

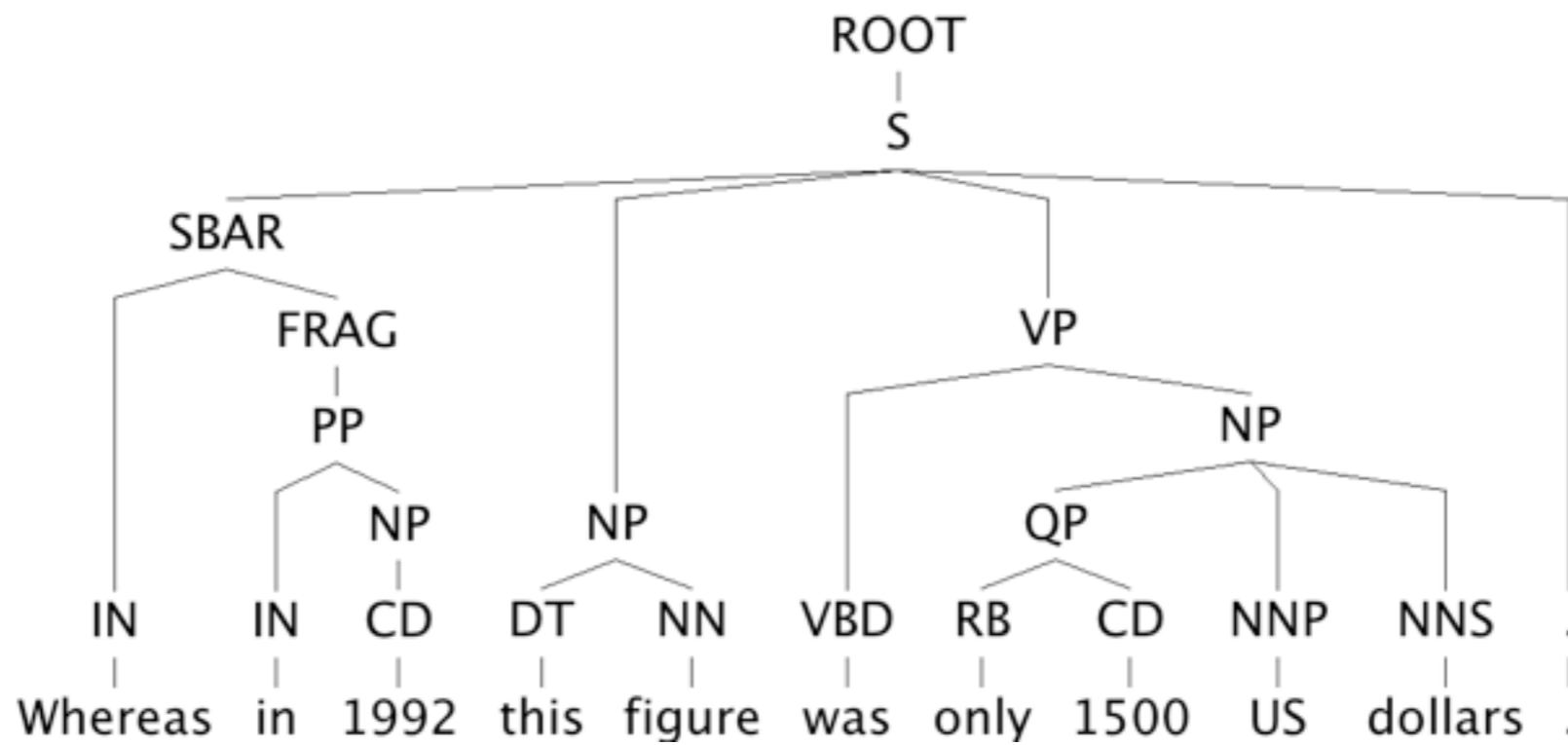
Two Languages are Better than One

(for Syntactic Parsing)

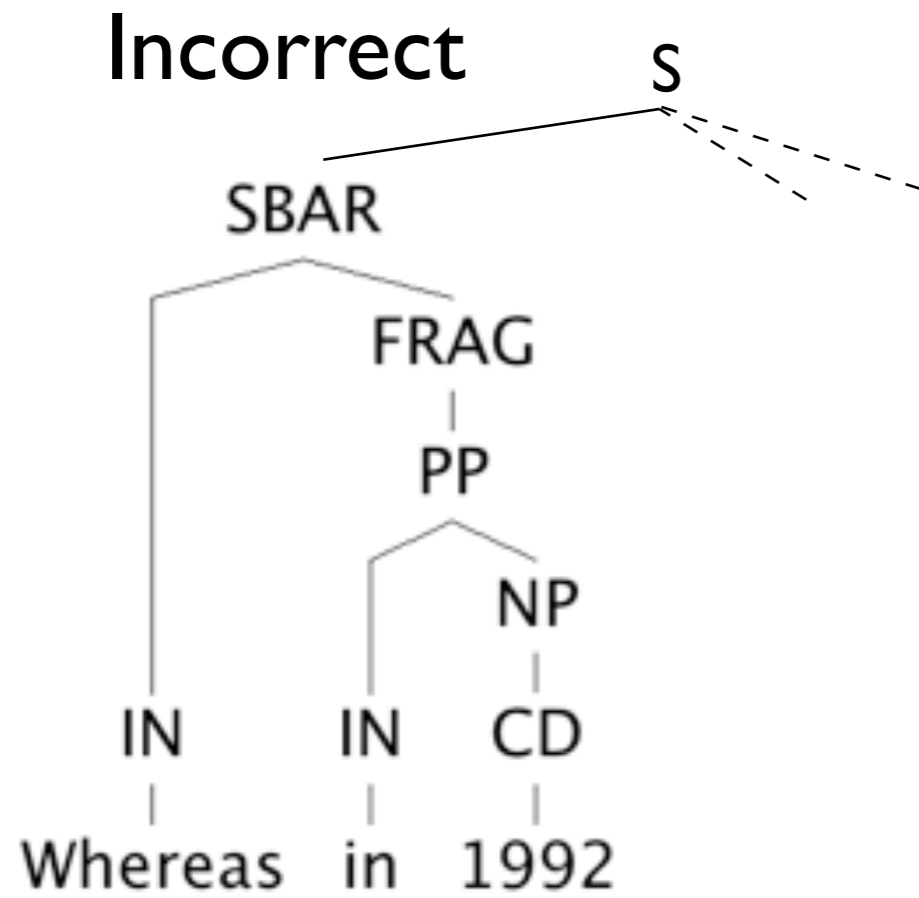


David Burkett and Dan Klein
{dburkett, klein}@cs.berkeley.edu

Motivating Example

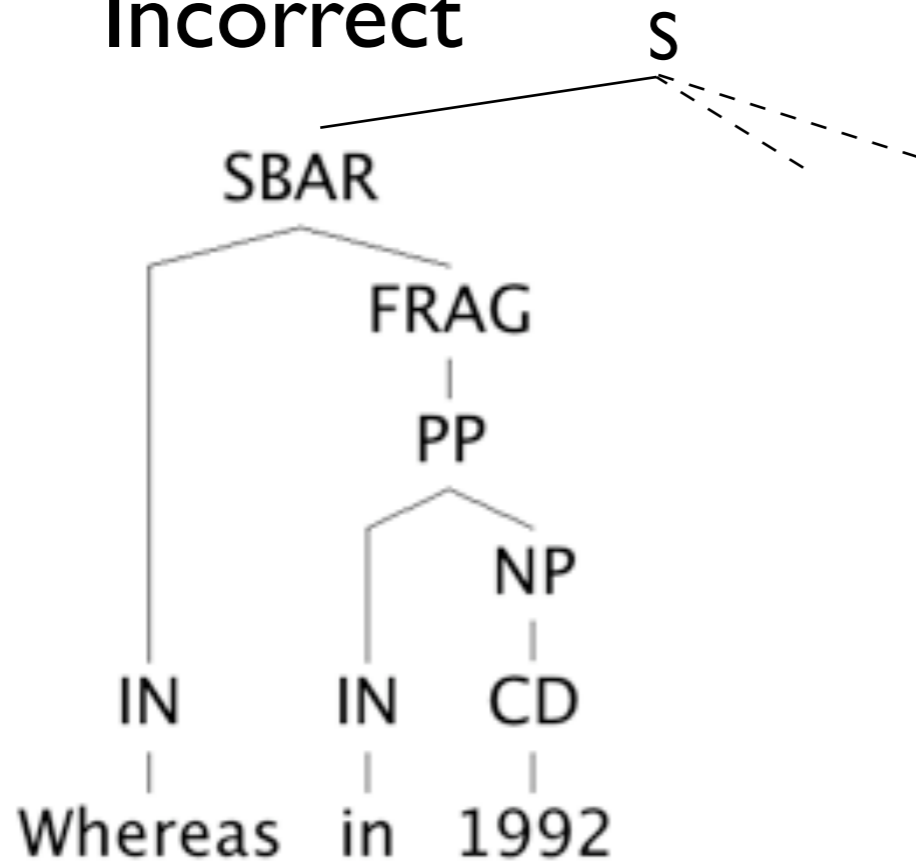


Motivating Example

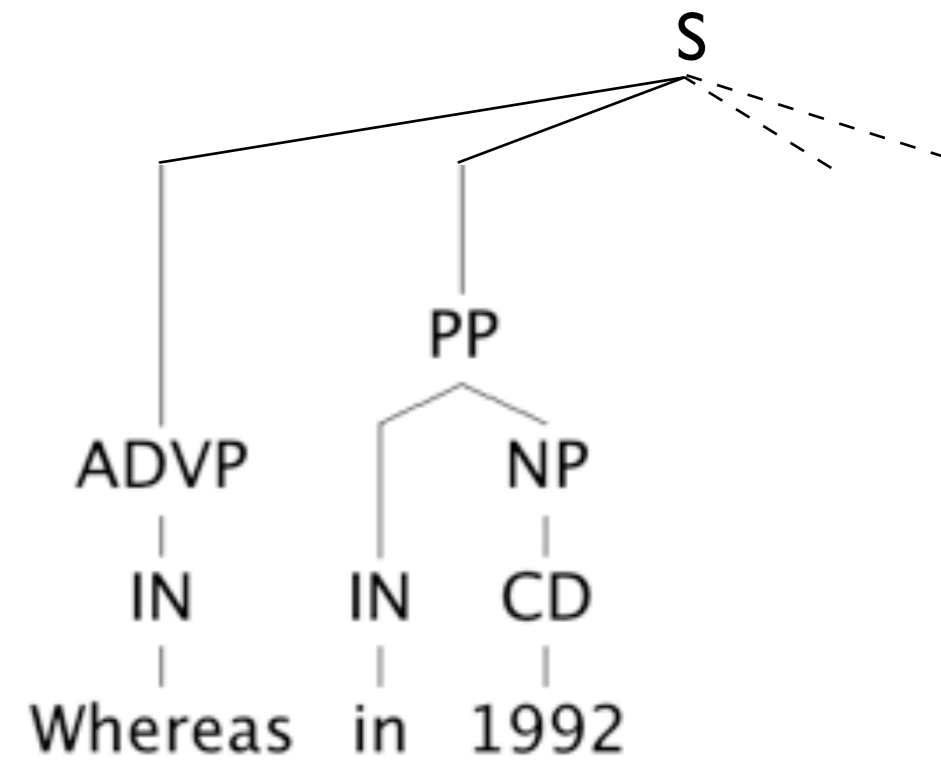


Motivating Example

Incorrect

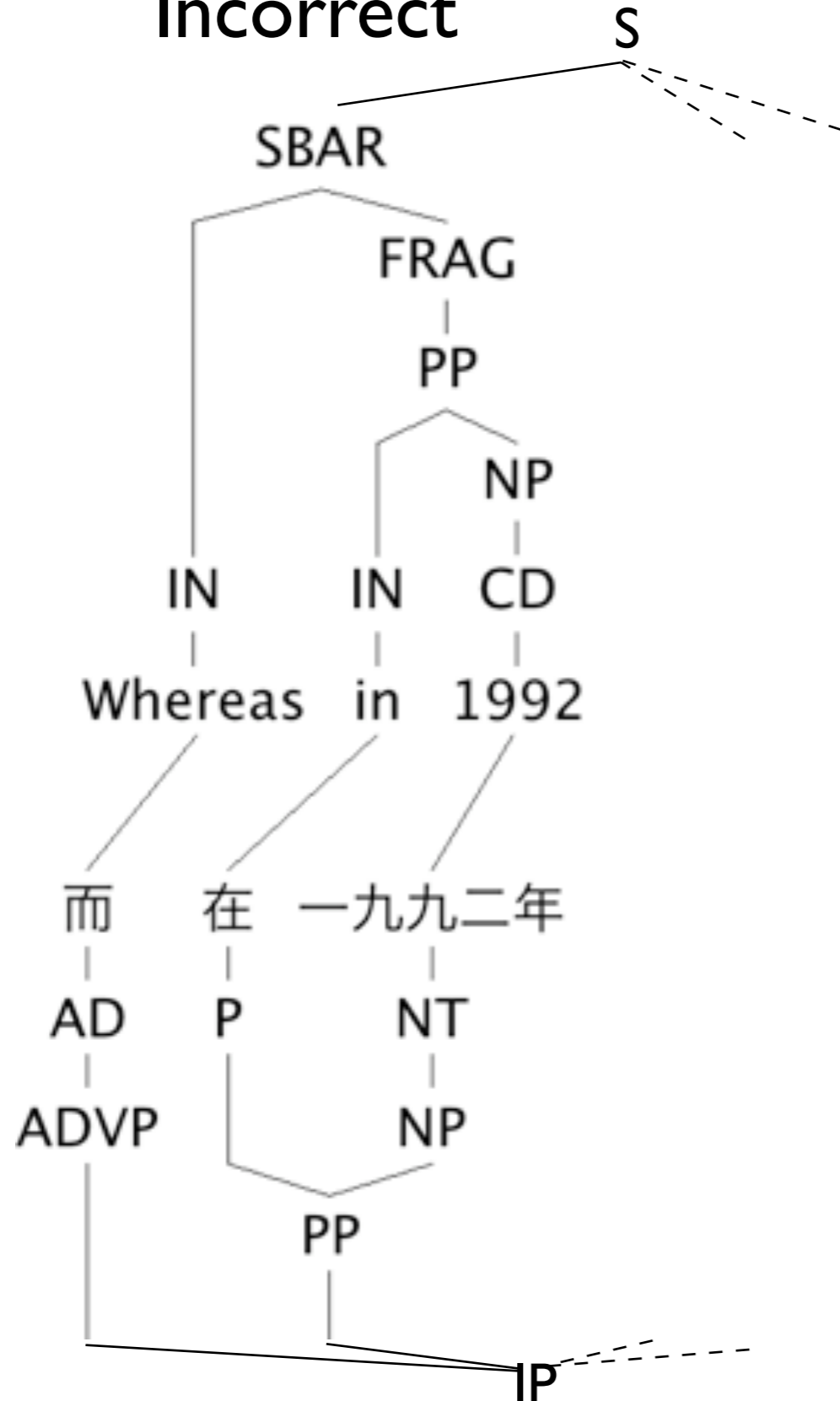


Correct

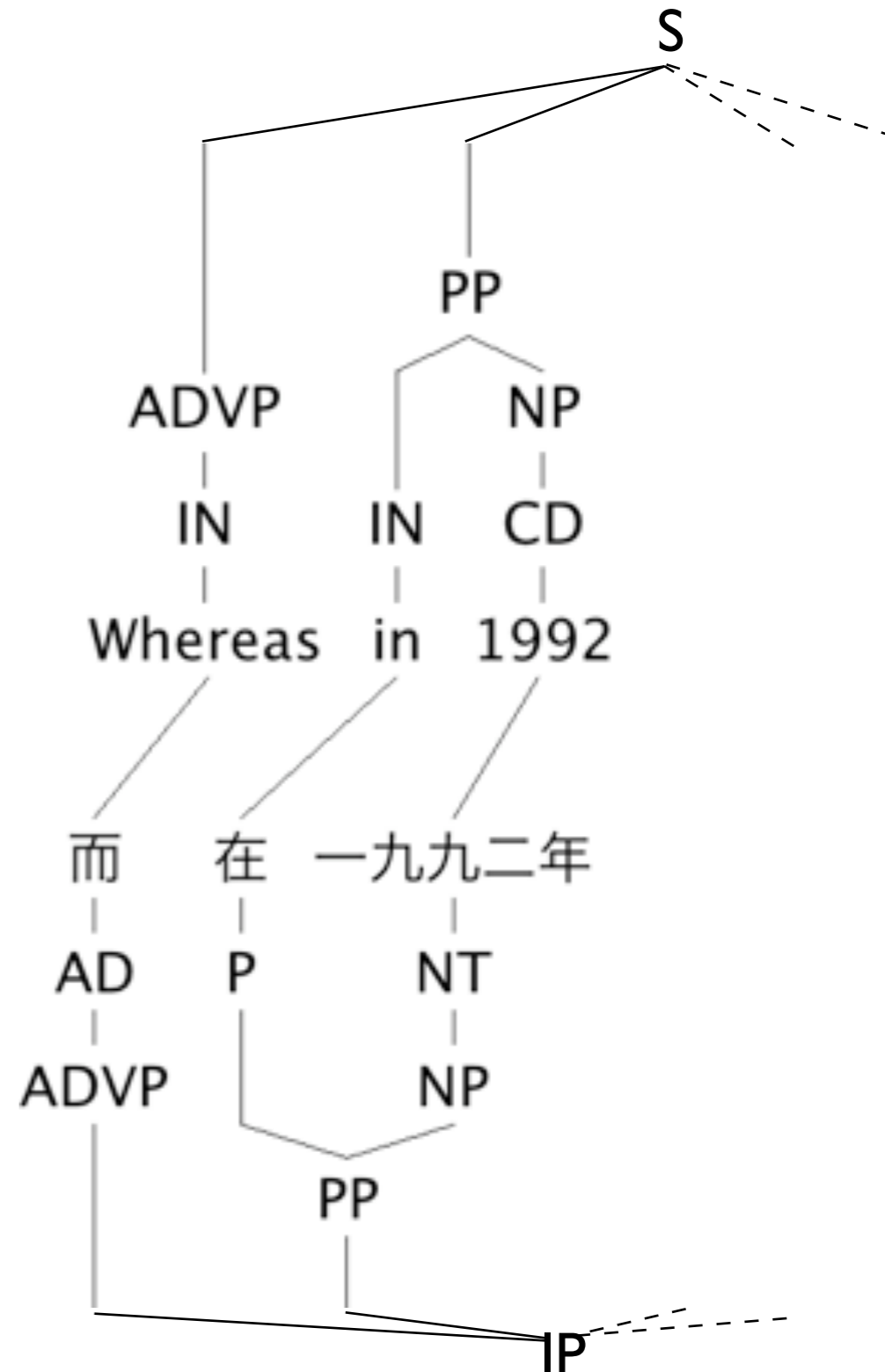


Motivating Example

Incorrect



Correct





Joint Parsing

Bitext

Joint Parsing

Target Sentence

Whereas in 1992

Bitext

而在一九九二

Source Sentence

Joint Parsing

Target Sentence

Whereas in 1992

Target
Parser

Bitext

而在一九九二

Source Sentence

Source
Parser

Joint Parsing

Target Parse Trees



Target Sentence

Whereas in 1992

Target
Parser

Bitext

而在一九九二

Source Sentence

Source
Parser

Source Parse Trees



Joint Parsing

Target Parse Trees



Target Sentence

Target
Parser

Whereas in 1992

Bitext

Word
Aligner

而在一九九二

Source Sentence

Source
Parser

Source Parse Trees



Joint Parsing

Target Parse Trees



Target Sentence

Whereas in 1992

Target
Parser

Bitext

Word
Aligner

Wh	in	1992	
area			而
			在
			一九
			二

而 在 一 九 九 二

Word Alignment

Source Sentence

Source
Parser



Source Parse Trees

Joint Parsing

Target Parse Trees



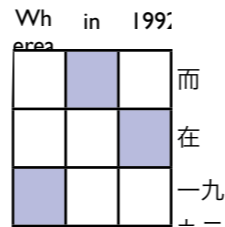
Target Sentence

Whereas in 1992

Target
Parser

Bitext

Word
Aligner



Bilingual
Reranker

而在一九九二

Word Alignment

Source Sentence

Source
Parser



Source Parse Trees

Joint Parsing

Target Parse Trees



Target Sentence

Whereas in 1992

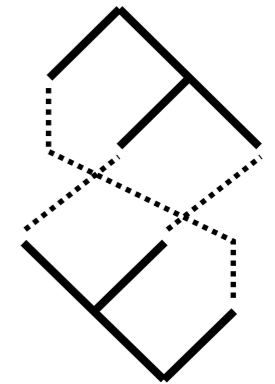
Target Parser

Bitext

Word Aligner

Whereas	in	1992	
			而
			在
			一九
			九二

Bilingual Reranker



而在一九九二

Word Alignment

Source Sentence

Source Parser

Jointly Parsed Trees



Source Parse Trees

Overview

- A log-linear model over aligned parse trees
- Training with latent tree alignments
- Improvements from joint parsing



Model

Model

Sentences

(s, s')

