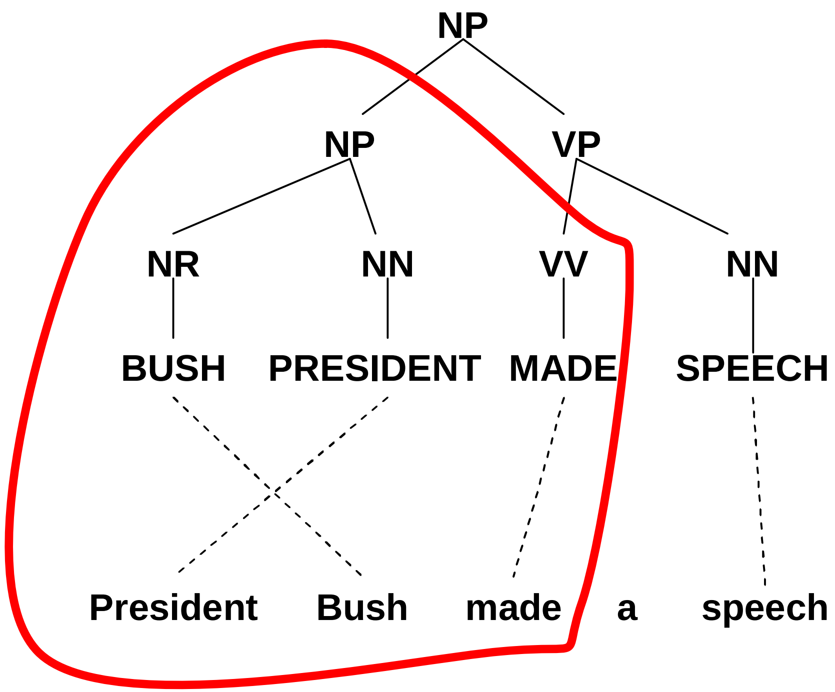
An Example

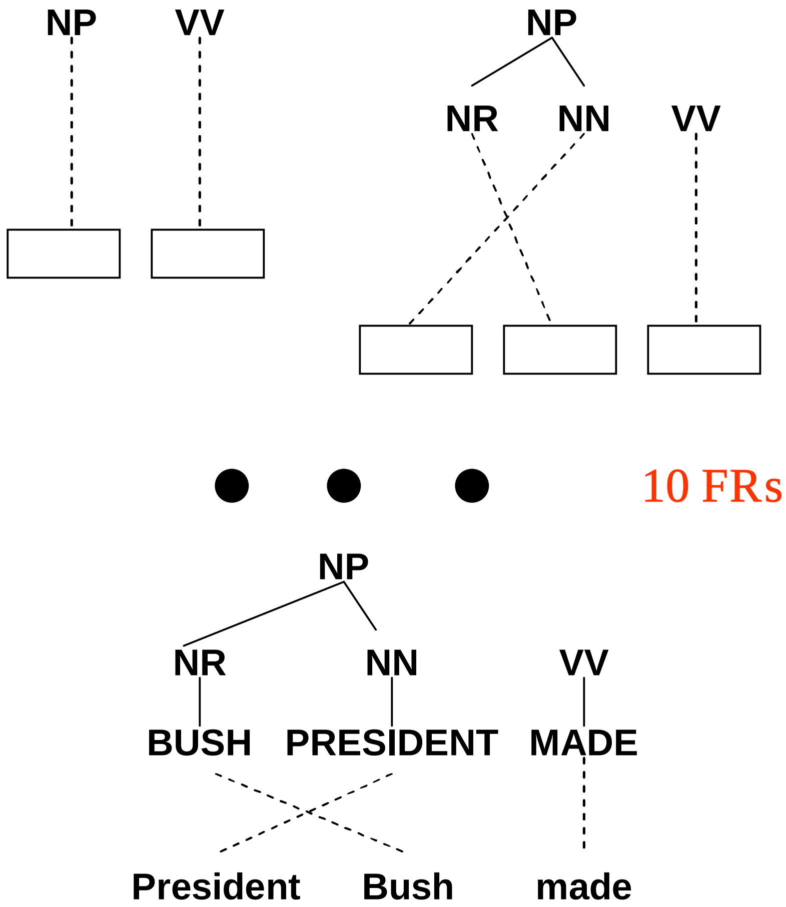

NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech

An Example

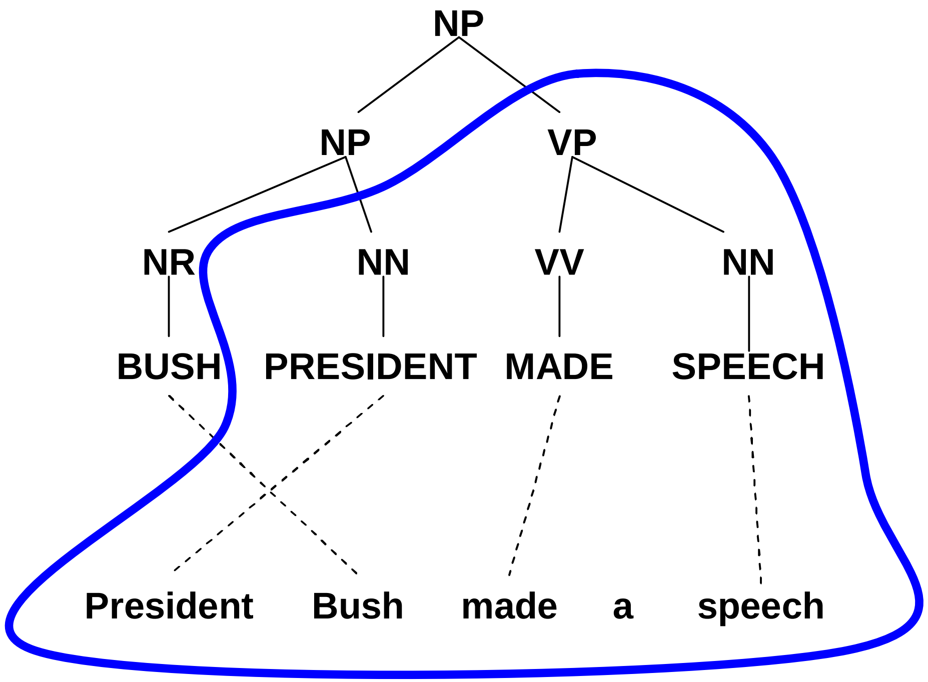

NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech
VP
VV
NN
a
VP
VV
NN
MADE
made
a
VP
VV
NN
SPEECH
a
speech
VP
VV
NN
MADE
SPEECH
made
a
speech

