

Joint Parsing and Alignment with Weakly Synchronized Grammars



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

Berkeley



Statistical Machine Translation Rules



Statistical Machine Translation Rules

Input

Whereas in 1992

而在一九九二

Xinhua News Agency

新华社

(End)

(完)

Beijing , February 13 -

北京 二月 十三日 电



Statistical Machine Translation Rules

Input

Whereas in 1992

而在一九九二

Xinhua News Agency

新华社

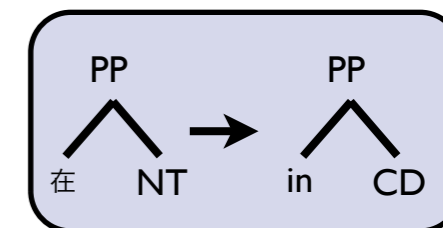
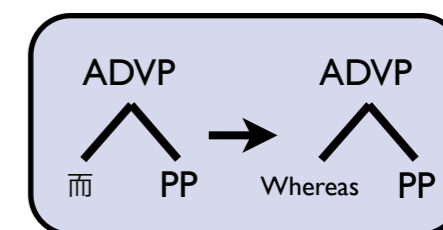
(End)

(完)

Beijing , February 13 -

北京 二月 十三日 电

Output





Statistical Machine Translation Rules

Input

Whereas in 1992

而在一九九二

Xinhua News Agency

新华社

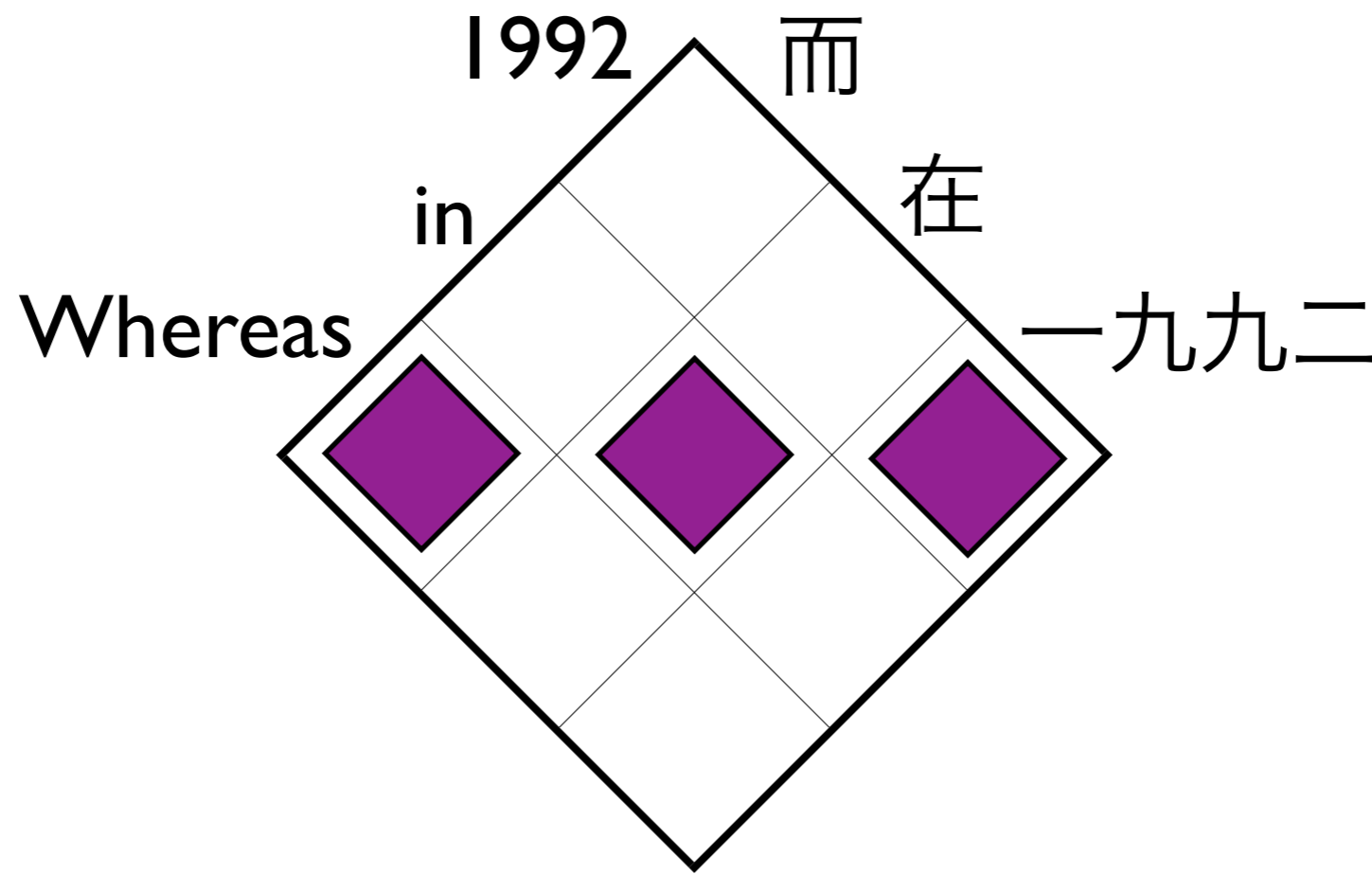
(End)

(完)

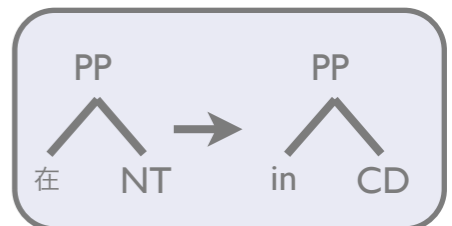
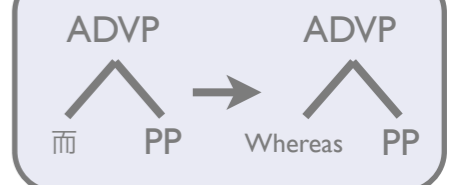
Beijing, February 13 -

北京二月十三日电

Intermediate



Output



Statistical Machine Translation Rules



Input

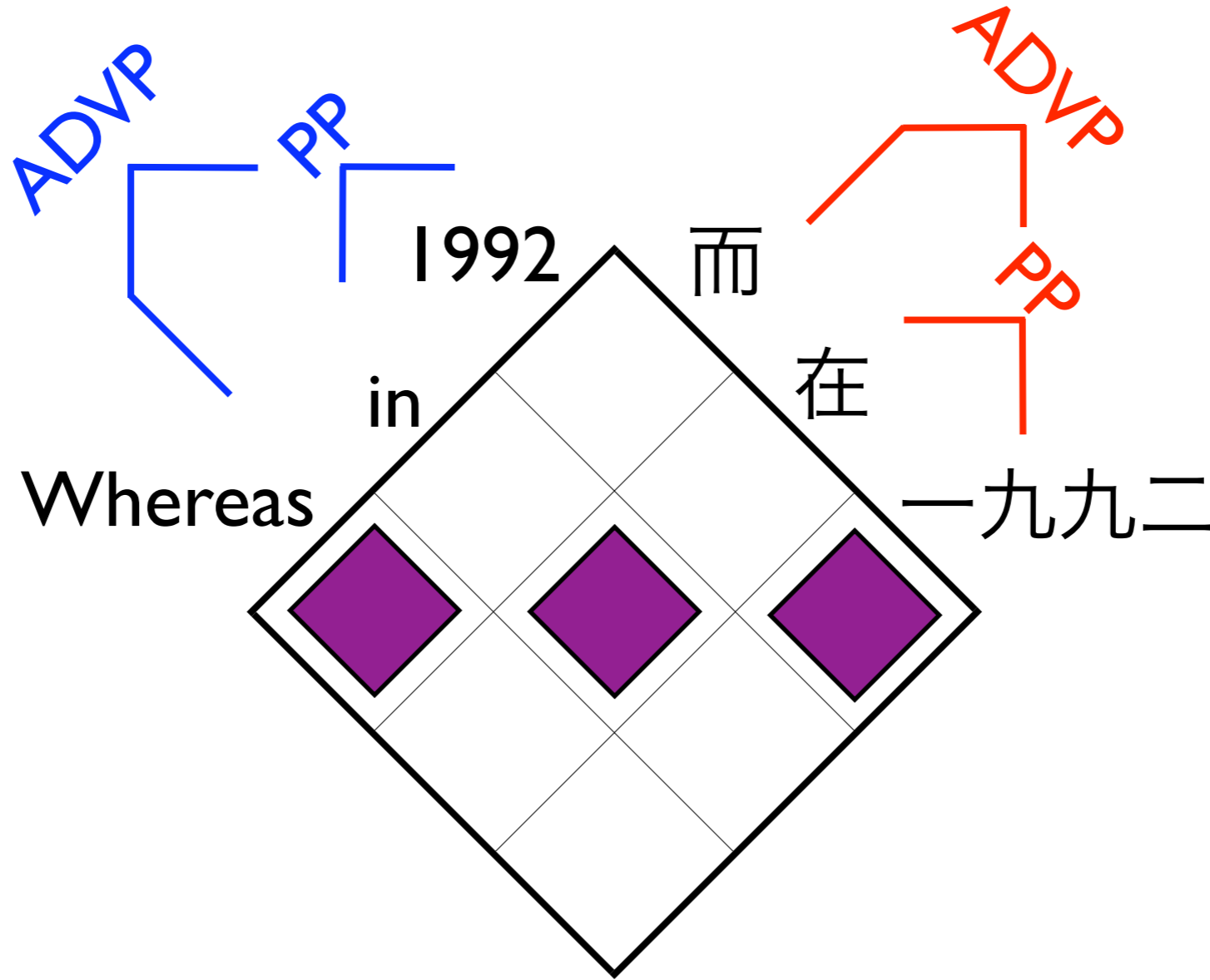
Whereas in 1992
而在一九九二

Xinhua News Agency
新华社

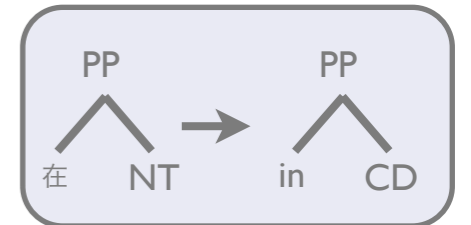
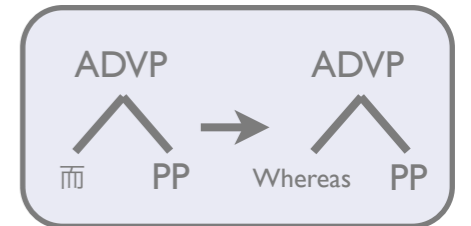
(End)
(完)

Beijing, February 13 -
北京二月十三日电

Intermediate

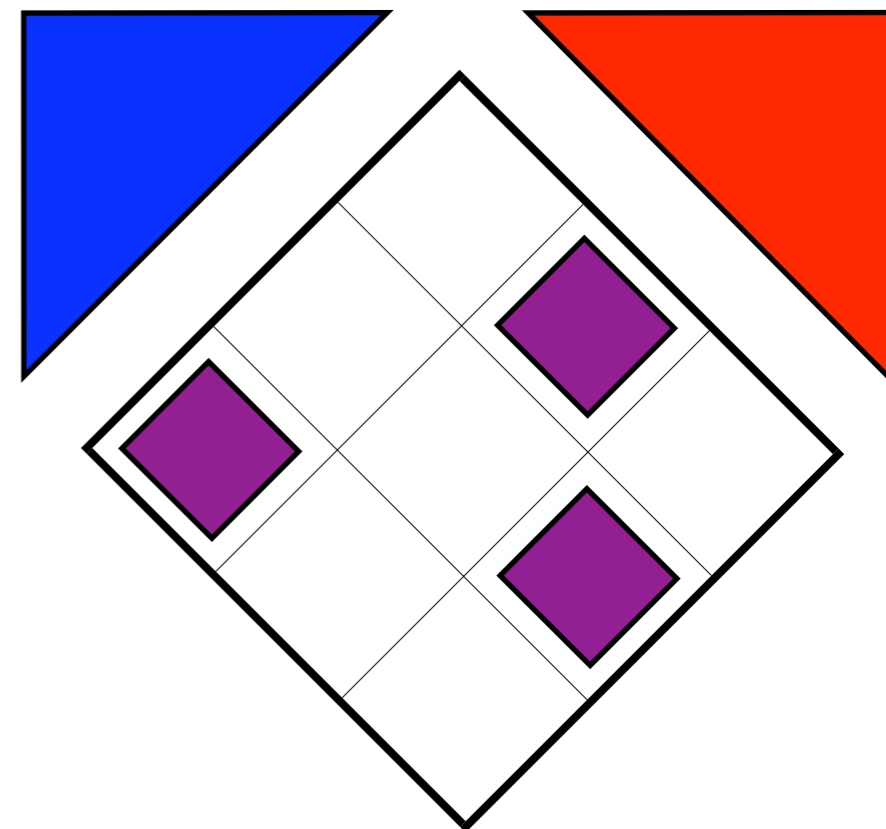


Output





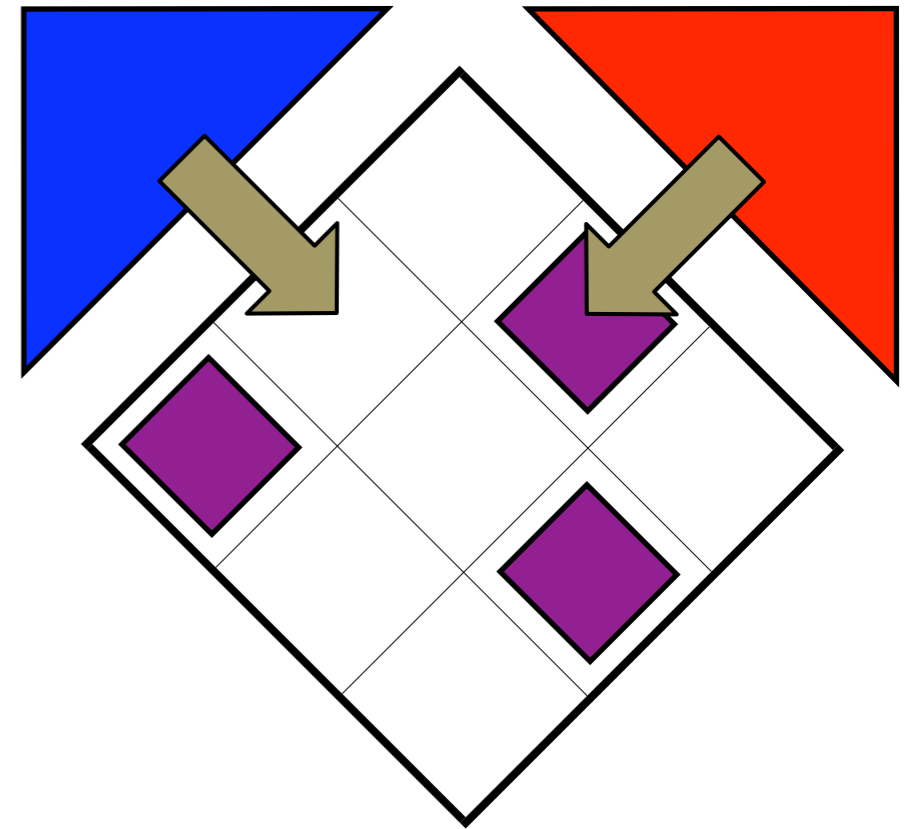
Parsing and Word Alignment





Parsing and Word Alignment

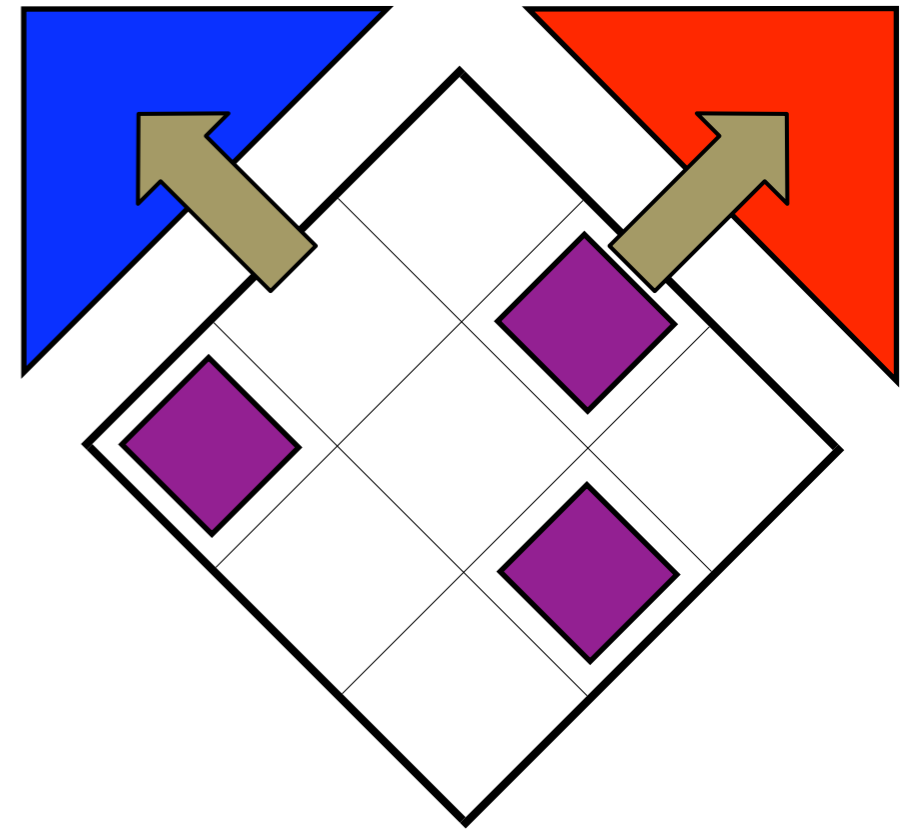
- Information about parses can improve alignment quality (DeNero and Klein, 2007; Fossum et al, 2008; Pauls et al, 2010).





Parsing and Word Alignment

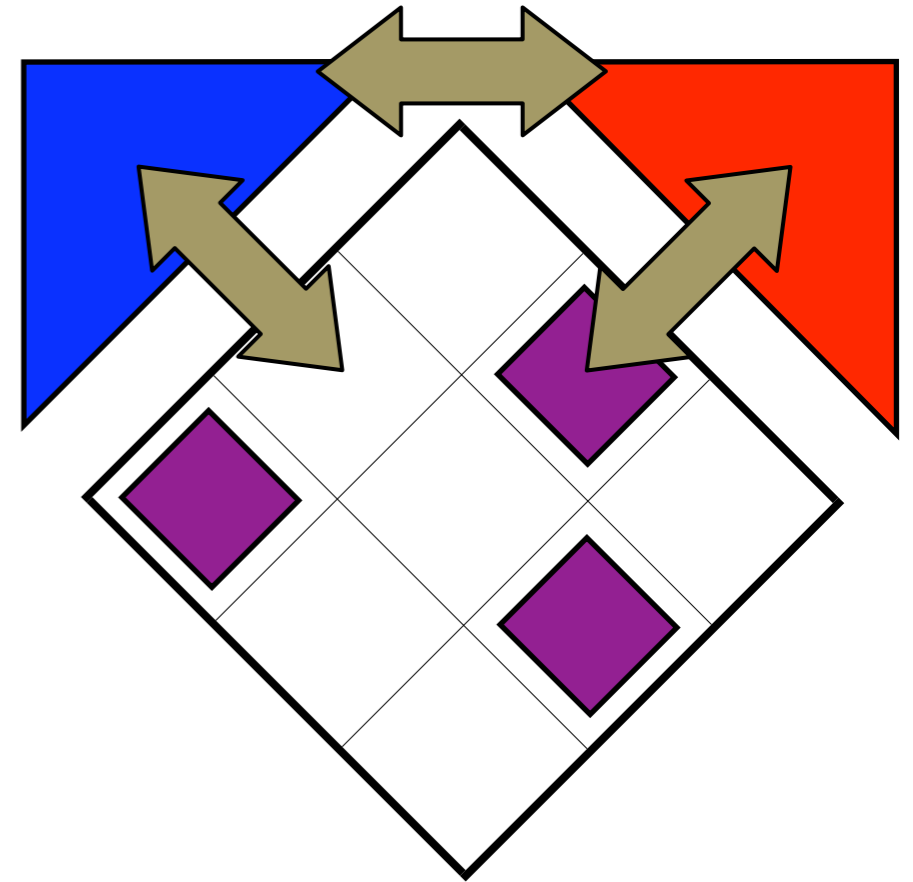
- Information about parses can improve alignment quality (DeNero and Klein, 2007; Fossum et al, 2008; Pauls et al, 2010).
- Information about alignment can improve parse quality (Smith and Smith, 2004; Burkett and Klein, 2008).





Parsing and Word Alignment

- Information about parses can improve alignment quality (DeNero and Klein, 2007; Fossum et al, 2008; Pauls et al, 2010).
- Information about alignment can improve parse quality (Smith and Smith, 2004; Burkett and Klein, 2008).
- This talk: performing both tasks jointly.