High levels of product and project

产品

、

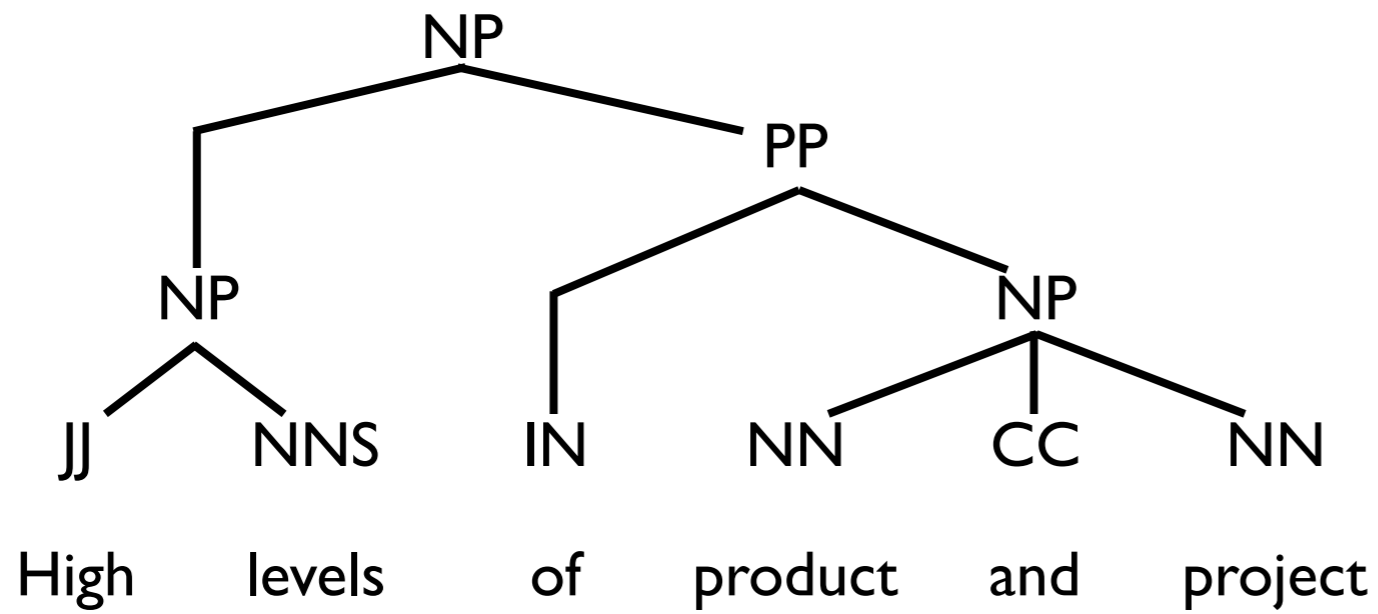
项目

水平

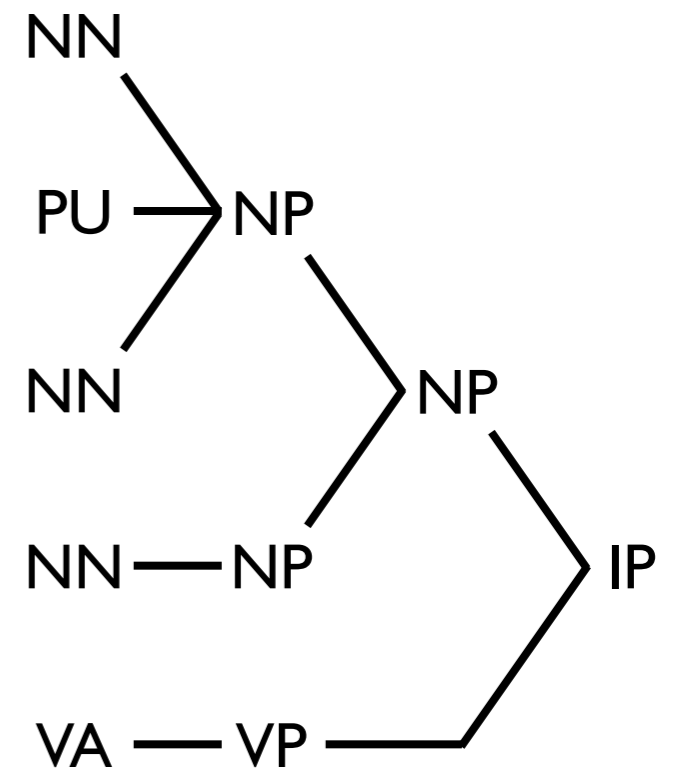
高

Model

Trees *(t, t')*



产品
、
项目
水平
高

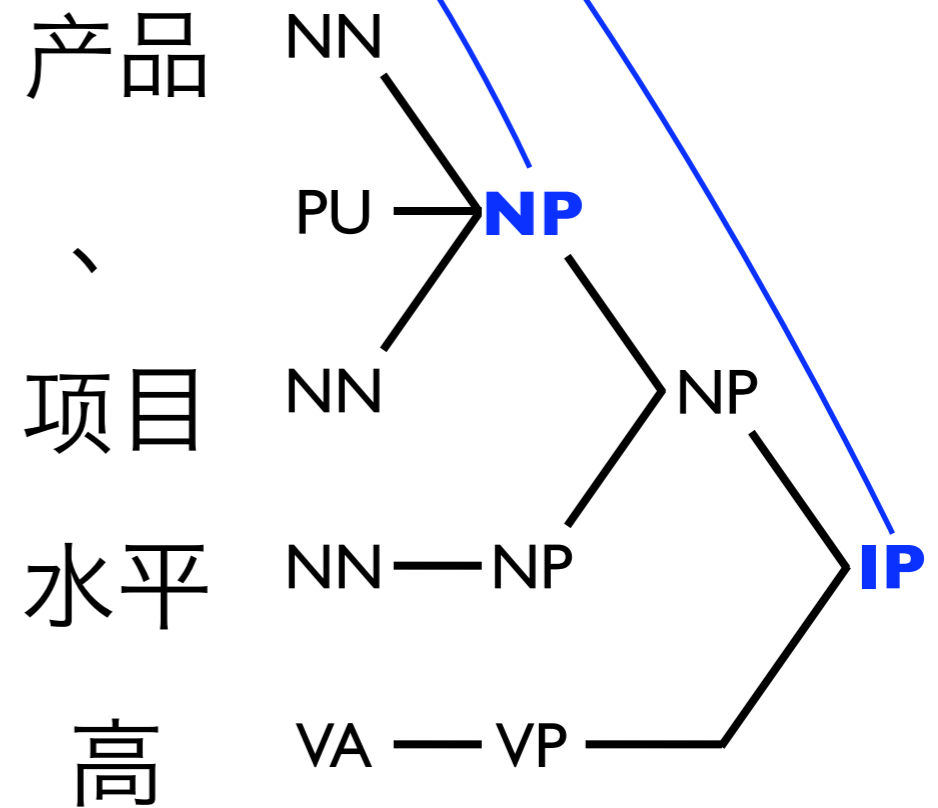
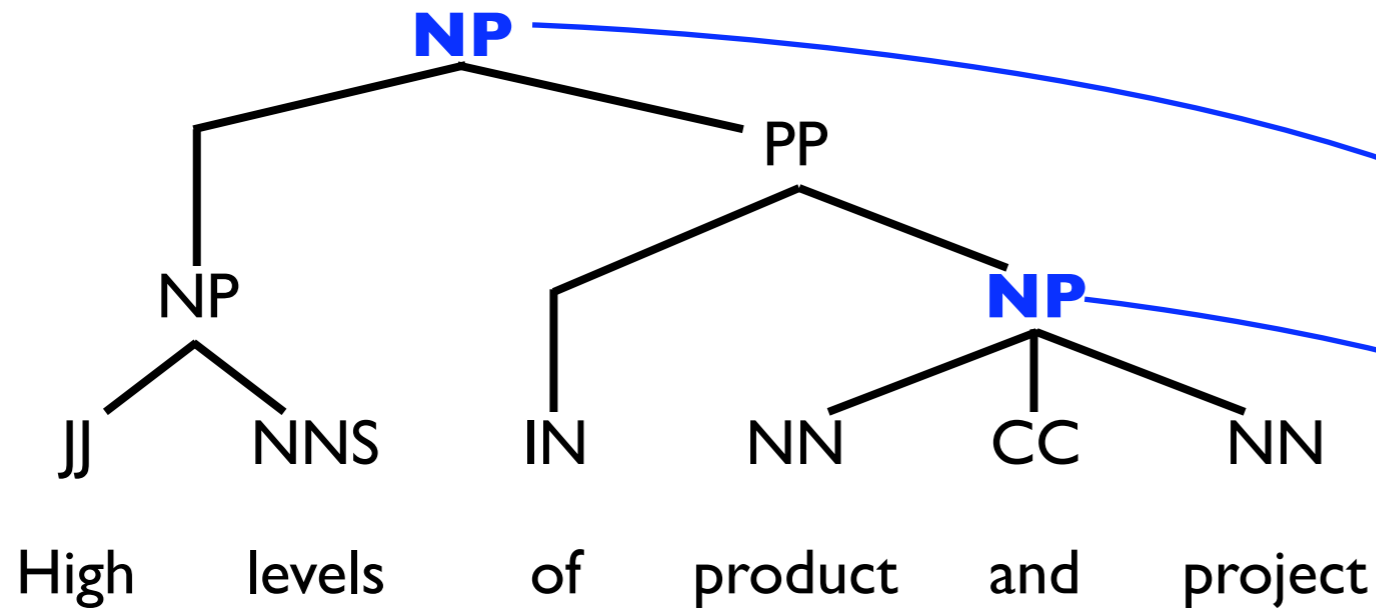


Model

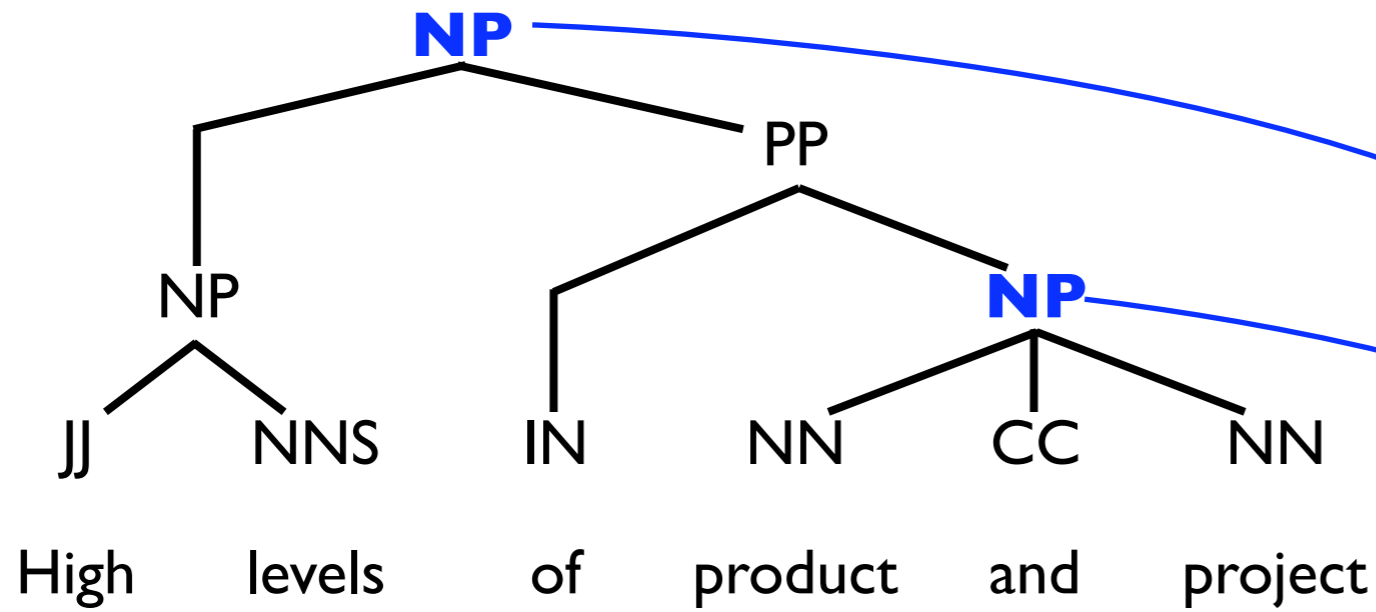
Alignments

a

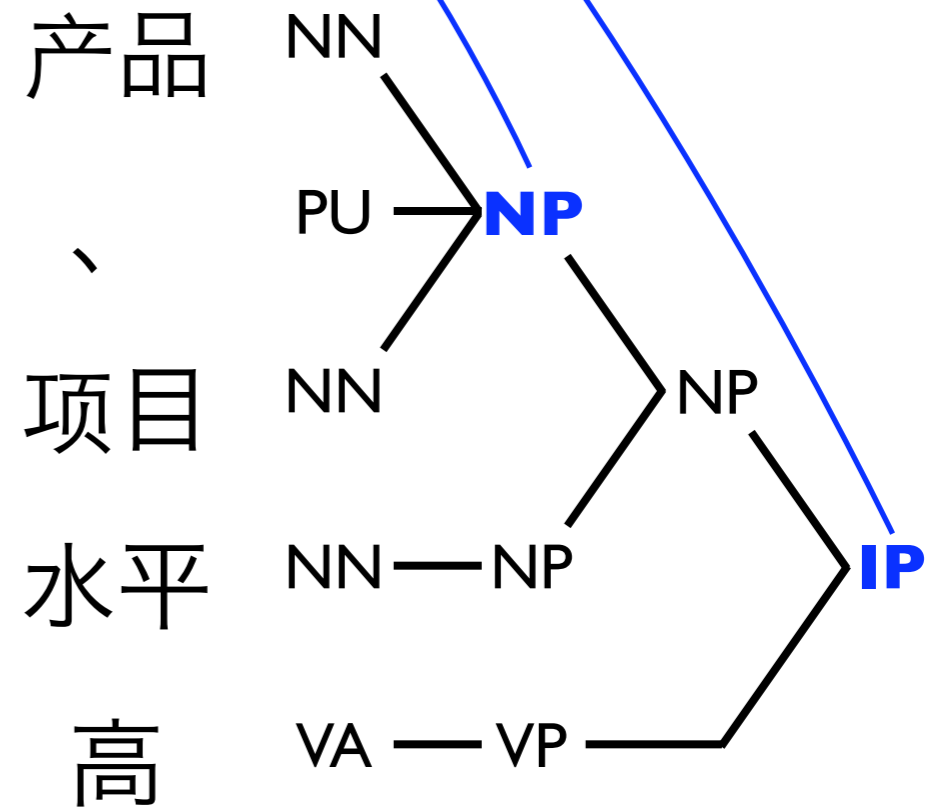
at most 1-1 matchings between tree nodes



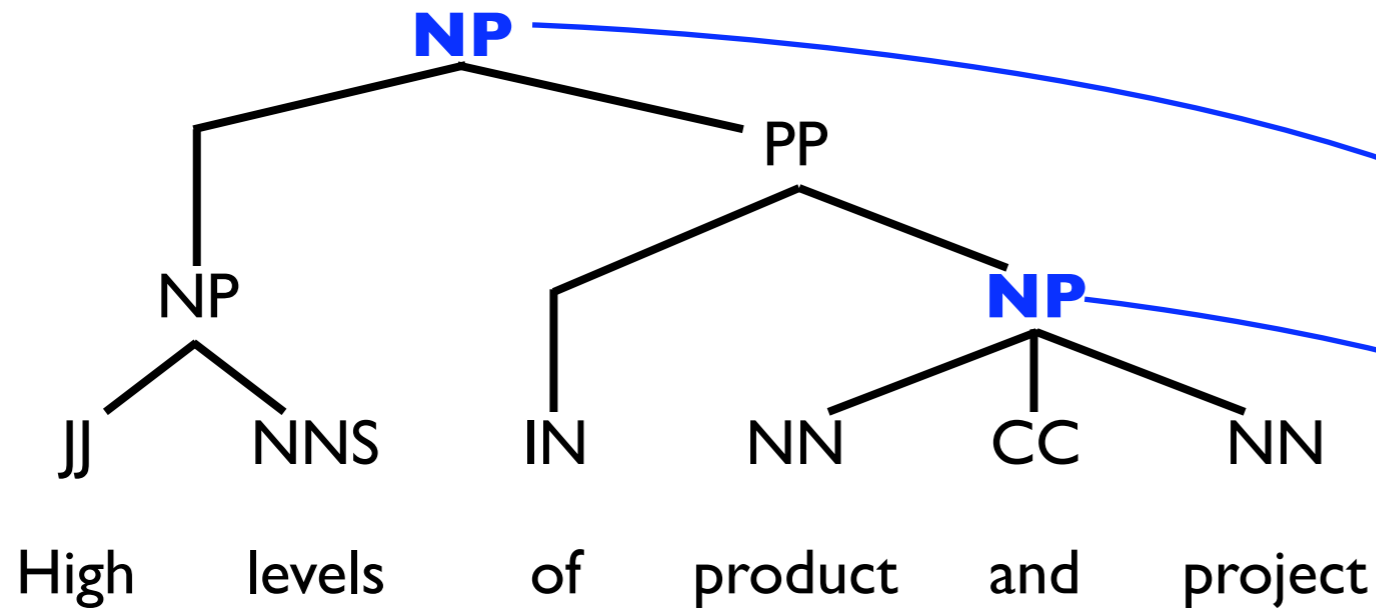
Model



Feature function: $\phi(t, a, t')$



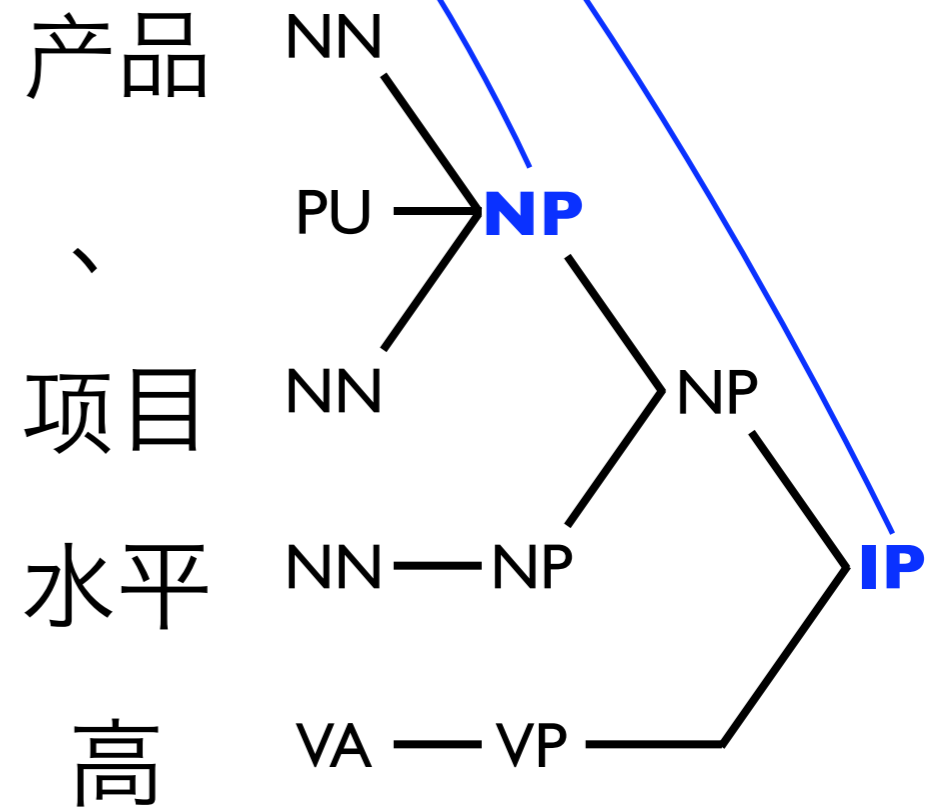
Model



Feature function: $\phi(t, a, t')$

Log-linear model:

$$P(t, a, t' | s, s') \propto \exp(w^\top \phi(t, a, t'))$$





Baseline Features

Baseline Features

Source tree log likelihood: $\log P(t|s)$

Target tree log likelihood: $\log P(t'|s')$

Baseline Features

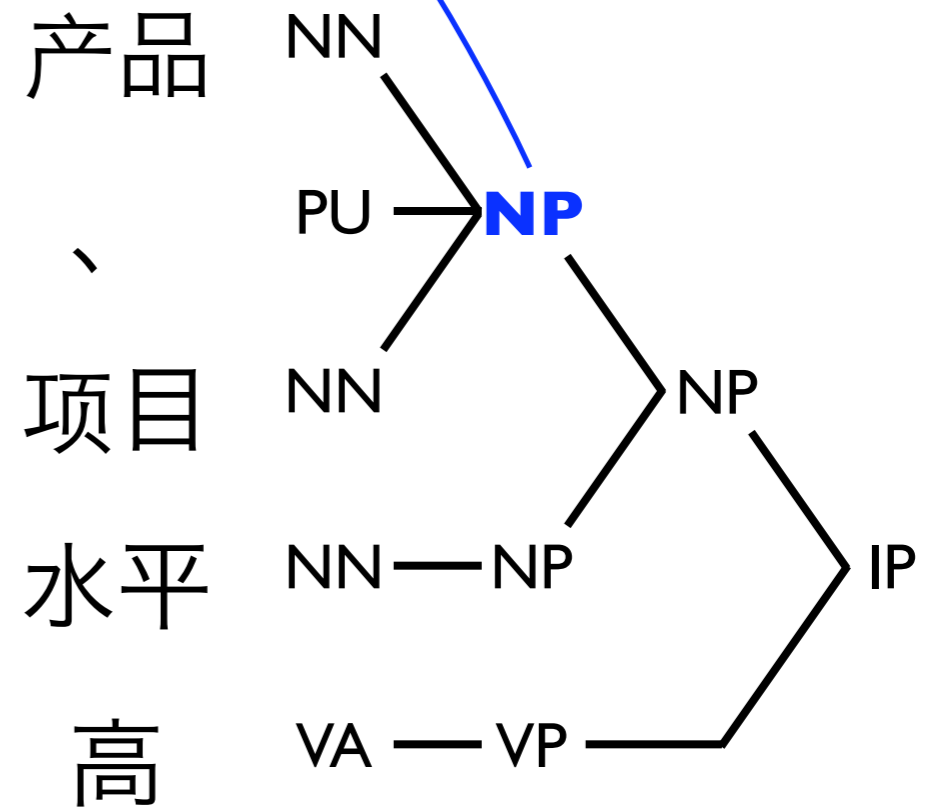
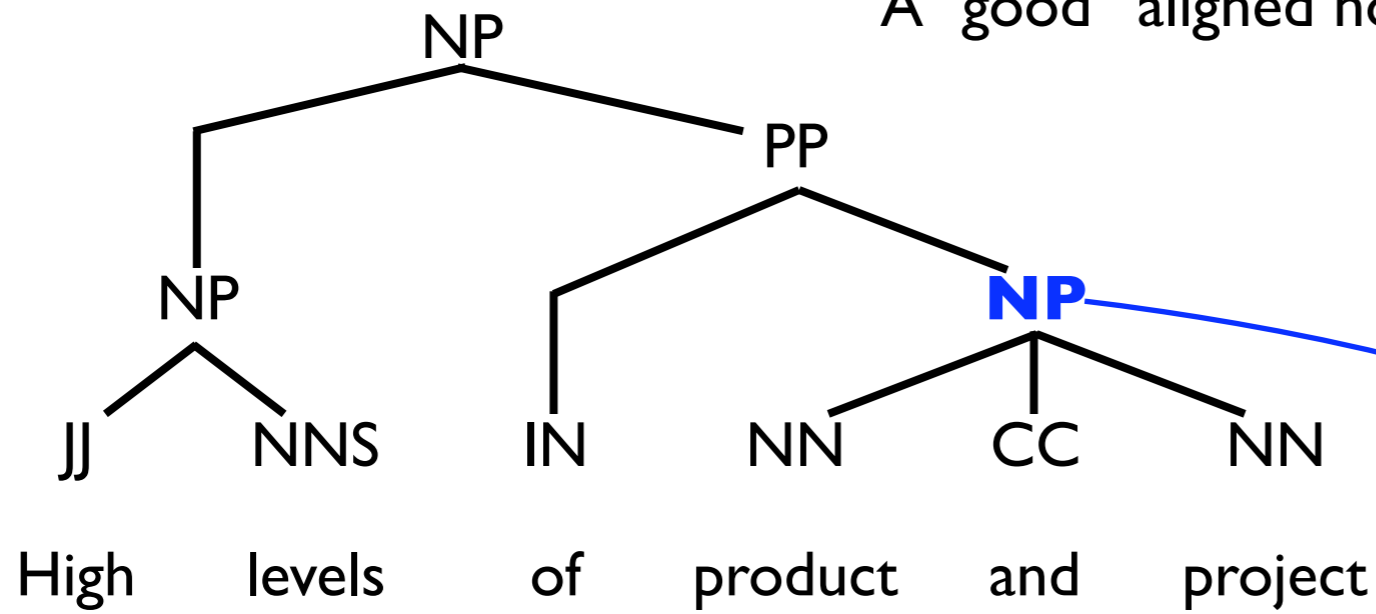
Source tree log likelihood: $\log P(t|s)$