An Example

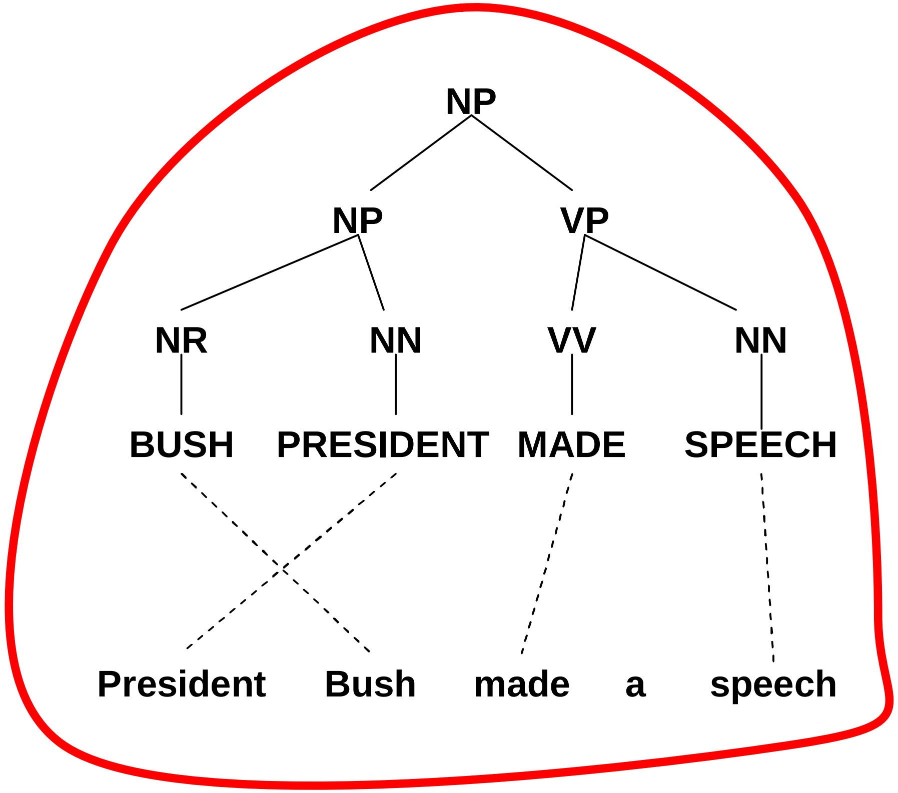

NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech

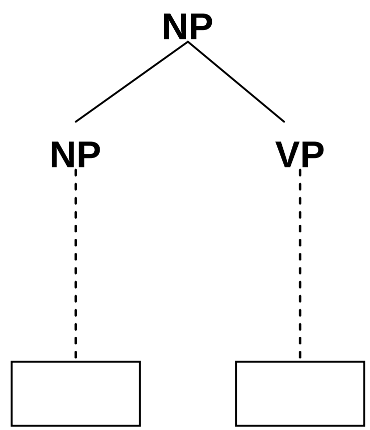

NP
VV
NP
NR
NN
VV
10 FRs
NP
NR
NN
VV
BUSH
PRESIDENT
MADE
President
Bush
made

An Example


NP
NP
VP
NR
NN
VV
NN
BUSH
PRESIDENT
MADE
SPEECH
President
Bush
made
a
speech

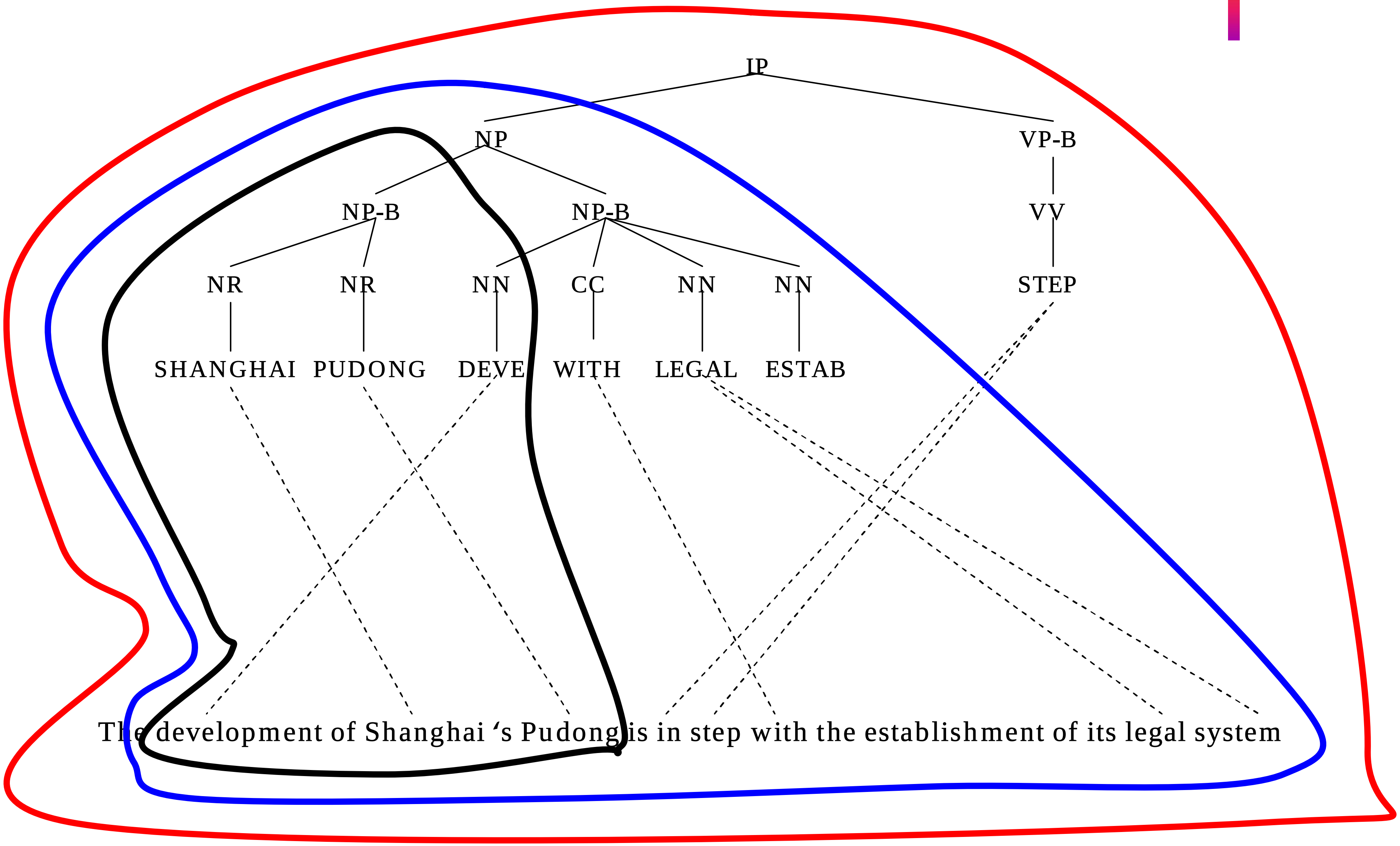
An Example





max_height = 2

Why We Don't Extract Auxiliary Rules ?



