Recent Progress in Syntax-based Machine Translation

Qun Liu

ADAPT Centre, Dublin City University

2016-05-18 Nanjing University

The ADAPT Centre is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

Overview of Syntax in SMT

- **Introduction to Syntax-based SMT**
- Dependency-to-String Translation
- Graph-based Translation
- Dependency-based MT Evaluation
- Conclusion and Future Work

Google Translate - An Example

日本和美国的关系

Relationship between Japan and the United States

日本外交政策和美国亚太再平衡的关系

Japan's foreign policy and the US Asia-Pacific rebalancing of relations

Google Translate - An Example

日本和美国的关系

Relationship between Japan and the United States

日本外交政策和美国亚太再平衡的关系

Japan's foreign policy and the US Asia-Pacific rebalancing of relations

When input sentences become longer, it is more difficult for the Google Translate to capture their syntax structures.

Google Translate - An Example

Aiken, Milam, and Shilpa Balan. "An analysis of Google Translate accuracy." Translation journal 16.2 (2011): 1-3.

Language Paris	→	←
English - French	91	92
English - German	77	86
English - Italian	87	89
English - Japanese	26	49
English - Chinese	17	49

Google Translate - An Example

Aiken, Milam, and Shilpa Balan. "An analysis of Google Translate accuracy." Translation journal 16.2 (2011): 1-3.

Language Paris	→	←
English - French	91	92
English - German	77	86
English - Italian	87	89
English - Japanese	26	49
English - Chinese	17	49

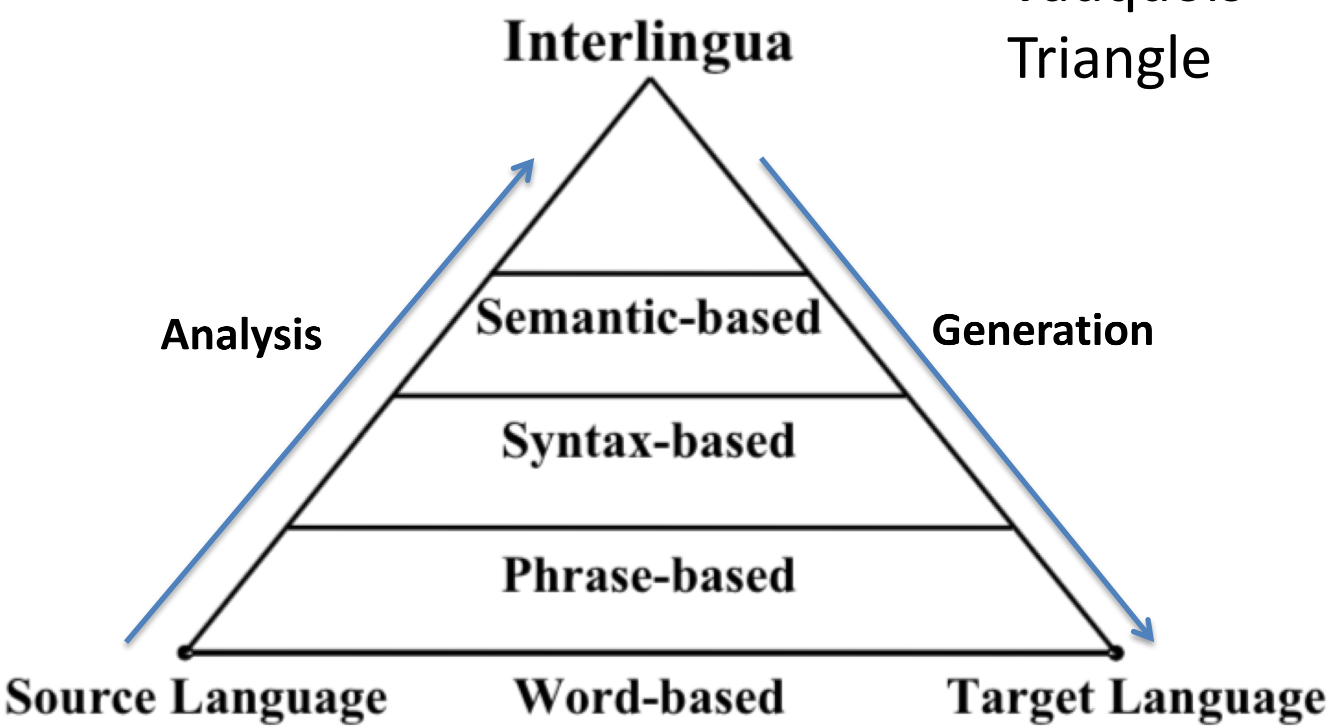
Google Translate performs worse for language pairs with bigger difference in syntax structures.

Overview of Syntax in SMT

- Introduction
- **Syntax-based Translation Models**
- Syntax-based Language Models
- Syntax-based Translation Evaluations
- Conclusion and Future Work

MT Pyramid

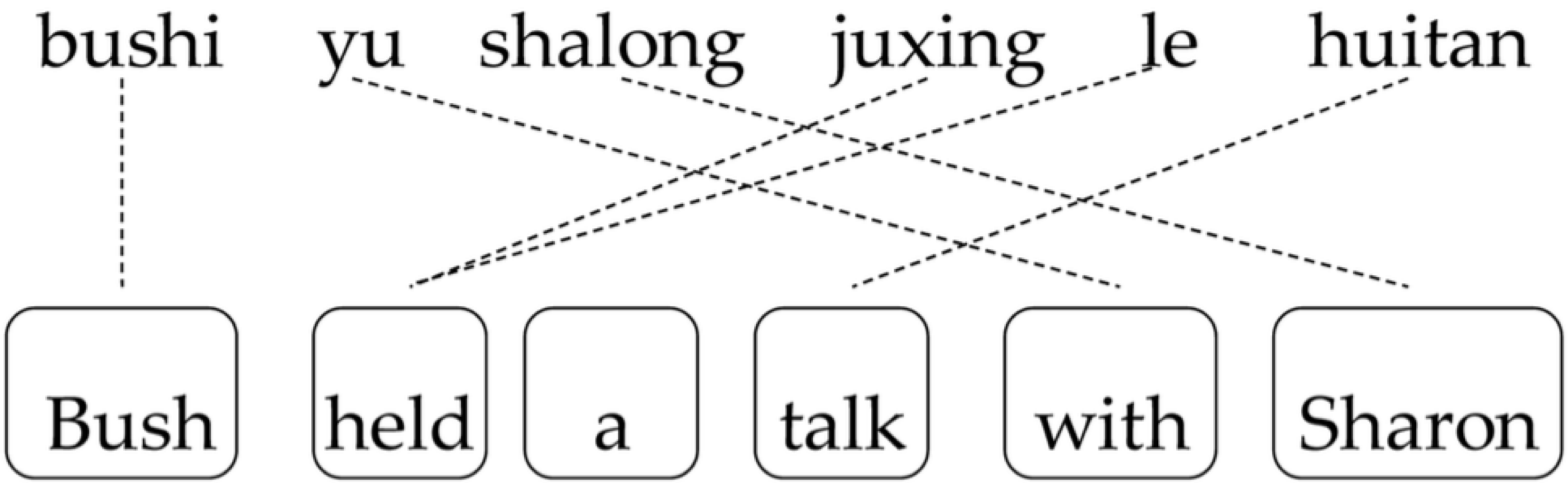
Vauquois' Triangle



IBM Model 1-5

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics, 19(2):263-311.

Word-based Models



IBM Models

Source	Target	Probability
Bushi （布什）	Bush	0.7
	President	0.2
	US	0.1
yu （与）	and	0.6
	with	0.4
juxing （举行）	hold	0.7
	had	0.3
le （了）	hold	0.01

Phrase-based Models

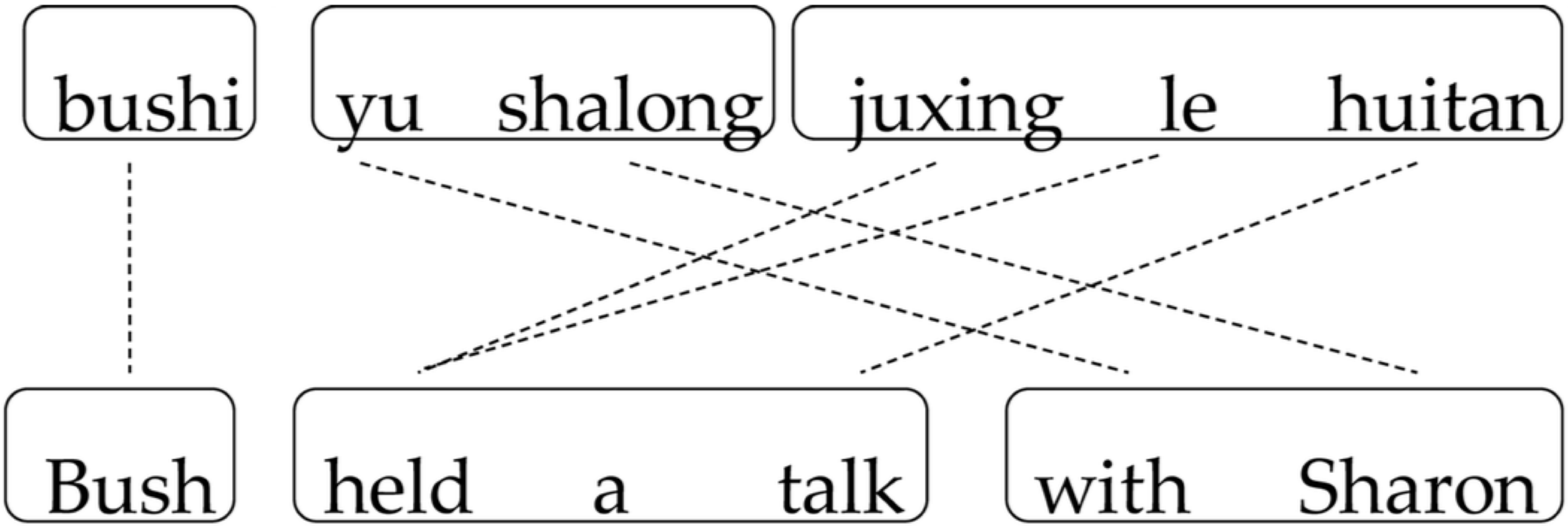
Phrase-based Model

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference, pages 127-133, Edmonton, Canada, May.

Alignment Template Model

Franz J. Och and Hermann Ney. 2004. The Alignment Template Approach to Statistical Machine Translation. Computational Linguistics, 30(4):417-449.

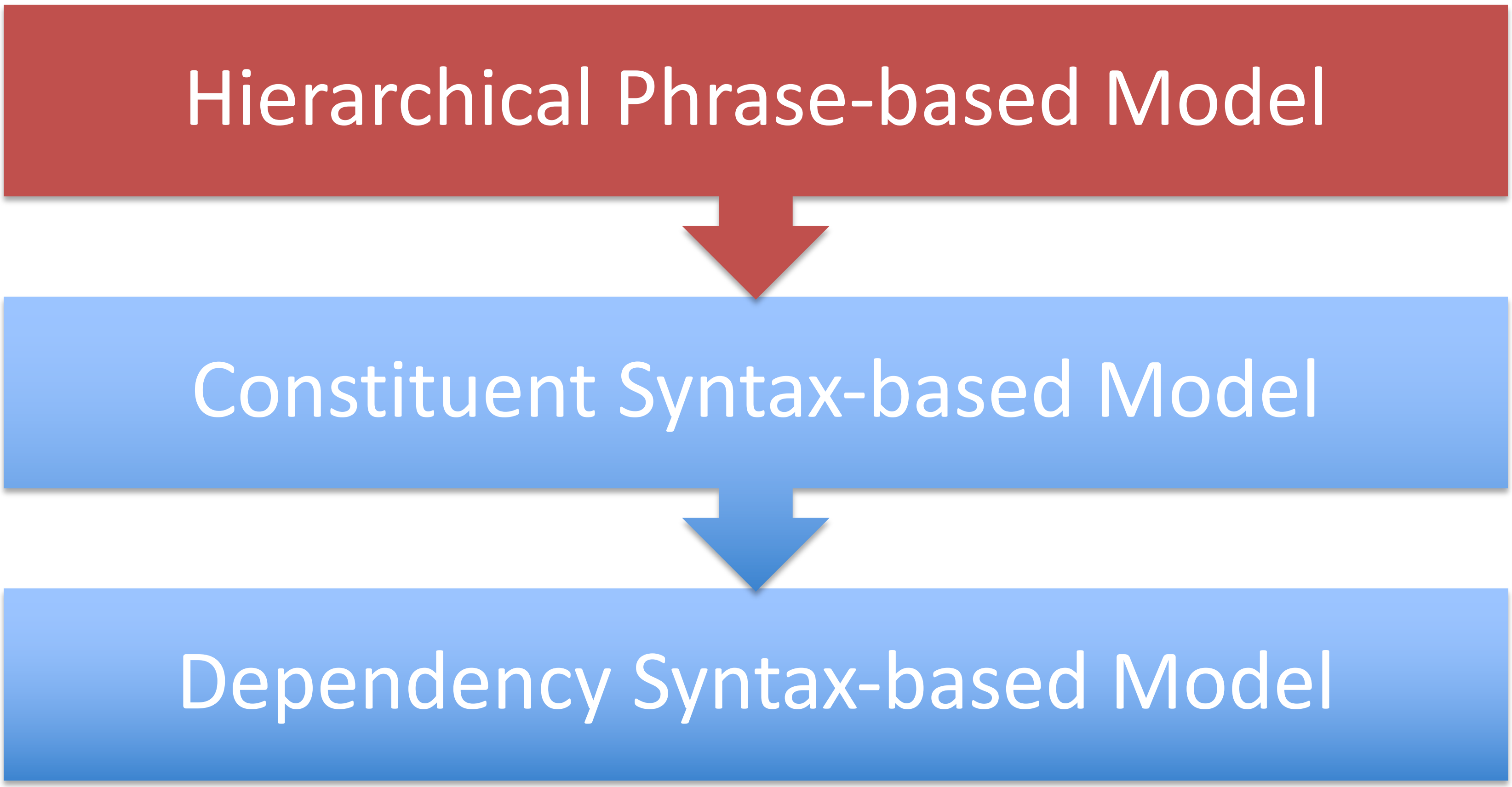
Phrase-based Models



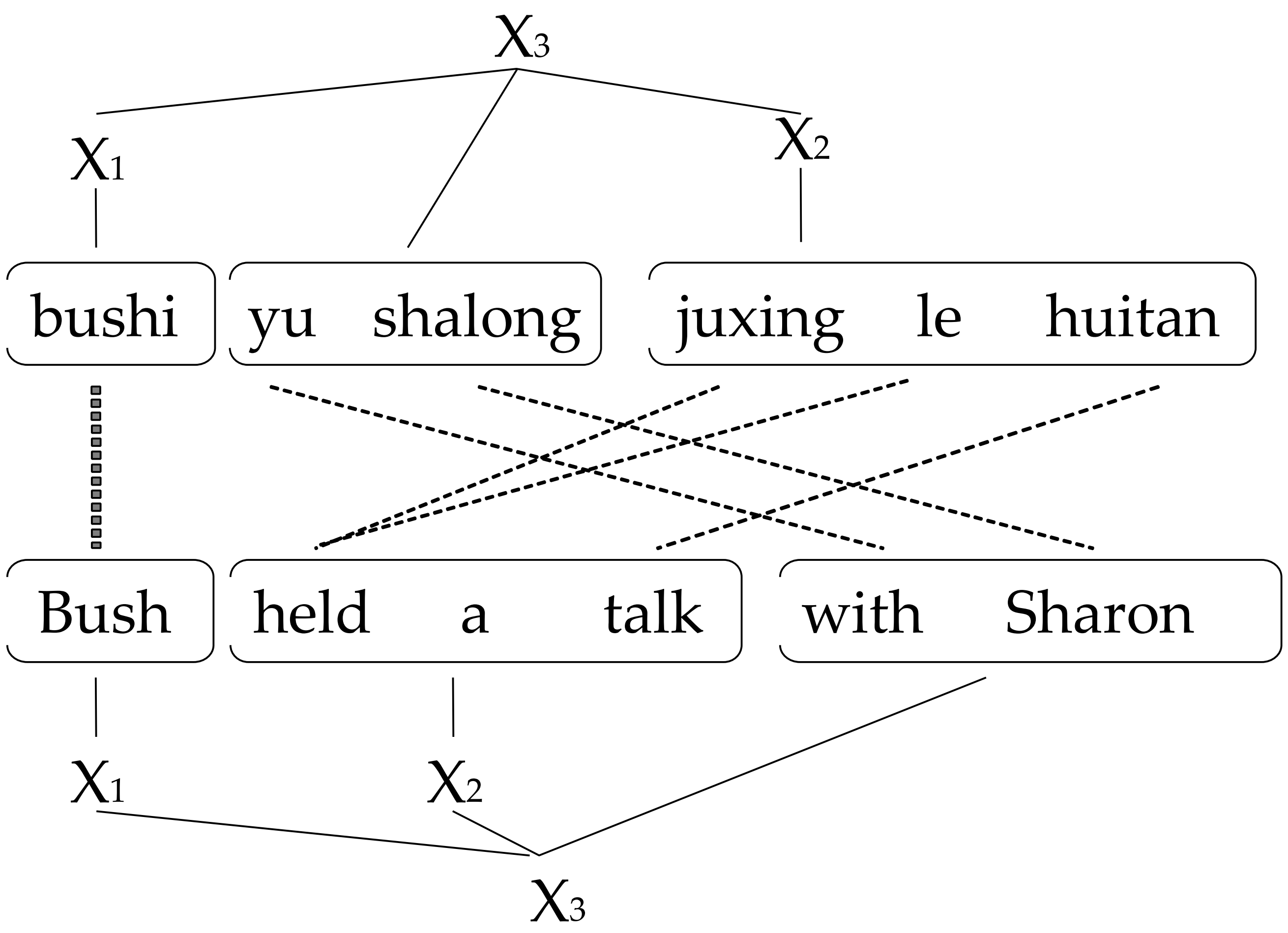
Phrase-based Model

Source	Target	Probability
Bushi (布什)	Bush	0.5
	president Bush	0.3
	the US president	0.2
Bushi yu (布什与)	Bush and	0.8
	the president and	0.2
yu Shalong (与沙龙)	and Shalong	0.6
	with Shalong	0.4
juxing le huiang (举行了会谈)	hold a meeting	0.7
	had a meeting	0.3