Target tree log likelihood: $\log P(t'|s')$

“Default” is to use output of baseline parsers

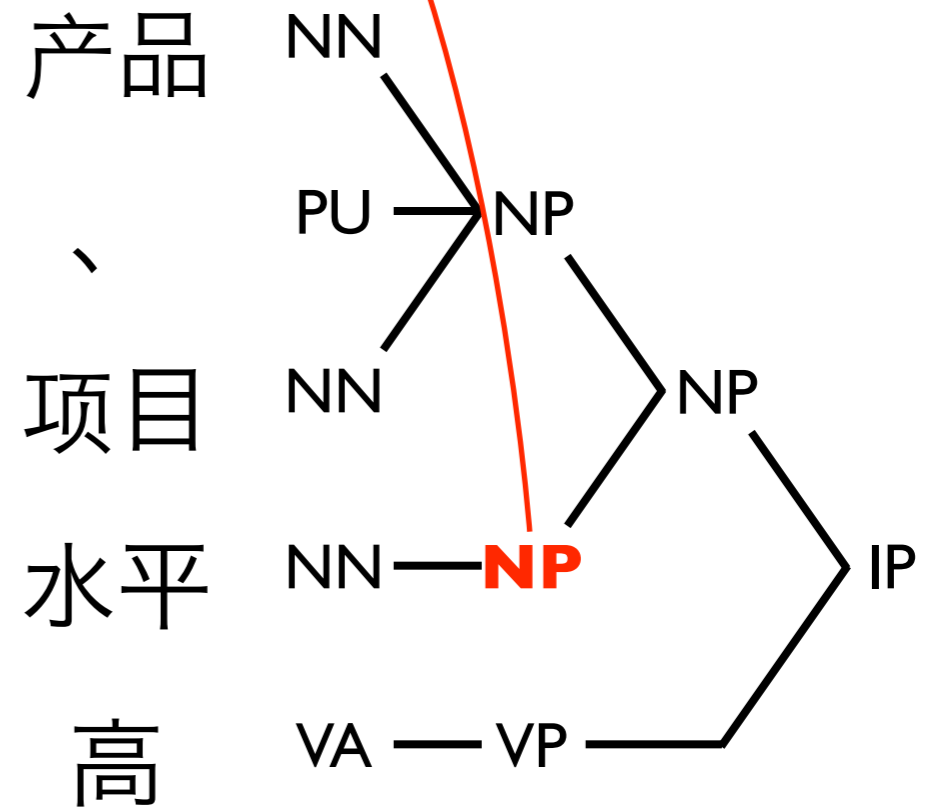
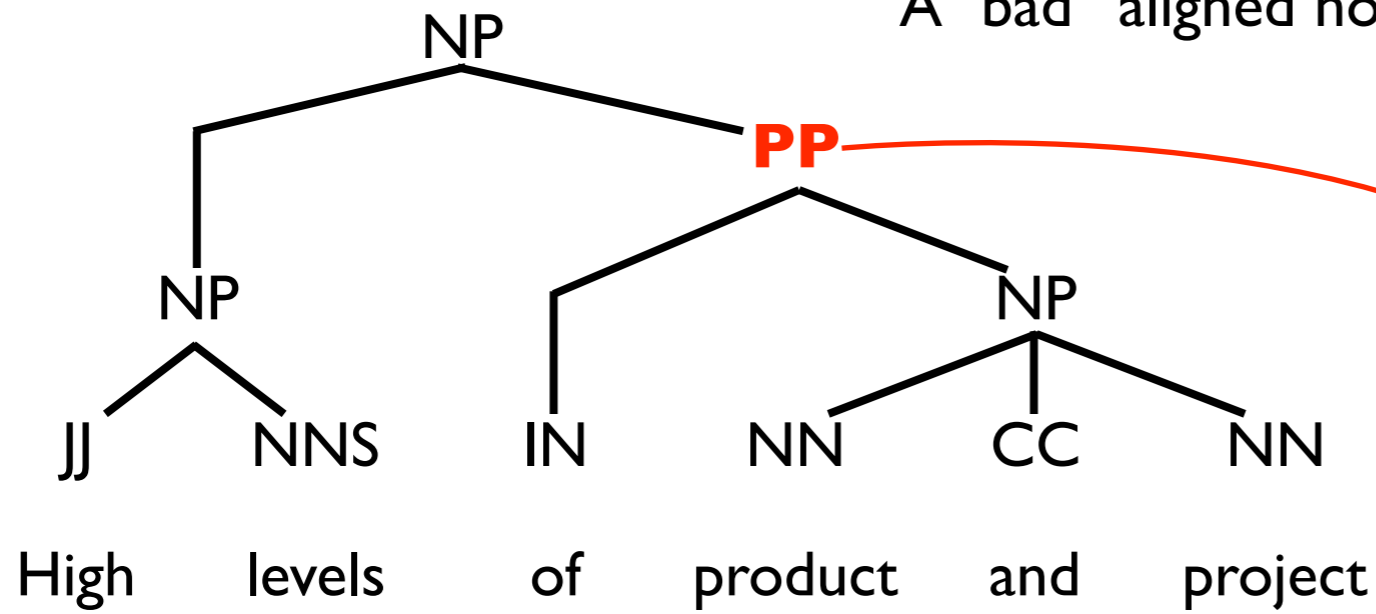
Bilingual Features

A "good" aligned node pair



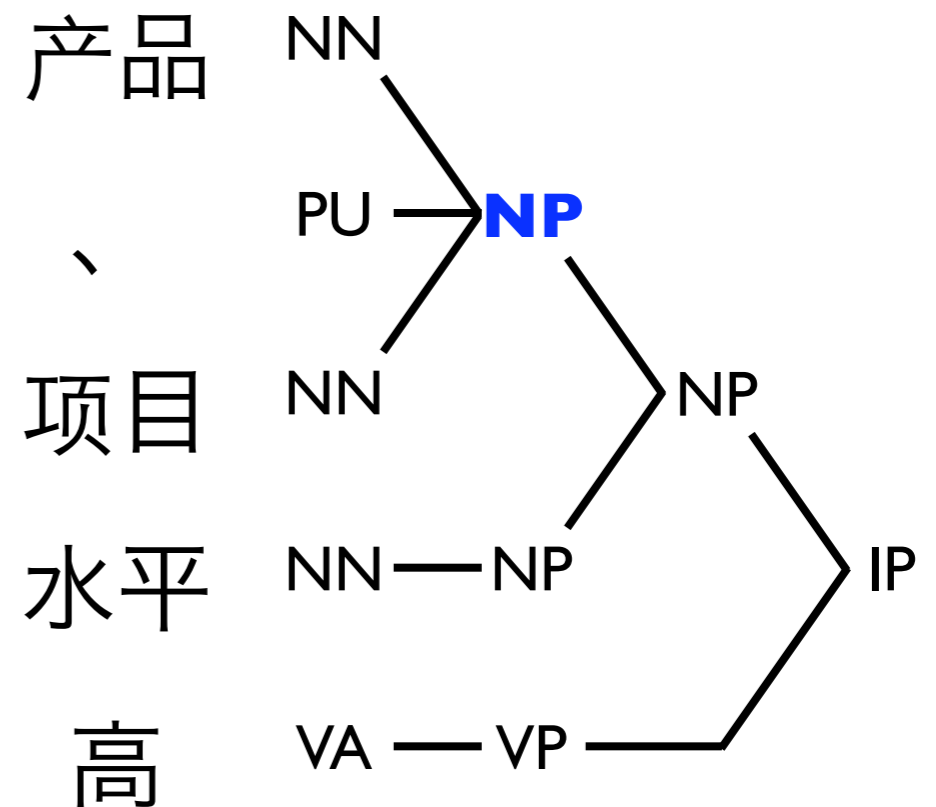
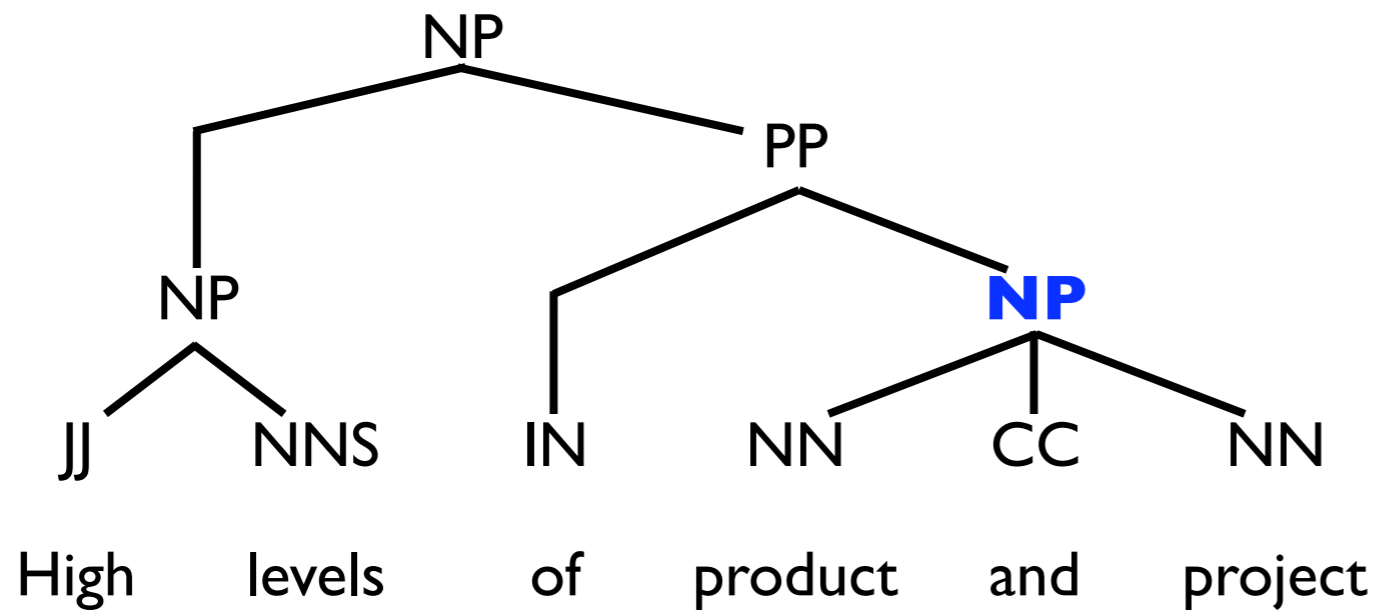
Bilingual Features

A "bad" aligned node pair



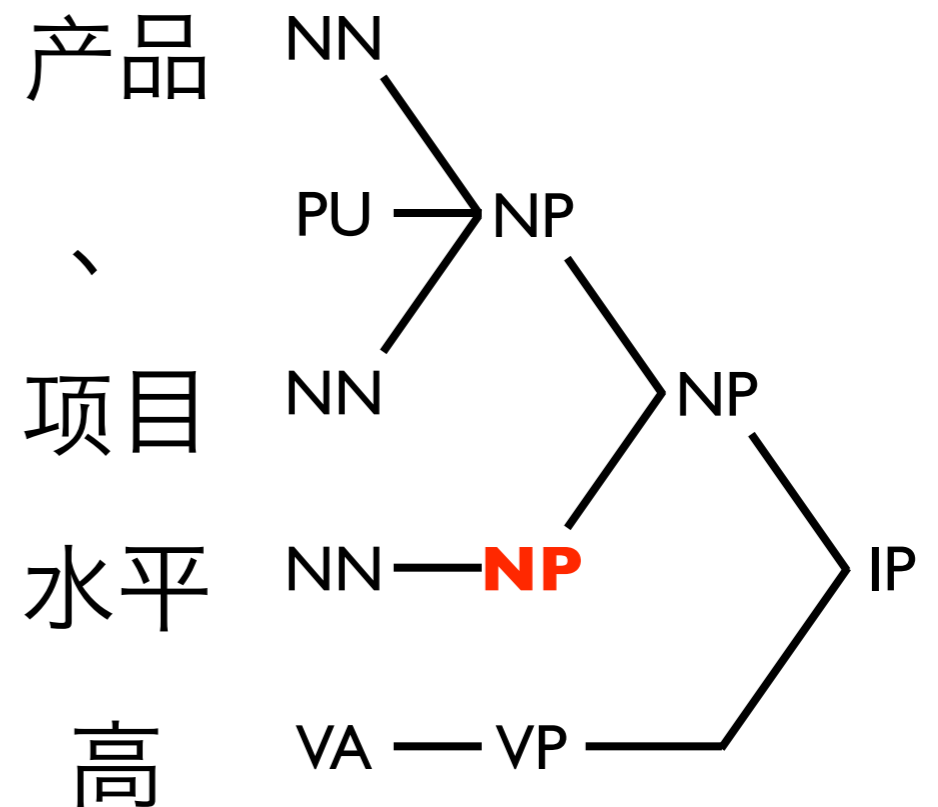
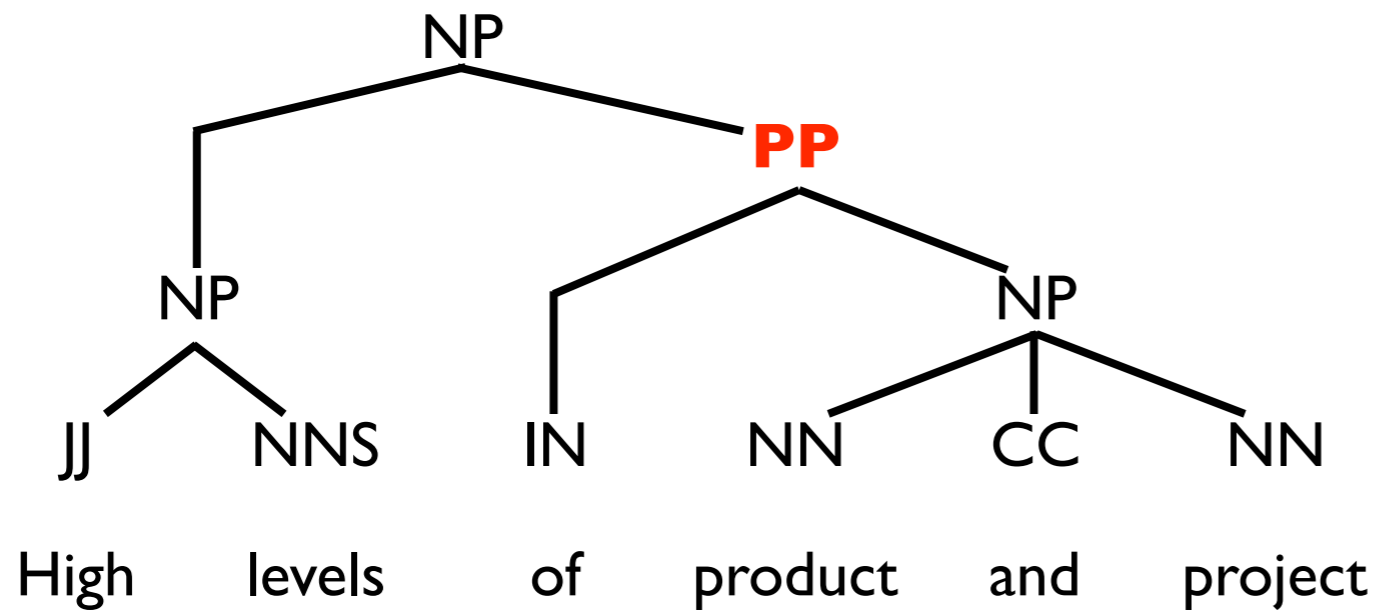
Features on Aligned Node Pairs