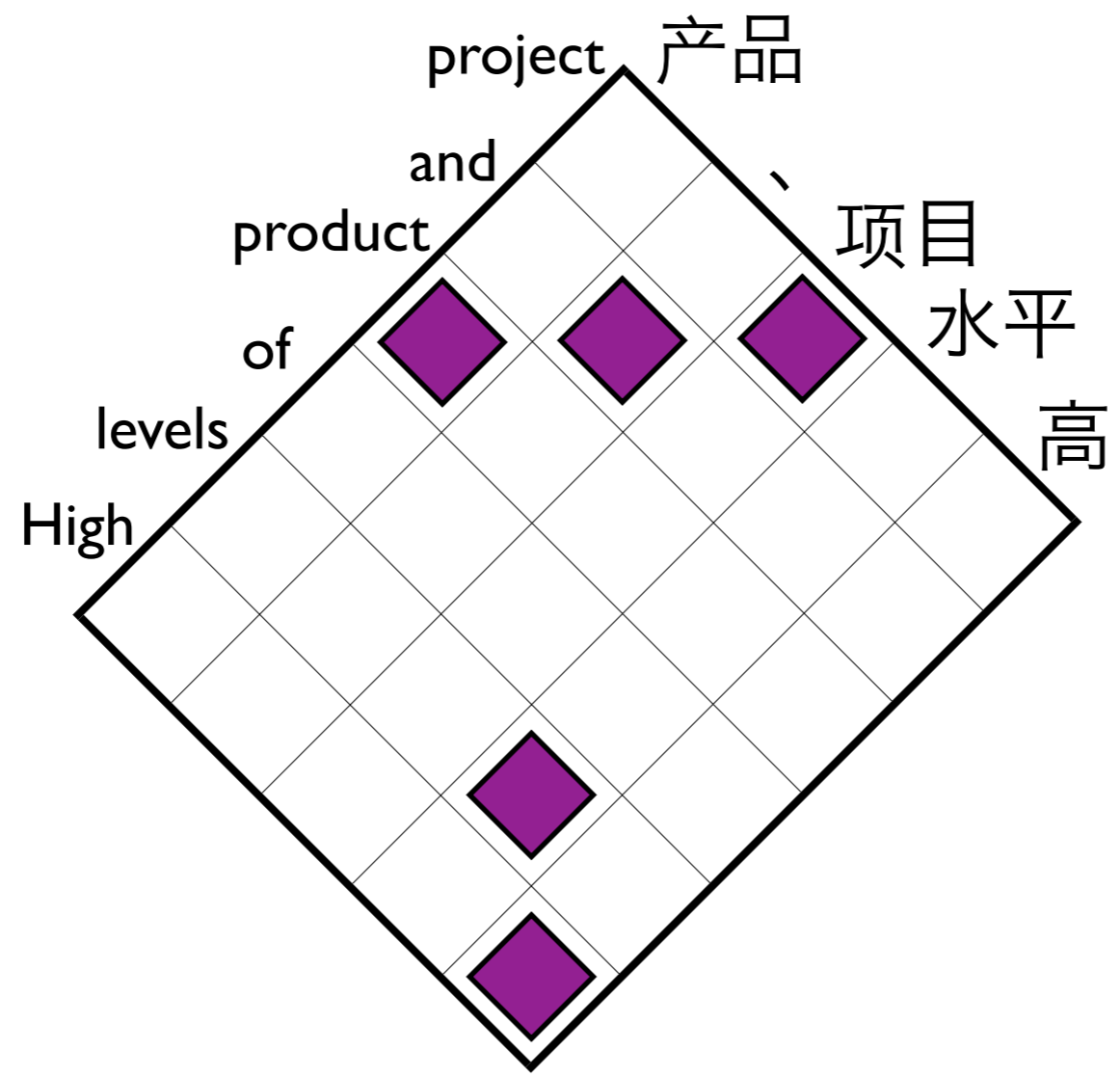


Weak Synchronization

High levels of product and project synchronization、项目水平高

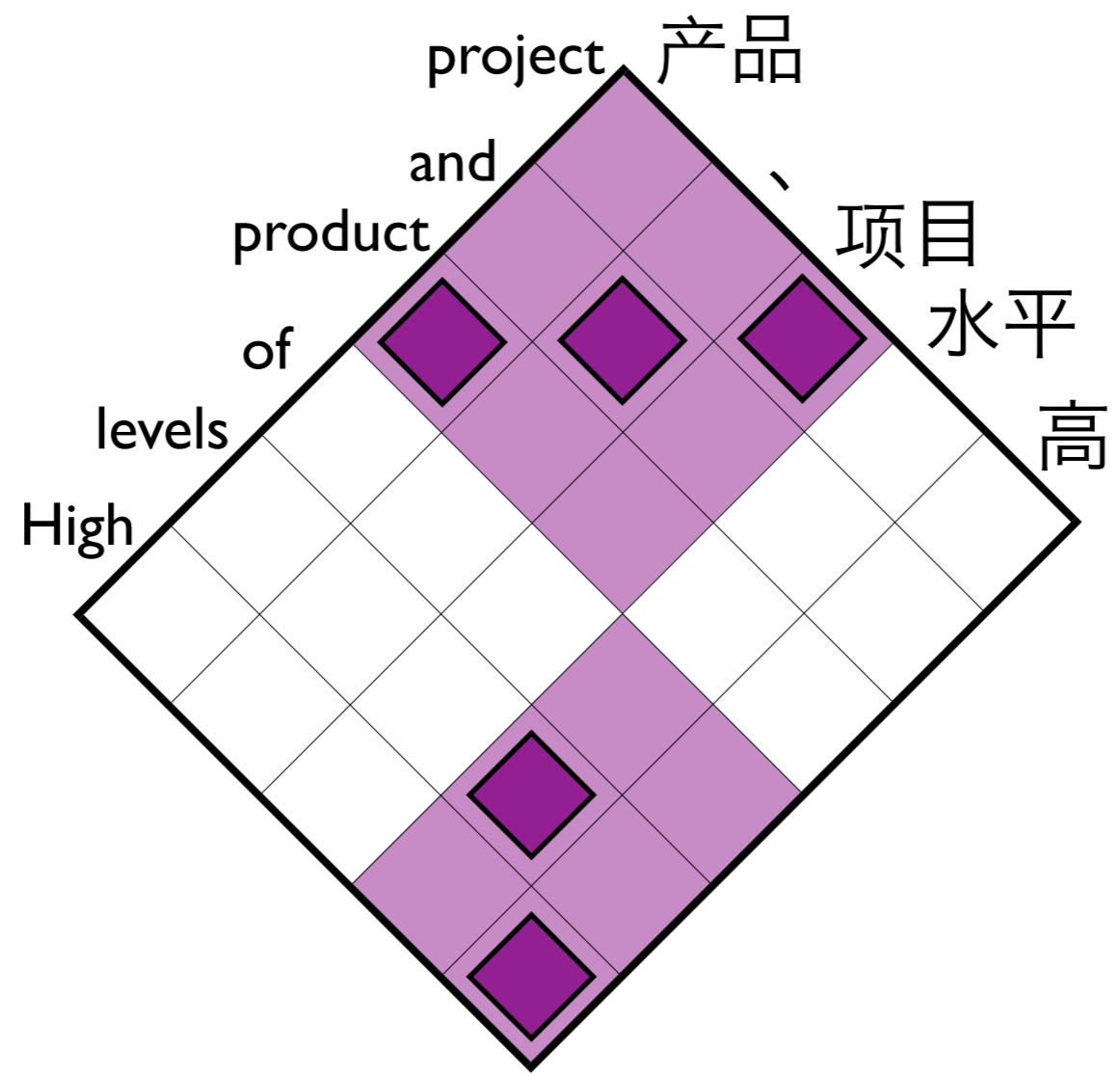


Weak Synchronization



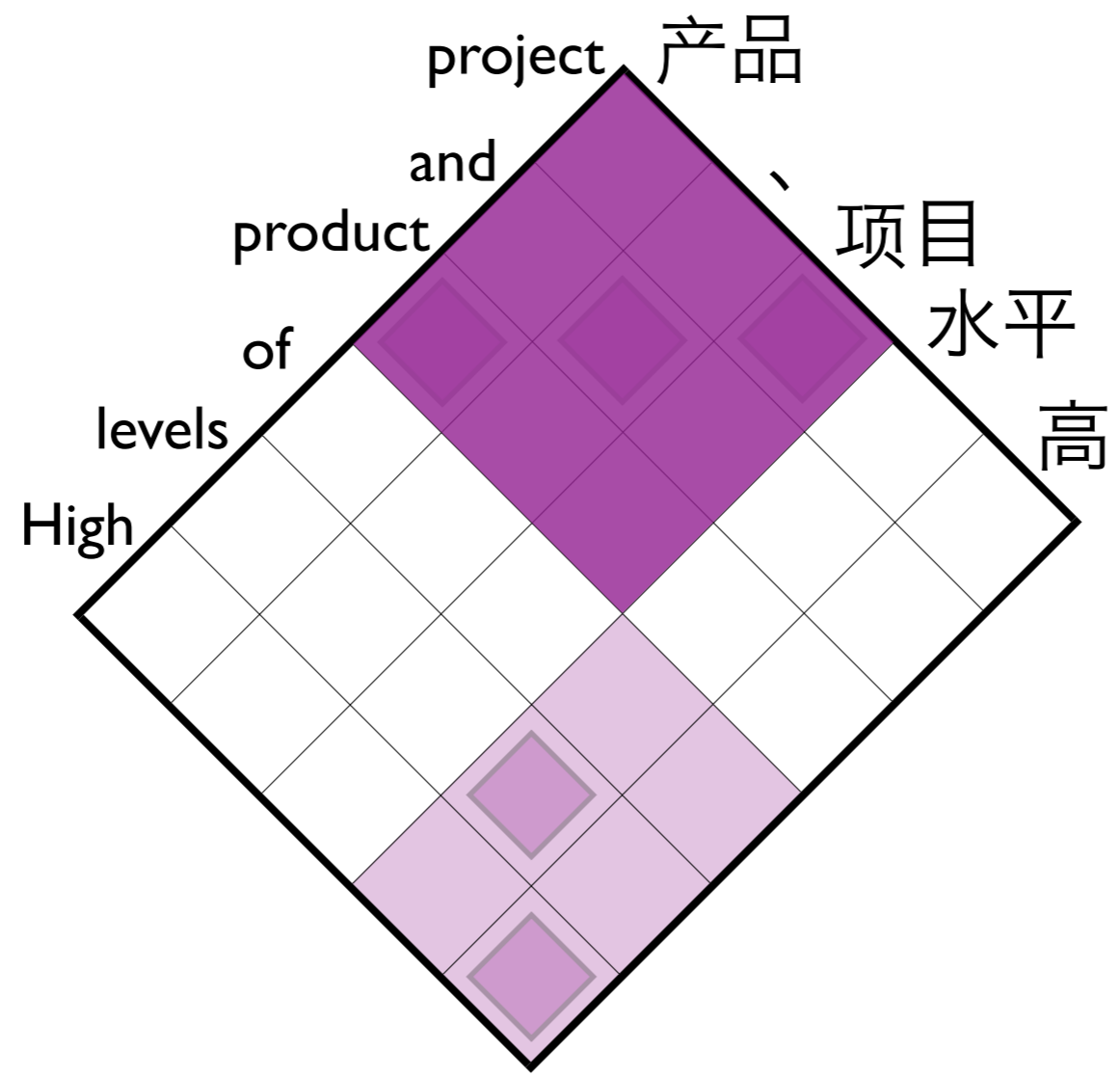


Weak Synchronization



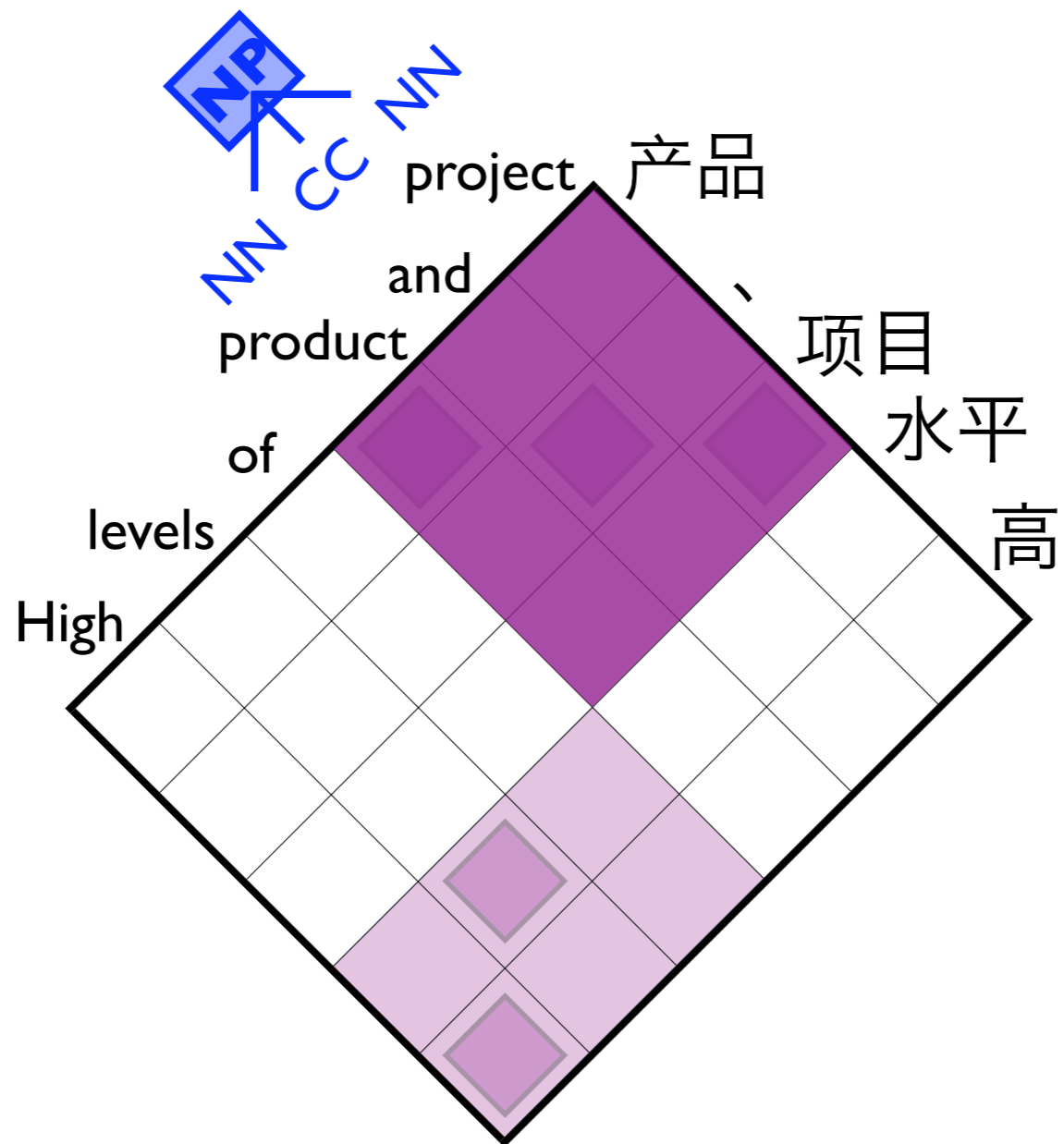


Weak Synchronization



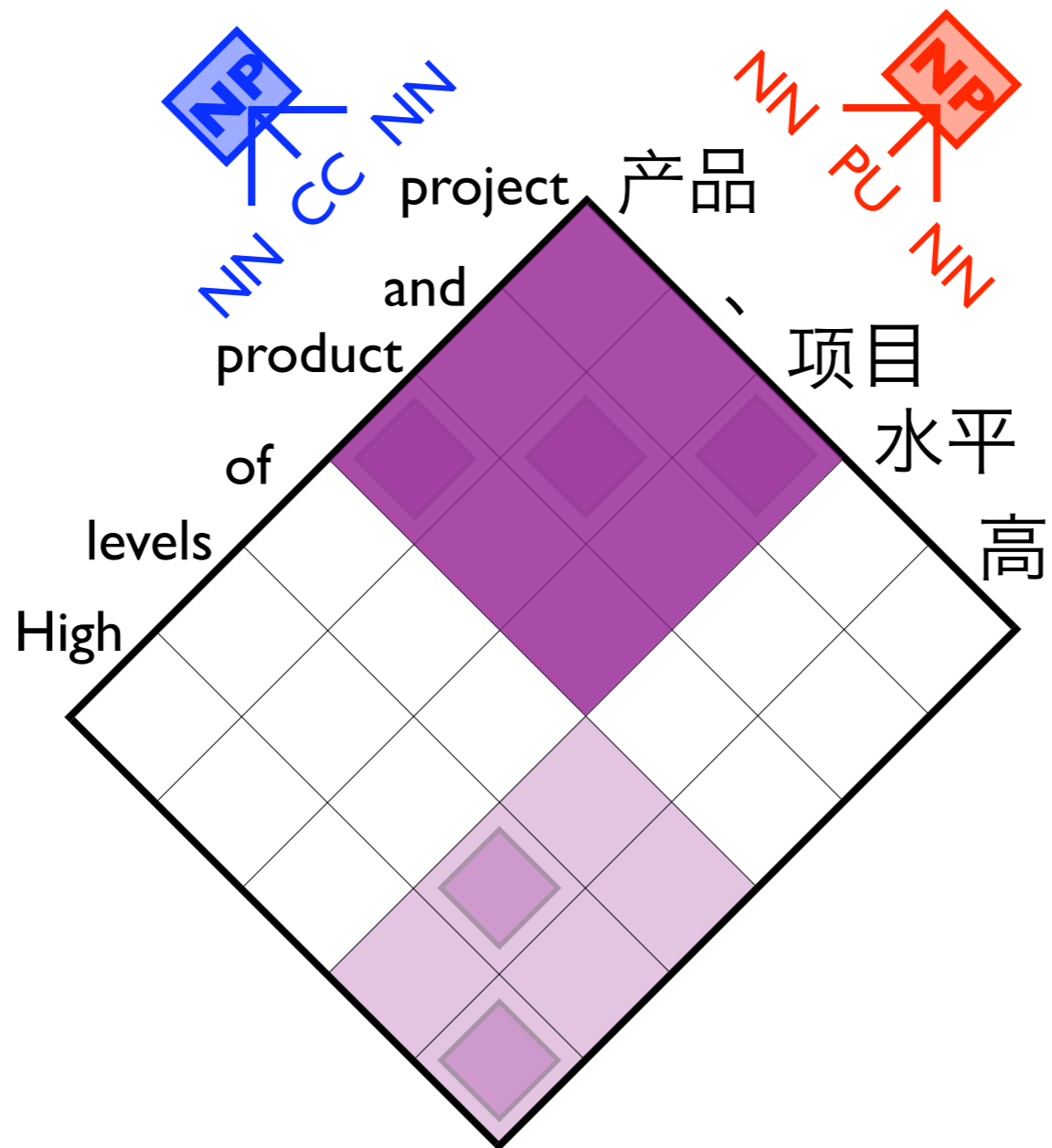


Weak Synchronization



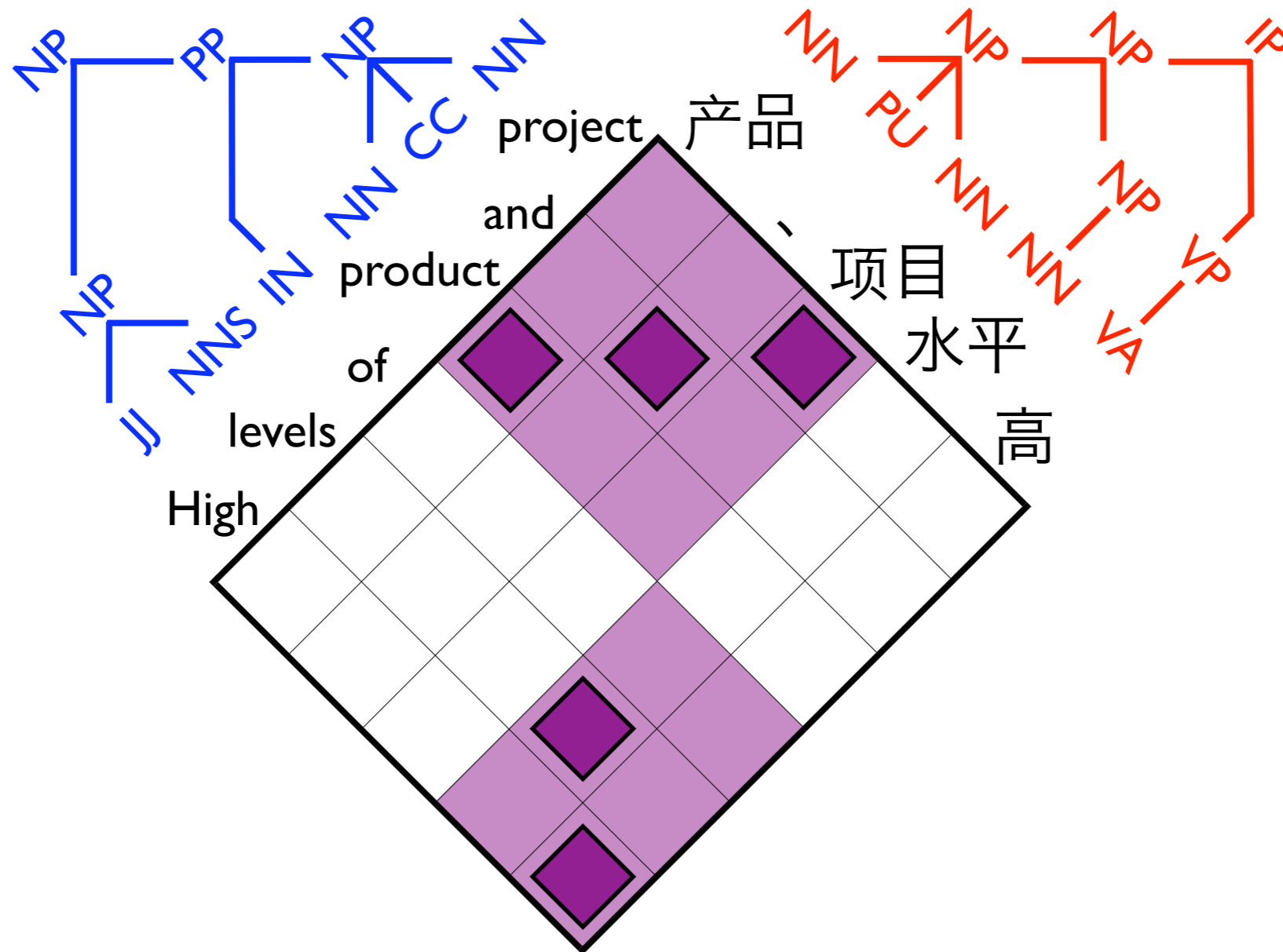


Weak Synchronization



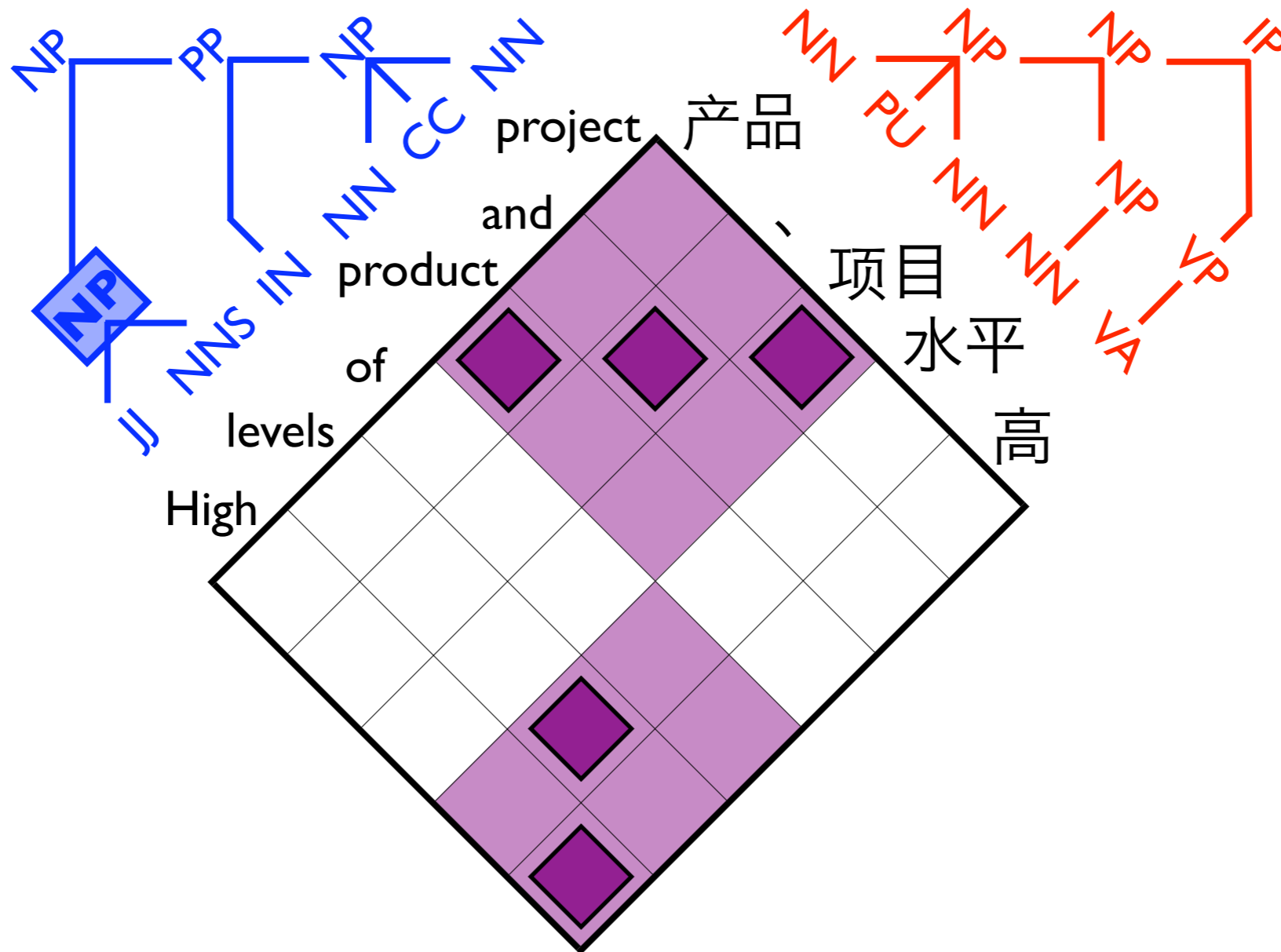


Weak Synchronization



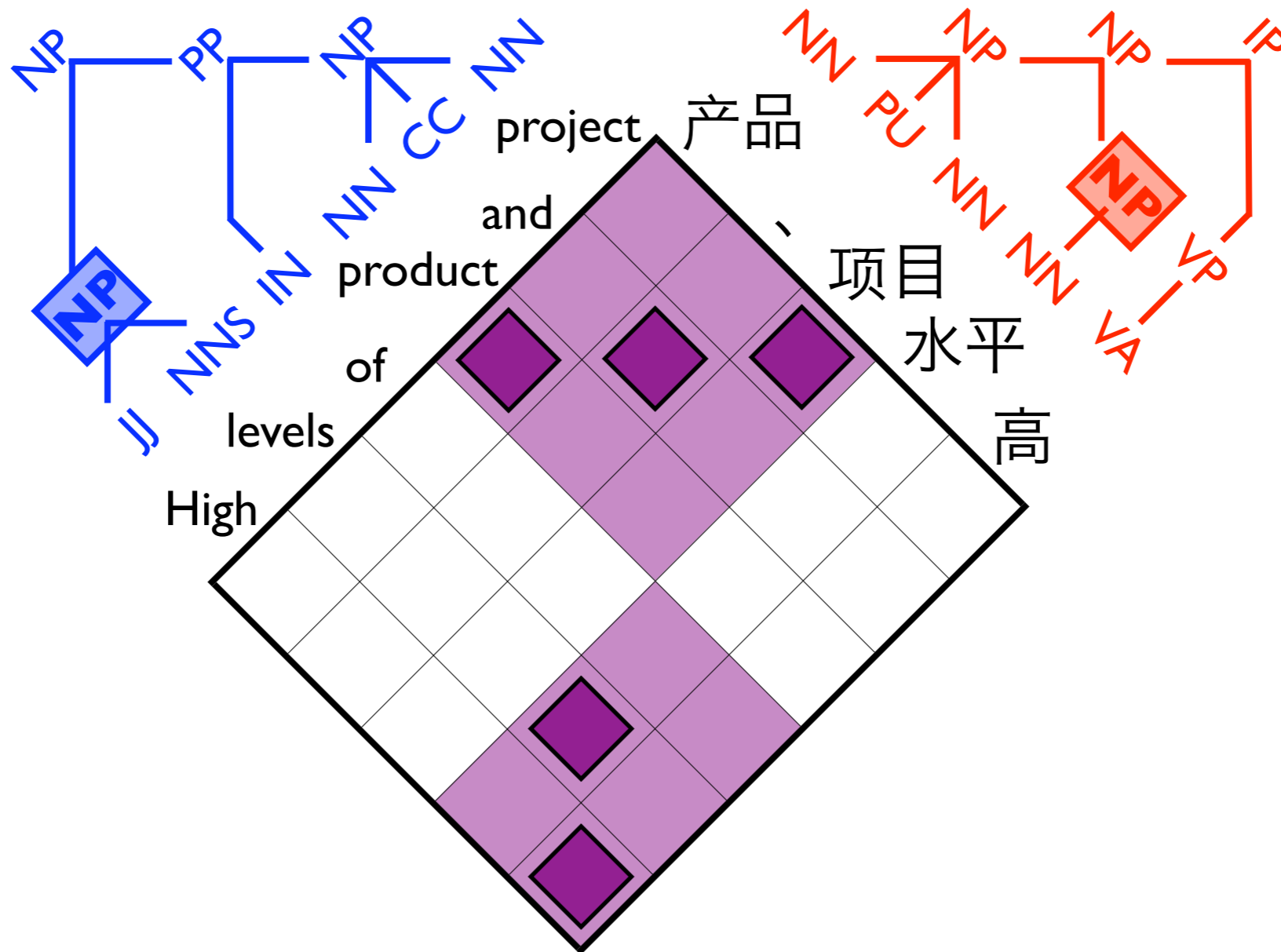


Weak Synchronization



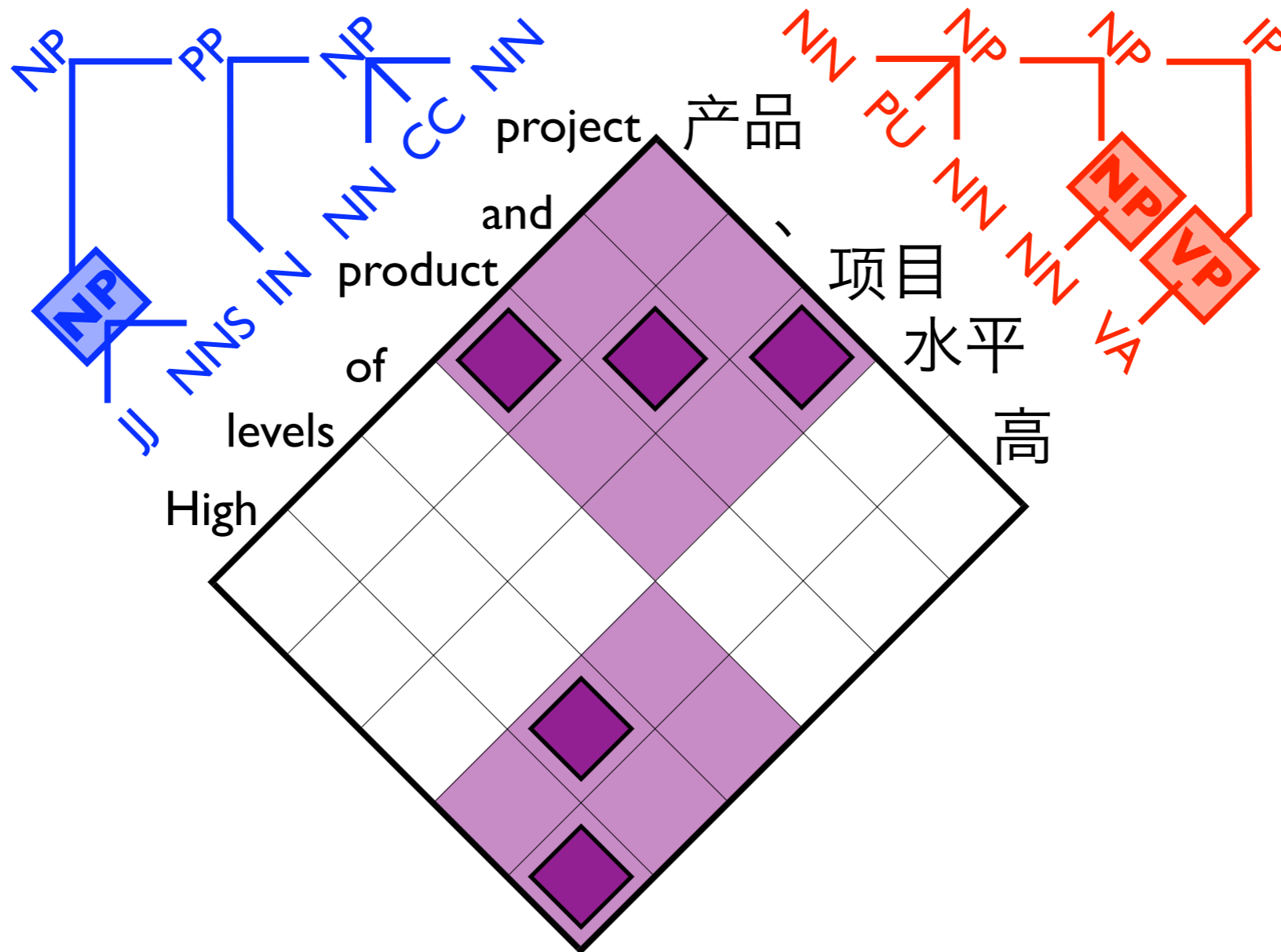


Weak Synchronization





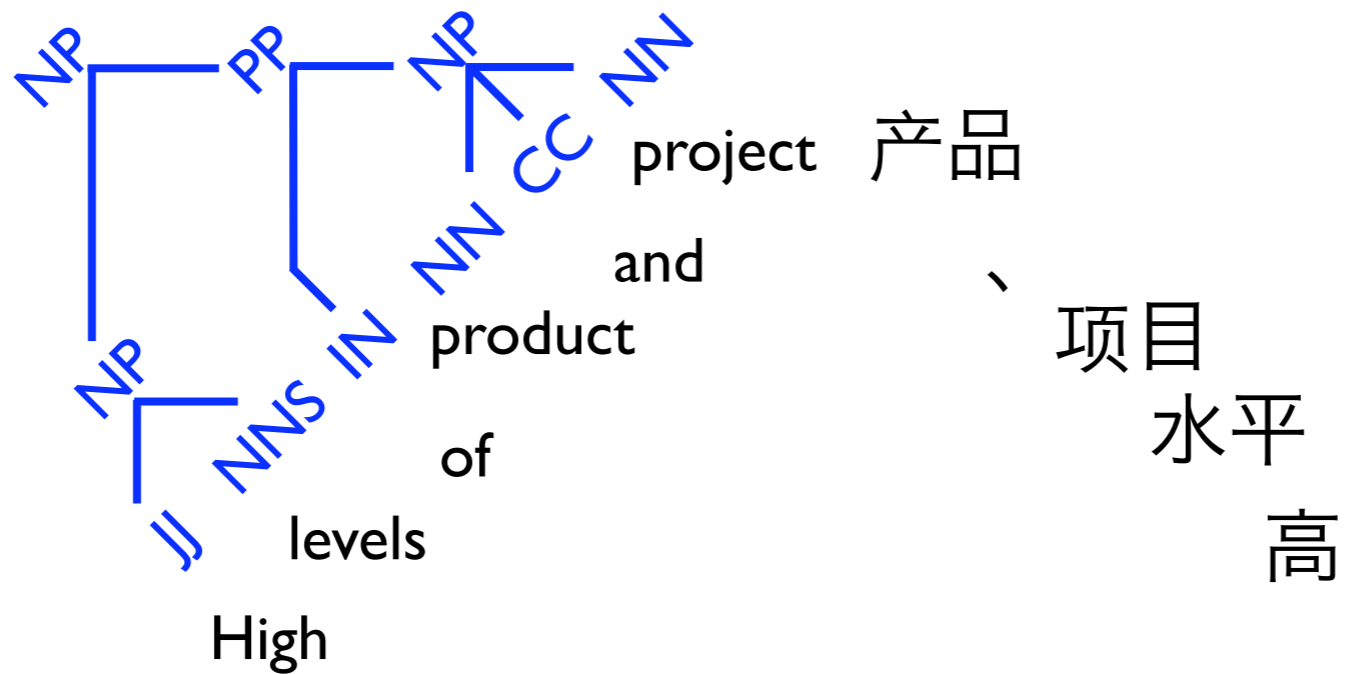
Weak Synchronization





Parsing and Alignment Output

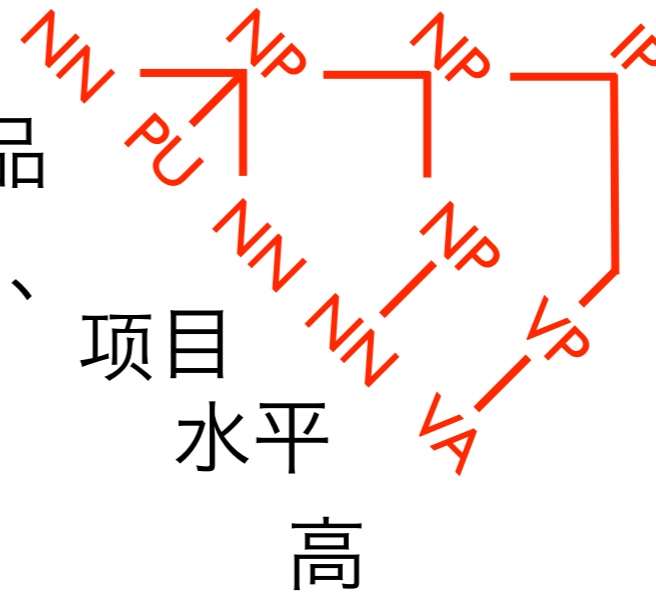
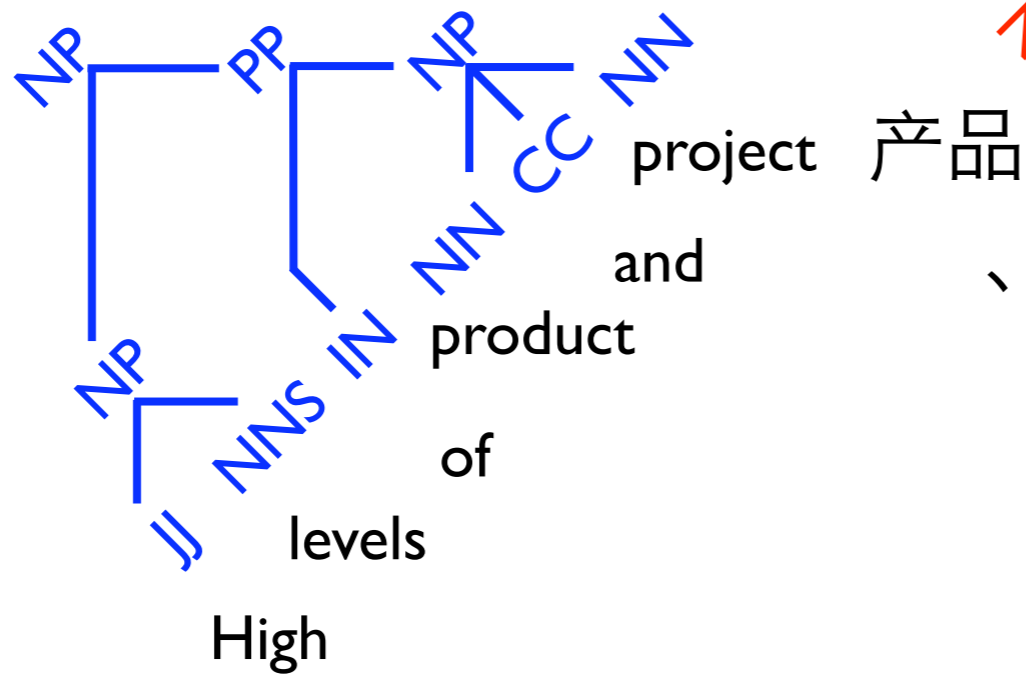
Trees *t*





Parsing and Alignment Output

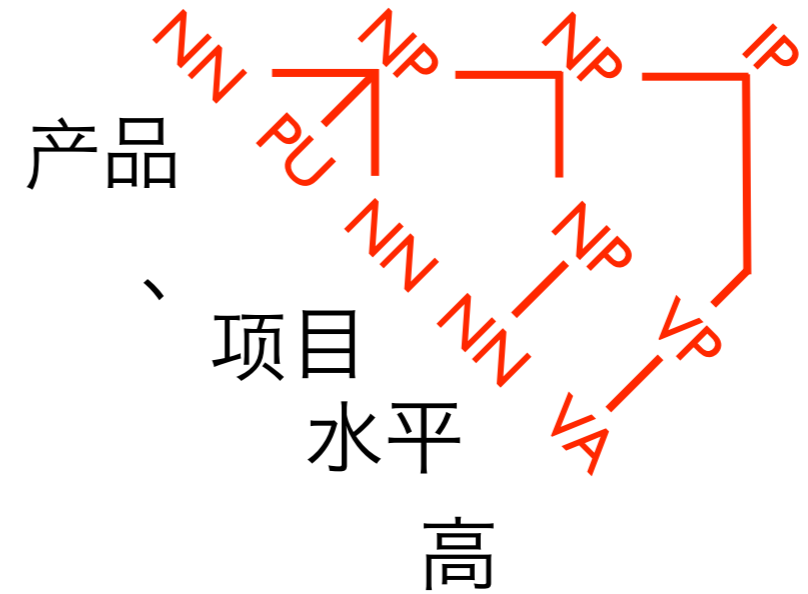
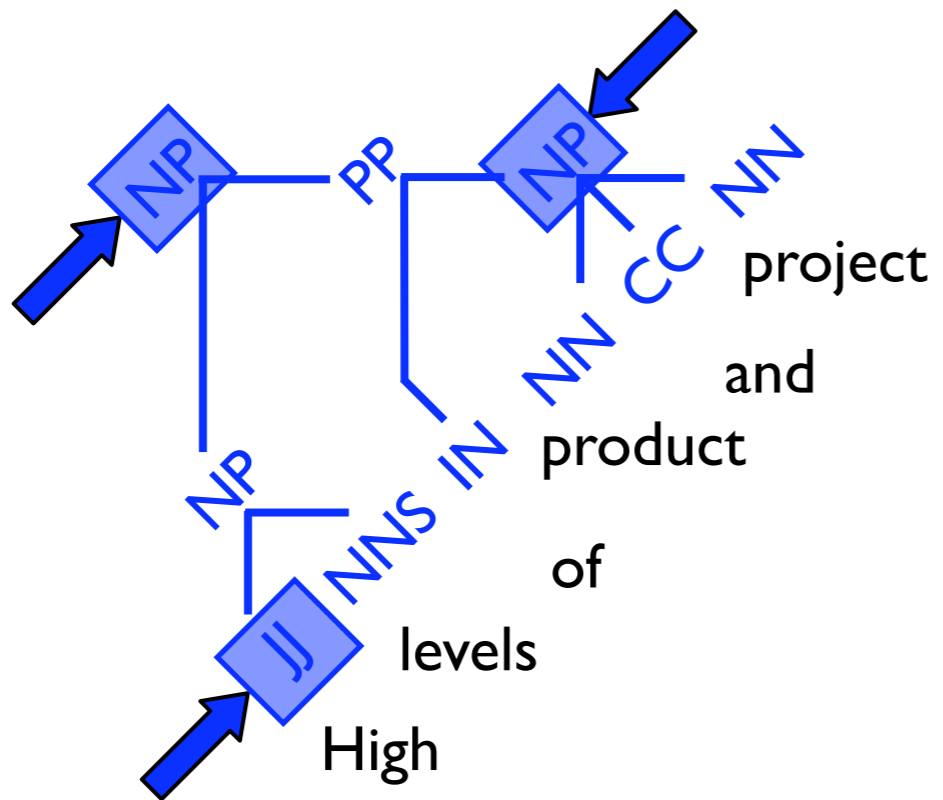
Trees t t'