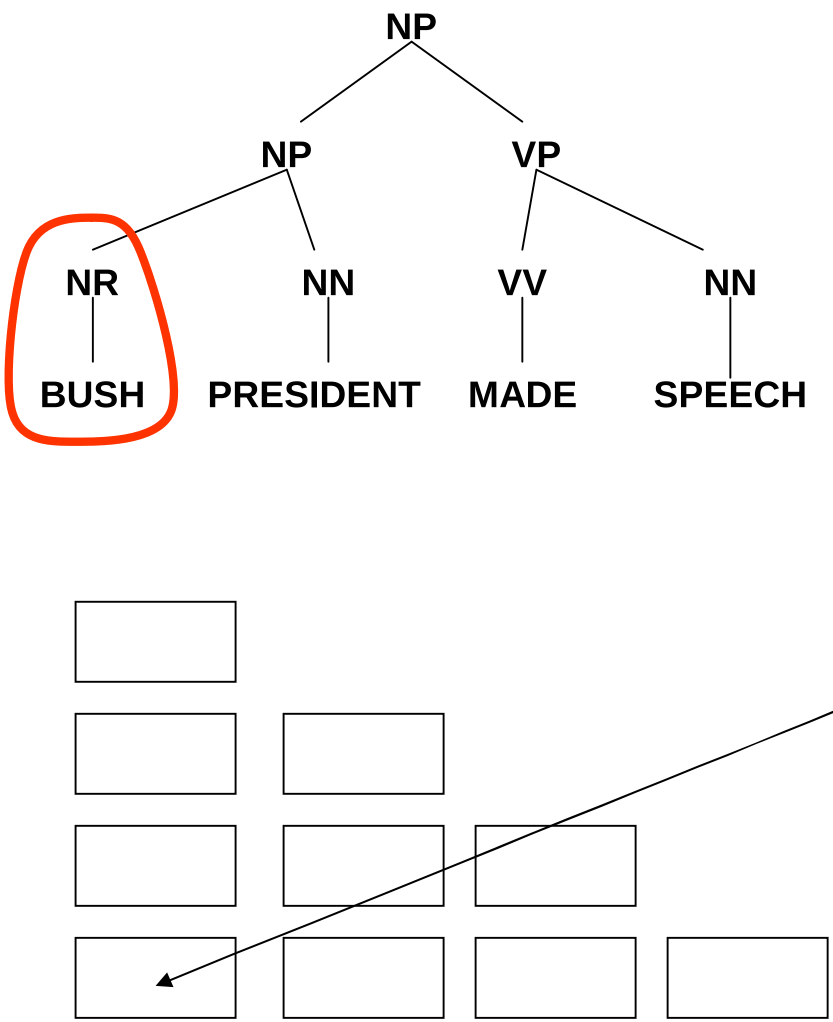
Outline

- *Introduction*
- *Forest-to-String Translation Rules*
- *Training*
- Decoding
- Experiments
- Conclusion

Decoding

- Input: a source parse tree
  Output: a target sentence
- Bottom-up strategy
- Build auxiliary rules while decoding
- Compute subcell divisions for building auxiliary rules