Syntax-based Model



Hierarchical Phrase-based Model



Hierarchical Phrase-based Model

Source	Target	Probability
juxing le huiang (举行了会谈)	hold a meeting	0.6
	had a meeting	0.3
X huitang (X会谈)	X a meeting	0.8
	X a talk	0.2
juxing le X (举行了X)	hold a X	0.5
	had a X	0.5
Bushi yu Shalong (布什与沙龙)	Bush and Sharon	0.8
Bushi X (布什X)	Bush X	0.7
X yu Y (X与Y)	X and Y	0.9

Hierarchical Phrase-based Model

Advantage:

- Non-linguistic knowledge used
  - Language Independent
- High Performance
  - Synchronous CFG

Disadvantage:

- Limitation in long distance dependency
  - Use of Glue Rules for long phrases

Advantage:

- Non-linguistic knowledge used
  – Language Independent
- High Performance
  – Synchronous CFG

Disadvantage:

- Limitation in long distance dependency
  – Use of Glue Rules for long phrases

Synchronous CFG

$$S \rightarrow \langle NP_{\boxed{1}}\ VP_{\boxed{2}},\ NP_{\boxed{1}}\ VP_{\boxed{2}} \rangle$$

$$VP \rightarrow \langle V_{\boxed{1}}\ NP_{\boxed{2}},\ NP_{\boxed{2}}\ V_{\boxed{1}} \rangle$$

$$NP \rightarrow \langle \text{i},\ \text{watashi wa} \rangle$$

$$NP \rightarrow \langle \text{the box},\ \text{hako wo} \rangle$$

$$V \rightarrow \langle \text{open},\ \text{akemasu} \rangle$$

The implementation of decoding algorithm is straightforward – just like a parsing procedure, either CYK or Chart algorithm works

Hierarchical Phrase-based Model

Advantage:

- Non-linguistic knowledge used
  - Language Independent
- High Performance
  - Synchronous CFG

Disadvantage:

- Limitation in long distance dependency
  - Use of Glue Rules for long phrases

Glue Rules

$$S \rightarrow \langle S_{\boxed{1}} X_{\boxed{2}}, S_{\boxed{1}} X_{\boxed{2}} \rangle$$

$$S \rightarrow \langle X_{\boxed{1}}, X_{\boxed{1}} \rangle$$

- Using Glue Rules means sequentially concatenating all the target phrases,
  which lead to a back-off to phrase based model
- Two cases to use Glue Rules:
  - ➢ No hierarchical rules applicable
  - ➢ The span to be covered by the hierarchical rule is longer than a threshold

Glue Rules

$$S \rightarrow \langle S_{\boxed{1}} X_{\boxed{2}}, S_{\boxed{1}} X_{\boxed{2}} \rangle$$

$$S \rightarrow \langle X_{\boxed{1}}, X_{\boxed{1}} \rangle$$

- Using Glue Rul
  the phrases wh
  model
- Two cases to use Glue
  - ➢ No hierarchical ru es applicable
  - ➢ The span to be covered by the hierarchical rule is longer than a threshold

Hierarchical Rules failed to capture dependency between words with a distance longer than a threshold

Syntax-based Model

Hierarchical Phrase-based Model

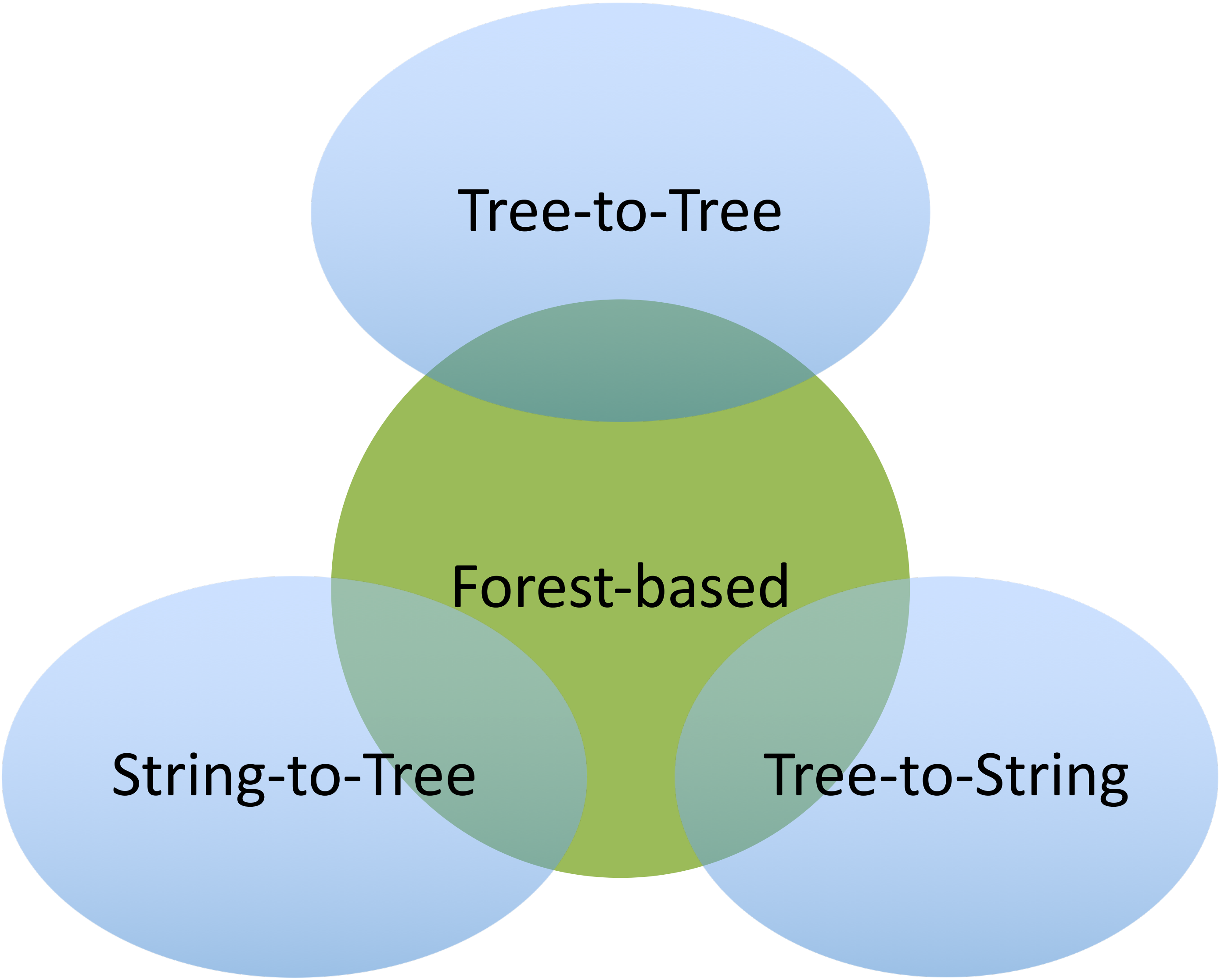
↓

Constituent Syntax-based Model

↓

Dependency Syntax-based Model

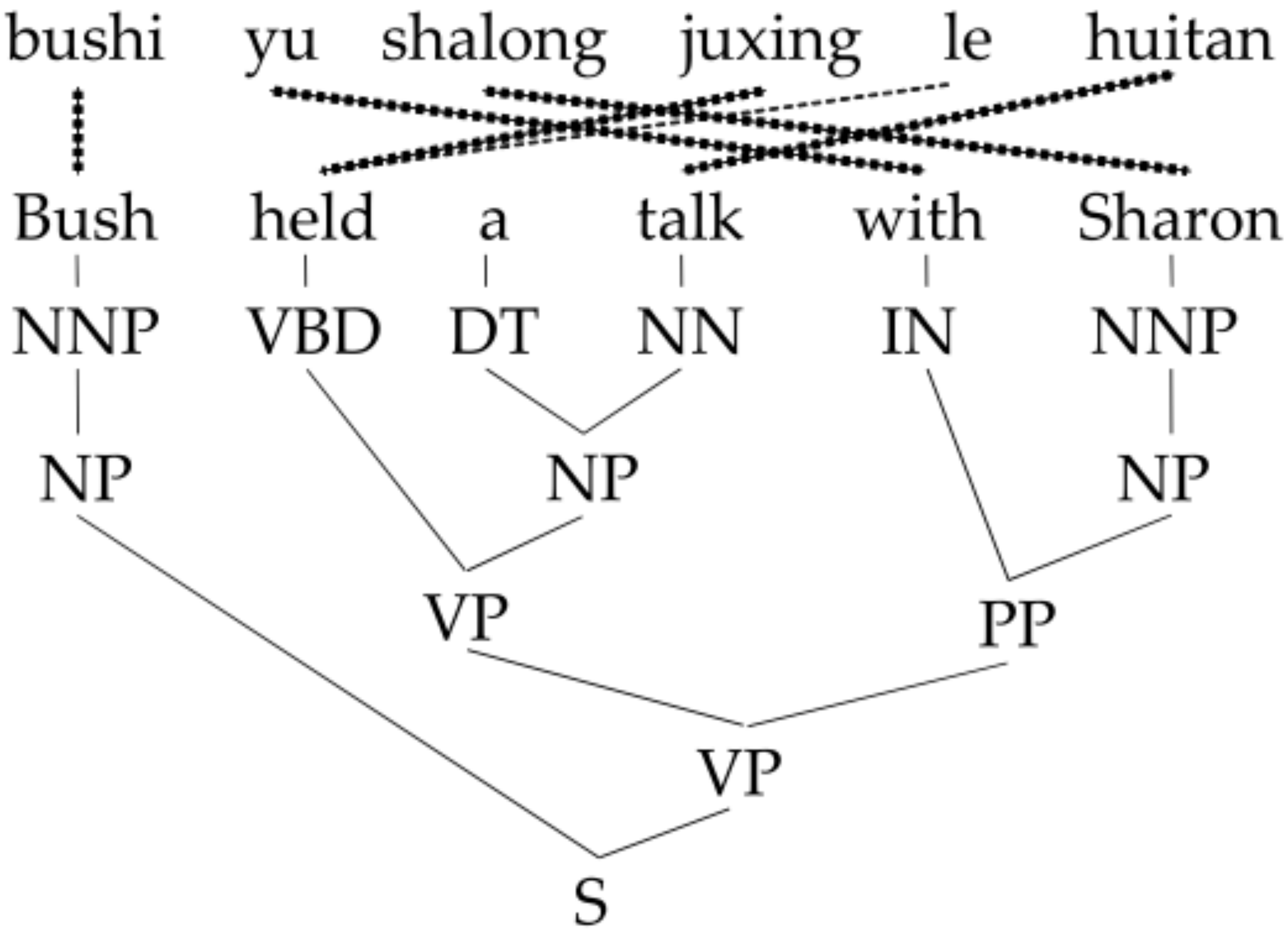
Constituent Syntax-based Models



String-to-Tree Model

- Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical machine translation model. In Proceedings of ACL 2001.
- Daniel Marcu, Wei Wang, Abdessamad Echihabi, and Kevin Knight. 2006. SPMT: Statistical machine translation with syntactified target language phrases. In Proceedings of EMNLP 2006.
- Michel Galley, Jonathan Graehl, Kevin Knight, Daniel Marcu, Steve DeNeefe, Wei Wang, and Ignacio Thayer. 2006. Scalable inference and training of context-rich syntactic translation models. In Proceedings of COLING-ACL 2006.

String-to-Tree Model



String-to-Tree Model

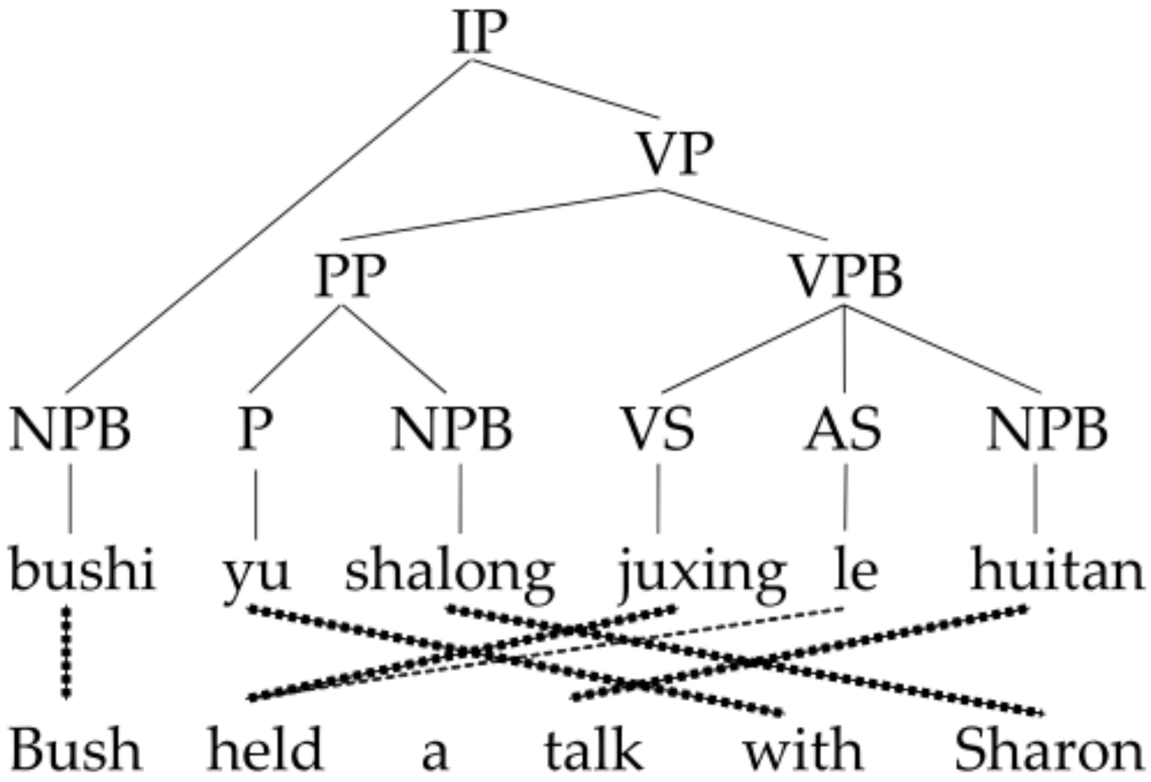
Source	Target	Probability
juxing le huiang（举行了会谈）	VP(VPD(hold) NP(DT(a) NN(meeting)))	0.6
	VP(VPD(had) NP(DP(a) NN(meeting)))	0.3
	VP(VPD(had) NP(DT(a) NN(talk)))	0.1
x_1 huitang（x_1会谈）	VP(x_1:VPD NP(DT(a) NN(meeting)))	0.8
	VP(x_1:VPD NP(DT(a) NN(talk)))	0.2
juxing le x_1（举行了x_1）	VP(VPD(hold) NP(DT(a) x_1:NN))	0.5
	VP(VPD(had) NP(DT(a) x_1:NN))	0.5
x_1 yu x_2（x_1与x_2）	NP(x_1:NNP CC(and) x_2:NNP))	0.9

Tree-to-String Model

- Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-String Alignment Template for Statistical Machine Translation. In Proceedings of COLING/ACL 2006, pages 609-616, Sydney, Australia, July.