Parsing and Alignment Output

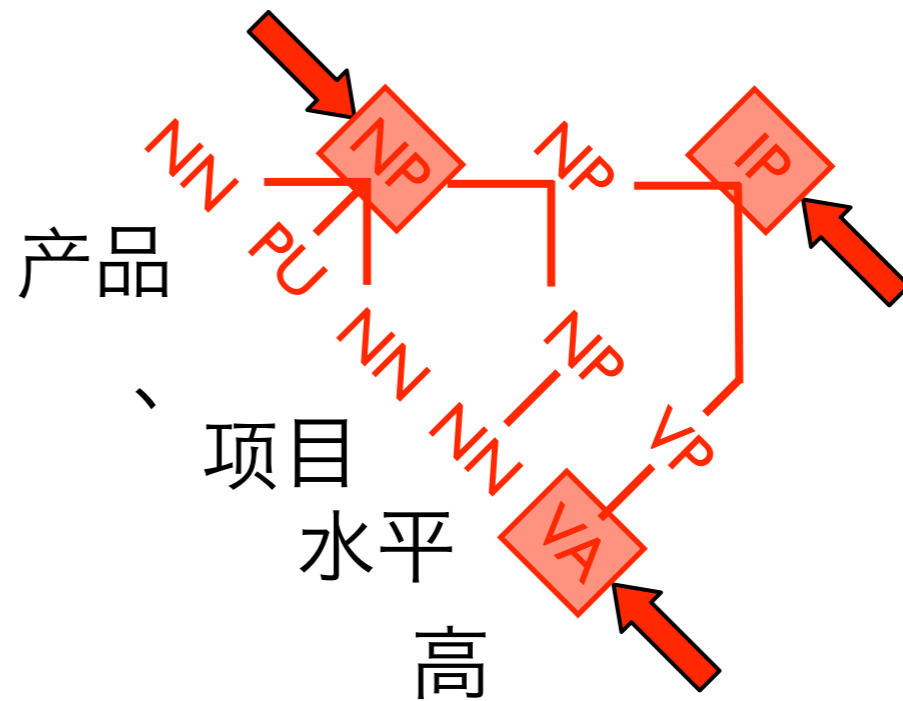
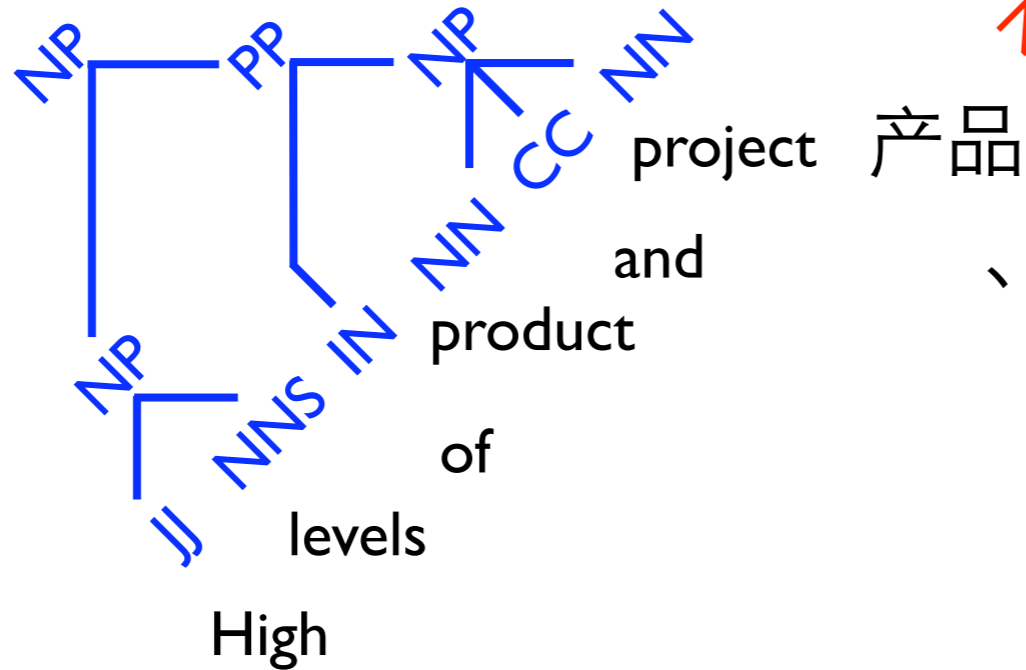
Trees t t'
contain Nodes n





Parsing and Alignment Output

Trees t t'
 contain Nodes n n'





Parsing and Alignment Output

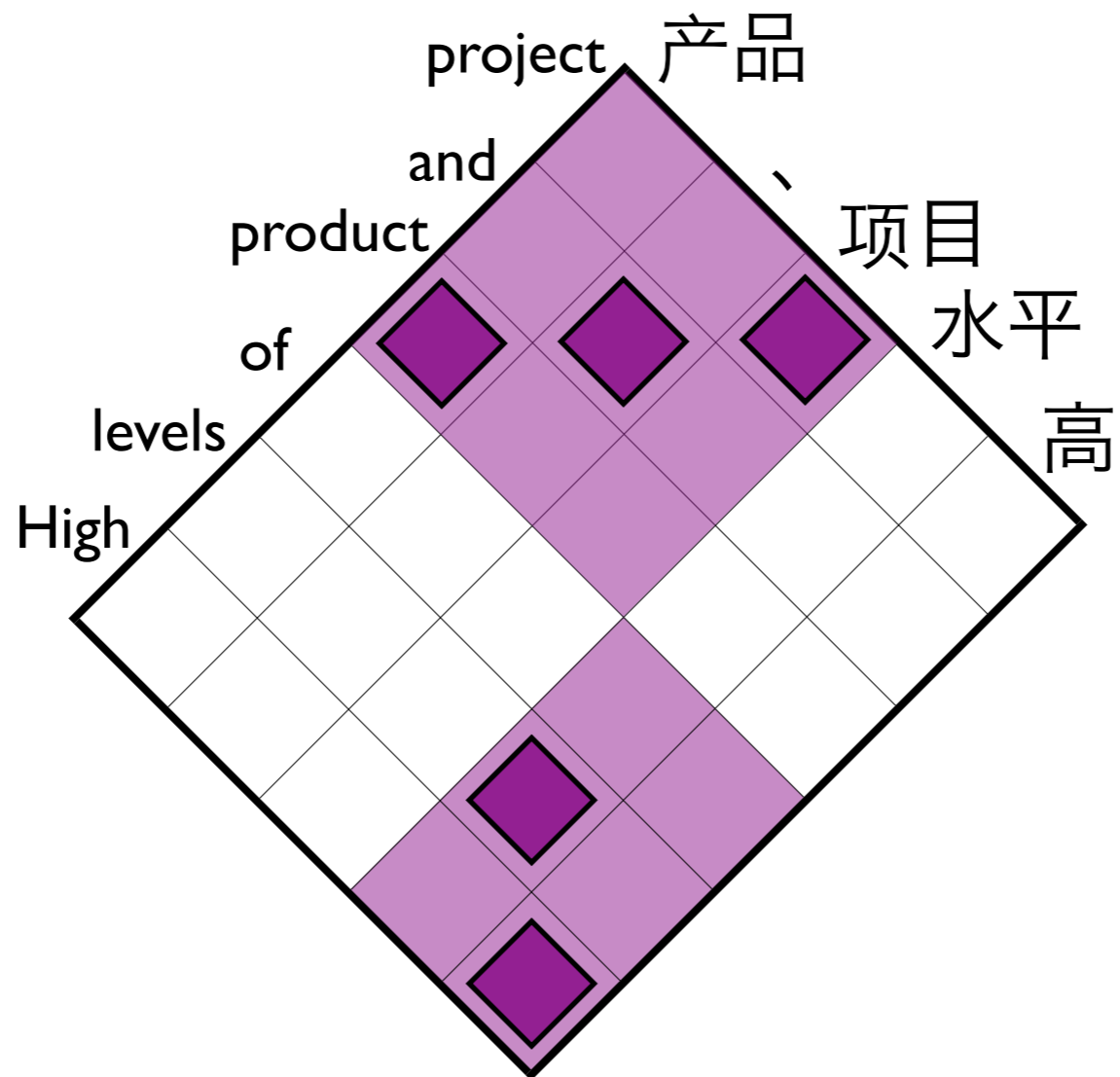
Alignments

	project	产品
	and	、
product		项目
of		水平
levels		高
High		



Parsing and Alignment Output

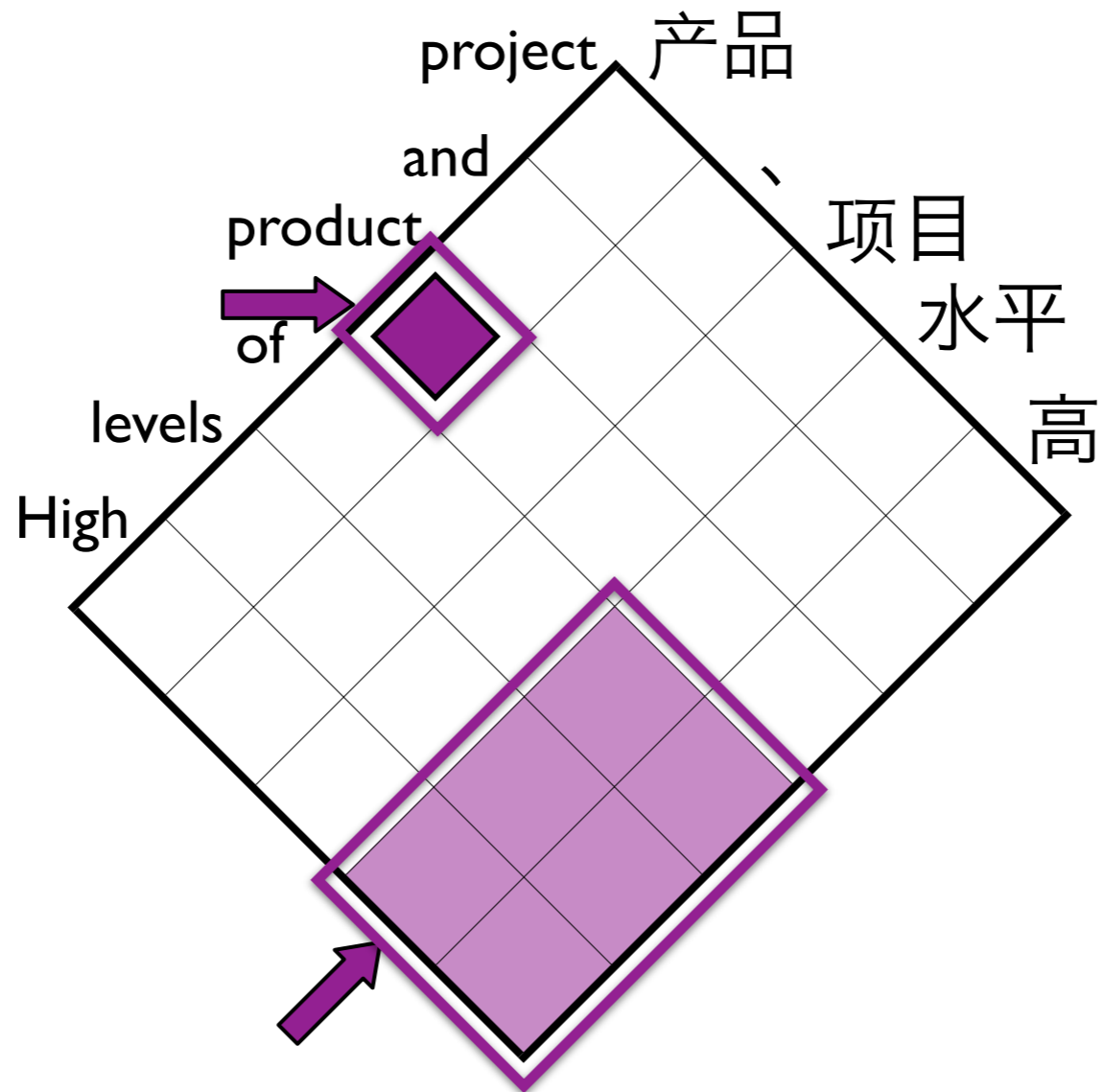
Alignments *a*





Parsing and Alignment Output

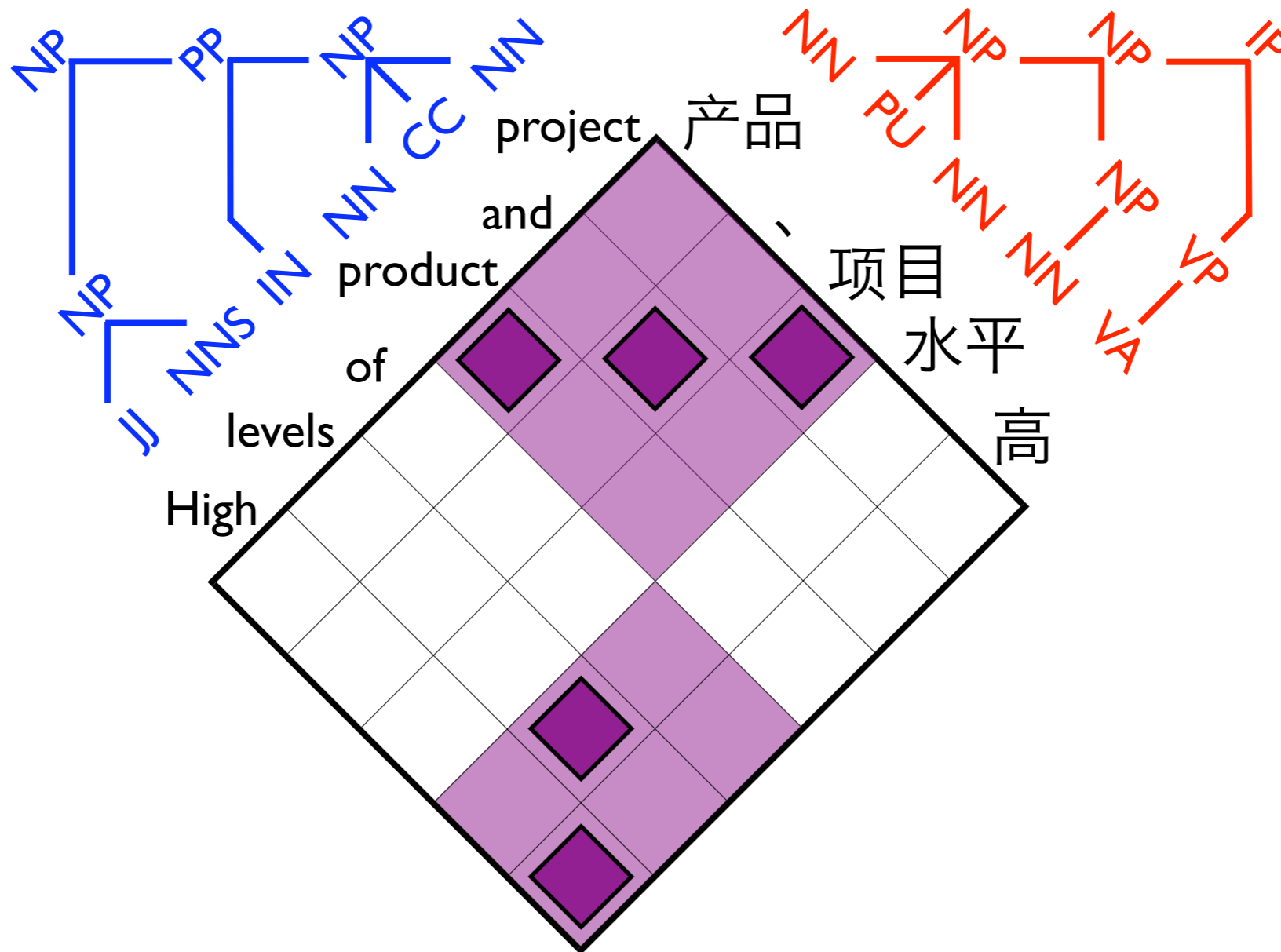
Alignments *a*
contain Bispans *b*





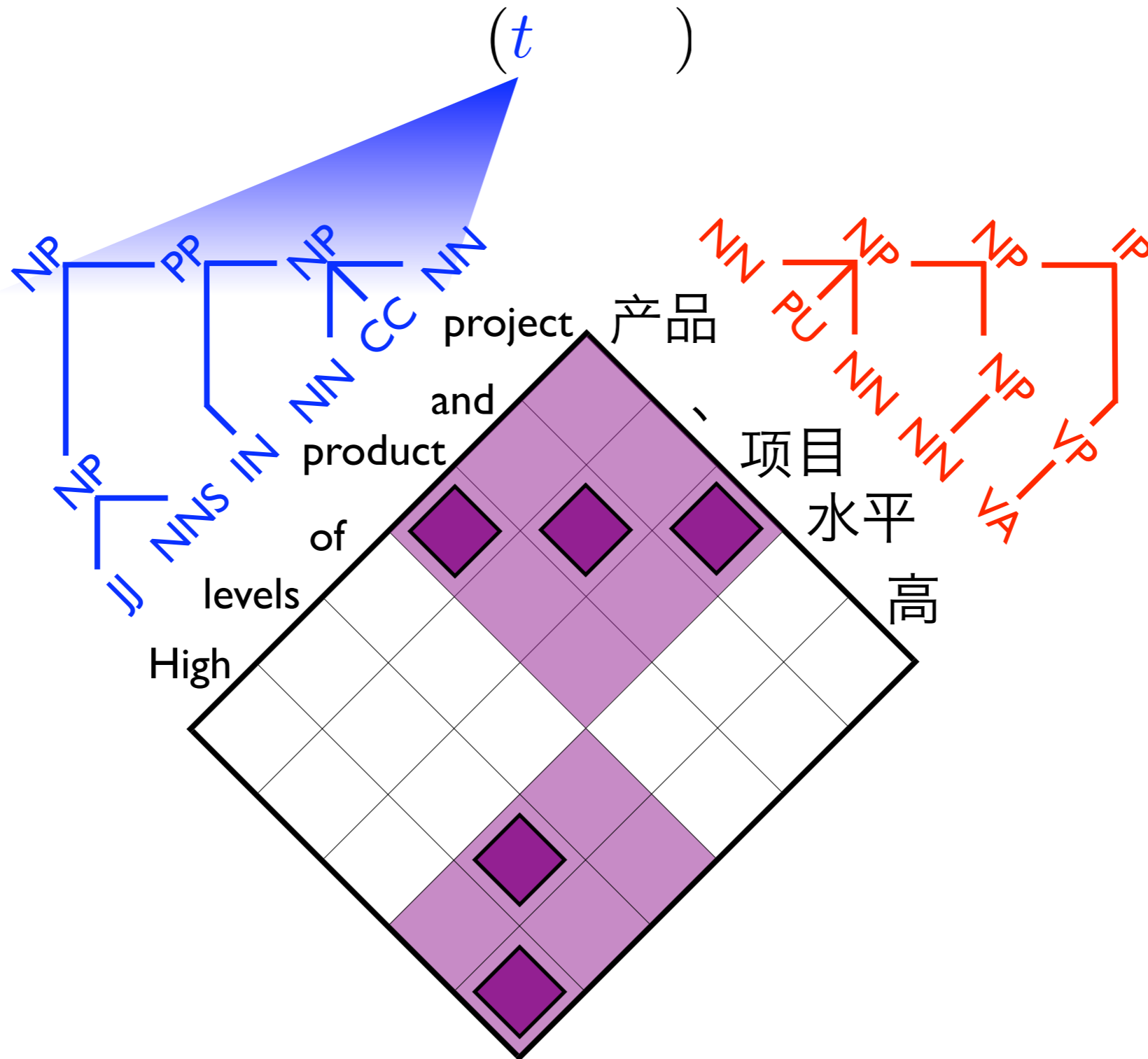
Parsing and Alignment Output

()



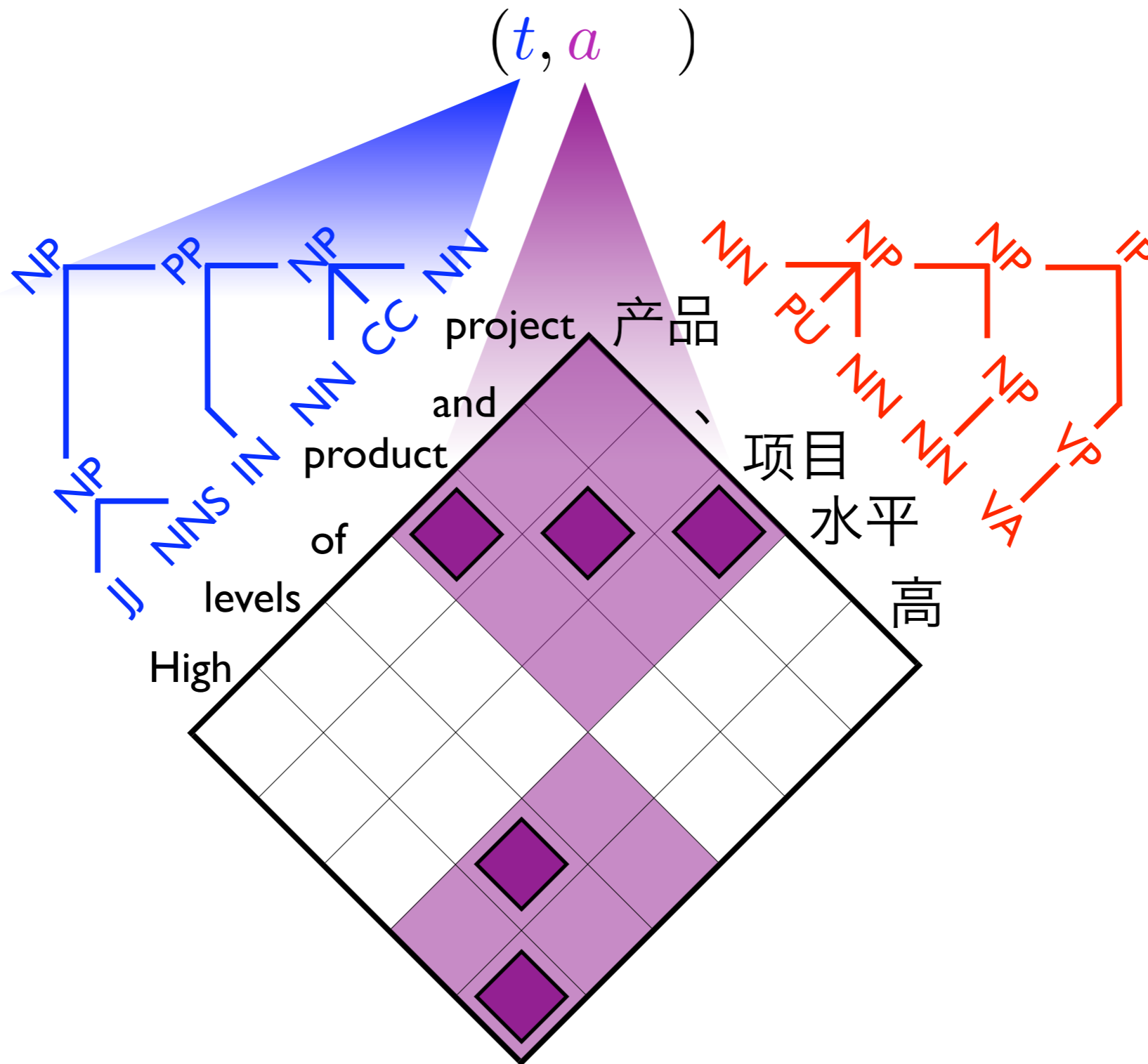


Parsing and Alignment Output



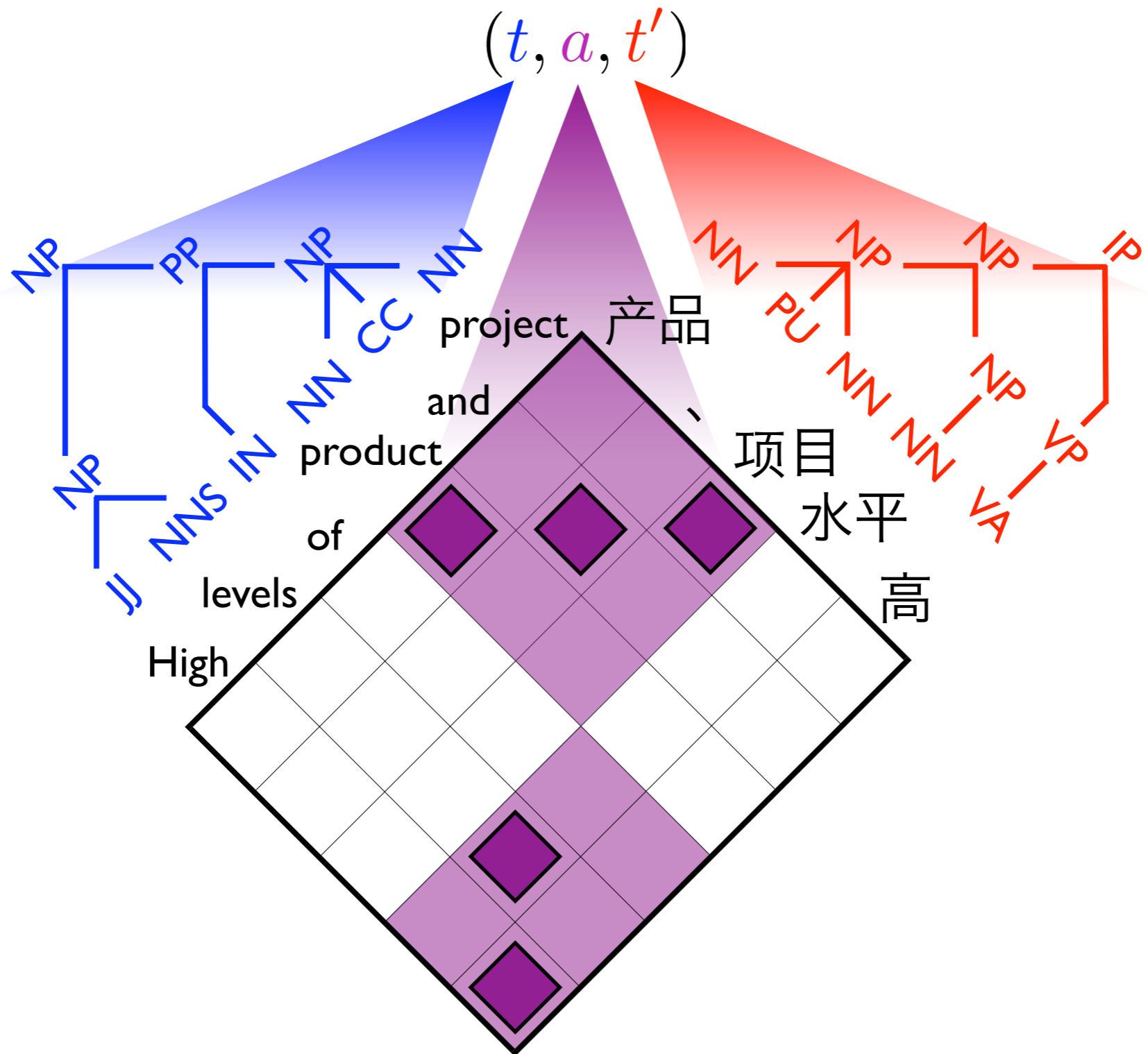


Parsing and Alignment Output





Parsing and Alignment Output

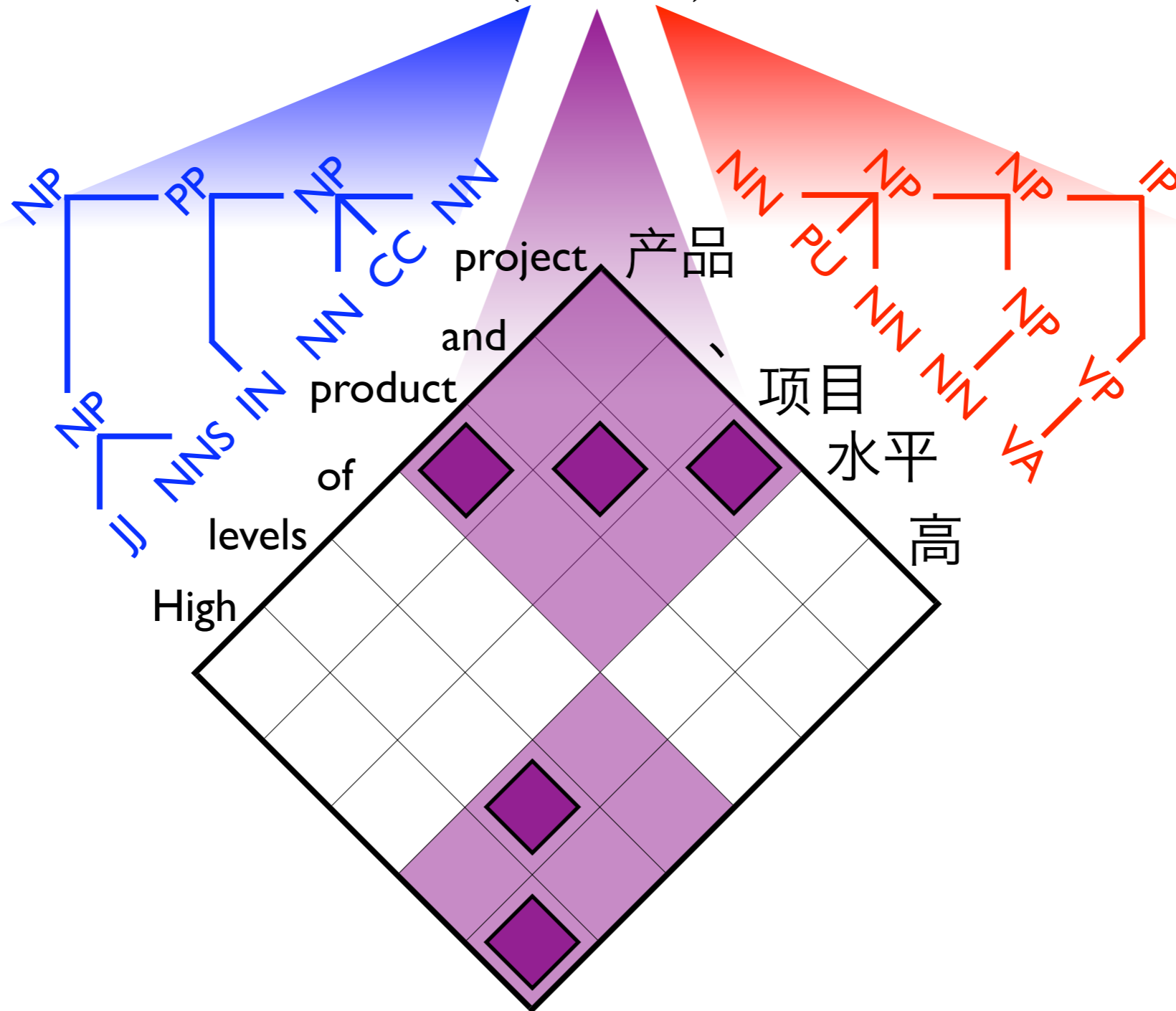




Parsing and Alignment Output

Tralignments

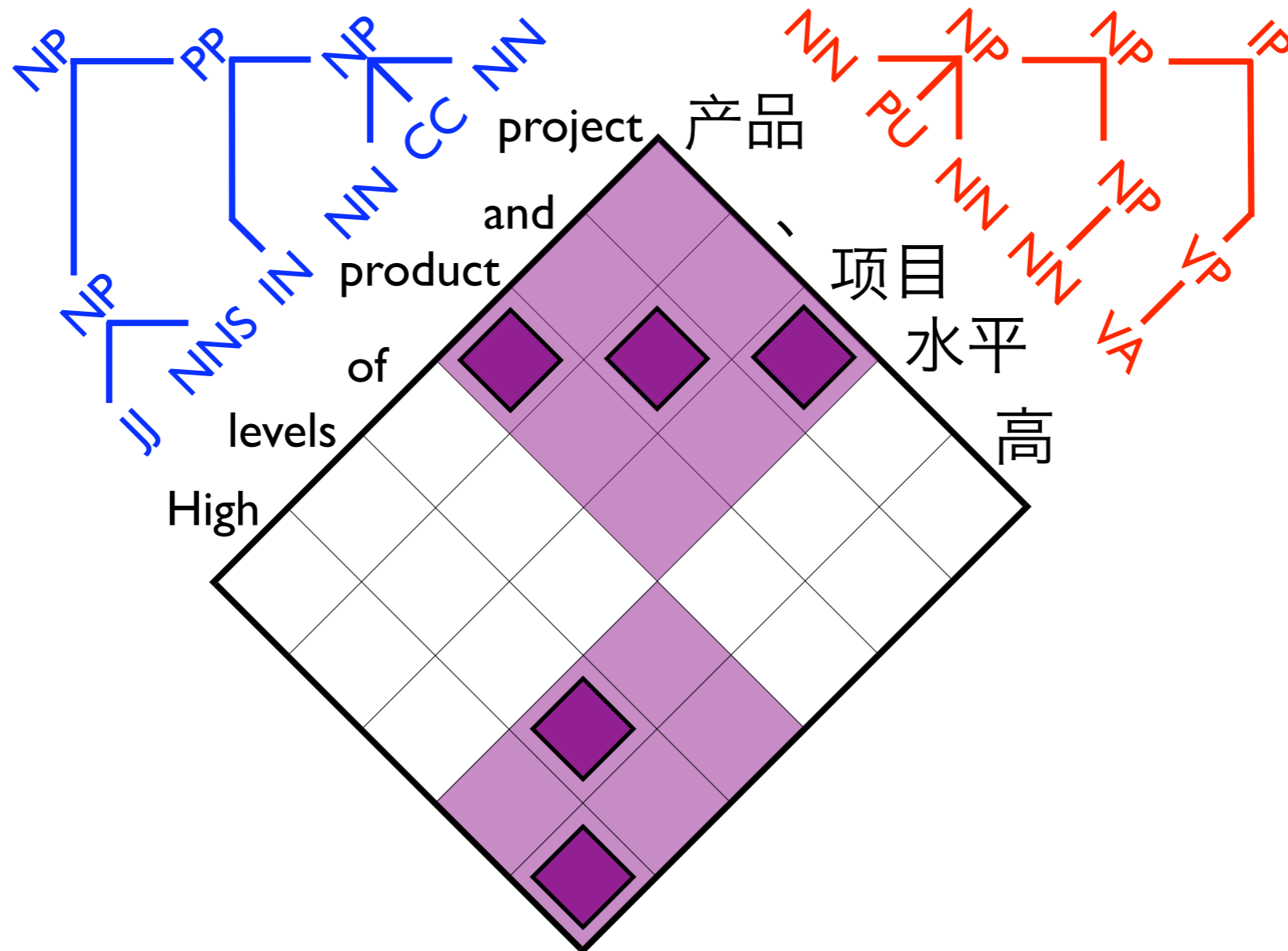
(t, a, t')