An Example



Rule

NR
BUSH
Bush

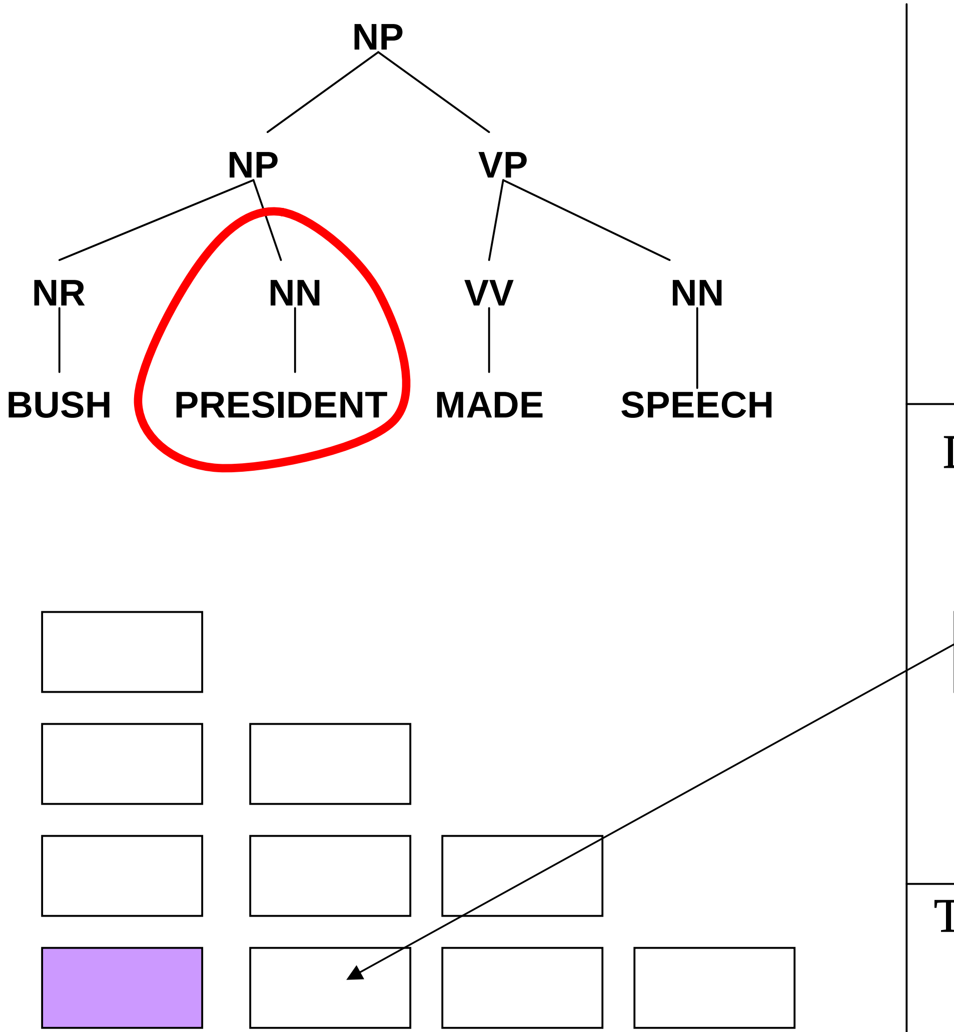
Derivation

(NR BUSH) ||| Bush ||| 1:1

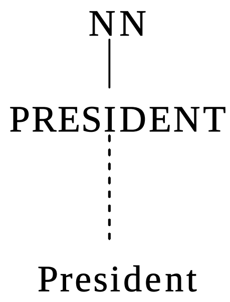
Translation

Bush

An Example



Rule



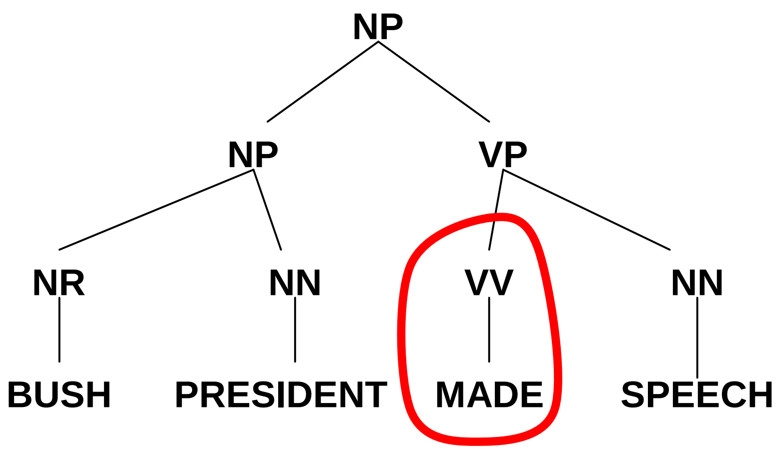
Derivation

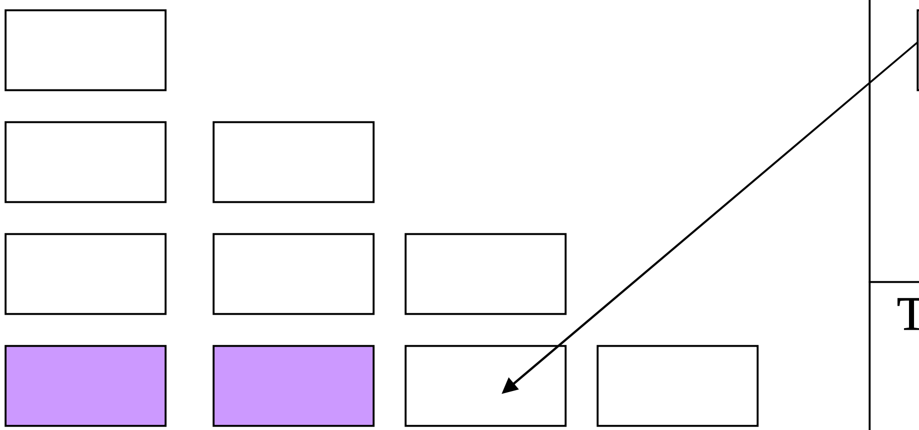
(NN PRESIDENT) ||| President ||| 1:1

Translation

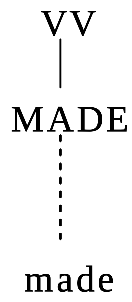
President

An Example





Rule



Derivation

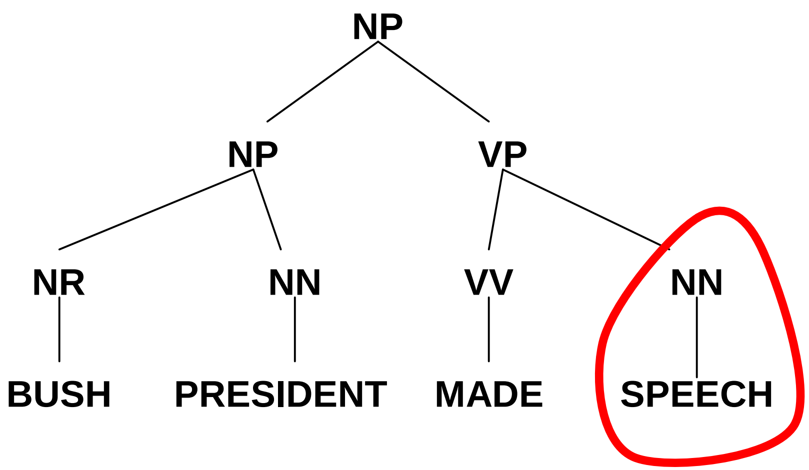
(VV MADE) ||| made ||| 1:1

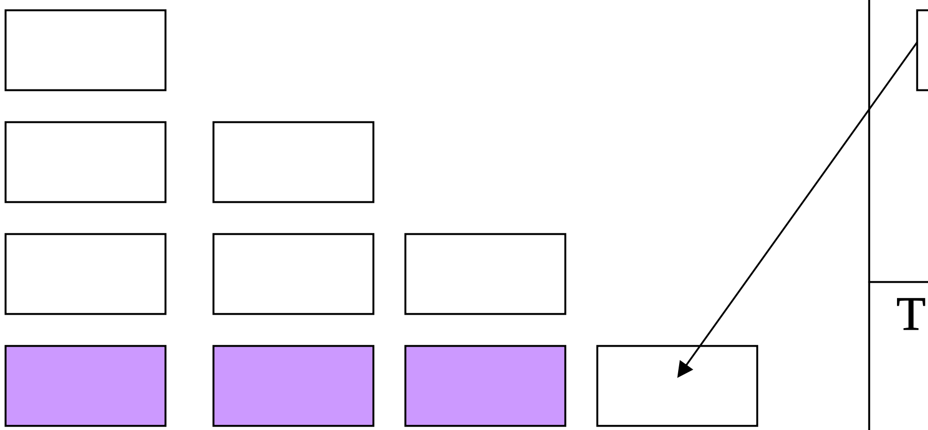