Node Label Indicator
 $\text{Indicator}[\text{NP}, \text{NP}] = 1.0$

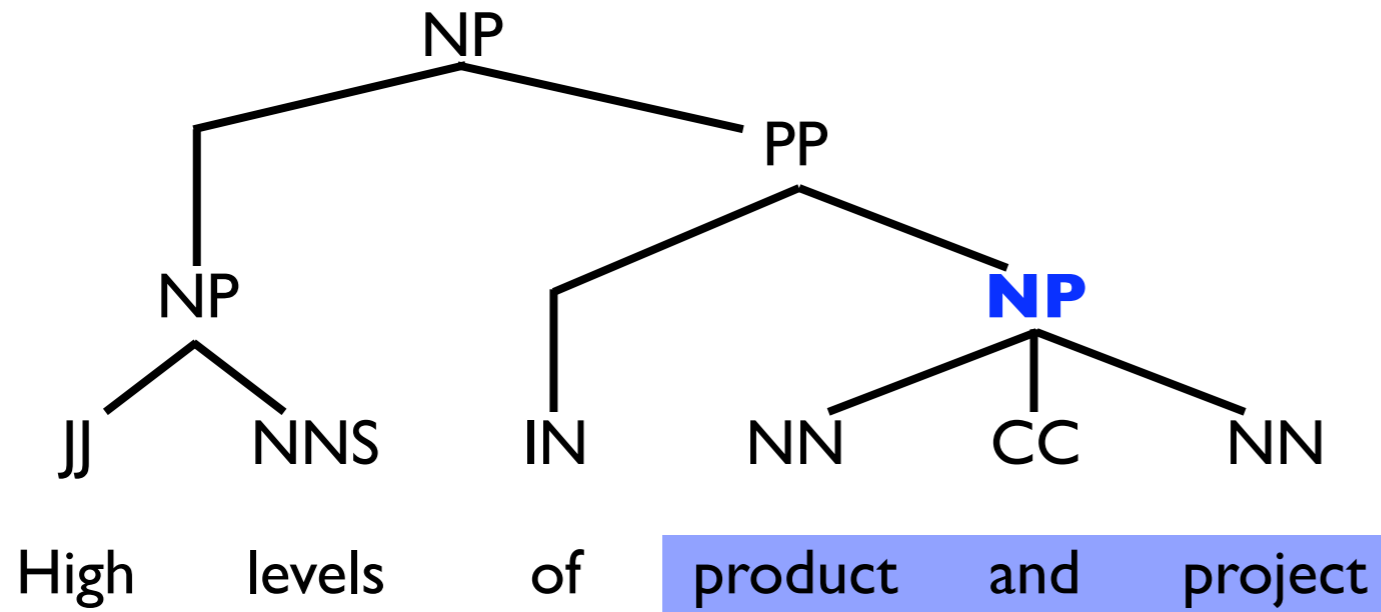


Features on Aligned Node Pairs

Node Label Indicator
Indicator[NP, PP] = 1.0

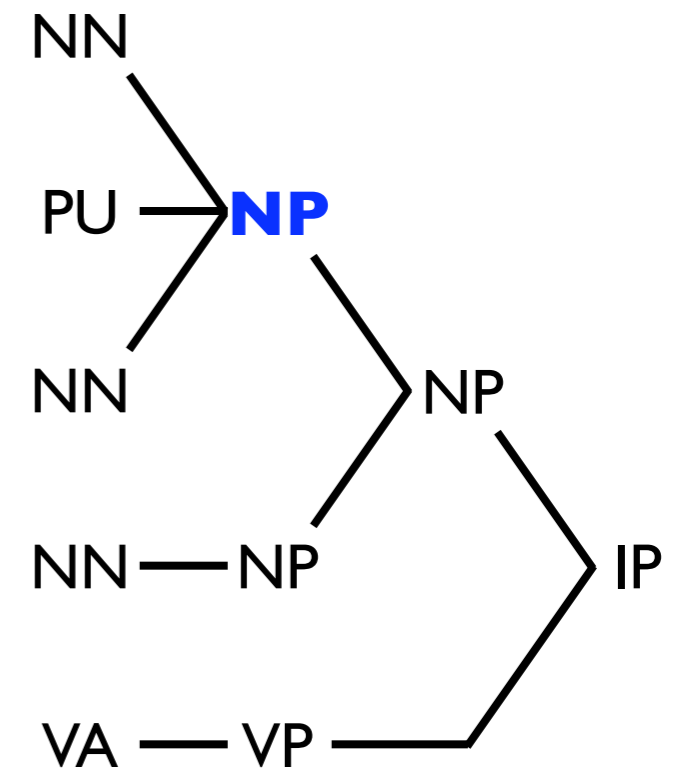


Features on Aligned Node Pairs

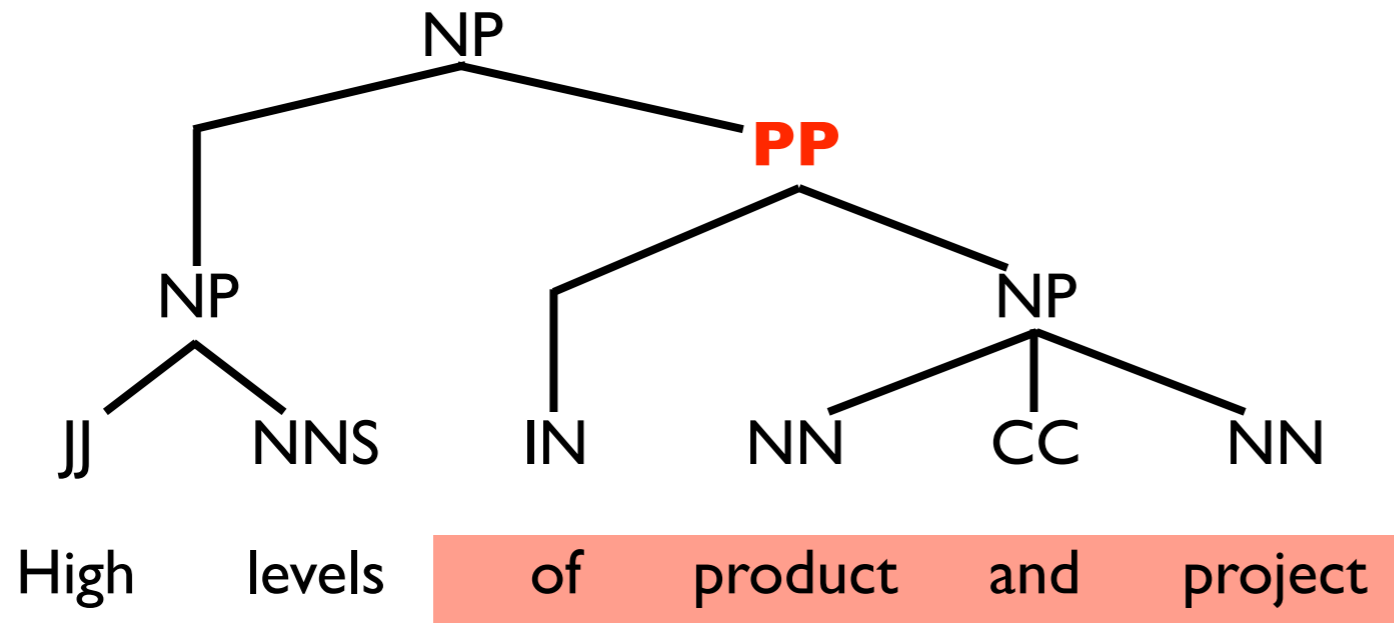


Span Difference
 SpanDiff = 0.0

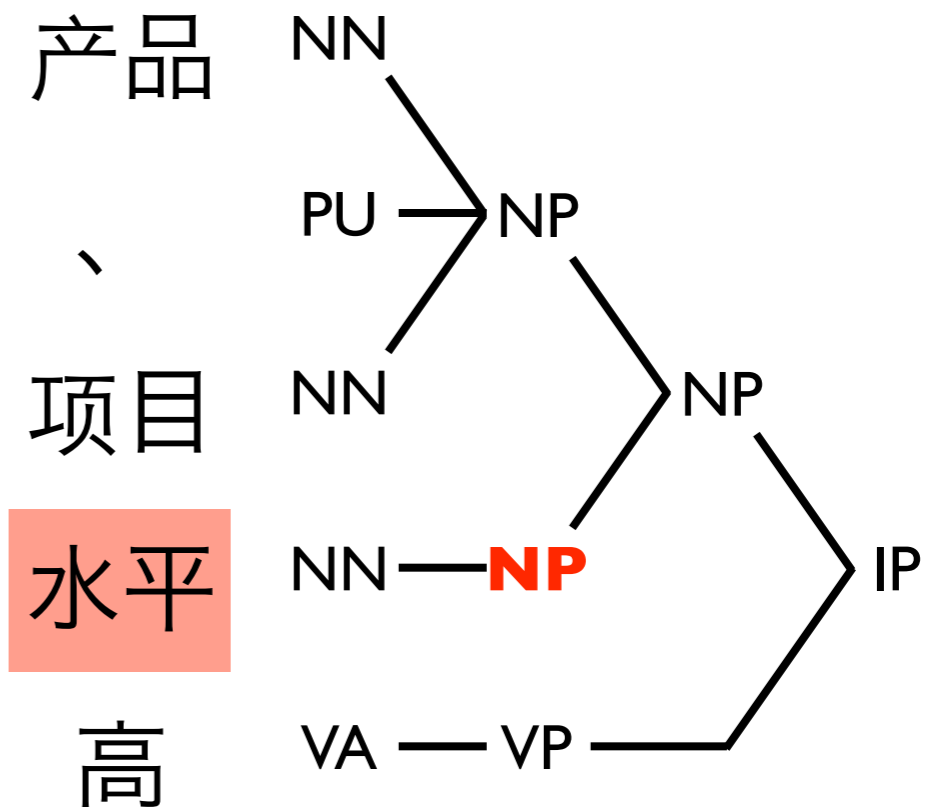
产品
 、
 项目
 水平
 高



Features on Aligned Tree Pairs

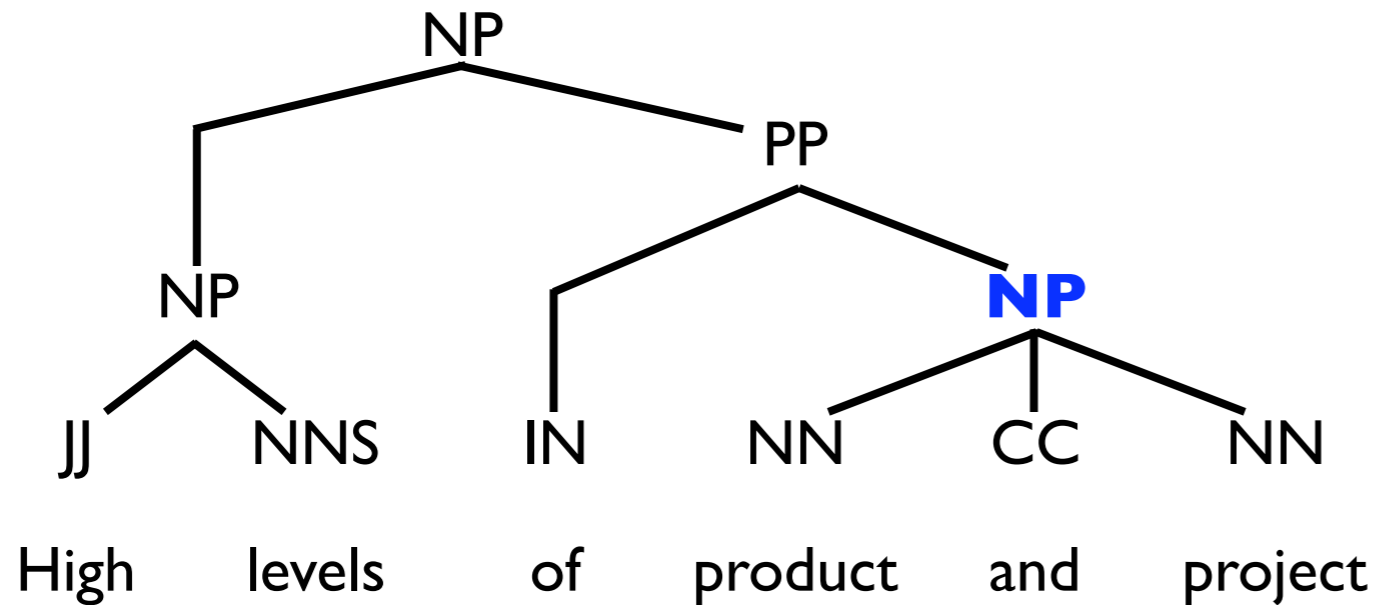


Span Difference
SpanDiff = 3.0

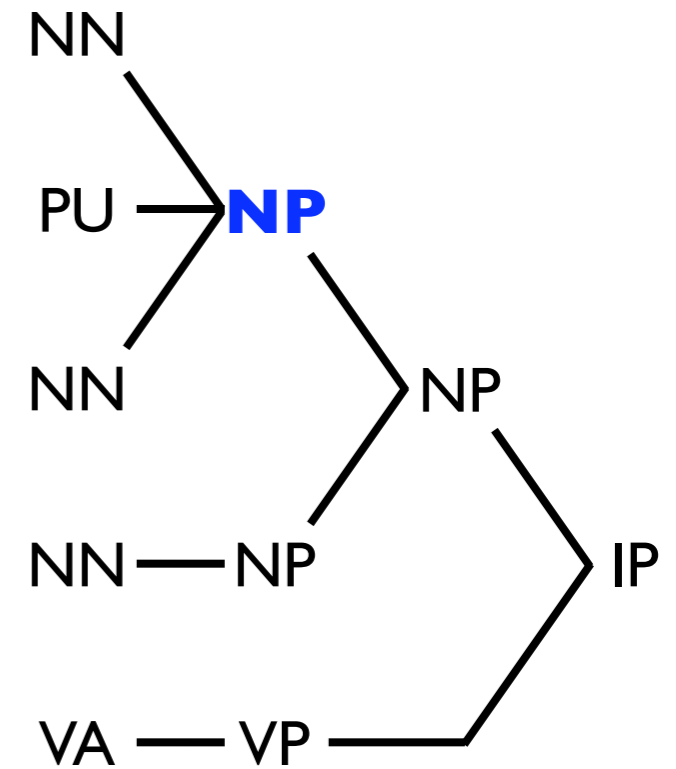


Features on Aligned Node Pairs

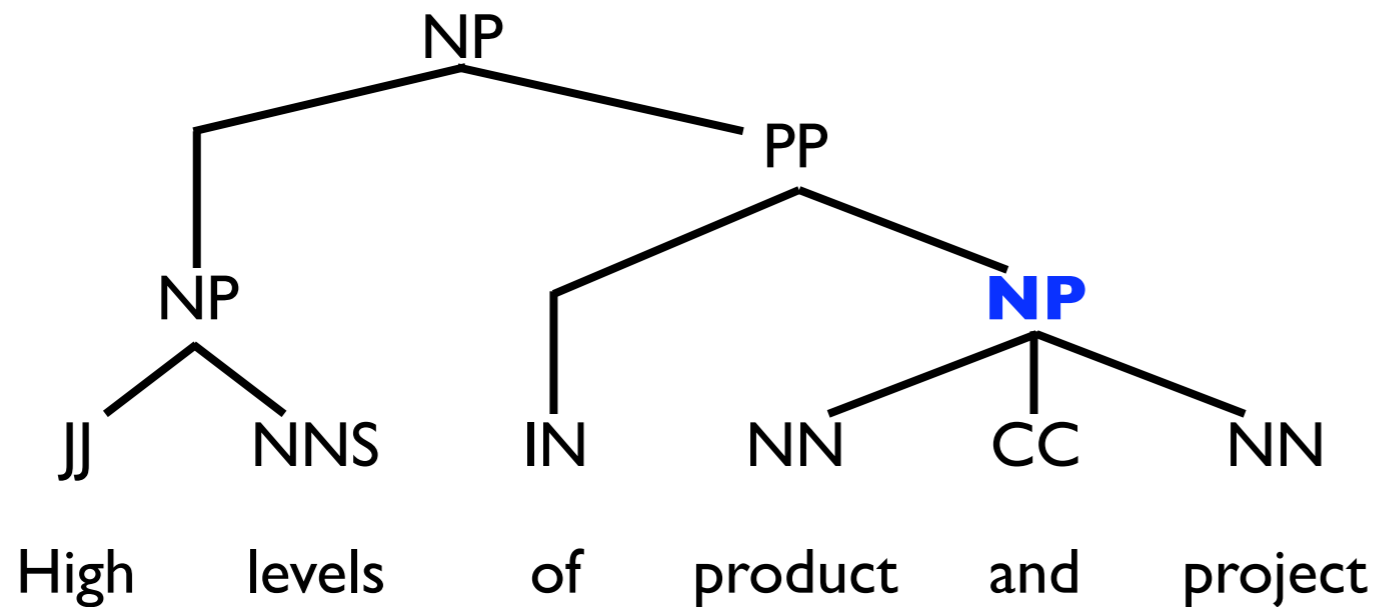
Word Alignments



产品
、
项目
水平
高



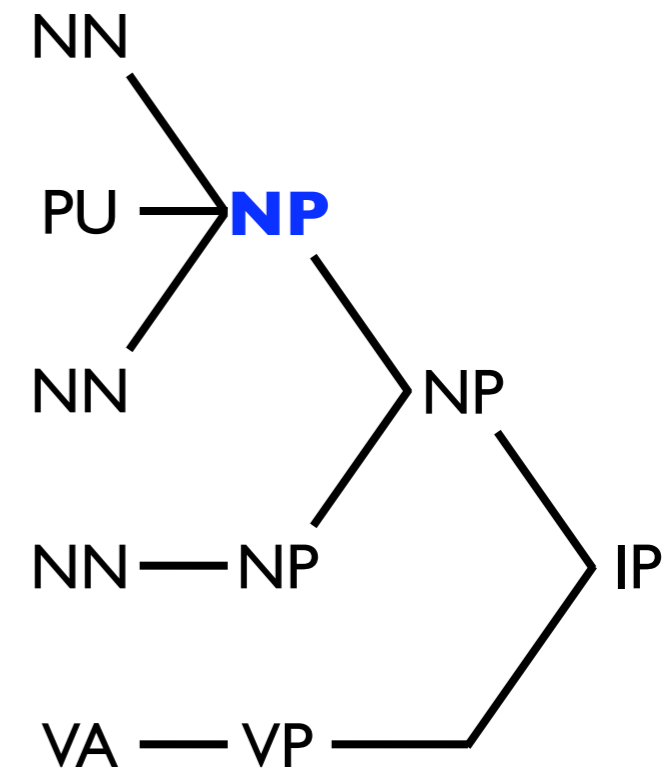
Features on Aligned Node Pairs



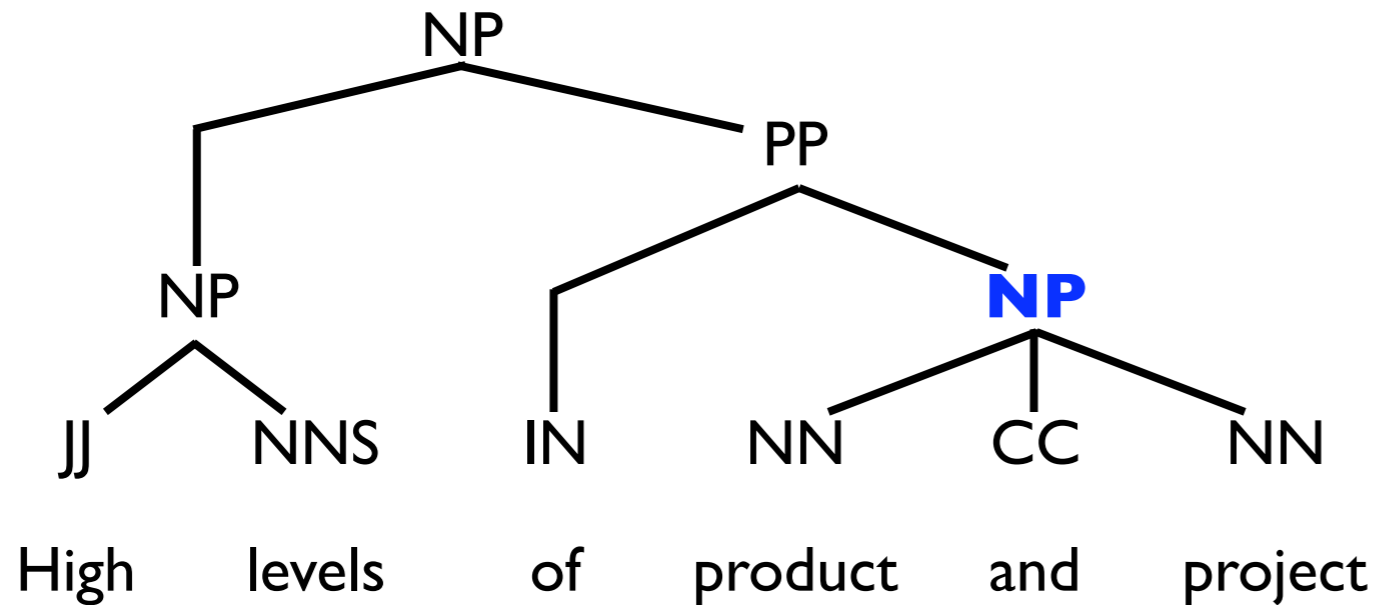
Word Alignments

InsideBoth	3.0
------------	-----

产品
、
项目
水平
高



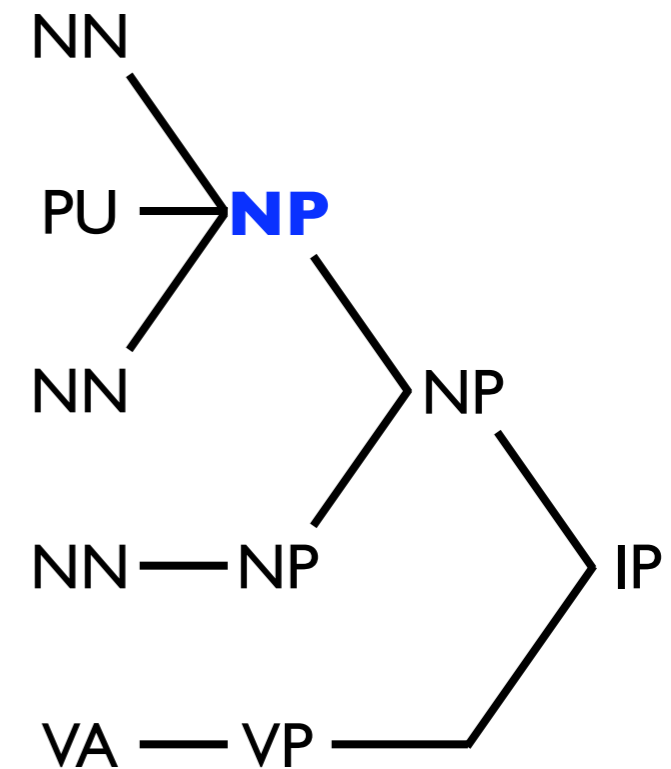
Features on Aligned Node Pairs



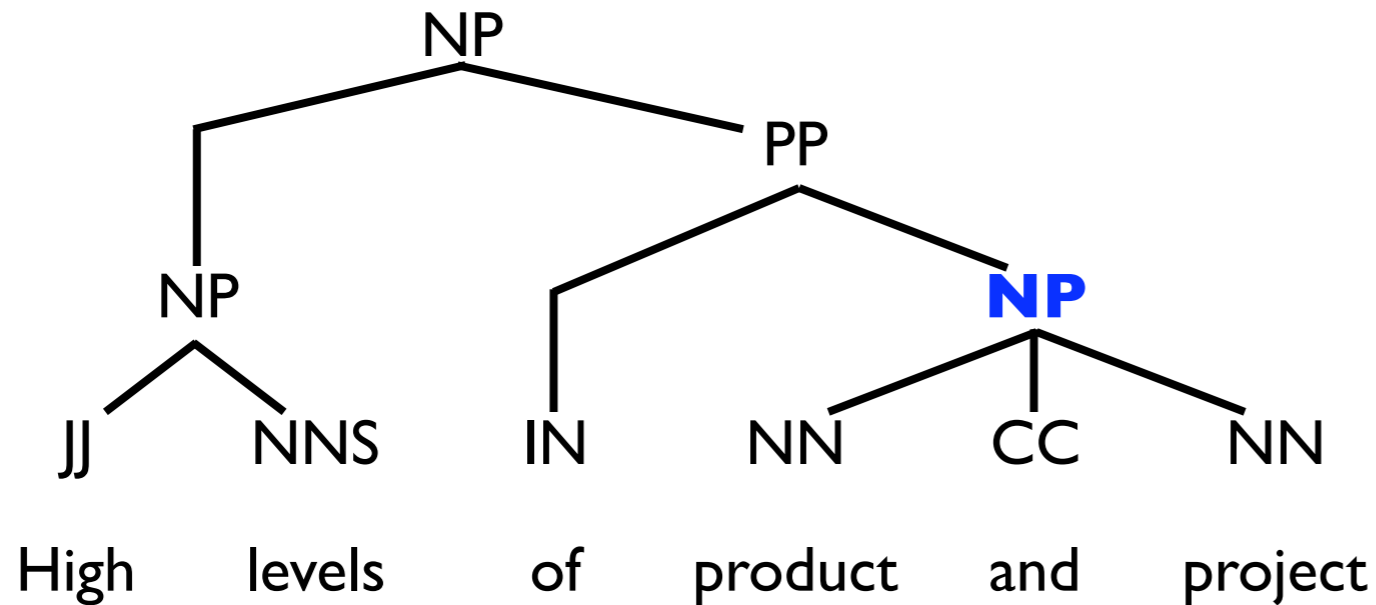
Word Alignments

InsideBoth	3.0
InSrcOutTrg	0.0

产品
、
项目
水平
高



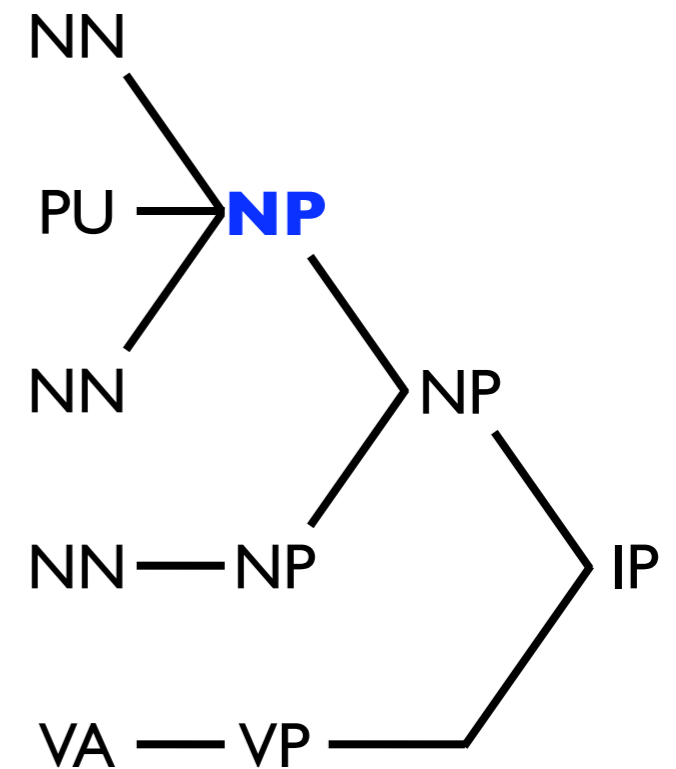
Features on Aligned Node Pairs



Word Alignments

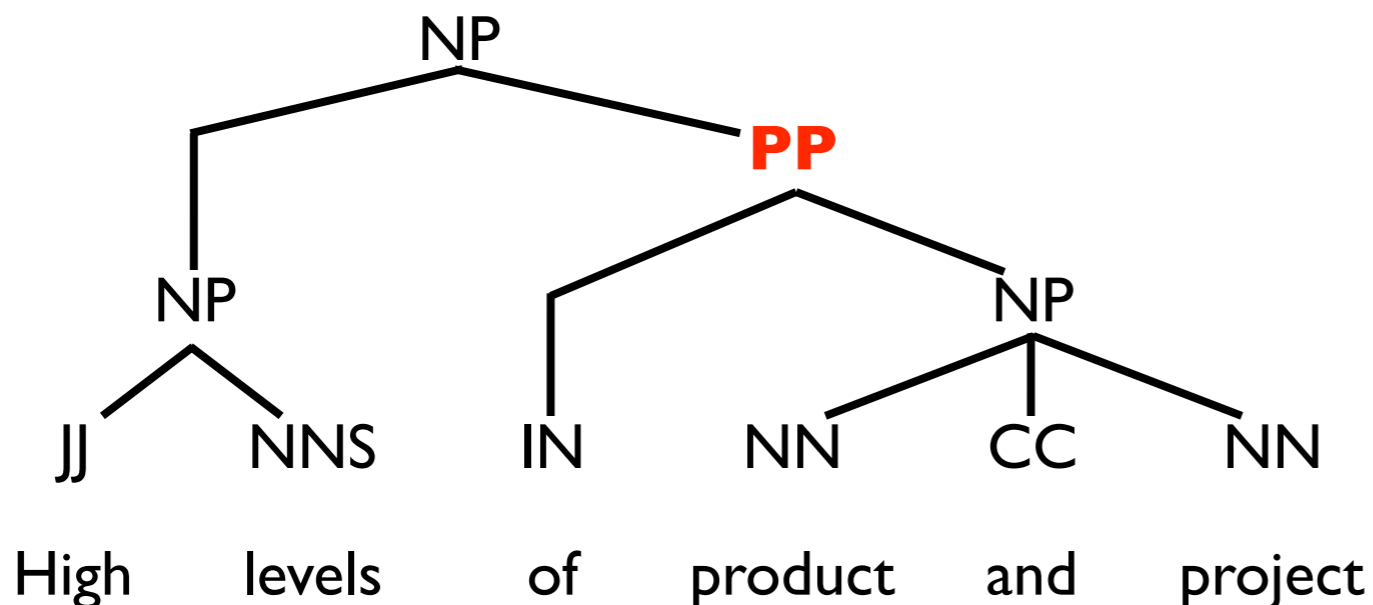
InsideBoth	3.0
InSrcOutTrg	0.0
InTrgOutSrc	0.0