  (Meritorious Asian NLP Paper Award)
- Huang, Liang, Kevin Knight, and Aravind Joshi. "Statistical syntax-directed translation with extended domain of locality." Proceedings of AMTA. 2006.

Tree-to-String Model



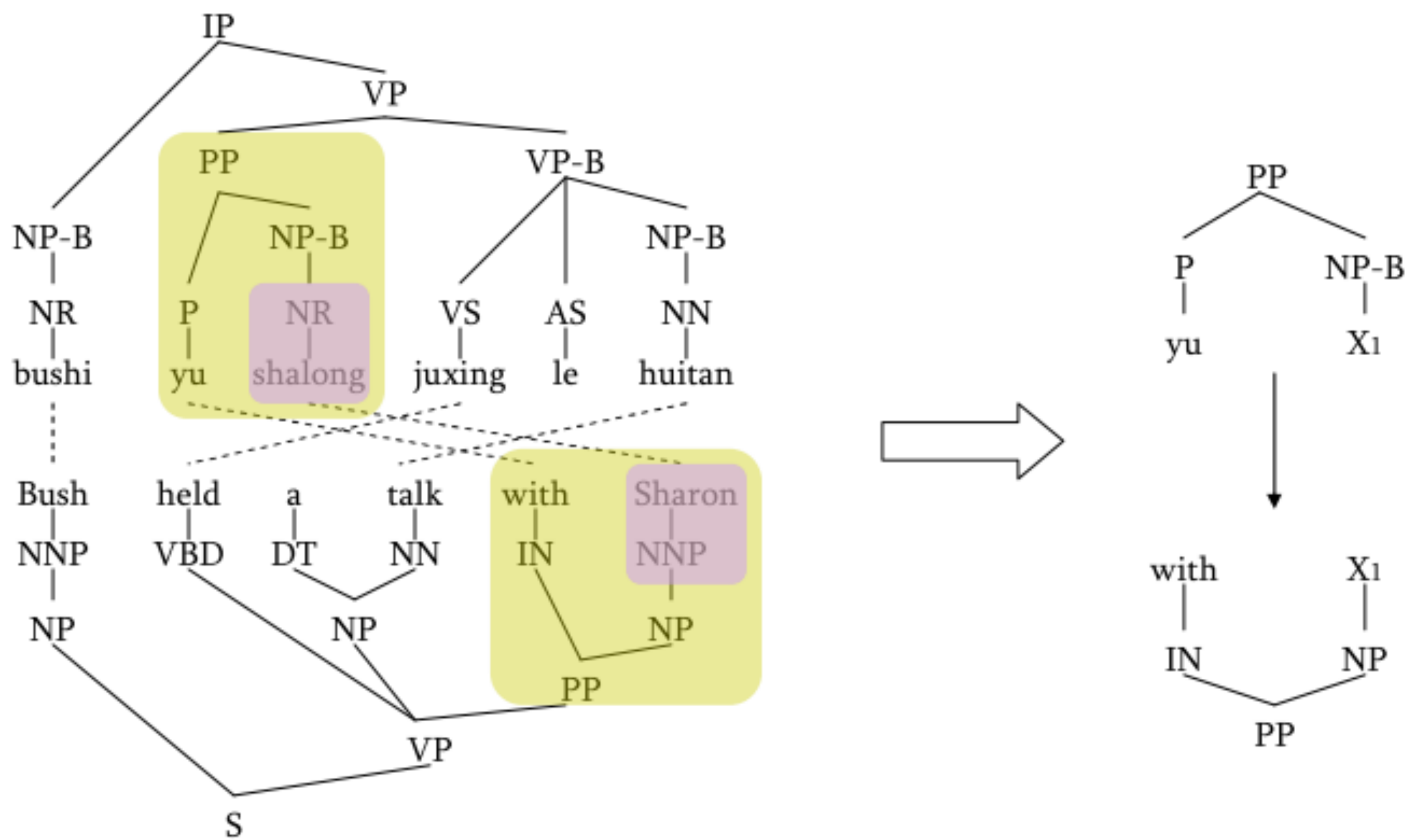
Tree-to-String Model

Source	Target	Probability
VPB(VS(juxing) AS(le) NPB(huiang)) （举行了会谈）	hold a meeting	0.6
	have a meeting	0.3
	have a talk	0.1
VPB(VS(juxing) AS(le) x_1) （举行了x_1）	hold a x_1	0.5
	have a x_1	0.5
VP(PP(P(yu) x_1:NPB) x_2:VPB) （与 x_1 x_2）	x_2 with x_1	0.9
IP(x_1:NPB VP(x_2:PP x_3:VPB))	x_1 x_3 x_2	0.7

Tree-to-Tree Model

- Jason Eisner. 2003. Learning non-isomorphic tree mappings for machine translation. In Proc. of ACL 2003
- Min Zhang, Hongfei Jiang, Aiti Aw, Haizhou Li, Chew Lim Tan, and Sheng Li. "A tree sequence alignment-based tree-to-tree translation model." *ACL-08: HLT* (2008): 559.
- Yang Liu, Yajuan Lü, and Qun Liu. 2009. Improving Tree-to-Tree Translation with Packed Forests. In Proceedings of ACL/IJCNLP 2009, pages 558-566, Singapore, August.

Tree-to-Tree Model



Constituent Syntax-based Models

Advantage:

- Linguistic knowledge used
  - Long distance dependency

Disadvantage:

- Ungrammatical phrases
- Syntactic Ambiguity
- Computational Complexity
  - Synchronous TSG

Constituent Syntax-based Models

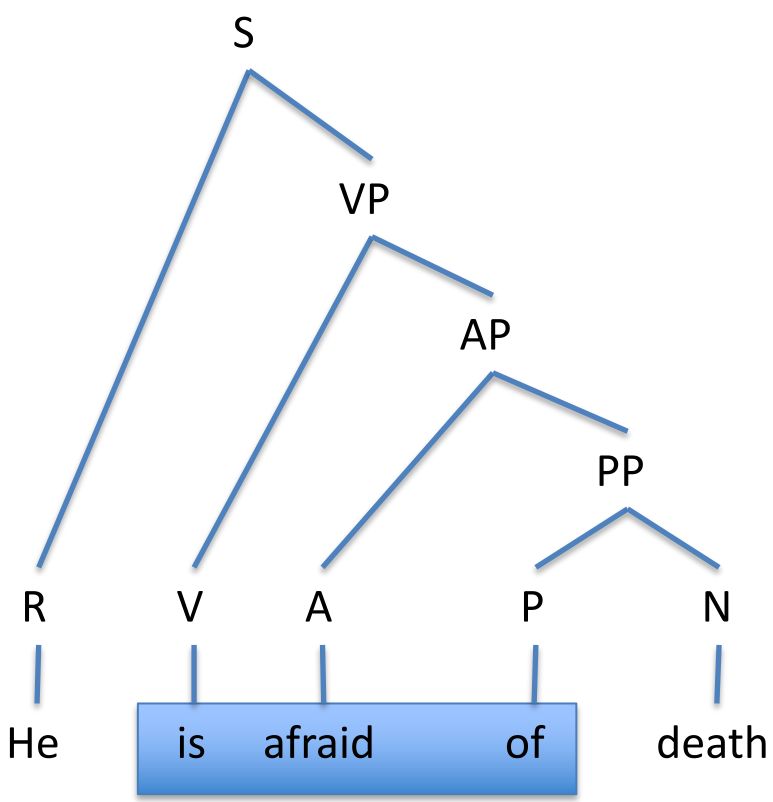
Advantage:

- Linguistic knowledge used
  - Long distance dependency

Disadvantage:

- Ungrammatical phrases
- Syntactic Ambiguity
- Computational Complexity
  - Synchronous TSG

Ungrammatical Phrases



- Pure tree-based models get very low performance, even lower than phrase-based models
- Various techniques are developed to incorporate ungrammatical phrases into tree-based models, which lead to an significant improvement on tree-based models

Constituent Syntax-based Models

Advantage:

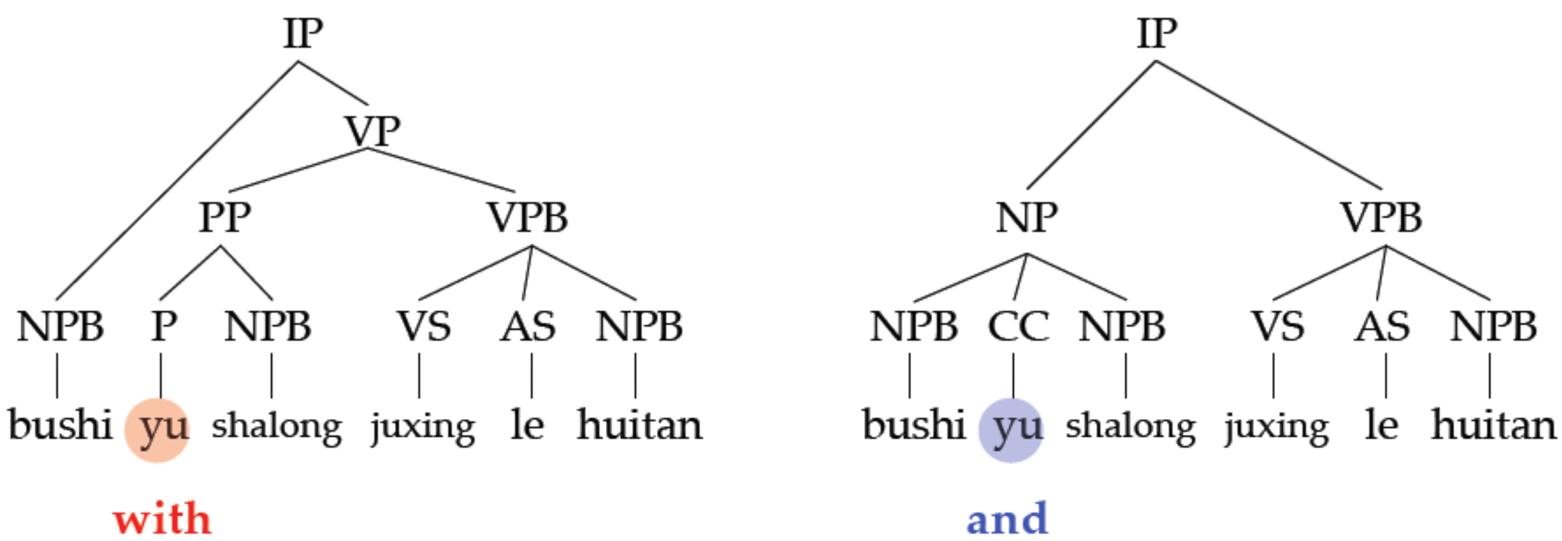
- Linguistic knowledge used
  - ➢ Long distance dependency

Disadvantage:

- Ungrammatical phrases
- Syntactic Ambiguity
- Computational Complexity
  - ➢ Synchronous TSG

Syntactic Ambiguity

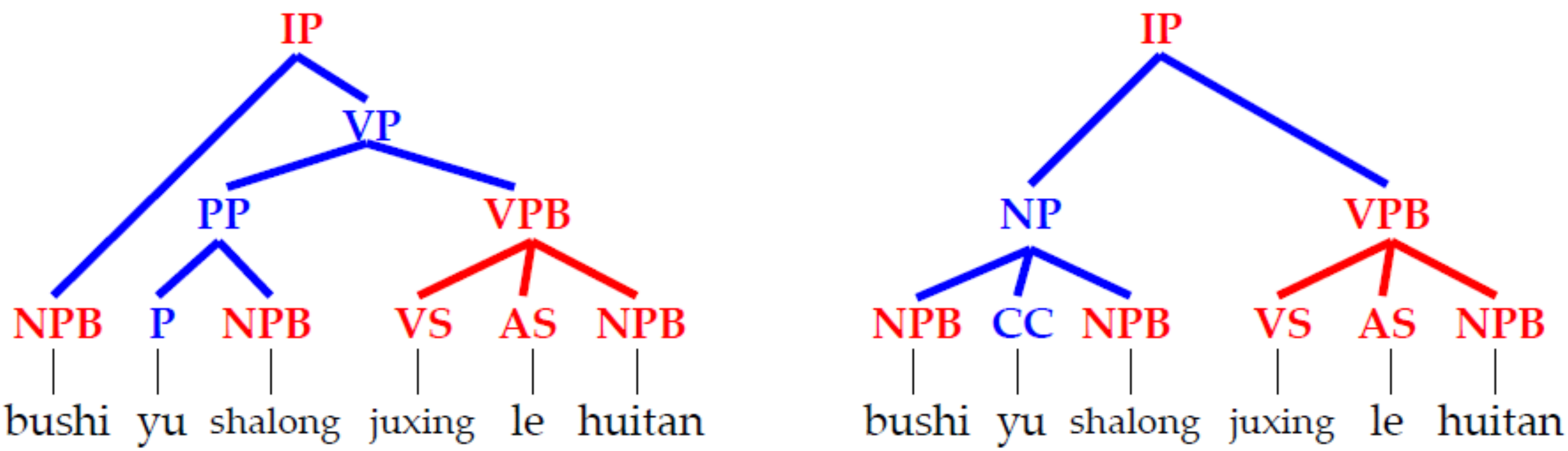
It is important to choose a correct tree for producing a good translation!



``Bush held a talk with Sharon''

``Bush and Sharon held a talk''

1-best → n-best trees?

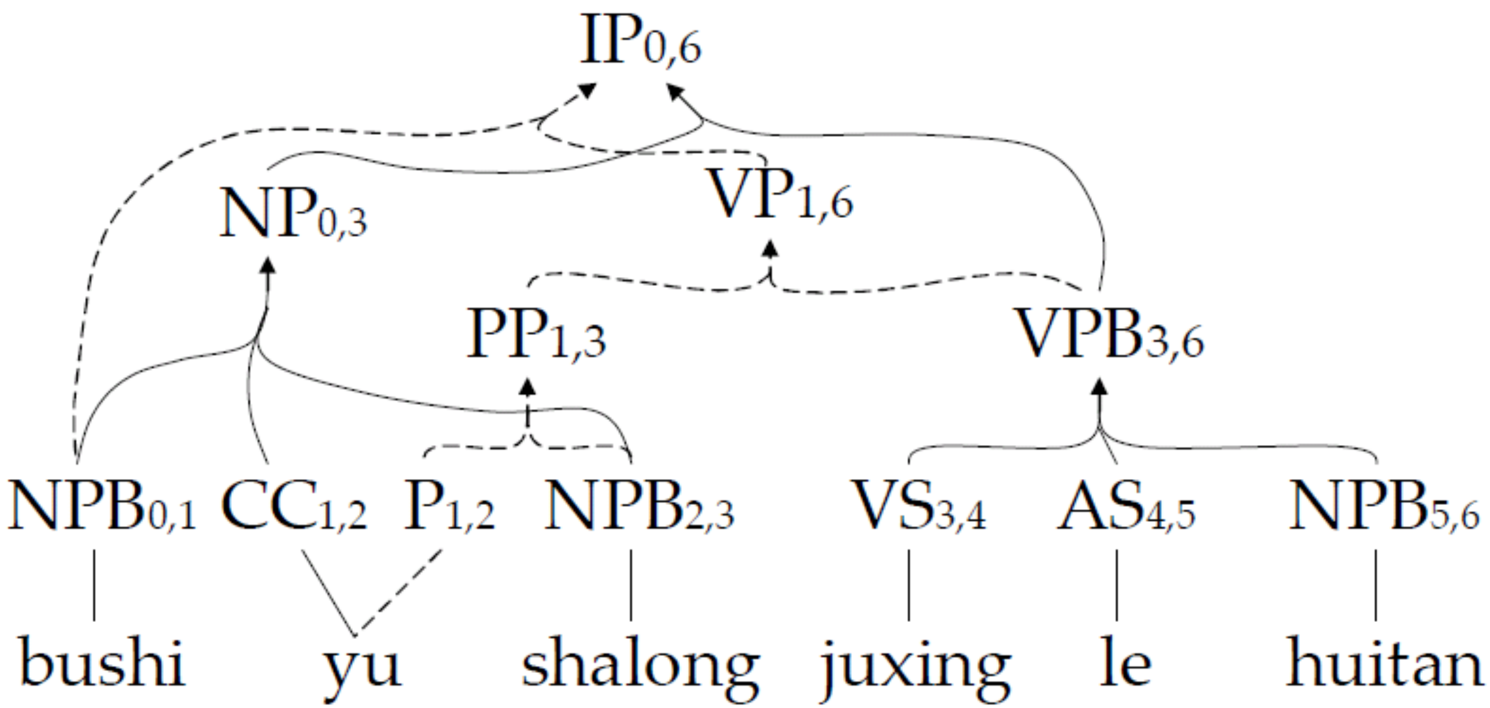


Very few variations among the *n*-best trees!