Translation

made

INSTITUTE OF COMPUTING TECHNOLOGY

An Example





Rule

NN

SPEECH

speech

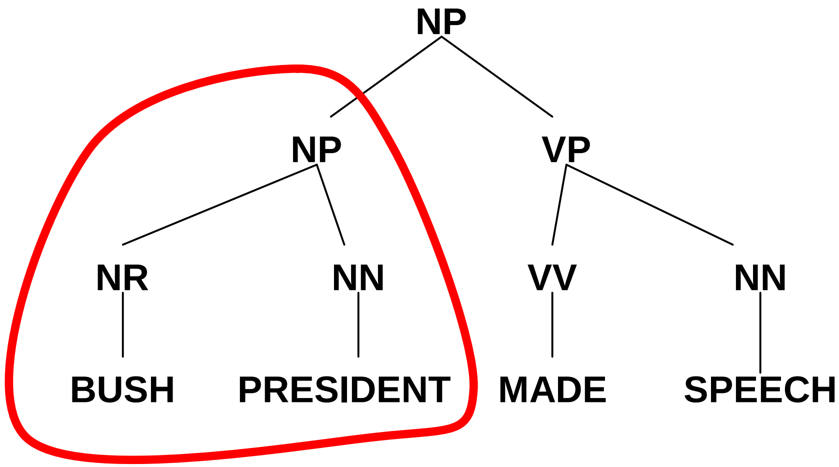
Derivation

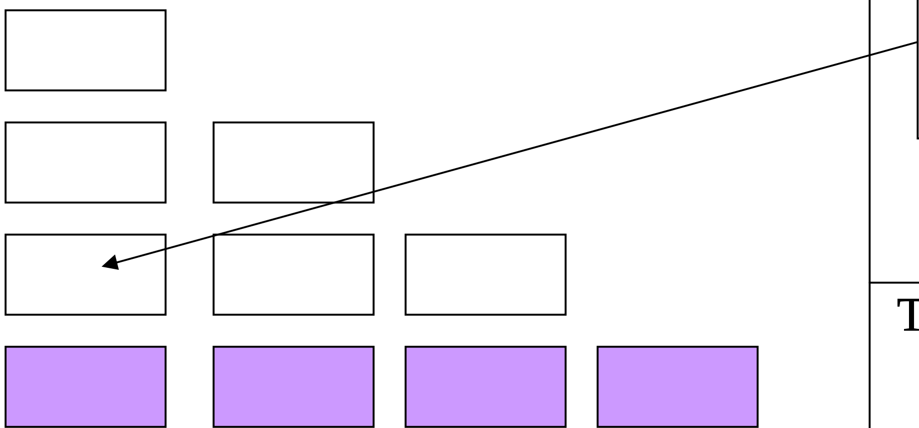
(NN SPEECH) ||| speech ||| 1:1

Translation

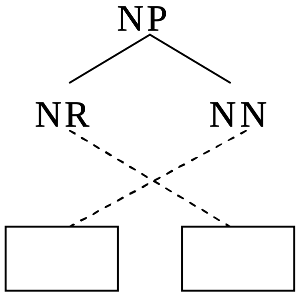
speech

An Example





Rule



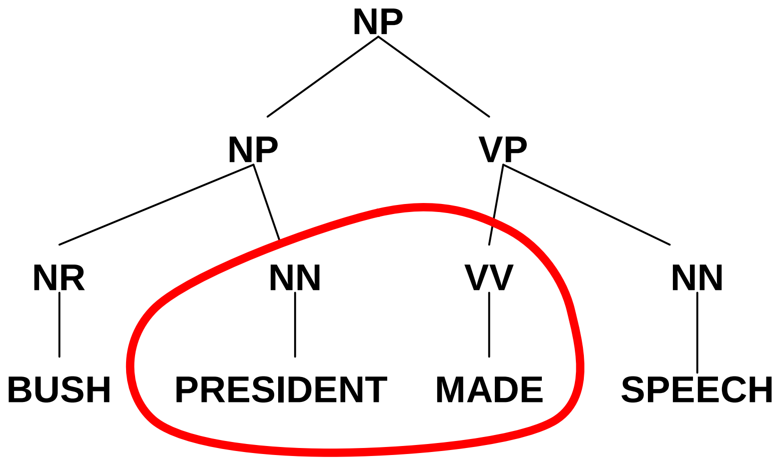
Derivation

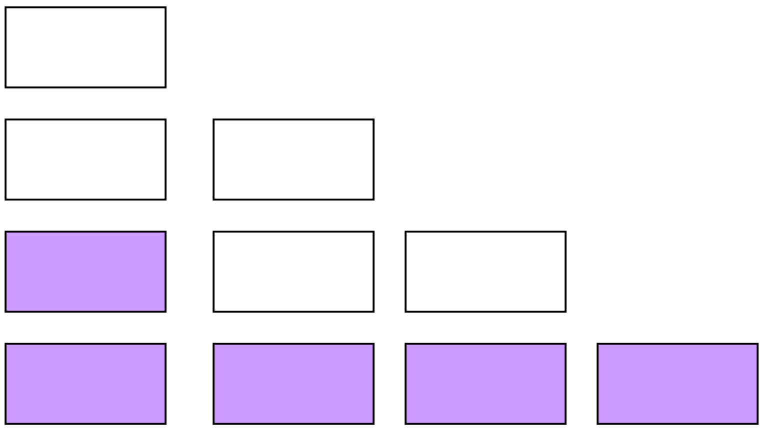
(NP (NR) (NN)) ||| $\boldsymbol{X_1}$ $\boldsymbol{X_2}$ ||| 1:2 2:1
(NR BUSH) ||| Bush ||| 1:1
(NN PRESIDENT) ||| President ||| 1:1