Loglinear Model

$$P(t, a, t' | s, s') \propto \exp(\langle \theta, \phi(t, a, t') \rangle)$$

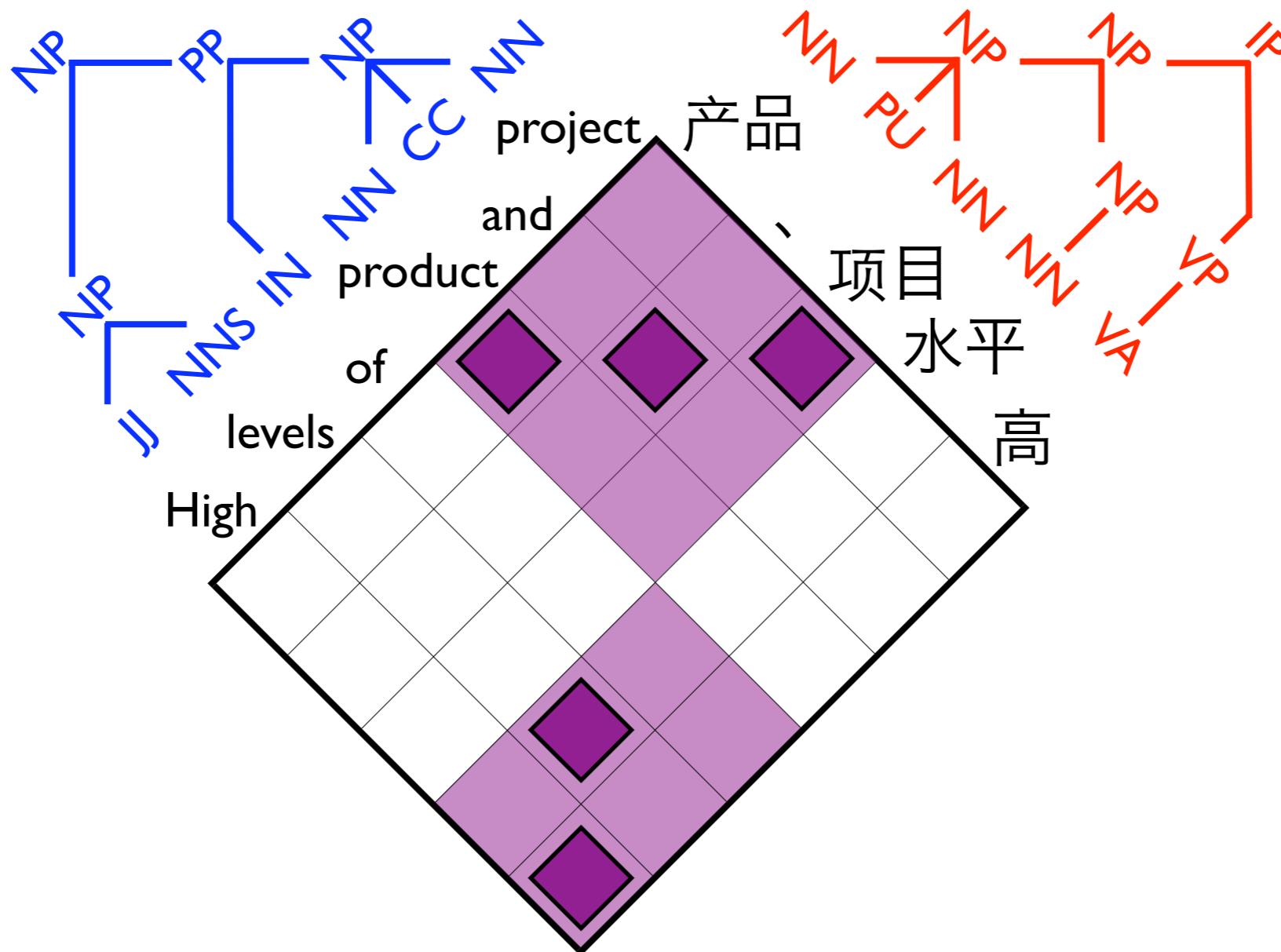




Loglinear Model

$$P(t, a, t' | s, s') \propto \exp(\langle \theta, \phi(t, a, t') \rangle)$$

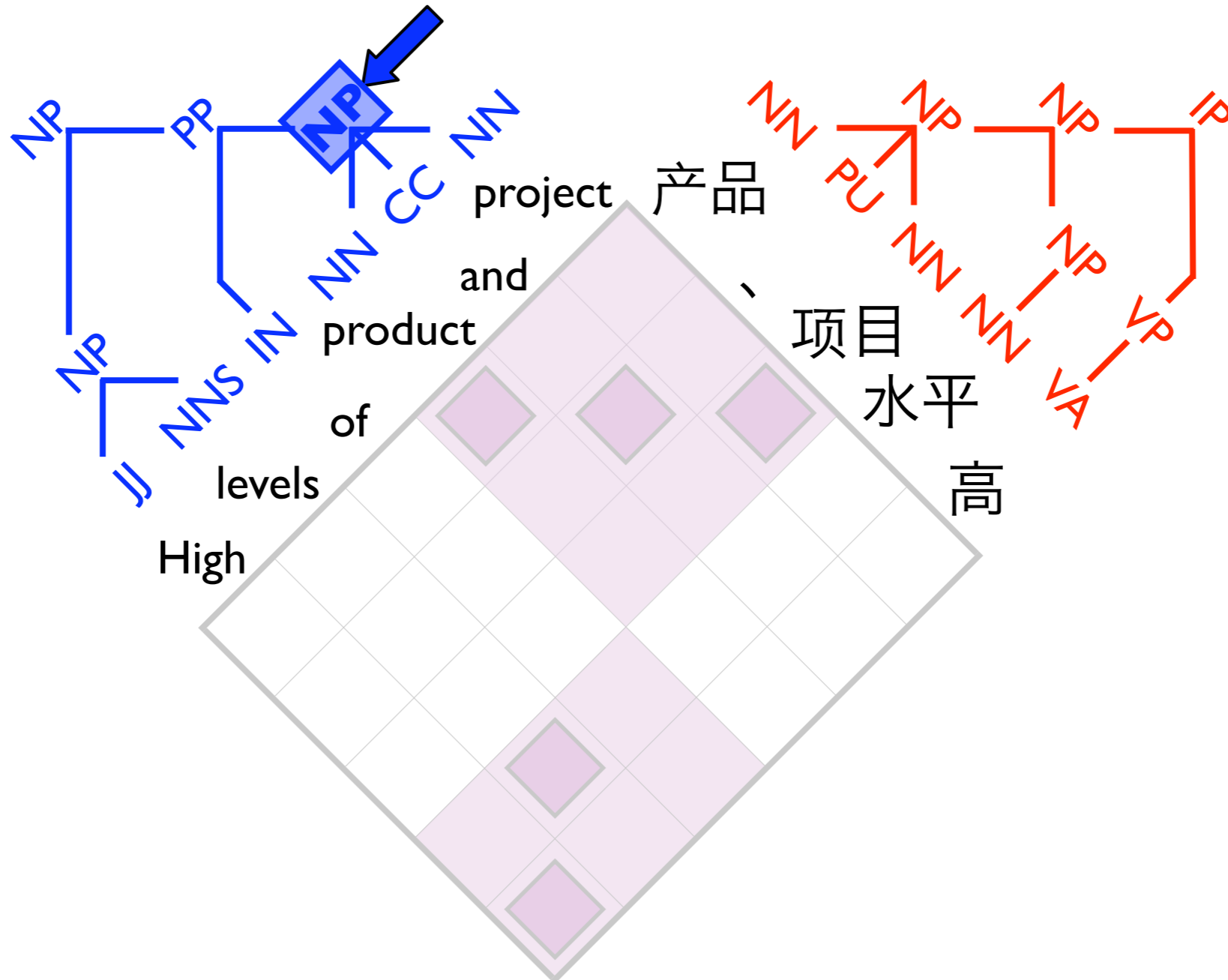
$$\phi(t, a, t')$$





Monolingual Parsing Features

$$\phi(n) \text{ TRGMONO} = \log P_{eng} \left(\begin{array}{c} \text{NP} \\ \text{NN} \quad \text{CC} \quad \text{NN} \end{array} \right)$$

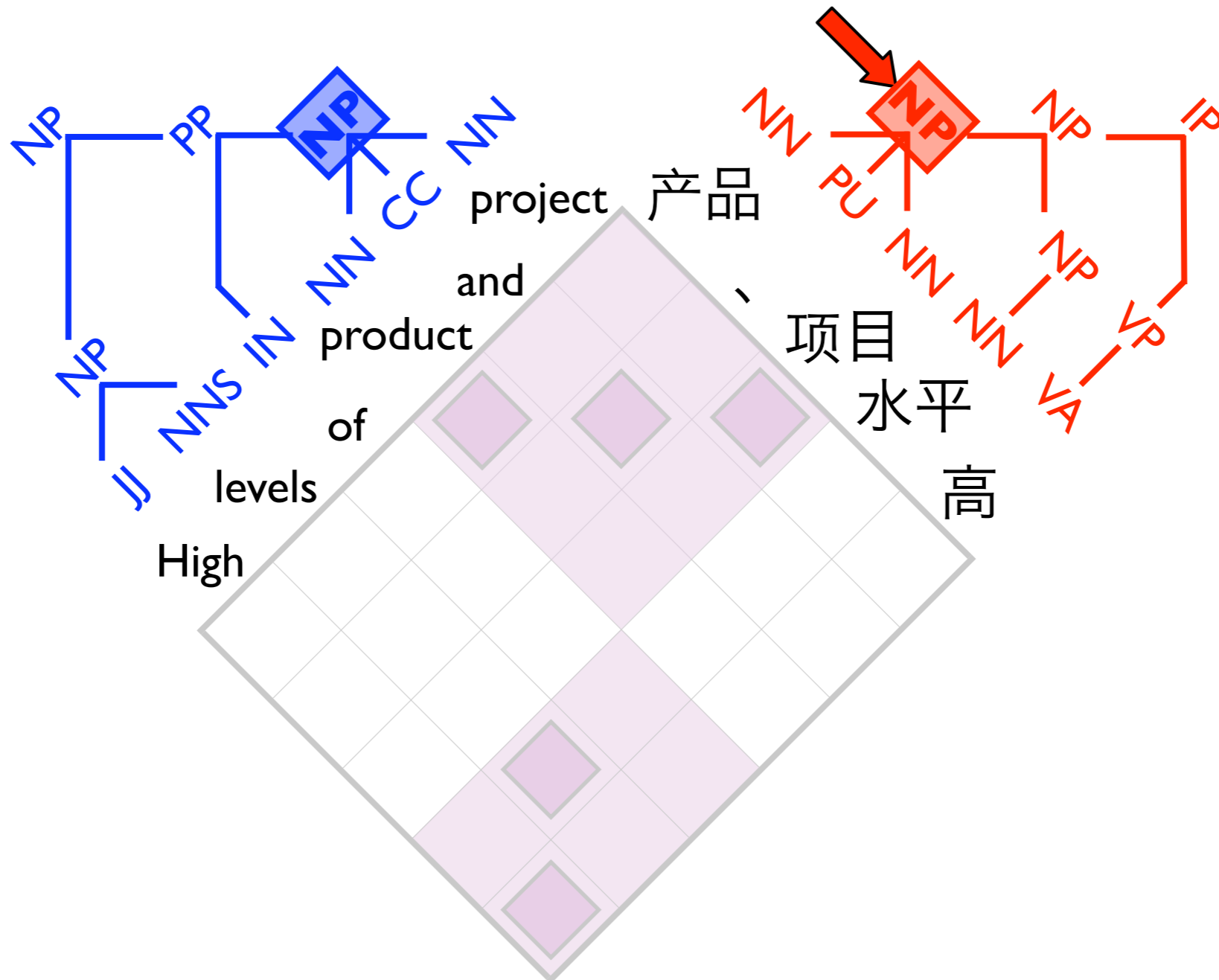




Monolingual Parsing Features

$$\phi(n) \text{ TRGMONO} = \log P_{eng} \left(\begin{array}{c} \text{NP} \\ \text{NN} \text{ CC} \text{ NN} \end{array} \right)$$

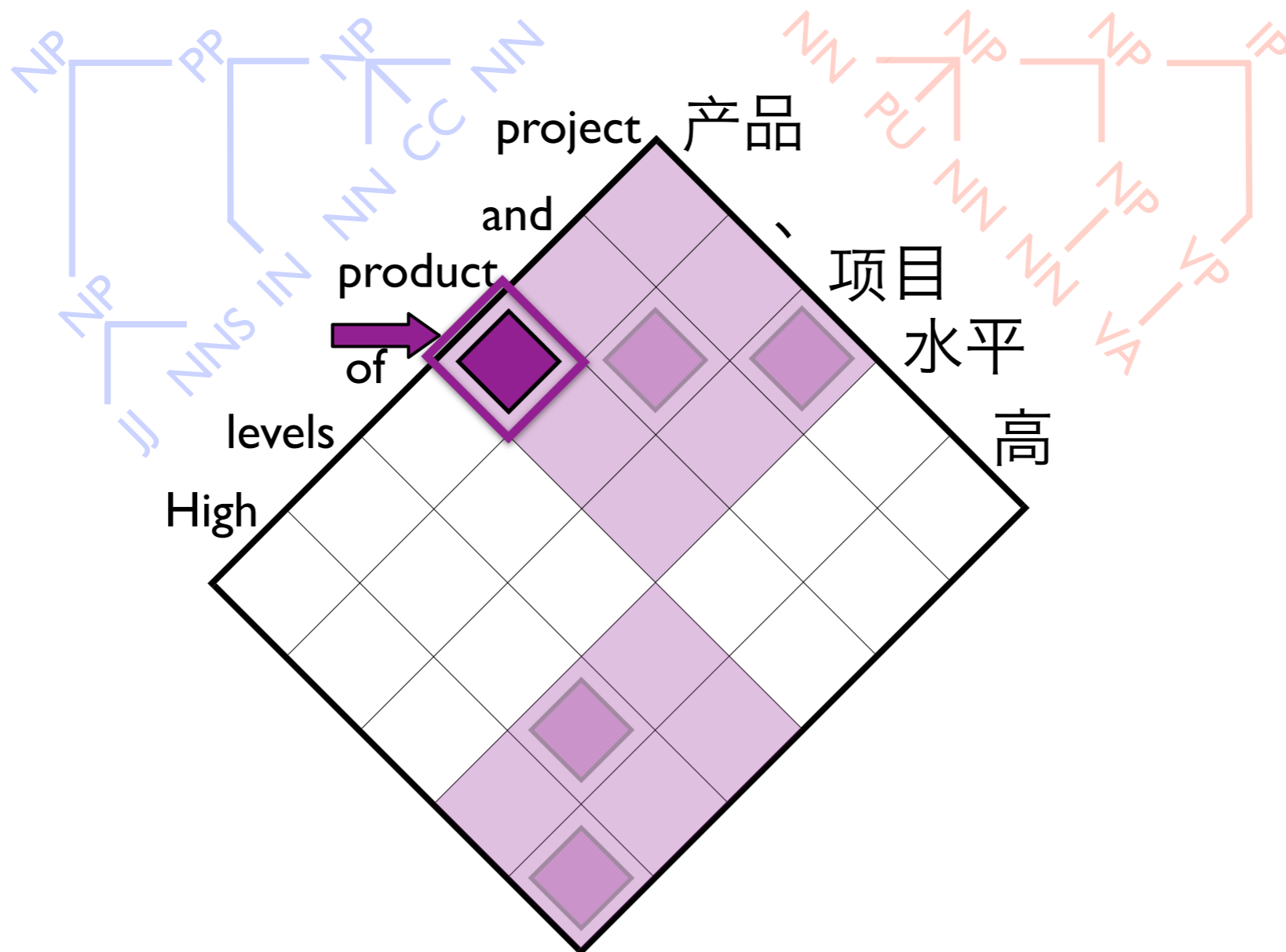
$$\phi(n') \text{ SRCMONO} = \log P_{chi} \left(\begin{array}{c} \text{NP} \\ \text{NN} \text{ PU} \text{ NN} \end{array} \right)$$





Alignment Features

$$\phi(b)$$

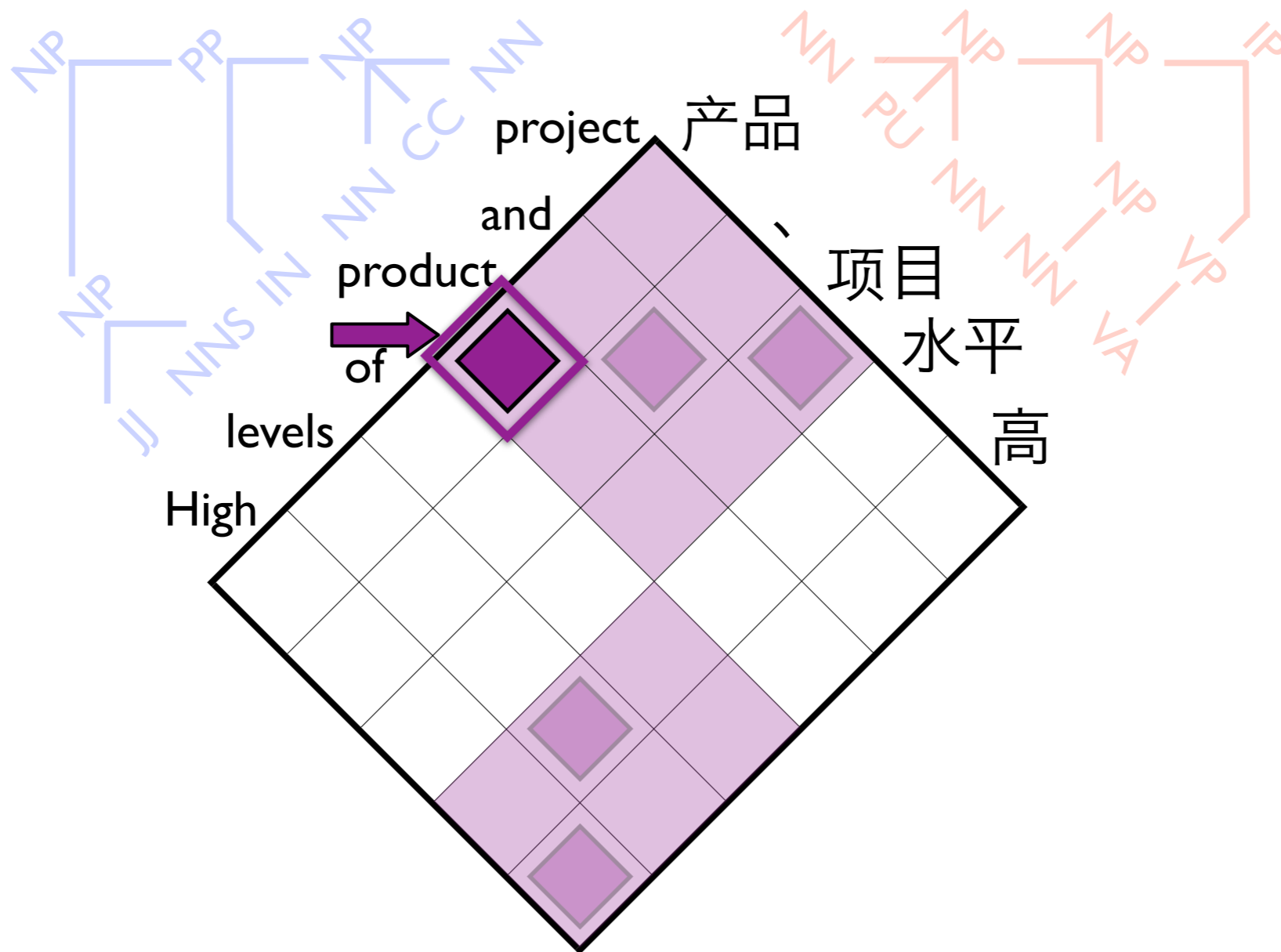




Alignment Features

$$\text{UNSUPSCORE} = \phi(b)$$

$$P_{uns}(\text{product} \diamond \text{产品})$$

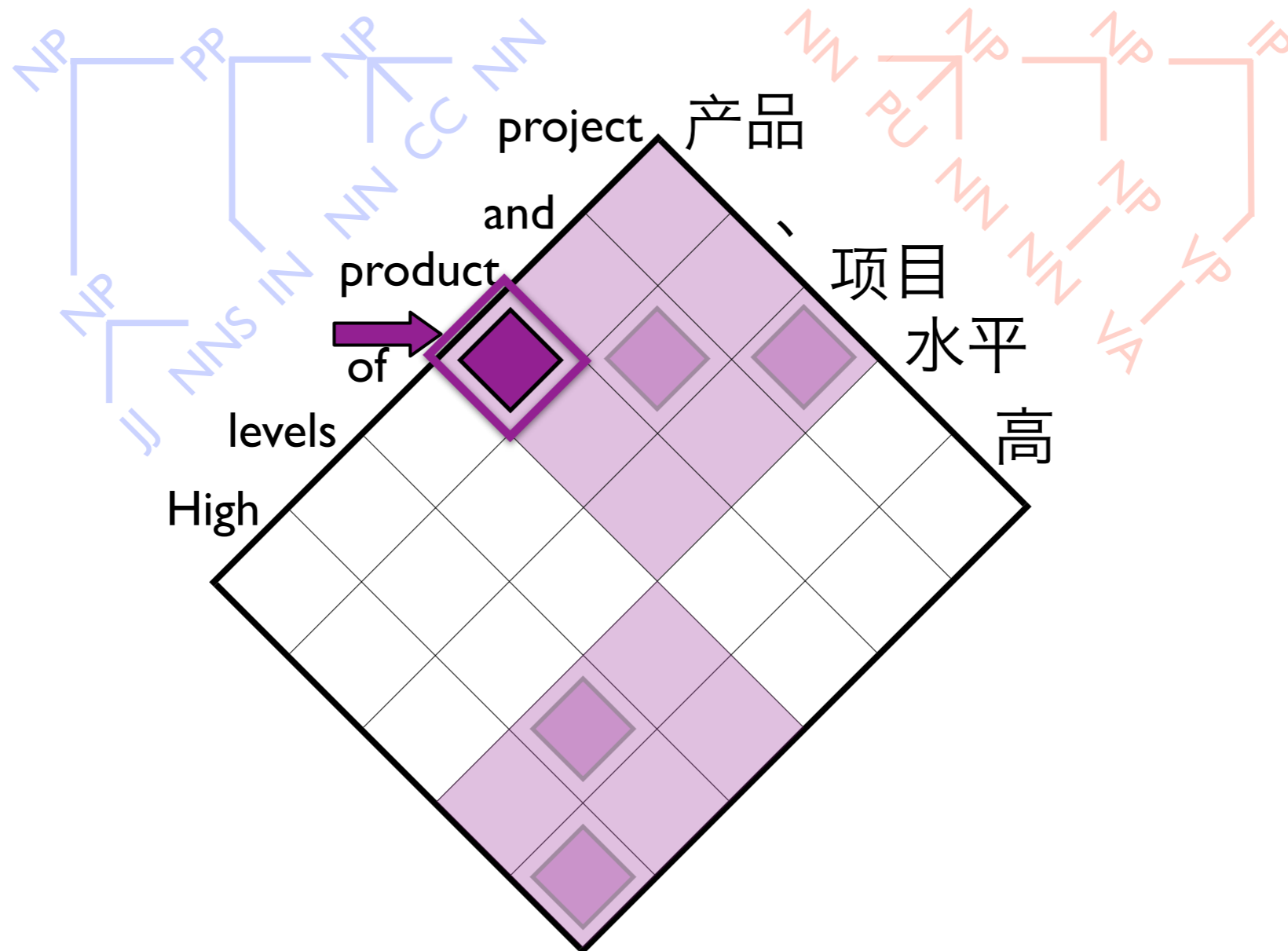




Alignment Features

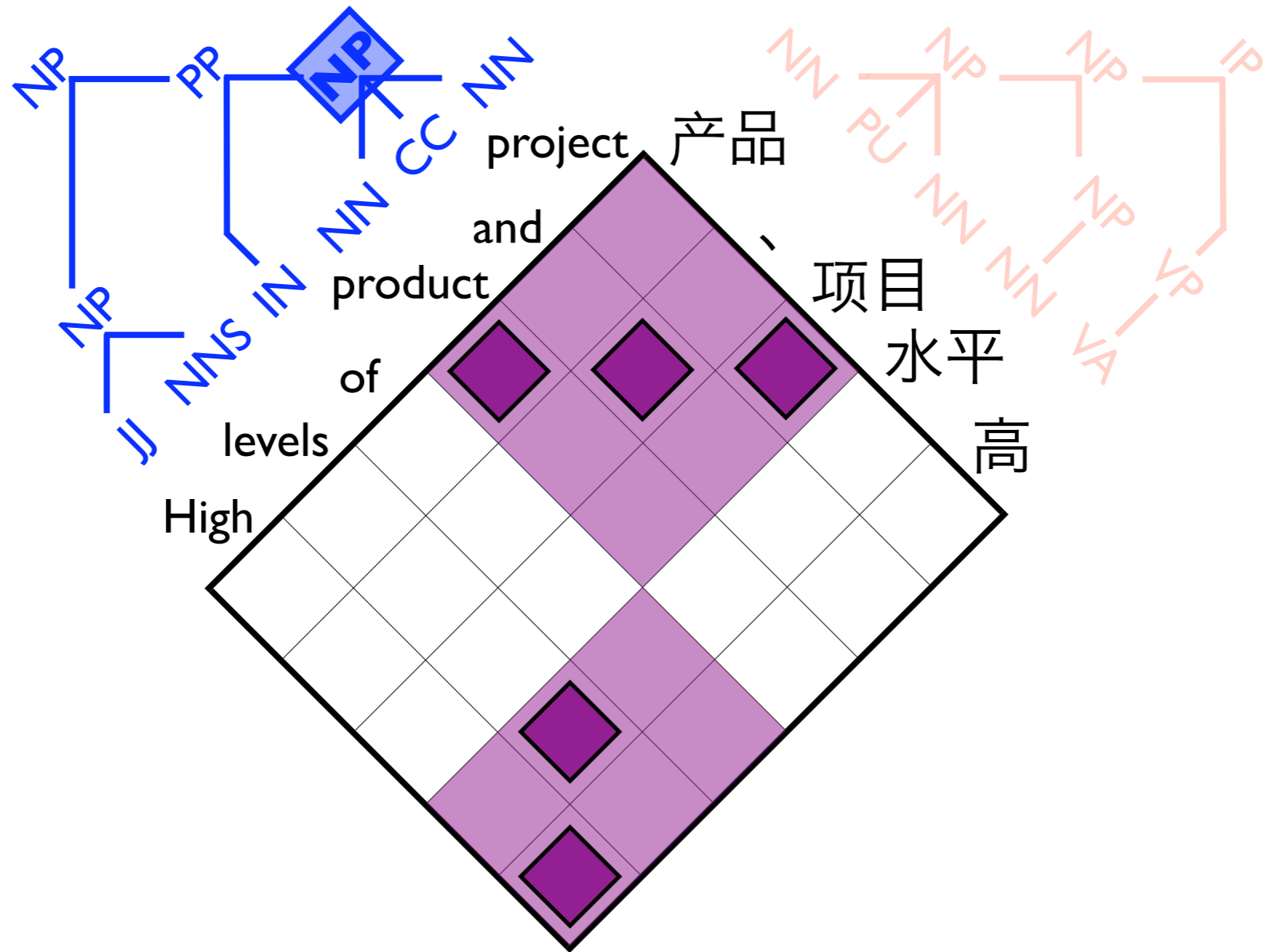
$$\text{UNSUPSCORE} = \phi(b) \quad \text{LEX} \langle \text{product}, \text{产品} \rangle = 1$$