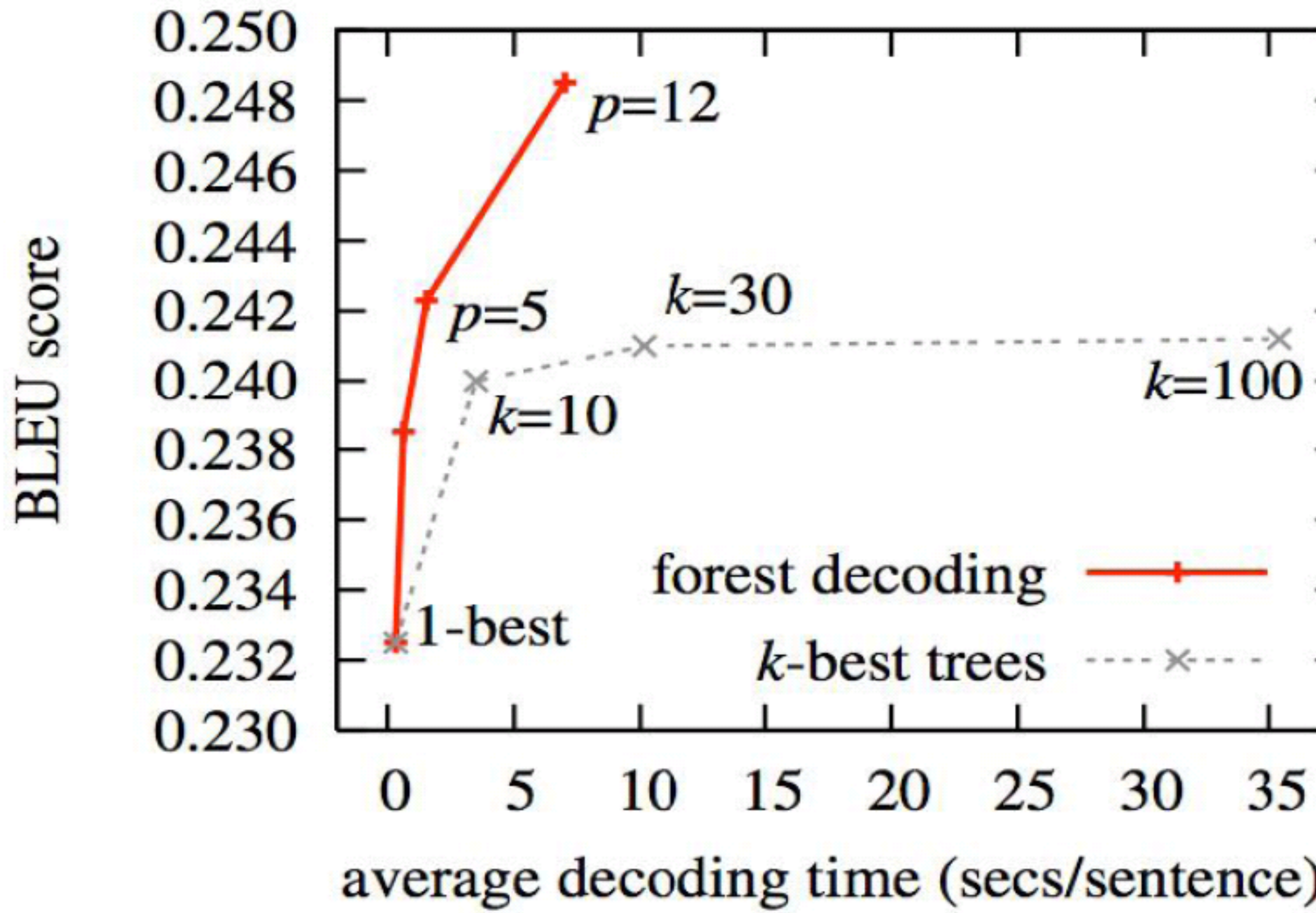
Forest-based Translation

- Mi, Haitao, Liang Huang, and Qun Liu. "Forest-Based Translation." Proceedings of ACL 2008.
- Mi, Haitao, and Liang Huang. "Forest-based translation rule extraction." Proceedings of the EMNLP 2008.

Packed Forest



N-best Trees vs. Forest



Constituent Syntax-based Models

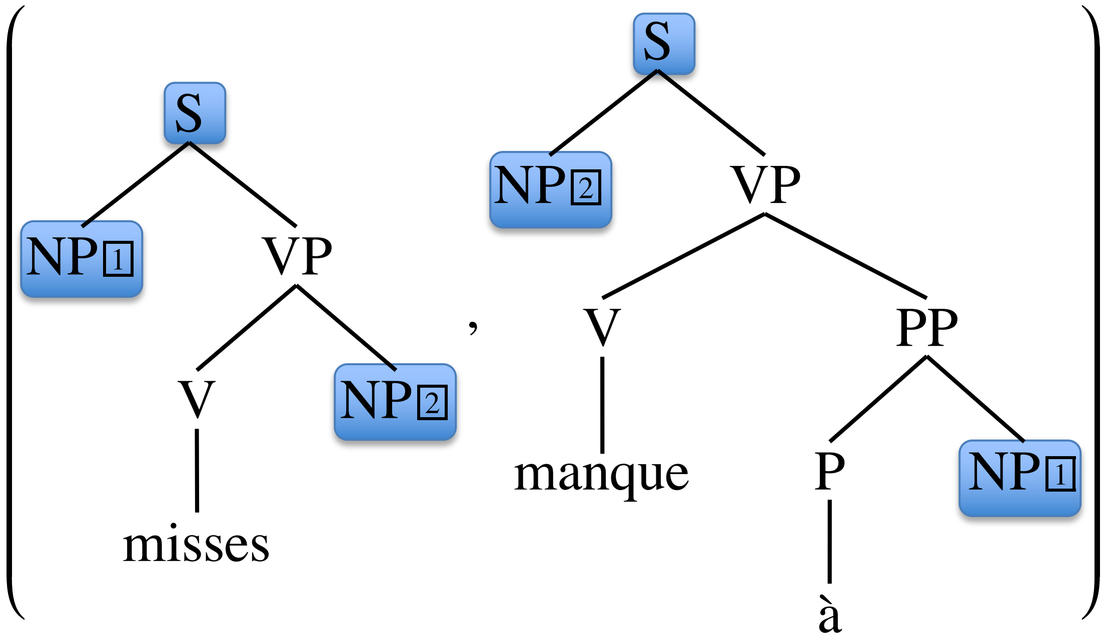
Advantage:

- Linguistic knowledge used
  - ➢ Long distance dependency

Disadvantage:

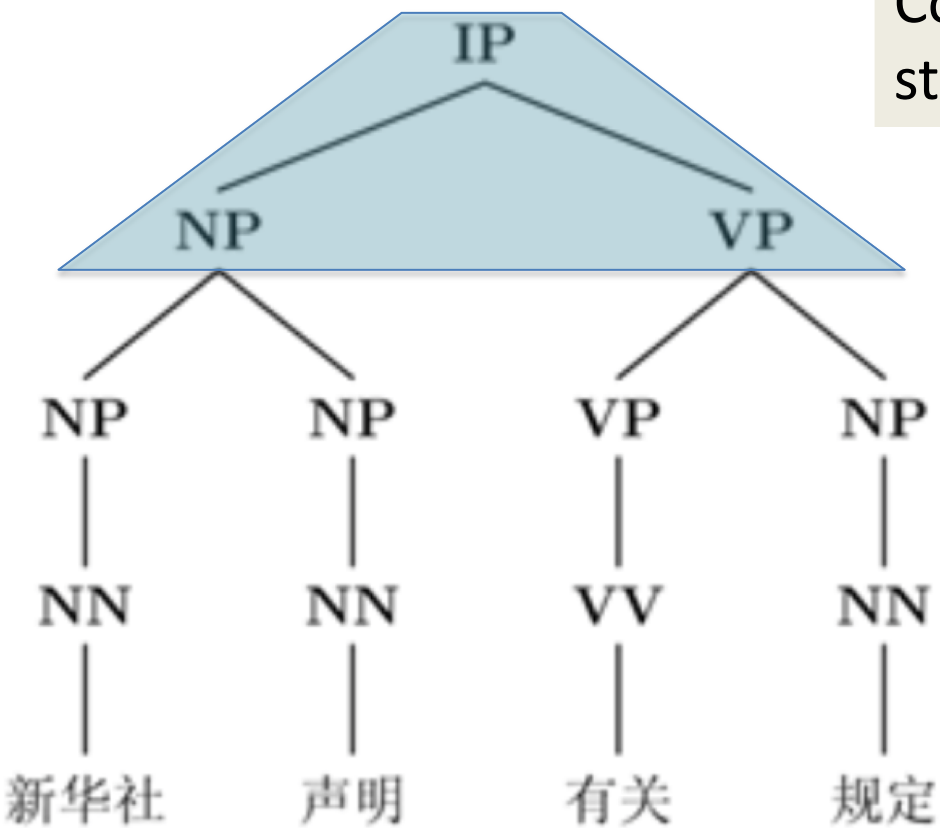
- Ungrammatical phrases
- Syntactic Ambiguity
- Computational Complexity
  - ➢ Synchronous TSG

Synchronous TSG



Synchronous CFG can be regarded as a special case of Synchronous TSG where the trees are limited to have only two layers of nodes

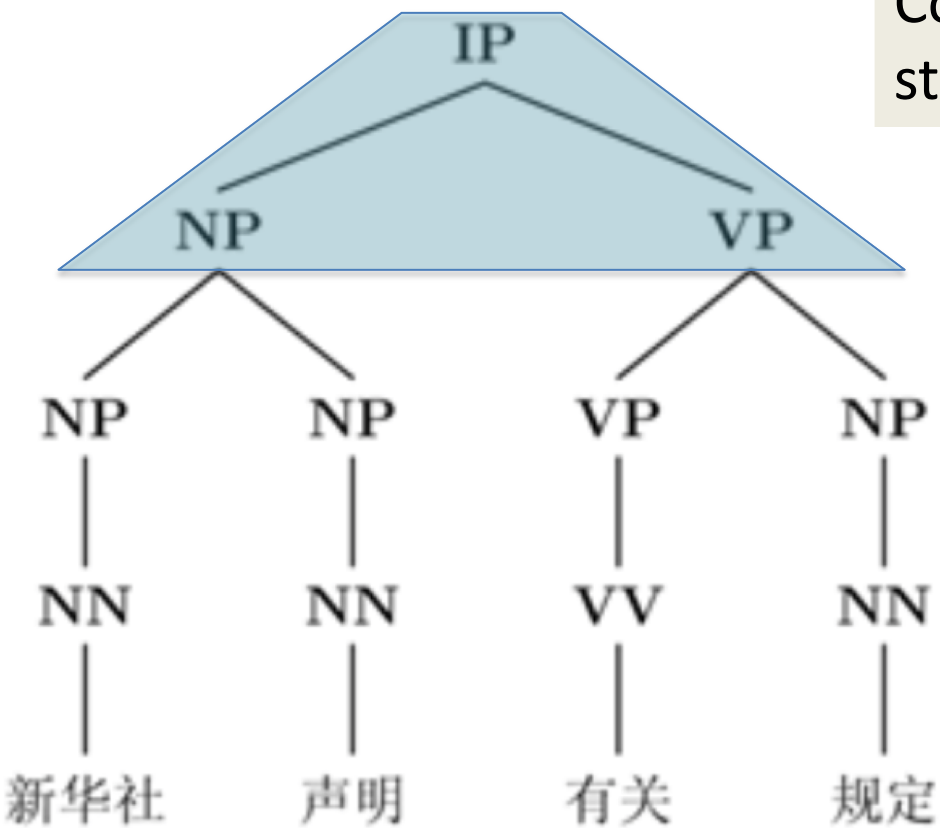
Synchr. TSG vs Synchr. CFG



Considering matching a rule starting from the root node

For Synchronous CFG, there is only one possible tree in the source side

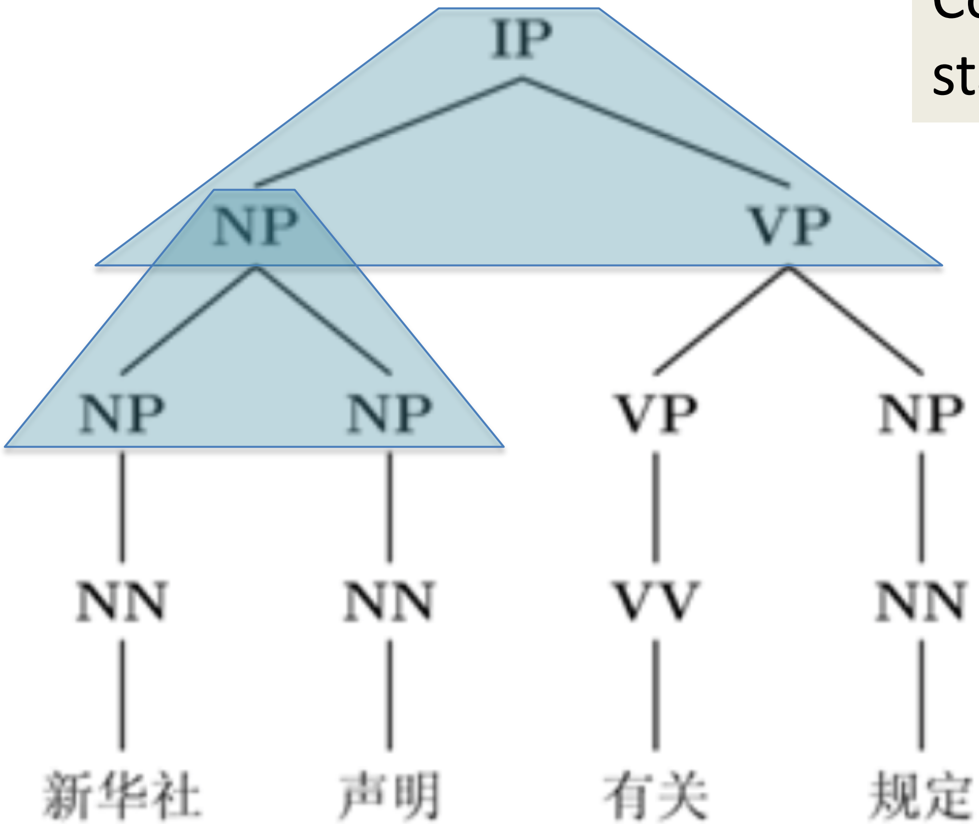
Synchr. TSG vs Synchr. CFG



Considering matching a rule starting from the root node

For Synchronous TSG, there are a large number of possible trees in the source side

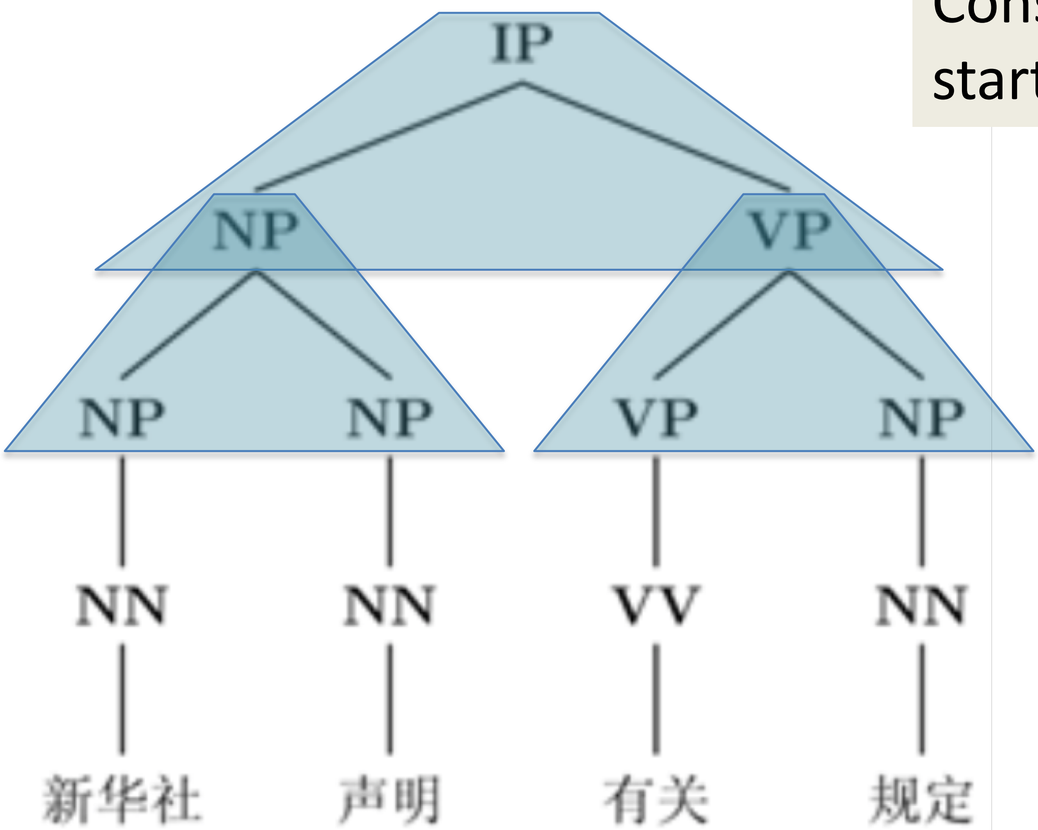
Synchr. TSG vs Synchr. CFG



Considering matching a rule starting from the root node

For Synchronous TSG, there are a large number of possible trees in the source side