Translation

President Bush

An Example



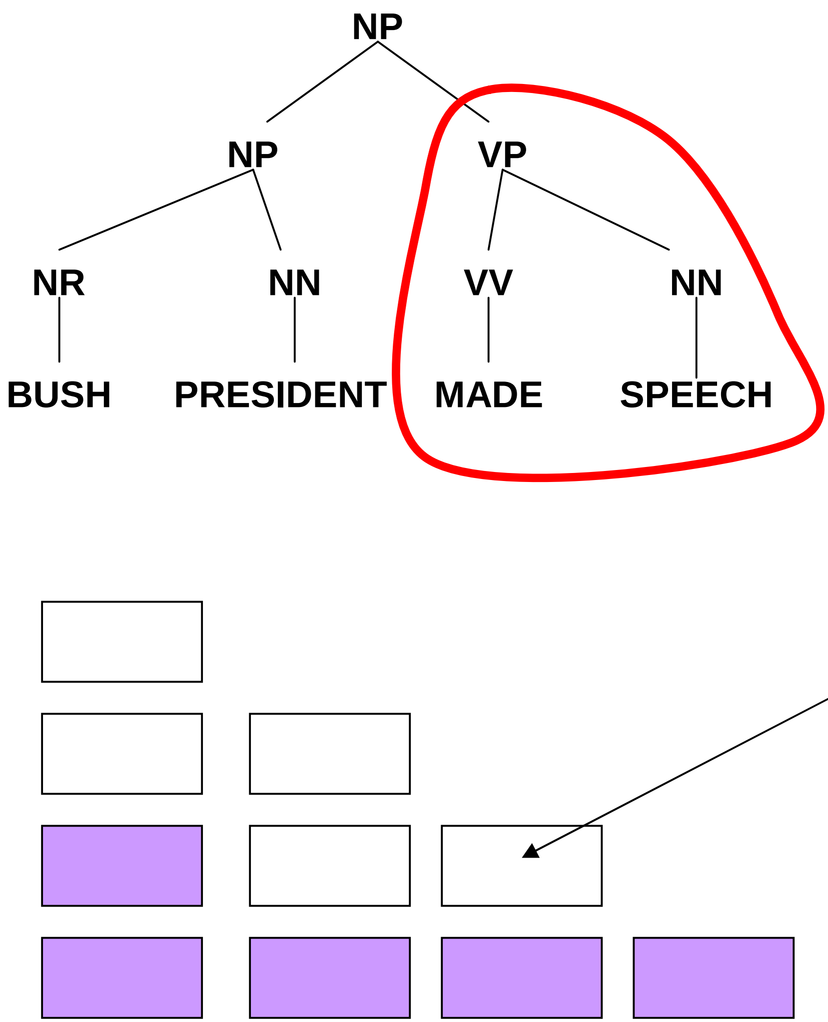


Rule

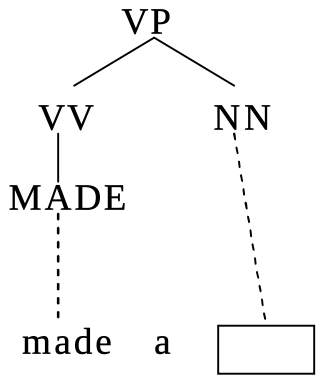
Derivation

Translation

An Example



Rule



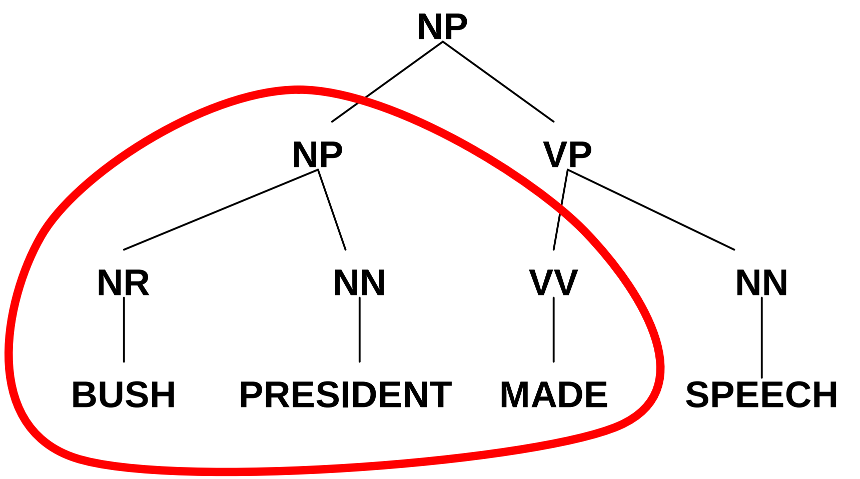
Derivation

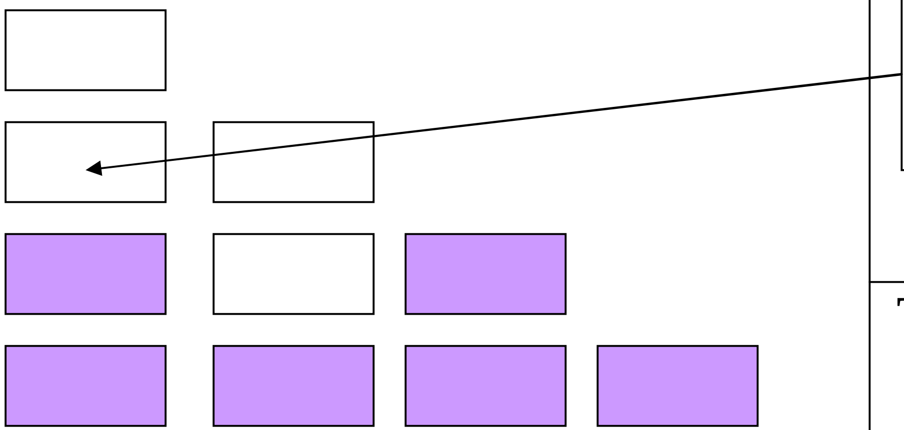
(VP (VV MADE) (NN)) ||| made a ***X*** ||| 1:1 2:3
(NN SPEECH) ||| speech ||| 1:1

Translation

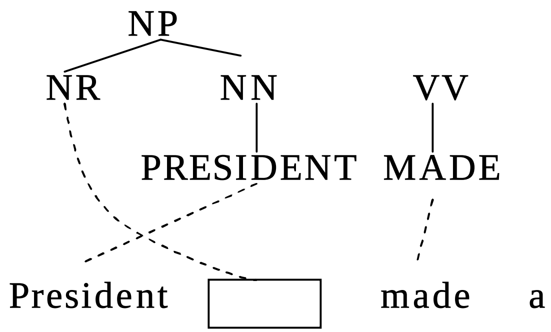
made a speech

An Example





Rule



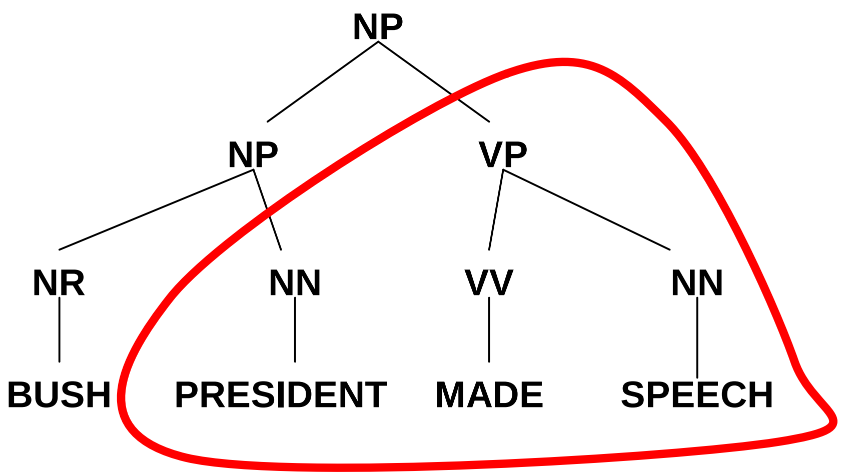
Derivation

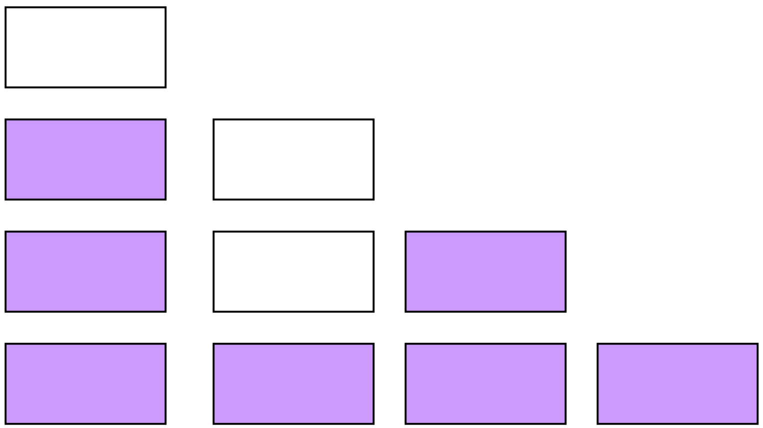
(NP (NN) (NN PRESIDENT)) (VV MADE)
||| President X made a ||| 1:2 2:1 3:3
(NR BUSH) ||| Bush ||| 1:1