$$P_{uns}(\text{product} \blacklozenge \text{产品})$$

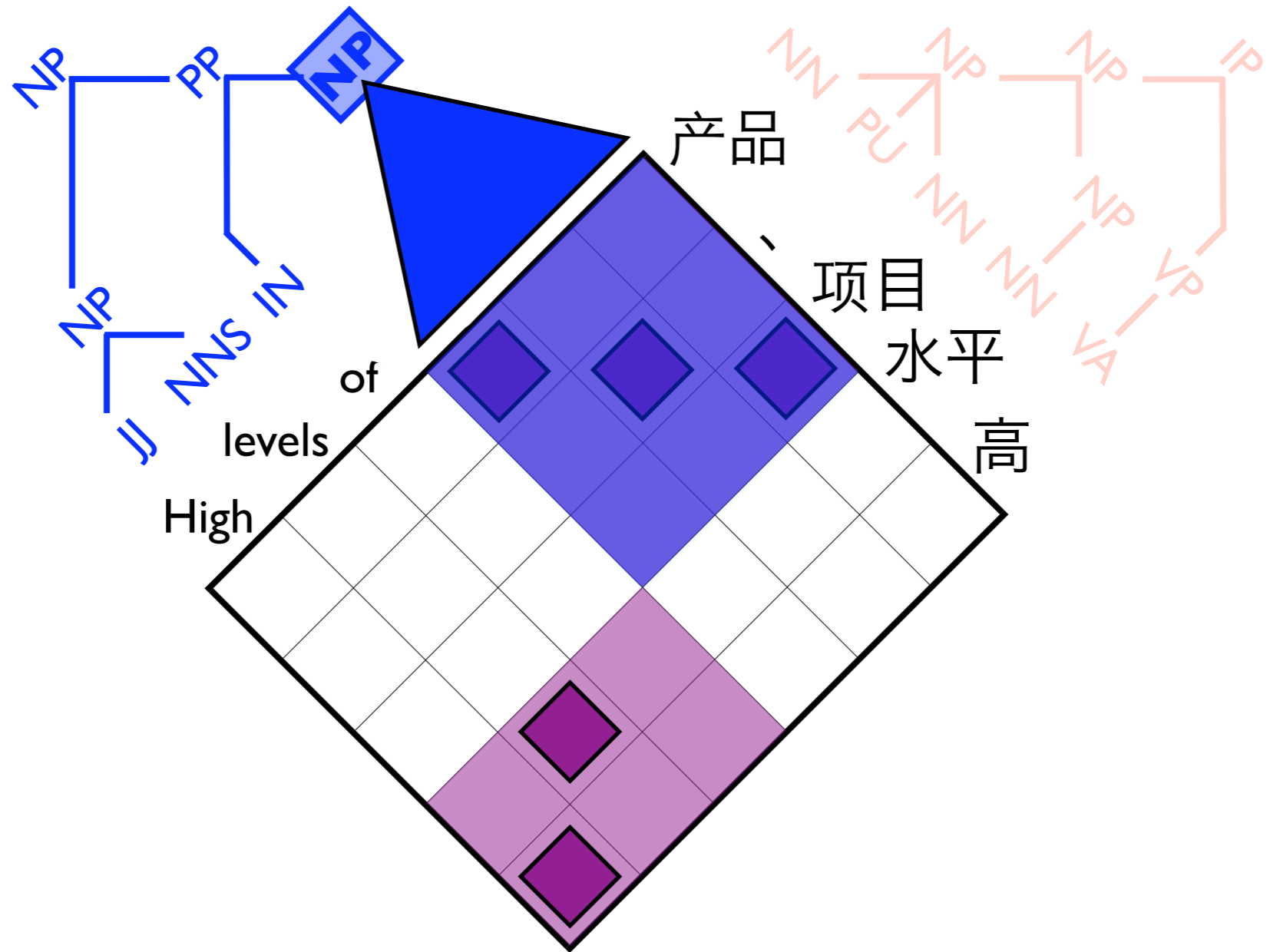




Target-Alignment Synchronization Features



Target-Alignment Synchronization Features

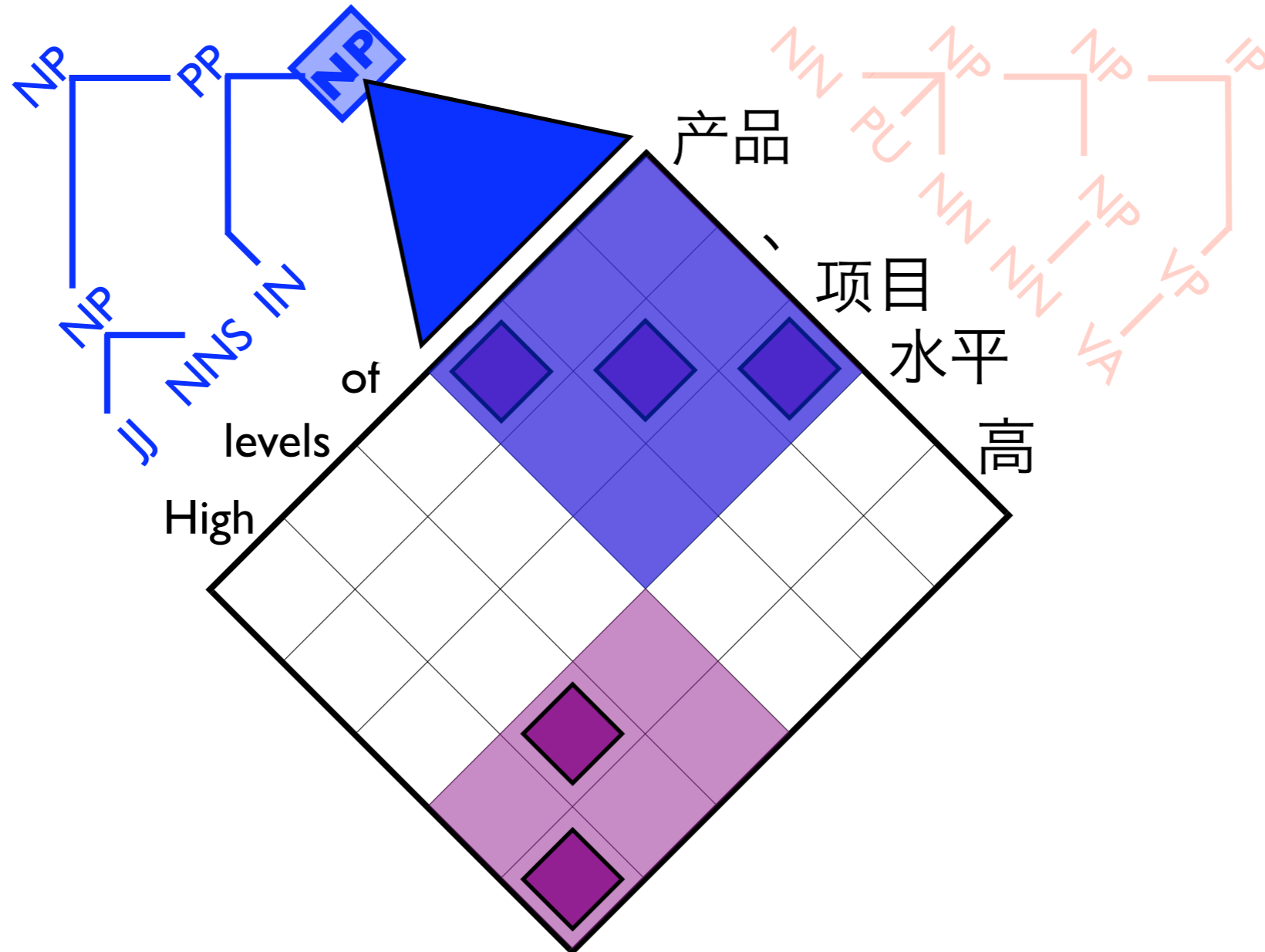




Target-Alignment Synchronization Features

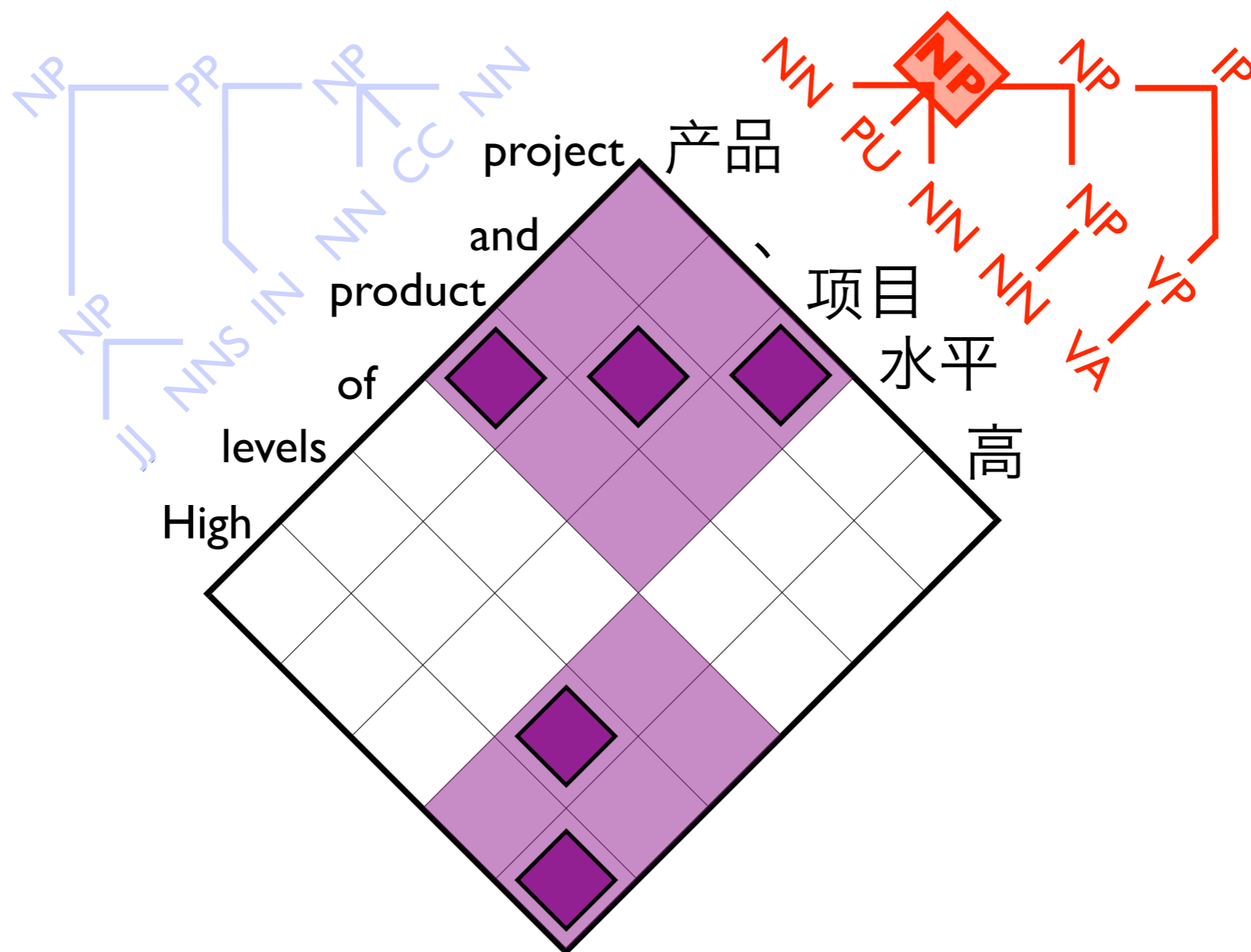
$$\phi(n, b)$$

$$\text{TRGALIGN}\langle \text{NP} \rangle = 1$$





Source-Alignment Synchronization Features

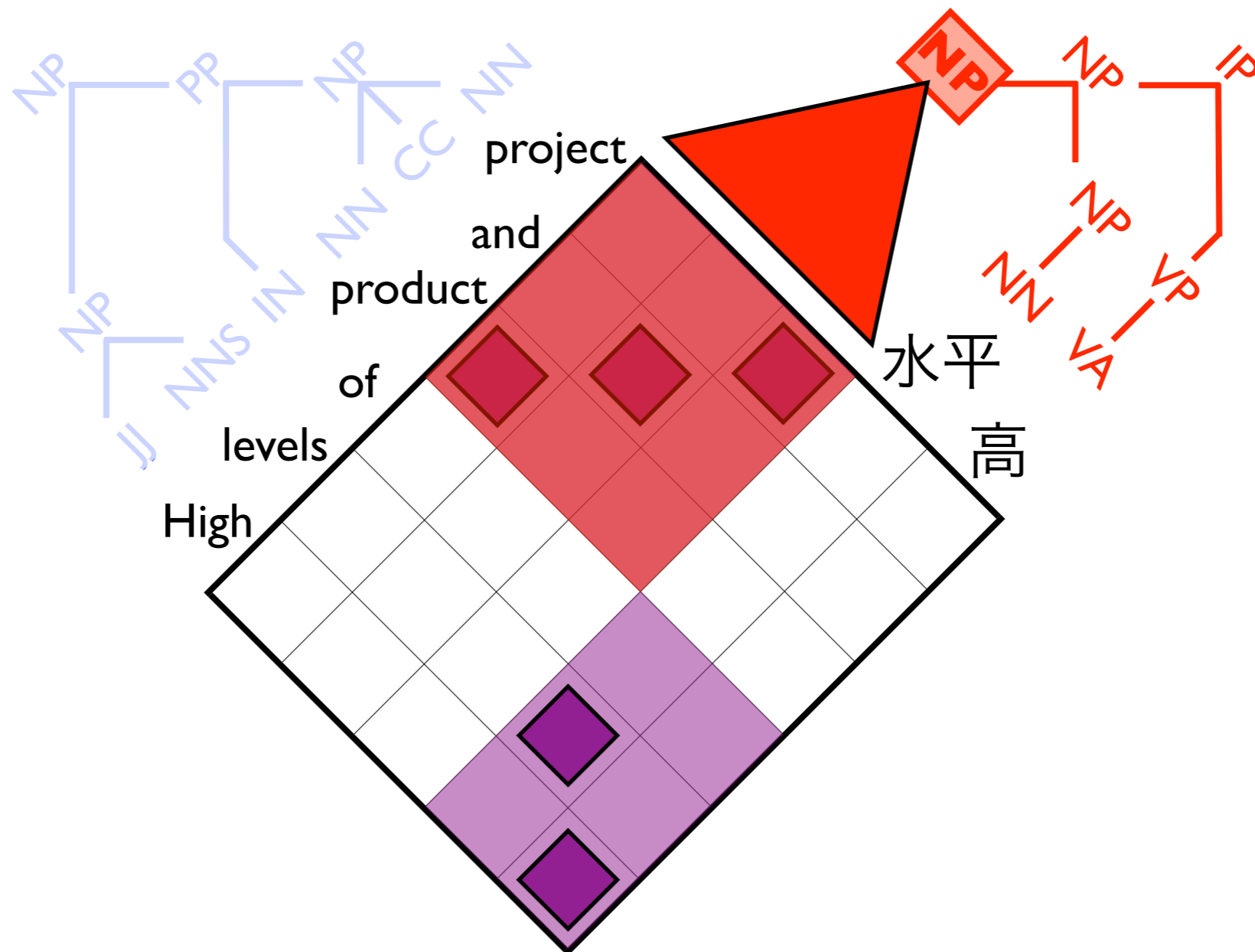




Source-Alignment Synchronization Features

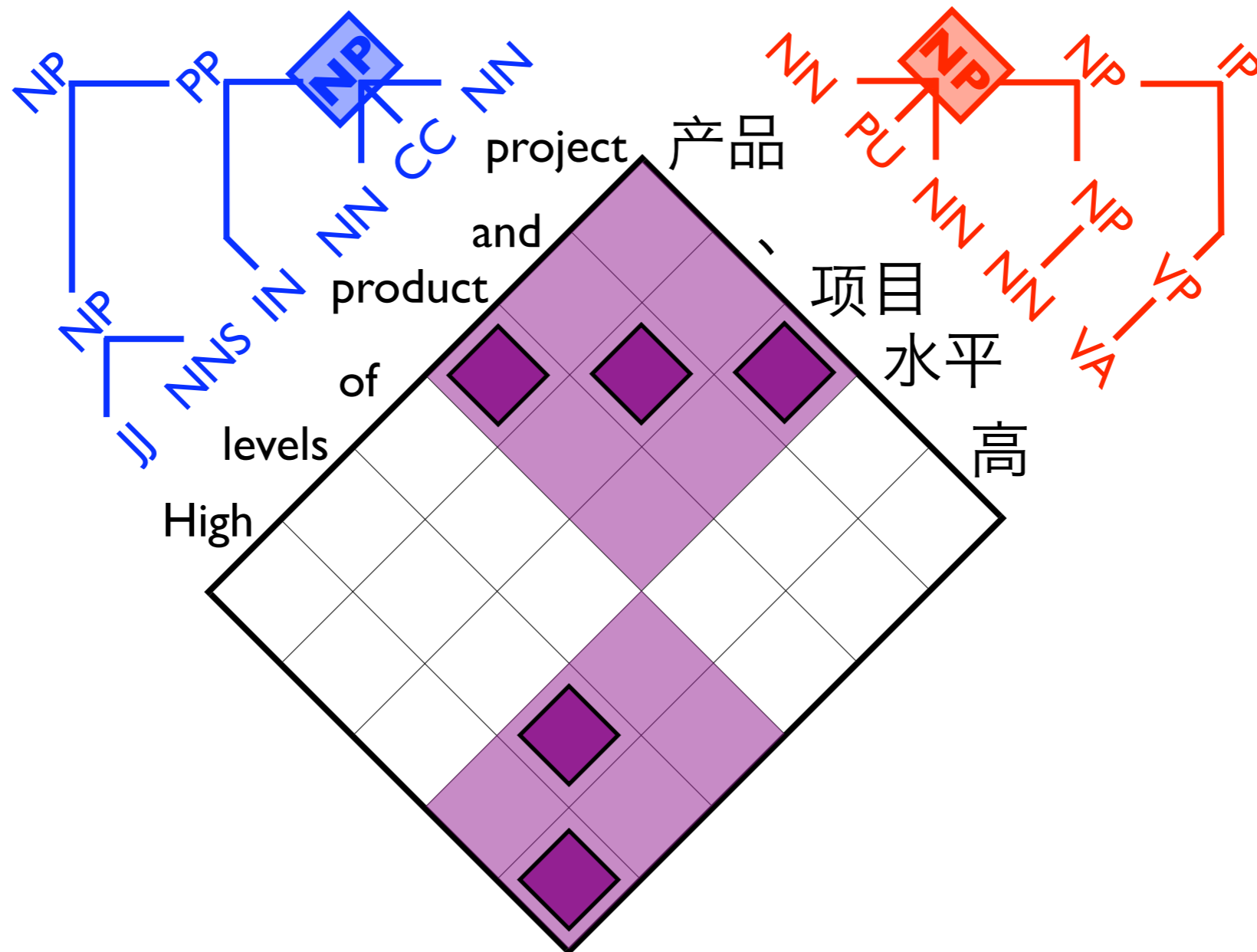
$$\phi(b, n')$$

$$\text{SRCALIGN} \langle \text{NP} \rangle = 1$$





Target-Source Synchronization Features

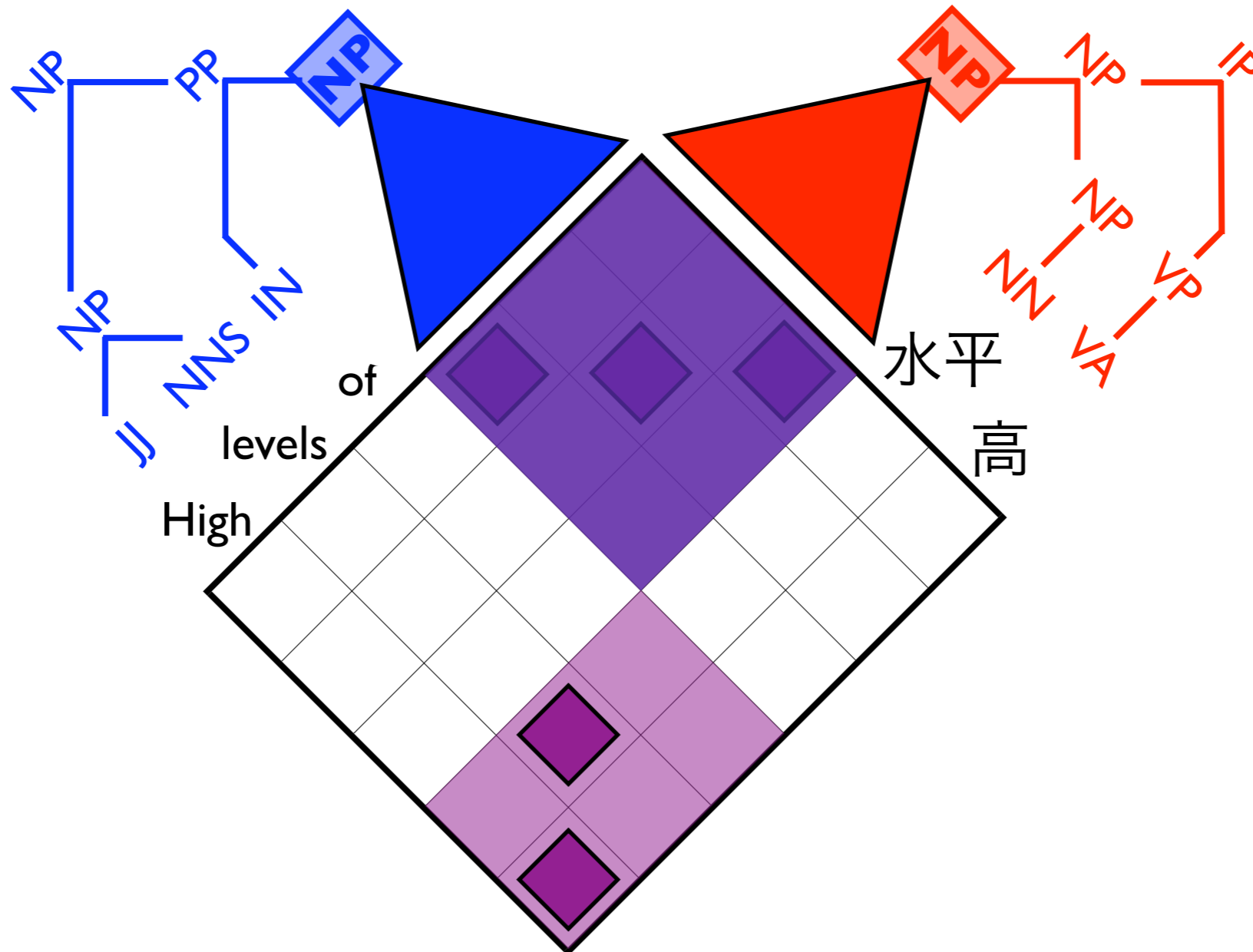




Target-Source Synchronization Features

$$\phi(n, b, n')$$

$$\text{SRCTRG} \langle \text{NP}, \text{NP} \rangle = 1$$



Berkeley



Approximate Viterbi Inference



Approximate Viterbi Inference

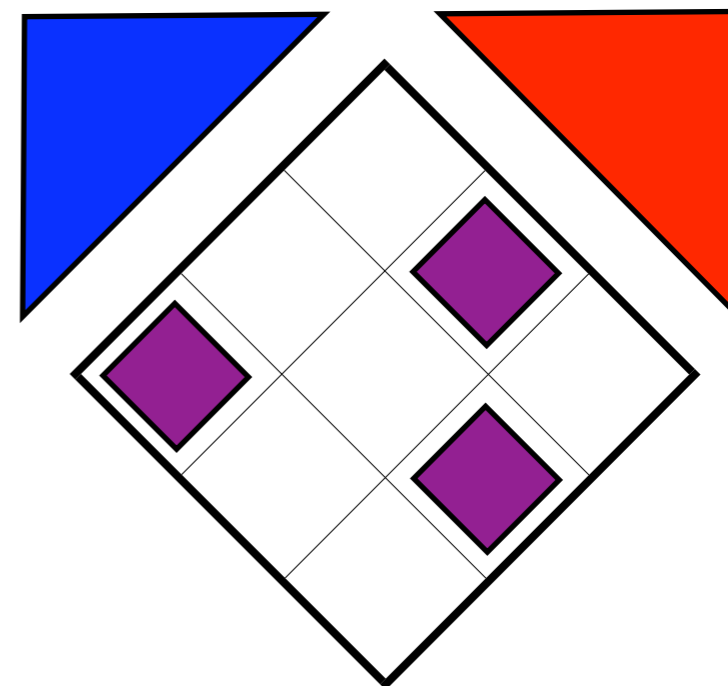
Given: s, s'



Approximate Viterbi Inference

Given: s, s'

Wanted: $\operatorname{argmax}_{(t, a, t')} P(t, a, t' | s, s')$



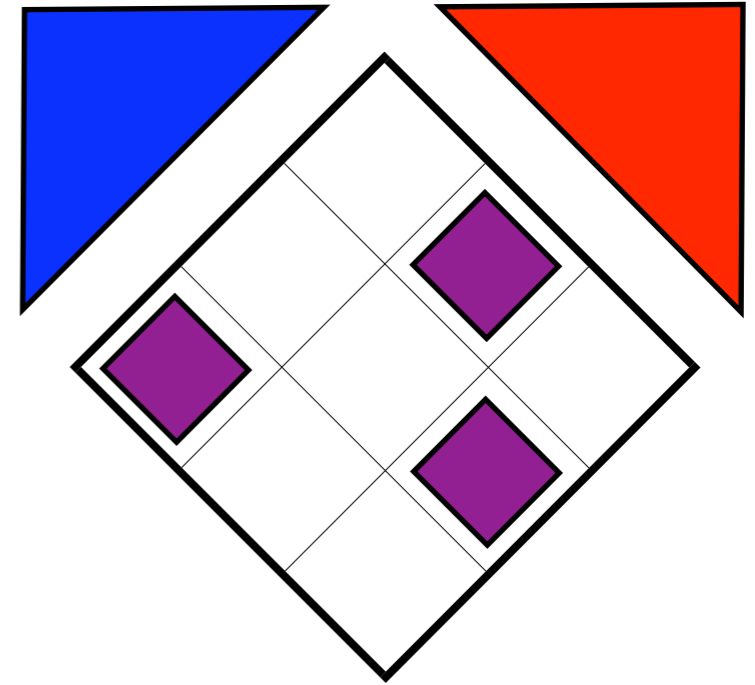


Approximate Viterbi Inference

Given: s, s'

Wanted: $\operatorname{argmax}_{(t, a, t')} P(t, a, t' | s, s')$

Problem: Synchronization features





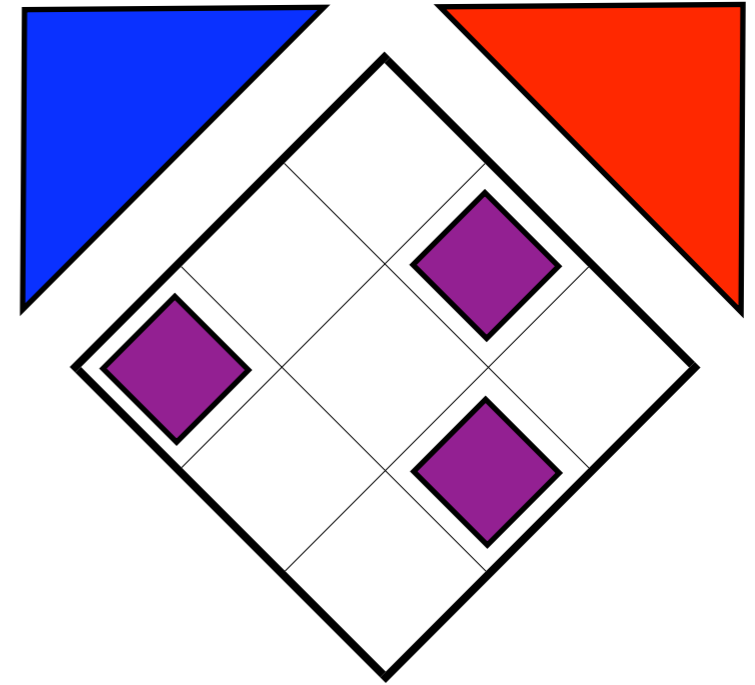
Approximate Viterbi Inference

Given: s, s'

Wanted: $\operatorname{argmax}_{(t, a, t')} P(t, a, t' | s, s')$

Problem: Synchronization features

Solution:





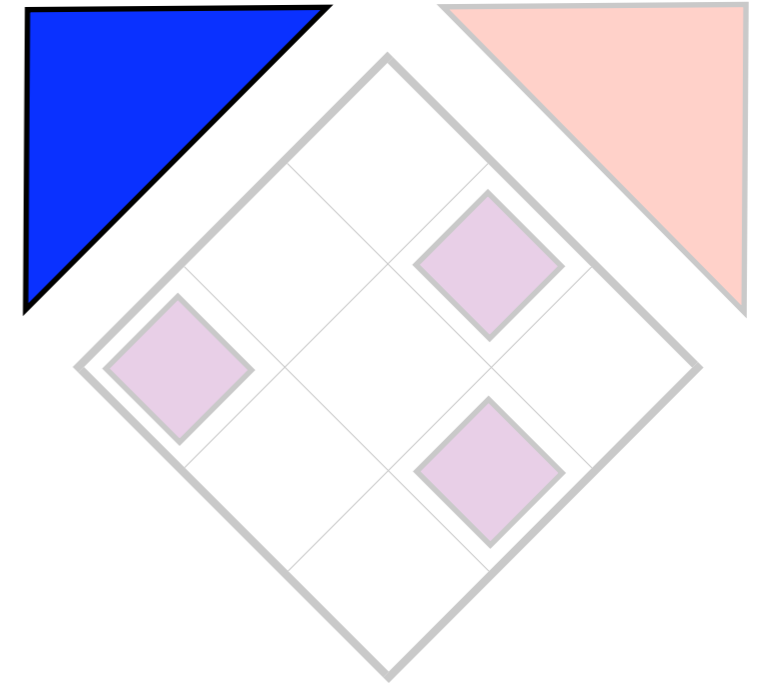
Approximate Viterbi Inference

Given: s, s'

Wanted: $\operatorname{argmax}_{(t, a, t')} P(t, a, t' | s, s')$

Problem: Synchronization features

Solution: I. Fix a, t' , find $\operatorname{argmax}_t P(t | a, t', s, s')$



Viterbi
Inside
Algorithm

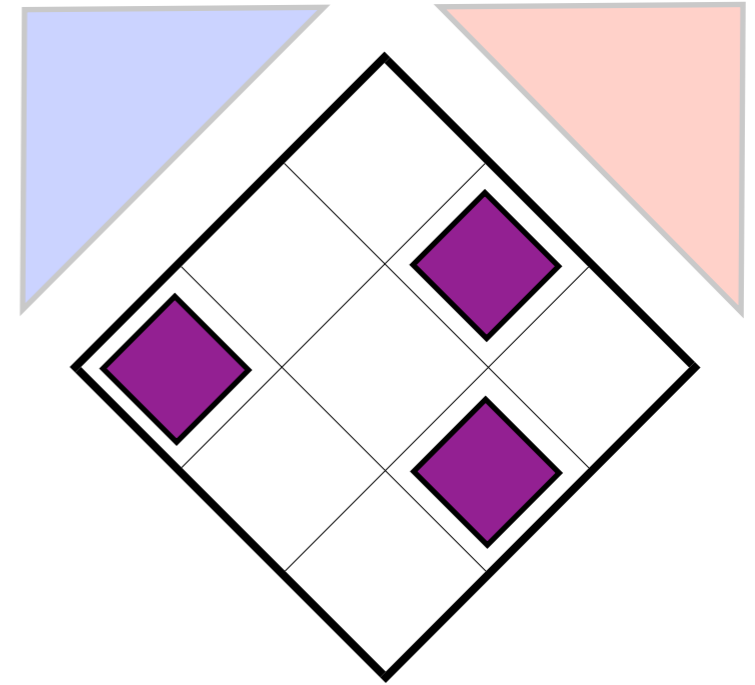


Approximate Viterbi Inference

Given: s, s'

Wanted: $\operatorname{argmax}_{(t, a, t')} P(t, a, t' | s, s')$

Problem: Synchronization features



Solution:

1. Fix a, t' , find $\operatorname{argmax}_t P(t | a, t', s, s')$
2. Fix t, t' , find $\operatorname{argmax}_a P(a | t, t', s, s')$

Viterbi
Bitext
Inside
Algorithm

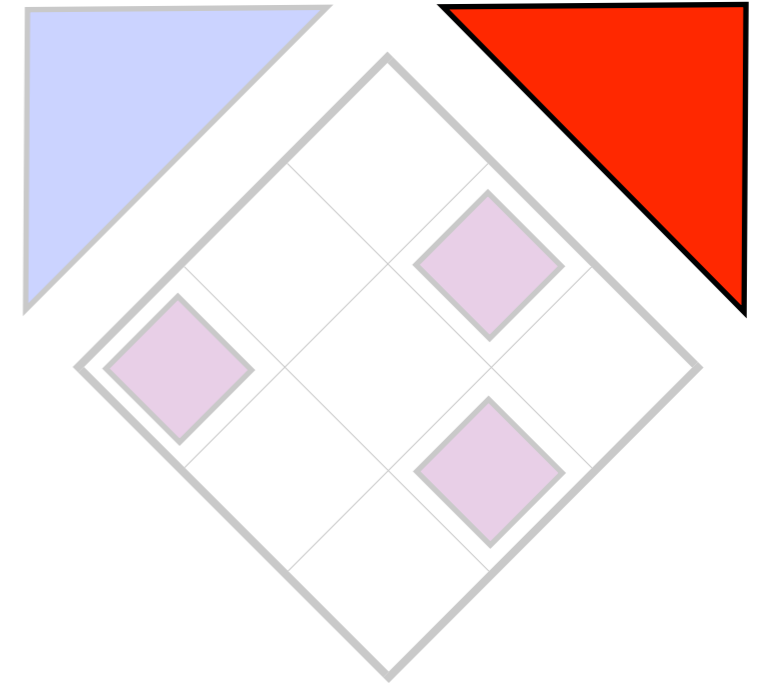


Approximate Viterbi Inference

Given: s, s'