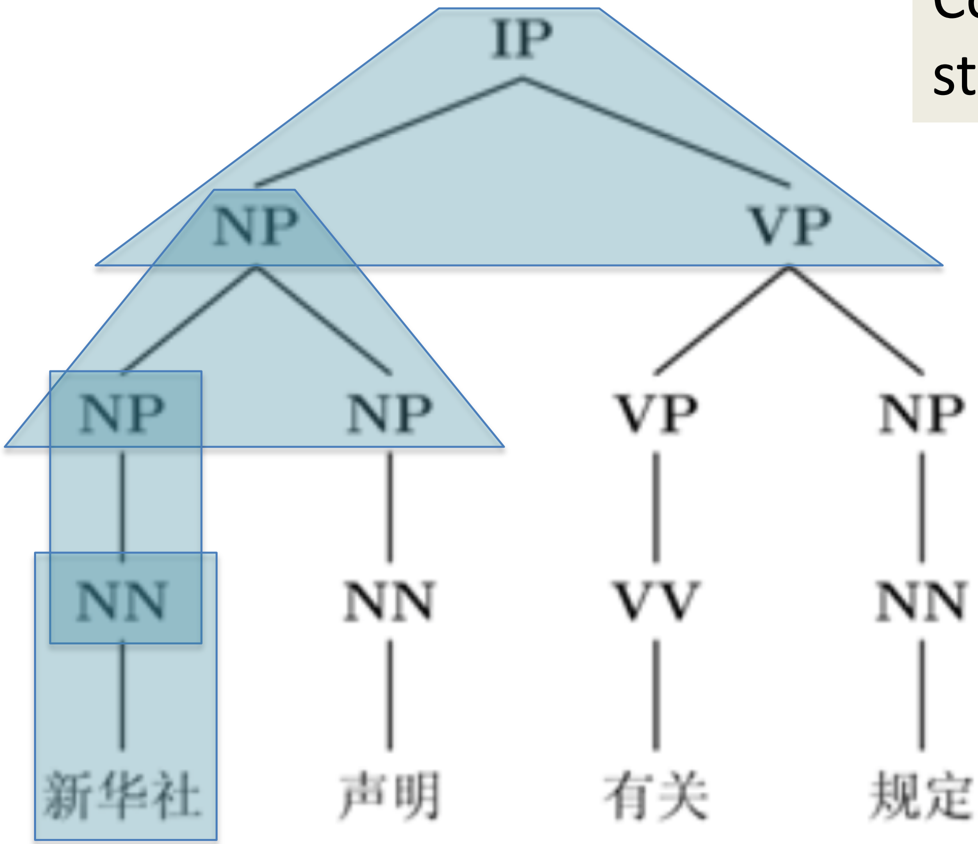
Synchr. TSG vs Synchr. CFG



Considering matching a rule starting from the root node

For Synchronous TSG, there are a large number of possible trees in the source side

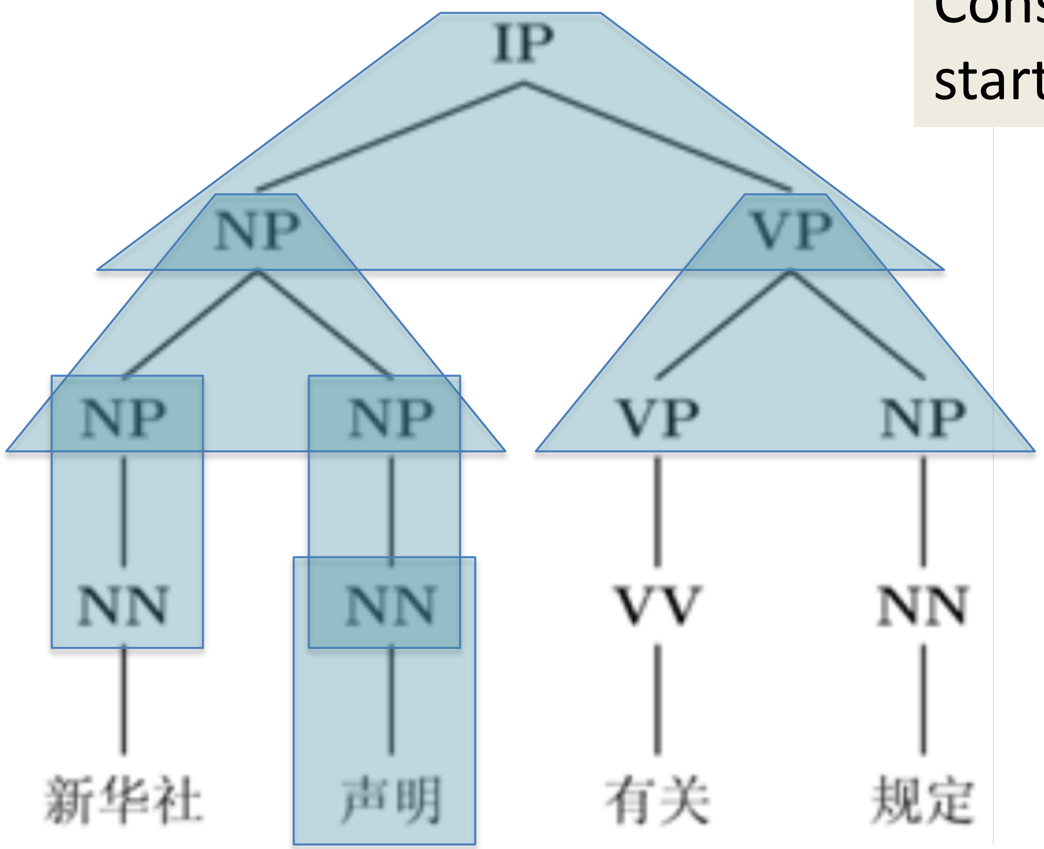
Synchr. TSG vs Synchr. CFG



Considering matching a rule starting from the root node

For Synchronous TSG, there are a large number of possible trees in the source side

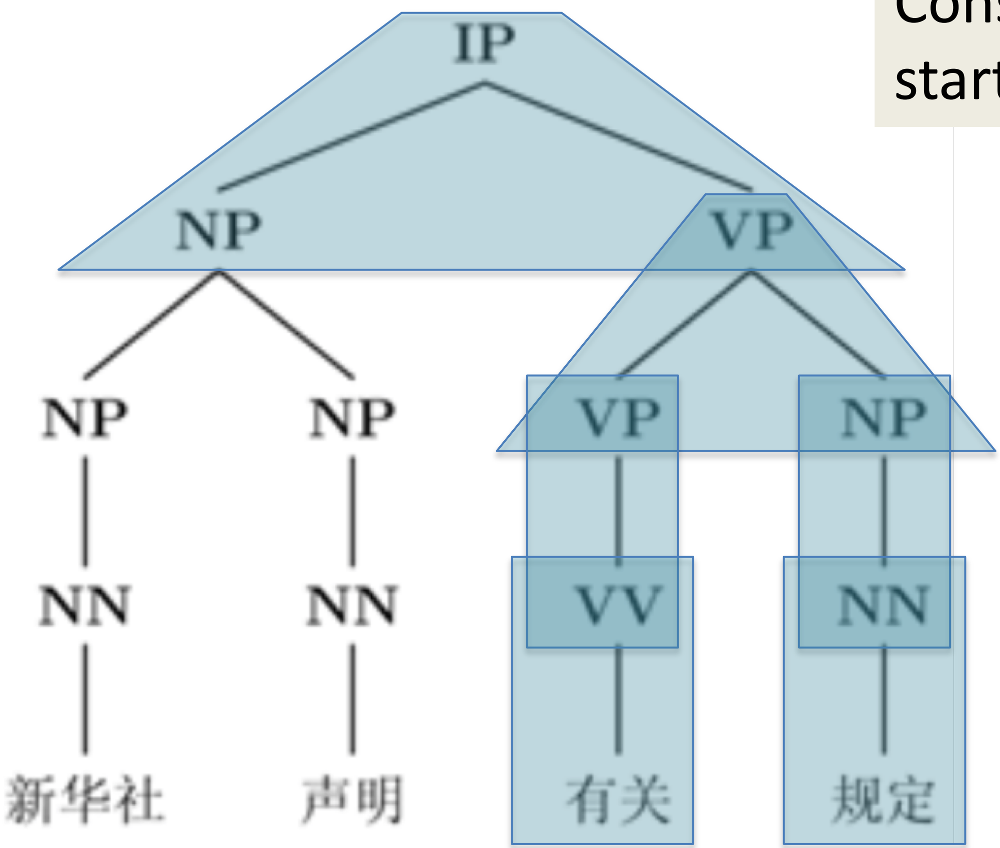
Synchr. TSG vs Synchr. CFG



Considering matching a rule starting from the root node

For Synchronous TSG, there are a large number of possible trees in the source side

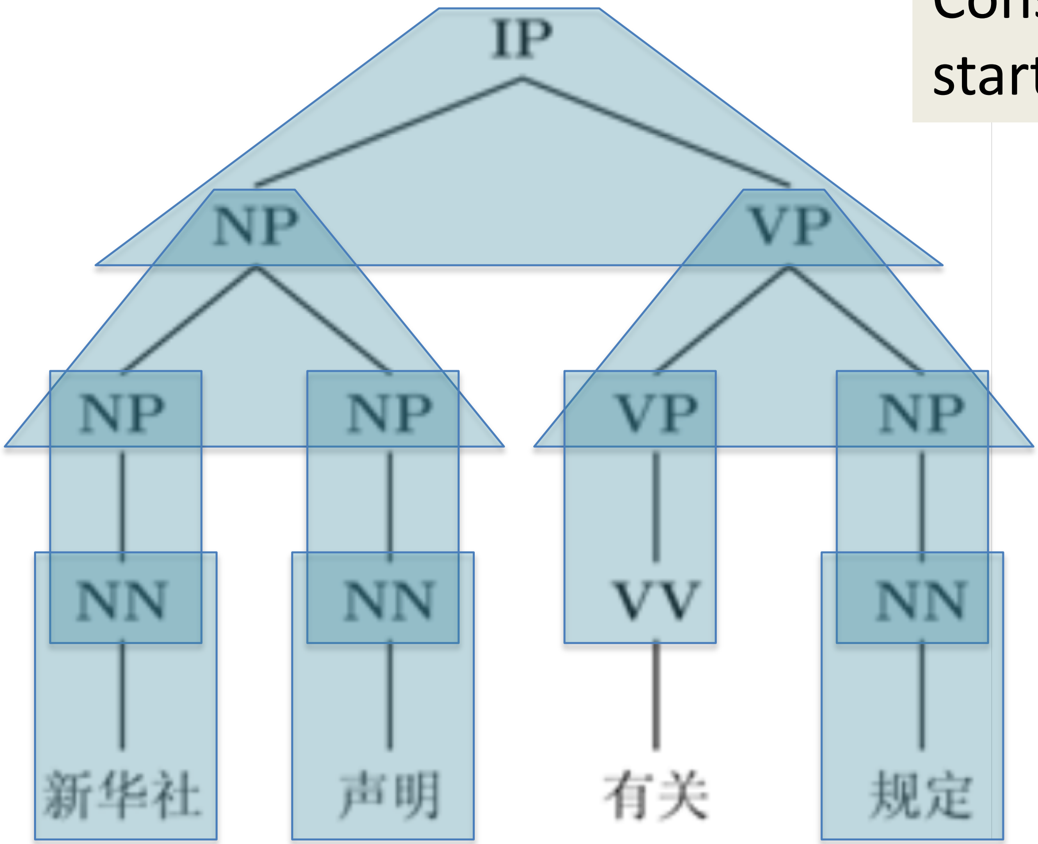
Synchr. TSG vs Synchr. CFG



Considering matching a rule starting from the root node

For Synchronous TSG, there are a large number of possible trees in the source side

Synchr. TSG vs Synchr. CFG



Considering matching a rule starting from the root node

For Synchronous TSG, there are a large number of possible trees in the source side

Synchr. TSG vs Synchr. CFG

- The implementation of Synchronous TSG is much more complex than Synchronous CFG, both in space and in time
- Technologies are developed to deal with the rule indexing problem for Synchronous TSG decoder [Zhang et al., ACL-IJCNLP 2009]
- The syntax based decoder implemented in Moses does not support Synchronous TSG model with rules having more than two layers.

Constituent Syntax-based Models

Advantage:

- Linguistic knowledge used
  - ➢ Long distance dependency

Disadvantage:

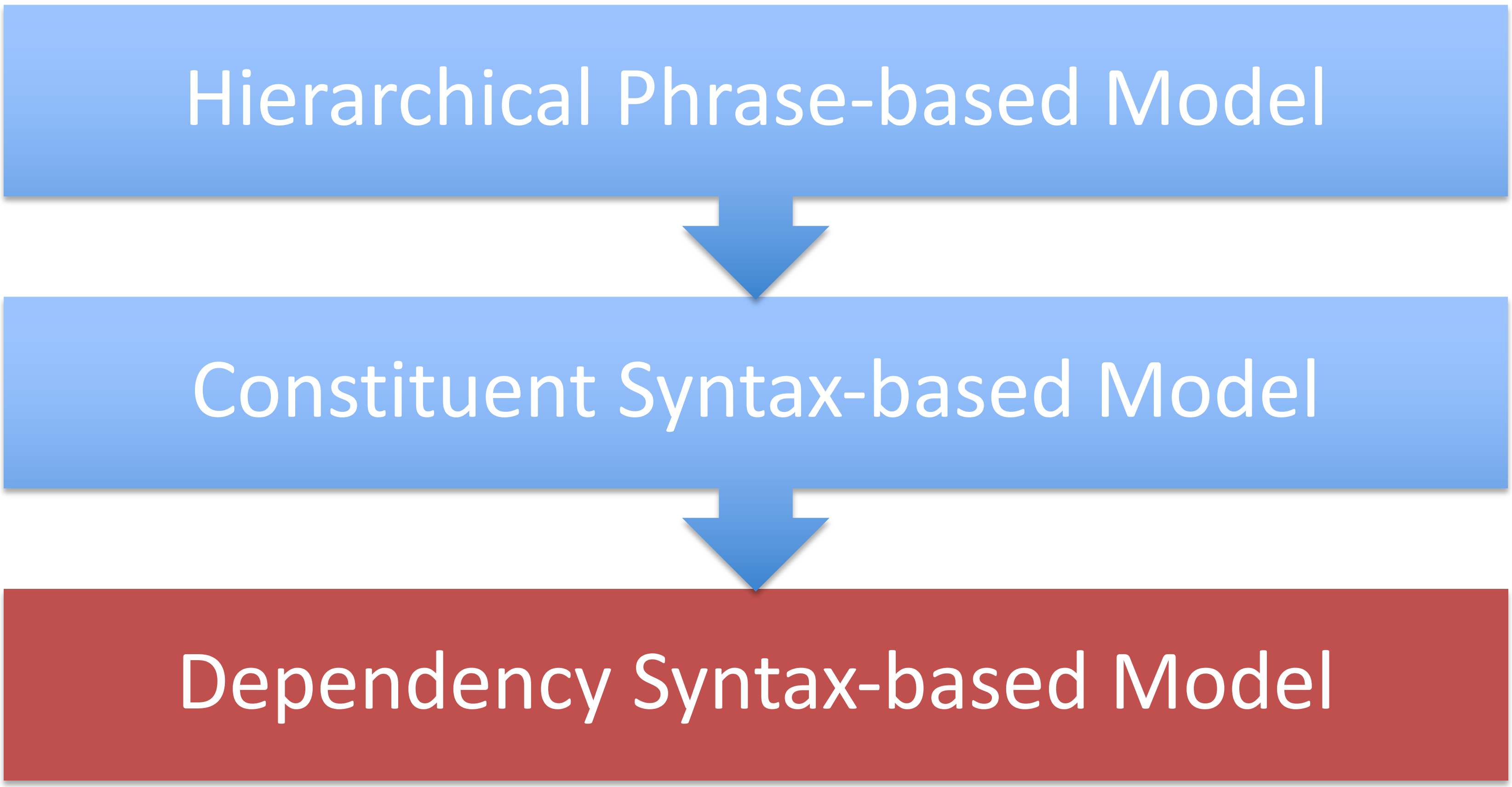
- Ungrammatical phrases
- Syntactic Ambiguity
- Computational Complexity
  - ➢ Synchronous TSG

Is it possible to build a linguistically syntax-based model with the complexity of Synchronous CFG?