Translation

President Bush made a

An Example



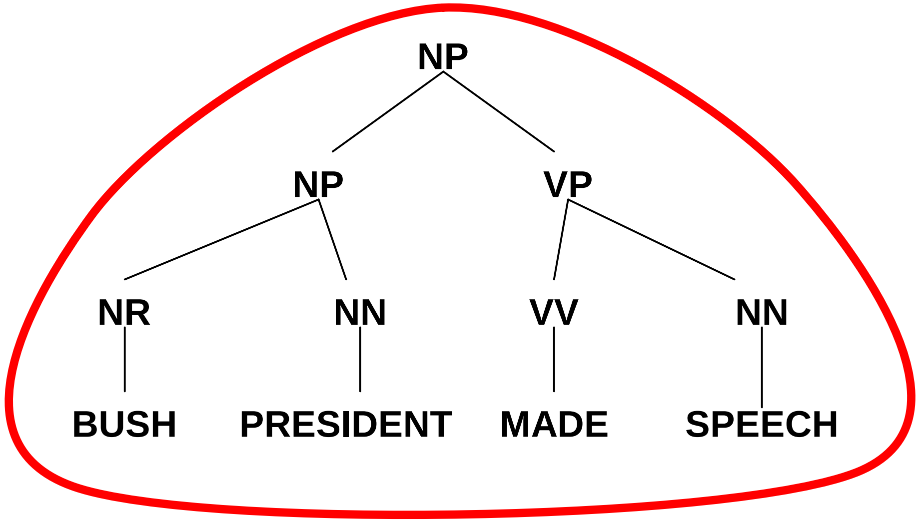


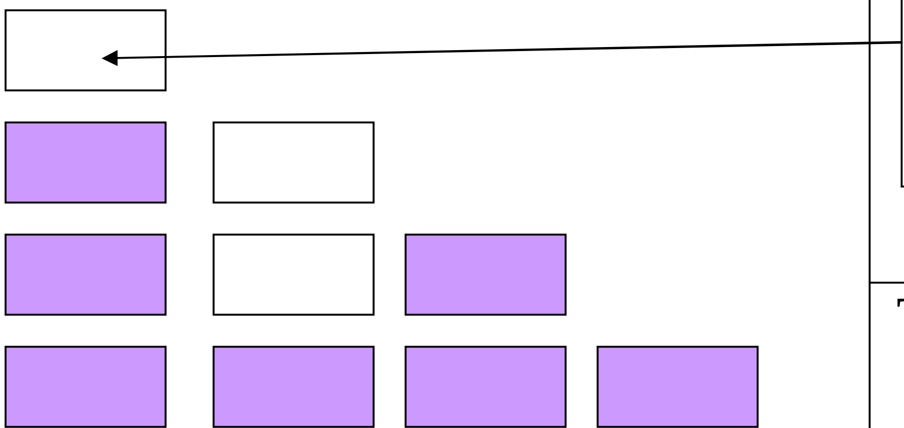
Rule

Derivation

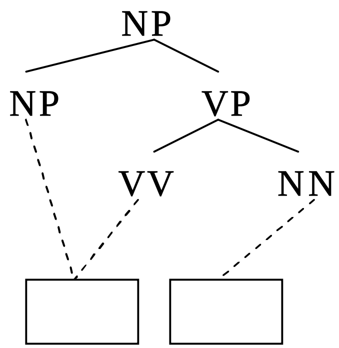
Translation

An Example





Rule



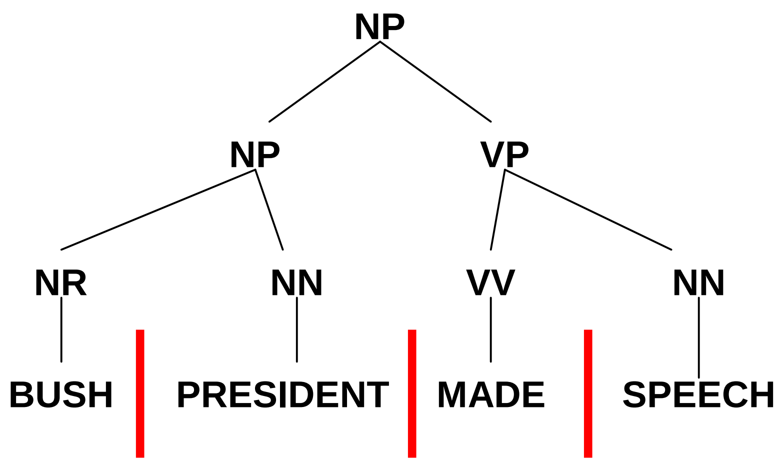
Derivation

(NP (NP) (VP (VV) (NN))) ||| ***X1 X2*** || 1:1 2:1 3:2
(NP (NN) (NN PRESIDENT)) (VV MADE)
||| President X made a ||| 1:2 2:1 3:3
(NR BUSH) ||| Bush ||| 1:1
(NN SPEECH) ||| speech ||| 1:1