产品
、
项目
水平
高

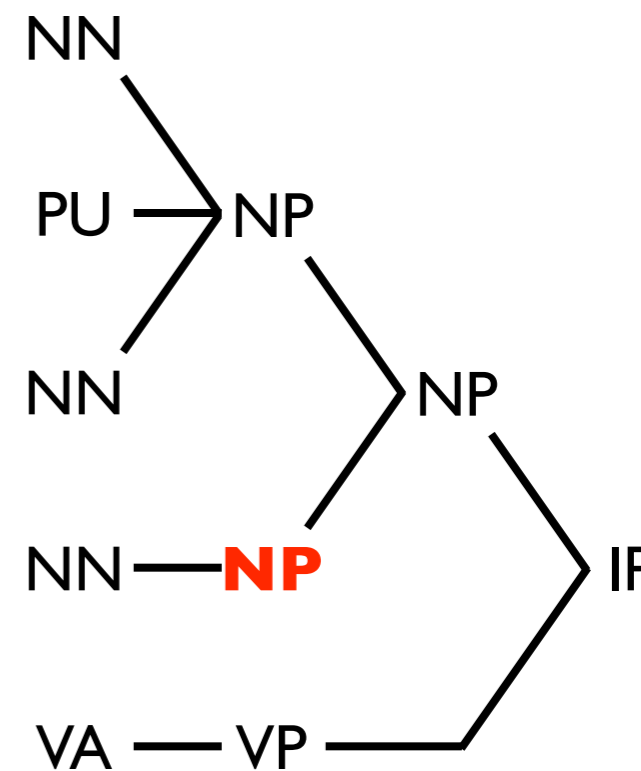


Features on Aligned Node Pairs

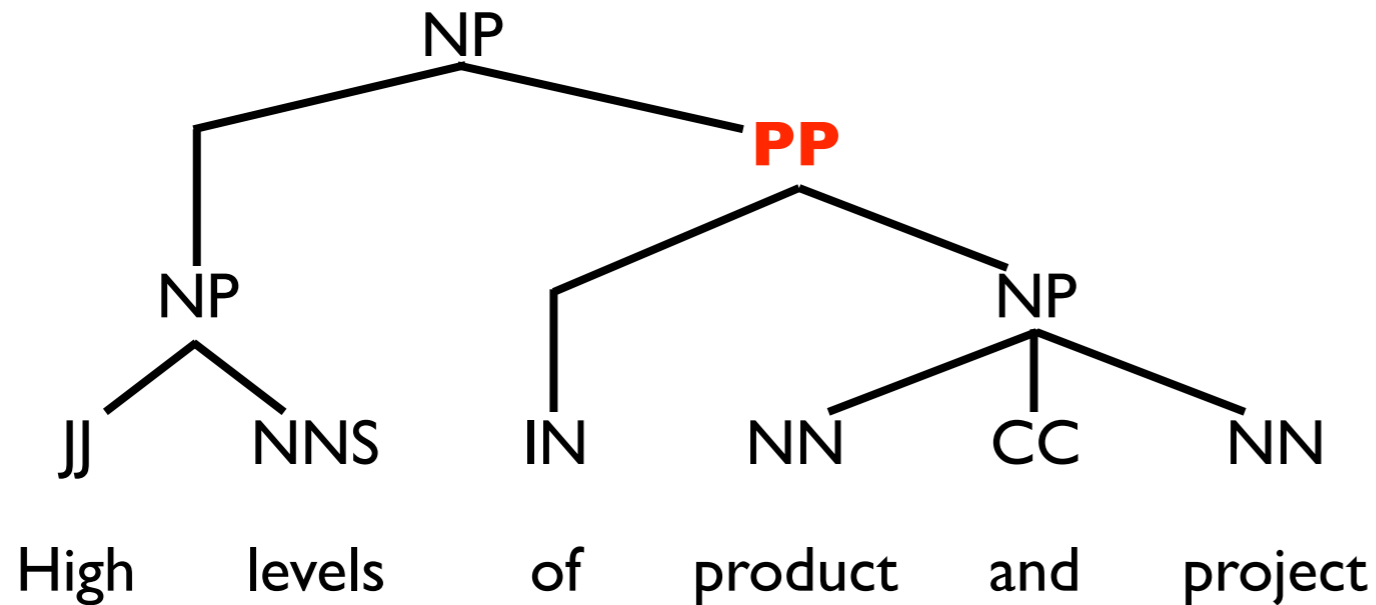
Word Alignments



产品
、
项目
水平
高



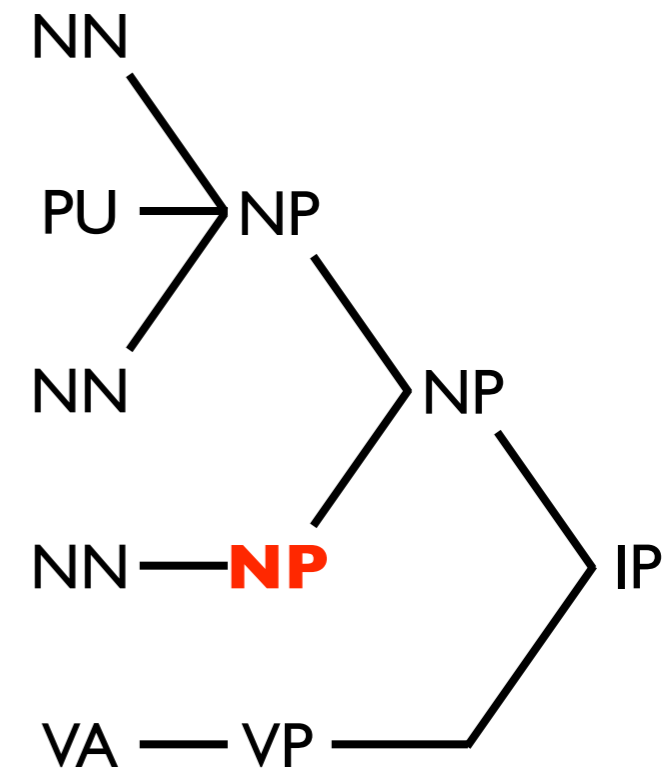
Features on Aligned Node Pairs



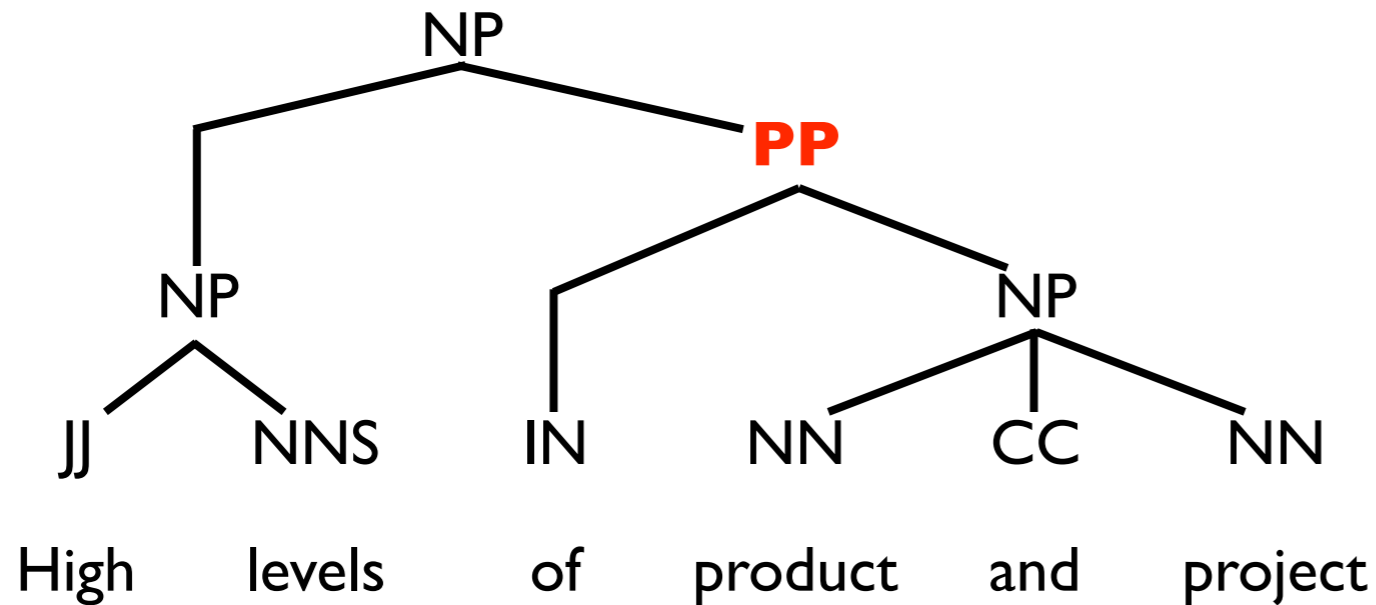
Word Alignments

InsideBoth	0.0
------------	-----

产品
、
项目
水平
高



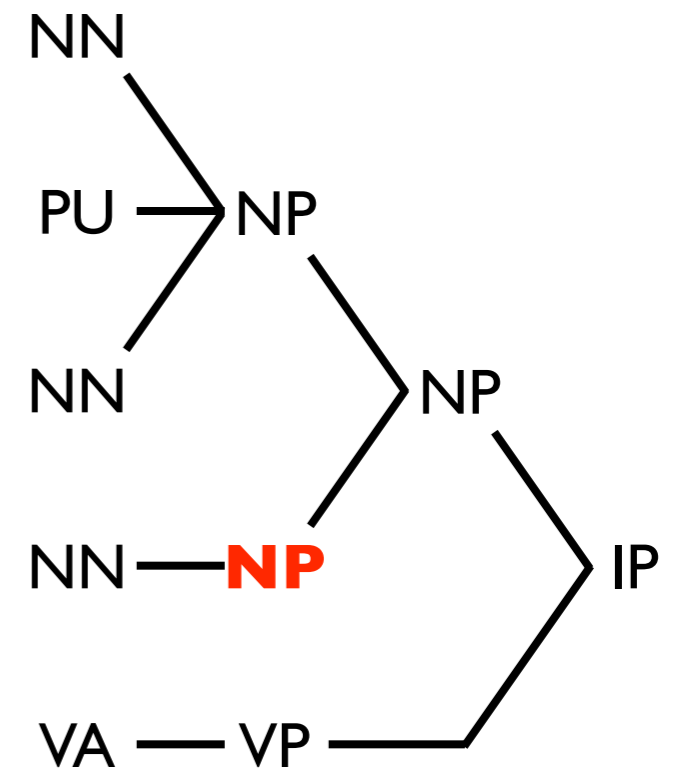
Features on Aligned Node Pairs



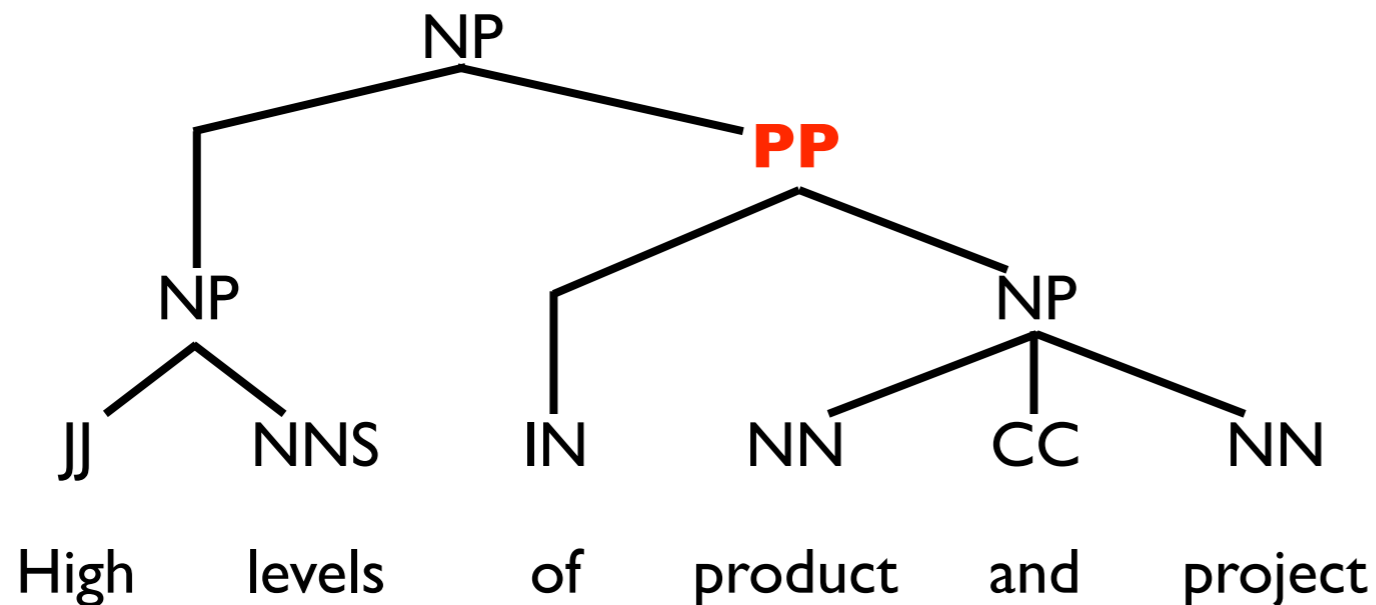
Word Alignments

InsideBoth	0.0
InSrcOutTrg	3.0

产品
、
项目
水平
高



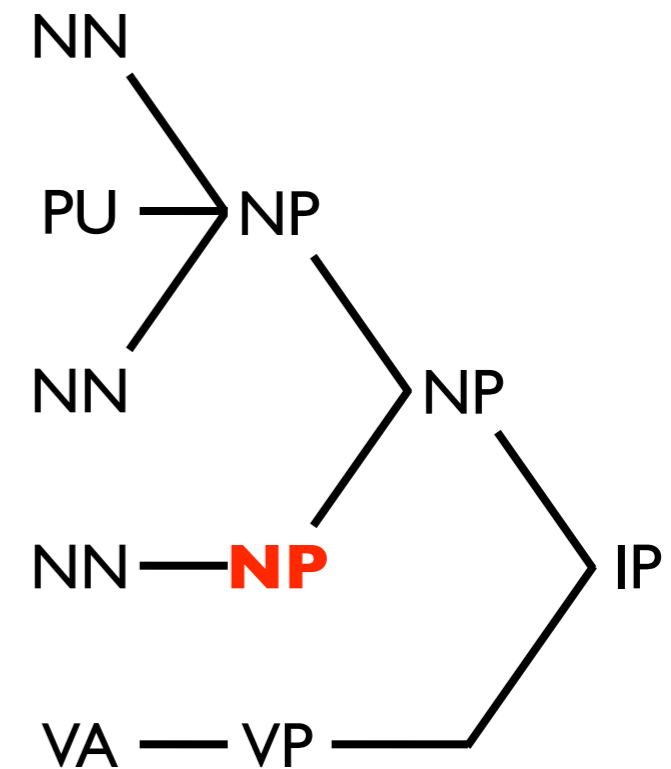
Features on Aligned Node Pairs



Word Alignments

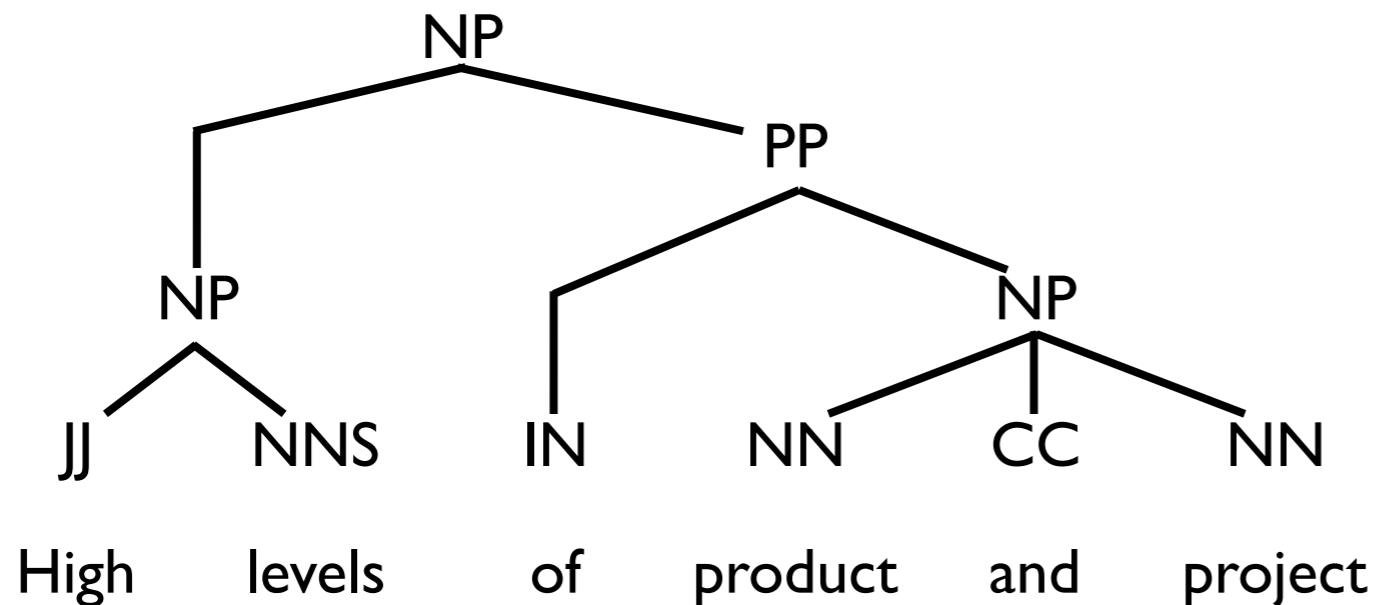
InsideBoth	0.0
InSrcOutTrg	3.0
InTrgOutSrc	1.0

产品
、
项目
水平
高



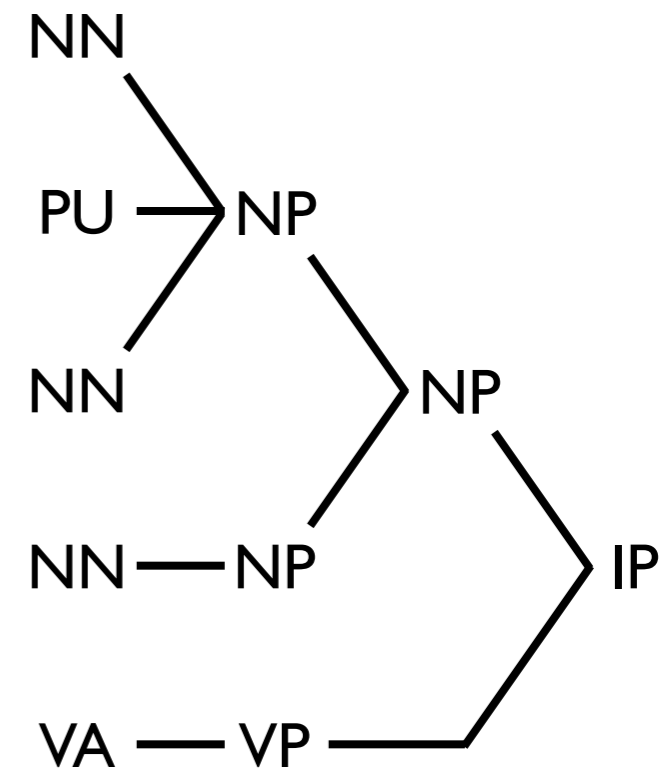
Features on Aligned Node Pairs

Word Alignment Posteriors

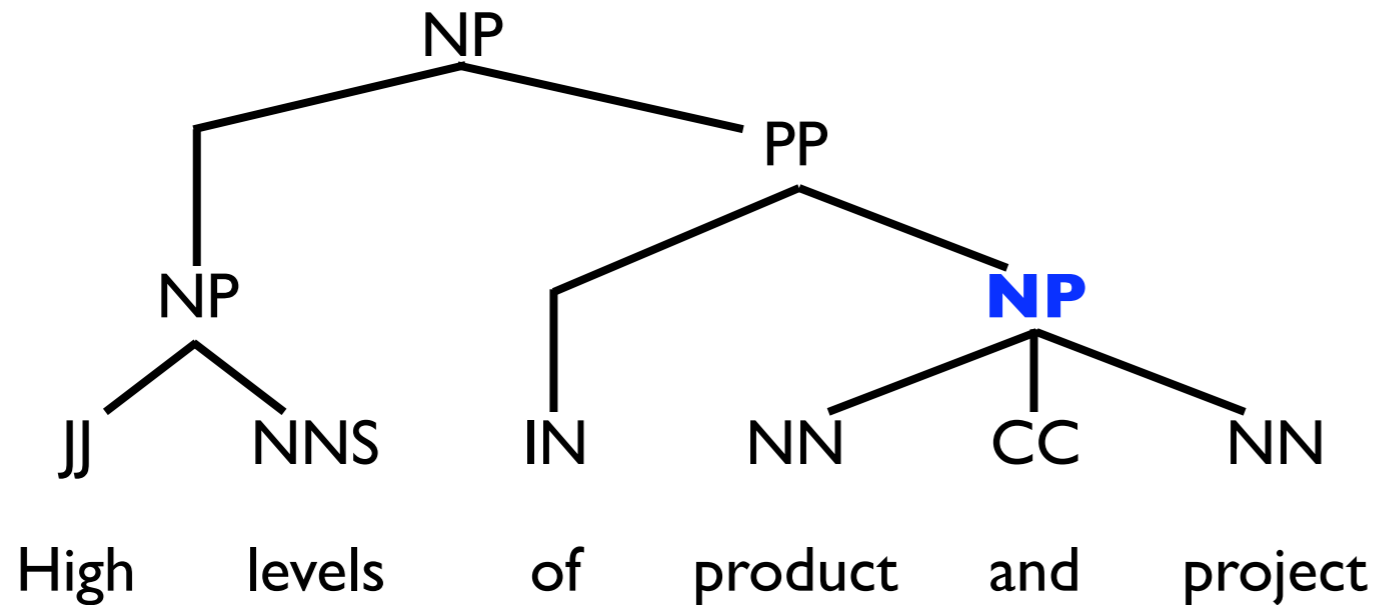


0	0.2	0	0.8	0	0.4
0	0	0	0	0.9	0.1
0.1	0	0.1	0.3	0	0.8
0	0.7	0.4	0.2	0	0
0.8	0.3	0.2	0	0.1	0.1

产品
、
项目
水平
高



Features on Aligned Node Pairs

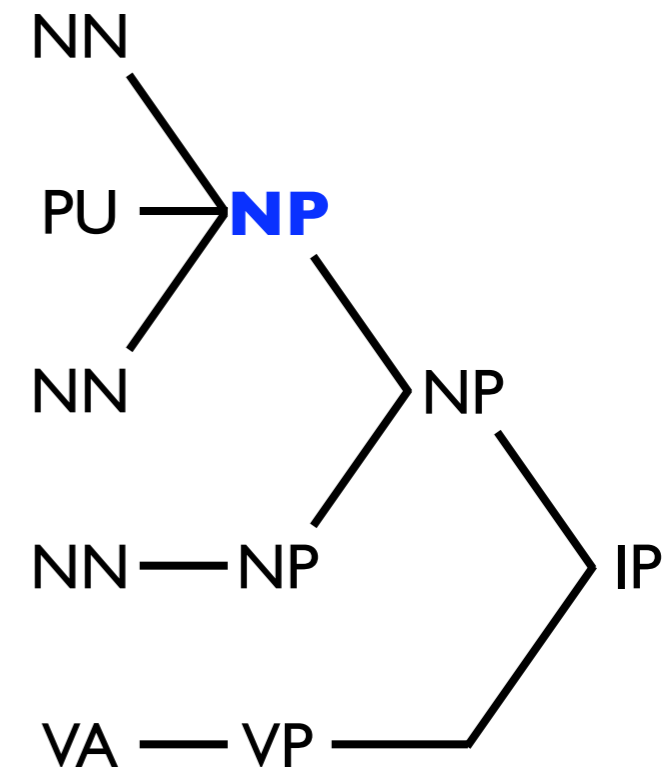


Word Alignment Posteriors

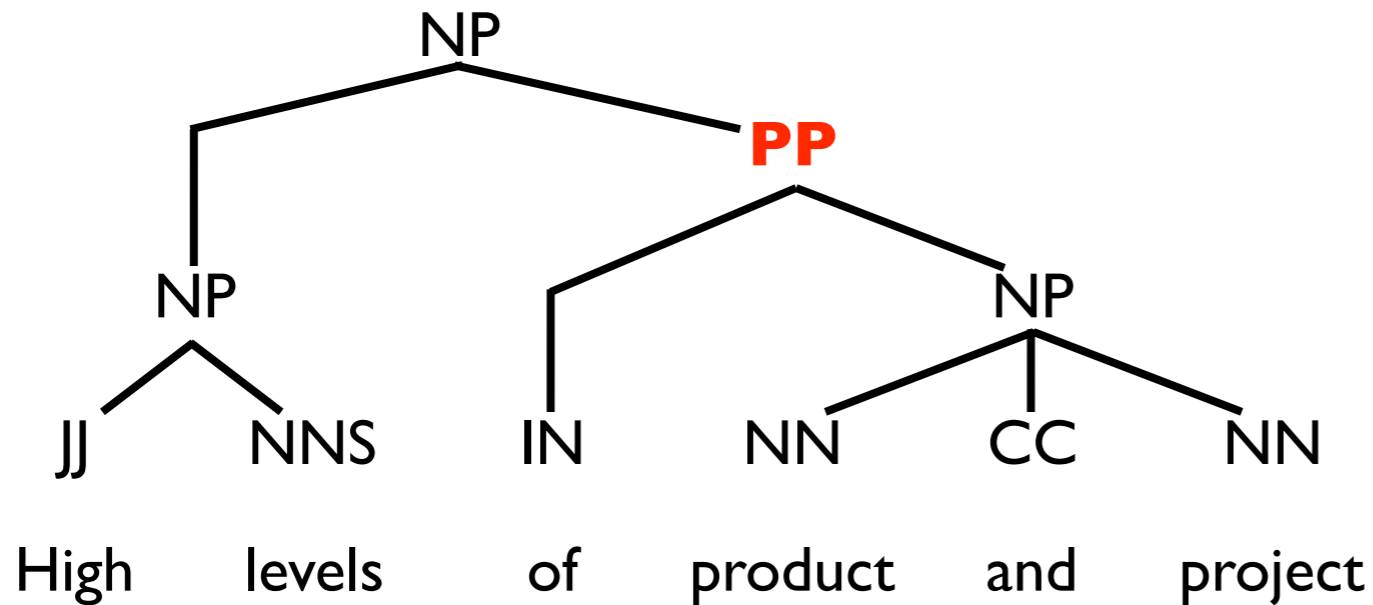
InsideBoth	3.3
InSrcOutTrg	0.4
InTrgOutSrc	0.4