Overview of Syntax in SMT

- Introduction to Syntax-based SMT
- **Dependency-to-String Translation**
- Graph-based Translation
- Dependency-based MT Evaluation
- Conclusion and Future Work

Syntax-based Model



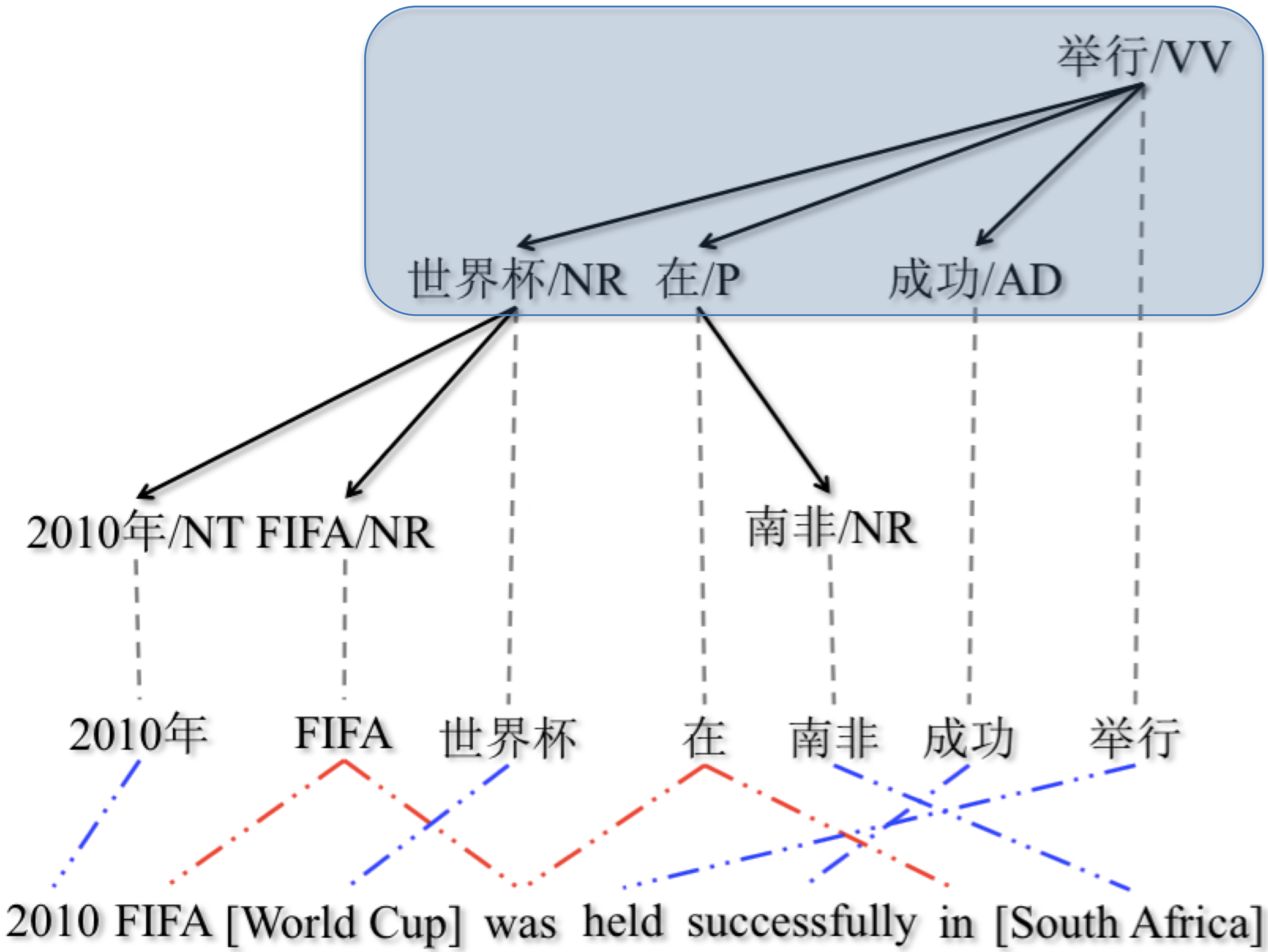
History

Ding Y. et al. 2003, 2004

Quick C. et al. 2005

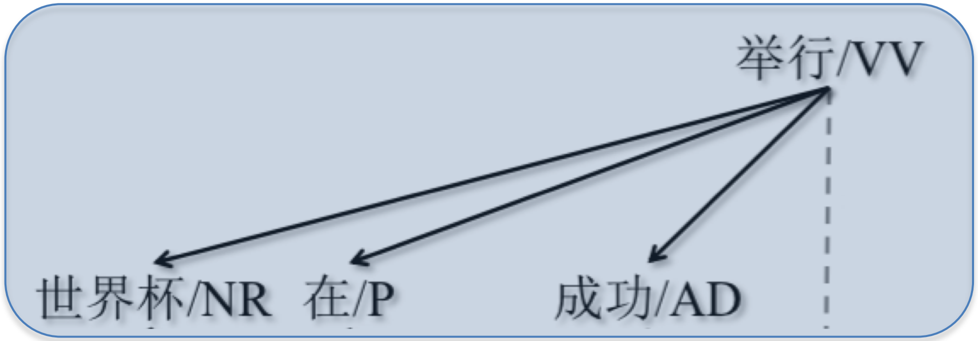
Xiong D. et al. 2007

Difficult of Dependency-based SMT



Difficult of Dependency-based SMT

A dependency translation rule:



…世界杯(World Cup)…在(in)…成功(Successfully) 举行(was held)

Problem: Low Coverage, Sparcity

History

Ding Y. et al. 2003, 2004

Quick C. et al. 2005

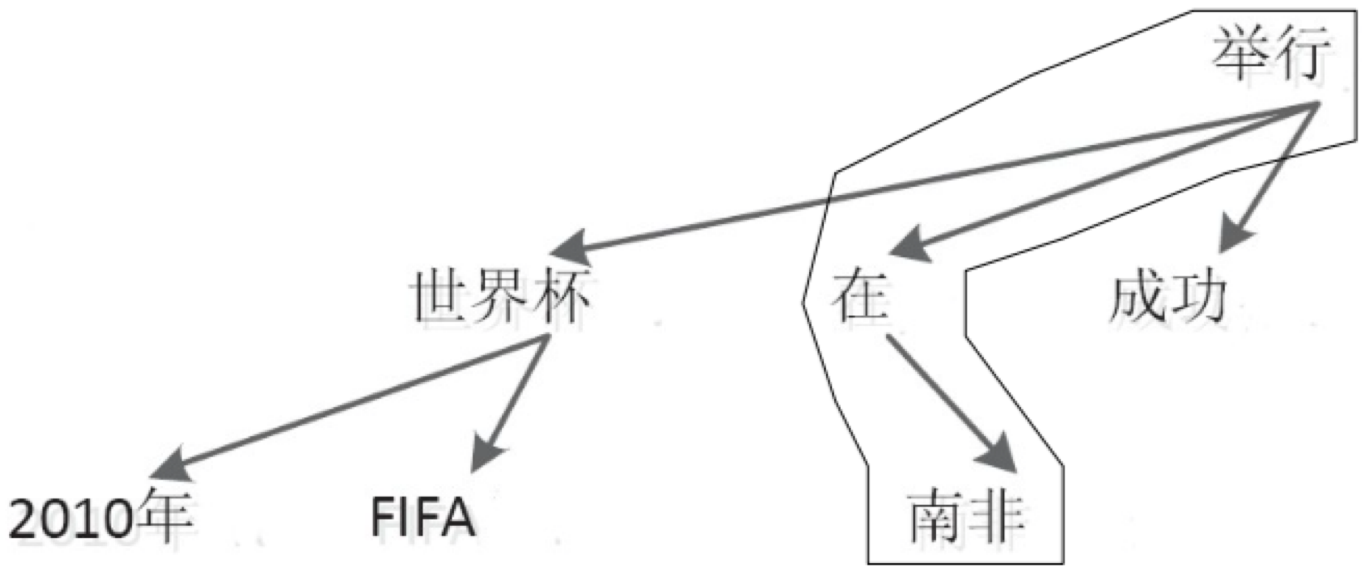
Xiong D. et al. 2007

} Dependency-Treelet-based Approach

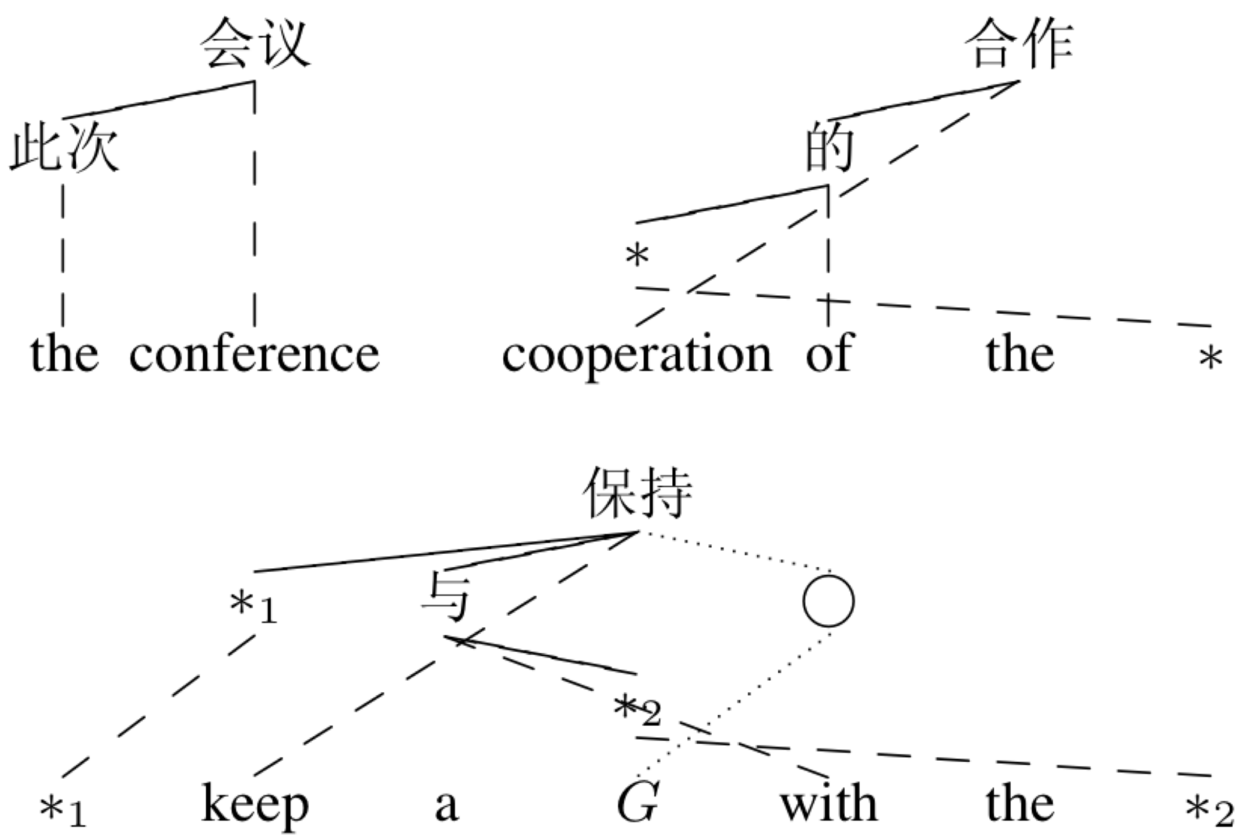
Dependency-Treelet-based Approach

Dependency Treelet:

Any connected subgraph of a dependency tree



Dependency-Treelet-based Rules



Problem of Dep-Treelet-based Approach

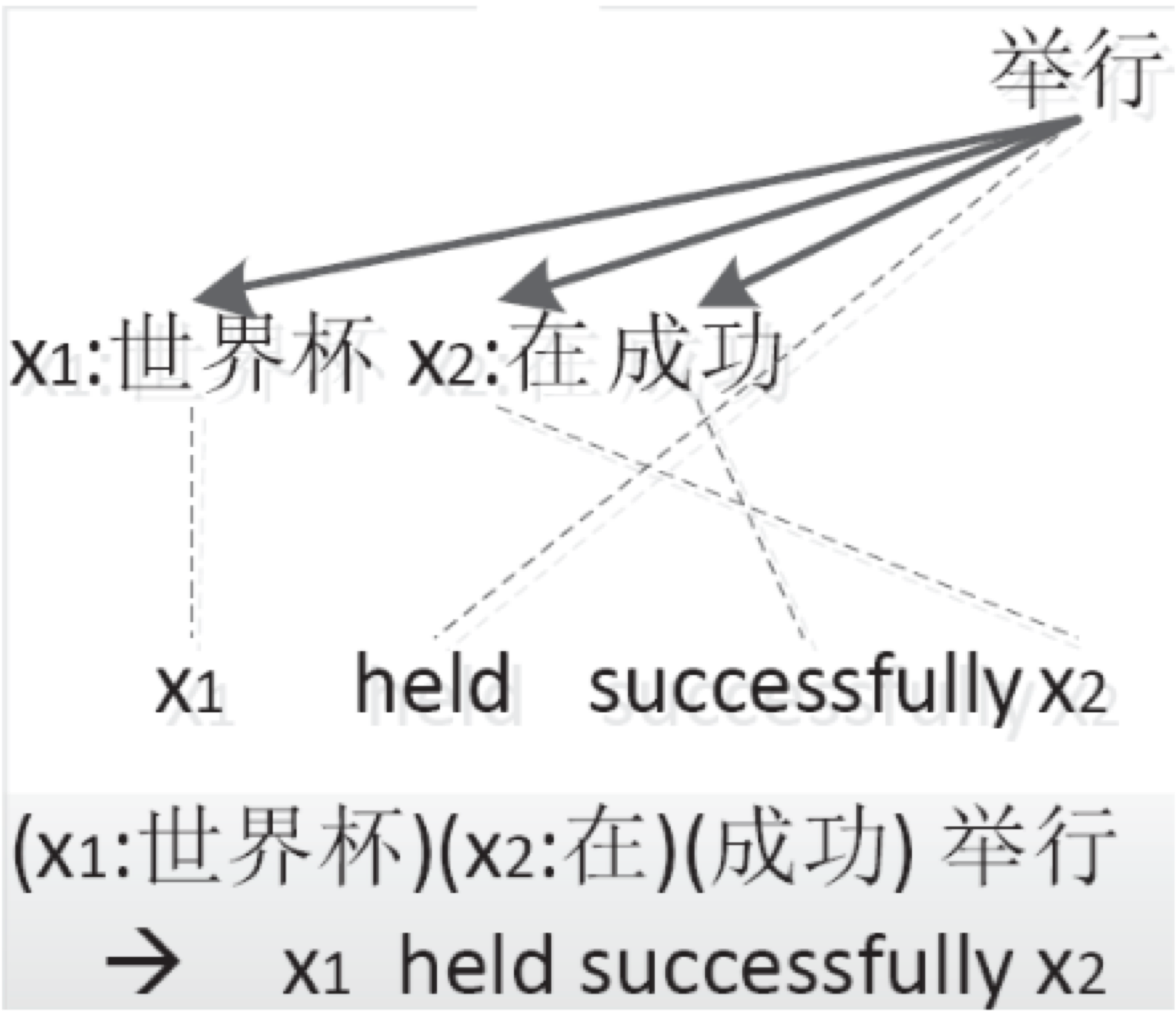
- The partition of a dependency tree to a set of treelets is too flexible (more flexible than the partition of a constituent tree in a tree-to-string model)
- The reordering is difficult in target side:
  - These are no sequential orders between treelets
  - The translation of a treelet is usually non-continuous