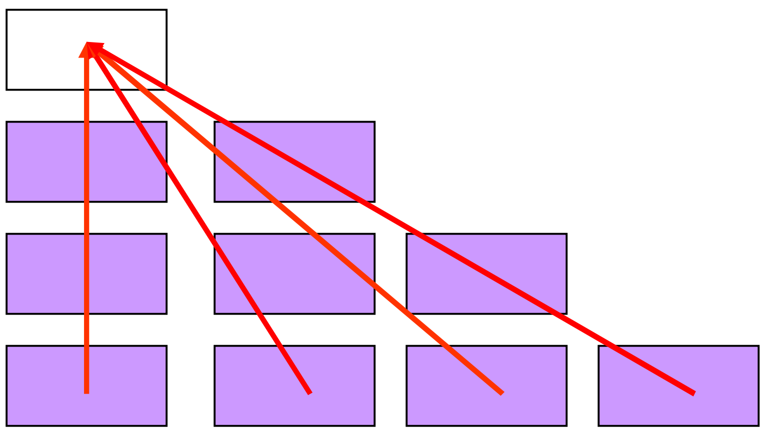
Translation

President Bush made a speech

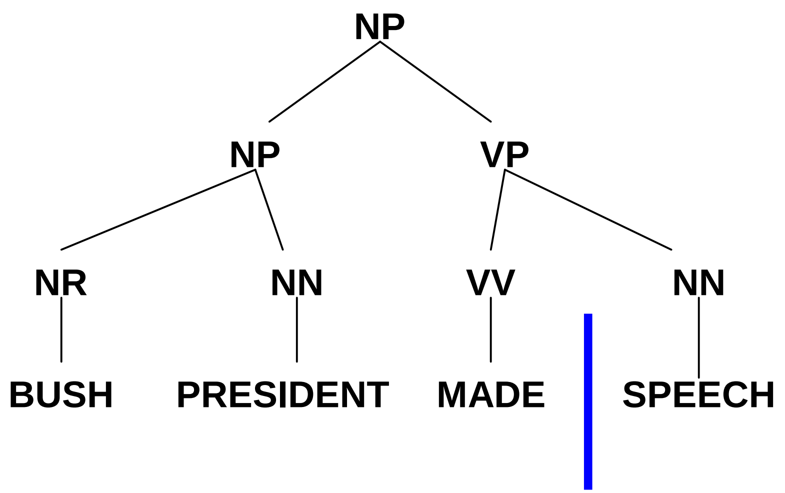
Subcell Division



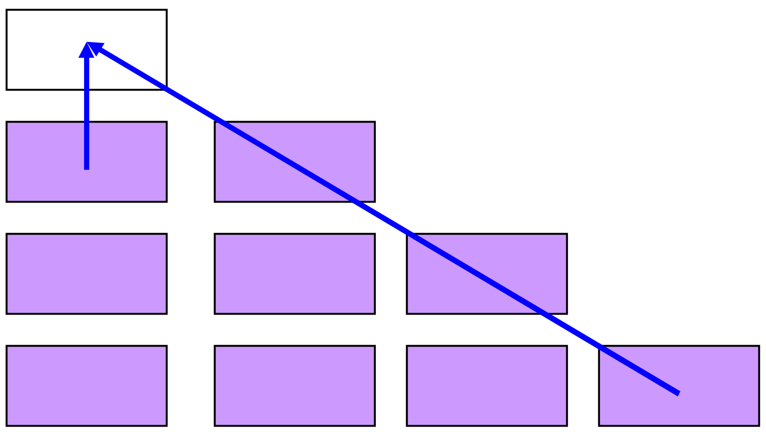
1:1 2:2 3:3 4:4



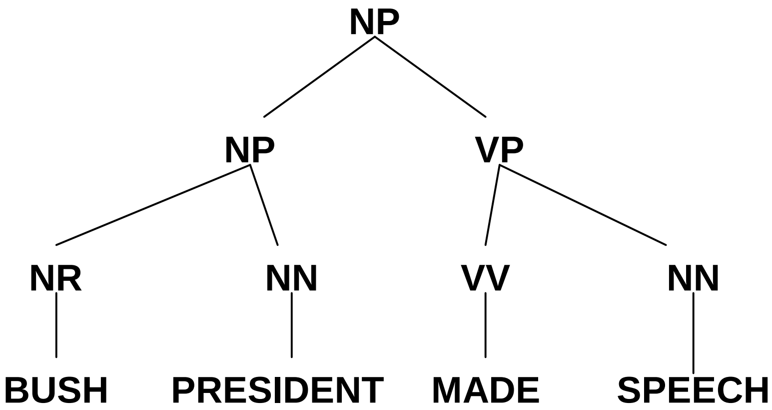
Subcell Division

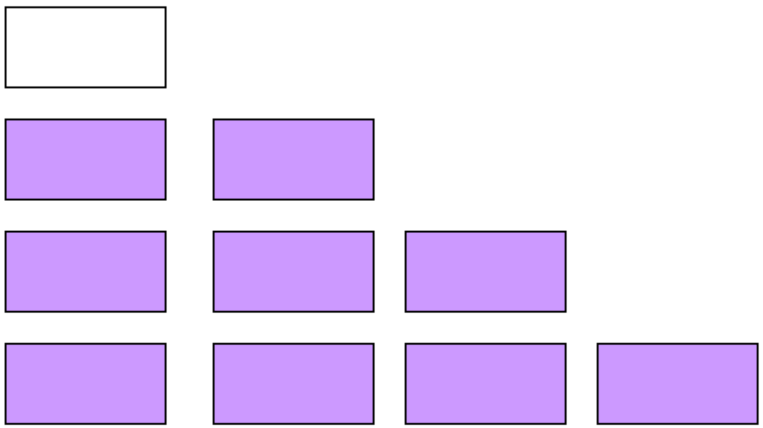


1:3 4:4



Subcell Division

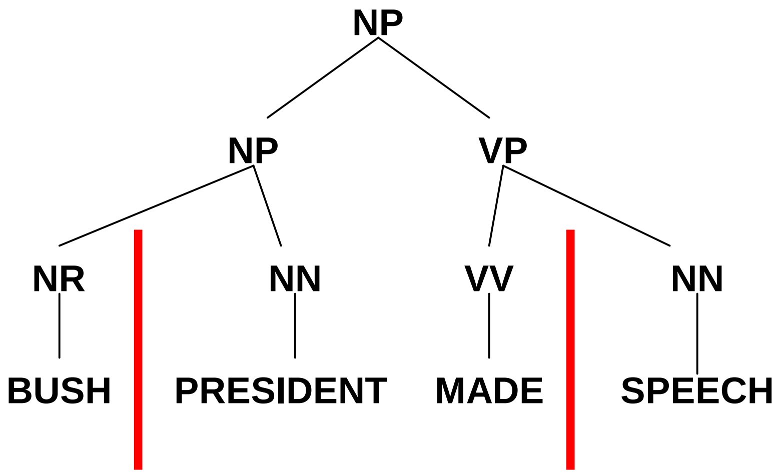


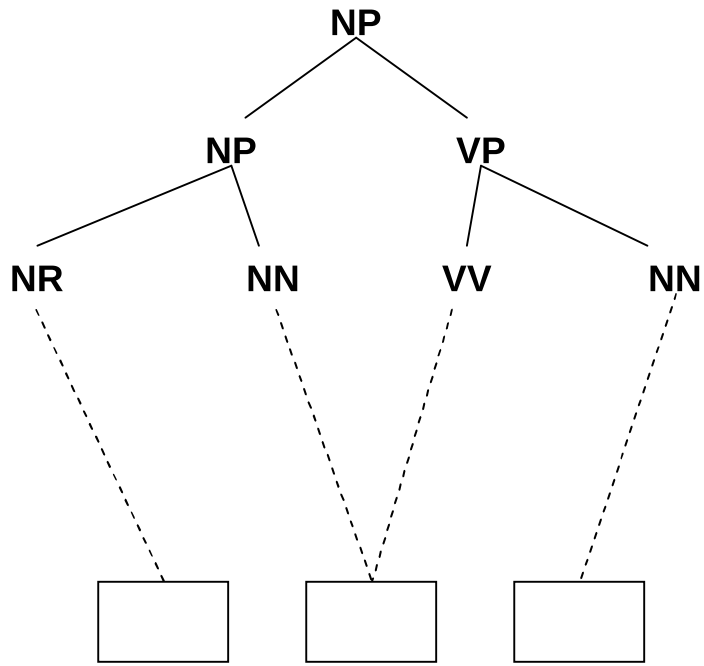


1:4
1:1 2:4
1:2 3:4
1:3 4:4
1:1 2:2 3:4
1:1 2:3 4:4
1:2 3:3 4:4
1:1 2:2 3:3 4:4

2^(n-1)

Build Auxiliary Rule





Penalize the Use of FRs and ARs

- Auxiliary rules, which are built rather than learnt, have no probabilities.
- We introduce a feature that sums up the node count of auxiliary rules to balance the preference between
  - conventional tree-to-string rules
  - new forest-to-string and auxiliary rules

Outline

- *Introduction*
- *Forest-to-String Translation Rules*
- *Training*
- *Decoding*
- Experiments
- Conclusion

Experiments

中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

- Training corpus: 31,149 sentence pairs with 843K Chinese words and 949K English words
- Development set: 2002 NIST Chinese-to-English test set (571 of 878 sentences)
- Test set: 2005 NIST Chinese-to-English test set (1,082 sentences)

Tools

- Evaluation: mteval-v11b.pl
- Language model: SRI Language Modeling Toolkits (Stolcke, 2002)
- Significant test: Zhang et al., 2004
- Parser: Xiong et al., 2005
- Minimum error rate training: optimizeV5IBMBLEU.m (Venugopal and Vogel, 2005)

Rules Used in Experiments

中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

Rule	L	P	U	Total
BP	251, 173	0	0	251,173
TR	56, 983	41, 027	3, 529	101, 539
FR	16, 609	254, 346	25, 051	296, 006

Comparison

System	Rule Set	BLEU4
Pharaoh	BP	0.2182±0.0089
Lynx	BP	0.2059±0.0083
	TR	0.2302±0.0089
	TR + BP	0.2346±0.0088
	TR + FR + AR	0.2402±0.0087

TRs Are Still Dominant

- To achieve the best result of 0.2402, Lynx made use of:
  - 26, 082 tree-to-string rules
  - 9,219 default rules
  - 5,432 forest-to-string rules
  - 2,919 auxiliary rules

Effect of Lexicalization

Forest-to-String Rule Set	BLEU4
None	0.2225±0.0085
L	0.2297± 0.0081
P	0.2279±0.0083
U	0.2270±0.0087
L + P + U	0.2312±0.0082

中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

Outline

- *Introduction*
- *Forest-to-String Translation Rules*
- *Training*
- *Decoding*
- *Experiments*
- Conclusion

Conclusion

中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

- We augment the tree-to-string translation model with
  - forest-to-string rules that capture non-syntactic phrase pairs
  - auxiliary rules that help integrate forest-to-string rules into the tree-to-string model
- Forest and auxiliary rules enable tree-to-string models to derive in a more general way and bring significant improvement.

Future Work

- Scale up to large data
- Further investigation in auxiliary rules

Thanks!