0	0.2	0	0.8	0	0.4
0	0	0	0	0.9	0.1
0.1	0	0.1	0.3	0	0.8
0	0.7	0.4	0.2	0	0
0.8	0.3	0.2	0	0.1	0.1

产品
、
项目
水平
高



Features on Aligned Node Pairs

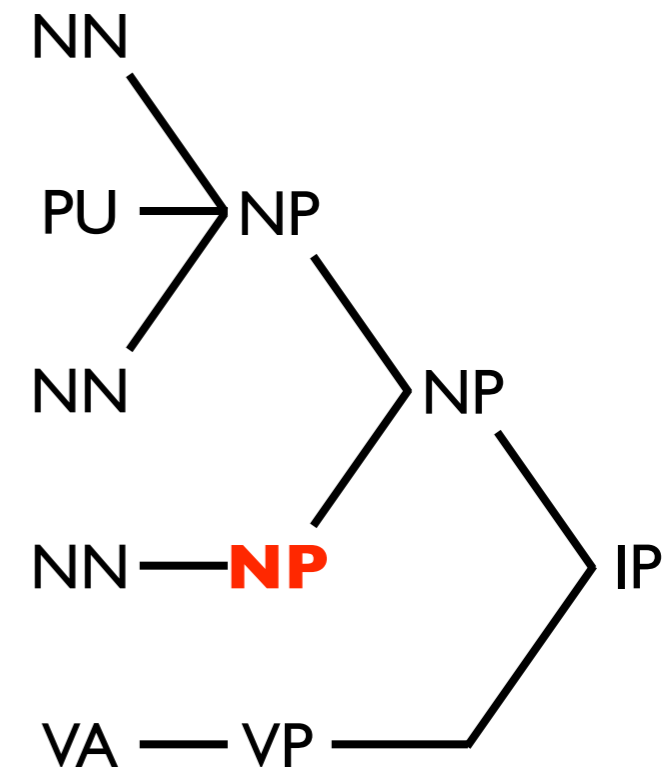


Word Alignment Posteriors

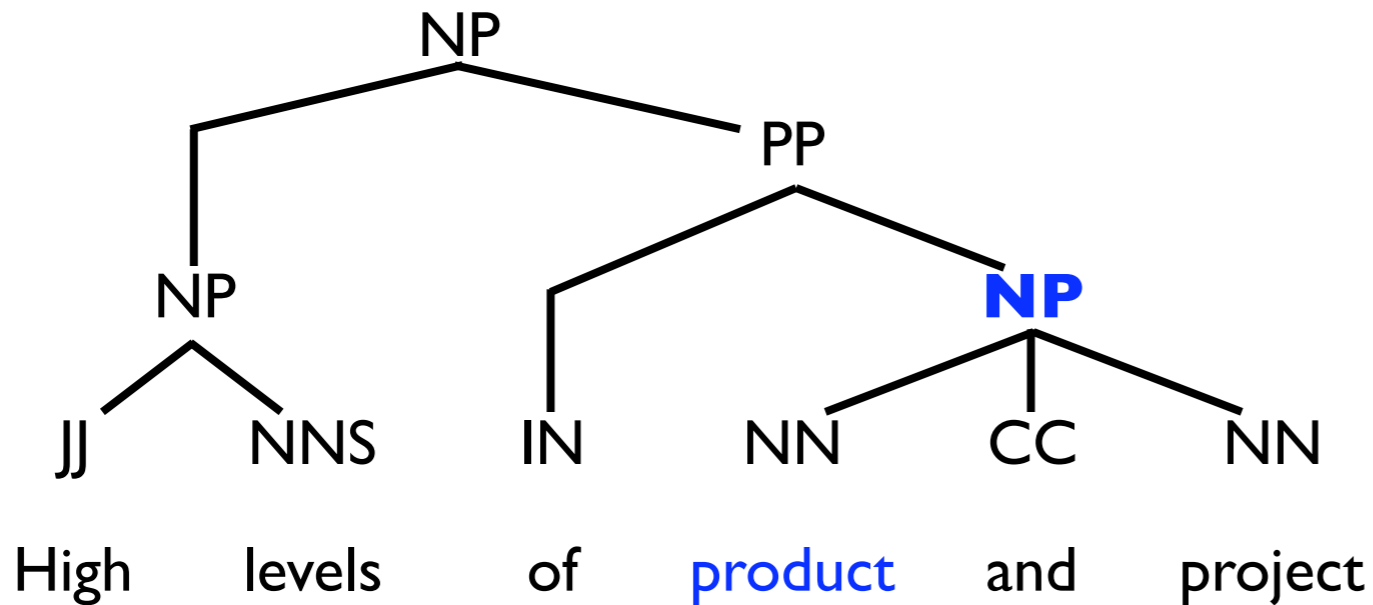
InsideBoth	0.6
InSrcOutTrg	3.8
InTrgOutSrc	0.7

0	0.2	0	0.8	0	0.4
0	0	0	0	0.9	0.1
0.1	0	0.1	0.3	0	0.8
0	0.7	0.4	0.2	0	0
0.8	0.3	0.2	0	0.1	0.1

产品
、
项目
水平
高



Features on Aligned Node Pairs

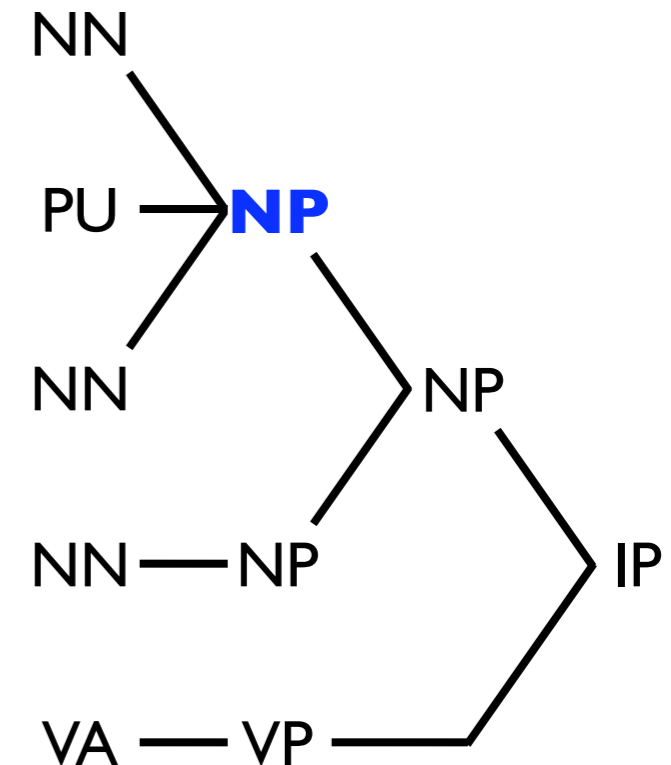


Head Word Alignments

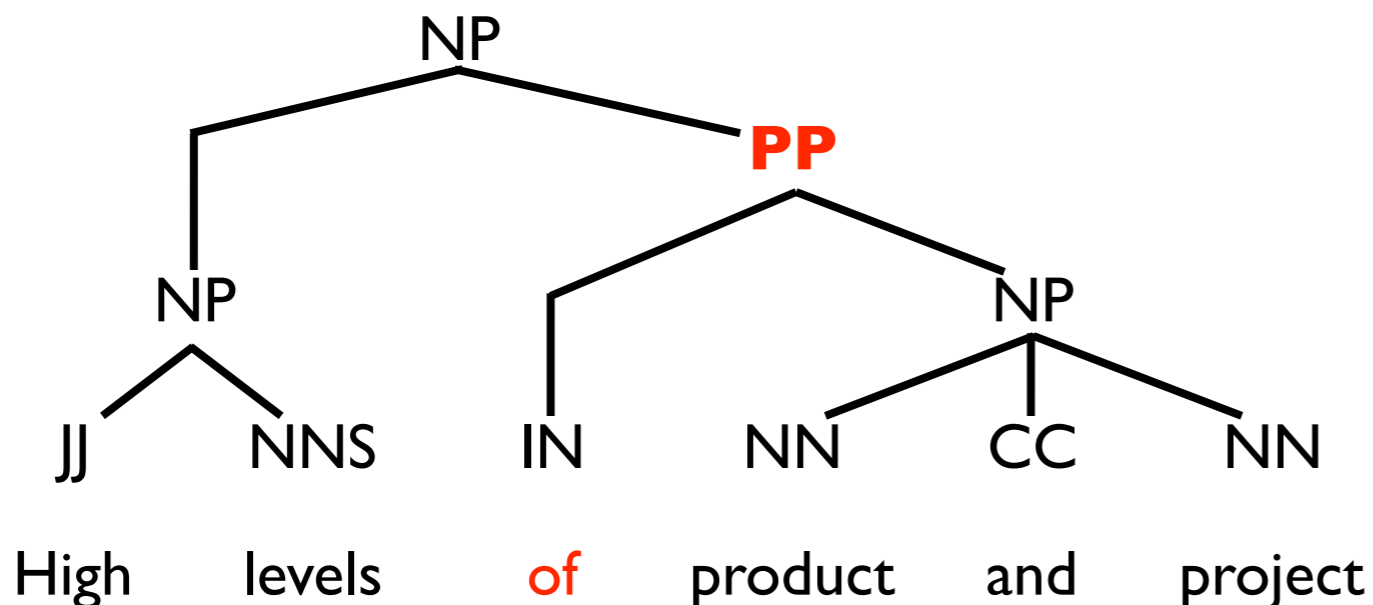
Hard	1.0
Soft	0.8

0	0.2	0	0.8	0	0.4
0	0	0	0	0.9	0.1
0.1	0	0.1	0.3	0	0.8
0	0.7	0.4	0.2	0	0
0.8	0.3	0.2	0	0.1	0.1

产品
、
项目
水平
高



Features on Aligned Node Pairs

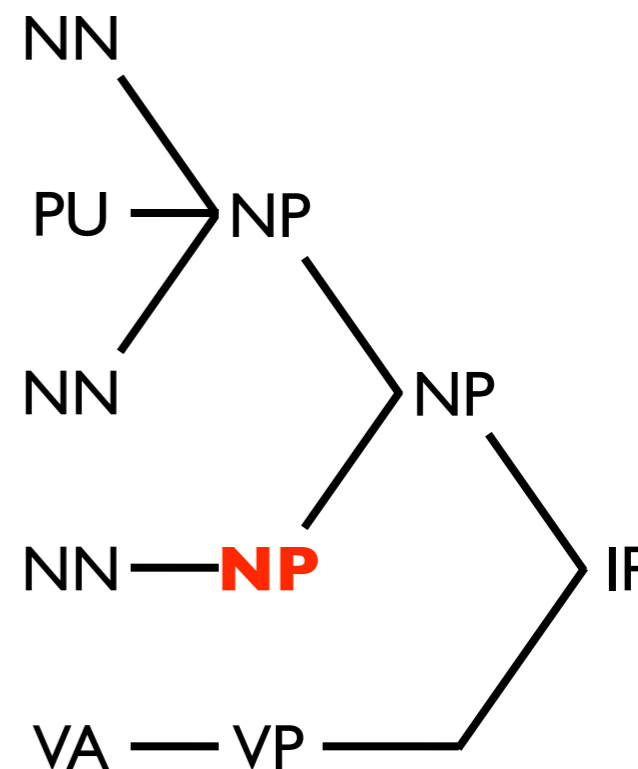


Head Word Alignments

Hard	0.0
Soft	0.4

0	0.2	0	0.8	0	0.4
0	0	0	0	0.9	0.1
0.1	0	0.1	0.3	0	0.8
0	0.7	0.4	0.2	0	0
0.8	0.3	0.2	0	0.1	0.1

产品
、
项目
水平
高



Overview

- A log-linear model over aligned parse trees

- Training with latent tree alignments

- Improvements from joint parsing



Training with Latent Alignments

Training with Latent Alignments

- Observed: (s, s', t, t')

Training with Latent Alignments

- Observed: (s, s', t, t')
- Optimizing weight vector:

$$w^* = \arg \max_w P(t, t' | s, s')$$

Training with Latent Alignments

- Observed: (s, s', t, t')
- Optimizing weight vector:

$$\begin{aligned} w^* &= \arg \max_w P(t, t' | s, s') \\ &= \arg \max_w \sum_a P(t, a, t' | s, s') \end{aligned}$$

Training with Latent Alignments

- Observed: (s, s', t, t')
- Optimizing weight vector:

$$\begin{aligned} w^* &= \arg \max_w P(t, t' | s, s') \\ &= \arg \max_w \sum_a P(t, a, t' | s, s') \\ &= \arg \max_w \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))} \end{aligned}$$

Training with Latent Alignments

- Observed: (s, s', t, t')
- Optimizing weight vector:

$$\begin{aligned}
 w^* &= \arg \max_w P(t, t' | s, s') \\
 &= \arg \max_w \sum_a P(t, a, t' | s, s') \\
 &= \arg \max_w \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}
 \end{aligned}$$

Problem 1: Infinite sum over tree pairs

Training with Latent Alignments

- Observed: (s, s', t, t')
- Optimizing weight vector:

$$\begin{aligned}
 w^* &= \arg \max_w P(t, t' | s, s') \\
 &= \arg \max_w \sum_a P(t, a, t' | s, s') \\
 &= \arg \max_w \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}
 \end{aligned}$$

Problem 2: #P-hard sum over alignments



Approximating Sum over Tree Pairs

Approximating Sum over Tree Pairs

- Problem:

$$P(t, t' | s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}$$

Approximating Sum over Tree Pairs

- Problem:

$$P(t, t' | s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}$$

- Solution:

Approximating Sum over Tree Pairs

- Problem:

$$P(t, t' | s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}$$

- Solution:
 - Train in reranking mode

Approximating Sum over Tree Pairs

- Problem:

$$P(t, t' | s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}$$

- Solution:
 - Train in reranking mode
 - Candidate lists (T, T') from baseline parsers

Approximating Sum over Tree Pairs

- Problem:

$$P(t, t' | s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}$$

- Solution:

- Train in reranking mode
- Candidate lists (T, T') from baseline parsers

$$P(t, t' | s, s') \approx \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \sum_a \exp(w^\top \phi(t, a, t'))}$$



Tree Candidates

Tree Candidates

k Source Candidates



Tree Candidates

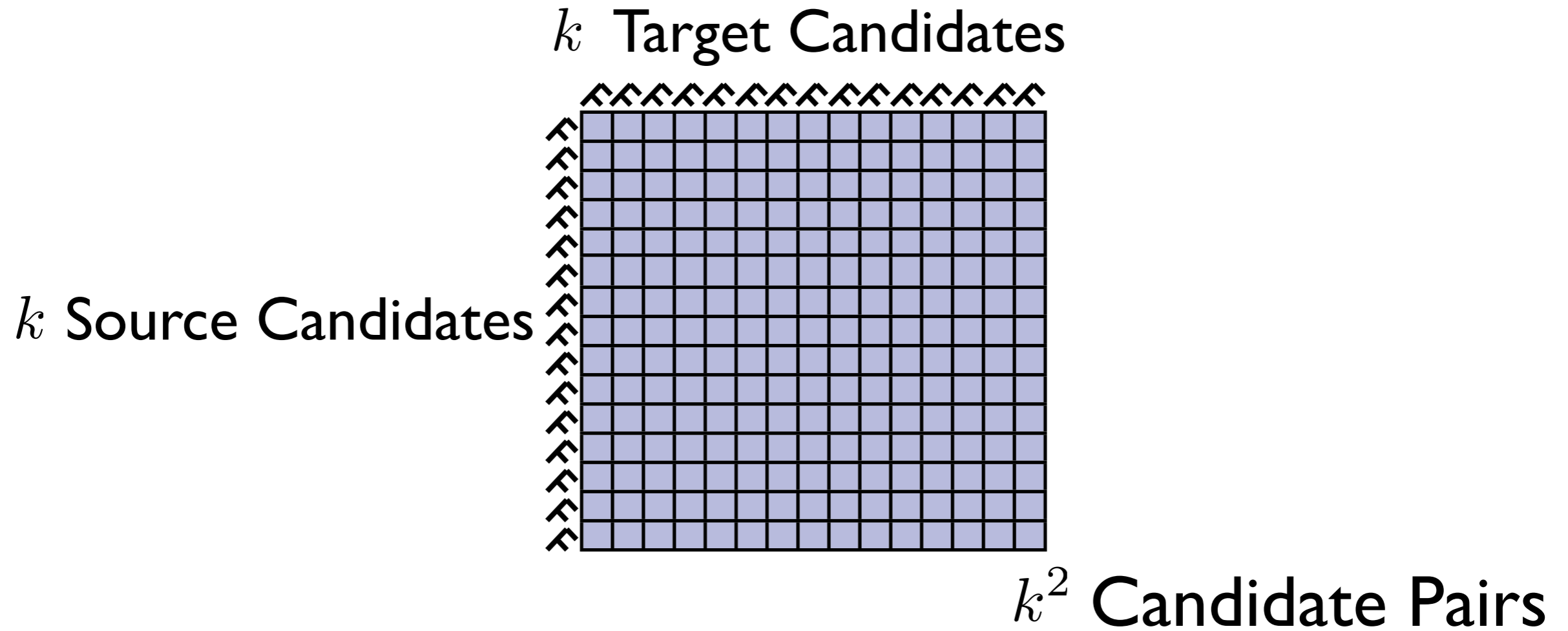
k Target Candidates



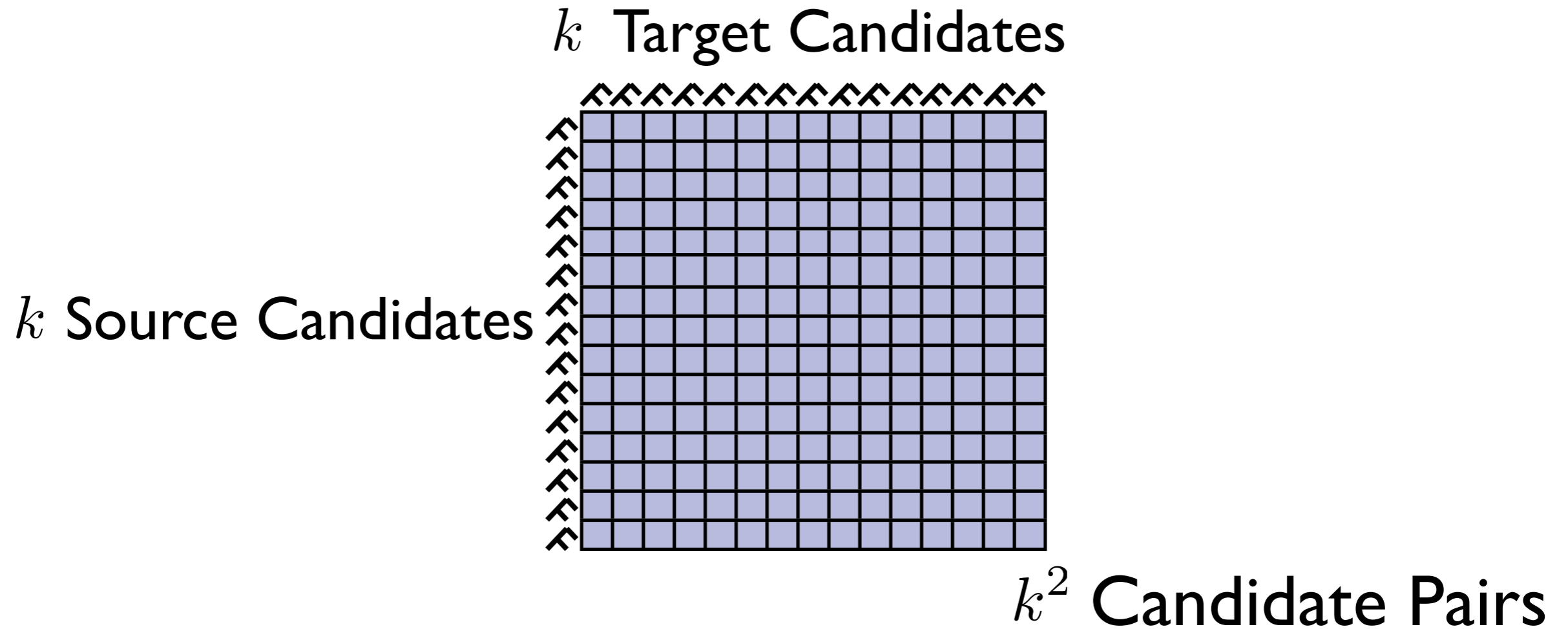
k Source Candidates



Tree Candidates

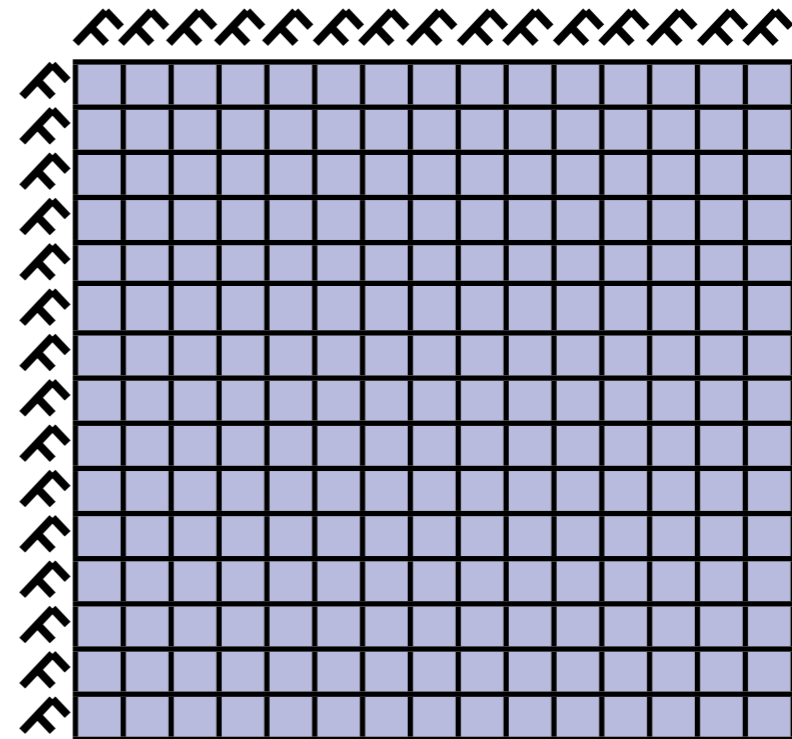


Tree Candidates



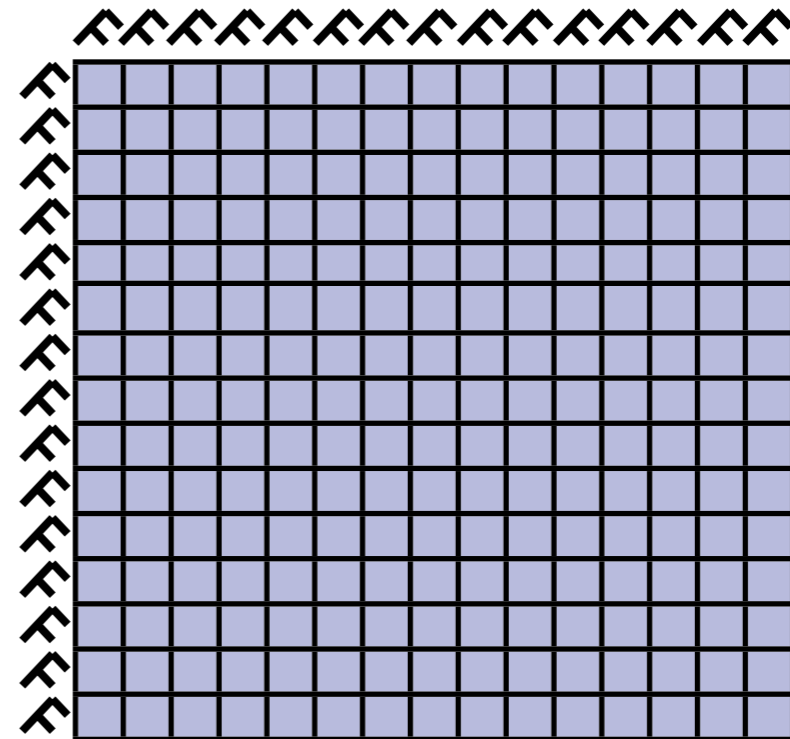
- New Problem:
 - Large number of candidate pairs