Wanted: $\operatorname{argmax}_{(t, a, t')} P(t, a, t' | s, s')$

Problem: Synchronization features



Solution: 1. Fix a, t' , find $\operatorname{argmax}_t P(t | a, t', s, s')$

2. Fix t, t' , find $\operatorname{argmax}_a P(a | t, t', s, s')$

3. Fix t, a , find $\operatorname{argmax}_{t'} P(t' | t, a, s, s')$

Viterbi
Inside
Algorithm

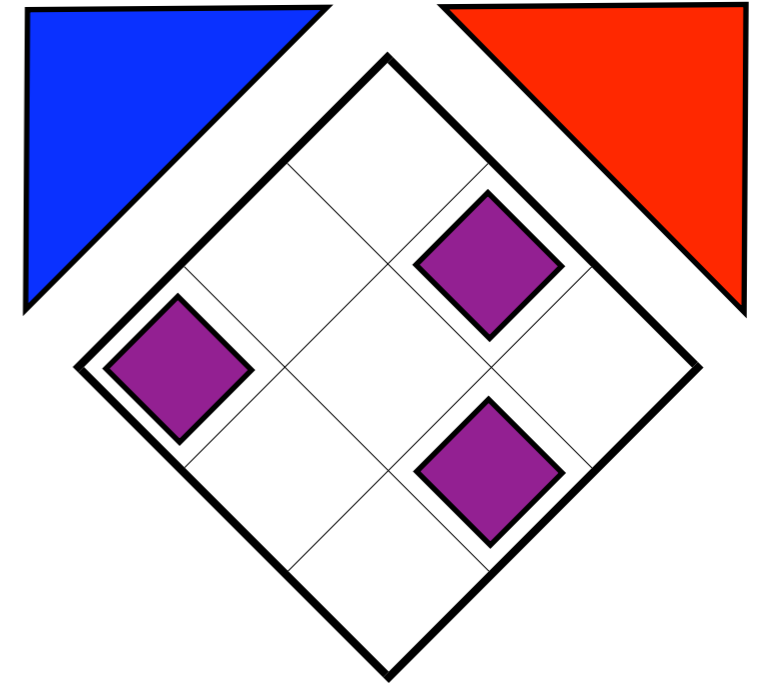


Approximate Viterbi Inference

Given: s, s'

Wanted: $\operatorname{argmax}_{(t, a, t')} P(t, a, t' | s, s')$

Problem: Synchronization features



Solution: 1. Fix a, t' , find $\operatorname{argmax}_t P(t | a, t', s, s')$

2. Fix t, t' , find $\operatorname{argmax}_a P(a | t, t', s, s')$

3. Fix t, a , find $\operatorname{argmax}_{t'} P(t' | t, a, s, s')$

4. Repeat until convergence

Berkeley



Structured Mean Field Inference



Structured Mean Field Inference

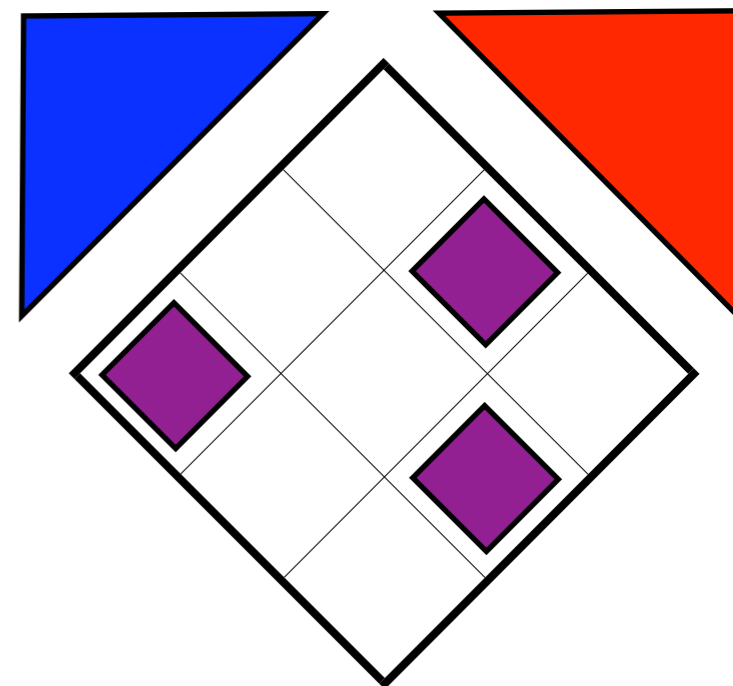
Given: s, s'



Structured Mean Field Inference

Given: s, s'

Wanted: $P(t, a, t' | s, s')$



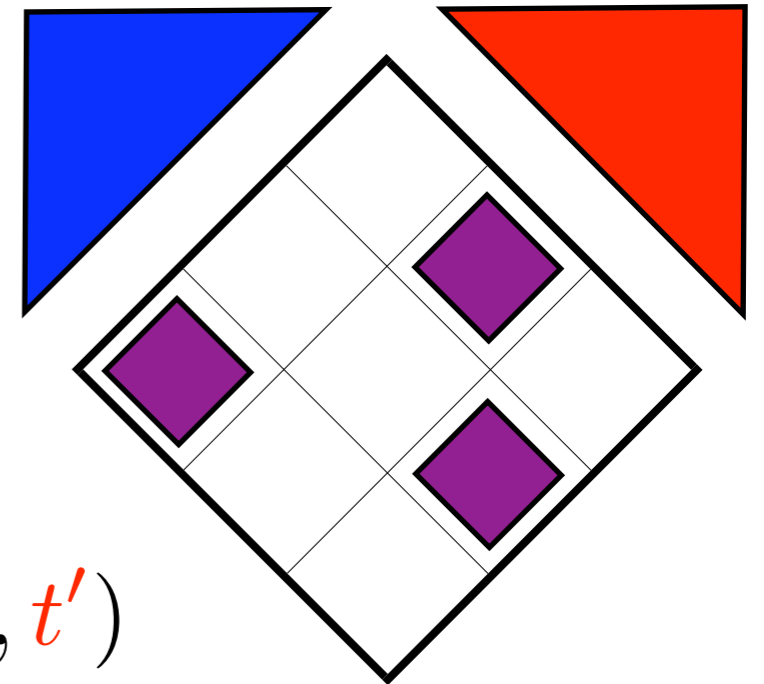


Structured Mean Field Inference

Given: s, s'

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t')$
 $= q(t)q(a)q(t')$



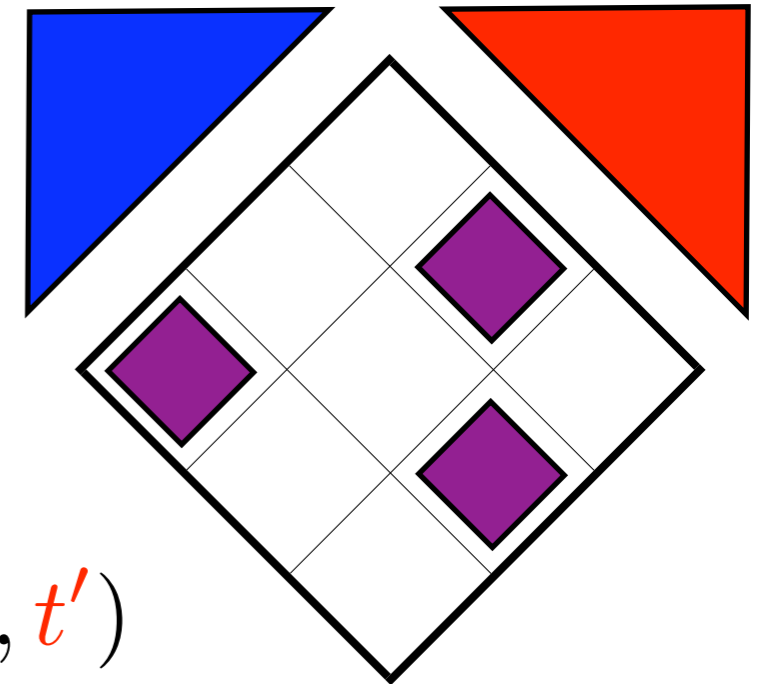


Structured Mean Field Inference

Given: s, s'

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t')$
 $= q(t)q(a)q(t')$



Algorithm:



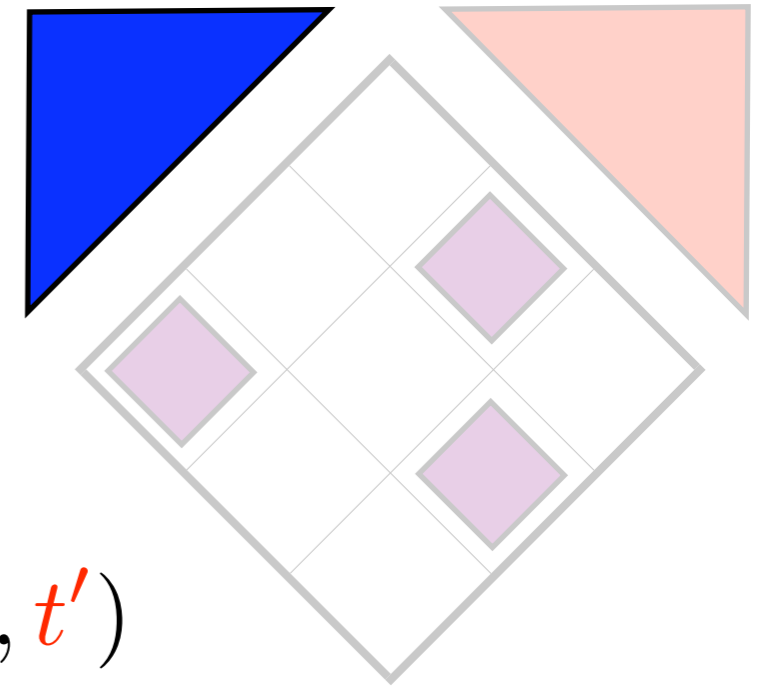
Structured Mean Field Inference

Given: s, s'

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t')$
 $= q(t)q(a)q(t')$

Algorithm: 1. $q(t) \propto \exp(\langle \theta, \phi(t, E_q(a), E_q(t')) \rangle)$



Inside-
Outside
Algorithm

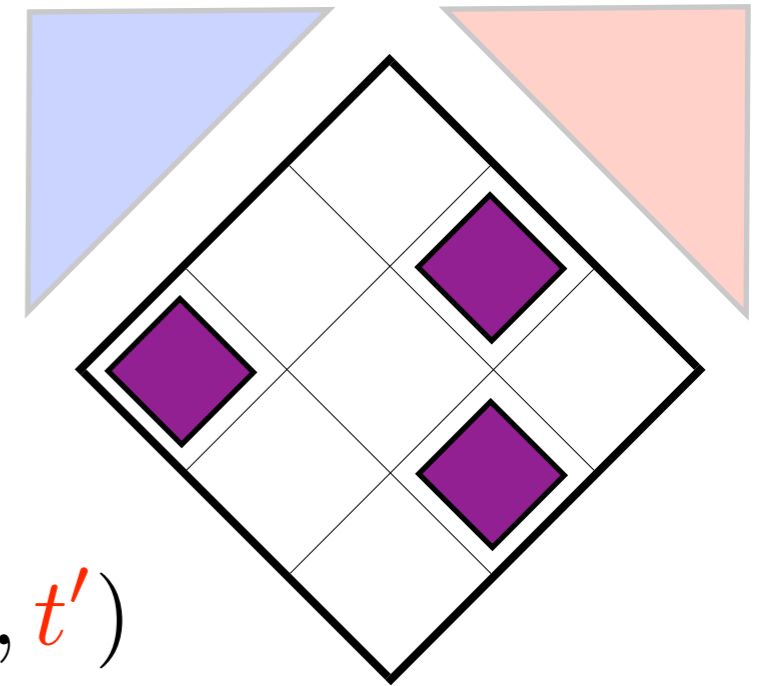


Structured Mean Field Inference

Given: s, s'

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t')$
 $= q(t)q(a)q(t')$



Algorithm:

1. $q(t) \propto \exp(\langle \theta, \phi(t, E_q(a), E_q(t')) \rangle)$
2. $q(a) \propto \exp(\langle \theta, \phi(E_q(t), a, E_q(t')) \rangle)$

Bitext
Inside-
Outside
Algorithm

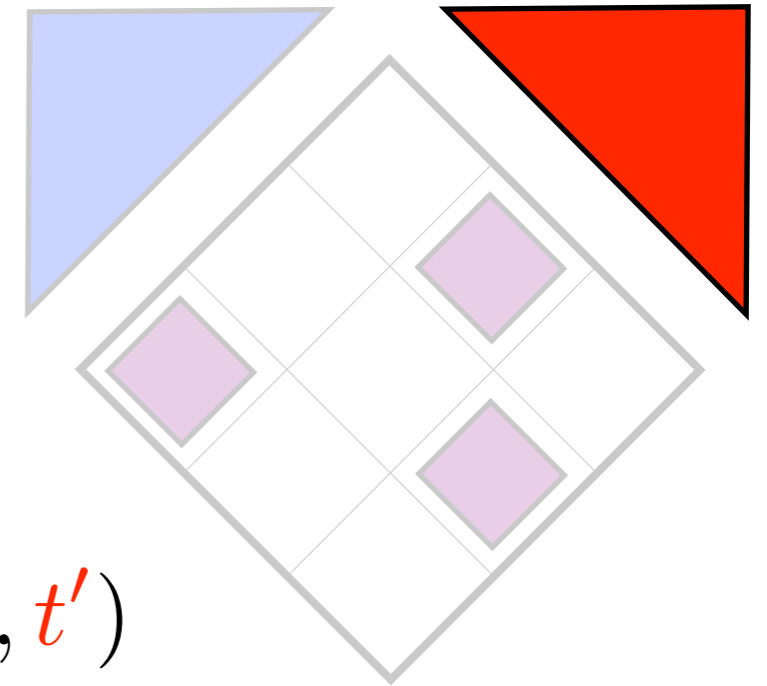


Structured Mean Field Inference

Given: s, s'

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t')$
 $= q(t)q(a)q(t')$



Algorithm: 1. $q(t) \propto \exp(\langle \theta, \phi(t, E_q(a), E_q(t')) \rangle)$

2. $q(a) \propto \exp(\langle \theta, \phi(E_q(t), a, E_q(t')) \rangle)$

3. $q(t') \propto \exp(\langle \theta, \phi(E_q(t), E_q(a), t') \rangle)$

Inside-
Outside
Algorithm

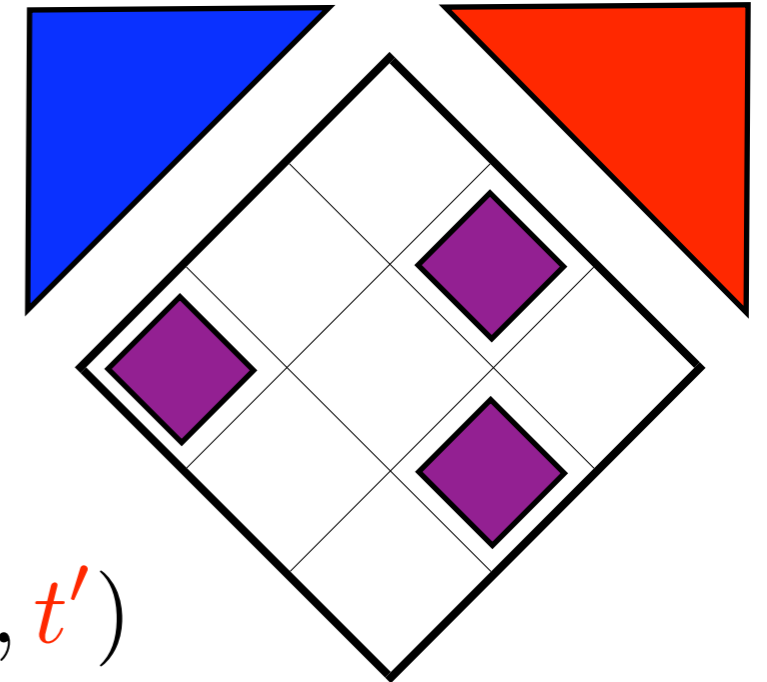


Structured Mean Field Inference

Given: s, s'

Wanted: $P(t, a, t' | s, s')$

Approximation: $P(t, a, t' | s, s') \approx Q(t, a, t')$
 $= q(t)q(a)q(t')$



- Algorithm:
1. $q(t) \propto \exp(\langle \theta, \phi(t, E_q(a), E_q(t')) \rangle)$
 2. $q(a) \propto \exp(\langle \theta, \phi(E_q(t), a, E_q(t')) \rangle)$
 3. $q(t') \propto \exp(\langle \theta, \phi(E_q(t), E_q(a), t') \rangle)$
 4. Repeat until convergence

Berkeley

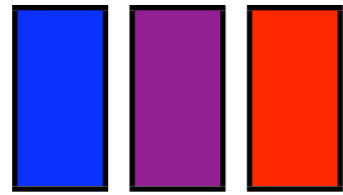


Training Data

Berkeley



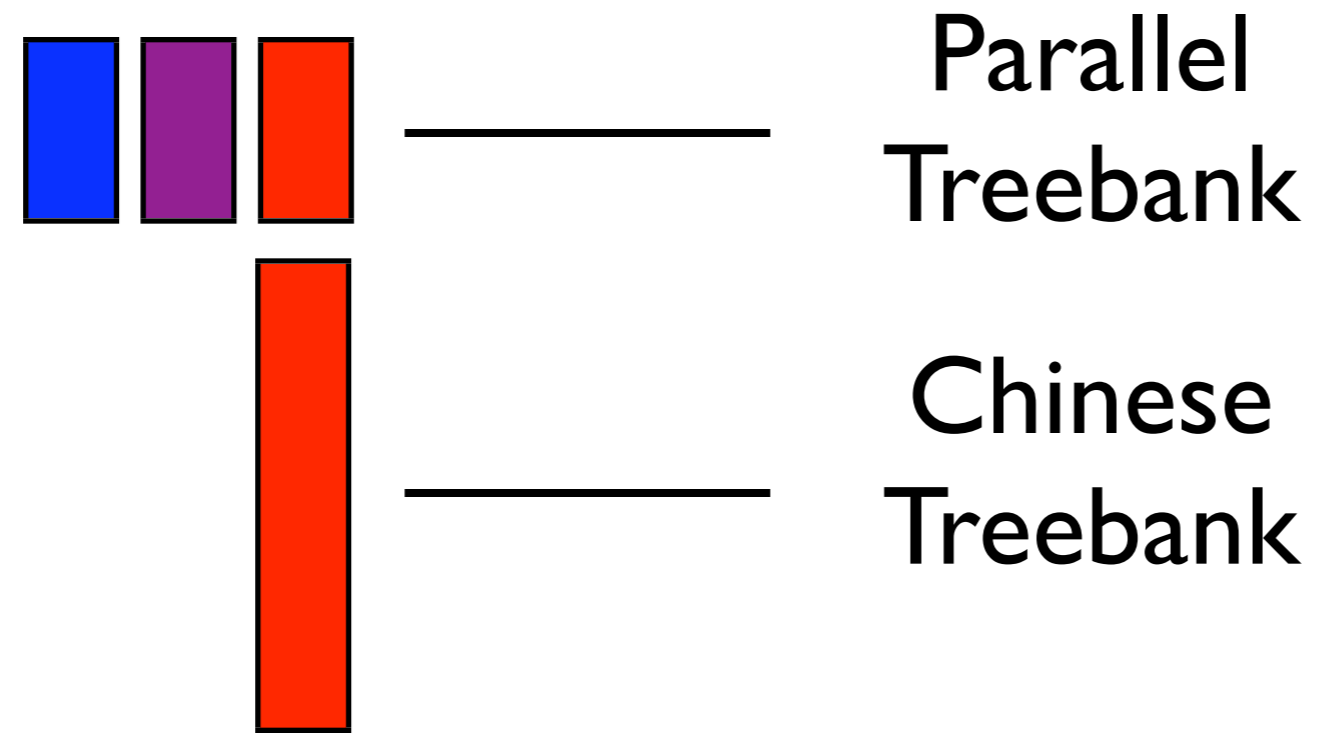
Training Data



Parallel
Treebank

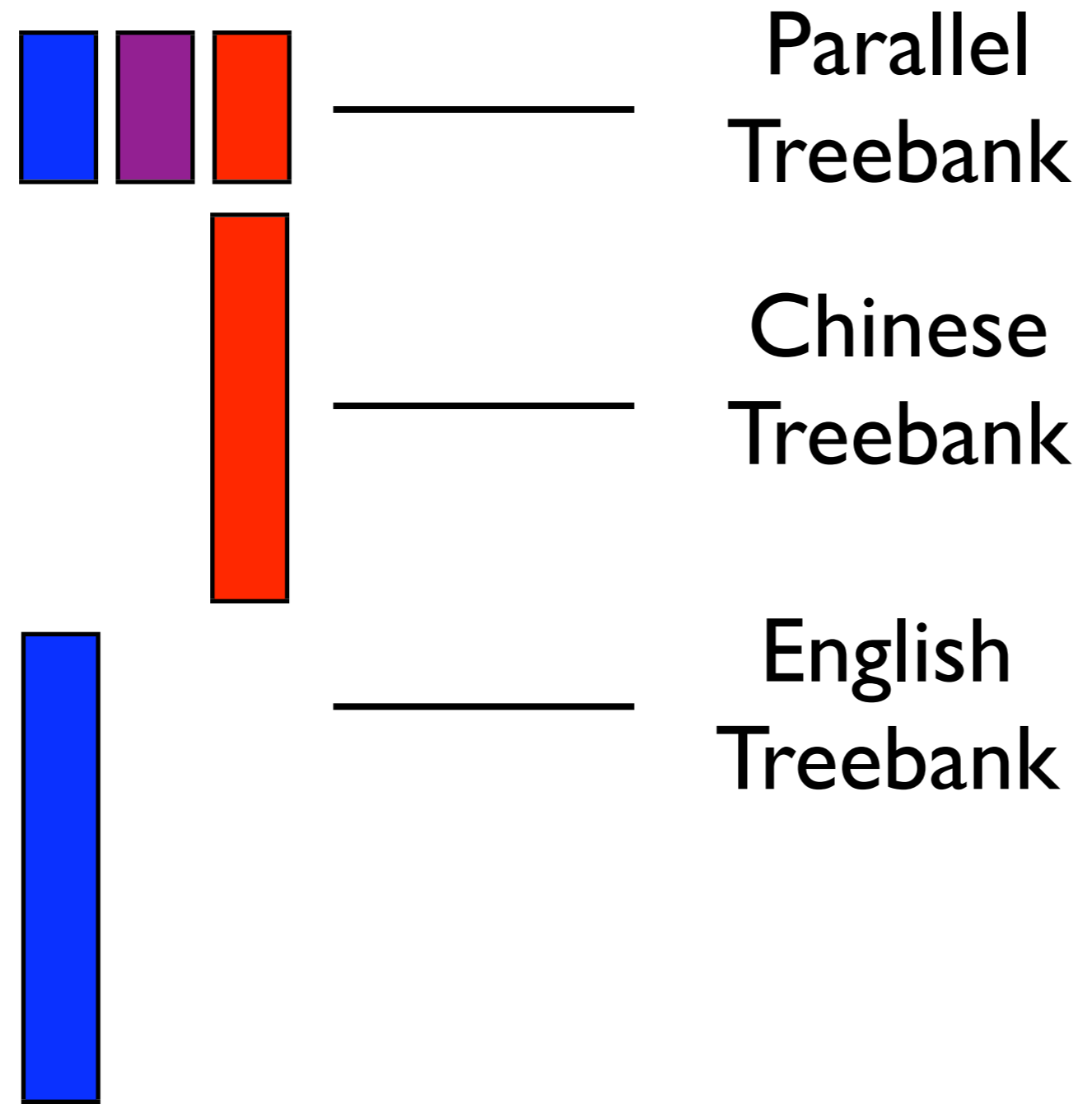


Training Data



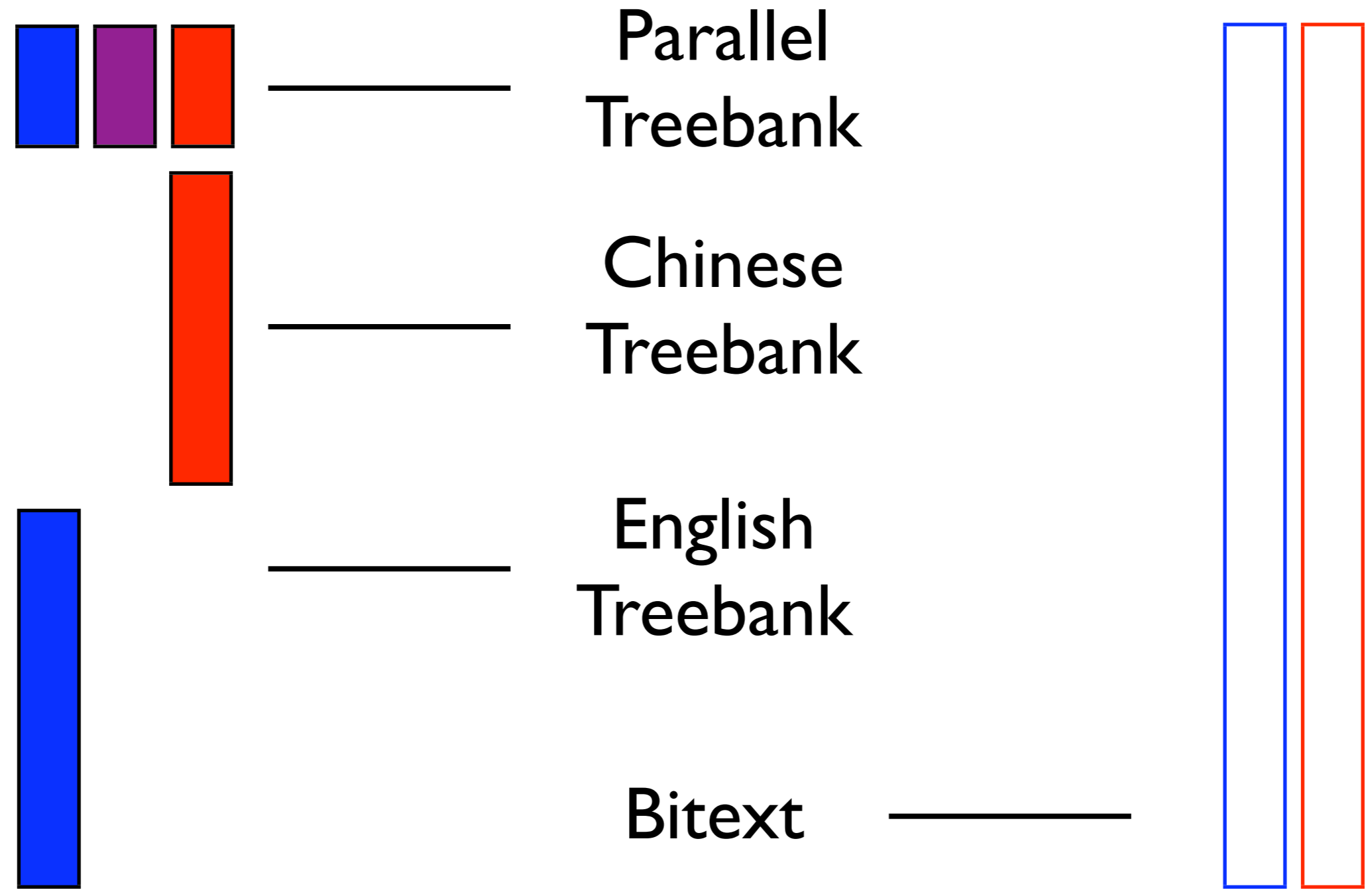


Training Data





Training Data



Parallel
Treebank

Chinese
Treebank

English
Treebank

Bitext

Berkeley

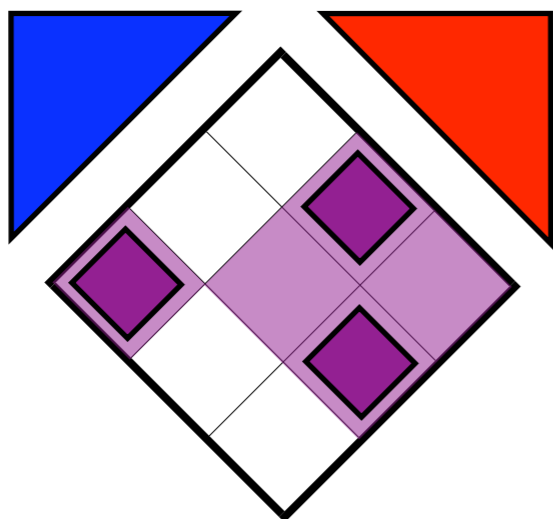


Incomplete Training Data



Incomplete Training Data

Required data:

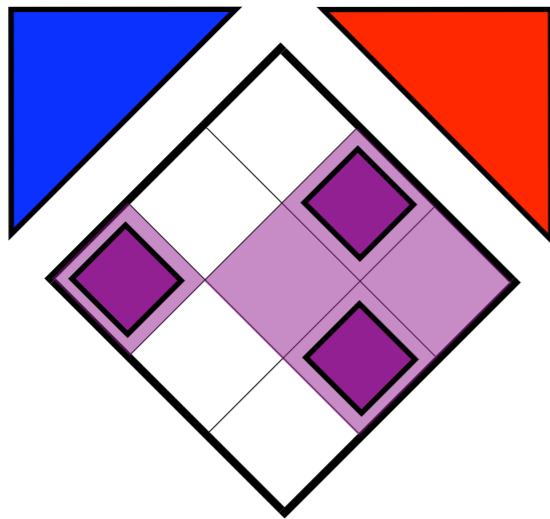


$$(t_i, a_i, t'_i)$$



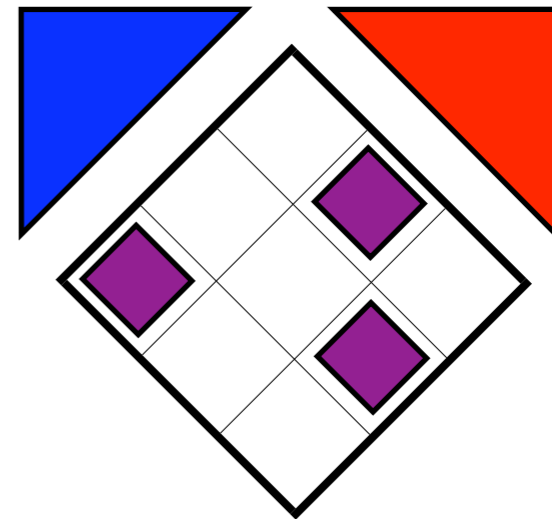
Incomplete Training Data

Required data:



$$(t_i, a_i, t'_i)$$

Actual data:



$$(t_i, w_i, t'_i)$$

Berkeley

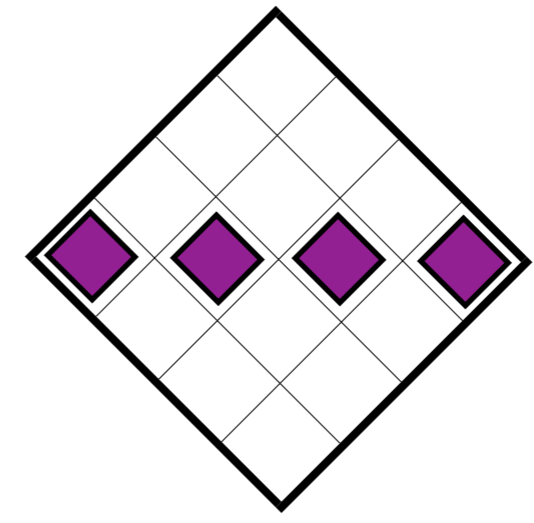


Incomplete Training Data



Incomplete Training Data

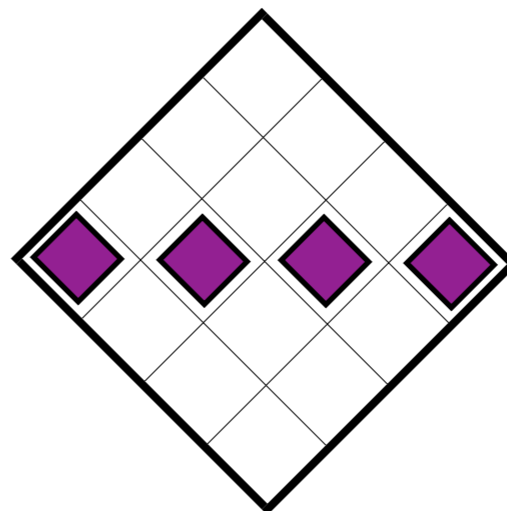
Word alignment w



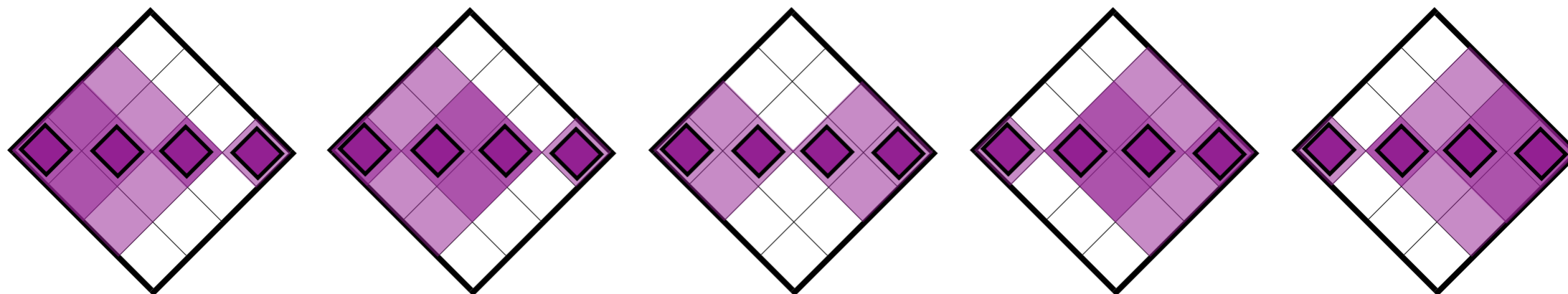


Incomplete Training Data

Word alignment w



Legal alignments a





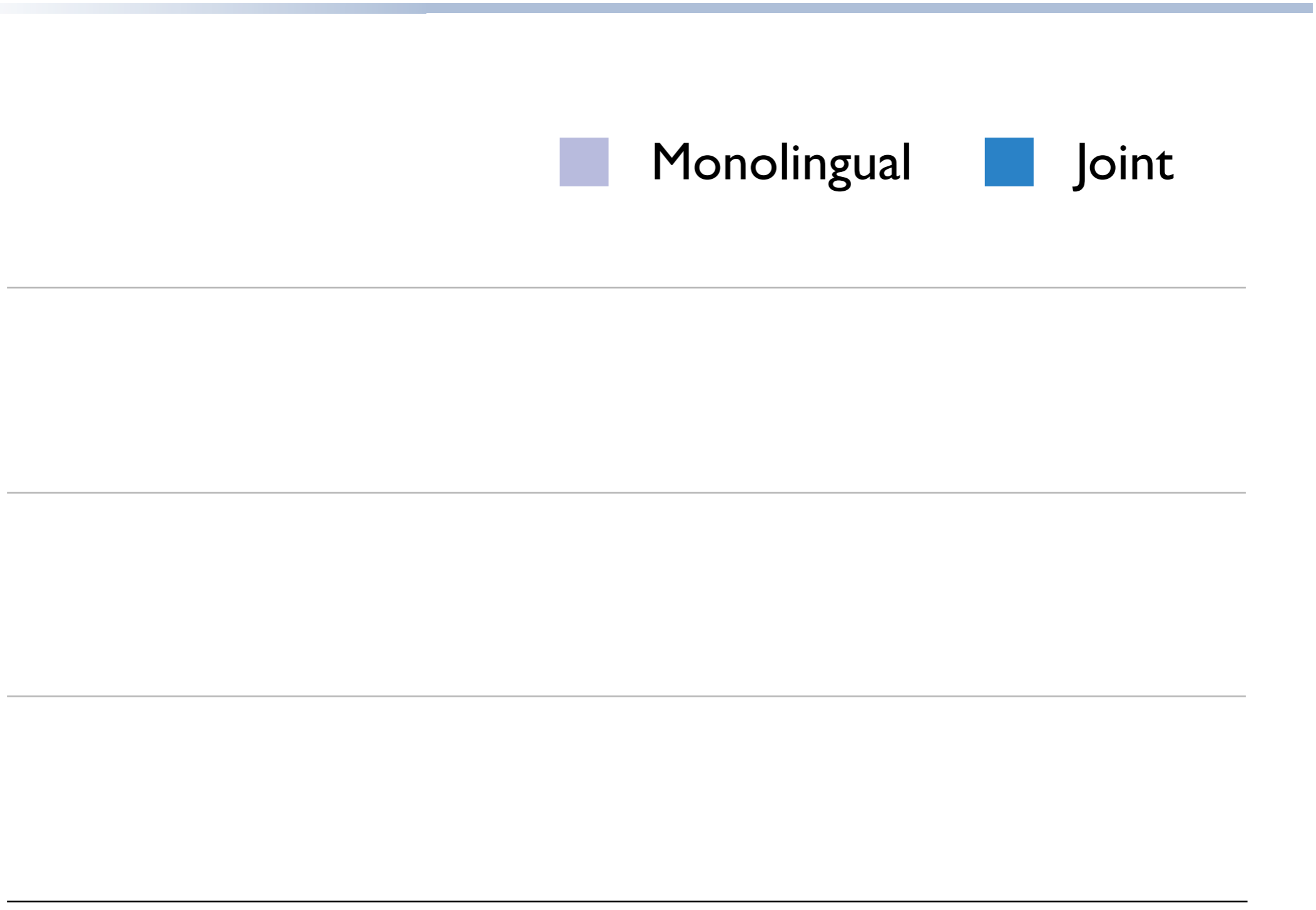
Parsing Improvement

■ Monolingual ■ Joint

86
84
82
80

Chinese

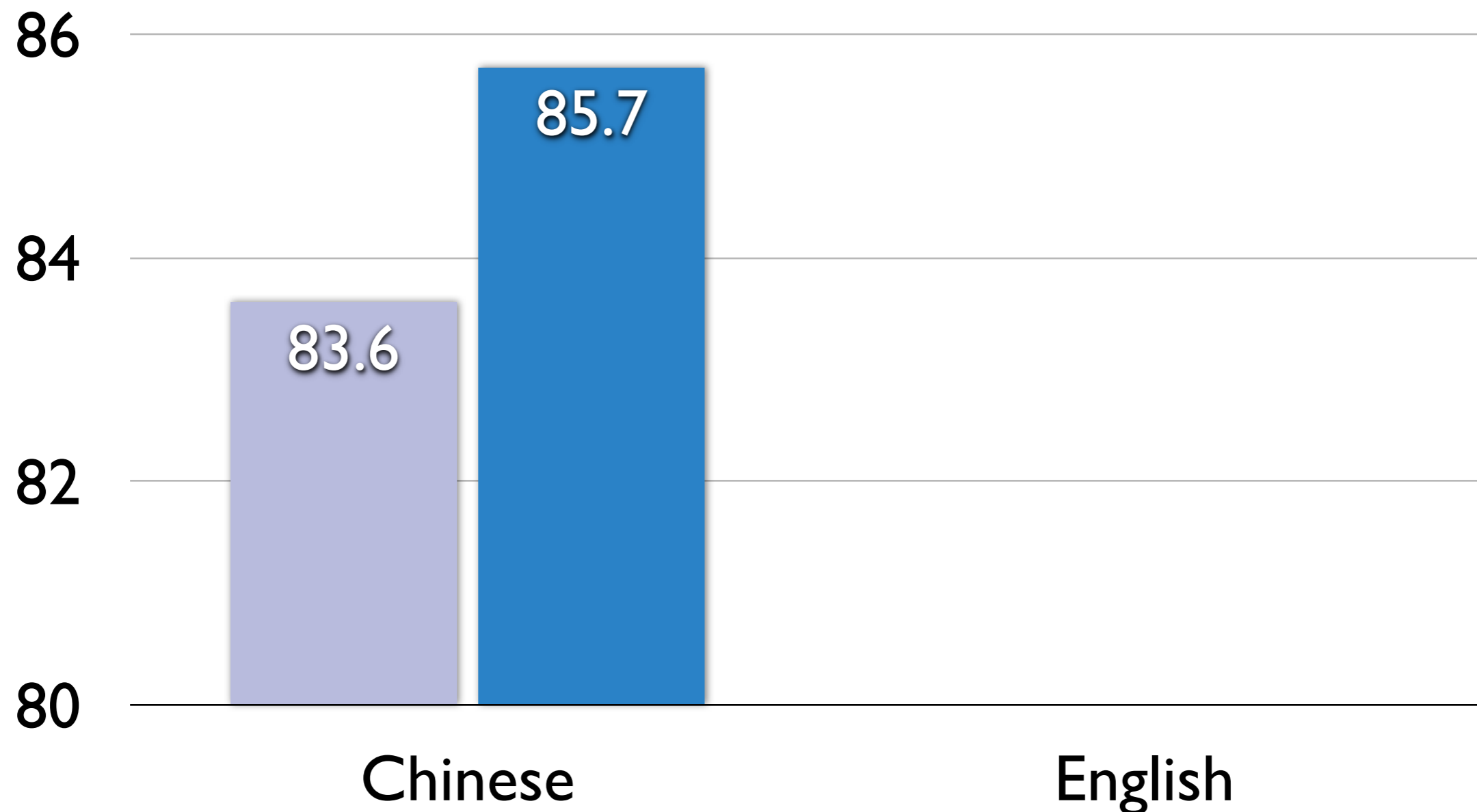
English





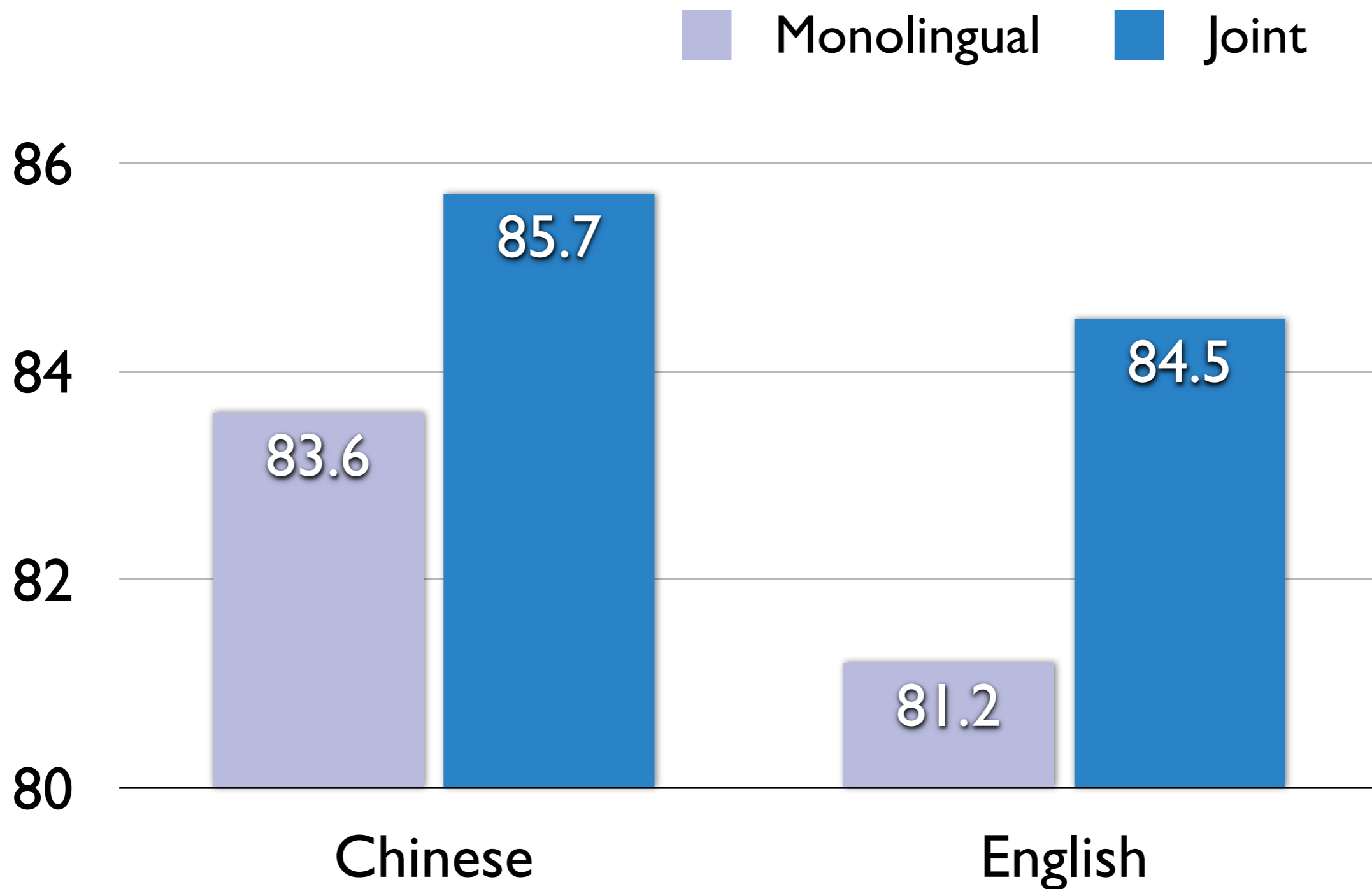
Parsing Improvement

■ Monolingual ■ Joint





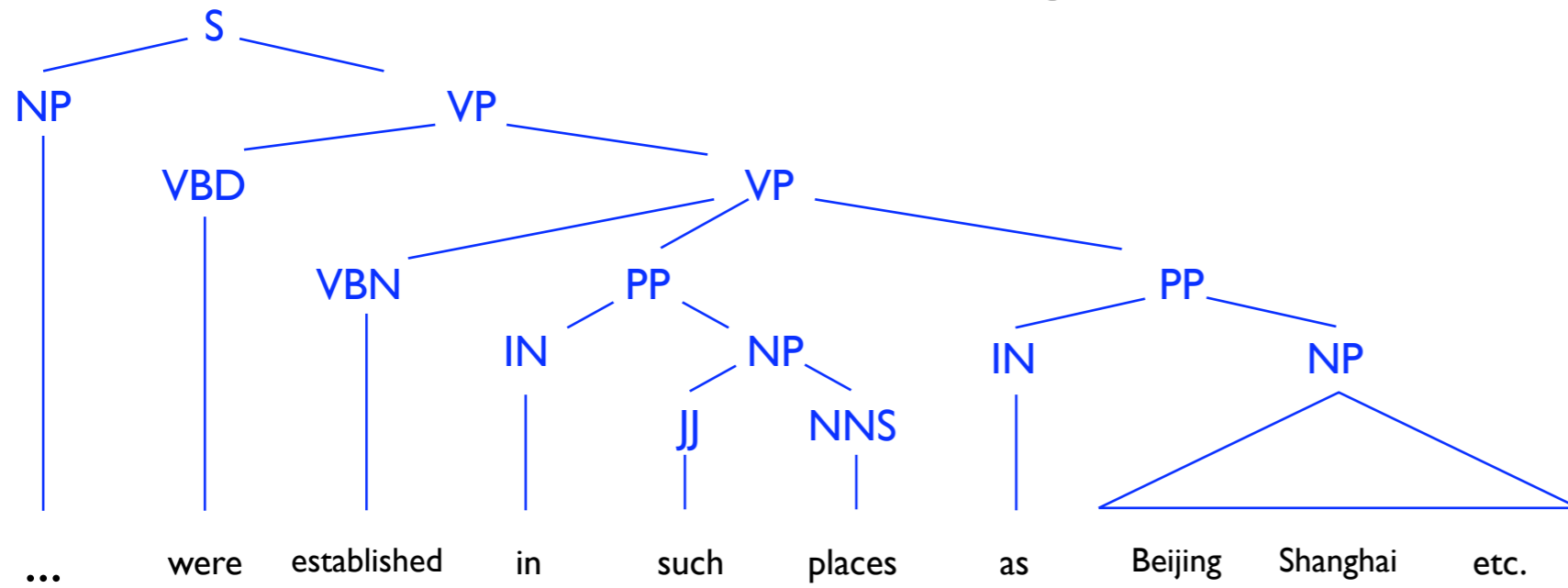
Parsing Improvement



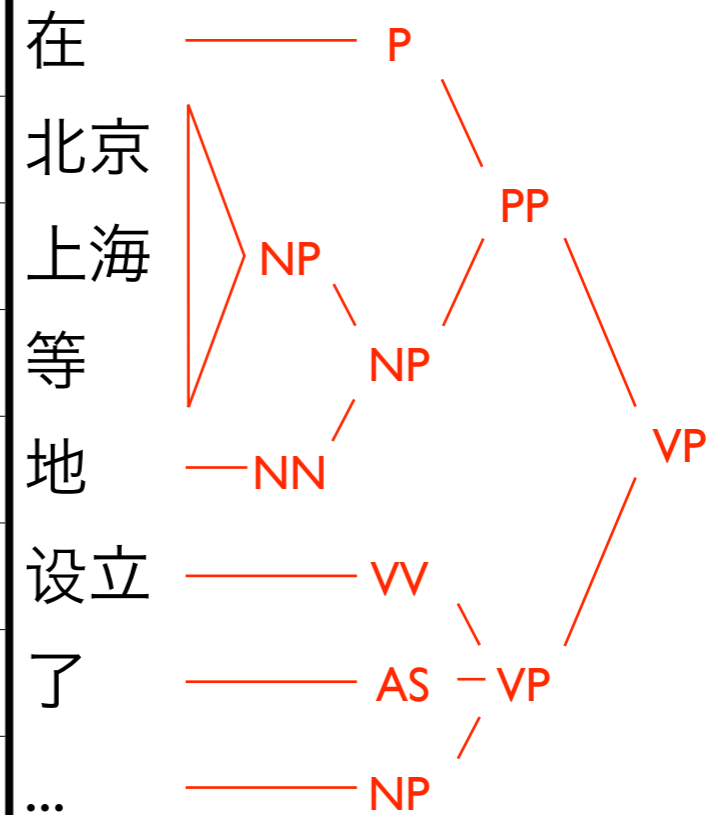


Parsing Improvement

Incorrect English PP Attachment



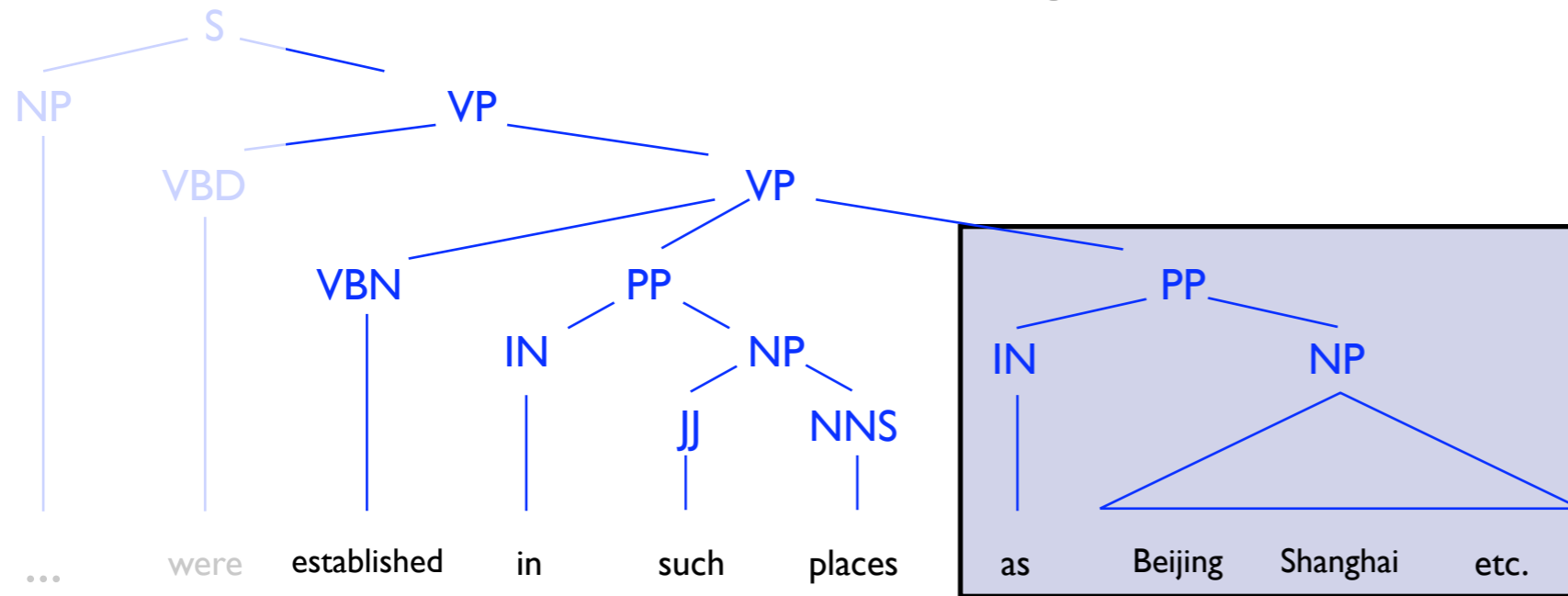
			在						
							北京		
								上海	
									等
									地
									设立
									了
									...



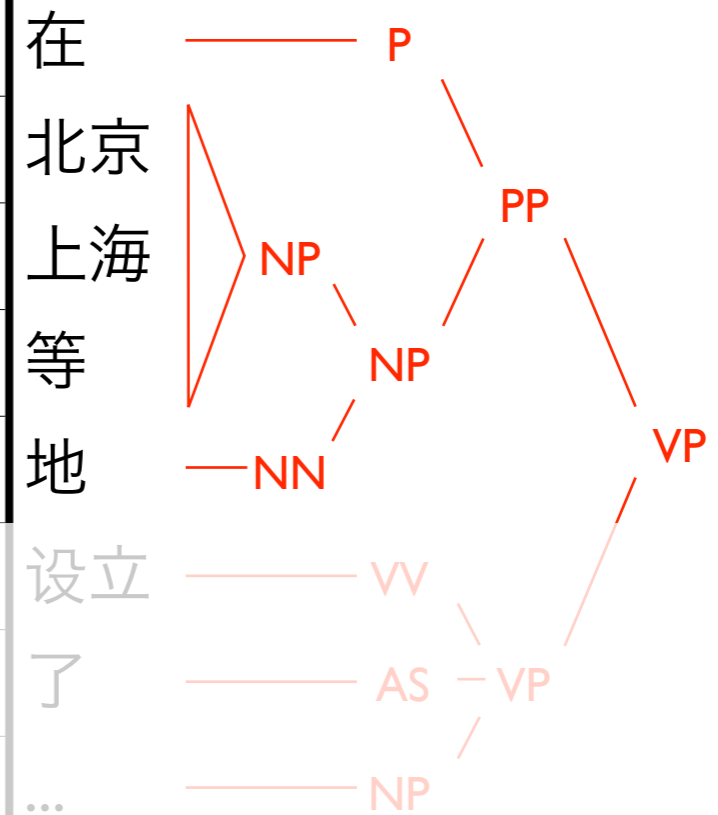


Parsing Improvement

Incorrect English PP Attachment



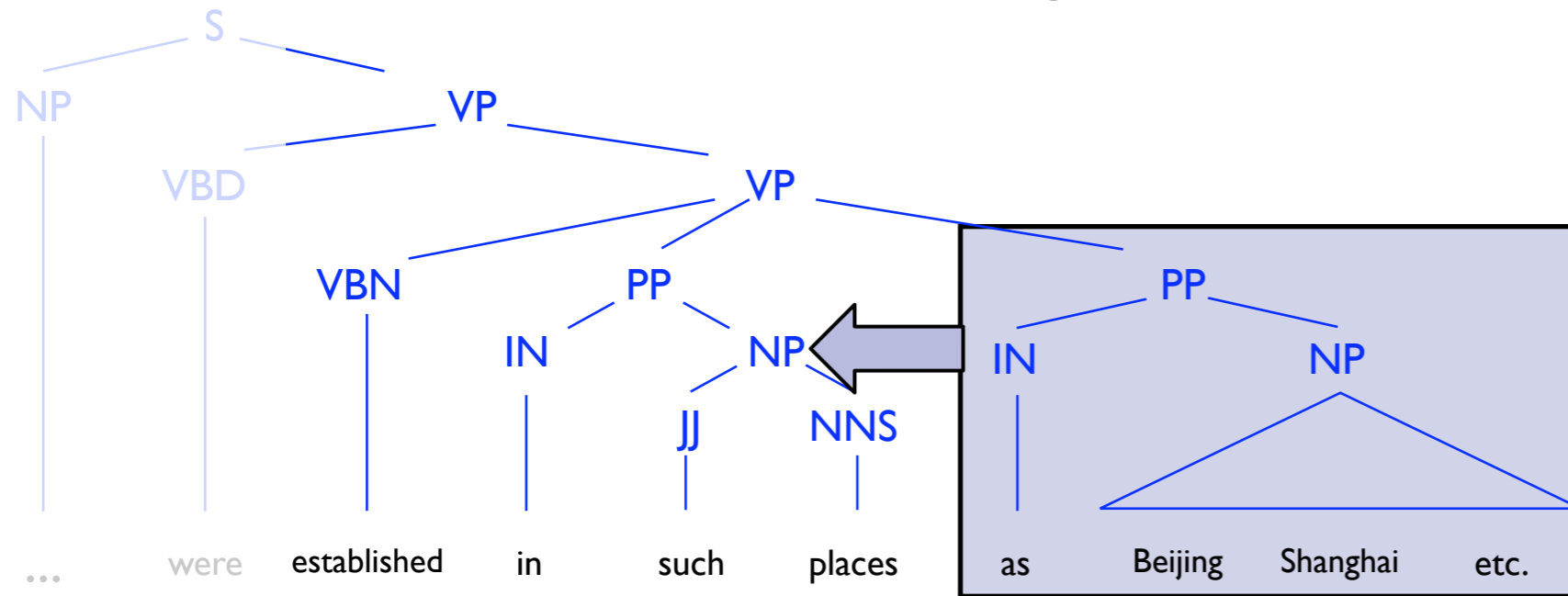
			在						
			北京						
			上海						
			等						
			地						
		设立							
		了							
		...							