Dependency-to-String Model

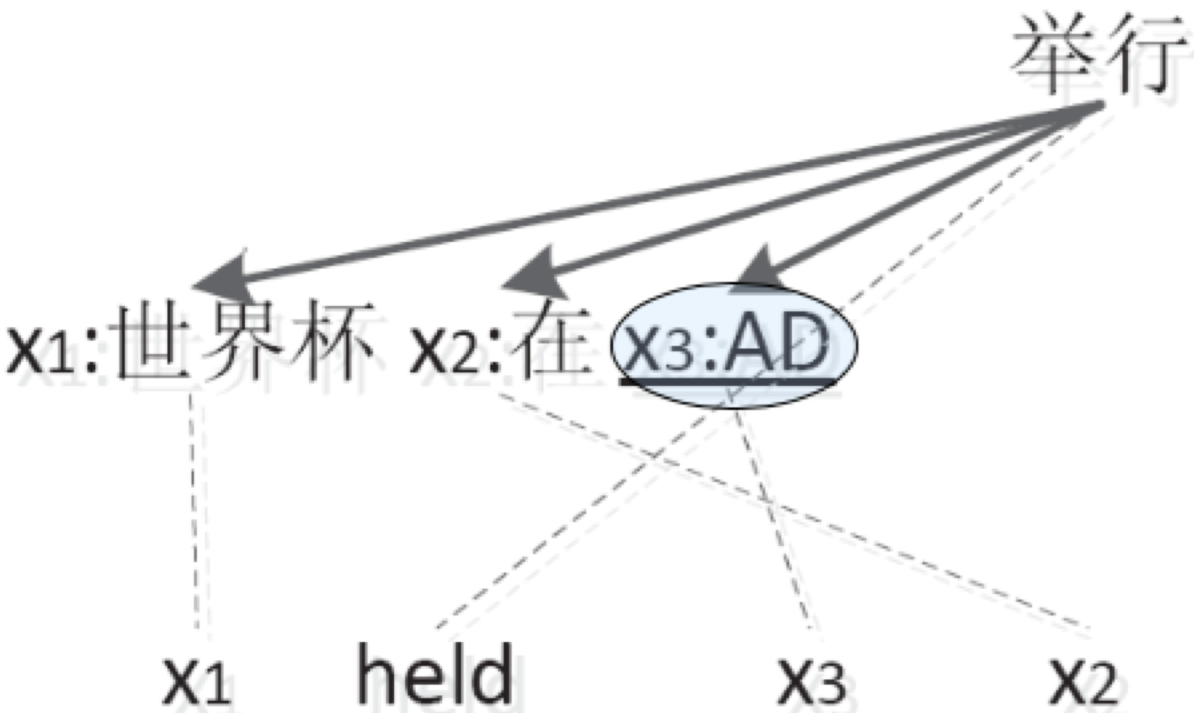
- Our Solution
  - One layer subtree (head-dependency)
  - Using POS for Smoothing

Jun Xie, Haitao Mi and Qun Liu, A novel dependency-to-string model for statistical machine translation, in the Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP2011), pages 216-226, Edinburgh, Scotland, UK. July 27–31, 2011