Training Set Pruning



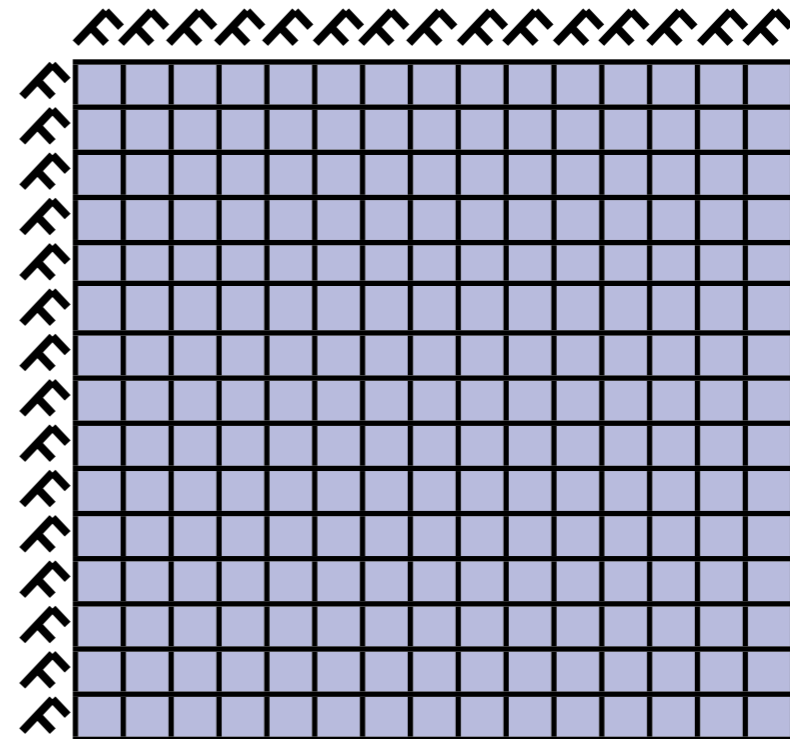
Training Set Pruning

- Solution: Pruning the candidate lists



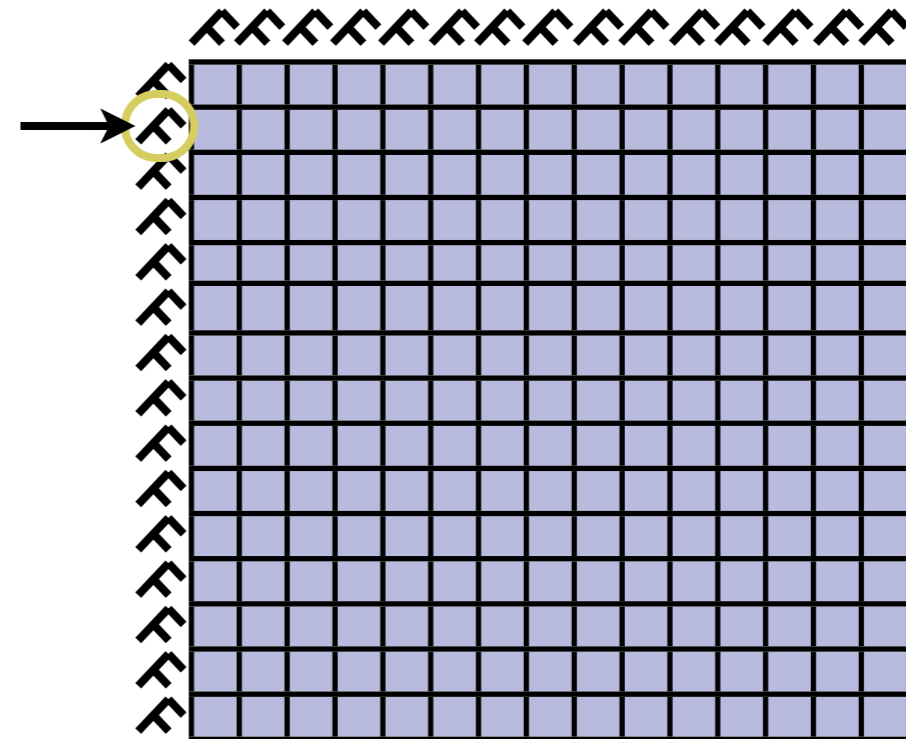
Training Set Pruning

- Solution: Pruning the candidate lists
- Start with ranked candidates from baseline parser



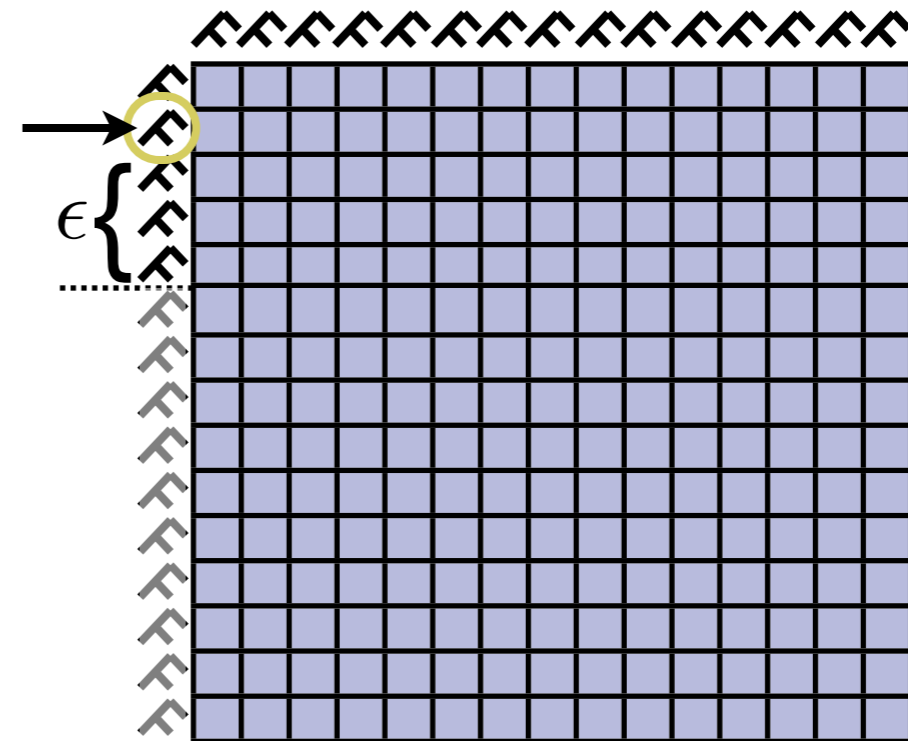
Training Set Pruning

- Solution: Pruning the candidate lists
- Start with ranked candidates from baseline parser
- Find gold tree



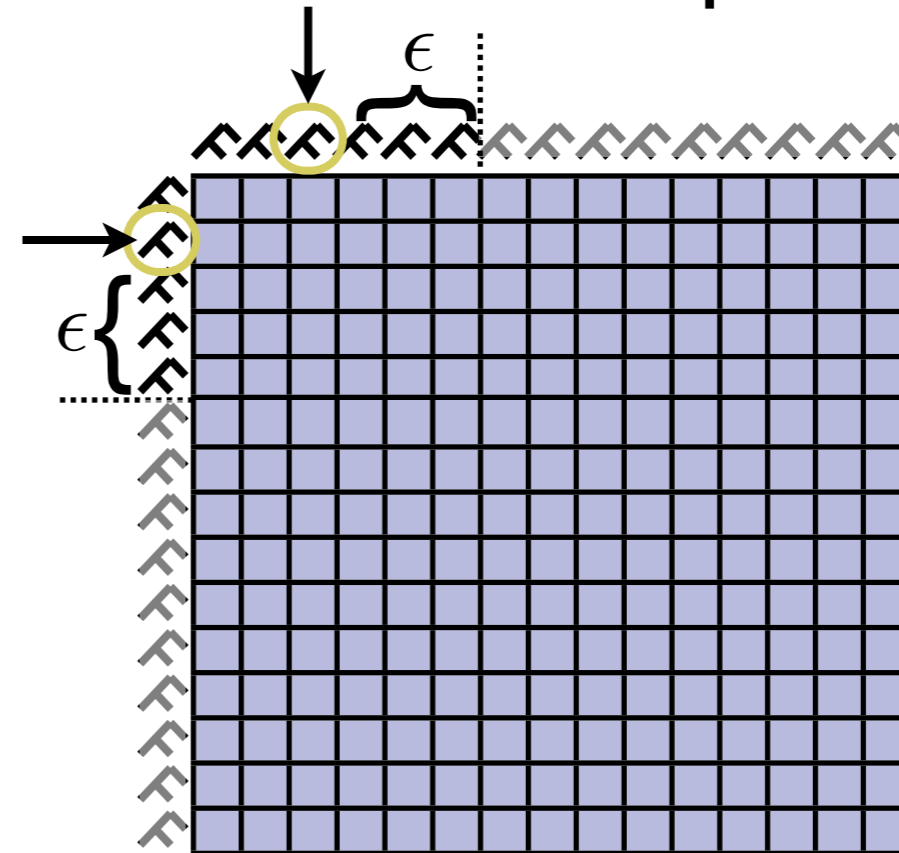
Training Set Pruning

- Solution: Pruning the candidate lists
- Start with ranked candidates from baseline parser
- Find gold tree
- Add ϵ more



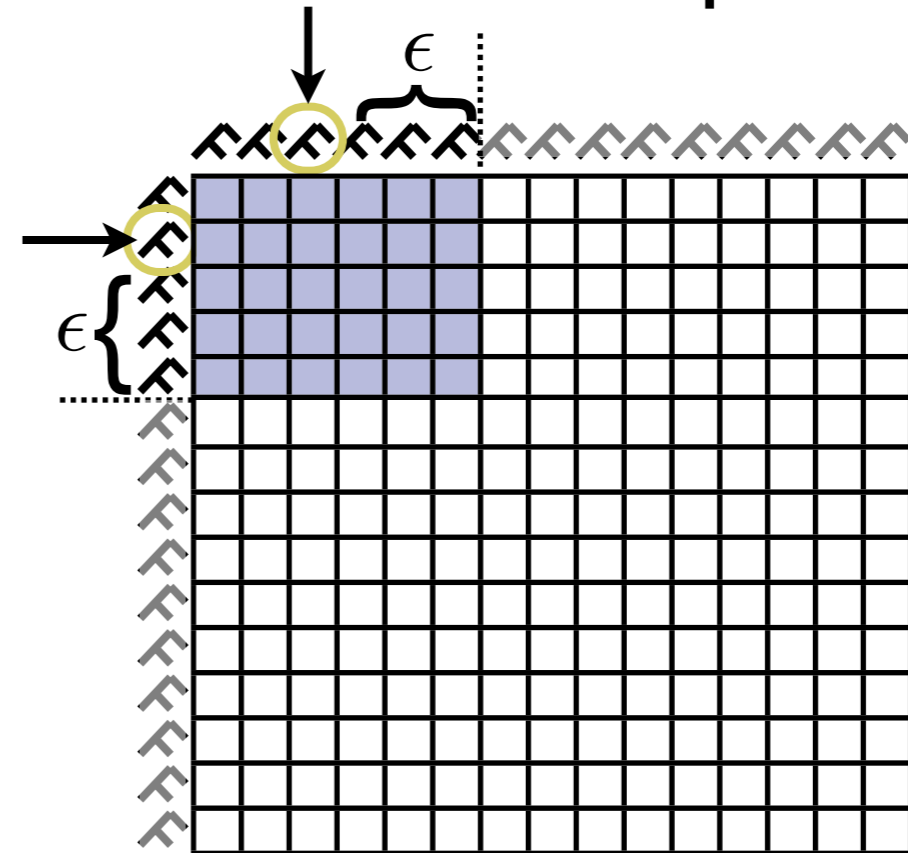
Training Set Pruning

- Solution: Pruning the candidate lists
- Start with ranked candidates from baseline parser
- Find gold tree
- Add ϵ more
- Do same on target side



Training Set Pruning

- Solution: Pruning the candidate lists
- Start with ranked candidates from baseline parser
- Find gold tree
- Add ϵ more
- Do same on target side
- Restrict sum to tree pairs in intersection





Approximating Sum over Alignments

Approximating Sum over Alignments

- Problem:

$$P(t, t' | s, s') \approx \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \sum_a \exp(w^\top \phi(t, a, t'))}$$

Approximating Sum over Alignments

- Problem:

$$P(t, t' | s, s') \approx \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \sum_a \exp(w^\top \phi(t, a, t'))}$$

- Solution:
 - Replace summation with maximization

Approximating Sum over Alignments

- Problem:

$$P(t, t' | s, s') \approx \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \sum_a \exp(w^\top \phi(t, a, t'))}$$

- Solution:

- Replace summation with maximization

$$P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}$$



Iterative Training

Iterative Training

- Problem: How to solve?

$$P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}$$

Iterative Training

- **Problem: How to solve?**

$$P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}$$

- **Solution: Iterative Procedure**

Iterative Training

- **Problem: How to solve?**

$$P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}$$

- **Solution: Iterative Procedure**

I. Fix weights, optimize alignments

$$a_{(t, t')}^{(\tau)} = \arg \max_a w^{(\tau)\top} \phi(t, a, t')$$

Iterative Training

- **Problem: How to solve?**

$$P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}$$

- **Solution: Iterative Procedure**

I. Fix weights, optimize alignments

Bipartite
Matching

$$a_{(t, t')}^{(\tau)} = \arg \max_a w^{(\tau)\top} \phi(t, a, t')$$

Iterative Training

- **Problem: How to solve?**

$$P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}$$

- **Solution: Iterative Procedure**

1. Fix weights, optimize alignments

Bipartite
Matching

$$a_{(t, t')}^{(\tau)} = \arg \max_a w^{(\tau)\top} \phi(t, a, t')$$

2. Fix alignments, optimize weights

$$w^{(\tau+1)} = \arg \max_w \frac{\exp(w^\top \phi(t, a_{(t, t')}^{(\tau)}, t'))}{\sum_{(t, t') \in (T, T')} \exp(w^\top \phi(t, a_{(t, t')}^{(\tau)}, t'))}$$

Iterative Training

- **Problem: How to solve?**

$$P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}$$

- **Solution: Iterative Procedure**

1. Fix weights, optimize alignments

Bipartite
Matching

$$a_{(t, t')}^{(\tau)} = \arg \max_a w^{(\tau)\top} \phi(t, a, t')$$

2. Fix alignments, optimize weights

Convex
Optimization

$$w^{(\tau+1)} = \arg \max_w \frac{\exp(w^\top \phi(t, a_{(t, t')}^{(\tau)}, t'))}{\sum_{(t, t') \in (T, T')} \exp(w^\top \phi(t, a_{(t, t')}^{(\tau)}, t'))}$$

Overview

- A log-linear model over aligned parse trees
- Training with latent tree alignments
- Improvements from joint parsing