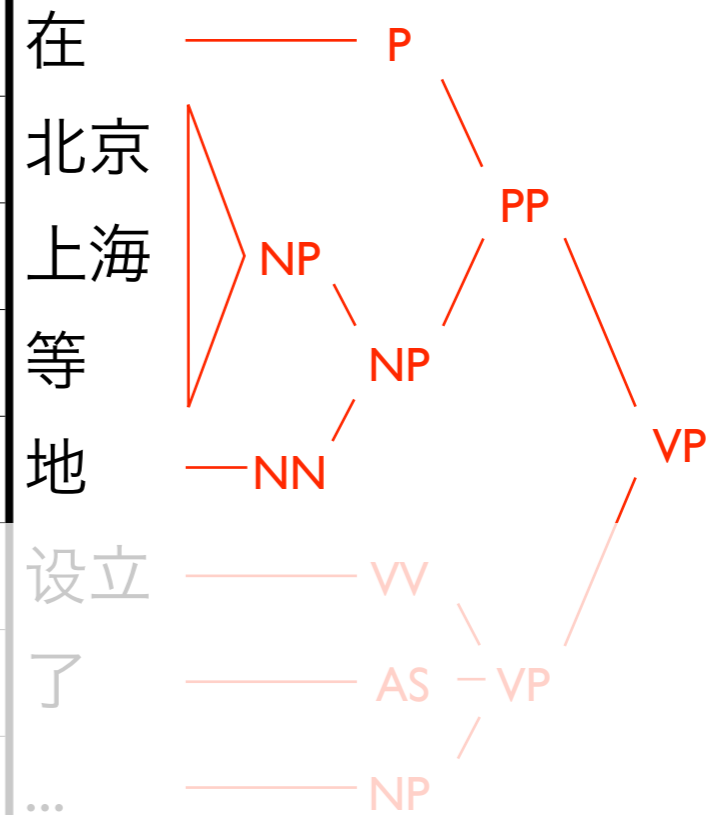




Parsing Improvement

Incorrect English PP Attachment

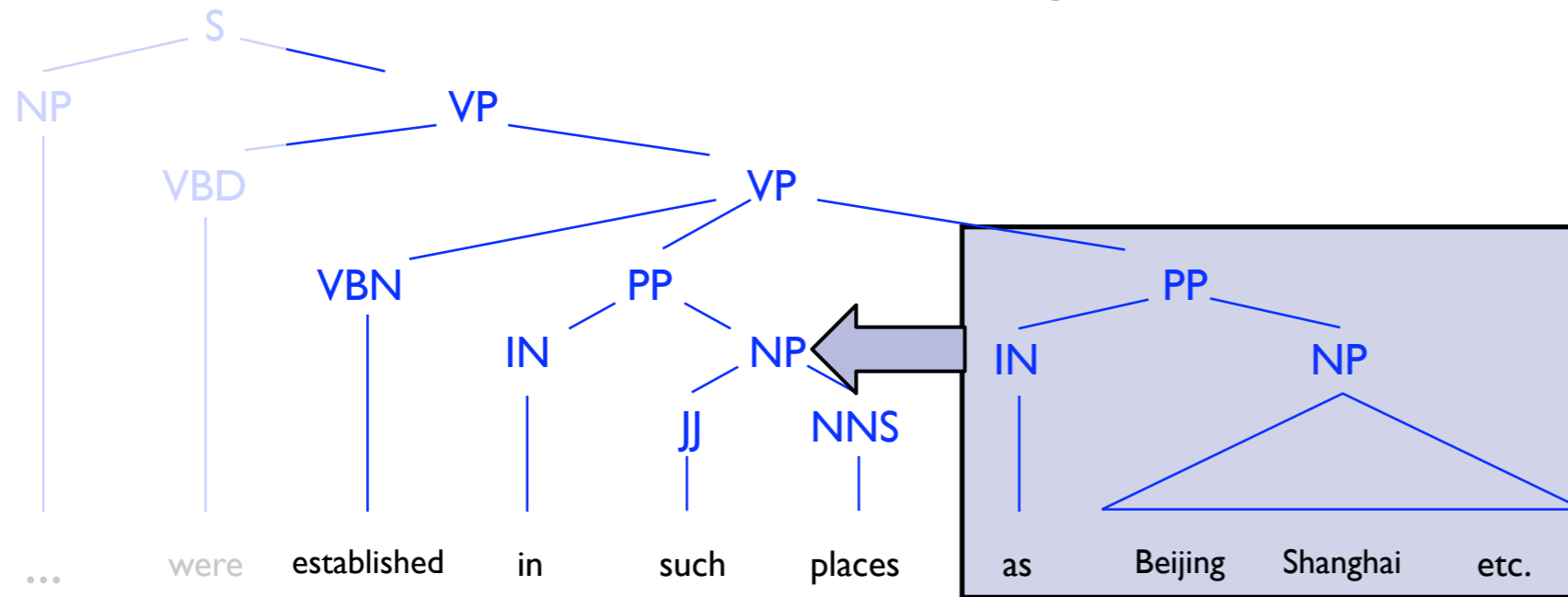


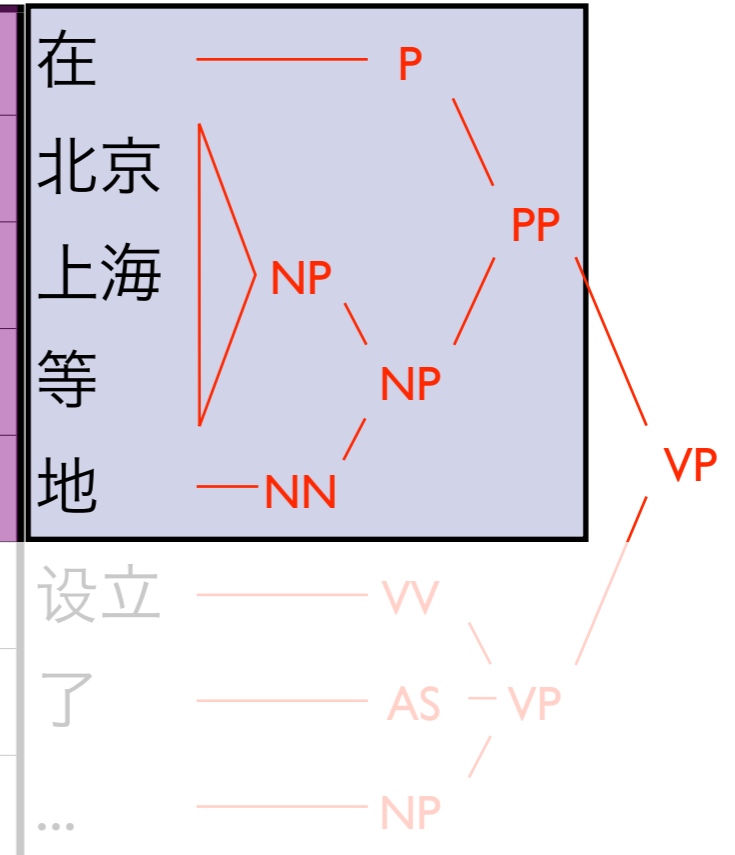




Parsing Improvement

Incorrect English PP Attachment

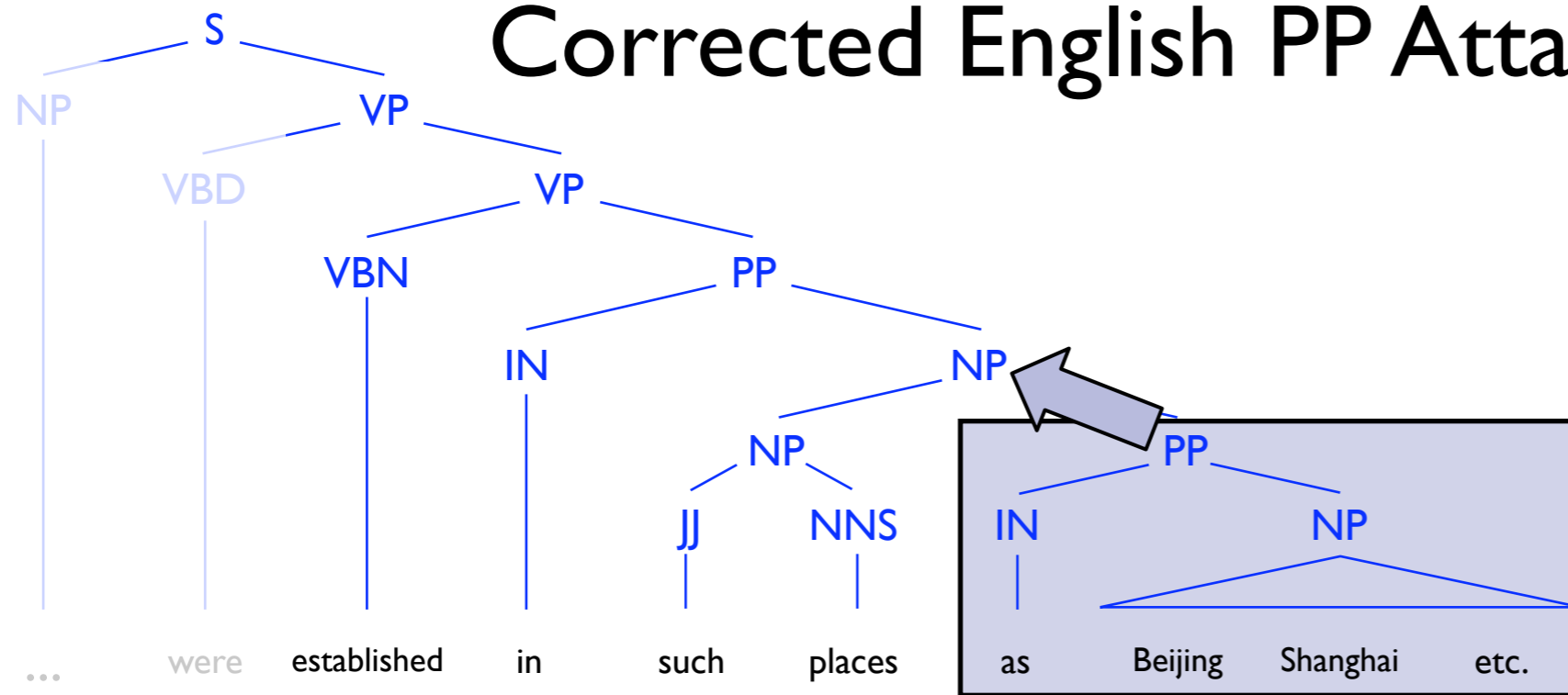




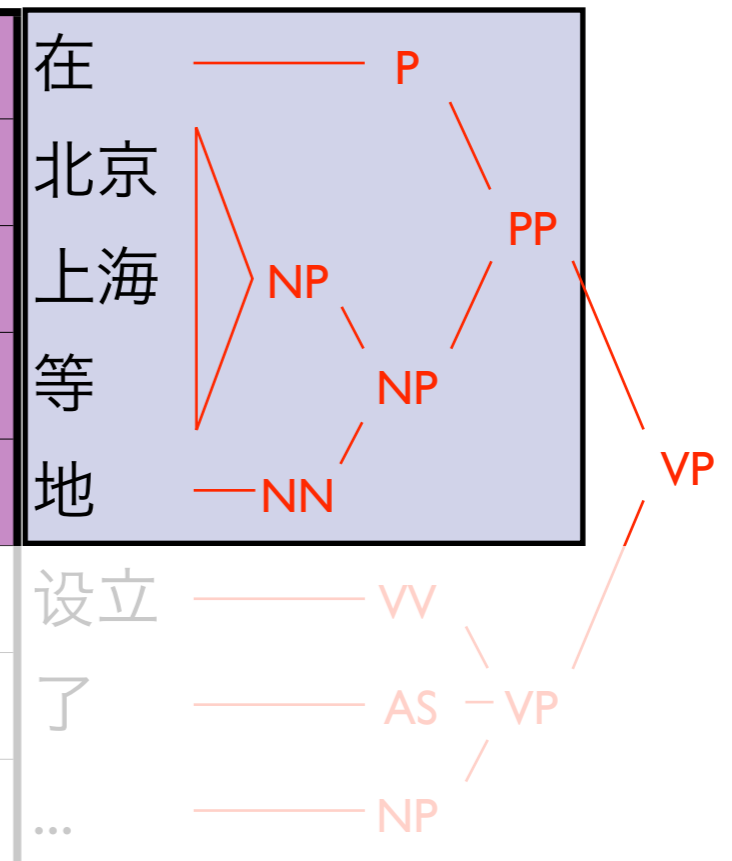


Parsing Improvement

Corrected English PP Attachment

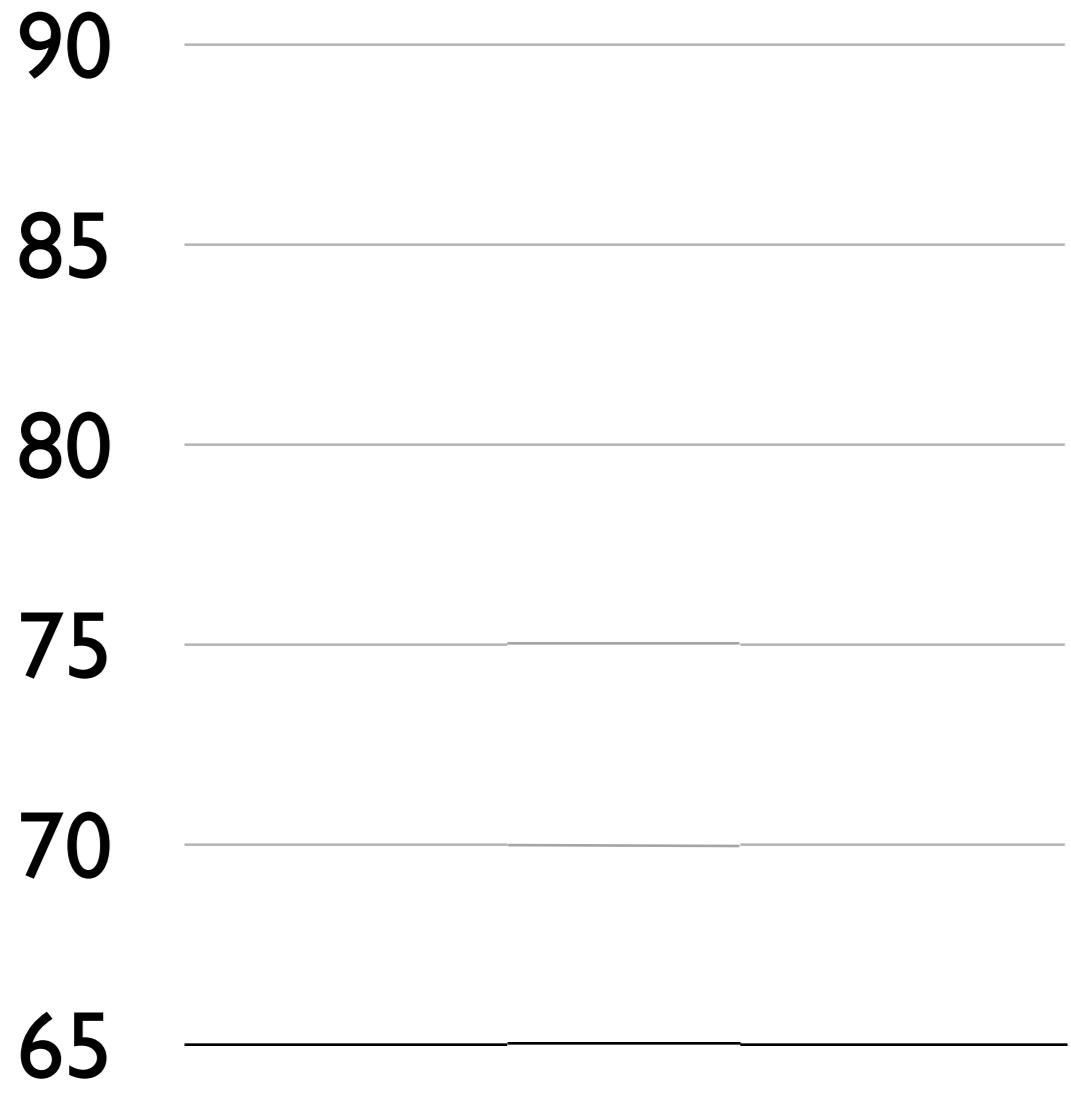


			█						
							█		
								█	
									█
					█				
		█							
		█							
█									





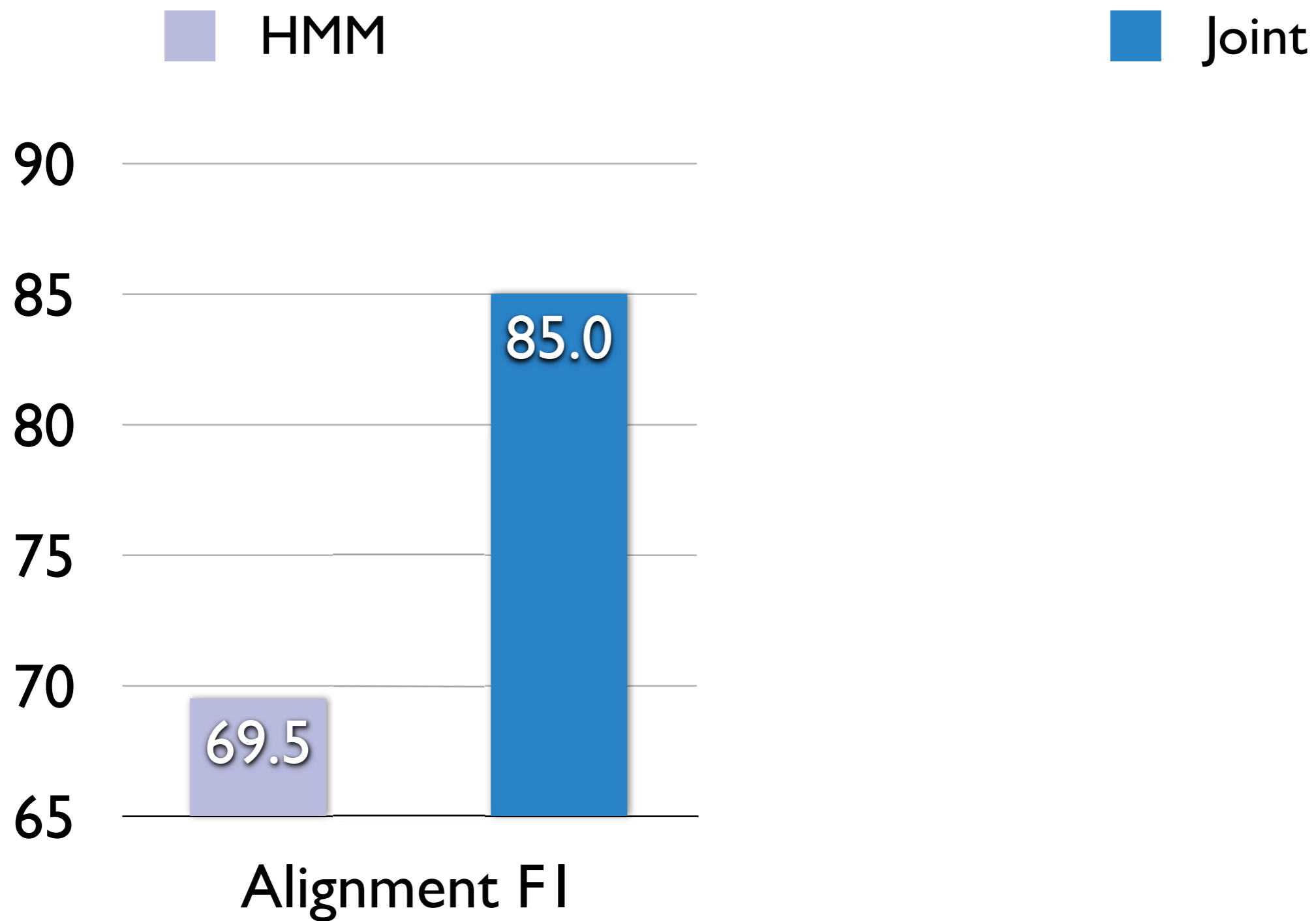
Word Alignment Improvement



Alignment F1

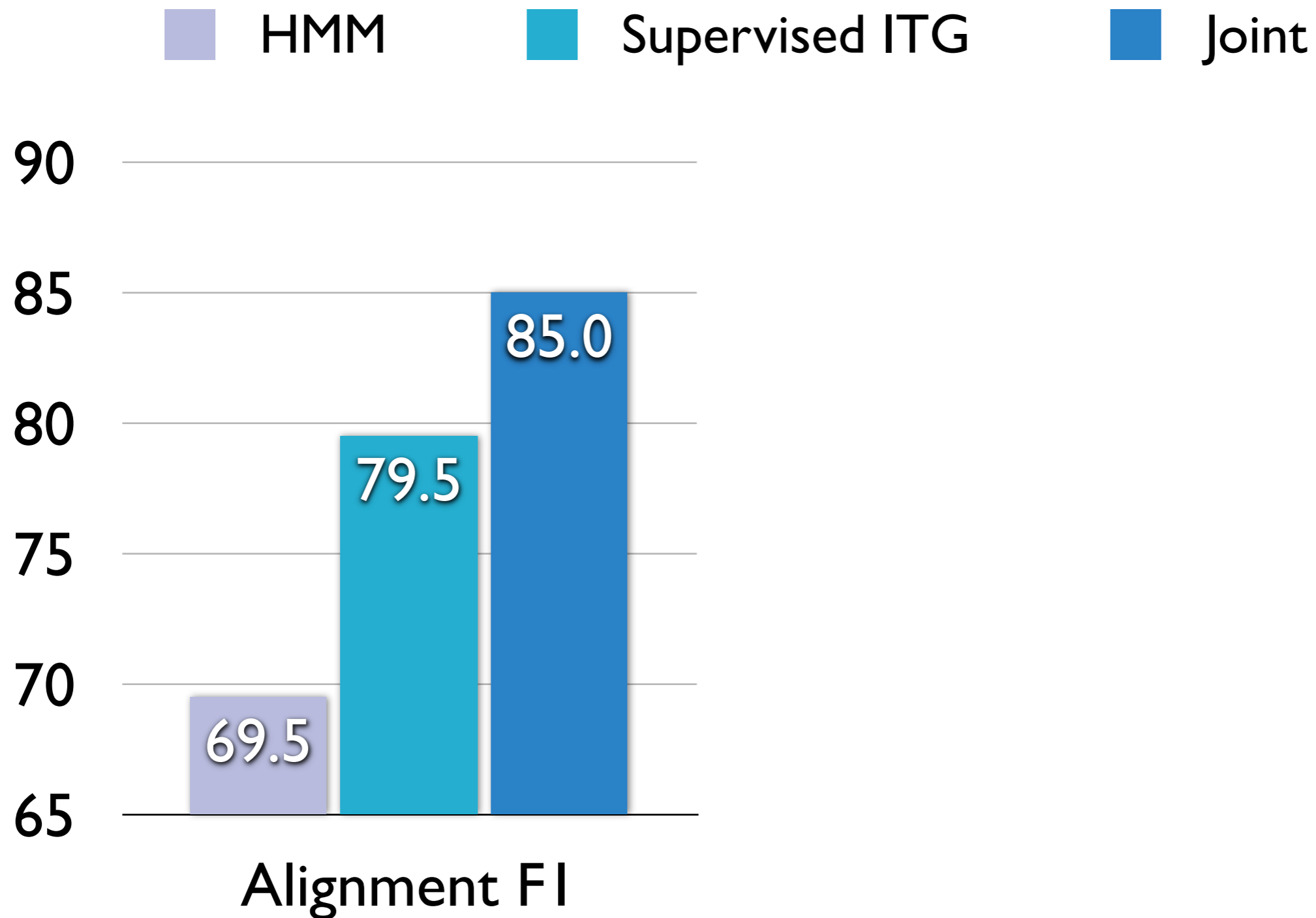


Word Alignment Improvement





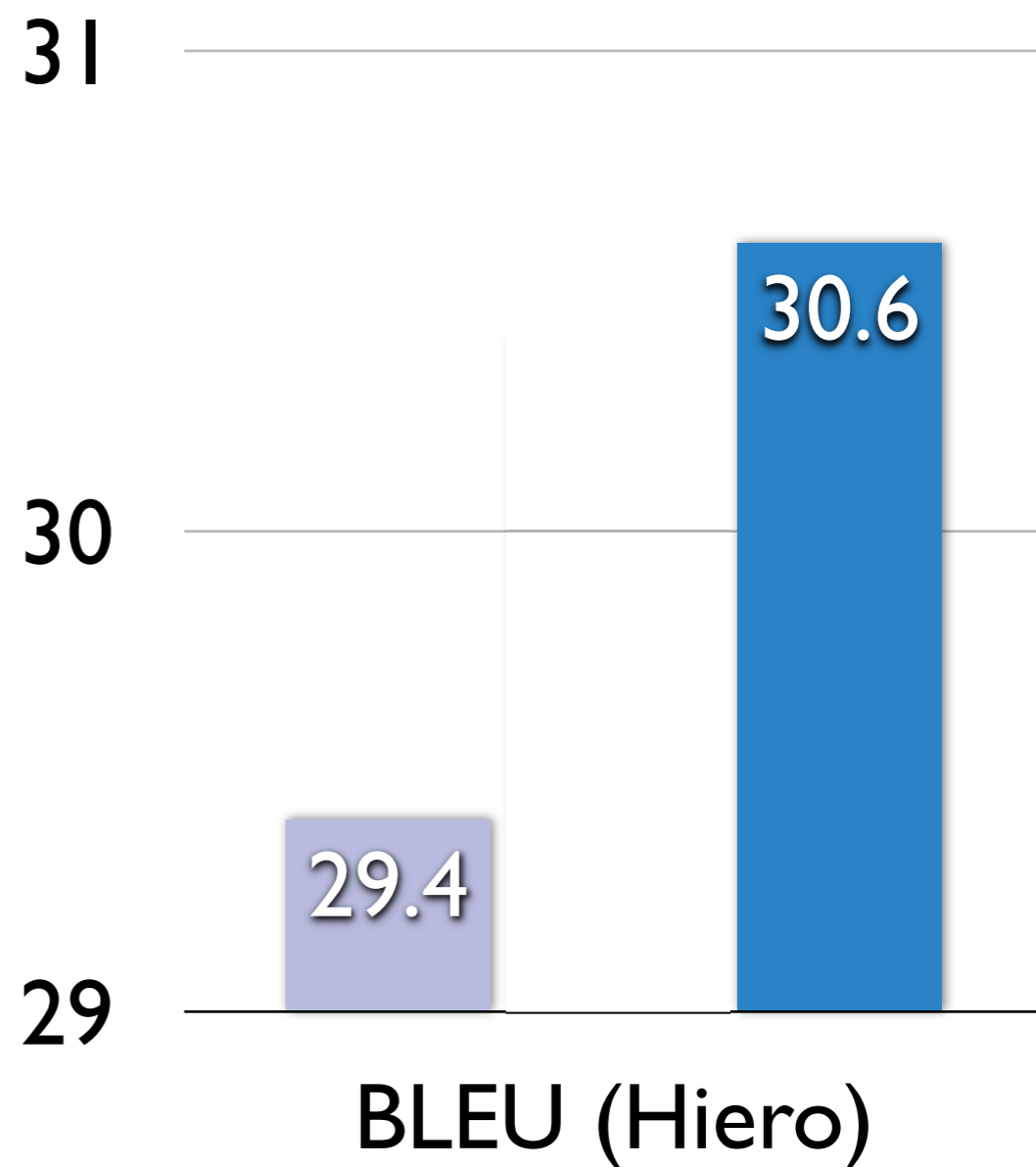
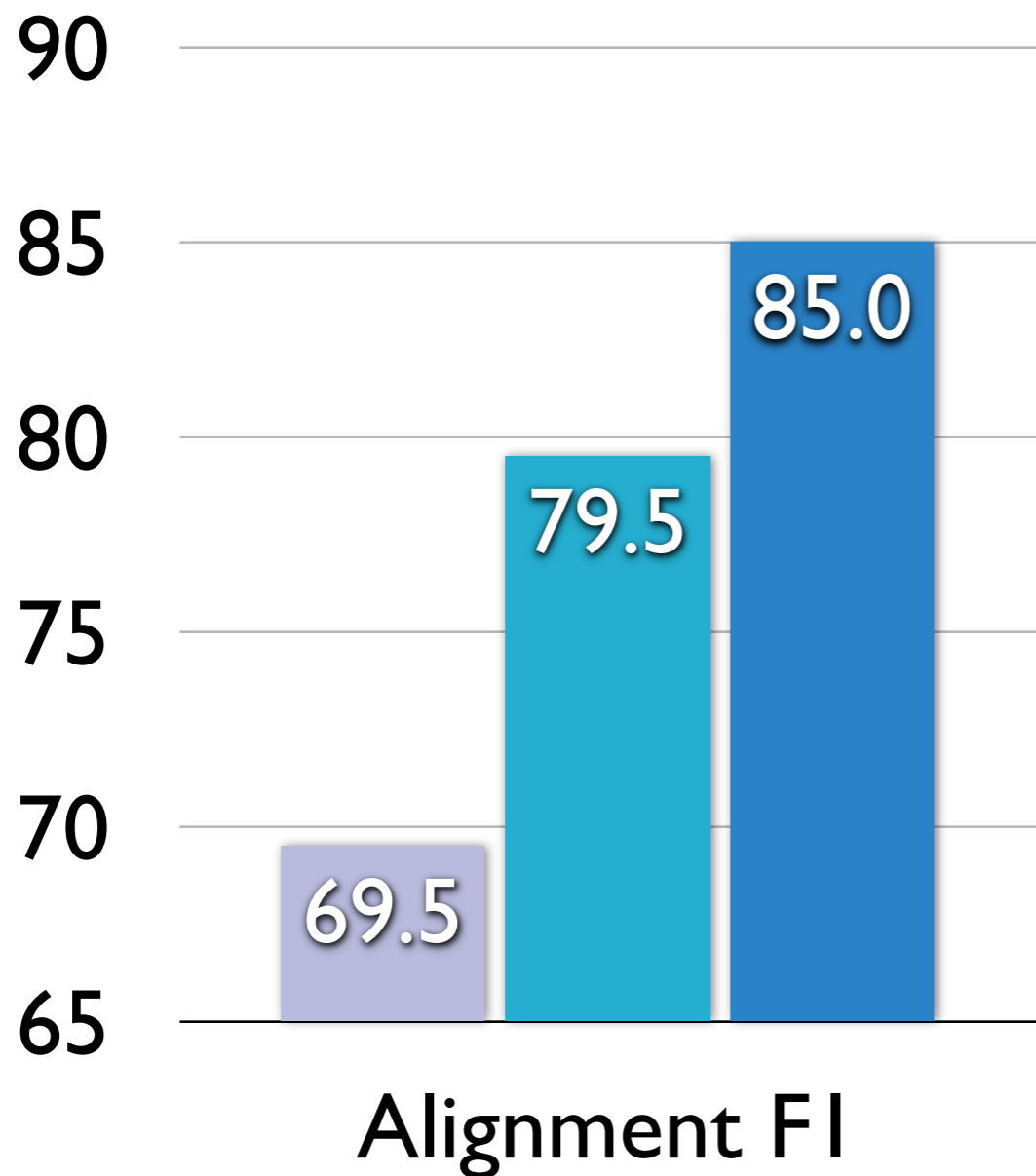
Word Alignment Improvement





Word Alignment Improvement

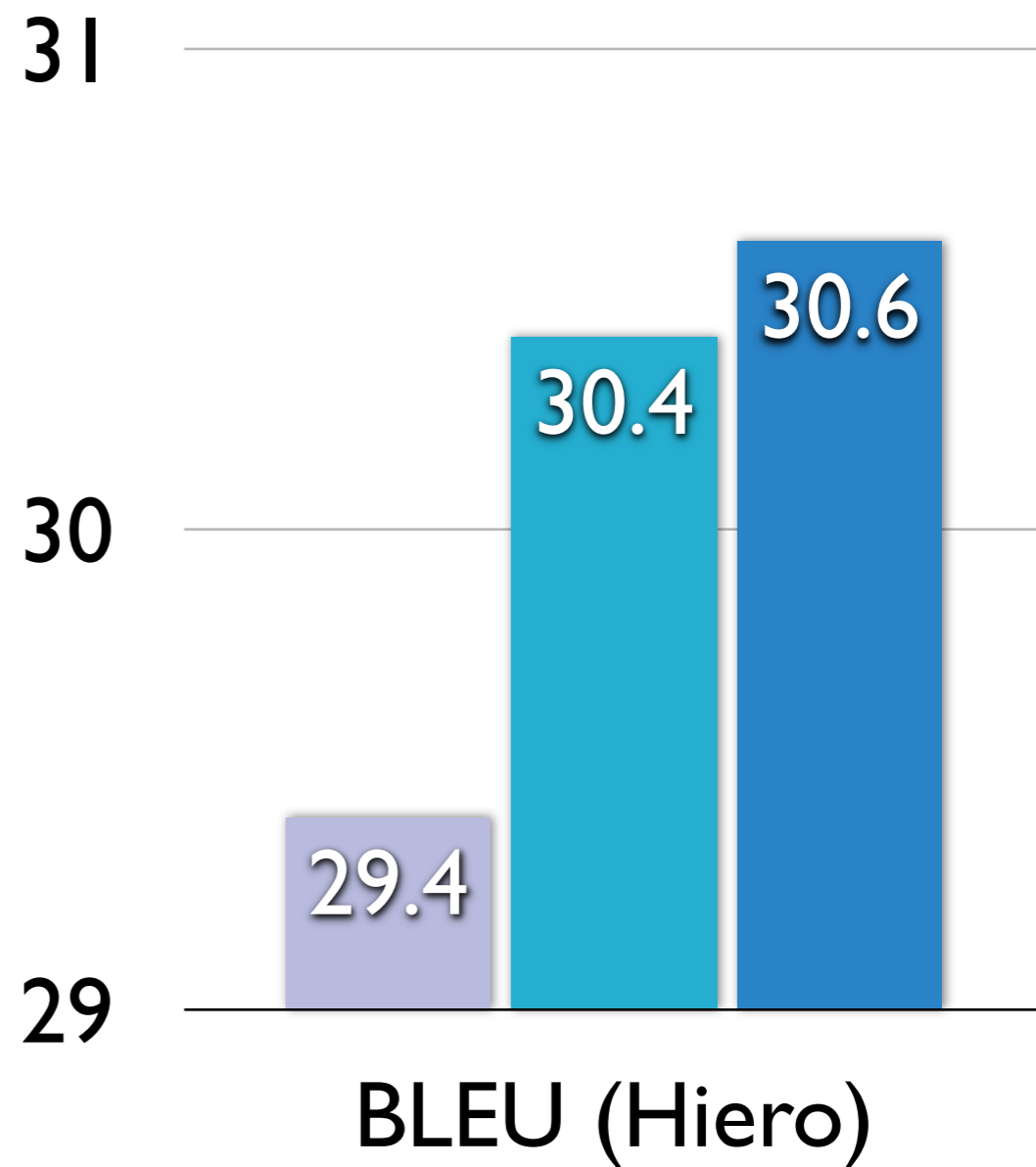
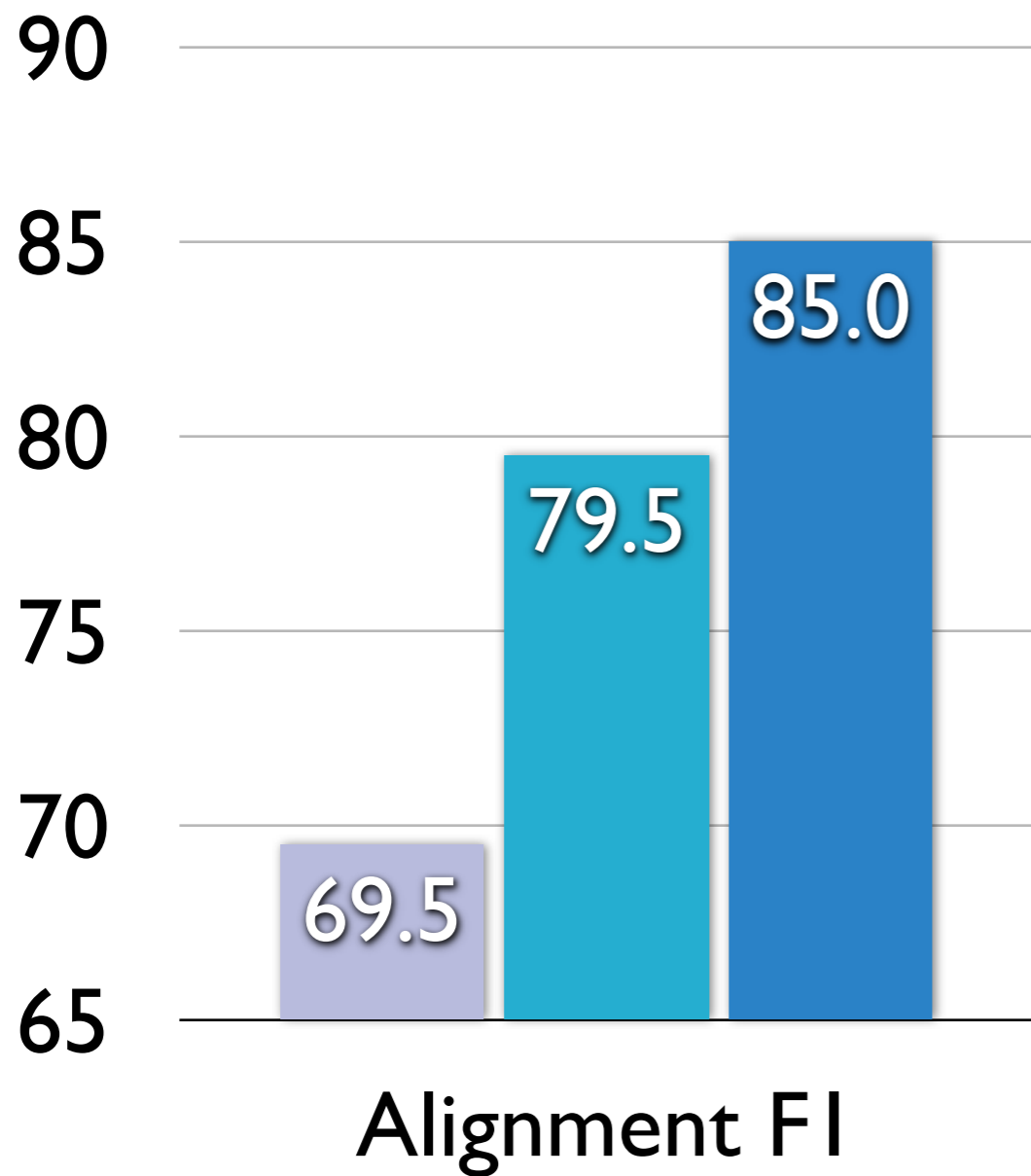
■ HMM ■ Supervised ITG ■ Joint





Word Alignment Improvement

HMM Supervised ITG Joint





Translation Improvement

目前 导致 飞机 相撞 的 原因 尚 不 清楚

Currently cause plane crash DE reason yet not clear

Reference

At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)

The cause of planes is still not clear yet.



Translation Improvement

目前 导致 飞机 相撞 的 原因 尚 不 清楚

Currently cause plane crash DE reason yet not clear

Reference

At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)

The cause of planes is still not clear yet.



Translation Improvement

目前 导致 飞机 相撞 的 原因 尚 不 清楚

Currently cause plane crash DE reason yet not clear

Reference

At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)

The cause of planes is still not clear yet.



Translation Improvement

目前 导致 飞机 相撞 的 原因 尚 不 清楚

Currently cause plane crash DE reason yet not clear

Reference

At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)

The cause of planes is still not clear yet.



Translation Improvement

目前 导致 飞机 相撞 的 原因 尚 不 清楚

Currently cause plane crash DE reason yet not clear

Reference

At this point the cause of the plane collision is still unclear.

Baseline (GIZA++)

The cause of planes is still not clear yet.

Joint Model

The cause of plane collision remained unclear.

Berkeley



Conclusions



Conclusions

- Alignments and syntax are mutually constraining.



Conclusions

- Alignments and syntax are mutually constraining.
- The tasks of parsing and word alignment both gain from joint, weakly synchronized modeling.



Conclusions

- Alignments and syntax are mutually constraining.
- The tasks of parsing and word alignment both gain from joint, weakly synchronized modeling.
- Our bitext parse-and-aligner will be available for download at <http://nlp.cs.berkeley.edu>.



Thank You