Dep-to-String Rule



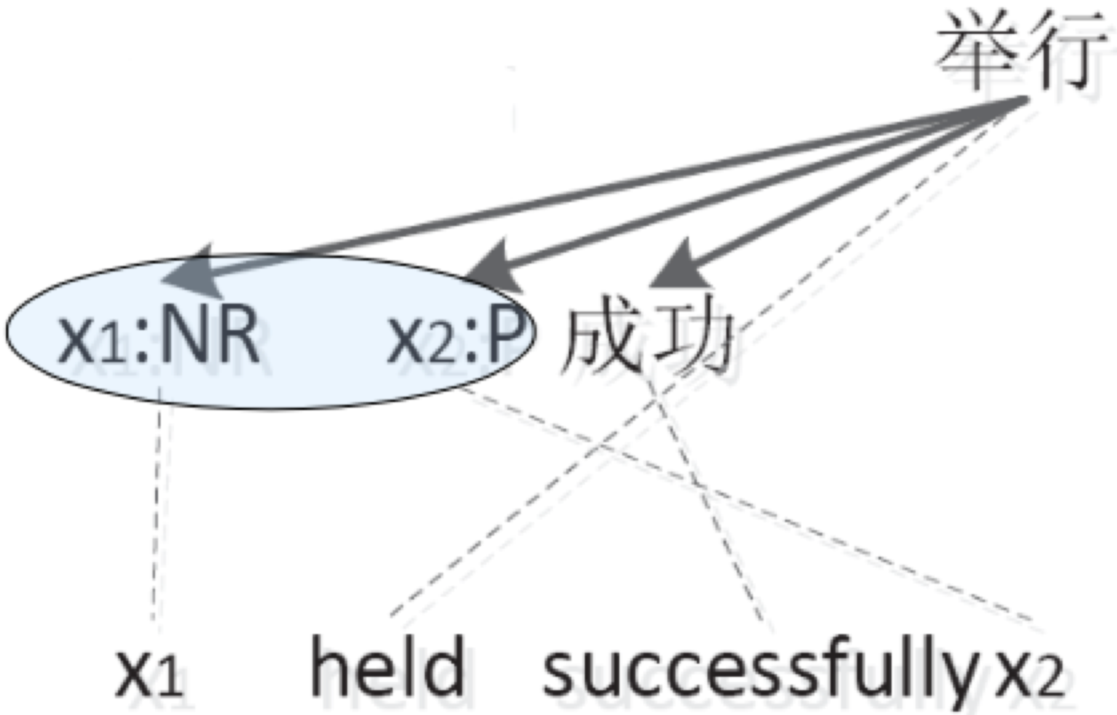
Smoothing with: Leaf nodes



(x_1:世界杯)(x_2:在)(<u>x_3:AD</u>) 举行

→ x_1 held x_3 x_2

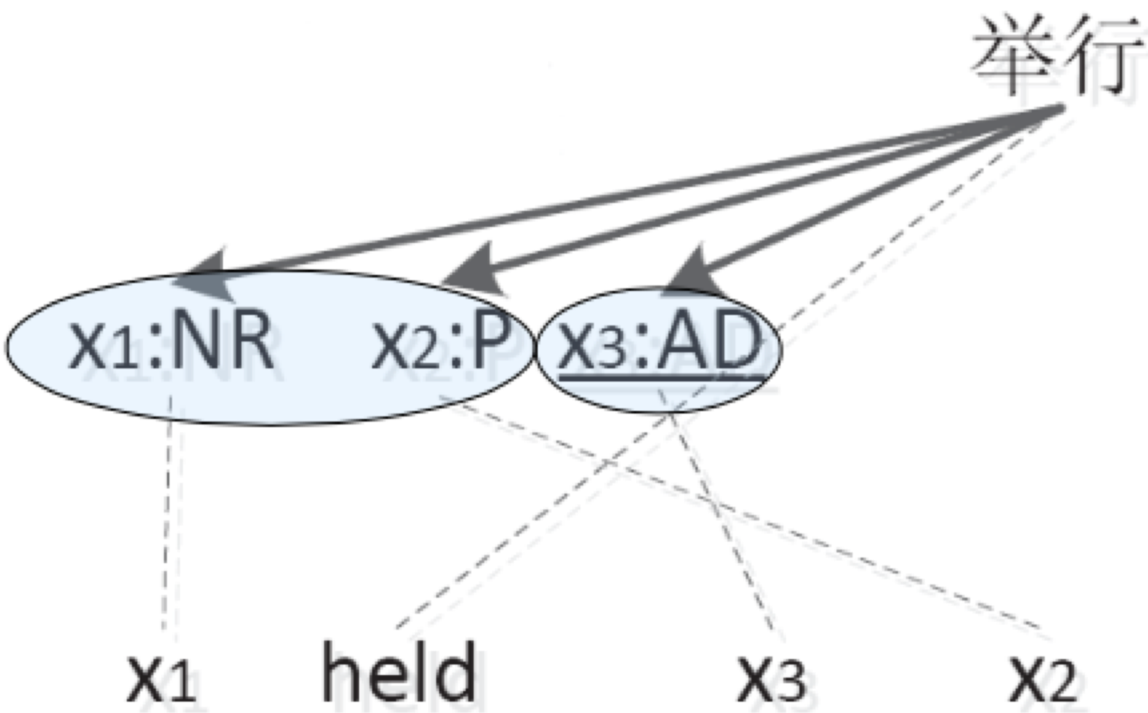
Smoothing with: Internal nodes



(x1:NR)(x2:P)(成功) 举行

→ x1 held successfully x2

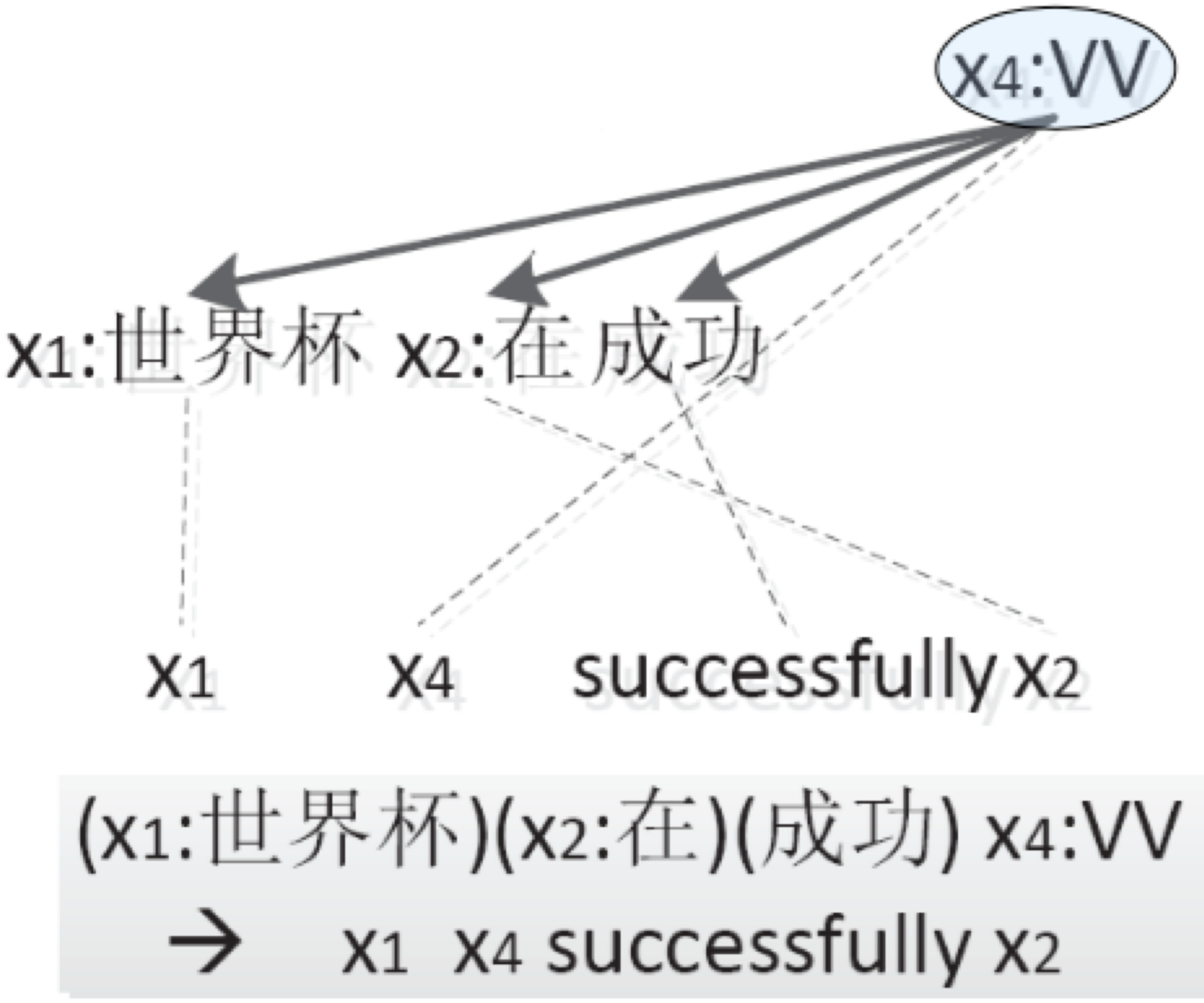
Smoothing with: Leaf & Internal node



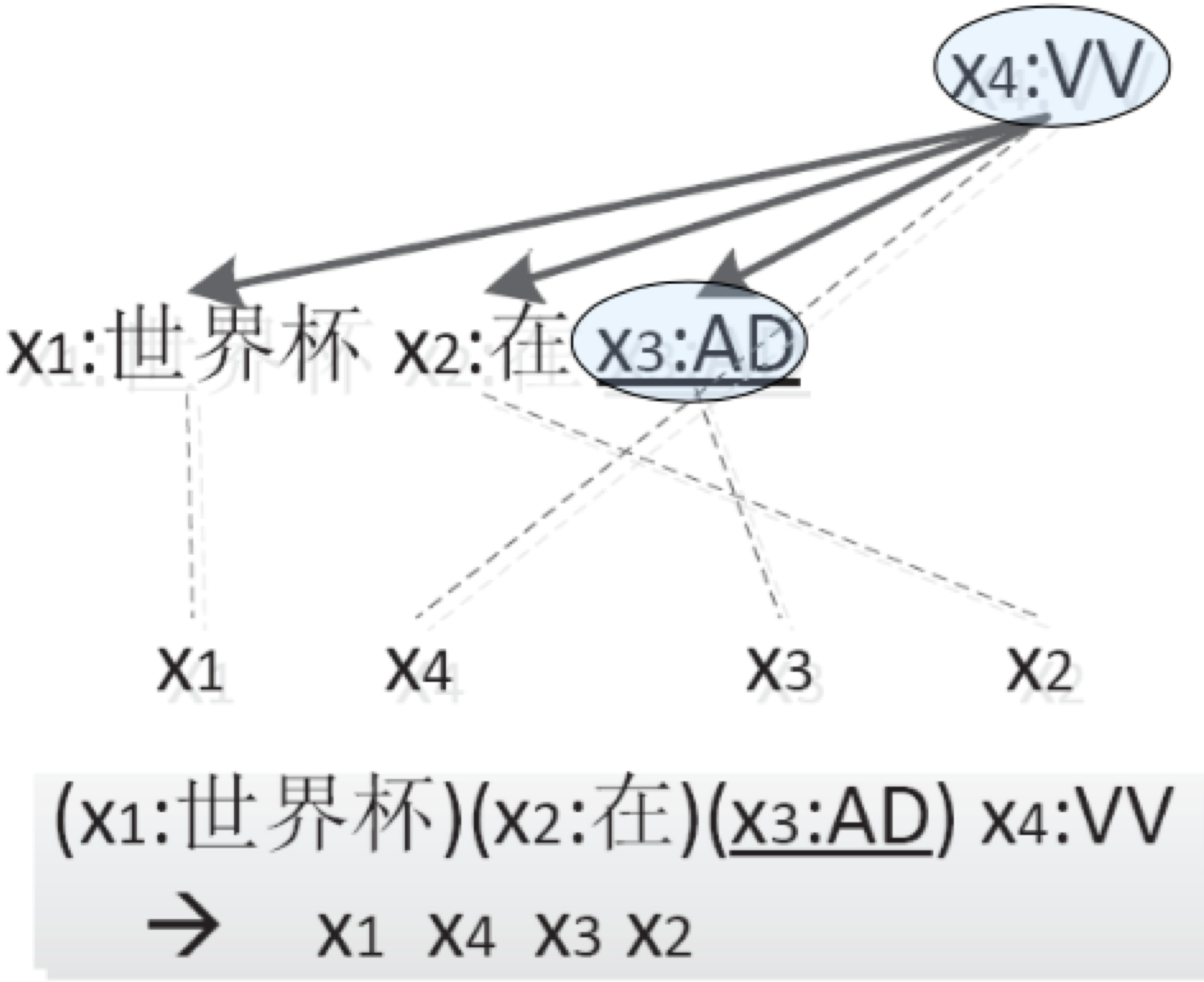
(x1:NR)(x2:P)(x3:AD) 举行

→ x1 held x3 x2

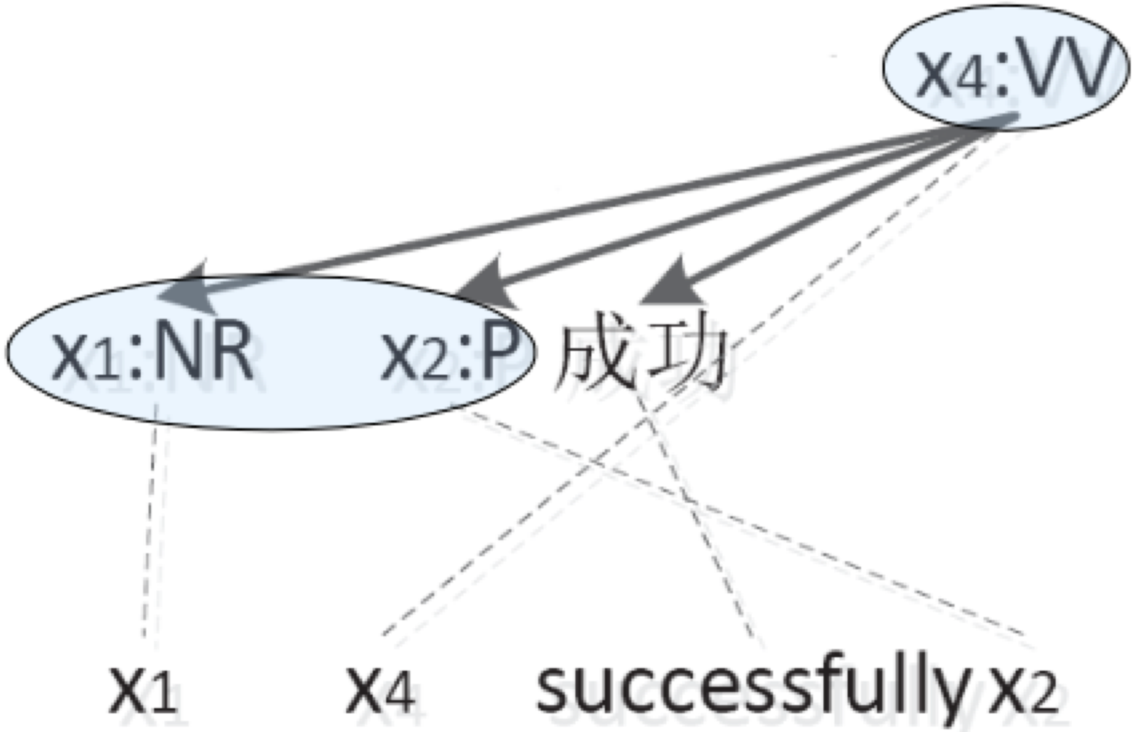
Smoothing with: Head node



Smoothing with: Head & Leaf nodes



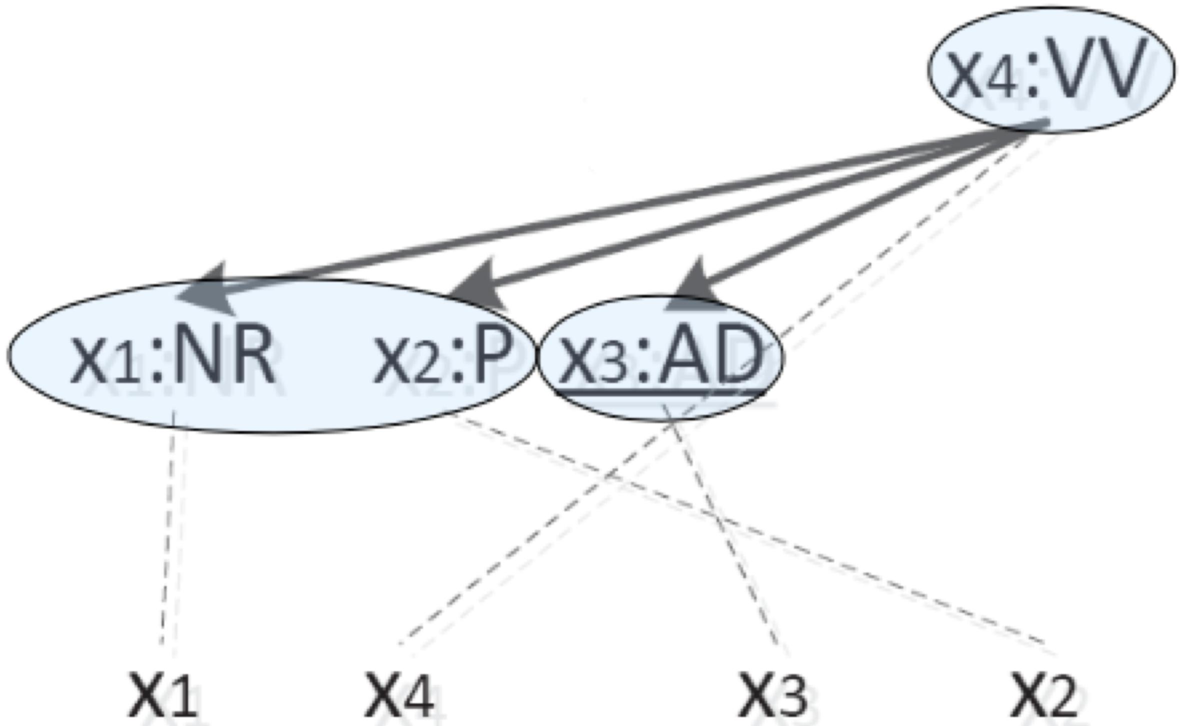
Smoothing with: Head & Internal nodes



(x1:NR)(x2:P)(成功) x4:VV

→ x1 x4 successfully x2

Smoothing with: All nodes



(x1:NR)(x2:P)(x3:AD) x4:VV

→ X1 X4 X3 X2

Experiments

System	Rule #	MT04(%)	MT05(%)
cons2str	30M	34.55	31.94
hiero-re	148M	35.29	33.22
dep2str	56M	**35.82**$^{+}$	**33.62**$^{+}$

Dependency-to-String Model

Advantage:

- Linguistic knowledge used
  - Long distance dependency
- Computational Complexity
  - Equivalent to: Synchronous CFG

Disadvantage:

- Ungrammatical phrases
- Syntactic Ambiguity

Dependency-to-String Model implemented as Synchronous CFG

Liangyou Li, Jun Xie, Andy Way, Qun Liu, Transformation and Decomposition for Efficiently Implementing and Improving Dependency-to-String Model In Moses, In Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Pages 122-131. Doha, Qatar. 2014.

- Implement Dependency-to-String in a Synchronous CFG which is compatible with Moses chart decoder
  - Open Source Tools: dep2str
- Implement pseudo-forest to support partially matched head-dependency structures

Overview of Syntax in SMT

- Introduction to Syntax-based SMT
- Dependency-to-String Translation
- **Graph-based Translation**
- Dependency-based MT Evaluation
- Conclusion and Future Work

Graph-based Translation

- Liangyou Li, Andy Way, Qun Liu, Dependency Graph-to-String Translation, In Proceedings of the EMNLP 2015, pages 33-43, Lisbon, Portugal, 17-21 September 2015.
- Liangyou Li, Andy Way, Qun Liu, Graph-Based Translation Via Graph Segmentation, submitted to ACL2015.

Overview of Syntax in SMT

- Introduction to Syntax-based SMT
- Dependency-to-String Translation
- Graph-based Translation
- **Dependency-based MT Evaluation**
- Conclusion and Future Work

Evaluation for Machine Translation

Candidate 1: It is a guide to action which ensures that the military always obeys the command of the party

Candidate 2: It is to insure the troops forever hearing the activity guidebook that party direct

Reference 1: It is a guide to action that ensures that the military will forever heed party commands

Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the party

Reference 3: It is the practical guide for the army to heed the directions of the party

Question: Given the human translations as references, how to evaluation the machine translation candidates automatically?

Existing MT Evaluation Metrics

- Lexicalized Metrics

  BLEU NIST Rouge WER PER METEOR AMBER

- Syntax-based Metrics

  STM HWCM

- Semantic-based Metrics

  MEANT HMEANT

- Combinational Metrics

  LAYERED DISCOTK

Existing MT Evaluation Metrics

Metrics	知识类型	模型	优点	缺点
基于词汇的方法	词汇	相似度错误率	善于捕捉词汇或短语	不能捕捉句法结构
基于句法的方法	句法信息	相似度	一定程度上捕捉句法信息	机器译文端句法分析正确率不能保证
基于语义的方法	语义信息	相似度	一定程度上捕捉语义信息	SRL准确率不理想 缺乏有效的语义表示方法
集合多种类型知识的方法	词汇 句法 语义	机器学习相似度	兼顾各类型知识性能最好	不适合没有训练语料的情况

RED: A MT Metric based on Reference Dependency

- Using dependency to measure the similarity between the candidate and the reference.
- The dependency similarity is calculated as a balance between head-dependency chain similarity and the float-fix structure similarity
- Dependency parser is applied only on the reference, to avoid the instability result of parsing result on the MT translation (candidate).

- Obtained good results in WMT 2014 metric tasks

- Tuning to RED ranked No.1 in WMT 2015 tuning task on English-Czech Translation

RED: A MT Metric based on Reference Dependency

Yu, H., Wu, X., Xie, J., Jiang, W., Liu, Q., & Lin, S. RED: A Reference Dependency Based MT Evaluation Metric. In COLING 2014, Vol. 14, pp. 2042-2051.

Liangyou Li, Hui Yu, Qun Liu, MT Tuning on RED: A Dependency-Based Evaluation Metric, In WMT 2015, pages 428-433, Lisboa, Portugal, 17-18 September 2015.

Tuning on RED

System Name	TrueSkill Score Tuning-Only	TrueSkill Score All	BLEU
DCU	0.320	-0.342	4.96
BLEU-MIRA-DENSE	0.303	-0.346	5.31
AFRL	0.303	-0.342	5.34
USAAR-TUNA	0.214	-0.373	5.26
BLEU-MERT-DENSE	0.123	-0.406	5.24
METEOR-CMU	-0.271	-0.563	4.37
BLEU-MIRA-SPARSE	-0.992	-0.808	3.79
USAAR-BASELINE-MIRA	—	—	5.31
USAAR-BASELINE-MERT	—	—	5.25

Results of WMT2015 Tuning Task on English-Czech translation

DPMF: Parsing as Evaluation

- We proposed a novel MT Evaluation Metrics based on Dependency Parsing Model
- We use the reference translations as the training corpus to train a parser
- The parser are used to parse the translation candidates
- The score of the parsing model obtained by the translation candidates are regarded as its quality score.

DPMF: Parsing as Evaluation

Hui Yu, Xiaofeng Wu, Wenbin Jiang, Qun Liu, ShouXun Lin, An Automatic Machine Translation Evaluation Metric Based on Dependency Parsing Model, arXiv:1508.01996 [cs.CL], August 2015

Hui Yu, Qingsong Ma, Xiaofeng Wu, Qun Liu, CASICT-DCU Participation in WMT2015 Metrics Task, In WMT 2015, Lisboa, Portugal, 17-18 September 2015.

WMT 2015 Metric Shared Tasks

Correlation coefficient Direction	Pearson Correlation Coefficient fr-en	fi-en	de-en	cs-en	ru-en	Average	Spearman's Average
DPMFcomb	.995 ± .006	.951 ± .013	**.949** ± .016	**.992** ± .004	.871 ± .025	**.952** ± .013	.879 ± .053
RATATOUILLE	.989 ± .010	.899 ± .019	.942 ± .018	.963 ± .008	**.941** ± .018	.947 ± .014	**.905** ± .047
DPMF	**.997** ± .005	.939 ± .015	.929 ± .019	.986 ± .005	.868 ± .026	.944 ± .014	.867 ± .050
METEOR-WSD	.982 ± .011	.944 ± .014	.914 ± .021	.981 ± .006	.857 ± .026	.936 ± .016	.797 ± .062
chrF3	.979 ± .012	.893 ± .020	.921 ± .020	.969 ± .007	.915 ± .023	.935 ± .016	.834 ± .068
BEER_Treepel	.981 ± .011	**.957** ± .013	.905 ± .021	.985 ± .005	.846 ± .027	.935 ± .016	.827 ± .064
BEER	.979 ± .012	.952 ± .013	.903 ± .022	.975 ± .006	.848 ± .027	.931 ± .016	.828 ± .061
chrF	.997 ± .005	.942 ± .015	.884 ± .024	.982 ± .006	.830 ± .029	.927 ± .016	.877 ± .051
LeBLEU-optimized	.989 ± .009	.895 ± .020	.856 ± .025	.970 ± .007	.918 ± .023	.925 ± .017	.857 ± .055
LeBLEU-default	.960 ± .015	.895 ± .020	.856 ± .025	.946 ± .010	.912 ± .022	.914 ± .018	.813 ± .071
BS	−.991 ± .008	−.904 ± .019	−.800 ± .029	−.961 ± .008	−.569 ± .042	−.845 ± .021	−.758 ± .054
USAAR-ZWICKEL-METEOR-MEDIAN	n/a	.934 ± .016	.935 ± .019	.973 ± .007	.891 ± .024	.933 ± .016	.849 ± .044
USAAR-ZWICKEL-METEOR-MEAN	n/a	.945 ± .014	.921 ± .020	.982 ± .006	.866 ± .026	.929 ± .016	.833 ± .041
USAAR-ZWICKEL-METEOR-ARIGEO	n/a	.945 ± .014	.921 ± .020	.982 ± .006	.866 ± .026	.929 ± .016	.833 ± .041
USAAR-ZWICKEL-METEOR-RMS	n/a	.949 ± .014	.895 ± .023	.982 ± .006	.815 ± .030	.910 ± .018	.821 ± .039
USAAR-ZWICKEL-COMET-RMS	n/a	.834 ± .023	.847 ± .027	.869 ± .014	.603 ± .041	.788 ± .026	.665 ± .069
USAAR-ZWICKEL-COMET-ARIGEO	n/a	.805 ± .025	.811 ± .030	.837 ± .016	.626 ± .040	.769 ± .028	.684 ± .063
USAAR-ZWICKEL-COMET-MEAN	n/a	.805 ± .025	.811 ± .030	.837 ± .016	.626 ± .040	.769 ± .028	.684 ± .063
USAAR-ZWICKEL-METEOR-HARMONIC	n/a	.542 ± .034	.553 ± .046	.712 ± .021	.407 ± .047	.554 ± .037	.770 ± .059
USAAR-ZWICKEL-COMET-HARMONIC	n/a	.463 ± .036	.511 ± .047	.614 ± .024	.406 ± .047	.498 ± .038	.596 ± .068
USAAR-ZWICKEL-COMET-MEDIAN	n/a	−.116 ± .044	.230 ± .051	.644 ± .025	.183 ± .054	.235 ± .043	.209 ± .092
PARMESAN	n/a	−.219 ± .043	.437 ± .047	.328 ± .035	.105 ± .055	.163 ± .045	.071 ± .080
USAAR-ZWICKEL-COSINE2METEOR-MEDIAN	n/a	−.236 ± .042	.014 ± .051	.509 ± .028	.102 ± .055	.097 ± .044	.048 ± .091
USAAR-ZWICKEL-COSINE2METEOR-MEAN	n/a	−.115 ± .044	−.337 ± .049	.450 ± .029	.318 ± .051	.079 ± .043	.086 ± .095
USAAR-ZWICKEL-COSINE2METEOR-ARIGEO	n/a	−.115 ± .044	−.337 ± .049	.450 ± .029	.318 ± .051	.079 ± .043	.086 ± .095
USAAR-ZWICKEL-COSINE2METEOR-RMS	n/a	−.093 ± .043	−.286 ± .052	.406 ± .031	.264 ± .052	.073 ± .045	.066 ± .087
USAAR-ZWICKEL-COSINE-MEDIAN	n/a	−.409 ± .039	−.502 ± .046	.817 ± .019	.072 ± .052	−.006 ± .039	−.082 ± .092
USAAR-ZWICKEL-COSINE2METEOR-HARMONIC	n/a	−.355 ± .040	−.117 ± .052	−.090 ± .033	.280 ± .053	−.070 ± .045	.099 ± .092
USAAR-ZWICKEL-COSINE-RMS	n/a	nan	.008 ± .052	.912 ± .013	nan	nan	.122 ± .079
USAAR-ZWICKEL-COSINE-MEAN	n/a	nan	−.048 ± .052	.908 ± .014	nan	nan	.111 ± .080
USAAR-ZWICKEL-COSINE-HARMONIC	n/a	nan	−.159 ± .052	.900 ± .014	nan	nan	.034 ± .077

Table 1: System-level correlations of automatic evaluation metrics and the official WMT human scores when translating into English.

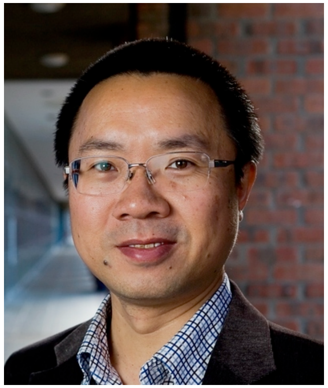
Overview of Syntax in SMT

- Introduction to Syntax-based SMT
- Dependency-to-String Translation
- Graph-based Translation
- Dependency-based MT Evaluation
- **Conclusion and Future Work**

Conclusion & Future Work

- Conclusion: recent work on syntax-based MT
  - Dependency-to-String Translation
  - Dependency-Graph-to-String Translation
  - Graph-based Translation by Graph Segmentation
  - Red: Dependency-based MT Metrics
  - Tuning on Red
  - DPMF: MT Evaluation by Parsing
- Future Work
  - Graph-based Translation by Graph Grammar
  - Graph-based Translation with Rich Linguistic Features

Q&A



Qun Liu, Professor, Dr.,PI

ADAPT Centre
Dublin City University

Institute of Computing Technology
Chinese Academy of Sciences

Email: qun.liu@dcu.ie