Parsing Experimental Setup

Chinese Treebank

--	--	--	--

--	--	--

English Translations

Penn Treebank

--	--	--	--

Parsing Experimental Setup

Chinese Treebank



English Translations

Penn Treebank



Baseline Parser Training

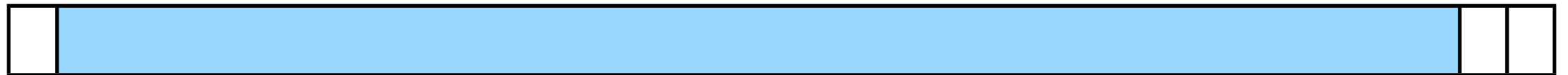
Parsing Experimental Setup

Chinese Treebank



English Translations

Penn Treebank



Baseline Parser Training

Bilingual Reranker Training

Parsing Experimental Setup

Chinese Treebank



English Translations

Penn Treebank

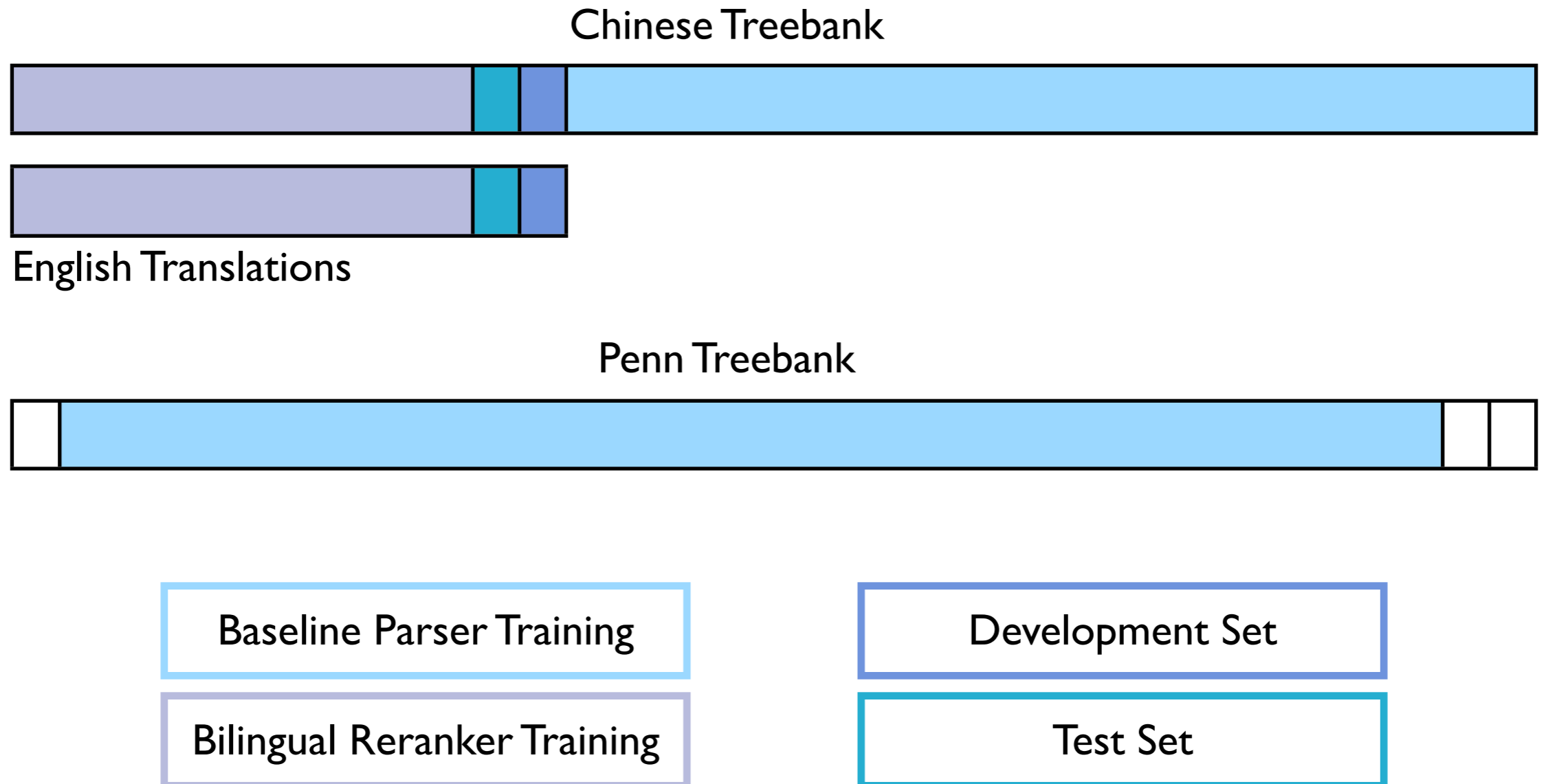


Baseline Parser Training

Development Set

Bilingual Reranker Training

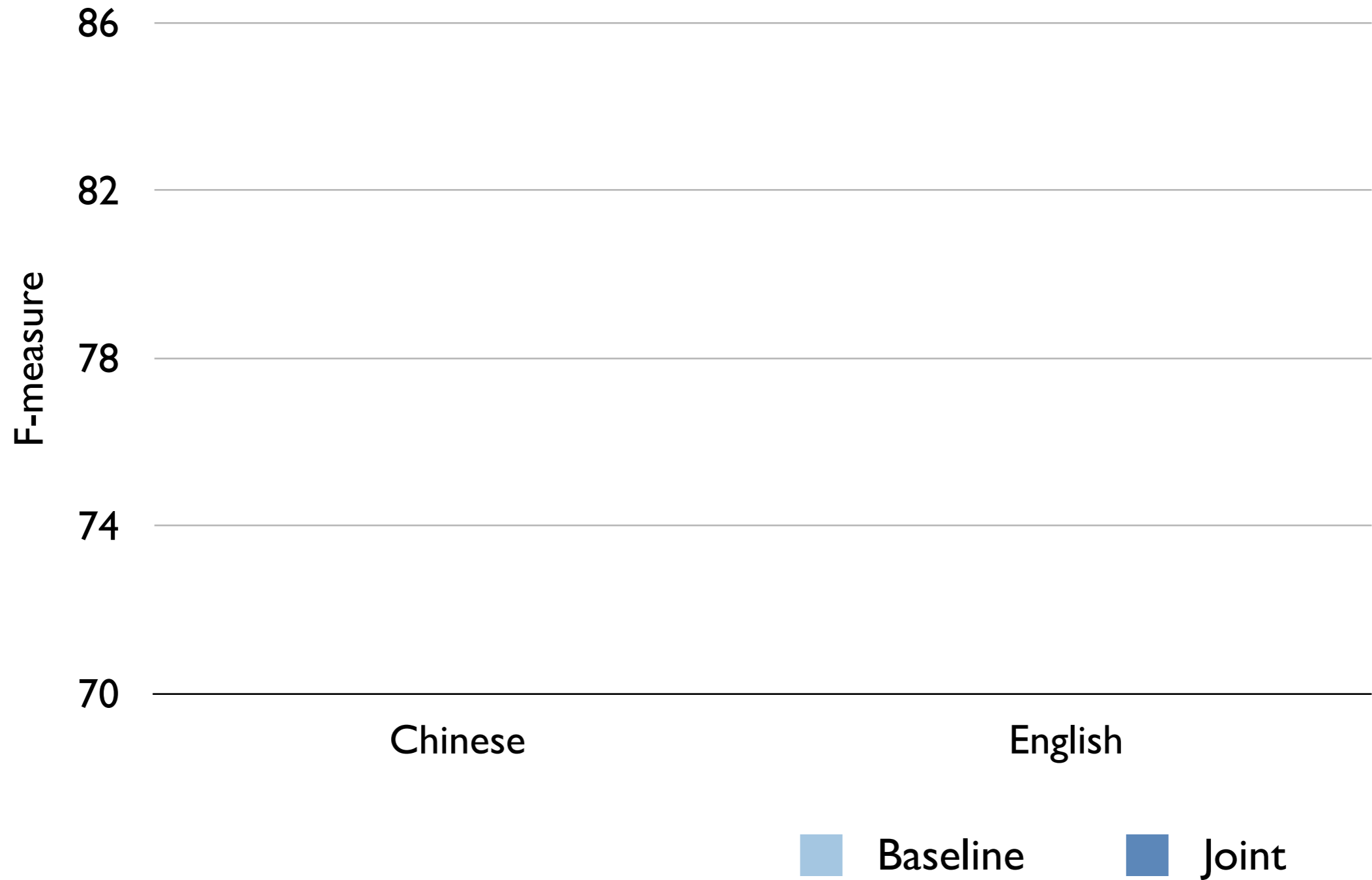
Parsing Experimental Setup



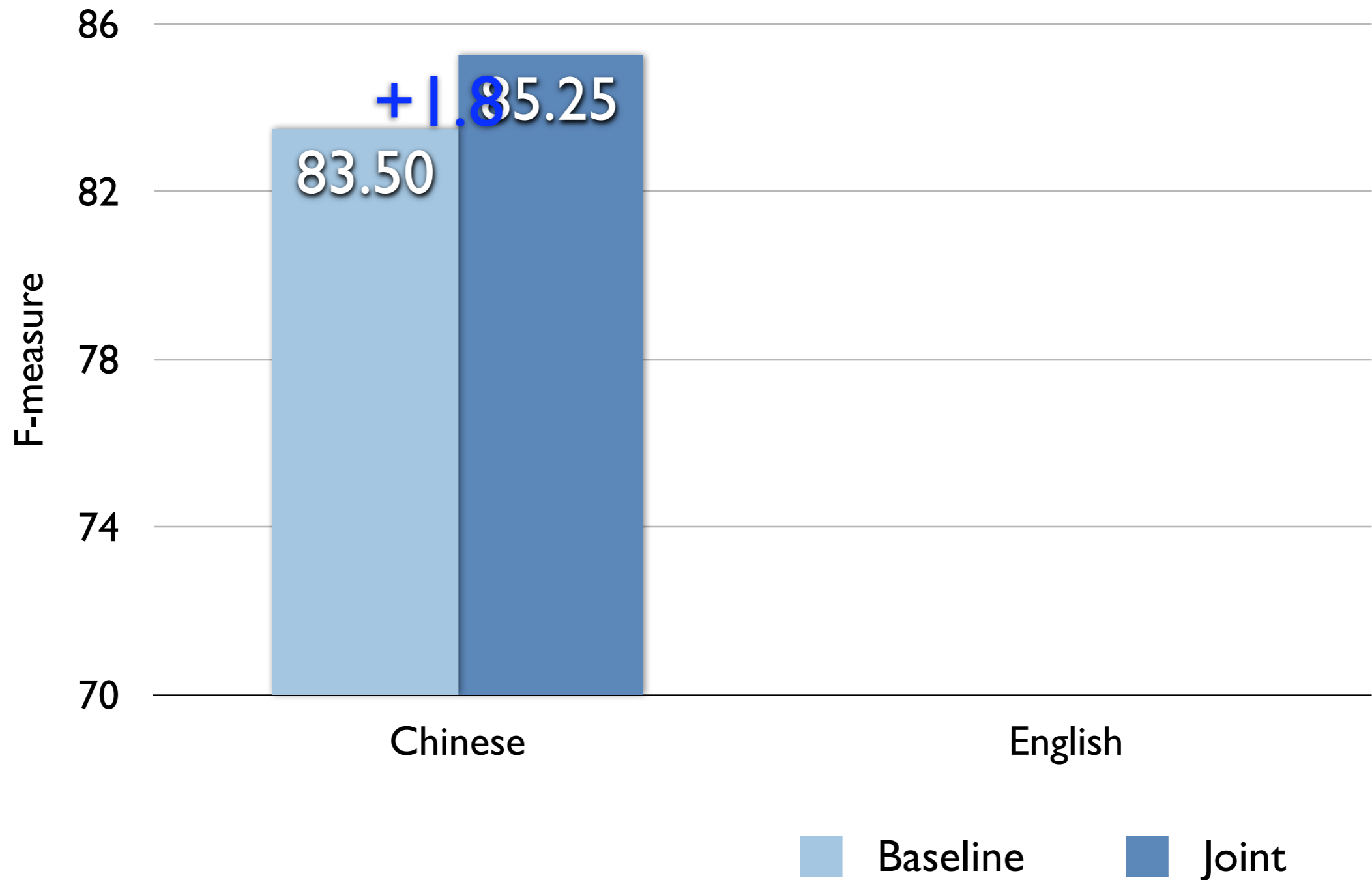
Parsing Results

■ Baseline ■ Joint

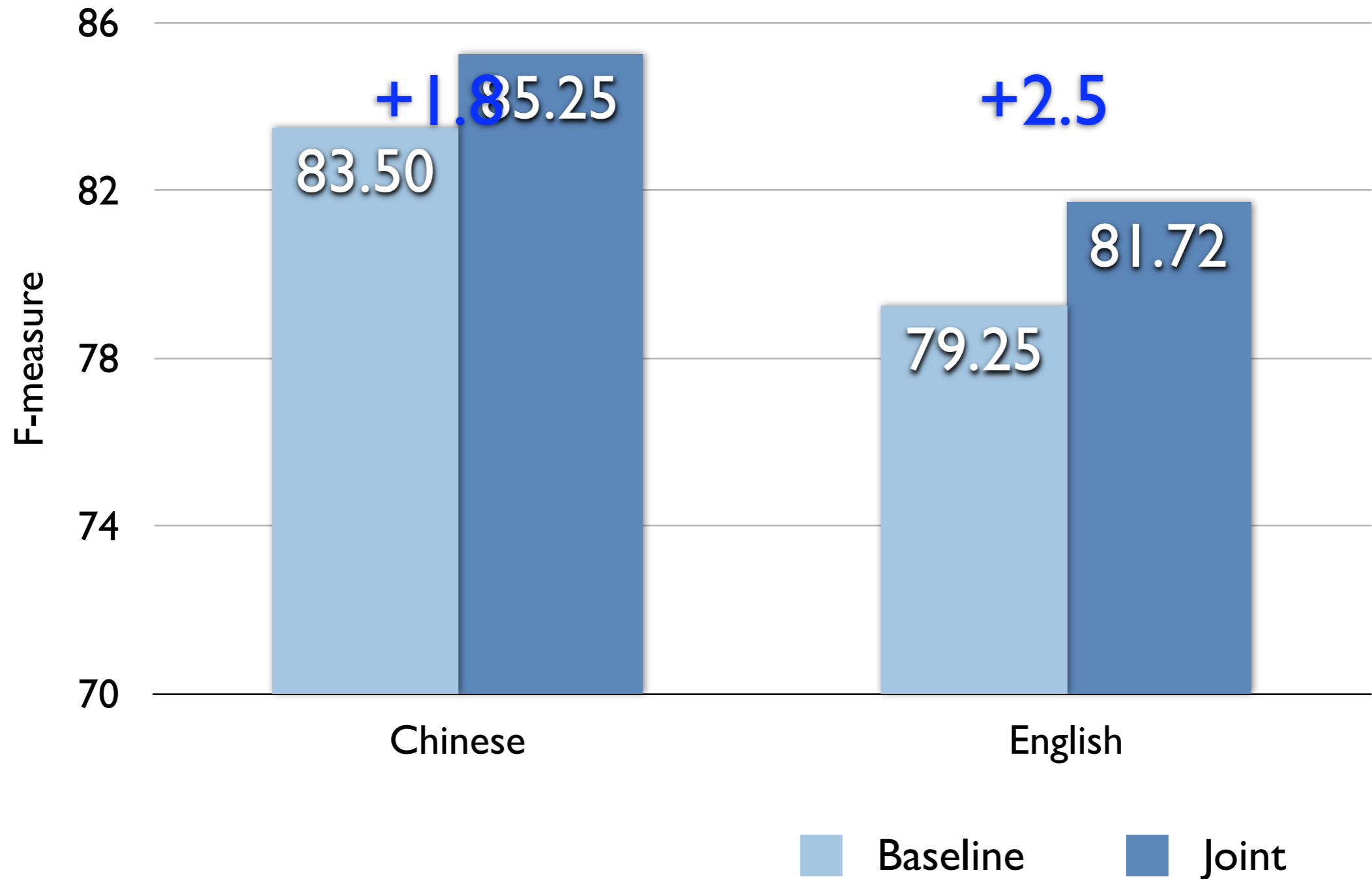
Parsing Results



Parsing Results



Parsing Results





Machine Translation Experimental Setup

Machine Translation Experimental Setup

- **Syntax-based MT:**
Translation rules derived from target-side trees
(Galley et al, 2004; 2006)

Machine Translation Experimental Setup

- **Syntax-based MT:**
 - Translation rules derived from target-side trees
(Galley et al, 2004; 2006)
- **Comparison:**
 - Target trees from monolingual parser
vs
Target trees from joint parser

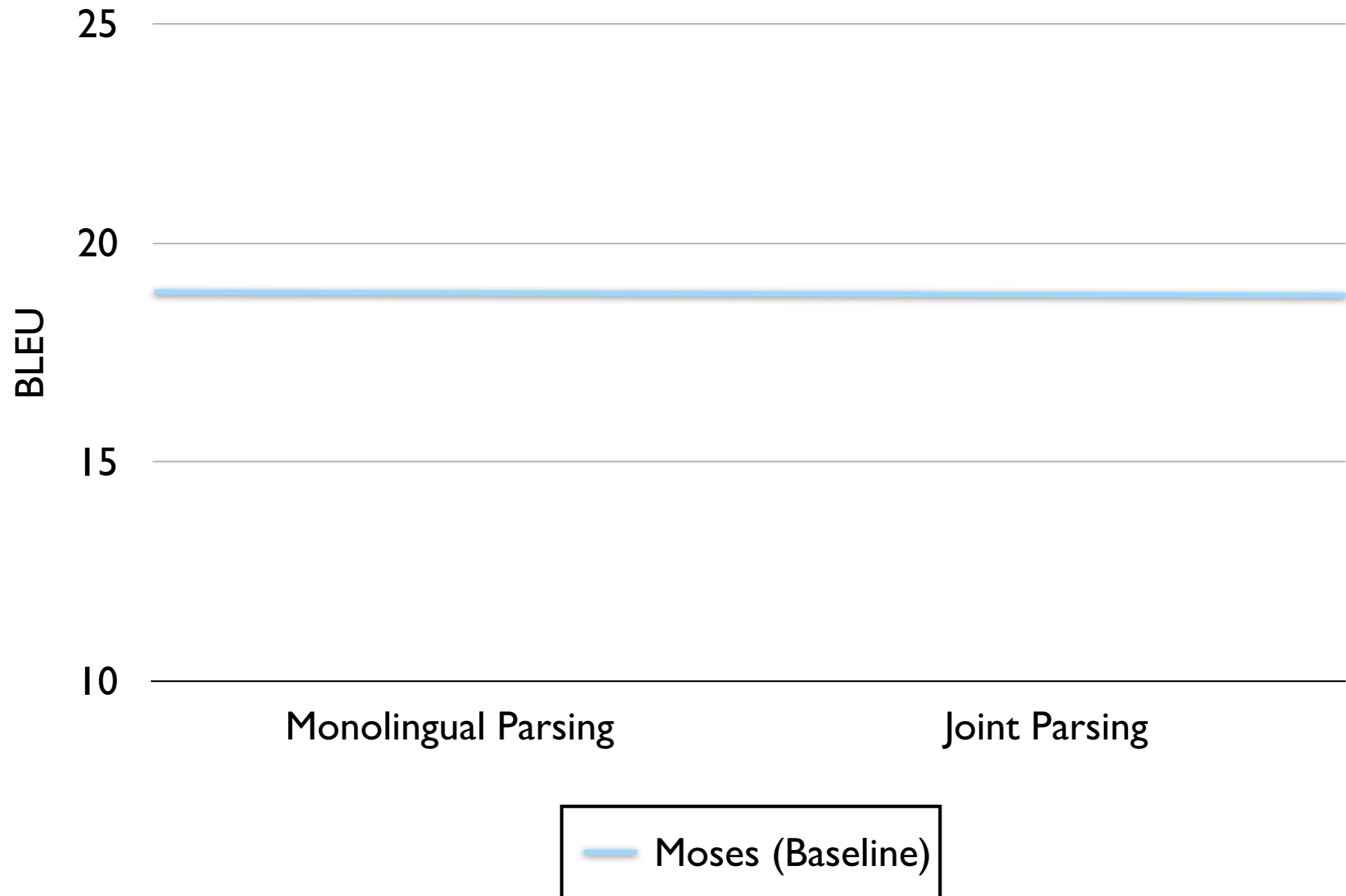
Machine Translation Experimental Setup

- **Syntax-based MT:**
 - Translation rules derived from target-side trees
(Galley et al, 2004; 2006)
- **Comparison:**
 - Target trees from monolingual parser
vs
Target trees from joint parser
- **Remaining data is identical**

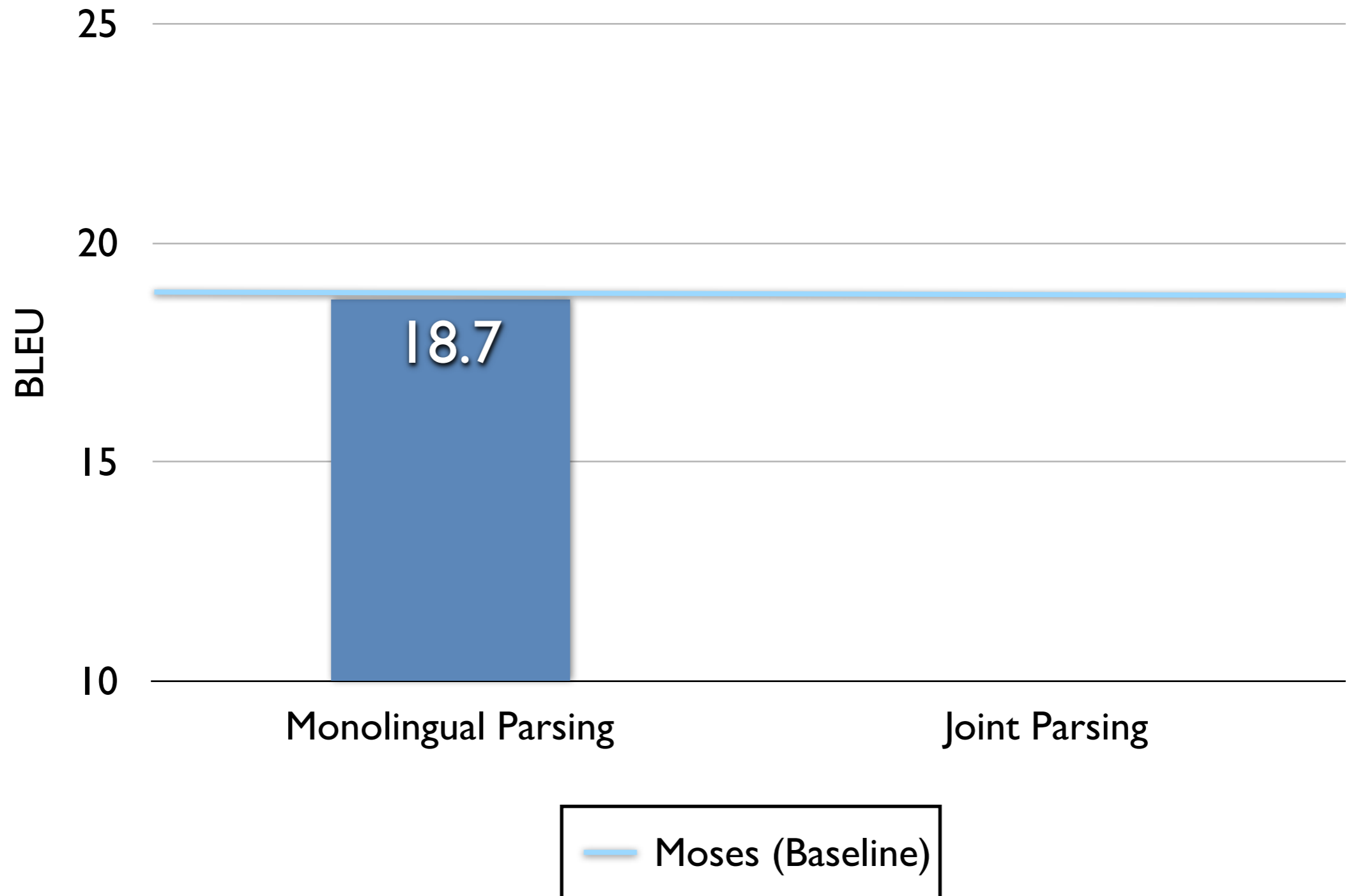
Machine Translation Results

— Moses (Baseline)

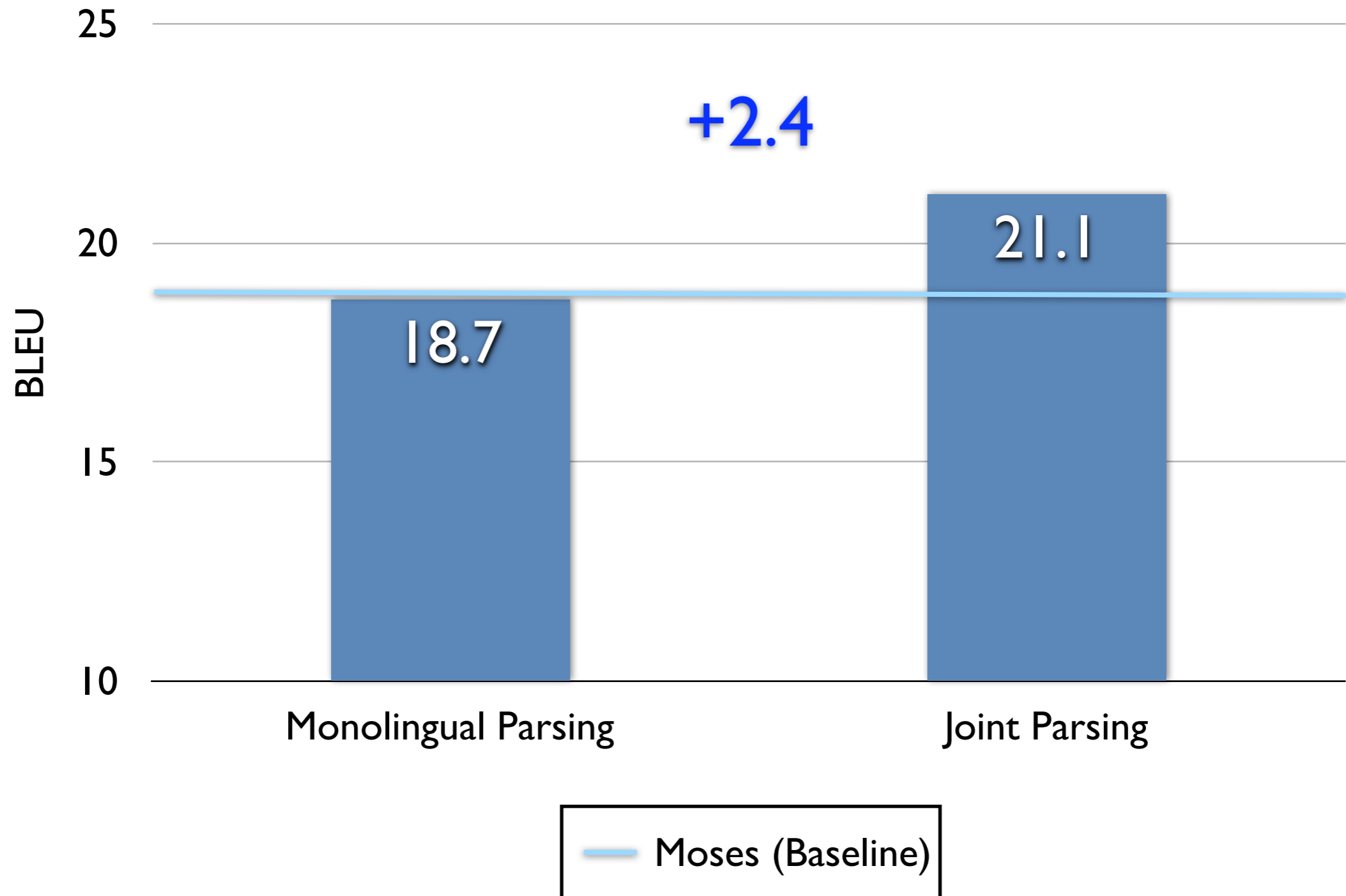
Machine Translation Results



Machine Translation Results



Machine Translation Results





Conclusions

Conclusions

- Bilingual constraints can be leveraged to improve parse quality

Conclusions

- Bilingual constraints can be leveraged to improve parse quality
- Parsing gets better
 - In-domain Chinese F_1 improves by 1.8
 - Out-of-domain English F_1 improves by 2.5

Conclusions

- Bilingual constraints can be leveraged to improve parse quality
- Parsing gets better
 - In-domain Chinese F_1 improves by 1.8
 - Out-of-domain English F_1 improves by 2.5
- Downstream machine translation gets better
 - BLEU improves by 2.4

Thank You

